

# scientific data



OPEN

DATA DESCRIPTOR

## The Canadian Fire Spread Dataset

Quinn E. Barber<sup>1</sup>✉, Piyush Jain<sup>1</sup>, Ellen Whitman<sup>1</sup>, Dan K. Thompson<sup>2</sup>, Luc Guindon<sup>3</sup>, Sean A. Parks<sup>4</sup>, Xianli Wang<sup>1</sup>, Matthew G. Hethcoat<sup>1</sup> & Marc-André Parisien<sup>1</sup>

Satellite data are effective for mapping wildfires, particularly in remote locations where monitoring is rare. Geolocated fire detections can be used for enhanced fire management and fire modelling through daily fire progression mapping. Here we present the Canadian Fire Spread Dataset (CFSDS), encompassing interpolated progressions for fires  $>1,000\text{ ha}$  in Canada from 2002–2021, representing the day-of-burning and 50 environmental covariates for every pixel. Day-of-burning was calculated by ordinary kriging of active fire detections from the Moderate Resolution Imaging Spectroradiometer and the Visible Infrared Imaging Radiometer Suite, enabling a substantial improvement in coverage and resolution over existing datasets. Day of burning at each pixel was used to identify environmental conditions of burning such as daily weather, derived weather metrics, topography, and forest fuels characteristics. This dataset can be used in a broad range of research and management applications, such as retrospective analysis of fire spread, as a benchmark dataset for validating statistical or machine-learning models, and for forecasting the effects of climate change on fire activity.

### Background & Summary

Wildfire science is undergoing a data revolution enabled by improvements in the resolution and coverage of satellite-based observations of wildfire<sup>1</sup>. There has been a steady increase in the number and quality of active-fire monitoring sensors, first and foremost being the Moderate Resolution Imaging Spectroradiometer (MODIS), available 2001–present<sup>2</sup>, and the Visible Infrared Imaging Radiometer Suite (VIIRS), available 2012–present<sup>3</sup>. These low-earth orbit systems capture active wildfire data several times daily with no cost to the end user, allowing scientists and wildfire managers to track wildfire ignitions and daily spread<sup>4–7</sup>. Historically, data access challenges and processing requirements have limited efforts to consolidate and use these large active fire datasets, but with the launch of open data platforms such as Google Earth Engine<sup>8</sup> (GEE) and the NASA Fire Information for Resource Management System<sup>9</sup> (FIRMS), these barriers have largely been removed.

Researchers have developed several automated fire-event-delineation algorithms based on active-fire detections, thermal anomalies detected by satellite observation (hereafter “hotspots”). For example, FRY<sup>10</sup> is a global database of fire patches built on the novel combination of MCD14A1 hotspots<sup>11</sup> (hereafter “MODIS hotspots”) and change-detection maps, which were built on the integration of Collection 6 MCD64A1 burned area<sup>12</sup> (hereafter “MODIS burned area”) and FIRECCI41<sup>13</sup>. The Global Fire Atlas<sup>14</sup> and GlobFire<sup>15</sup> go beyond fire delineation by using MODIS burned area<sup>12</sup> to identify daily fire perimeters, providing information such as daily fire size and estimates of fire front location and direction of fire spread. The Fire Events Delineation<sup>16</sup> (FIRED) algorithm provides further improvement through automated temporal clustering of fire events, producing a continuous estimate of fire spread rate and direction for the conterminous United States. However, all four algorithms (FRY<sup>10</sup>, Global Fire Atlas<sup>14</sup>, GlobFire<sup>15</sup>, and FIRED<sup>16</sup>) are reliant on the change-detection algorithm of MODIS burned area<sup>12</sup> instead of fire hotspot interpolation, and thus they are subject to a multi-day temporal uncertainty and a 500-m or coarser resolution. For example, when MODIS burned area<sup>12</sup> products were compared against MODIS hotspots, 32% of all MODIS burned area pixels reported fire arrival three or more days after the first hotspot detection<sup>12</sup>. Newer global burned area mapping algorithms such as FIRECCI50<sup>17</sup> and FIRECCI51<sup>18</sup> have reported finer spatial resolution and accuracy of day-of-burning estimates over MODIS burned area<sup>12</sup>, but still suffer from temporal uncertainty, which reduces their usefulness for identifying daily weather conditions during fire spread.

<sup>1</sup>Northern Forestry Centre, Canadian Forest Service, Natural Resources Canada, 5320-122 Street NW, Edmonton, AB, T6H 3S5, Canada. <sup>2</sup>Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, 1219 Queen Street, Sault Ste. Marie, ON, P6A 2E5, Canada. <sup>3</sup>Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada, 1055 Rue du Peps, P.O. Box 10380, Station Sainte-Foy, Quebec City, QC, G1V 4C7, Canada. <sup>4</sup>USDA Forest Service, Rocky Mountain Research Station, Aldo Leopold Wilderness Research Institute, 790 E. Beckwith Ave., Missoula, MT, United States of America. ✉e-mail: quinn.barber@nrcan-rncan.gc.ca

Interpolation of day-of-burning or fire arrival time (hereafter “progression mapping”) can be used to answer complex ecological and fire management questions. Hotspots and hand- or GPS-delineated fire progressions from aerial and ground observations have long been an essential part of wildfire operational management<sup>19</sup>. However, near-real-time fire progression mapping from satellite observations is gradually being practical<sup>20</sup>, given that continuous aerial observations by aircraft are no longer necessary. The Fire Event Data Suite<sup>7</sup> (FEDS) algorithm uses spatiotemporal aggregation of VIIRS hotspots with an alpha hull to automate fire progression mapping at a 12-hr timescale. However, to better address missing hotspots, most automated fire progression mapping estimates fire arrival date using interpolation as opposed to spatiotemporal aggregation<sup>5</sup>. When assessed against high-quality infrared fire burned area observations, pooling VIIRS and MODIS and interpolating using kriging or natural neighbour interpolation leads to the best estimates of fire arrival time<sup>21</sup>. Although these methods are limited to approximately daily-scale accuracy by the limited satellite overpass frequency, methodological and sensor-technology advancements suggest that fire progression mapping on a subdaily scale is becoming possible. For example, coarse-resolution subdaily fire mapping using the Geostationary Operational Environmental Satellites (GOES) is possible for the Continental United States<sup>22</sup>; however, the satellites’ high view angles and the products’ coarse spatial resolution and high omission errors limit their usefulness at Canadian latitudes.

Improved fire progression mapping represents an opportunity to address a variety of wildfire research questions. For example, progression mapping provides an estimate of the date and thereby the weather conditions under which each pixel burned, necessary information for modelling fire spread under various fire weather conditions<sup>6,23,24</sup>. When coupled with maps of bottom-up factors such as land cover and topographic characteristics, progression mapping can serve as a foundation for improving the knowledge base of large-scale wildfire spread phenomena, with applications involving wildfire in unconventional settings, such as novel disturbances or treed peatlands. Given that over twenty years of satellite observations of wildfire are available, even relatively rare circumstances or phenomena such as wildland-urban interface fires and high-latitude fires constitute hundreds or thousands of gridded data points for model building.

