

Response to Referee # 2

Authors are grateful to the reviewer for constructive comments and suggestions. In authors' response, we have responded point-by-point to each comment (reviewer's comments are in *blue* and authors' responses are in *black*), and have included the revisions in the text with track-change.

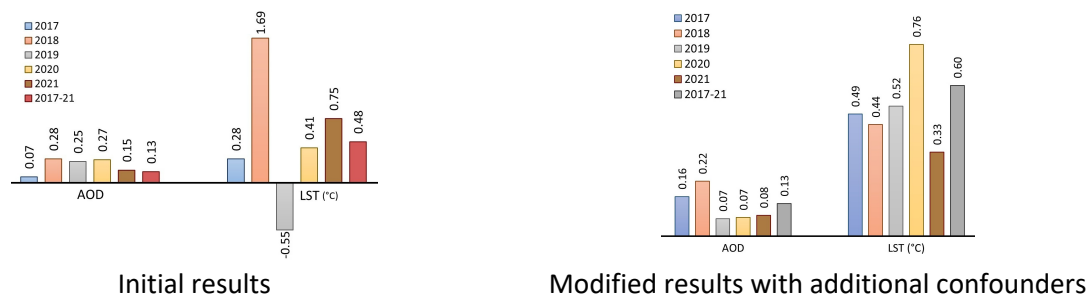
Pandey et al.'s study "Spatial influence of agriculture residue burning and aerosols on land surface temperature" presents an observation-driven study of how crop-residue fires in northwest India influence land surface temperature (LST) and aerosol loading. They identify year-specific intensive fire zones using VIIRS FRP, retrieve VIIRS AOD and MODIS Aqua daytime LST, and use AgERA5 meteorology to control for meteorological context. They apply a space-for-time grid comparison to estimate Δ LST and Δ AOD associated with fire, compute Hurst exponents for persistence, and develop two Random-Forest (RF) regression scenarios (broad fire season and high-correlation windows) to quantify relative feature importance. Finally, they run a Geographically Weighted Regression (GWR) of FRP and LST to map spatial heterogeneity. The paper reports an average fire-induced Δ LST $\approx +0.48^\circ\text{C}$ (range -0.55 to 1.69°C) and Δ AOD $\approx +0.19 \text{ yr}^{-1}$ during the peak season, and finds FRP is the top RF predictor of LST in both scenarios (with much higher RF performance in the "scenario 2" windows). Crop-residue burning in NW India and other parts of South Asia has major air-quality and climate implications, this study's focus on crop-burning and LST is important for this region. The use of VIIRS FRP, VIIRS AOD, MODIS LST, MODIS LC data and AgERA5 meteorology enables a multi-angle observational assessment. The space-for-time comparison, Hurst analysis, random forest for non-linear attribution, and GWR for spatial heterogeneity form a coherent methodological ensemble. However, there are some major concern and queries that needs to be properly addressed at this stage:

Thank you for the valuable comments and suggestions. All recommended changes have been incorporated, including improvements to the space-for-time methodology, Random Forest analysis, and GWR. Additional parameters including meteorological covariates (PBLH, AT, SR, RH, and PR), physical environment (elevation), vegetation and soil characteristics (NDVI, soil moisture), climatological mean LST and AOD, and surface properties (albedo), were included in the selection of "fire" and "no-fire" grids to strengthen the causal attribution of fire to Δ AOD and Δ LST. These parameters were also included in Random Forest and GWR. A nonparametric bootstrapping is performed to estimate uncertainty in Δ AOD and Δ LST. A new supplementary table (Table S1) summarizing all datasets and a workflow diagram (Figure S1) have been added. The interpretation of results has also been revised to improve coherence and scientific clarity.

