Responses to Reviewer

General comments：

The manuscript does not have a clearly stated scientific hypothesis or a sufficiently focused model evaluation question. Instead, it combines three rather different activities: comparison of meteorological and land cover inputs, one at a time parameter perturbations, and HPC/runtime optimization.

I am not convinced that the experiments support several of the main conclusions. In particular, the model simulations do not reproduce the dominant observed direction and geometry of fire spread but the manuscript repeatedly describes some configurations as having good agreement or improved representation.

The observational basis for evaluation is weak. The "observed perimeter" appears to be a burn scar derived from pre-fire and post-fire satellite imagery, not a time-resolved perimeter. This raises a fundamental question about comparisons made to the purported observed fire extent at 13:00 UTC.

The sensitivity experiments largely demonstrate responses that are already built into the model formulation, and the claimed ranking of "dominant" controls is difficult to interpret because the perturbations of wind speed, RH, and temperature are not normalized or otherwise made comparable.

The computational optimization section seems only loosely connected to the fire behavior science. Some of the I/O and decomposition results may be useful computationally but benchmarking specific Intel compiler versions on one platform does not by itself constitute a substantial advance in modeling science.

Overall, I do not yet see a clear new scientific result. The paper contains potentially useful case-specific observations, but the central evaluation and sensitivity conclusions are substantially stronger than the evidence presented.

Response to General comments:

We thank the reviewer for reading the manuscript in full and for the careful assessment. The study examines one real grassland fire. With the ignition location, terrain, and model held fixed, it asks how the dominant direction of spread, and the overlap with the final burn scar, change when the meteorological forcing is replaced, when two-way fire–atmosphere coupling is switched on or off, and when the fuel distribution derived from land cover is changed, and which meteorological quantities and fuel parameters account for those differences. The experiments are interpreted through relative differences. The computational section is not part of that comparison. It only quantifies the reduction in runtime achieved, for the same high-resolution configuration, by the choice of compiler, parallelization, and I/O. In the Introduction, the three aims previously listed side by side will be restated as this comparison. In the Abstract, the evaluation against the satellite burn scar and the runtime tests will be stated separately. Further detail is given in the responses to Comments 2 and 6.

The images of 26 April and 10 May provide the shape of the burn scar, and the area of

1299 km² was obtained from the post-fire survey by the local forestry and grassland authority. Both describe the final extent after the fire, not an observed perimeter at 13:00 UTC. The time 13:00 UTC is a common comparison time. Ignition was at 07:30 UTC, so 5.5 h had elapsed. It is used to examine how much of the simulated fire then overlaps the final burn scar and to which side the simulated fire is displaced. In the FD experiments, two-way coupling has by then been active for several hours. Two-way coupling is not enabled in the noFD experiments. Figure 5 will be revised in this sense: the black line is the final burn scar, and the colored lines are the simulated perimeters at 13:00 UTC. The three terms burn scar, burned area, and fire perimeter will be used separately. Section 2.1 will state that the imagery supplies the shape and the survey supplies the area. Further detail is given in the response to Comment 3 and in technical points 4 and 5.

The eastward lobe does not appear in any experiment. The final boundary was also limited by firebreaks and fire suppression, neither of which is represented in the model. The simulations therefore indicate the area that the fire can reach under the prescribed meteorological fields, fuels, and terrain. The revision will retain only the relative differences supported by the experiments: the major axis of FNL 6 h (FD) lies closer to the north–south body of the burn scar, whereas the ERA5 simulations are displaced toward the northwest and extend over a large area outside the scar. No configuration will be described as having reproduced the geometry of the observed fire. The difference between FNL and ERA5 will no longer be interpreted as showing that a higher temporal resolution is inherently preferable. The land-cover results will be described separately by overlap area, northward expansion, and non-burnable patches, and the undefined phrase “best overall performance” will not be used. Two-way coupling accelerates the growth of burned area after approximately 12 h. After approximately 17 h, the feedback modifies the local wind and thermal structure, further expansion in FD is limited, and the burned area in noFD can become larger. The Abstract and Conclusions will follow these two stages. Further detail is given in the responses to Comments 5, 7, 12, and 13.

An increase in wind speed increases the rate of spread, fuel load changes flame intensity, and fuel depth changes spread. These directions of response follow from the Rothermel formulation used in SFIRE. The experiments quantify the changes in each metric for this grassland fire under the perturbations in Tables 4 and 5. In Table 4, the perturbations of wind speed, relative humidity, and air temperature are all applied to the input meteorological fields and integrated by WRF. The fuel-moisture model is off in these experiments, and the fuel moisture content is prescribed. The three quantities follow different pathways, and 1 m s^{-1} cannot be converted into an equivalent change in temperature or humidity, so that normalization is not applied. At the amplitudes in Table 4, only wind speed clearly changes the direction of the fire head, so wind speed is still identified as the meteorological factor with the clearest effect on spread direction. Relative humidity produces the larger response in spread rate and flame length, and the overall response to air temperature is the weakest. The original statement that air temperature played a secondary role refers to these magnitudes in Fig. 8. The magnitudes and pathways will be added in Section 3.2, and the existing conclusions

will be retained within the corresponding perturbation ranges. Further detail is given in the responses to Comments 8, 9, and 10.