Here we present the Canadian Fire Spread Dataset<sup>25</sup> (CFSDS), an event-based daily fire progression dataset at a resolution of 180 m, built on the combination of high-precision wildfire boundaries and active fire detections from multiple sensors. The CFSDS provides a research-oriented fire growth dataset that is useful for applications such as building statistical and machine-learning fire spread models. In this first release we provide a collection of interpolated wildfire progression maps and associated environmental covariates for all wildfires larger than 1,000 ha that occurred in the forested areas of Canada from 2002 to 2021. Hotspots<sup>9</sup> were identified with co-located wildfire perimeters from the National Burned Area Composite<sup>26</sup> (NBAC). We combined VIIRS and MODIS hotspots<sup>9</sup> to produce daily-scale fire progression maps for each fire with at least 6 observed hotspots<sup>6</sup>. We attributed each burned point (i.e. 180-m pixel) with 50 environmental covariates such as daily weather, derived weather metrics (e.g. drought metrics), topography, and forest attributes. The final product is a dataset of 3,269 fires comprising 70,895 days of burning, each with an estimate of the daily fire spread distance and area of growth.

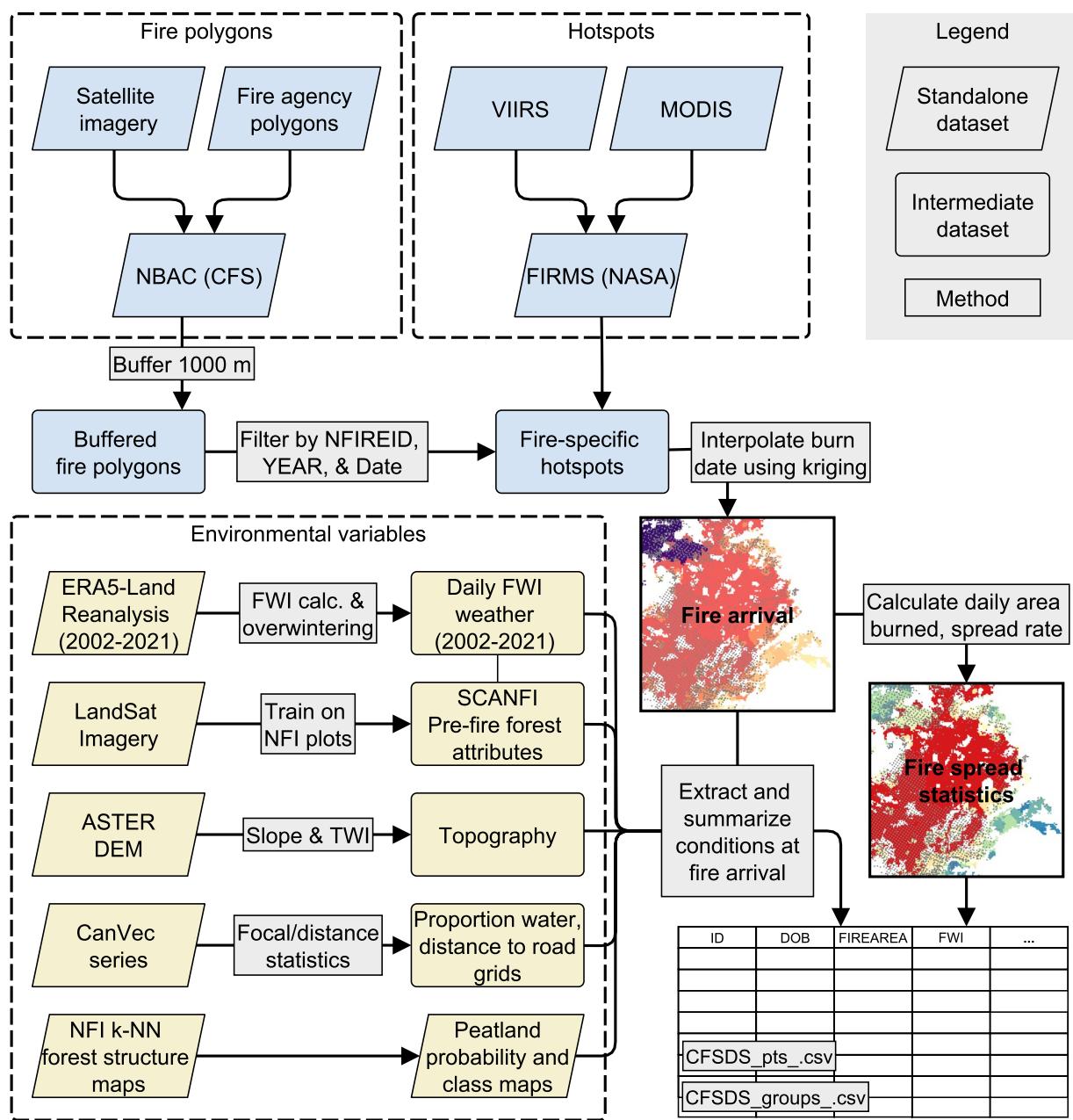
## Methods

We constructed the CFSDS<sup>25</sup> from three complementary groups of spatial and temporal data: final fire perimeter polygons, hotspots, and associated environmental covariates. In brief, hotspots were first filtered to predetermined fire perimeters. The detection times of these hotspots were then interpolated using ordinary kriging to provide an estimate of the date at which fire reached any given pixel within the fire perimeter. This raster of fire arrival times (i.e. “day of burning”) was then used to identify environmental covariates associated with the fire (Fig. 1).

**Fire event delineation and burned area polygons.** We obtained wildfire perimeters from NBAC<sup>26</sup>, an annual, national fire dataset mapping individual wildfire events in Canada (Fig. 1). NBAC<sup>26</sup> integrates three Canadian burned-area products, listed in order of decreasing precedence: pre- and post-fire Landsat imagery processed into normalized burn ratio using the Multi-Acquisition Fire Mapping System<sup>27</sup> (MAFiMS); the Canadian National Fire Database<sup>28</sup> (CNFDB), created from provincial/territorial wildfire agency polygon data; and the Hotspot and Normalised Difference Vegetation Index Differencing Synergy (HANDS) algorithm<sup>29</sup>. NBAC<sup>26</sup> also includes non-spatial attributes for each fire, such as fire start date and fire end date, used here for subsetting hotspots. From the full NBAC<sup>26</sup> record, we selected all fires that: (i) occurred from 2002–2021, and (ii) attained a final area of at least 1,000 ha (Fig. 2).

**Hotspots.** We downloaded all MODIS and VIIRS hotspots within Canada from 2002 to 2021 from the NASA FIRMS<sup>9</sup> (Fig. 1). The Collection 6 MODIS hotspots<sup>2,30</sup> (MCD14ML, V006/V0061) includes all Terra (MOD14) and Aqua (MYD14) satellite detections at a spatial resolution of 1 km and four daily overpasses. VIIRS hotspots include both the Suomi-NPP<sup>31</sup> (VNP14IMGT) and NOAA-20<sup>32</sup> (VJ114IMGTDL\_NRT) satellite detections at a spatial resolution of 375 m and four daily overpasses<sup>3</sup>.

