remote sensing

MDPI

Article

Wildfire Time-of-Arrival Reconstruction from Satellite Active-Fire Detections: External Evaluation of Variance-Weighted Kriging and Application to South Korea

Chulhyun Choi [1], Hyunjin Seo [2] and Jaegyu Cha [3,*]

[1] Safety Management Department, National Institute of Ecology, 1210 Geumgang-ro, Maseo-myeon, Seocheon 33657, Republic of Korea
[2] ESG Management Department, National Institute of Ecology, 1210 Geumgang-ro, Maseo-myeon, Seocheon 33657, Republic of Korea
[3] Climate Change Adaptation Team, National Institute of Ecology, 1210 Geumgang-ro, Maseo-myeon, Seocheon 33657, Republic of Korea
* Correspondence: flow@nie.re.kr

Highlights

What are the main findings?

- With all available observations, differences in time-of-arrival (ToA) reconstruction among interpolation configurations were small; in polar-orbiting observation gaps, variance-weighted kriging that also used geostationary data reconstructed ToA best by median.
- Applying the selected method to 21 South Korean large wildfires of 100 ha or more, we built a 375 m arrival-time and standard-deviation dataset for 15 events together with event-level evaluation information.

What are the implications of the main findings?

- The dataset can support pixel-level validation of spread models and analysis of spread drivers in regions where independent validation is difficult for want of aerial or field observations.

Abstract

Large wildfires cause casualties, ecosystem damage, and property loss. Reducing them requires predicting spread; validating a spread model requires pixel-level time of arrival (ToA) from real fires. Few regions have both such data and an independent reference. We combine polar-orbiting (VIIRS) and geostationary (GOES, Himawari) active-fire detections by kriging to produce ToA inside a known final perimeter with a per-pixel kriging standard deviation. For 29 U.S. wildfires with aerial infrared perimeters, we compared configurations differing in whether geostationary data were added and whether pixel size entered the variance, against an inverse-distance-weighting baseline. With all observations available, differences among configurations were small. In polar-orbiting gaps, configurations using geostationary data reconstructed ToA better, and variance-weighted kriging was best by median. Weighing reconstruction performance and the per-pixel standard deviation, we selected variance-weighted kriging. Applying it to 21 South Korean wildfires $\geq$ 100 ha, we built a 375 m arrival-time and standard-deviation dataset for 15 events with event-level evaluation information. This is an application, not an independent accuracy validation, and reconstruction stays inside the final perimeter. Even so, the dataset supports pixel-level validation of spread models and analysis of spread drivers where aerial or field observations are unavailable.

Check for updates

Academic Editors: Xingwen Quan and Binbin He

Received: 4 August 2026
Revised: 15 September 2026
Accepted: 21 September 2026
Published: 26 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: wildfire spread; VIIRS; GOES; Himawari; geostationary satellite; measurement-error kriging; spatial block cross-validation; NIROPS

1. Introduction

1.1. Background

Wildfires cause loss of life, ecosystem damage, and property loss [1]. Reducing that damage requires predicting spread, and validating a prediction model requires knowing how actual wildfires spread. The final shape of a wildfire perimeter alone does not reveal in which direction and how fast combustion progressed. Pixel-level time of arrival (ToA) reconstructs the spread front over time within the same final perimeter. From these data, spread rate and direction can be computed. They can also serve as basic data for analyzing the response to meteorological, topographic, and fuel conditions. Moreover, observed ToA allows the arrival order and timing predicted by wildfire-spread models to be compared spatially, making it an evaluation target that provides more information than a single final area [2–5].

South Korea's wildfire statistics and final damage perimeters are useful for identifying the size and location of events, but they do not directly provide continuous pixel-level ToA. In recent years, the frequency and damage of large wildfires have increased in South Korea, and the combined role of climatic, forest, and wildland–urban interface factors has been emphasized [6–8]. A quantitative understanding of spread drivers is the basis for prevention that curbs the escalation of wildfires and for early-response strategies. However, public data that provide the event-level spread process, the starting point of that understanding, on a consistent grid and with quality information are limited. Under conditions where field or aerial observation cannot be applied repeatedly to every wildfire, satellite active-fire detection is a realistic input that can reconstruct the progression of many events with the same observation system.

1.2. Satellite-Based Reconstruction of Wildfire Progression

National-, continental-, and global-scale wildfire datasets provide event-level size and duration, spread direction, and temporal bounds, broadening the basis for comparative studies [9–12]. However, because satellites differ in orbit, spatial resolution, cloud and smoke, view angle, and detection threshold, the detection time itself cannot readily be taken as the actual arrival time.

A first approach spatially interpolates observed active-fire times to construct a continuous ToA surface. Reconstruction of daily or per-survey progression perimeters [13,14], construction of a Canadian wildfire-spread dataset [15], and kriging-based reconstruction of the growth of large boreal wildfires [16] show that the interpolation family can process large events with transparent rules. This family makes the relationship between inputs and outputs easy to trace. However, when observation gaps are wide, the interpolated surface depends on a few points. Treating sensors of different resolution with the same reliability can also produce local distortion.

A second approach combines machine learning or physical models with satellite data. Proposed methods include machine-learning estimation of arrival time from active fires [17], a generative model that fuses a wildfire-spread model with VIIRS observations [18], a reconstruction using multimodal active fires and terrain [19], deep learning that downscales geostationary imagery to high resolution or refines perimeters using VIIRS as the teacher [20,21], and a learning-based method that reconstructs savanna wildfire events [22]. However, the results of this family depend on the distribution of the wild-

fire cases used for training. Applying them to wildfires that differ in region, scale, or observation conditions requires retraining or separate validation data.

