

Thank you for your valuable, detailed comments. Based on these identified defects we were able to trace specific errors in our processing code – a corrupted ignition layer, land cover lost to projection errors, and duplicated LAI frames. These findings helped us inspect our pipeline end to end and rebuild the dataset from source under a single version.

Now, the second version of the dataset (WSDS-CAN v2) contains 19,468 fire events across all 14 Canadian agencies ranging from 2001 to 2025. The new version separates the two prediction tasks that you correctly noted we had mixed.

We will also correct one figure from the submitted manuscript: the number of fire events in the first version is 2,565 not 2,656, which was a typographical error.

General comments:

1. we regenerated, renormalized the ignition layer and now all values fall in [0,1] interval.

2. fully accepted. In the first version we only relied on polygons from CNFDB and filtered fire events based on the availability of their reported start/end dates and a 60-day limit on the fire durations, that resulted in much lower number of records. However, in the second version the National Burned Area Composite (NBAC) is the primary perimeter source for the whole country with NFDB records linked to it spatially. This allowed us to increase the number of fire records to 19,468. The breakdown of fire events is as follows:

BC: 4,084 – MB: 2,571 – NT: 2,505 – AB: 2,425 – QC: 2,337 – SK: 1,989 – ON: 1,421 – YT: 1,187 – Parks Canada: 400 – NS: 239 – NL:168 – DND: 54 – NU: 46 – NB: 42

3,692 events are contributed by Yukon and Northwest Territories. A national distribution map will be added as new figure to the revised manuscript.

3. fair point. The reason for stating 2001-2020 was that we used a year filter to match the fire records to land cover availability but oversaw the resulting fire ranges per province/territory. In the second version, the dataset spans 2001-2025 and every year is filled: from 216 events in 2001 to 1,696 in 2023. Out of the fourteen agencies, nine span all 25 years and the remaining five – NB, NU, NS, DND, NL – that do not, contribute 549 events in total. This is due to their genuine scarcity of fires above 1 ha rather than a processing choice. Since the number of events vary by year and agency, a random split would let neighbouring fires from the same season fall on both sides of the split; the pre-computed spatial, temporal and spatio-temporal block splits described in Specific comment 10 are our response to that imbalance.

4. we used day-of-burning with a hotspot-density criterion: ≥ 3 days with detections, ≥ 10 detections within 1 km of the perimeter, and ≥ 2 days observed duration. These thresholds are a data-sufficiency rule not a fire behaviour rule. Interpolating a day-of-burning surface needs arrival times distributed in both space and time; below these counts the surface would describe the interpolator rather than the fire. The supporting counts are included as columns (hs_n, hs_days) so users can tighten or relax the criterion.

7,826 of fire events satisfy this criterion and can be used in a spread prediction task. The target is a day-of-burning (DOB) raster that is interpolated from MODIS and VIIRS hotspot arrival times from which daily masks follow as $\text{mask}(t) = \text{DOB} \leq t$. we will describe these fires as hotspot-derived rather than observed as they are inferred from detection timing not measured fire fronts, and the hotspot density supporting each event will be reported in the revised manuscript. Figure 5 will also be recomputed against the daily masks for the spread prediction demonstration.

18,992 events have a usable final burned mask and can be used for final perimeter prediction. The 7,826 spread events mentioned above, are a subset of these. Therefore, 11,166 events support the final perimeter prediction task only. The remaining 476 are same-day fires larger than 100 ha and have at most one day of hotspot detection making the weather window undeterminable. These fires support neither prediction task.