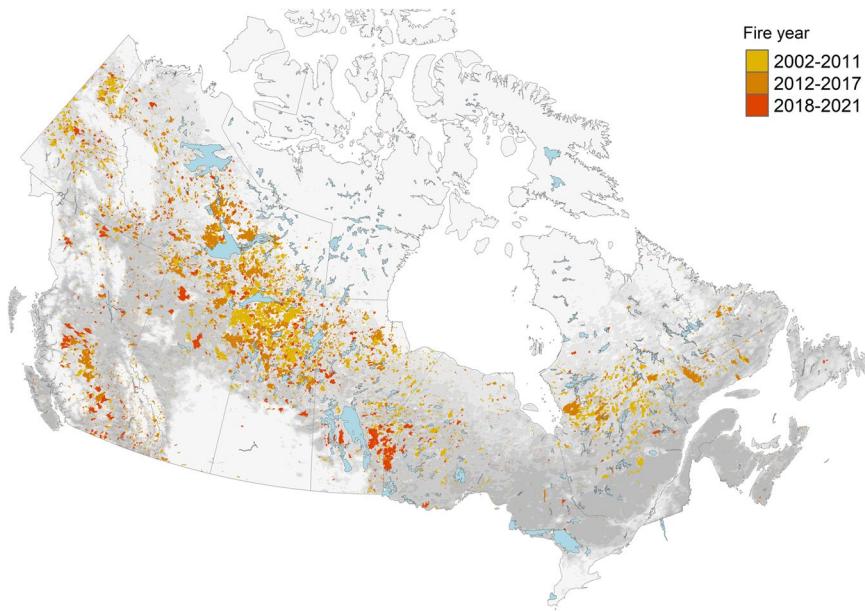
In total, this consisted of 3,731,070 hotspots, including all pooled MODIS and VIIRS hotspots<sup>9</sup>. The temporal resolution of the data depends on the year it was collected, due to satellite launches: four times daily from 2002 to 2011, six times daily from 2012 to 2017, and eight times daily from 2018 to 2021. We filtered hotspots to individual fire events using NBAC<sup>26</sup> polygon date and spatial location (Fig. 1). To account for hotspot precision limitations, all hotspots within 1,000 m of a fire polygon were considered for interpolation. Any hotspots occurring more than 30 days before the recorded fire start date or more than 30 days after the recorded fire end date were removed, to eliminate spatially coincident burning (from agricultural or silvicultural debris burning, for example) while still allowing for possible errors in fire start and end date attribution. We also applied a filter based on hotspot “confidence”, as reported by FIRMS<sup>9</sup>; MODIS hotspots with “confidence” less than or equal to 10 (range 0–100) and VIIRS hotspots with “confidence” of “LOW” (range “LOW” to “HIGH”) were removed, given that these hotspots are commonly associated with sun glint<sup>9</sup>. Hotspots were also adjusted to represent local standard time based on geographic location.



**Fig. 1** CFSDS production process. See Table 3 for covariate descriptions.

**Fire arrival time interpolation and processing.** For each individual fire, fire arrival times were interpolated using kriging on a 180-m resolution base grid (Fig. 1). This resolution was selected to produce outputs that were a practical size for storage and upload, while still being a multiple of our covariate datasets' resolution. Using all hotspots corresponding to an individual fire, we fit a variogram on hotspot detection times, as represented by the decimal numerical day of year (i.e. Feb. 1, 1200 hr as 32.5). We fit the experimental variogram using an exponential, spherical, and Matern model, and subsequently selected the best-fitting model for each fire. We used the default nugget, sill, and range values as determined by the *gstat* R package<sup>33</sup>. Using ordinary kriging, we interpolated the fire arrival time based on the fire-specific variogram fit. The resultant fire arrival time raster was then masked to burned areas as identified in NBAC<sup>26</sup>. The fire arrival date was rounded to the closest (temporally) date that was observed in the 6 nearest hotspots; this ensured that the predicted burning date occurred during detected fire growth and was not an artefact of averaging<sup>6</sup>.

The phenomenon of “peak burning” describes the tendency for wildfires in Canada to burn most intensely in the late afternoon or early evening<sup>34</sup>, a time of day during which the MODIS and VIIRS satellites are not present overhead. This artificially delays wildfire detection; for example, the VIIRS satellite overpasses occur at approximately 0100–0230 hr and 1300–1430 hr for the Canadian boreal zone. Therefore, the uncorrected day-of-burning rasters are biased towards late, nighttime detections, which may not always represent true fire arrival time. To limit the impact of these missing observations, we assigned any hotspot observed between



**Fig. 2** CFSDS fire polygons from NBAC. Fire year indicates the era of hotspot collection: 2002–2011 includes MODIS only, 2012–2016 includes MODIS and Suomi-NPP (VIIRS), 2017–2021 includes MODIS, Suomi-NPP, and NOAA-20 (VIIRS). Grey indicates mean canopy cover<sup>48</sup>.

1900–0800 hr to the previous day. We assigned an observed time of 1900 hr to these hotspots, immediately after diurnal “peak burning” as determined by peak Fire Radiative Power (FRP) using GOES-16 and GOES-17<sup>35</sup>. In this case, the temporal rounding of hotspots (i.e. kriging inputs) is not to be confused with temporal rounding of the day-of-burning grid (i.e. kriging outputs).

**Zero-growth days.** Days with no observable fire growth were included in the dataset as zero-growth days, defined as any day with less than three pixels (9.72 ha) of fire growth. Environmental covariates were defined using the spatial location of pixels from the next day of non-zero fire growth (a “fire growth day”). The subsequent fire growth day was taken as the next sequential day in which the fire grew more than three pixels. False zero-growth days are possible, but less likely due to the high detection efficiency of nighttime active fire detections<sup>30–32</sup>.

**Environmental covariates.** Daily fire arrival time interpolations were intersected with spatial datasets of 50 environmental covariates (Table 1) (Fig. 1), providing a record of the conditions under which each pixel burned. Environmental covariates were resampled to the same 180-m resolution grid used for interpolation of fire progression, where needed. We resampled using a mean aggregate function for continuous variables and a mode aggregate function for categorical variables.

**Fire weather.** We obtained weather data from the ERA5-Land dataset<sup>36</sup>, produced globally, hourly at 0.1 deg resolution. Specifically, 2-m air temperature, dewpoint temperature, total precipitation, 10-m u and v components of wind and snow cover variables were downloaded over Canada for the temporal period of the study (2002–2021). The data were further processed to provide the necessary inputs to the Canadian Forest Fire Weather Index (FWI) System<sup>37</sup>: the noon LST values of temperature, relative humidity, wind speed and the 24-hr accumulated precipitation ending at noon LST each day. The FWI System produces six indices: Fine Fuel Moisture Code (“FFMC”), Duff Moisture Code (“DMC”), Drought Code (“DC”), Initial Spread Index (“ISI”), and Buildup Index (“BUI”), and Fire Weather Index (“FWI”). The FWI System variables were calculated only for the snow-free period corresponding to the main fire season each year; season start was determined by the first day following three consecutive days of less than 50% snow cover. The DC is adjusted at startup by applying an overwintering correction based on the amount of overwinter precipitation<sup>38</sup>.

In addition to the FWI System outputs, the following daily weather variables were also included in the dataset: the maximum daily temperature (“Tmax”; °C), precipitation summed over 24 hr, from noon the prior day (“Prec”; mm), noon wind speed at 10-m elevation (“WS”; km/h), noon relative humidity (“RH”; %), and maximum vapour pressure deficit over a 24 hr period (“VPD”; hPa). Fire weather data is provided for all fires at the CFSDS data repository<sup>25</sup>.