Numerical studies of grassland wildfire remain few. Existing work relies largely on post-fire remote sensing, or on experimental burns and idealized simulations. High-resolution, two-way coupled simulations of the full course of a real grassland fire are scarce, and there is little against which the present results can be compared. How meteorological forcing, the fuel distribution derived from land cover, and fire–atmosphere feedback together control the direction of spread is still poorly understood in several respects. This study does not claim to resolve those processes with a single case. It places the relative differences that can be compared for one real fire under a common configuration, in order to invite further discussion, and to provide a reference for understanding grassland-fire processes, for risk assessment, and for decisions on prevention and control. The fire perimeter obtained here and the present parameter settings are not put forward directly as an operational forecast. Further detail is given in the response to Comment 15.

Specific comments:

Comment 1: *Opening motivation / grassland fire risk*

The first sentence of the abstract states that "under global warming, grassland wildfire risk has increased substantially." I do not think this claim is necessary to motivate the study and the cited material does not appear to establish that general statement for global grassland wildfire risk. The paper is about evaluation of WRF SFIRE for one Inner Mongolia event; the motivation should be tied directly to that problem.

Response to comment 1:

We thank the reviewer for this comment. We will narrow the opening sentence, but the background will not be removed entirely. The paper is not a self-contained recounting of one case; the background explains why this grassland-fire simulation was undertaken. First, wildfire is a global problem, and the frequency and extremity of fire weather have increased overall during recent decades (Jolly et al., 2015; Jones et al., 2022). Second, existing observations and simulations are concentrated largely on forest fires, and the increase in extreme wildfire events has occurred mainly in boreal and temperate conifer forests (Cunningham et al., 2024). Studies of fires in grassland alone, especially coupled fire–atmosphere simulations, remain limited, and relatively few publications are available to cite. Third, the specific conditions in China are not developed at length in the manuscript because of space. China has a large grassland area. Across the grasslands of northern China, climate anomalies, inbound cross-border fires, and anthropogenic ignitions occur together, and recent years have seen repeated grassland fires that ignited abruptly, spread rapidly, and affected large areas. Examples include the cross-border fire from Mongolia on 21 May 2009, the Daofu grassland fire in Sichuan on 5 December 2010, the major grassland fire in East Ujimqin Banner of Inner Mongolia on 24 May 2011, the border fire in Yumin, Xinjiang, in 2014, the extreme grassland fire in East Ujimqin Banner on 29 March 2016, the Hongyuan grassland fire in Sichuan on 12 February 2020, the Xilin Gol grassland fire in Inner Mongolia in 2023,

and the extreme grassland fire in Chenbarhu Banner, Inner Mongolia, on 1 May 2023, which is simulated in this study. These events indicate the practical need for simulations of this kind.

References:

Jolly, W.M., Cochrane, M.A., Freeborn, P.H., Holden, Z.A., Brown, T.J., Williamson, G.J., Bowman, D.M.J.S.: Climate-induced variations in global wildfire danger from 1979 to 2013. Nat. Commun. 6, 7537, https://doi.org/10.1038/ncomms8537, 2015.

Jones, M.W., Abatzoglou, J.T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., Smith, A.J.P., Burton, C., Betts, R.A., van der Werf, G.R., Sitch, S., Canadell, J.G., Santín, C., Kolden, C., Doerr, S.H., Le Quéré, C.: Global and regional trends and drivers of fire under climate change. Rev. Geophys. 60, e2020RG000726, https://doi.org/10.1029/2020RG000726, 2022.

Cunningham, C.X., Williamson, G.J., Bowman, D.M.J.S.: Increasing frequency and intensity of the most extreme wildfires on Earth. Nat. Ecol. Evol. 8, 1420–1425, https://doi.org/10.1038/s41559-024-02452-2, 2024.

Comment 2: *First sentence of the abstract*

The statement that WRF SFIRE was applied using satellite-derived perimeters "to evaluate the effects of meteorological forcing, land cover, and key driving factors … as well as to optimize model performance" does not follow logically. Satellite observations can be used for model evaluation but they have no direct connection to compiler selection, domain decomposition, or I/O optimization. The abstract combines two substantially different objectives.

Response to comment 2:

We appreciate this comment. The satellite burn scar is used only to evaluate the meteorological forcing, the land-cover-to-fuel mapping, and two-way coupling. Compiler choice, process decomposition, and I/O are assessed by the wall-clock time of the same simulation. The two analyses do not share an evaluation metric.

In the revised manuscript, the Abstract will separate the objectives into two sentences. The first will state that the final burn scar is used to compare burned area, the overlap fraction, and the directional displacement under different inputs and coupling settings. The second will state separately that parallel and I/O configurations were tested on the same hardware, and that PNetCDF combined with asynchronous I/O reduced the runtime by 29.3% relative to the serial NetCDF baseline. That second sentence will no longer say that model performance was optimized by means of the satellite perimeter.