5. Counting the sine and cosine components separately and counting the four daily meteorological variables as separate variables from the noon values, the second version now carries 33 covariates which is a significant increase compared to v1's eight. 12 of these variables are already built; the 21 meteorological and fire weather variables are produced by the same pipeline once the ERA5-Land retrieval finishes. The 33 covariate breakdown is as follows:

Terrain (5): elevation, slope, aspect sin, aspect cos, topographic wetness index

Fuel and land (4): land cover (17 IGBP classes, fire's own year), fuel type, non-fuel fraction, water fraction

Vegetation (2): LAI and LAI quality flag

Noon meteorology (11): u10, v10, wind speed, wind direction sin/cos, T2m, skin T, dewpoint, RH, precipitation, VPD

Daily statistics (4): maximum temperature, minimum RH, maximum wind, 24 h precipitation

Fire weather (6): FFMCI, DMC, DC, ISI, BUI, FWI

Ignition (1): Gaussian likelihood

Fire weather system is defined on local noon observations which means noon temperature, noon relative humidity, noon 10 m wind, and 24 h precipitation to noon whereas we used ERA5-Land daily means. On the other hand, FWI cannot be computed from daily means in a way that is comparable to CWFIS or CFSD. Thus, we redownload ERA5-Land hourly retaining a window that spans local noon through mid-afternoon (00 UTC + 15-23 UTC) from Newfoundland to Yukon while the 00 UTC step carries the previous day's precipitation accumulation. Based on this, the FWI inputs and conventional daily statistics (maximum temperature, minimum RH, maximum wind, 24 h precipitation) are derived as separate variables. The convention for every meteorological variable will be stated in the revised manuscript.

6. In the second version we offer a land cover composition report and a boreal_eligible flag. In the composition report every fire carries the fraction of each of the 17 IGBP classes taken from MCD12Q1 for the fire's year itself and computed twice: inside the perimeter and over the surrounding tile.

The boreal-eligible flag needs two things: first, the ignition point must fall inside the boreal or hemiboreal zone mapped by Brandt (2009) and second, no more than half of the burned area is non-forest (grassland, cropland, urban, cropland-natural mosaic, snow and ice, barren, water). The ignition point for 14,884 fires (76.5% of the events) fall inside the boreal zone, 1,748 (9.0%) in the hemiboreal transition, and 2,836 (14.6%) fall outside of both zones.

Based on the composition report, 847 fires from 18,165 that have usable land cover for their own year, exceed the non-forest threshold being dominated by grassland and cropland.

Taking both criteria (composition report and boreal_eligible) together gives 15,026 of the 18,388 events for which the combined flag could be decided, which is 81.7% of those and 77.2% of all 19,468 events. The two denominators differ because an event lying outside the boreal and hemiboreal zones is ineligible regardless of its land cover, so 223 events without a usable own-year land cover are still decided by location alone. For the remaining 1,080 events, almost all from 2025 where land cover is not yet released, we left the composition criterion undetermined rather than inferring it from another year. We use distinct columns for the zone label, the non-forest fraction, the stricter boreal-only variant, and each criterion so a user can reproduce our cohort with a filter or impose their own.

The flag does not remove fire events from either prediction task. A user who filters to boreal_eligible retains 14,612 of the 18,992 final-perimeter events and 6,226 of the 7,826 spread events.

We considered your suggestion for using a fuel-type filter and decided to use land cover instead. The reason is that the national FBP fuel-type grid is published as a single map dated May 2024 making it the landscape after the fire for 86.8% of the events from before 2024. In addition, inside the average 2023 fire perimeter, 85% of pixels have one fuel class that matches the burned area of that season, compared with only 1-6% in other years. Therefore, using this map to select fire events could introduce bias since the selection would partly depend on the fire outcome itself.

According to NBAC, fire event ON_AYL1_2010 is a prescribed burn (44 ha, April 2010). We exclude prescribed burns (318 records) from the second version, so this record no longer exists in the dataset. Fire event QC_20181080442 currently identified as WSDSCAN_2018_QC_0121 is 1.6 ha fire located in the agricultural St. Lawrence lowland south of Montreal. It has two 500 m land-cover cells under its perimeter that are woody savanna and savanna while half of its tile is cropland or cropland mosaic making it a woodlot fire in a farming landscape. This example passes the composition test and fails the location one. The zone join puts it outside the boreal zone so boreal_eligible flag is false for it.