1. LST is strongly influenced by near-surface air temperature, PBL height, soil moisture, recent precipitation, cloud cover, surface albedo and vegetation state (NDVI/LAI). Although AgERA5 meteorology

(At, Sr, Pr, RH) is included as one of the predictors, the manuscript does not convincingly demonstrate that the estimated ΔLST (and RF / GWR results) are not driven by meteorological covariates or systematic land-cover differences between “fire” and “no-fire” grids. Without stronger control for these confounders, the causal attribution “fire to AOD and LST” remains tentative. In the space-for-time comparison, conducting matched comparisons, for instance for each fire grid choose one or more no-fire grids matched by NDVI, elevation, distance to major urban areas, and climatological mean LST. This reduces bias from non-random spatial placement of fires. Propensity-score matching or simple stratified matching would help. Additional proxies including but not limited to PBL height, surface soil moisture, and in-situ atmospheric radiative impacts induced by the fire-emitted aerosols themselves used in the predictor set may help clarify this relationship and strengthen the findings. However, I welcome the authors to instead post a rationale on why not including these variables and this suggested approach may still suffice in relationship quantification.

Thank you for this suggestion. Accordingly, authors have considered additional parameters, including meteorological covariates (PBLH, AT, SR, RH and PR), physical environment (elevation), vegetation and soil characteristics (NDVI, soil moisture), climatological mean LST and AOD, and surface property (albedo), into the selection of “fire” and “no-fire” grids to strengthen the causal attribution of fire in ΔAOD and ΔLST . As suggested, we also applied a stratified matching technique using combinations of major confounders and conducted comparisons within strata to estimate the difference in LST and AOD between fire and no-fire grids. This refinement substantially strengthened the robustness of our estimates, revealing a consistent and statistically meaningful increase in both LST and AOD in every year due to recurrent fire.



A detailed and modified space-for-time approach is now included in section 2.6. Briefly, it states:

‘....To ensure that changes in LST and AOD were attributable solely to fire activity, grids with similar characteristics in terms of topography, climate, and physical environment were compared (Liu et al., 2019). To achieve this, daily datasets including meteorological covariates (PBLH, AT, SR, RH and PR), physical environment (elevation), vegetation and soil characteristics (NDVI, soil moisture), climatological mean LST and AOD, and surface property (albedo) were extracted over both fire and no-fire grids at a spatial resolution of 10 × 10 km². Fire and no-fire grids with comparable spatial characteristics were grouped into a single stratum, and a stratified matching technique was applied to generate multiple strata based on combinations of the selected confounders. Grids were retained only when differences in their physical environment, vegetation and soil characteristics, climate and land cover between fire and no-fire

conditions were smaller than the defined thresholds (Δ elevation < 50 m; Δ NDVI <0.05; Δ soil moisture <0.05; Δ albedo <0.05; Δ LST <10.0; Δ AOD <0.80). Comparisons were then made within strata containing grids of similar attributes to ensure that the observed variations in LST and AOD could be attributed solely to fire activity’.

Authors would also like to emphasize that the entire residue-burning zone in northwestern India follows similar agronomic practices, with comparable land characteristics, vegetation dynamics, and climatic conditions, as it lies within a single composite climatic zone. Consequently, only subtle variations in meteorological covariates (At, SR, RH, PT) and PBLH (SD: ± 10 m to ± 33 m, yearly) were observed across the grids. All selected grids were representative of croplands within the extended geographical region; therefore, distance from urban centers was not incorporated as an additional constraint. This choice is justified by the fact that agricultural emissions overwhelmingly dominate over anthropogenic urban sources in the post-monsoon season when major residue burning occurs. Columnar aerosol loading was included in the analysis; however, fire-emitted aerosols were not considered separately, as segregating fire-derived aerosols from background loading could introduce additional uncertainty. Authors have included a rationale on variable selection criteria in section 2.6:

‘It is noteworthy that the grids were not classified based on meteorological covariates, as only insignificant variations were noted among the grids. The entire northwestern cropland experiences a relatively uniform background climate during October–November, including comparable boundary layer heights, with PBLH standard deviations ranging from ± 10 m to ± 33 m within a single fire season. The climatological mean LST and AOD were computed only for the pre-fire season, during which none of the grids experienced residue-burning activity. Furthermore, grids were not differentiated by slope or aspect, given the minimal topographic variation across the Gangetic Plain.’