Polar-orbiting and geostationary sensors have complementary observing characteristics. VIIRS provides relatively fine spatial location through its 375 m active-fire product but observes the same location a limited number of times. GOES and Himawari observe at short time intervals but have larger pixels, and their detection location and time can be mixed. In South Korea as well, Himawari-based fire detection [23], GK2A-based real-time AI monitoring [24], and monitoring of spread extent combining GK2A and VIIRS [25] have been studied. However, data are scarce that separately evaluate, at the event level and against external aerial perimeters, the inclusion of the two sensor families and the weighting for resolution differences.

The contribution of this study lies not in the sensor combination itself but in the separate evaluation of design decisions. We evaluate separately the roles of geostationary inclusion and empirical variance weighting through external-reference evaluation and spatial block cross-validation. In the application to South Korea, we adopt a quality-screening structure that distributes together the producibility, a diagnosis of the spatial self-predictability of the input data, and the reasons for evaluation withheld and exclusion from production.

1.3. Study Objectives

The objectives of this study are to quantitatively evaluate how combining polar-orbiting and geostationary active-fire detections affects the performance of wildfire time-of-arrival reconstruction, and to propose a quality-diagnosis-based ToA production procedure applicable to regions without external reference data. Specifically, first, we evaluate how the inclusion of geostationary detections and empirical variance weighting affects the reconstruction of fire progression inside the final perimeter and the prediction performance for polar-orbiting observation gaps. Second, we analyze the association between the spatial block cross-validation metric and the external-reference evaluation results to examine its potential as an internal quality indicator. Third, we apply the selected procedure to South Korean wildfires of 100 ha or larger to build ToA and kriging standard-deviation rasters, and present a quality-screened dataset that provides production feasibility and internal quality-diagnosis results together.

2. Materials and Methods

2.1. Study Workflow

The overall procedure consists of four parts: input-data construction, external-reference evaluation in the United States, method selection, and application to Korea (Figure 1). The first test in the U.S. stage is an external-reference evaluation that produces per-configuration ToA rasters from the same active-fire detection inputs and computes their accuracy against the NIROPS perimeters. The second test is an operational-condition spatial block cross-validation that sequentially holds out blocks occupied by polar-orbiting observations while retaining the geostationary inputs. In both tests, a control (PS-IDW) was also compared in order to isolate the effect of adding data. Combining the results of the two tests, the default method was selected, and the per-fire r from the same block cross-validation was compared against the external accuracy and then used as the Korean reliability metric.

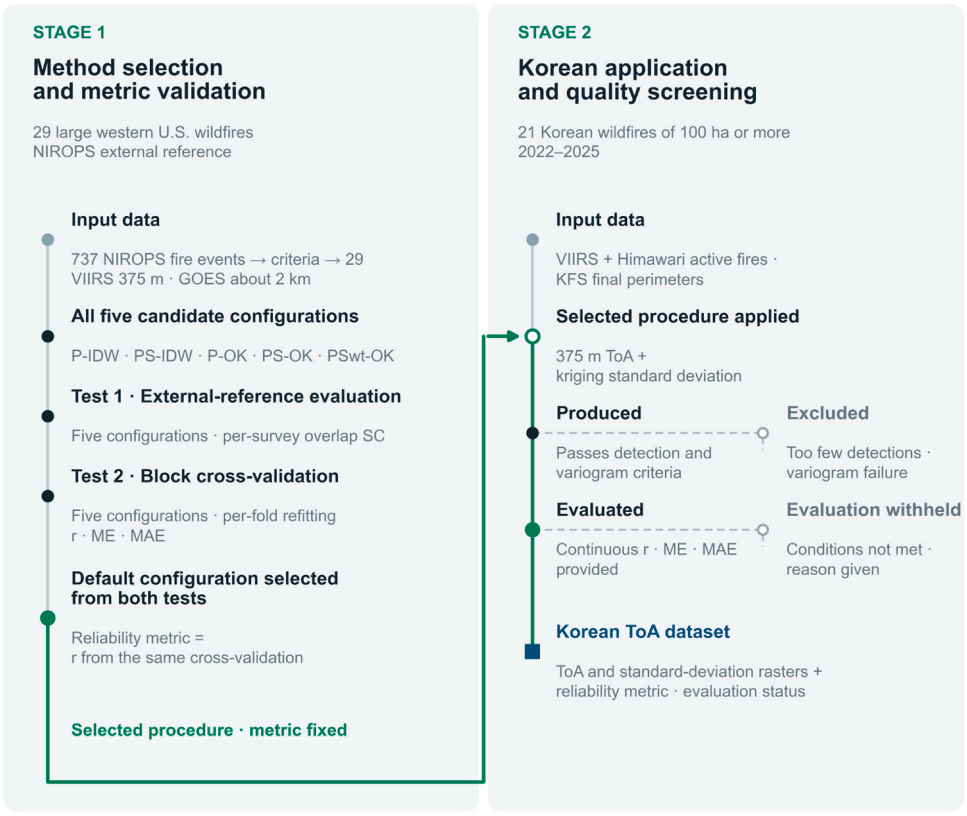


Figure 1. Overall study workflow: in Stage 1 (29 U.S. fires), five configurations including the control (PS-IDW) were evaluated by external-reference evaluation (Test 1) and spatial block cross-validation (Test 2), and the default configuration was selected. The per-fire reliability metric was defined as the r from the same cross-validation. In Stage 2 (21 Korean fires), the fixed procedure was applied and cases were classified as evaluated, evaluation withheld, or excluded from production according to the production criteria and whether cross-validation was valid. For evaluated products, r, ME, and MAE were reported.