Comment 3: *Nature of the satellite observations*

The manuscript describes an image from 26 April and a post-fire image from 10 May from which the burned area was extracted using BAI and then visually refined. That is a final burn scar not an observed time sequence of fire perimeters. Yet Fig. 5 compares simulations with an "observed burned area perimeter at 13:00 UTC." What evidence establishes that this final burn scar represents the fire extent at 13:00 UTC? If there is an independent perimeter observation at that time, it needs to be described. Otherwise

this comparison is not a valid evaluation of fire progression.

Response to comment 3:

We thank the reviewer for this comment. This was not a prescribed experimental burn, and no densely observed sequence of fire perimeters is available. The usable data are a pre-fire GF-6 image from 26 April and a post-fire GF-1 image from 10 May. Because of satellite overpass times and viewing geometry, clear images that both cover this fire and can be used to delineate the burned extent are scarce. The spatial extent of the final burn scar was therefore extracted with the Burned Area Index (BAI) and refined by visual interpretation. The total area of 1299 km² comes from the field survey conducted by the local forestry and grassland authority after the fire. The imagery provides the shape, and the survey provides the final area. Both are post-fire references, not observations at 13:00 UTC.

Evaluation of a real wildfire often has to proceed in this way. Information on the progression is commonly limited to a final perimeter obtained from post-fire analysis, and suppression is rarely recorded in a form that can be introduced into the model (Kochanski et al., 2013; Hu et al., 2024). For the Xintian fire, Hu et al. (2024) state explicitly that, in the absence of a time series of burned area, only the final perimeter could be used for comparison, and that suppression was omitted from the simulation because no suppression data were available. Liu et al. (2024) likewise compared their simulation of the Xichang fire with the final burned extent.

The time 13:00 UTC is a comparison time shared by all simulations. It is not the time at which the fire reached its final extent. Ignition was at 07:30 UTC, so 5.5 h had elapsed by 13:00 UTC. The adjustment associated with the initial ignition had ended, and two-way coupling had been active for several hours. Suppression was applied after the real fire started, but no suppression data are available for assimilation, and the model cannot represent that process. As shown in Fig. 4, without intervention the burned area in every simulation reaches several times the surveyed final area by 18:00 UTC. The area at the end of the simulation is therefore not treated as a prediction of the final burn scar. The time 13:00 UTC is used to compare the position and shape during the unsuppressed stage of spread. In the revision, the caption of Fig. 5 and the methods will state that the black line is the final burn scar and the colored lines are the simulated perimeters at 13:00 UTC. The comparison shows the spatial overlap between the simulated fire and the final burn scar. It does not represent the actual fire extent at 13:00 UTC.

References:

Kochanski, A.K., Jenkins, M.A., Krueger, S.K., Mandel, J., Beezley, J.D.: Real time simulation of 2007 Santa Ana fires. For. Ecol. Manage. 294, 136–149, https://doi.org/10.1016/j.foreco.2012.12.014, 2013.

Hu, H., Deng, X., Zhang, G., Feng, L., Long, J., Li, Z., Zhu, Y., Wang, Y.: Fire behavior simulation of Xintian forest fire in 2022 using WRF-fire model. Front. For. Glob. Change 7, 1336716, https://doi.org/10.3389/ffgc.2024.1336716, 2024.

Liu, Y., Miao, S., Zhang, H.: Sensitivity simulations of the 30 March 2020 Xichang wildfire in southwest China based on the WRF-Fire model. J. Meteorol. Res. 38, 826–843, https://doi.org/10.1007/s13351-024-3171-5, 2024.

Comment 4: *Recall = 39.1%*

The abstract gives a Recall value of 39.1% without defining what that is. More importantly, Recall is later defined as overlap area divided by observed burned area. This doesn't penalize simulated burned area outside the observed scar. Recall alone is not a sufficient metric of spatial agreement. Calling the 39.1% case the "best agreement in both extent and location" seems stronger than this metric supports. Some measure of false positive area / precision or an intersection-over-union type measure would be more informative.

Response to comment 4:

We agree with the reviewer. In the revised manuscript, the value of 39.1% will be described as the fraction of the final burn scar that is overlapped by the simulation. The methods will define the overlap fraction (the present Recall) as the overlap area divided by the area of the final burn scar, precision as the overlap area divided by the simulated burned area, and the intersection over union as the overlap area divided by the sum of the simulated area and the scar area minus the overlap area. Figure 5c will report all three quantities for every experiment.

Comment 5: *The simulations do not reproduce the observed spread direction*

Fig.5 is, to me, a major problem. The observed burned area has a substantial lateral/easterly extension, whereas the simulations predominantly spread northward/northwestward. Even the best scoring simulation does not reproduce the dominant observed geometry. This should be treated as a model failure to understand, not mainly as a basis for ranking the simulations. In a coupled fire-atmosphere model, getting the fire-driving flow wrong is central to the evaluation. The statement that the FD simulations "more closely reproduce the observed fire geometry and spread pathway" needs to be reconsidered in light of Fig.5.