2. Provide details on RF hyperparameter tuning (max_depth, max_features, min_samples_leaf). The manuscript uses n_estimators=100 with a fixed seed — please show whether you tuned parameters (grid search / CV) or at least show sensitivity to n_trees and max_features. To further improve RF model validation, spatial and temporal block cross-validation (e.g., leave-one-year-out, or K-fold blocking by contiguous spatial clusters) and report cross-validated R2, RMSE, MAE. This may provide more robust predictive skill.

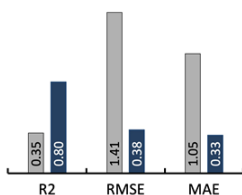
Thank you for this note and guidance. This has indeed improved model performance and creates a statistically rigorous and computationally efficient modelling outcome.

In the revised manuscript, authors have incorporated additional predictors, including FRP, AOD, regional meteorology, surface properties, and vegetation characteristics, into the Random Forest (RF) model to establish a non-linear statistical association between LST and multiple predictors. Accordingly, Section 2.8 (in methods) and Section 3.5 (in results and discussions) have been updated and expanded.

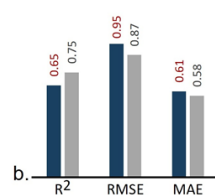
Key RF hyperparameters (`n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf`, and `max_features`) were optimized using Bayesian optimization implemented via `BayesSearchCV` in `scikit-optimize`. Bayesian optimization provides an adaptive and probabilistic search strategy that efficiently explores high-dimensional hyperparameter spaces, outperforming traditional grid and random search approaches in both accuracy and computational efficiency (Snoek et al., 2012; Shahriari et al., 2016; Frazier, 2018).

To ensure robust model evaluation and minimize temporal dependence, authors adopted temporal block cross-validation using a 3-fold `GroupKFold` strategy in `scikit-learn`, in which all samples from the same year were assigned to the same fold, following the blocking principles recommended by Roberts et al. (2017) and Valavi et al. (2019) for temporally structured datasets. This approach provides temporally independent estimates of predictive skill and mitigates information leakage across folds.

Following reviewer’s recommendations, cross-validated R^2 , RMSE, and MAE is now reported in manuscript (Fig. 8), averaged across folds to provide an unbiased estimate of predictive accuracy. This combined framework, Bayesian hyperparameter optimization and temporally independent cross-validation, offers a statistically rigorous and computationally efficient modelling strategy. Details of the RF hyperparameter tuning procedure are included in Table S6.



Initial results



Modified results with RF hyperparameter tuning

3. The GWR model for scenario 2 is using only FRP, SR and AOD as predictor for LST, I do not understand the rationale of leaving out other local factors, included but limited to those mentioned in point 1 above. Are the authors testing the concept of using these specific variables exclusively in relationship to LST? However, I am confused if other meteorological variables and aerosol types (their optical variability in terms of scattering and absorption, and how these may influence atmospheric heating/radiative forcing and near-surface based cooling/radiative forcing (Freychet et al 2019; Tiwari et al. 2023) and surface albedo (Hou et al. 2025) when running GWR could bias the local coefficients. Local coefficients maybe absorb the effect of omitted spatially-varying covariates. I am confused why scenario 2 is missing out important variables. Additionally, please also include bandwidth and kernel details of the AICc minimization you mention.

Thank you for this suggestion. In the revised manuscript, all predictor variables used in the Random Forest model (AOD, PBLH, AT, RH, SR, PT, NDVI, elevation, albedo, and FRP) were also incorporated into the GWR framework. However, aerosol types were not included. Aerosols over South Asia exhibit substantial compositional diversity and are influenced by multiple mixed sources, and classifying aerosol type using AOD–fine-mode fraction–SSA combinations can introduce considerable uncertainty. Moreover, based on

our earlier trials using satellite datasets, carbonaceous smoke aerosols were the only type that could be identified with reasonable confidence over the northwestern region during the biomass-burning period. Including aerosol type as a predictor would therefore risk adding noise and misleading spatial patterns.