2.2. Data

2.2.1. U.S. External-Reference Sample and Stepwise Reference ToA

The U.S. reference data are the NIROPS aerial infrared survey records from the high-resolution daily wildfire perimeter dataset for the western United States (2020–2024) released by Magstadt et al. [26,27]. Mendeley Data v1 contains 737 fire events and 12,705 unique survey records defined by event–UTC time [26]. After an in-house quality check of the source timestamps (Table S1), the times were used as an interval-censored reference that represents the cumulative area at each survey time rather than an absolute per-pixel time of arrival.

Each NIROPS survey perimeter was interpreted as the burned area accumulated up to that survey time. The polygons of each event were sorted chronologically and unioned with the preceding area to prevent non-monotonic shrinkage of the perimeter. The survey time at which the center of a 375 m grid cell was first enclosed was then recorded as that cell's reference ToA. This value is not a point observation of the actual moment of arrival. If a cell is outside the perimeter at survey $k-1$ and first enclosed at survey k, one can only know that the actual time of arrival lies between the two surveys; therefore the reference ToA is interval-censored data.

The sample was selected by three criteria. First, events with at least 25 growth-phase surveys, defined as surveys whose cumulative area is below 90% of the final area, were retained. This condition secures a stable per-event median of the per-survey overlap metric, which is the event score. Second, events whose first-survey area was at most 20% of the final

area were retained. This condition excludes events first surveyed after most of the spread had ended, and no event was actually excluded by it (Table S1). Third, the per-survey overlap evaluation requires a sufficiently large evaluation area. In this study, the lower bound was set to 1000 area-equivalent pixels of 375 m of the final perimeter area. Where the same incident number was recorded as multiple polygon sets, only the record with the largest area was kept. All criteria used only the attributes of the reference records. From the 737 source events, all 29 events that passed the three criteria were taken as the sample. The stepwise screening flow is given in Table S1. The number of growth-phase surveys for the 29 events was 26–59, and the final areas were 14,861.3–418,079.8 ha. The number of events by year was 7 in 2020, 13 in 2021, 3 in 2022, and 6 in 2024 (Figure 2a; Tables 1 and S2).