Response to comment 5:

We thank the reviewer for this comment. The final burn scar is not the boundary of spread in the absence of intervention. Firebreaks and other barriers were used during suppression, and these measures directly limited the outer edge of the fire. The model is integrated forward from the ignition point and includes only the meteorological fields, the fuel fields, the terrain, and fire–atmosphere feedback. It does not include firebreaks, and no suppression data can be introduced into the model. The simulation therefore indicates the extent and the dominant direction of spread that are possible under those conditions. It does not reconstruct the final boundary after human intervention. Without fire perimeters or firebreak locations that can be assimilated, a forward simulation cannot reproduce the dominant geometry shaped by that intervention. Kochanski et al. (2013) and Hu et al. (2024) likewise note that, when suppression is not recorded quantitatively, the final perimeter can serve only as a reference and the model cannot be expected to reproduce it segment by segment.

The meteorological-forcing and coupling experiments compare configurations for the same fire, the same fuels, and the same terrain. They compare the direction of spread

and the overlap with the final burn scar. They are not an absolute forecast verification. In this comparison, the major axis of the fire in FNL 6 h (FD) lies closer to the north–south body of the burn scar. The ERA5 experiments are clearly displaced toward the northwest and fall largely outside the scar. The eastward lobe is absent from all experiments. The revision will retain this relative difference and will not describe any configuration as having reproduced the geometry or the spread pathway of the observed fire.

Comment 6: *What is the actual hypothesis or model evaluation question?*

The introduction states three aims: identify suitable input datasets, identify controlling factors, and improve computational efficiency. These are tasks rather than a clearly stated scientific hypothesis or evaluation question. What, specifically, is being tested? For example, is the hypothesis that one meteorological forcing represents the observed fire-driving flow better? That two-way coupling improves simulation of the observed spread? That a particular land cover representation improves the simulation for a physically identifiable reason? The experiment design should follow from a stated question.

Response to comment 6:

We thank the reviewer for raising this point. The question addressed for this grassland fire is as follows. With the ignition location, terrain, and model held fixed, how do the dominant direction of the simulated fire and its overlap with the final burn scar change when a different meteorological forcing is used, when two-way fire–atmosphere coupling is switched on or off, and when the fuel distribution converted from a different land-cover product is used; and which meteorological quantities and fuel parameters are mainly responsible for those differences? The experiments are compared through relative differences. The computational experiments are not part of this test. They ask only how much the compiler, parallelization, and I/O can shorten the runtime of the same high-resolution configuration. In the Introduction, the three aims previously listed side by side will be drawn together and stated directly as the comparison above. The computational experiments will be described separately as runtime tests of the same simulation and will not be combined with the fire-behavior results into a single conclusion.

Comment 7: *FNL versus ERA5*

The manuscript correctly states that both FNL and ERA5 are used at 0.25 ° spatial resolution; their principal stated resolution difference is temporal, 6 h versus 1 h. However, differences between FNL and ERA5 cannot simply be attributed to temporal resolution. These are different analysis/reanalysis systems with different atmospheric states and data assimilation. The FNL – ERA5 comparison a not a clean temporal resolution experiment. The manuscript states that ERA5 produces larger burned areas and higher spread rates and concludes that higher temporal resolution is "more favorable for representing fire spread." That conclusion does not follow from the

comparison. A larger or faster simulated fire is not necessarily a better simulation.

Response to comment 7:

We agree with the reviewer. The difference between FNL and ERA5 cannot be treated as an experiment on temporal resolution. Both datasets have a spatial resolution of 0.25°, but they come from different analysis and assimilation systems, their atmospheric states differ, and their update intervals also differ. In Fig. 4, the burned area and spread rate driven by ERA5 are indeed larger than those driven by FNL. In Fig. 5, however, the overlap of every ERA5 experiment with the final burn scar is below 20%, and the simulated fires are clearly displaced toward the northwest, with a large fraction lying outside the scar. A larger and faster simulated fire is less consistent with this burn scar. The results therefore cannot be used to conclude that a higher temporal resolution is more favorable for representing fire spread. That sentence will be removed from the manuscript.

In the revised text, FNL at a 6 h interval and ERA5 will be compared only as two driving datasets. The manuscript will report the wind fields they produce, the fire areas, and the differences in position relative to the burn scar. Those differences will not be interpreted as a comparison of a 6 h update with a 1 h update. Update frequency is examined within ERA5: the same 1 h fields are supplied once at a 1 h interval and once at a 3 h interval. At 13:00 UTC, their overlap fractions with the final burn scar are 19.5% and 17.9%, respectively. By 18:00 UTC, the burned area in both experiments greatly exceeds the surveyed area of 1299 km^2. On the present results, the 1 h update does not bring the simulated fire into closer agreement with the final burn scar.