We favor flagging over removing as the dataset's value for small fires depend on keeping events near settlement and agriculture margins where suppression often has a greater effect on final size than fuel type.

Specific comments:

1. in the revised manuscript we will add source products with version resolution and temporal range; and delivered variables with units, data type, valid range, fill convention, temporal resolution and the transformation applied to the table. Additionally, we will add categorical grouping column in the same table as per the second referee's comment.

2. during investigation of the ignition layer, we realized a few defects. First, we had joined ignition coordinates to perimeters by attribute rather than by geometry which resulted in ignition points lay inside their own perimeters for only 47.6% of the 2,565 events. The joined point was then used as it landed, with no check that it fell inside the perimeter. For the instances that the joined point was badly wrong, its pixel index fell far outside of the tile. Hence the kernel evaluated to zero everywhere producing NaN and infinite values after normalization. Second, the kernel width was bound between 0.01 and 5 pixels so above approximately 25 pixels of fire radius the width saturated making identical ignition blobs for a 100 ha fire and a 100,000 ha one. In the second version, we joined the point and perimeter spatially, expressed the

kernel width in meters which scales with fire size, computed distances from map coordinates in floating point rather than in integer pixel indices, normalized the layer and ensured values remain in [0,1] interval.

We also developed a QC code to verify the second version rebuild. The code rejected eight events due to an empty ignition layer. Further inspection showed that these events were part of a larger group of 195 fires whose ignition seed (the first hotspot detection) fell outside the fire perimeter, despite having an offset of zero and confidence rating of “high”. We identified two causes for this: first, the seed was chosen from the earliest detections near the perimeter instead of from detections inside it. Second, in about one-third of these cases, the history of detection belonged to a different fire due to the duplication defect described in response to Specific comment 11.

After repairing both issues, all 19,468 fire events now have their ignition points inside their own perimeter. All corrections are recorded. For 122 events, the ignition coordinates were moved to the closest interior location and the confidence level downgraded to “medium”. For another 15 events, where the original seed was more than 1 km away from the perimeter, it was rejected and replaced with an interior point from the perimeter and the confidence level was marked as “low”. We store the original ignition points in `ign_seed_x/ign_seed_y`. the two repairs mentioned above, account for 137 of the downgrades other than that the confidence level is set by how the ignition point was obtained: 13,783 events take it from an NFDB point already inside the perimeter and 765 from a first hotspot detection, both rated high; 4,710 take it from an NFDB point within 2 km that had to be snapped to the perimeter, rated medium (3,331 records), or low where the point had to move more than 250 m, which applies to 1,379 of them; and 73 fall back to the perimeter's interior representative point, rated low. Thus, the final statistics are as follows: 14,548 as “high”, 3,453 as “medium”, and 1,467 as “low”.

In the second version the ambiguity is resolved by centering the Gaussian distribution on the ignition coordinate which now carries the confidence level.

3. this was a wording issue describing one-hot encoding in eight channels in the previous manuscript. The model used `nn.Embedding(num_landcover_classes=18, embedding_dim=8)`, a learned 8-dimensional embedding over the 17 IGBP classes plus a nodata index, so the published v1 landcover rasters carry the full IGBP codes. An issue was that landcover rasters were reprojected without an explicit destination CRS which happened to be harmless as the source product was already in the target projection. The second version explicitly sets the destination CRS, ships the landcover as a single categorical raster carrying the IGBP class code per pixel, with no encoding imposed, together with per-class area fractions computed inside the perimeter and over the tile. The encoding is left to the model. The revised manuscript will include the class legend and will also report the full training configuration, data splits, and baseline comparisons.

4. we found three reasons for this issue:

First, in the previous version we used MOD15A2H that provides 8-day composites. Since the median fire length is 6 days, most fire events overlap only one to two composites. Thus, this is the main reason that requesting five frames resulted in duplications.