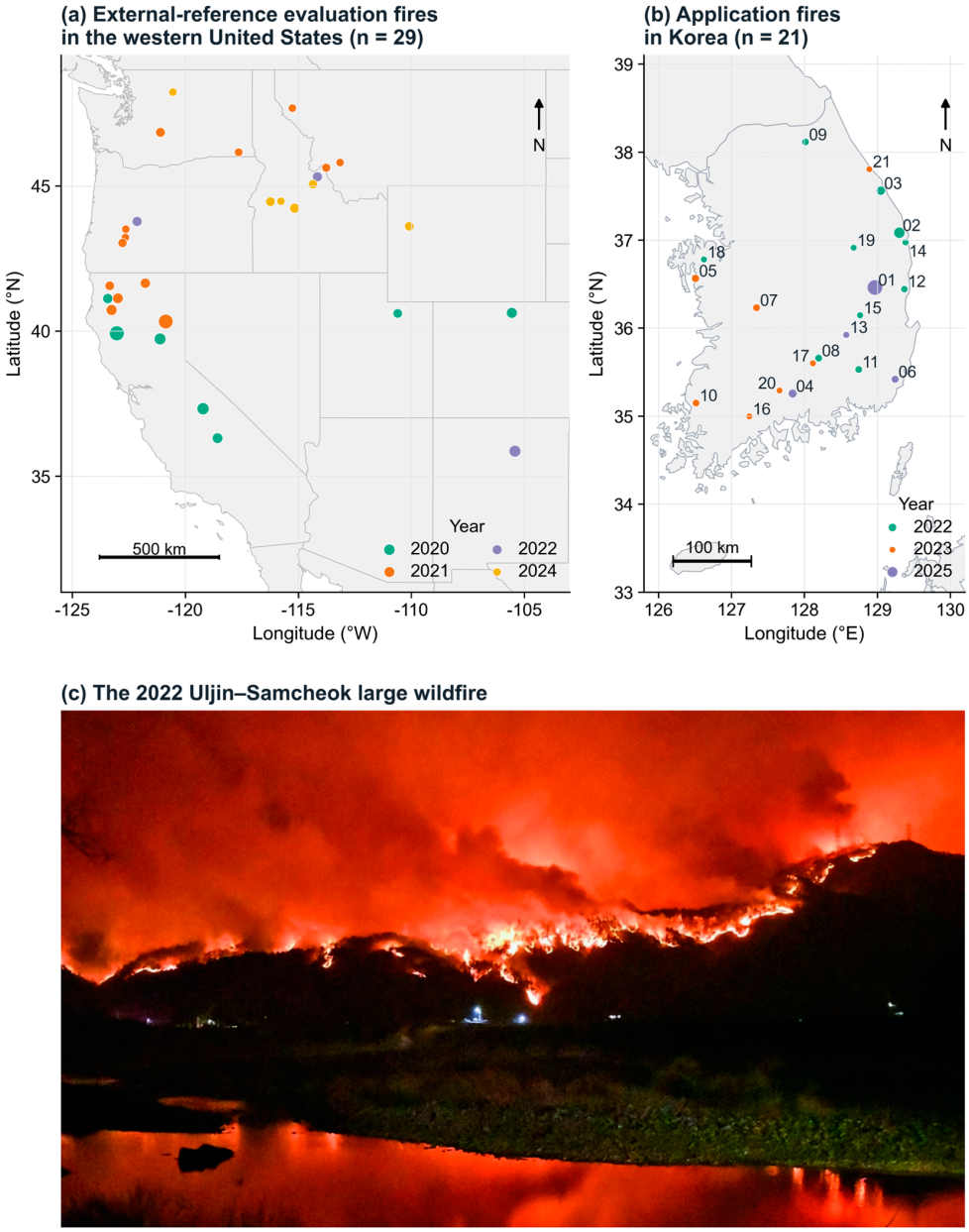


Figure 2. Locations of the study fires and a representative photograph of a large Korean wildfire. (**a**) The 29 external-reference evaluation fires in the western United States. (**b**) The 21 application fires in Korea. Point size is proportional to the square root of the final burned area. Color indicates the year of occurrence. The numbers beside the Korean points are the KFS event IDs. (**c**) The night-time fire line of the 2022 Uljin–Samcheok large wildfire. Photograph: National Institute of Forest Science (2025) [28], p. 30, Korea Open Government License Type 1.

Table 1. Characteristics of the 29 U.S. external-reference evaluation fires. Growth surveys are the number of surveys at the selection stage whose cumulative area is below 90% of the final; only surveys at or above 1% among these were used in the evaluation. The first-survey ratio is the cumulative area of the first valid survey divided by the final area. The VIIRS cell count is the value after 375 m grid deduplication summary. The geostationary pixel count is the number of GOES pixels used in the analysis.

No.	Fire (Event Number)	Year	Total Surveys	Growth Surveys	First-Survey Ratio	Final Area (ha)	VIIRS Cells	Geostationary Pixels
1	Red Salmon Complex (CA-SRF-000656)	2020	90	59	0.003	58,602.7	4267	174
2	Cameron Peak (CO-ARF-000636)	2020	68	59	0.026	84,544.3	5248	288
3	Trail Creek (MT-BDF-006272)	2021	63	51	0.015	25,095.7	1697	78
4	Jack (OR-UPF-000265)	2021	67	51	0.065	14,977.2	1042	65
5	Dixie (CA-BTU-009205)	2021	75	50	0.002	390,119.7	26,666	1252
6	Alder Creek (MT-BDF-006269)	2021	65	49	0.002	14,960.4	1089	55
7	Moose (ID-SCF-022105)	2022	70	46	0.094	52,692.0	3328	158
8	McCash (CA-SRF-000651)	2021	65	45	0.007	38,479.2	2854	128
9	Goat (ID-BOF-000904)	2024	52	44	0.010	14,861.3	1134	68
10	Green Ridge (OR-UMF-000659)	2021	73	43	0.003	17,698.6	1281	59
11	SQF Complex (CA-SQF-002622)	2020	75	43	0.036	72,840.6	5437	312
12	Hermits Peak (NM-SNF-000027)	2022	87	43	0.001	139,521.1	8228	504
13	Cedar Creek (OR-WIF-220180)	2022	64	42	0.008	51,521.6	3463	151
14	Pioneer (WA-SES-000173)	2024	63	41	0.051	15,736.4	1041	65
15	Monument (CA-SHF-001187)	2021	73	39	0.004	90,691.2	6556	266
16	Rough Patch Complex (OR-UMP-000373)	2021	51	38	0.014	15,507.3	1098	45
17	Red Rock (ID-SCF-240202)	2024	47	38	0.006	32,074.2	2218	114
18	River Complex (CA-KNF-006385)	2021	73	38	0.039	78,394.9	5842	257
19	Creek (CA-SNF-001391)	2020	59	36	0.120	154,151.0	11,119	503
20	Fish Creek (WY-BTF-002416)	2024	47	36	0.087	36,393.2	2591	145
21	Antelope (CA-KNF-006454)	2021	43	36	0.017	58,777.8	3118	186

Table 1. *Cont.*

No.	Fire (Event Number)	Year	Total Surveys	Growth Surveys	First-Survey Ratio	Final Area (ha)	VIIRS Cells	Geostationary Pixels
22	Wapiti (ID-BOF-000683)	2024	63	34	0.001	52,229.8	3654	186
23	Lava (ID-BOF-001080)	2024	58	34	0.006	39,651.3	2081	121
24	Devils Knob (OR-UPF-000450)	2021	49	32	0.003	28,092.9	1998	101
25	PNF North Complex (CA-PNF-001308)	2020	73	30	0.002	129,008.8	8239	394
26	August Complex (CA-MNF-000753)	2020	62	30	0.063	418,079.8	28,553	1180
27	Schneider Springs (WA-OWF-000453)	2021	42	28	0.003	43,469.7	3128	125
28	West Lolo (MT-LNF-000893)	2021	29	27	0.021	15,658.5	1060	65
29	East Fork (UT-ASF-100340)	2020	43	26	0.028	36,956.2	2222	163