Comment 8: *Meteorological sensitivity experiments*

Wind speed is perturbed by ±2 and ±4 m s^{-1}, RH by ±5 and ±10 percentage points, and temperature by ±2 and ±4°C throughout the entire domain and forcing period. The manuscript then compares the resulting percentage response of the fire metrics and labels the largest response the "dominant" control. But these perturbations are arbitrary and are not normalized to comparable climatological variability, uncertainty, variance, or nondimensional sensitivity. Thus, conclusions such as "wind speed was the dominant meteorological factor" are conditional on the perturbations chosen. They should not be presented as general physical rankings.

Response to comment 8:

We understand the reviewer's concern. Wind speed, relative humidity, and air temperature are different physical quantities. This study comprises only one fire, and the record near the fire line is not long enough to define the variability of each quantity, so the three perturbations are not normalized by climatological variability and then ranked. In Table 4, all three perturbations are applied to the meteorological fields that drive WRF and are then integrated by the model. The fuel-moisture model remains off in these experiments, and the fuel moisture content is prescribed. Only when that model is enabled separately in Table 5 are the air temperature and relative humidity from WRF used to calculate the equilibrium fuel moisture content (Mandel et al., 2014). After the

integration, the Rothermel rate of spread uses the midflame wind speed interpolated from the near-surface wind, the slope, and the prescribed fuel moisture content. Atmospheric temperature and relative humidity do not themselves enter that equation (Mandel et al., 2011). A wind-speed perturbation changes the input wind field, and the direction of the fire head changes after the integration. Perturbations in relative humidity and air temperature change the moisture and thermal state, and they affect spread only after WRF has transformed them into the near-surface wind and stratification. Because the pathways differ, 1 m s^{-1}, a change of several humidity percentage points, and a change of several degrees Celsius cannot be converted to a common unit.

At the perturbation amplitudes in Table 4, wind speed is still identified as the meteorological factor with the clearest effect on spread direction. When the wind speed is increased, the fire head turns toward the northeast; when it is decreased, the major axis shifts toward the northwest and the fire expands westward. Under the humidity and temperature perturbations, the fire perimeter remains comparatively compact. Spread direction determines the side toward which the head fire moves and the side on which a firebreak is placed. Relative humidity has the larger effect on spread rate and flame length, 85.4% and 54.5%, and these responses are reported separately in Fig. 8. They are the magnitudes obtained in this experiment; they do not mean that the humidity perturbation in Table 4 changed the fuel moisture content. The overall response of burned area, spread rate, spread distance, and flame length to air temperature is the weakest. The original statement that air temperature played a secondary role refers to this comparison of magnitudes in Fig. 8. The revision will state that the judgment applies only to the perturbation amplitudes in Table 4.

Comment 9: *Air temperature sensitivity*

I do not understand the interpretation that air temperature "played a secondary role." What physical pathway is being tested here? SFIRE uses the Rothermel spread formulation, and the manuscript separately enables/disables a fuel moisture model. The authors need to explain how perturbing atmospheric temperature affects fire spread in this configuration and whether that response occurs through fuel moisture, atmospheric dynamics, stability, surface fluxes, or some other pathway. Without that explanation, comparing the magnitude of a temperature perturbation response with wind or RH does not establish the relative physical importance of temperature.

Response to comment 9:

We thank the reviewer for asking how the air-temperature perturbation affects fire spread. In Table 4, the temperature perturbation is applied to the meteorological fields that drive WRF. The Rothermel rate of spread is determined by the midflame wind speed, the slope, and the fuel moisture content. Atmospheric temperature does not itself enter that equation (Mandel et al., 2011). The offset temperature field is integrated by WRF. Advection, boundary-layer mixing, and surface fluxes redistribute the perturbation, and the near-surface wind, stability, and surface heat flux change accordingly. By the time the perturbation reaches the fire, it is no longer a fixed

temperature increment. The sensible and latent heat released by combustion are calculated from the mass of fuel consumed and the fuel moisture content. Ambient air temperature does not appear in those expressions. After the fluxes are returned to WRF, the environmental stratification and the background wind determine the convergence that they induce and the resulting change in the wind along the fire line. The pathway from air temperature to fire spread is therefore as follows: WRF first transforms the input temperature field into the near-surface wind and stratification, which then enter the rate of spread and the fire-atmosphere feedback. The same temperature offset does not correspond to a fixed change in the spread rate.

The fuel-moisture model is not part of this pathway. Only when that optional module is enabled are the air temperature and relative humidity from WRF used to calculate the equilibrium fuel moisture content (Mandel et al., 2014). In this manuscript it is enabled only in the fuel experiment in Table 5. Both the temperature and relative-humidity experiments in Table 4 use a prescribed fuel moisture content.

The original statement that air temperature played a secondary role is based on the magnitudes in Fig. 8, not on the pathway itself. Under the offsets in Table 4, the overall response of burned area, spread rate, spread distance, and flame length to air temperature is weaker than the responses to wind speed and relative humidity. The explanation above is the pathway corresponding to that result. Both wind speed and air temperature are modified in the input meteorological fields, and both are integrated by WRF. After the integration, it is the near-surface wind that enters the spread equation; the temperature perturbation is first transformed by WRF into changes in the wind field and the stratification. The revision will add this pathway and will state that the judgment applies only to the offset amplitudes in Table 4.