Second, MODIS LAI products use 249-254 as landcover flags rather than LAI measurements, whereas only 255 is declared as fill value. Therefore, scaling these codes allowed invalid values as big as 25.4 (corresponding to code 254, inland water) into a layer whose maximum value is 10.

Third, we had selected the nearest available composite without recording its actual date.

In the second version, we addressed all three issues. We use MCD15A3H (4-day, Terra and Aqua) as the primary product with MOD15A2H covering 2001 to mid 2002. We masked flag codes and quality-flagged pixels before reprojection and recorded the actual composite date for every frame. Across the 19,460 events with an LAI stack, the median fire event contains 18 composites, 17 of which are distinct. Eight events have no LAI stack at all because their entire window falls outside the MODIS LAI record for that location, which is what happens to a winter fire. Furthermore, 81.0% of all events have no identical LAI frames. The frames are marked based on their timing relative to the fire: before ignition, during the burn, and after the last recorded burning day so users can select frames based on their relationship to the fire. Any repeated frames are now visible in the data.

5. will be added to the revised manuscript. We used Python with GeoPandas/Shapely for vector operations, rasterio/GDAL for raster handling, xarray for netCDF stacks and PyTorch for the baseline.

6. this is resolved in v2. All 19,468 events have start and end dates with source of each recorded:

For start dates, 17,848 are taken from agency records (NBAC's AG_SDATE field for 17,683 events and from NFDB point records for the 165 where NBAC carries none) and 1,620 are derived from hotspot detections (the CWFIS national hotspot archive, which is MODIS and VIIRS satellite detections).

For end dates, both agency-reported and hotspot-derived end dates are retained when they were available. They are both available for 8,715 events. The primary end date is the last hotspot detection if one exists and is not earlier than the start date; otherwise, it is the agency end date. On that rule, hotspot end dates are primary for 14,580 fires and agency-reported for 4,883. For the remaining 5 events neither condition held meaning the archive's last detection preceded the recorded start date and no agency end date existed; so the end date already carried in the source record was kept and labelled accordingly.

The two sources have a systematic difference. For the 8,715 events that have both dates available, hotspot-derived end dates are earlier by a median of 23 days (interquartile range 10 to 44 days earlier) and are earlier in 94% of cases. Instead of choosing one source, we keep both and record their provenance. We also add a `duration_is_lower_bound` flag that is set to true for the 14,580 events with hotspot-derived end dates, as satellite detections may stop before the fire is fully extinguished. Additionally, each event carries a three-level provenance label for its date pair, computed on the start date and the originally selected end date rather than on `end_primary`, which ships as a separate column with its own source field: 13,190 agency, 1,617 hotspot and 4,661 combined, where agency means both the start and end date come from agency records, hotspot means both come from hotspot detections, and combined means one of each.

7. this is solved in v2. All masks are written as uint8 with values in {0,1}. This is verified by the QC code for every file. This check is part of the pipeline released with the dataset.

8. we have applied the QC code to all the second version data except the meteorological data from ERA5-Land. At the time of writing this response, that part of data is still being downloaded. Based on the QC code, all declared layers are present and correctly shaped, there are no all-NaN layers, values sit within the stated ranges, masks are binary, ignitions are in [0,1] summing to a finite value, and time axes strictly increase without duplications. Once the download for meteorological data finishes, we will ensure the coverage spans the event window. Events that fail any check are retained and marked ineligible with the reason recorded.

Across all fire-size classes, the rasterised burned masks recover 95–98% of the reported burned area at the median. Only 86 fires (0.44%) fall below 80% area recovery, all of them between 1 and 4 ha. In 84 cases, the small perimeter covers only 1–32 pixels on the 31 m grid, that makes edge quantisation the main source of error. In the remaining two cases, the burned area consists of separate patches kilometres apart, while the raster tile is sized using the bounding box. These 86 events are flagged, and both rasterised and reported areas are provided for every event. The QC criteria and per-event results are included with the dataset and will be described in the revised manuscript.