**Fuels.** We represented modeled pre-fire fuels with the Spatialized Canadian National Forest Inventory data product<sup>39</sup> (SCANFI). In brief, Landsat multispectral imagery was trained on national forest inventory plots and used to map pre-fire fuels at a 30-m resolution, including percentage deciduous component (“prcD”), percentage coniferous component (“prcC”), biomass in tonnes/ha (“Biomass”), and crown closure percentage (“Closure”).

Peatland class	peatCl code
No peatland	0
Open bog	1
Open poor fen	2
Open rich fen	3
Treed bog	4
Treed poor fen	5
Treed rich fen	6
Forested bog	7
Forested poor fen	8
Forested rich fen	9

**Table 1.** Peatland classes<sup>41</sup> used in “peatCl” covariate.

Zone name	“ecozone” code
Arctic Cordillera	1
Northern Arctic	2
Southern Arctic	3
Taiga Plains	4
Taiga Shield West	5
Boreal Shield West	6
Atlantic Maritime	7
Mixedwood Plains	8
Boreal Plains	9
Prairies	10
Taiga Cordillera	11
Boreal Cordillera	12
Pacific Maritime	13
Montane Cordillera	14
Hudson Plains	15
Taiga Shield East	25
Boreal Shield East	26

**Table 2.** Ecozone names and associated “ecozone” code<sup>45</sup>.

**Topography.** All topographic variables were derived from the NASA ASTER Global Digital Elevation Model V003<sup>40</sup> 30-m digital elevation model (“DEM”). This DEM was subsequently used to calculate elevation (“elev”), slope (“slope”), aspect (“aspect”), and topographic wetness index (“TWI”).

**Peatlands and surface water.** Peatland presence (“peatland”) and peatland class (“peatCl”) are the modeled probability that any pixel is a peatland, and the modeled peatland type from 9 possible classes (e.g., bog, fen, etc.) (Table 1). These estimates were based on the combination of MODIS reflectance data trained on National Forest Inventory plots to estimate the distribution of treed and open peatlands at a resolution of 250 m, supplemented with Landsat-based land-cover datasets<sup>41</sup>.

Proportion of surface water (“hydrodens”) within a 2 km, 5 km, 10 km, and 25 km of each pixel was calculated using the 50 K CanVec Series Hydrographic Features dataset<sup>42</sup>.

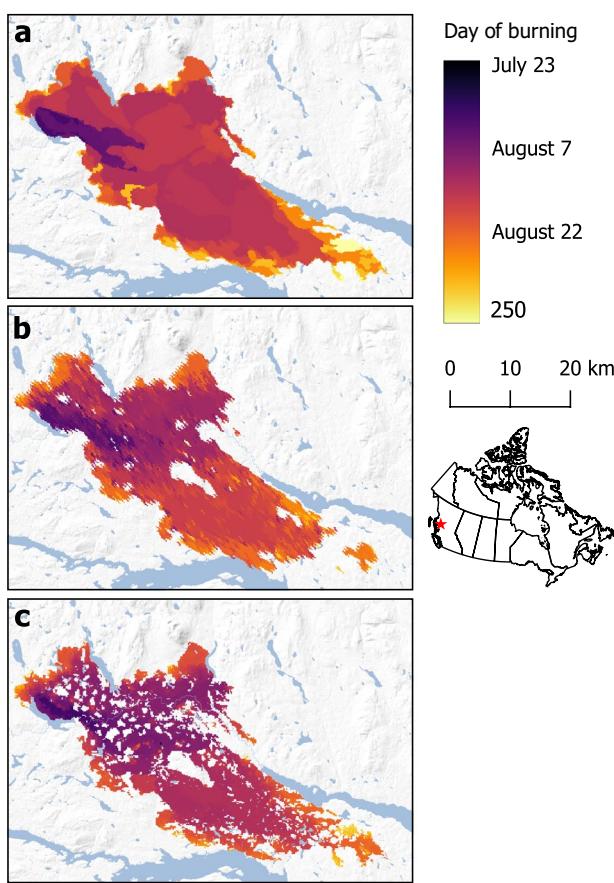
**Anthropogenic.** We represented road densities (“roaddens”) and distance to road (“roaddist”) using the 50 K CanVec Transport Features dataset<sup>43</sup>. Mean road density within 2 km, 5 km, 10 km, and 25 km of each pixel was calculated as the total length of roads per km<sup>2</sup>. Distance to road is based on linear distance to major roads or rail lines, not including temporary roads such as logging or access roads.

**Fire characteristics.** We calculated fire daily growth characteristics from the day-of-burn raster (*NFIREID\_YEAR\_krig.tif*). This includes daily area burned in hectares (“firearea”), cumulative area burned (“cumuarea”), and percent growth over the prior day’s total area (“pctgrowth”). The day since the start of fire (“fireday”) was calculated until extinguishment, including zero-growth days. Fire spread (metres per day) was estimated from the cumulative area burned relative to the cumulative area burned for the previous day. This assumes circular growth in a constant direction, according to the Eq. 1:

$$sprdistm = 2(\sqrt{cumuarea/\pi} - \sqrt{(cumuarea - firearea)/\pi}) \quad (1)$$