References:

Mandel, J., Beezley, J.D., Kochanski, A.K.: Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. Geosci. Model Dev. 4, 591–610, https://doi.org/10.5194/gmd-4-591-2011, 2011.

Mandel, J., Amram, S., Beezley, J.D., Kelman, G., Kochanski, A.K., Kondratenko, V.Y., Lynn, B.H., Regev, B., Vejmelka, M.: Recent advances and applications of WRF-SFIRE. Nat. Hazards Earth Syst. Sci. 14, 2829–2845, https://doi.org/10.5194/nhess-14-2829-2014, 2014.

Comment 10: *Are some of the sensitivity results simply consequences of the equations?*

Several reported findings appear largely to reflect the model formulation itself: increasing wind increases spread; modifying fuel load changes intensity; modifying fuel depth changes behavior according to the Rothermel fuel representation. The manuscript needs to distinguish between a genuinely emergent coupled-model result and a response already prescribed by the fire spread equations. Otherwise, it is difficult to determine what was learned from the experiments.

Response to comment 10:

The reviewer notes that an increase in wind speed increases spread, that fuel load changes flame intensity, and that fuel depth changes fire behavior according to the Rothermel equation. These directions of response come from the spread formulation

adopted in SFIRE. We agree, and the manuscript does not present these directions as newly identified mechanisms. The equation specifies the direction of the response. It does not specify how much each metric changes for this fire under the perturbations in Tables 4 and 5, nor which factor dominates which metric. Section 3.2 and the Conclusions report the latter. On the FNL 6 h (FD) and FROM_GLC30 baseline, the wind-speed offsets clearly change the direction of the fire head and affect burned area and spread distance by 30.9% and 13.8%, respectively. Relative humidity affects the spread rate and flame length by 85.4% and 54.5%, respectively. The overall response to air temperature is the weakest. Fuel load mainly changes flame length, with a range of 81.8%. Fuel depth has the strongest effect on burned area, 94.6%; when the depth is reduced to 0.25 times the baseline, burned area decreases by 62.5%. When the fuel-moisture model is enabled, the spread rate decreases by 20.0%, whereas burned area decreases by only 4.7%. These differences indicate which class of input should be constrained first in simulation and in fire control, and they cannot be read directly from the spread equation.

At high fuel load and large fuel depth, the increases in some metrics weaken. This is consistent with the dependence, in the equation, of the response on departure of the packing ratio from its optimum. Because the fuel perturbations were performed only with two-way coupling enabled, the present results are the net response of the coupled simulation. The additional contribution of coupling is not isolated. One sentence will be added in Section 3.2 to state that the directions of the responses follow the spread formulation used here, and that the conclusions of this study are the magnitudes and the differences among metrics given above. The original statements on which fire-behavior metrics are controlled by wind speed, relative humidity, fuel load, and fuel depth will be retained and restricted to the perturbation ranges in Tables 4 and 5.

Comment 11: *Land cover comparison is confounded with fuel model mapping*

The three land cover products are not merely alternative spatial maps. They are converted to different numbers and distributions of Anderson fuel classes: two produce four mapped fuel classes, whereas GLC_FCS30D produces nine. Consequently, differences between the simulations are not solely "effects of the land cover dataset." They also reflect the chosen land cover to fuel mapping. This distinction is especially important because the paper draws conclusions about which land cover dataset performs best.

Response to comment 11:

The reviewer notes that the three land-cover products were converted into different numbers of Anderson fuel classes, four for FROM_GLC30 and GlobeLand30 and nine for GLC_FCS30D, so the differences among the simulations cannot be attributed entirely to the land-cover datasets themselves. This distinction is correct and is a limitation of the present study. SFIRE uses Anderson's 13 fuel models as its fuel library. Those parameters were developed for North American vegetation and are not consistent with the grass height, fuel load, and moisture characteristics of Chinese grasslands. No localized fuel-parameter library that can be called directly by WRF-SFIRE is yet

available for China. In the absence of such a library, the only practicable approach was to use the class descriptions of the three land-cover products and map them to Anderson's 13 classes according to their fuel characteristics. Grassland in a product, for example, was assigned to short grass or tall grass according to the grass height and fuel characteristics implied by its class description. Establishing a Chinese grassland fuel standard that is independent of Anderson's 13 classes and calibrated against field measurements requires a dedicated fuel-database effort. It cannot be completed within this case simulation.