We also tested the QC code against actual failures. When applied to the initial static stack (the per-event file holding the layers that do not change through the fire), it rejected eight events. Inspecting those failures revealed the defects described in Specific comments 2 and 11. After the underlying defects were corrected and the affected files rebuilt, the check passed for all 19,468 events.

9. our previous claim stood on the visible LAI drop compared to the final burned mask of some sample fires and we accept it could not be a reliable claim for the role and impact of LAI. The fact is LAI is not a fire-progression observable and we will no longer portray it as one. In the second version, LAI is used only as pre-fire fuel-state covariate and a post-fire severity indicator. It is documented as `lai_pre`, `lai_post` and intermediate frames. 99.5% of the events carry at least one composite falling within the burn window, with a median of three. They are provided as inputs, not as evidence of progression.

Instead, the progression signal comes from hotspot-derived day-of-burning that is a more appropriate observable. We also attempted to address the following. To compensate the slow LAI variation, we used the 4-day product that halves the compositing interval. In addition, the quality layer is now retrieved and applied, so smoke and cloud contaminated pixels are masked rather than averaged.

Within this narrower role, the LAI signal is measurable. Out of the 19,460 events with an LAI stack, 95.2% have both a usable (have at least 25% of the perimeter's pixels valid) pre-fire and post-fire composite. These frames are elected based on coverage inside the fire perimeter instead of date alone, with the post-fire frame targeted at a fixed lag (median 15 days) to improve comparability between events. Median LAI inside the perimeter decreases from 1.44 before the fire to 0.91 after it. LAI begins to recover after a fire, so a post-fire frame taken five days after the end and one taken sixty days after would measure different things. Selecting the frame nearest a fixed target lag of 16 days (which gives an achieved median of 15 days), rather than simply the next available composite, makes the pre-to-post change comparable between events instead of being confounded by how soon the next clear composite happened to fall. The achieved lag is 11 to 18 days for the middle 80% of events.

The spatial limitation is also stated explicitly: one 500 m cell represents 25 ha, so for fires smaller than 25 ha, the perimeter mean reflects the surrounding area more than the burned area. 13,243 (68.1%) events exceed this threshold and are flagged accordingly.

10. agreed. In the second version, we adopt fixed, pre-computed dataset splits for spatial block (where geographic blocks are assigned entirely on one fold), temporal block (where whole years are held out), and spatio-temporal (that applies both constraints). The previous random split is also retained for comparison to v1. Baseline results will be reported under all four protocols, and we expect the more realistic splits to result lower scores than v1. The split assignments are included directly in the dataset, so all users evaluate the models on the same folds and results are directly comparable.

11. this is resolved in v2. We assign a uniform synthetic identifier to every event in the format of WSDSCAN_year_agency_seq. the original agency identifiers are kept as separate columns so events are still traceable to their source records without the naming collisions.

Our event builder originally linked hotspot histories using GID, which combines the year with the agency's fire number. Since fire numbers are unique only within each agency, this key was not unique nationally. For instance, Manitoba and Saskatchewan both recorded a fire numbered 176 in 2001. This caused a many-to-many join that duplicated 411 records across 399 fire perimeters. More seriously, some fires were assigned another fire's hotspot history, affecting first and last detection dates, end dates, spread eligibility, and the ignition seed (in about one-third of the out-of-perimeter ignition cases described in Specific comment 2).

We rebuilt the hotspot summaries for every affected perimeter using the detection archive and recalculated all dependent dates and flags. We retained one row per perimeter and marked the surplus records as ineligible, with a pointer to the corresponding kept record. The corrected total count used throughout author responses is 19,468 events; the 411 surplus rows remain in the table rather than being deleted. The released pipeline now uses the row identifier to join hotspot histories, as it is the only key that uniquely identifies each perimeter.

Technical corrections:

All accepted and will be corrected in the revised manuscript.