attribute	description	mean ± sd	source
ID	Fire ID	NA	NBAC <sup>26</sup>
DOB	Day of year burning	203 ± 28	CFSDS <sup>25</sup>
year	Year	2012 ± 5.7	NBAC <sup>26</sup>
fireday	Day of fire (ignition day = 1)	21.9 ± 21.1	CFSDS <sup>25</sup>
firearea	Fire growth this day (ha)	571 ± 2383	CFSDS <sup>25</sup>
ecozone	Ecozone (see Table 4)	NA	ESWG (1996) <sup>45</sup>
fwi	Fire weather index	11.1 ± 9.7	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
isi	Initial spread index	4.1 ± 3.4	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
ffmc	Fine fuel moisture code	79.0 ± 15.8	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
dmc	Duff moisture code	36.7 ± 26.5	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
dc	Drought code	285 ± 127	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
bui	Buildup index	51.6 ± 31.4	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
prec	24-hr precipitation (mm)	1.5 ± 3.2	ERA5-Land <sup>36</sup>
vpd	Daily max vapour pressure deficit (hPa)	15.8 ± 6.4	ERA5-Land <sup>36</sup>
tmax	Maximum daily temperature (°C)	20.9 ± 4.9	ERA5-Land <sup>36</sup>
ws	Noon windspeed at 10 m elevation (km/h)	9.4 ± 4.7	ERA5-Land <sup>36</sup>
rh	Noon relative humidity (%)	45.8 ± 11.9	ERA5-Land <sup>36</sup>
fwi_prev1	Fire weather index from previous day	11.0 ± 9.6	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
fwi_prev2	Fire weather index from 2 days prior	10.9 ± 9.6	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
rh_prev1	Noon relative humidity from previous day (%)	47.3 ± 12.6	ERA5-Land <sup>36</sup>
rh_prev2	Noon relative humidity from 2 days prior (%)	48.4 ± 13.3	ERA5-Land <sup>36</sup>
d_fwi	Fire weather index anomaly	5.1 ± 8.8	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
d_isi	Initial spread index anomaly	1.4 ± 3.2	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
d_ffmc	Fine fuel moisture code anomaly	7.8 ± 15.7	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
d_dmc	Duff moisture code anomaly	16.4 ± 22.3	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
d_dc	Drought code anomaly	83.7 ± 70.5	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
d_bui	Buildup index anomaly	21.6 ± 25.5	CFSDS <sup>25</sup> , ERA5-Land <sup>36</sup>
d_prec	Precipitation anomaly (mm)	-1.0 ± 3.2	ERA5-Land <sup>36</sup>
d_vpd	Daily max vapour pressure deficit anomaly (Pa)	6.5 ± 4.7	ERA5-Land <sup>36</sup>
d_tmax	Maximum daily temperature anomaly (°C)	6.0 ± 4.5	ERA5-Land <sup>36</sup>
d_ws	Noon windspeed anomaly (km/h)	2.2 ± 4.2	ERA5-Land <sup>36</sup>
d_rh	Noon relative humidity anomaly (%)	6.5 ± 21.8	ERA5-Land <sup>36</sup>
Biomass	Total live aboveground biomass (t/ha)	53 ± 46	SCANFI <sup>39</sup>
Closure	Crown closure (%)	41 ± 16	SCANFI <sup>39</sup>
prcD	Deciduous component (%)	11 ± 13	SCANFI <sup>39</sup>
prcC	Conifer component (%)	89 ± 13	SCANFI <sup>39</sup>
peatCl	Peatland class	NA	Thompson <i>et al.</i> <sup>41</sup>
peatland	Modeled probability of peatland presence (0–100)	29 ± 19	Thompson <i>et al.</i> <sup>41</sup>
hydrodens2k	Proportion surface water within 2 km window	0.09 ± 0.12	CanVec Hydro <sup>40</sup>
hydrodens5k	Proportion surface water within 5 km window	0.11 ± 0.12	CanVec Hydro <sup>40</sup>
hydrodens10k	Proportion surface water within 10 km window	0.12 ± 0.12	CanVec Hydro <sup>40</sup>
hydrodens25k	Proportion surface water within 25 km window	0.13 ± 0.12	CanVec Hydro <sup>40</sup>
roaddens2k	Road density within 2 km (km/km <sup>2</sup> )	0.03 ± 0.12	CanVec Transport <sup>43</sup>
roaddens5k	Road density within 5 km (km/km <sup>2</sup> )	0.03 ± 0.10	CanVec Transport <sup>43</sup>
roaddens10k	Road density within 10 km (km/km <sup>2</sup> )	0.03 ± 0.10	CanVec Transport <sup>43</sup>
roaddens25k	Road density within 25 km (km/km <sup>2</sup> )	0.03 ± 0.09	CanVec Transport <sup>43</sup>
roaddist	Distance to major road or railway line (m)	39000 ± 39180	CanVec Transport <sup>43</sup>
dem	Elevation (m)	557 ± 384	ASTER Global Digital Elevation <sup>40</sup>
slope	Slope (degrees)	5.0 ± 6.4	ASTER Global Digital Elevation <sup>40</sup>
twi	Topographic wetness index	5.6 ± 0.5	ASTER Global Digital Elevation <sup>40</sup>
aspect	Aspect (degrees, 0–359)	181 ± 103	ASTER Global Digital Elevation <sup>40</sup>
cumuarea	Cumulative fire area (ha, NBAC <sup>26</sup> )	14450 ± 39510	CFSDS <sup>25</sup>
pctgrowth	Percent growth this day (%) (NBAC <sup>26</sup> )	113 ± 1950	CFSDS <sup>25</sup>
prevgrow	Fire growth on prior fireday (ha)	617 ± 2463	CFSDS <sup>25</sup>
spredistm	Daily fire spread (m)	453 ± 1063	CFSDS <sup>25</sup>
lat	Latitude (m) (EPSG 4269)	-112 ± 16	CFSDS <sup>25</sup>
lon	Longitude (m) (EPSG 4269)	58 ± 5	CFSDS <sup>25</sup>

**Table 3.** Environmental covariates and other associated fields.



**Fig. 3** Estimated fire arrival time for the 2018 Nadina Lake Fire in British Columbia. Fire arrival estimated from the agency polygons (a), FIREID (b), and CFSDS (c).

using area in  $\text{m}^2$ . This calculation is an approximation of frontal fire spread, and not true linear spread<sup>44</sup>. This is an approximation of spread distance that has lower precision than the underlying hotspot data, and this value is not rounded based on the underlying resolution of hotspot data.

**Ecozones.** Ecozones, which are large geographic areas sharing biophysical characteristics, were identified from the National Ecological Framework for Canada<sup>45</sup> (Table 2). The Boreal Shield and Taiga Plains ecozones have been split into “East” and “West” subzones<sup>46</sup>.

### Data Records

The complete CFSDS<sup>25</sup> for 2002–2021 is provided open access at the Centre for Open Science OSF data repository (<https://doi.org/10.17605/OSF.IO/F48RY>). This includes two datasets for quantitative analysis and one collection of fire arrival time interpolations:

*Firegrowth\_pts\_v1\_YEAR.csv* – a data table where each row indicates one pixel (180 m) that burned on a specific day. Each row representing a pixel is attributed with the 50 covariates.

*Firegrowth\_groups\_v1.csv* – a summarised data table where each row indicates a single fire growth day. Environmental covariates are summarised over each specific day of a fire by taking the mean value of continuous covariates and the mode of categorical covariates. Latitude and longitude values represent the centroid of fire spread on each specific day.

*YEAR\_NFIREID\_krig.tif* – raster dataset depicting fire arrival time for a single fire, with values indicating the day of year at which the fire arrived in any given 180-m pixel. For example, a pixel burned on February 10 would have an integer value of 41. NFIREID and YEAR are replaced with attributes as identified in NBAC<sup>26</sup>. Rasters are grouped by year for download.

In the *Firegrowth\_pts* and *Firegrowth\_groups* datasets, data columns indicate the covariates under which the fire burned, by pixel and fire growth day respectively (Fig. 1). These variables are listed in Table 3. These datasets are intended for use in model building and statistical analysis.

Province	FIRED vs Agency	CFSDS vs Agency	n
Alberta	0.46 ± 0.33	0.73 ± 0.32	98
British Columbia	0.49 ± 0.32	0.72 ± 0.32	442
Manitoba	0.42 ± 0.29	0.76 ± 0.26	73
Quebec	0.40 ± 0.33	0.77 ± 0.26	85
Saskatchewan	0.47 ± 0.35	0.72 ± 0.31	23
Average	0.47 ± 0.32	0.73 ± 0.31	721

**Table 4.** Provincial summary of Sørensen–Dice index of agency polygons compared against FIRED<sup>16</sup> and CFSDS<sup>25</sup> perimeters. Each sample represents a single day of fire spread for which an agency polygon is available. The reported average is weighted by growth day.