The mapping is comparatively unambiguous for non-burnable classes. Water, wetland, marsh, built-up land, and other classes without fuel correspond to the non-burnable class in the Anderson classification and do not require a compromise on grass height or fuel load. Section 3.1 and Figs. 6 and 7 show that the later differences in the fire arise mainly from the location and connectivity of these non-burnable patches. FROM_GLC30 does not identify the border wetlands and marshes as non-burnable, and the fire expands northward. The non-burnable patches in GlobeLand30 are more fragmented, and the fire head is deflected repeatedly. GLC_FCS30D retains fine non-burnable patches near the fire head, and spread is limited. The revision will state that the three results reflect the land-cover products together with their fuel crosswalks. The crosswalk for non-burnable classes is unambiguous, whereas the crosswalk for burnable classes is limited by the differences between Anderson's 13 classes and the fuel parameters of Chinese grasslands.

Comment 12: *Inconsistent characterization of the best land cover product*

The manuscript says FROM_GLC30 gives better overall spatial agreement with the observed burn scar, while GLC_FCS30D better identifies fine-scale nonburnable patches. The abstract nevertheless states that GLC_FCS30D achieved the "best overall performance." What metric defines "overall performance"? That needs to be stated consistently.

Response to comment 12:

We thank the reviewer for identifying this inconsistency. The contradiction between the Abstract and the main text was our error. In the revision, the Abstract, Section 3.1, and the Conclusions will use the same description, and each statement will be tied to a metric. FROM_GLC30 has the largest overlap area, but the fire tends to cross the wetlands and marshes on the northern side. The fire in GlobeLand30 is more fragmented and its later burned area is smaller. GLC_FCS30D retains non-burnable patches near the fire head, and the spread rate decreases markedly between 12 and 13 h of the simulation. The value of 37.1% shown for GlobeLand30 in Fig. 6 will be corrected to 27.1%, from 352.1/1299.

Comment 13: *Fire-atmosphere feedback*

The manuscript concludes that fire-atmosphere feedback significantly enhances spread and burned area. But in the time series the FD and noFD cases cross, and later the

noFD burned area can exceed the FD area. The physical interpretation therefore seems more complicated than simply "feedback enhances spread." The paper should distinguish time-dependent and process-dependent effects rather than collapsing them into a single statement.

Response to comment 13:

We thank the reviewer for this comment. At each time step, WRF-SFIRE returns the sensible and latent heat released by combustion to the atmosphere and thereby modifies the local wind and thermal structure (Section 2.2; Mandel et al., 2011). In Fig. 4, after approximately 12 h the burned area in FD grows more rapidly, whereas the increase in noFD is more gradual. This stage corresponds to the enhancement of spread by the feedback. After approximately 17 h, the burned area in FD becomes smaller than that in noFD. The explanation in Section 3.1 is that the continuing feedback in FD modifies the local wind and thermal structure and limits further outward expansion of the fire. The noFD simulation does not include this feedback, so the background meteorological forcing continues to act and its burned area can keep increasing and exceed that of FD. The later crossing therefore reflects the action of the same coupling at a different stage. It does not remove the earlier enhancement.

The Abstract and Conclusions summarize the whole period by stating that the feedback significantly enhanced the spread rate and burned area. They do not separate the acceleration after approximately 12 h from the crossing of the area curves after 17 h. We accept this point. The revision will state that the feedback enhances spread and burned area during the earlier stage, whereas after 17 h further expansion in FD is limited because the local wind and thermal structure have been modified, and the burned area in noFD can become larger under the background forcing. The explanation already given in Section 3.1 will be retained.

Comment 14: *Computational optimization*

I understand what is being done: compiler comparisons, MPI versus MPI+OpenMP configurations, grid decomposition, and NetCDF/PNetCDF/asynchronous I/O tests. The question is what scientific or generally transferable result comes from these experiments. The compiler differences are only a few percent and are specific to the Intel versions and hardware tested. The conclusion that compiler versions should be benchmarked on the target architecture is unsurprising.

The PNetCDF/asynchronous-I/O and decomposition results may be more useful. If these are intended as a contribution to GMD, the manuscript should make clearer what is specific to WRF SFIRE, why the behavior occurs, and how transferable the recommended configurations are to other domains, processor counts, and architectures.

Response to comment 14:

We thank the reviewer for raising the question of what these computational experiments provide and how far the results can be transferred. The differences among compilers are small. On the Xeon Gold 6148 platform used here, the one-hour runtime is 17 min 39 s with v18, and v21 and v22 are slower by only 2.4% and 1.9%. This result indicates

only that a short benchmark is advisable on this hardware. It is not a principal conclusion, and it will be shortened in the revised manuscript. The more substantial results concern I/O and process decomposition. In addition to the atmospheric fields, WRF-SFIRE must write the high-resolution fire grid. In these tests, serial NetCDF requires 17 min 39 s and PNetCDF requires 15 min 40 s. Asynchronous I/O performs better when its processes account for about 7.5%–12.5% of the total. Combined with PNetCDF, the runtime is 12 min 20 s, which is 29.3% less than the serial baseline. Too few I/O processes leave the computation waiting on disk writes, whereas too many take ranks away from the computation, so the appropriate share depends on the output frequency. For the grid decomposition with 120 processes, the configuration 6×20, with more processes in the north–south direction, requires 22 min 02 s, whereas 20×6, with more processes in the east–west direction, requires 29 min 20 s. This differs from the WRF documentation, which recommends keeping the subdomains close to square. It is consistent with the distribution of computation in the present fire, which spreads mainly in the north–south direction.