2.2.2. Satellite Active-Fire Inputs

The polar-orbiting inputs for the United States are the Suomi NPP and NOAA-20 VIIRS Collection 2 active-fire detections provided by NASA FIRMS [29–32]. Each detection is provided as the pixel-center coordinates of a 375 m pixel observation together with the detection time. These detections are aggregated into 375 m grid cells at the analysis stage. A 2024 NOAA-21 product also exists [33], but it was not consistently available in the FIRMS archive over the entire analysis period (2020–2024) and was therefore excluded from the inputs. The 1 km MODIS detections were not included because the VIIRS 375 m product is the successor product with improved detection performance [30]. The collection extent for each event was the bounding box of the NIROPS final perimeter expanded by 0.05° in all directions. Because the widest extent of detections used in the analysis is the 2 km buffer of the final perimeter, a wider margin was applied so that detections near the perimeter would not be missed at the collection stage. The extent actually used is determined by the clipping rules (Section 2.3.1). The time window was set from 24 h before the first survey to 24 h after the last survey. This window is a margin to collect the initial spread before the first survey and the progression after the last survey.

The geostationary inputs for the United States are the detection times of the GOES-R Advanced Baseline Imager Fire/Hot Spot Characterization product [34]. GOES-16 and GOES-17 or GOES-18 were used according to the event year. The data were collected from Google Earth Engine.

The polar-orbiting inputs for Korea also used the same VIIRS series, and the geostationary inputs used the Wild Fire L2WLF product of Himawari-8/9 [35,36]. For the events in 2025, NOAA-21 VIIRS detections were also included [33]. For both the United States and Korea, duplicate detection rows with all attributes identical were removed first. Analysis times were unified to UTC, and the local time of each event was noted only when interpreting the results. The coordinate, time-origin, and grid-alignment conventions are given in Supplementary Files S1–S3. The input detections and NIROPS reference distributions for the 29 U.S. fires are provided in Figure S1.

2.2.3. Korean Study Fires and Final Perimeters

The Korean application targets are 21 fires of at least 100 ha during 2022–2025, based on the event information of the Korea Forest Service National Forest Fire Information System and the final damage perimeters obtained [37]. Large wildfires are difficult to suppress once they spread (Figure 2c), so they are the primary targets for understanding the spread process that underpins prevention and early response. Article 21 of the Enforcement Decree of the Forest Disaster Prevention Act of Korea classifies wildfires with a damage area of at least 1,000,000 m^2 as large wildfires. Since 100 ha corresponds to about seven 375 m pixels, fires smaller than this are hardly targets for 375 m ToA production. The targets comprise 10 fires in 2022, 7 in 2023, and 4 in 2025, with final areas of 120.7–99,416.7 ha (Figure 2b). For the Uiseong event, the records of three fires (Angye, Anpyeong, and Geumseong) linked by the same spread process were merged into a single event (Table 2).

Table 2. Characteristics of the 21 Korean application fires. The VIIRS cell count is the value after 375 m grid deduplication summary inside the perimeter. The geostationary pixel count is the number of Himawari pixels used in the analysis.

ID	Date	Province	City/County/District	Detailed Location	Final Area (ha)	VIIRS Cells in Perimeter	Geostationary Pixels
KFS_01	22 March 2025	Gyeongbuk	Uiseong	Angye, Anpyeong, and Geumseong combined	99,416.7	3108	419
KFS_02	4 March 2022	Gyeongbuk	Uljin	Bungmyeon Ducheon-ri	16,302.0	1117	93
KFS_03	5 March 2022	Gangwon	Gangneung	Okgye-myeon Namyang-ri	4190.4	239	31
KFS_04	21 March 2025	Gyeongnam	Sancheong	Sicheon-myeon Sincheon-ri	3397.5	204	28
KFS_05	2 April 2023	Chungnam	Hongseong	Seobu-myeon Jung-ri	1337.0	98	15
KFS_06	22 March 2025	Ulsan	Ulju	Onyang-eup Unhwa-ri	1117.4	70	14
KFS_07	2 April 2023	Chungnam	Geumsan	Boksu-myeon Sindae-ri	889.4	42	0
KFS_08	28 February 2022	Gyeongnam	Hapcheon	Yulgok-myeon Noyang-ri	813.9	3	6
KFS_09	10 April 2022	Gangwon	Yanggu	Yanggu-eup Songcheong-ri	716.3	35	12
KFS_10	3 April 2023	Jeonnam	Hampyeong	Daedong-myeon Yeonam-ri	681.7	3	0
KFS_11	31 May 2022	Gyeongnam	Miryang	Bubuk-myeon Chunhwa-ri	660.8	46	10
KFS_12	15 February 2022	Gyeongbuk	Yeongdeok	Jipum-myeon Samhwa-ri	405.7	35	11
KFS_13	28 April 2025	Daegu	Buk-gu	Nogok-dong	259.6	20	9
KFS_14	28 May 2022	Gyeongbuk	Uljin	Geunnam-myeon Haenggok-ri	228.8	6	6
KFS_15	10 April 2022	Gyeongbuk	Gunwi	Samgukyusa-myeon Hwabuk-ri	225.7	25	6
KFS_16	3 April 2023	Jeonnam	Suncheon	Songgwang-myeon Bongsan-ri	188.3	5	0
KFS_17	8 March 2023	Gyeongnam	Hapcheon	Yongju-myeon Wolpyeong-ri	179.1	5	5
KFS_18	9 April 2022	Chungnam	Seosan	Unsan-myeon Gopung-ri	158.1	2	4

Table 2. *Cont.*

ID	Date	Province	City/County/District	Detailed Location	Final Area (ha)	VIIRS Cells in Perimeter	Geostationary Pixels
KFS_19	5 April 2022	Gyeongbuk	Bonghwa	Bonghwa-eup Hwacheon-ri	130.0	3	5
KFS_20	11 March 2023	Gyeongnam	Hadong	Hwagae-myeon Beomwang-ri	128.5	4	5
KFS_21	11 April 2023	Gangwon	Gangneung	Nangok-dong	120.7	9	0

2.3. ToA Production

2.3.1. Detection Preprocessing and Cell Summary

All active-fire detection times were converted to elapsed time (h) from the per-event analysis time origin.