## Technical Validation

Validating day-of-burning rasters is difficult because scant “ground truth” validation data is available, making it hard to assess with confidence the accuracy of our day-of-burn rasters. The best data available are fire progression perimeters from wildfire agencies (“agency perimeters”); however, the regular production of agency polygons is limited by resource availability and they are known to slightly underpredict fire arrival time<sup>6</sup>. Therefore, instead of quantifying absolute error, we validated CFSDS<sup>25</sup> perimeters by assessing the degree to which they were an improvement over daily FIRED<sup>16</sup> progressions for all of Canada, when compared against agency perimeters (Fig. 3). FIRED<sup>16</sup> is the most recent and precise automated fire arrival time dataset with global coverage, compared against FRY<sup>10</sup>, Global Fire Atlas<sup>14</sup> or GlobFire<sup>15</sup>. Where all three datasets (FIRED, CFSDS<sup>25</sup>, and agency perimeters<sup>16</sup>) were available we assessed CFSDS and FIRED against the agency perimeters using two methods: daily perimeter similarity using Sørensen–Dice index and daily area-burned correlation. Wildfire agency perimeters were provided by the provinces of British Columbia (84 fires, n = 497), Alberta (18 fires, n = 96), Saskatchewan (13 fires, n = 31), Manitoba (17 fires, n = 59), and Quebec (12 fires, n = 58).

**Preprocessing.** The three datasets were substantially different and it was therefore necessary to subset them for a fair comparison. First and most importantly, the agency perimeters vary substantially in quality, including high-quality sources such as infrared aerial platforms and medium- or low-quality sources such as aerial GPS observations of fire perimeters. We only used agency perimeters with higher spatial and temporal precision as a validation dataset, defined as perimeters derived from: aerial photography; GPS aerial observations; processed infrared aerial imagery; Sentinel-2 satellite observations; and Landsat satellite observations. Both CFSDS<sup>25</sup> and FIRED<sup>16</sup> were subset to include only perimeters from dates which had representative observations in the agency polygons.

Second, CFSDS<sup>25</sup> is unfairly advantaged against FIRED<sup>16</sup> because CFSDS is constrained with an *a priori* fire perimeter, whereas FIRED is not. CFSDS<sup>25</sup> often matches the final agency perimeter very closely due to similar data sources being used for fire delineation in NBAC<sup>26</sup>; therefore, we excluded the final perimeter as a data point. Instead all validation stats were calculated on fire progression polygons from the period while the fire was still expanding and we compared area burned as a percentage of final perimeter size instead of absolute area burned. Using percentage area burned prevents FIRED<sup>16</sup> from being penalised by a high false-negative rate due to the lack of an *a priori* perimeter. Finally, to further reconcile the three datasets, we introduced unburned islands to FIRED<sup>16</sup> and the agency perimeters by masking them with the fire NBAC<sup>26</sup> polygon, as was done for CFSDS<sup>25</sup>.

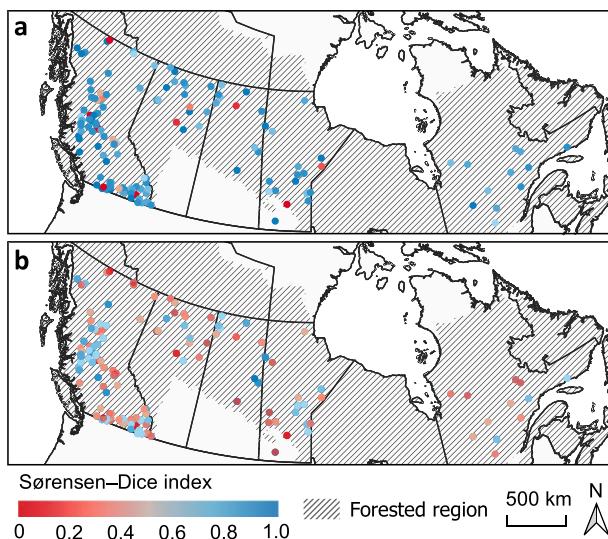
**Perimeter similarity using Sørensen–Dice index.** The Sørensen–Dice index is broadly used in image segmentation and more broadly to examine classification problems. It is equal to twice the area of overlap between perimeters divided by the total number of pixels in both perimeters (Eq. 2).

$$\text{DSC} = \frac{2TP}{2TP + FP + FN} \quad (2)$$

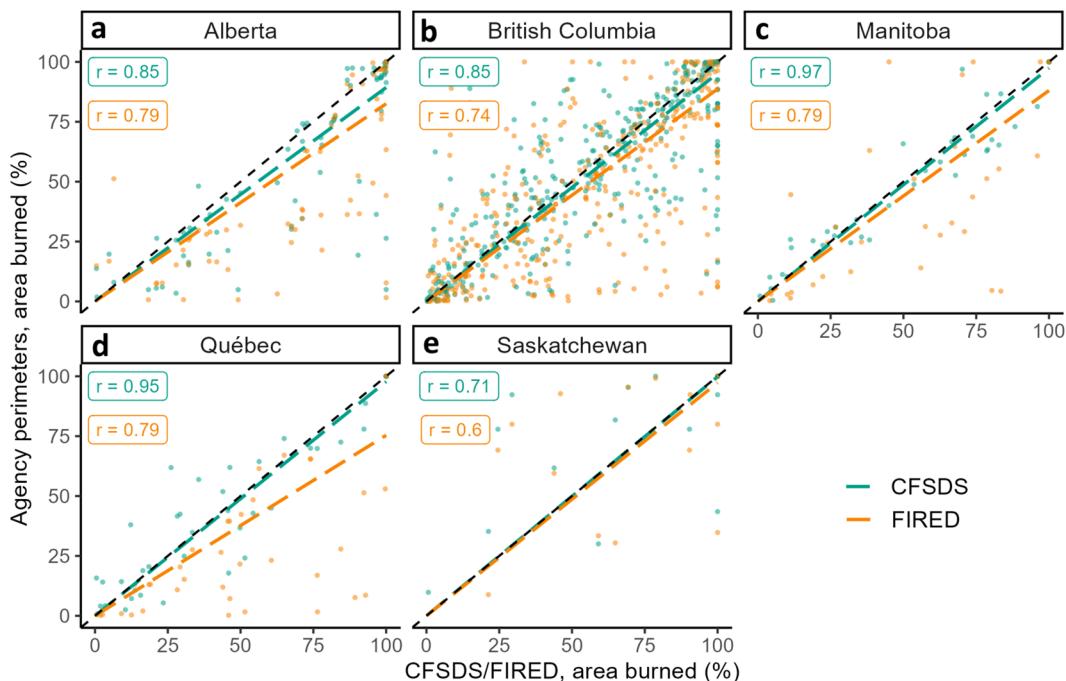
Where TP is true positive, FP is false positive, and FN is false negative. The index varies from 0 to 1, with 0 indicating no spatial alignment and 1 indicating perfect spatial alignment. We assessed perimeter similarity by calculating the Sørensen–Dice index between CFSDS-Agency and FIRED-Agency pairs for the same date. We used this index as an indicator of mean absolute improvement of CFSDS<sup>25</sup> over FIRED<sup>16</sup>.

Validation by Sørensen–Dice index indicates that CFSDS<sup>25</sup> agrees more closely with agency polygons than FIRED. Averaged across all provinces, CFSDS<sup>25</sup> has an index of  $0.73 \pm 0.31$ , while FIRED<sup>16</sup> has an index of  $0.47 \pm 0.32$  (Table 4, Fig. 4). This is likely due to the lower spatial resolution of FIRED<sup>16</sup> and the temporal lag commonly associated with MODIS burned area products<sup>12</sup>.