Accordingly, the local coefficient maps have been updated, the revised FRP–LST and AOD–LST GWR outputs for both scenarios are now presented in Fig. 9, while coefficient maps for the remaining predictors are provided in Fig. S7. The GWR model exhibited strong explanatory performance, with global R^2 values exceeding 0.74, indicating that the selected predictors effectively captured spatial variability in LST. The optimal bandwidth was determined via an iterative optimization procedure that minimizes the corrected Akaike Information Criterion (AICc). A new table (Table S8) has been added to the Supplementary Material, summarizing the GWR simulation setup, performance metrics, kernel structure, and bandwidth parameters used in the AICc minimization.

4. I am also confused with the description of scenario 2, specifically if the reported relative feature importance (RFI) is normalized in the right way? As you mention this is a normalized metric. But for scenario 2 FRP was 0.503 SR was 0.143 and Aerosol loading was 0.68. For these three predictors the normalized RFI sum more than 1. Is this a typographical error, a misunderstanding on my part, or is there some calculation mistake?

The authors apologize for this typographical error. The reported relative feature importance (RFI) values were, in fact, normalized.

In the revised manuscript, the RFI scores have been updated following the inclusion of additional parameters, and all reported values have been thoroughly rechecked to ensure accuracy.

5. Δ LST is reported as $+0.48^\circ\text{C}$ (mean) with range, but it's unclear whether this difference is statistically significant after accounting for temporal variability and dependence, and how many grid cells underpin the estimates. Provide confidence intervals (e.g., bootstrapped CIs) for Δ LST and Δ AOD. Additionally, consider how comparison of pre-post events within the same grid for fire vs. similar non-fire grids) could help strengthen causal inference.

Thank you for this suggestion. In the revised manuscript, we applied nonparametric bootstrapping to assess whether the Δ LST and Δ AOD attributable to fire remained statistically significant after accounting for temporal variability and dependence (L606–L615). The grid selection criteria were further refined to ensure that the estimated changes in LST and AOD could be attributed solely to fire. A total of 7,489 paired fire and no-fire grid cells from 2017–2021 were used to quantify relative differences. All grids, regardless of fire status, were selected from within the extended geographical region to capture localized variations in surface temperature and aerosol loading.

To quantify uncertainty, we computed 95% confidence intervals (CIs) for Δ LST and Δ AOD using nonparametric bootstrapping. For each variable, 10,000 bootstrap samples were generated by resampling

grid pairs with replacement, and the mean difference was recalculated for each iteration. The 2.5th and 97.5th percentiles of the resulting sampling distribution were used to define the 95% CI. Bootstrapping revealed a statistically significant increase in both Δ LST (0.57 °C; 95% CI: 0.33–0.81 °C) and Δ AOD (0.13; 95% CI: 0.08–0.17) in fire-affected regions. As both confidence intervals exclude zero, the estimated differences are statistically robust and unlikely to arise from sampling variability.

6. Justify selection of FRP density threshold (>5 MW grid⁻¹), the 1500 MW threshold and the 50% growth/decline rule for scenario 1, and the $r \geq 0.5$ threshold for scenario 2. Add rationale and sensitivity checks (e.g., try thresholds (+20%, -20%).

The onset, peak, and duration of post-harvest residue burning vary substantially from year to year, and numerous small, intermittent fires occur throughout the harvesting period. Unlike forest fires, residue-burning events are highly sporadic and exhibit strong spatial and temporal heterogeneity (Fig. S2 and Fig. S4) due to fragmented landholdings, where individual fields may experience multiple low-intensity fires. Such small fires release insufficient radiative energy to meaningfully alter evapotranspiration or surface albedo and therefore have limited impact on LST. However, during peak burning periods, fire intensity increases markedly and has the potential to modify the regional radiative balance.