For polar-orbiting detections, only points whose center lay strictly inside the final perimeter of the event were retained. The retained detections were aggregated by 375 m grid cell. When several detections fell in the same cell, the median of the detection times was used as the observed ToA for that cell.

For geostationary detections, given the pixel size of about 2 km, GOES and Himawari pixels within a 2 km buffer of the final perimeter were retained. The repeated detection history of each pixel was summarized to a single point per pixel by the same median rule.

2.3.2. Interpolation Configurations and Empirical Variance Weighting

ToA surfaces were computed with ordinary kriging of an exponential covariance [38–40]. The variogram was fitted to the coordinates and ToA of the polar-orbiting cells using an empirical variogram of 20 lags. The nugget lower bound was set to 0.001 times the partial sill. When the nugget is near zero the kriging matrix approaches singularity, so the lower bound was imposed. The nugget that enters Equation (1) is the value after this lower bound. The nearest 50 observations were used for each prediction cell. For the numerical stability of the kriging equations, 10^{-4} times the partial sill was added to the diagonal of the covariance matrix (Tikhonov regularization).

The comparison configurations were defined along two axes: input data (polar-orbiting only or polar-orbiting together with geostationary) and interpolation method (inverse distance weighting or ordinary kriging). This yielded five configurations (Table 3). Both the external-reference evaluation and the spatial block cross-validation used all five configurations including PS-IDW. In the names, P denotes polar-orbiting, S denotes geostationary, and wt denotes variance weighting. PSwt-OK adds the sensor-specific σ_i^2 of Equation (1) to the observation diagonal of the covariance matrix in the form of measurement-error kriging [41].

Table 3. The five production configurations compared. P denotes polar-orbiting (VIIRS) and S denotes geostationary (GOES or Himawari). The inverse distance weighting (IDW) configurations use the nearest 20 points and a distance exponent of 2. The variance-weighting coefficient σ_i^2 of PSwt-OK is determined by Equation (1).

Configuration	Method	Input	Geostationary Weighting	Role
P-IDW	inverse distance weighting	polar-orbiting	—	baseline
PS-IDW	inverse distance weighting	polar-orbiting + geostationary	distance weighting	control
P-OK	ordinary kriging	polar-orbiting	—	candidate

Table 3. *Cont.*

Configuration	Method	Input	Geostationary Weighting	Role
PS-OK	ordinary kriging	polar-orbiting + geostationary	equal weighting	candidate
PSwt-OK	ordinary kriging	polar-orbiting + geostationary	variance weighting	candidate

$$\sigma_i^2 = \begin{cases} 0, & i \in \text{VIIRS}, \\ 27.4 \max\left(c_0, 1\text{h}^2\right), & i \in \text{GOES or Himawari}. \end{cases} \quad (1)$$

where c_0 is the fitted nugget, subject to the lower bound of 0.001 times the partial sill. The coefficient 27.4 is an empirical variance inflation determined from the area ratio $(2000/375)^2 - 1 = 27.444\ldots$ by which the 2 km geostationary support additionally exceeds the 375 m output cell. This value is not the measured error variance of a geostationary pixel. The coefficient does not assign a fixed weight of $1/27.4$ to the geostationary observations; the actual weights are determined by the full kriging system. This procedure is therefore not formal area-to-point kriging [42], which integrates the covariance over differing support areas, but an approximation that keeps large pixels from excessively dominating the kriging weights. The kriging weight of each observation is determined by the whole covariance structure, which includes this variance-inflation term.

The sensitivity of the coefficient choice was checked at $1/4$, $1/2$, 1, 2, and 4 times the default value. At none of the coefficients did the results of the external-reference evaluation change appreciably (Table S10).

For events without geostationary inputs, the observation vector and covariance matrix of PSwt-OK become identical to those of P-OK, so the two surfaces coincide exactly.

2.3.3. Output Rasters

The product for each event is two layers, the ToA and the ordinary-kriging standard deviation recorded at cells inside the final perimeter, with units of hours (h). The area outside the perimeter was left as NoData. The final Korean products were recorded on a 375 m grid in EPSG:5179. Kriging, block cross-validation, and the statistical tests were implemented in Python (version 3.11.15) with NumPy (version 2.3.5), SciPy (version 1.17.1), and scikit-learn (version 1.8.0). Spatial data processing and map production used the Python packages GeoPandas (version 1.1.3), Shapely (version 2.1.2), Rasterio (version 1.4.4), and Matplotlib (version 3.10.8). The national boundaries in the map backgrounds were read from the world map data bundled with QGIS (version 3.40.12). The versions and roles of the software are given in Supplementary File S6.