These experiments were carried out because the same fire case ran very slowly with the default compiler and I/O settings, which made repeated high-resolution two-way simulations difficult. Compiler options, parallel mode, process decomposition, and I/O were therefore compared in order to identify settings that reduce the runtime on this platform. The same I/O and process configuration was subsequently used by our group in fire-smoke simulations with WRF-SFIRE coupled to WRF-Chem. Those simulations write a larger volume of output, and the same changes reduced the runtime more noticeably. The one-hour results are reported to show that, for coupled simulations of this kind, it is worth examining PNetCDF, the fraction of processes assigned to asynchronous I/O, and the direction of the decomposition, rather than retaining the default settings throughout. The specific fractions depend on the hardware, the core count, and the output frequency. The manuscript reports the results for the present cluster. The purpose of the experiments and the range over which they apply will be added in the revision.

Comment 15: *Claim of operational relevance*

The manuscript repeatedly connects these results to early warning, emergency response, and operational simulation. That claim seems premature when the simulated fire does not reproduce a major feature of the observed spread direction and the observational evaluation itself is uncertain. Computational speed is only useful operationally if the forecast has demonstrated physical and predictive skill.

Response to comment 15:

We thank the reviewer for drawing attention to the statements on early warning and emergency response. Existing studies in China rely largely on remote sensing for post-fire detection, or they concentrate on experimental burns and idealized settings. High-resolution simulation of the full course of a real grassland fire, and discussion of fire behavior itself, remain limited. This paper brings together, for one real grassland fire, the comparison of meteorological forcing, fuel distribution, two-way coupling, and

computational efficiency. Its purpose is to invite further discussion of this problem and to provide a reference for understanding grassland-fire processes in China and for risk assessment and decisions on prevention and control.

The simulated perimeter and the present parameter settings are not advanced directly as an operational forecast. The closer alignment of FNL 6 h (FD) with the north–south body of the burn scar, the effect of non-burnable patches on the later extent, the acceleration of area growth by two-way coupling during the earlier stage, and the reduction in runtime from the I/O and process settings are configuration experience obtained for this case. The final burn scar was also affected by suppression and firebreaks, and a single simulation cannot replace operational verification. In the revision, the phrase "provide more reliable support for fire-risk warning and management" will be changed to state that the results are used to discuss high-resolution simulation of a real grassland fire and to provide a reference for process understanding, risk assessment, and decisions on prevention and control.

Technical corrections / presentation:

These are mostly presentation issues rather than scientific ones:

1. Define Recall at first use in the abstract or preferably replace it there with a more readily interpretable description.

2. Remove most experiment/configuration names from the abstract. Names such as FNL 6 h (FD), FROM_GLC30, and GLC_FCS30D obscure the scientific message.

3. "with the computational optimization" in the title is awkward English and should be revised.

4. Use consistent terminology for burn scar, burned area extent, and fire perimeter. They are not interchangeable.

5. Clarify exactly what is meant by an "observed perimeter at 13:00 UTC."

6. The abstract and conclusions should avoid "best overall performance" unless a specific aggregate performance criterion is defined.

Response to Technical Corrections:

We thank the reviewer for identifying these points of presentation individually. All six are handled as follows.

1. Recall will not be used alone in the Abstract. The value of 39.1% will be replaced by the statement that the area where the simulated fire overlaps the final burn scar is 39.1% of the scar area. The definition of Recall will be given in Fig. 5 and at the first occurrence of the formula in the text, where it will be stated as the overlap area divided by the area of the final burn scar.

2. Experiment names such as FNL 6 h (FD), FROM_GLC30, and GLC_FCS30D will be removed from the Abstract.

3. The phrase "with the computational optimization" will be removed from the title. The title will be changed to "Meteorological and Land-Cover Controls on Grassland Fire Behaviour by WRF v4.4-SFIRE v0.1 and Its Computational Optimization".

4. Three terms will be used separately throughout the manuscript. The post-fire extent

obtained from the satellite imagery and the survey will be called the final burn scar. Its area, or the area obtained from a simulation, will be called the burned area. The outer boundary of the simulated fire at a given time will be called the fire perimeter. The final burn scar will no longer be described as the observed perimeter at that time.

5. The caption of Fig. 5 will be changed to state that the black line is the final burn scar and the colored lines are the simulated fire perimeters at 13:00 UTC. The caption will note that no independent observed perimeter is available at that time, and that the comparison shows the spatial overlap between the simulated fire and the final burn scar.

6. The phrase “best overall performance” will be removed from the Abstract and the Conclusions. The land-cover results will be reported only by individual measures: FROM_GLC30 has a larger area of overlap with the final burn scar, but the fire tends to expand beyond the scar on the northern side; GLC_FCS30D identifies fine non-burnable patches more effectively and limits expansion near the fire head.