To examine the temporal dynamics of fire intensity and its implications for regional LST, two scenarios were defined using distinct thresholds. Both Scenario 1 and Scenario 2 were used to capture days with substantially elevated FRP across the region (Table S5). In Scenario 1, a relatively larger number of fire days were selected within each burning season, beginning from the initial rise in FRP and continuing until a marked decline in aggregate FRP was observed. All small, sporadic, and low-intensity fires occurring during the early and late stages of the burning season were deliberately excluded. Intermittent cases in which aggregate FRP increased by $\geq 50\%$ relative to the preceding day but subsequently declined were also omitted. To avoid the inclusion of small-scale fire clusters, an additional criterion of cumulative FRP >1500 MW was applied.

Scenario 2, by contrast, included only periods characterized by persistently high FRP values that exhibited a strong positive association with regional mean LST. It represented days with a steady increase in aggregate FRP over time, indicating intensifying fire activity and energy release, accompanied by a positive correlation with regional mean LST. The requirement of >95% data completeness across FRP, LST, AOD, and meteorological variables restricted the number of eligible days per year but improved the robustness of the results.

Authors acknowledge that the major conclusions of this study are not sensitive to the exact threshold choices or to sensitivity checks. However, variations in thresholds can shift the yearly temporal window, leading to differences in the absolute magnitude of LST change across northwestern India. Accordingly, as stated in the abstract, both scenarios consistently identify FRP as a dominant driver of LST variability, although the precise magnitude of LST perturbation remains sensitive to the domain and the selection of fire-affected days. Overall, this analysis provides robust evidence that residue burning across

northwestern India significantly influences LST and alters the regional radiative budget. Nevertheless, the exact magnitude of fire-induced LST perturbations depends on both the intensity of burning and concurrent meteorological conditions.

7. The Hurst exponent computed and interpreted as persistence (> 0.5), is relevant when there is large number of data points which are specifically not impacted by seasonal trends, however, in this case, with only 5-year dataset and strong seasonality, Rescaled Range (R/S) analysis for Hurst estimation can be sensitive to trend and seasonality. This is an important featured previously determined by various observational studies in this part of the world where both inter- and intra-annual variability is common (Lin et al. 2020; Liu et al. 2024 etc.). Did the authors conduct detrended fluctuation analysis (DFA) or remove seasonal cycle before computing Hurst. Furthermore, the author's interpret values of $H > 0.5$ as indicating persistence and suggest that anomalies may "remain stable in the near future." While $H > 0.5$ indeed indicates statistical persistence or long-term data analysis, this interpretation could overstate the predictive implications of the Hurst exponent, especially given the relatively short five-year data record and the presence of strong seasonal cycles (such as monsoon and agricultural seasonality) inherent in the dataset. I recommend the authors temper the predictive language by replacing claims that anomalies "will" persist with the more cautious and appropriate statement that $H > 0.5$ indicates statistical persistence. Additionally, the authors are encouraged to clarify whether seasonal cycles were accounted for or removed prior to computing the Hurst exponent, as this can significantly affect estimates derived from R/S analysis.

Thank you for this valuable comment. The authors did not apply detrended fluctuation analysis because the dataset used to estimate the Hurst exponent represents a single season (October–November) from 2017 to 2021. All residue-burning events examined in this study occur exclusively during the post-monsoon period, which exhibits distinct characteristics compared with the monsoon (JJAS) and winter (DJF) seasons. Therefore, the retrieval and analysis of FRP, AOD, and LST were based on a single-season dataset, and seasonal decomposition was not intended.

As suggested, the interpretation of the Hurst exponent for LST, FRP, and AOD in Section 3.2 has been revised. We no longer refer to "certainty" in predicting anomalies and instead emphasize statistical persistence within the dataset. This clarification has been incorporated into the revised manuscript.

8. There are several small typos/grammatical slips (e.g., "Dring" typo of "During" (Page 19), "reginal" typo of "regional" (Page 19 3.4), please go through the manuscript carefully and correct these and similar mistakes.

Thank you for this comment. Accordingly, extensive language editing has been made throughout the manuscript to improve readability, grammar, and conciseness.