**Daily burned area correlation.** Validation by comparing daily area burned is an established method for intercomparison of remotely sensed fire progressions against agency polygons<sup>47</sup>. We calculated the Spearman correlation between each dataset’s prediction of daily area burned for individual fires, showing that the CFSDS<sup>25</sup> consistently outperforms FIRED (Fig. 5), with CFSDB having a Spearman’s  $\rho$  0.06–0.28 higher than FIRED. On a daily basis, both datasets tended to underpredict total area burned when compared against the agency perimeters, although this could also be attributed to agency perimeters overpredicting daily area burned. While MODIS burned area products<sup>12</sup>, such as FIRED<sup>16</sup>, have been shown to be poorly suited for burned area assessments of smaller ( $<100$  ha) fires<sup>47</sup>, our fire size threshold of 1,000 ha minimised this bias. It is notable that even with this advantage to FIRED<sup>16</sup>, CFSDS<sup>25</sup> was the strongest performer.



**Fig. 4** Sørensen–Dice index of agency polygons compared against CFSDS (a) and FIRED (b) polygons. Each point represents the mean index of a single fire, averaged across all days of burning.



**Fig. 5** Daily area-burned validation by province. Agency polygons from the provincial wildfire agencies of Alberta (a), British Columbia (b), Manitoba (c), Quebec (d), and Saskatchewan (e) provide the validation dataset to compare against the CFSDS or FIRED. Points represent a single agency polygon burning day compared against FIRED or CFSDS using percentage burned of the final perimeter. Dashed black line represents 1:1 fit, or perfect agreement with agency polygons. Dashed, coloured lines represent a linear regression fit of all points. Spearman correlation coefficients are indicated with inset boxes.

### Usage Notes

Users interested in examining the spread of a specific fire or set of fires should identify the fire IDs of these fires using NBAC<sup>26</sup>, available at <https://cwfis.cfs.nrcan.gc.ca/datamart/metadata/nbac>. Individual fire events were assigned a unique ID by the concatenation of the ‘YEAR’ and ‘NFIREID’ fields from NBAC<sup>26</sup>. For example, a fire from 2007 with an NFIREID of 346 would have an ID of “2007\_346” and a corresponding interpolation filename of “2007\_346\_krig.tif”.

CFSDS v1.0<sup>25</sup> will be expanded annually with corresponding covariates as better data become available, and will be made available at the same data repository (<https://osf.io/f48ry/>) (<https://doi.org/10.17605/OSF.IO/F48RY>). This data descriptor was peer reviewed in 2024 on v1.0 of the dataset, and based on the data available at the time.

## Code availability

R code used for producing CFSDS<sup>25</sup> has been made available at the Centre for Open Science OSF data repository at <https://osf.io/f48ry/> (<https://doi.org/10.17605/OSF.IO/F48RY>). Four code sections are provided, grouped under one code file “CFSDS\_example.R”. Code subsections: 1. ‘1.0 Processing’ - Preprocesses a single NBAC perimeter and FIRMS hotspots<sup>9</sup>. 2. ‘2.0 Interpolation’ - Interpolates fire arrival time for a single NBAC perimeter using kriging. 3. ‘3.0 Covariates - Uses the fire arrival time raster to extract environmental covariates associated with each burning day. 4. ‘4.0 Summarize’ - an example of how an end-user might summarize point-level spread data for use in fire modelling.

Received: 30 November 2023; Accepted: 28 May 2024;

Published online: 11 July 2024

## References

- Crowley, M. A. *et al.* Towards a whole-system framework for wildfire monitoring using Earth observations. *Glob. Chan. Biol.* **29**(6), 1423–1436 (2023).
- Giglio, L., Descloitres, J., Justice, C. O. & Kaufman, Y. J. An enhanced contextual fire detection algorithm for MODIS. *Remote Sens. Environ.* **87**(2–3), 273–282 (2003).
- Schroeder, W., Oliva, P., Giglio, L. & Csiszar, I. A. The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sens. Environ.* **143**, 85–96 (2014).
- de Groot, W. J. *et al.* Estimating direct carbon emissions from Canadian wildland fires. *Int. J. Wildland Fire* **16**, 593–606, <https://doi.org/10.1071/WF06150> (2007).
- Veraverbeke, S. *et al.* Mapping the daily progression of large wildland fires using MODIS active fire data. *Int. J. Wildland Fire* **23**(5), 655–667 (2014).
- Parks, S. A. Mapping day-of-burning with coarse-resolution satellite fire-detection data. *Int. J. Wildland Fire* **23**(2), 215–223 (2014).
- Chen, Y. *et al.* California wildfire spread derived using VIIRS satellite observations and an object-based tracking system. *Sci. Data* **9**(1), 249 (2022).
- Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
- NASA. Fire Information for Resource Management System, NASA. <https://firms.modaps.eosdis.nasa.gov/>. Accessed August 11, 2022.
- Laurent, P. *et al.* FRY, a global database of fire patch functional traits derived from space-borne burned area products. *Sci. Data* **5**(1), 1–12 (2018).
- Giglio, L., Schroeder, W. & Justice, C. O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **178**, 31–41 (2016).
- Giglio, L., Boschetto, L., Roy, D. P., Humber, M. L. & Justice, C. O. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sens. Environ.* **217**, 72–85 (2018).
- Chuvieco, E. *et al.* A new global burned area product for climate assessment of fire impacts. *Global Ecol. Biogeogr.* **25**(5), 619–629 (2016).
- Andela, N., Morton, D. C., Giglio, L. & Randerson, J. T. Global fire atlas with characteristics of individual fires, 2003–2016. ORNL DAAC (2019).
- Artés, T. *et al.* A global wildfire dataset for the analysis of fire regimes and fire behaviour. *Sci. Data* **6**(1), 296 (2019).
- Balch, J. K. *et al.* FIRED (Fire Events Delineation): an open, flexible algorithm and database of US fire events derived from the MODIS burned area product (2001–2019). *Remote Sens.* **12**(21), 3498 (2020).
- Chuvieco, E. *et al.* ESA Fire Climate Change Initiative (Fire\_cci): MODIS Fire\_cci Burned Area Pixel product, version 5.0. Centre for Environmental Data Analysis. <https://doi.org/10.5285/9c666602b89e468493e1c907a4de62ff> (2018).
- Lizundia-Loiola, J., Otón, G., Ramo, R. & Chuvieco, E. A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. *Remote Sens. Environ.* **236**, 111493 (2020).
- Noonan-Wright, E. K. *et al.* Developing the US wildland fire decision support system. *J. Combust.* **2011** (2011).
- Briones-Herrera, C. I. *et al.* Near real-time automated early mapping of the perimeter of large forest fires from the aggregation of VIIRS and MODIS active fires in Mexico. *Remote Sens.* **12**(12), 2061 (2020).
- Scaduto, E., Chen, B. & Jin, Y. Satellite-based fire progression mapping: A comprehensive assessment for large fires in northern California. *IEEE J-STARS* **13**, 5102–5114 (2020).
- Liu, T. *et al.* Systematically tracking the hourly progression of large wildfires using GOES satellite observations. *Earth Syst. Sci. Data* [in press], <https://doi.org/10.5194/essd-2023-389> (2024).
- Wang, X. *et al.* The potential and realized spread of wildfires across Canada. *Glob. Chan. Biol.* **20**(8), 2518–2530 (2014).
- Wang, X. *et al.* Critical fire weather conditions during active fire spread days in Canada. *Sci. Total Environ.* **869**, 161831 (2023).
- Barber, Q. E. *et al.* The Canadian Fire Spread Dataset. OSF <https://doi.org/10.17605/OSF.IO/F48RY> (2024).
- Skakun, R. *et al.* Extending the National burned area Composite Time Series of Wildfires in Canada. *Remote Sens.* **14**(13), 3050 (2022).
- Hall, R. J. *et al.* Generating annual estimates of forest fire disturbance in Canada: The National burned area Composite. *Int. J. Wildland Fire* **29**(10), 878–891, <https://doi.org/10.1071/WF19201> (2020).
- Natural Resources Canada. Canadian National Fire Database [online]. Available from <http://cwfis.cfs.nrcan.gc.ca/datamart> [accessed 2 September 2022].
- Fraser, R. H., Li, Z. & Cihlar, J. Hotspot and NDVI differencing synergy (HANDS): A new technique for burned area mapping over boreal forest. *Remote Sens. Environ.* **74**(3), 362–376 (2000).
- NASA. MODIS Collection 6 Hotspot / Active Fire Detections MCD14ML distributed from NASA FIRMS. Available on-line <https://earthdata.nasa.gov/firms>. <https://doi.org/10.5067/FIRMS/MODIS/MCD14ML> (2023).
- NASA. NRT VIIRS 375 m Active Fire product VNP14IMGMT distributed from NASA FIRMS. Available on-line <https://earthdata.nasa.gov/firms>. [https://doi.org/10.5067/FIRMS/VIIRS/VNP14IMGMT\\_NRT.002](https://doi.org/10.5067/FIRMS/VIIRS/VNP14IMGMT_NRT.002) (2023).
- NASA. NRT VIIRS 375 m Active Fire product VJ114IMGTDL\_NRT distributed from NASA FIRMS. Available on-line <https://earthdata.nasa.gov/firms>. [https://doi.org/10.5067/FIRMS/VIIRS/VJ114IMGT\\_NRT.002](https://doi.org/10.5067/FIRMS/VIIRS/VJ114IMGT_NRT.002) (2023).
- Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* **30**, 683–691 (2004).
- Beck, J. A., Alexander, M. E., Harvey, S. D. & Beaver, A. K. Forecasting diurnal variations in fire intensity to enhance wildland firefighter safety. *Int. J. Wildland Fire* **11**(4), 173–182 (2002).
- Li, F. *et al.* Hourly biomass burning emissions product from blended geostationary and polar-orbiting satellites for air quality forecasting applications. *Remote Sens. Environ.* **281**, 113237 (2022).
- Muñoz Sabater, J. ERA5-Land hourly data from 1981 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.e2161bac> (2019).