2.4. Evaluation and Quality Screening

2.4.1. Per-Survey Area Overlap

To preserve the interval-censored nature of the reference ToA, at each NIROPS survey time t_k the prediction raster was truncated to $P_k = \{s : \hat{T}(s) \leq t_k\}$ and overlaid with the reference area R_k accumulated up to the same time (Figure 3). The reference area R_k was constructed as the set of cells first included in a survey of order k or lower. The evaluation domain D was fixed to the cells inside the reference final burn perimeter. This evaluation therefore measures the internal progression order and time under the condition that the final extent is known, and does not include the performance of predicting whether the fire will spread beyond the perimeter.

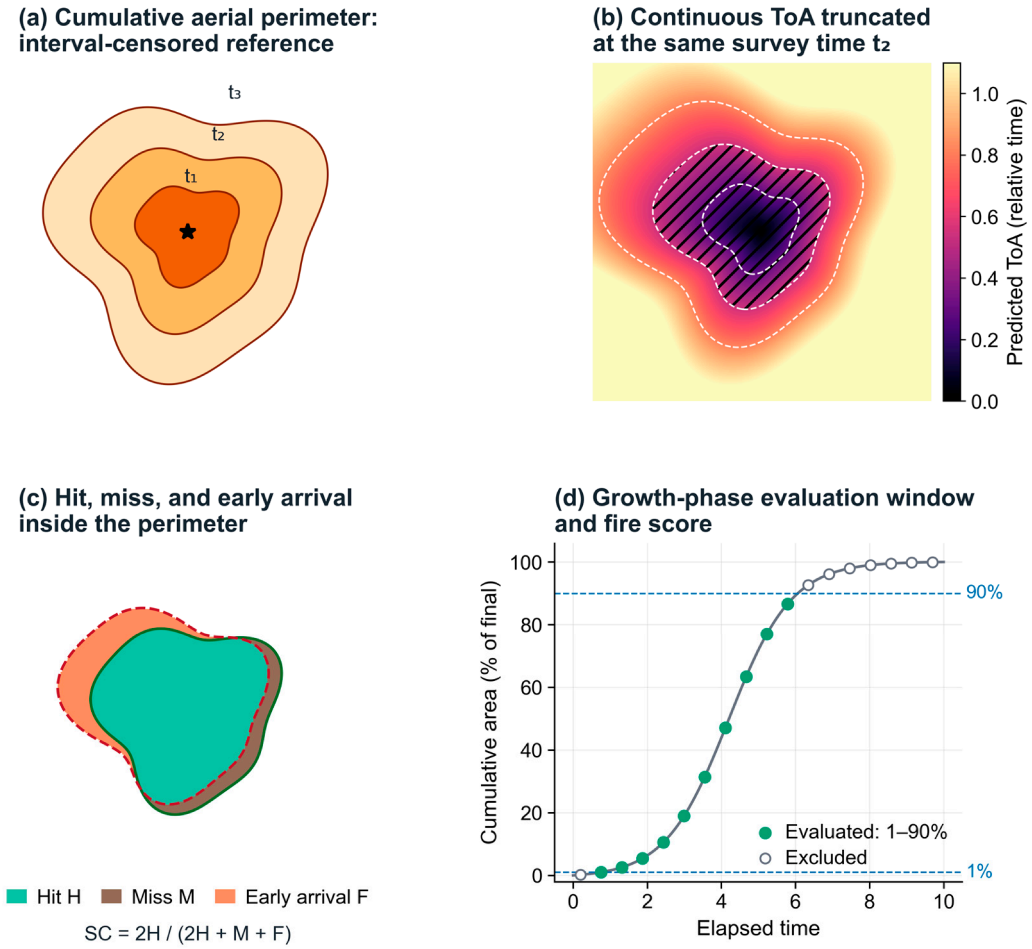


Figure 3. Structure of the accuracy evaluation. (**a**) Because the NIROPS reference records the cumulative area per survey time, the actual arrival at a cell is constrained only to lie between two surveys. (**b**) The continuous predicted ToA is truncated at the same survey time. (**c**) Inside the reference final perimeter, the prediction is partitioned into hit, miss, and early arrival to compute SC. (**d**) Only surveys between 1% and 90% of the final area are evaluated, and the median of the per-survey SC is used as the fire score.

Let H be the number of cells included in both the reference and the prediction, M the number of cells included only in the reference, and F the number of cells included only in the prediction. The Sørensen coefficient (SC) was computed by Equation (2), following the precedent of ToA evaluation against aerial infrared perimeters [17,18].

$$\mathrm{SC} = \frac{2H}{2H + M + F} \tag{2}$$

The evaluation was restricted to growth surveys whose reference cumulative area was between 1% and 90% of the final. Below the initial 1%, the positional error of a few cells dominates the metric. After 90%, both the reference and the prediction converge to the final perimeter, so the overlap rises regardless of interpolation ability. Owing to this restriction, the number of evaluation surveys per event may be smaller than the number of growth surveys in Table 1. The sensitivity of this evaluation interval (1–90%) and of the fire-exclusion re-test is given in Tables S3–S5.

The independent unit of aggregation was defined as the fire event. The per-survey metric was computed first, the median was taken over the growth surveys of that event, and then the scores of the 29 events were compared. The per-event SC was recorded to four decimal places, and this value was used for the aggregation and the tests. Differences

between methods were tested for all pairs of the five configurations by a two-sided exact sign test on the nonzero per-event differences [43]. The surveys of the same event were not treated as independent samples because the cumulative perimeters overlap in time and thus have strong autocorrelation. The paired comparison was presented as the distribution of the per-event differences and the median difference. The 95% confidence interval of the median difference was computed from the percentiles of 10,000 event-resampling bootstraps. The raw p and the Holm-corrected p [44] of the sign test are given in Table S4. The comparison of kriging and IDW was distinguished as an exploratory secondary comparison.