37. Van Wagner, C. Development and Structure of the Canadian Forest Fire Weather Index System Forestry Technical Report 35. Canadian Forestry Service (1987).
38. McElhinny, M., Beckers, J. F., Hanes, C., Flannigan, M. & Jain, P. A high-resolution reanalysis of global fire weather from 1979 to 2018—overwintering the Drought Code. *Earth Syst. Sci. Data* **12**(3), 1823–1833 (2020).
39. Guindon, L. *et al.* A new approach for Spatializing the Canadian National Forest Inventory (SCANFI) using Landsat dense time series. *Can. J. Forest Res.* <https://doi.org/10.1139/cjfr-2023-0118> (2014).
40. NASA/METI/AIST/Japan Spaceystems and U.S./Japan ASTER Science Team, 2019. ASTER Global Digital Elevation Model V003 [Data set]. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/ASTER/ASTGTM.003>. Accessed June 10, 2021.
41. Thompson, D. K., Simpson, B. N. & Beaudoin, A. Using forest structure to predict the distribution of treed boreal peatlands in Canada. *Forest Ecol. Manag.* **372**, 19–27 (2016).
42. Natural Resources Canada. Lakes, Rivers & Glaciers in Canada - CanVec Series - Hydrographic Features. Available from open.canada.ca/data/en/dataset/2dac78ba-8543-48a6-8f07-faeef56f9895 [accessed May 20, 2021] (2019).
43. Natural Resources Canada. Transport Networks in Canada — CanVec Series — Transport Features. Available from open.canada.ca/data/en/dataset/2dac78ba-8543-48a6-8f07-faeef56f9895 [accessed May 20, 2021] (2019).
44. Parisien, M. A. *et al.* Considerations for modeling burn probability across landscapes with steep environmental gradients: an example from the Columbia Mountains. *Canada. Nat. Hazards* **66**, 439–462 (2013).
45. Ecological Stratification Working Group. A National Ecological Framework for Canada. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research, and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, Ottawa/Hull. Report and national map at 1:7,500,000 scale (1996).
46. Stocks, B. J. *et al.* Large forest fires in Canada, 1959–1997. *J. Geophys. Res.-Atmos.* **107**(D1), pp. FFR-5 (2002).
47. Galizia, L. F., Curt, T., Barbero, R. & Rodrigues, M. Assessing the accuracy of remotely sensed fire datasets across the southwestern Mediterranean Basin. *Nat. Hazard Earth Sys.* **21**(1), 73–86 (2021).
48. M. Hansen, M. & Song, X. P. MEaSUREs Vegetation Continuous Fields (VCF) Yearly Global 0.05 Deg V001. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MEaSUREs/VCF/VCF5KYR.001> (2017).

## Acknowledgements

We would like to thank Morgan Crowley (Canadian Forest Service) and Mark de Jong (Canadian Forest Service) for assisting with algorithm development and hotspot validation. Jonathan Boucher (Canadian Forest Service), Dave Schroeder (Government of Alberta), and Amber Sigurdson (Government of Alberta) provided fire perimeters for interpolation validation. Philippe Villemaire (Canadian Forest Service) and Francis Manka (Canadian Forest Service) assisted with SCANFI forest fuels data processing. This research was supported in part by the USDA Forest Service, Rocky Mountain Research Station, Aldo Leopold Wilderness Research Institute. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

## Author contributions

Q.E.B.: Conceptualization, Methodology, Software, Validation, Data Curation, Writing - Original Draft, Visualization. P.J.: Conceptualization, Methodology, Validation, Visualization, Writing - Review and Editing. E.W.: Conceptualization, Methodology, Validation, Writing - Review and Editing. D.K.T.: Conceptualization, Methodology, Supervision, Writing - Review and Editing. L.G.: Resources, Writing - Review and Editing. S.A.P.: Conceptualization, Methodology, Writing - Review and Editing. X.W.: Software, Writing - Review and Editing. M.H.: Validation, Writing - Review and Editing. M.-A.P.: Conceptualization, Methodology, Supervision, Writing - Review and Editing

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Q.E.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© His Majesty the King in Right of Canada as represented by the Minister of Natural Resources. Parts of this work were authored by US Federal Government authors and are not under copyright protection in the US; foreign copyright protection may apply 2024