Because the same 29 fires were used for selection and reporting, this evaluation is not a separate hold-out.

2.4.2. Spatial Block Cross-Validation: Method Comparison and Reliability Diagnostic

A spatial block cross-validation that separates adjacent observations by block [45,46] was used as the second test for configuration selection and as a per-fire reliability diagnostic.

The polar-orbiting cells of each event were divided into square spatial blocks. The block side is half of the variogram range fitted to the whole polar-orbiting data of the event, clamped to 5–30 km. One block at a time was held out for validation, and the ToA of the held-out cells was predicted from the remaining polar-orbiting cells and all geostationary summary points of the event. Because the geostationary summary points are observations actually used by the product, they were retained in the training data of every fold. The variogram was refitted using only the training polar-orbiting cells of each fold. The σ_i^2 of the geostationary observations was also computed from the nugget c_0 of that fold. Only folds with at least 5 training polar-orbiting cells were used. A comparison with cross-validation in which the variogram fitted to the whole polar-orbiting data of the event is held fixed is given in Table S7.

The per-fire performance of each configuration was summarized by the Pearson r combining the held-out observations and predictions of the valid folds. Differences between methods were compared by a per-event two-sided exact sign test over all pairs of the five configurations. The 95% confidence interval of the median Δr was computed by the same bootstrap as in Section 2.4.1. The raw p and the Holm-corrected p of the sign test are given in Table S7. For the selected default configuration, the mean error ME and the mean absolute error MAE were also reported. The two statistics were computed by Equation (3) and Equation (4), respectively.

$$\mathrm{ME} = n^{-1} \sum (\hat{T}_i - T_i) \tag{3}$$

$$\mathrm{MAE} = n^{-1} \sum |\hat{T}_i - T_i| \tag{4}$$

where n is the number of held-out cells, and $\hat{T}_i$ and T_i are the predicted and observed ToA of cell i.

ME and MAE are auxiliary statistics that show the direction and time scale of the error.

Events for which the cross-validation conditions were not met or for which the variogram degenerated were released with evaluation withheld. The r is a spatial self-prediction diagnostic under the operational inputs. Its relationship to the external accuracy was evaluated in Section 3.2. The r is not a time accuracy directly calibrated against an external reference.

2.4.3. Korean Production and Quality Status

The selected default configuration was applied to the 21 Korean fires without change. Only events with at least 5 deduplication-summarized VIIRS cells inside the final perimeter and a converged exponential variogram were subjected to ToA and standard-deviation

production. Events that did not pass these criteria were not arbitrarily filled with a raster; the excluded-from-production status and the reason were recorded.

The Korean block cross-validation applied the same block definition with up to 10 folds. The r, ME, and MAE were defined only for events that satisfied all of the minimum conditions on the number of blocks, training cells, validation cells, and valid predictions (Supplementary File S5). If the variogram range of the whole data was below 0.5 km, the event was left with evaluation withheld even when a raw r could be computed. Evaluation withheld means not that the accuracy was low but that the data structure was insufficient to define a valid internal cross-validation.

3. Results

3.1. Comparison of Configurations in External-Reference Evaluation

In the external-reference evaluation, the accuracy of the five configurations was broadly similar (differences in the between-event median SC below 0.01; P-OK 0.9398, PSwt-OK 0.9394, PS-OK 0.9374, P-IDW 0.9340, PS-IDW 0.9336; Table 4). PSwt-OK and PS-OK were lower than P-OK by per-event medians of 0.0007 and 0.0015, respectively, and these differences remained detectable after Holm correction (Figure 4a; Table S4). The control PS-IDW was lower than the three kriging configurations, and this difference remained detectable after Holm correction (median ΔSC +0.0048 for PSwt-OK versus PS-IDW; Figure 4a; Table S4).

Table 4. Event-level comparison of the five configurations in the external-reference evaluation. Values are the medians over 29 events after reducing the per-survey metric to a per-event median. The distribution of per-event paired differences and the median difference are shown in Figure 4a.

Configuration	Input Data and Weighting	SC
P-IDW	VIIRS, distance weighting	0.9340
PS-IDW	VIIRS + GOES, distance weighting	0.9336
P-OK	VIIRS, ordinary kriging	0.9398
PS-OK	VIIRS + GOES, equal weighting	0.9374
PSwt-OK	VIIRS + GOES, empirical variance weighting	0.9394

3.2. Spatial Block Cross-Validation: Method Selection and Validity of the Reliability Metric

In the spatial block cross-validation, the between-event median r of the five configurations was compared. The between-event median r of PSwt-OK was the highest at 0.880 (Table 5). After Holm correction in the exact sign test, PSwt-OK was significantly higher than P-IDW and PS-OK (Figure 4b; Table S7). The differences from P-OK and PS-IDW were not detected (p = 0.711 and 0.265 before correction).

The control decomposition distinguishes the data effect from the method effect. The gain obtained from adding the geostationary data alone was a median +0.036 [+0.014, +0.094] (PS-IDW versus P-IDW). No additional gain beyond this was detected for variance-weighted kriging either (PSwt-OK versus PS-IDW; p = 0.265 before correction, Holm p = 1.000; Table S7). No difference was detected between equal-weighted kriging and PS-IDW either.

In the external-reference evaluation, the SC of PSwt-OK was significantly lower than that of P-OK (per-event median −0.0007, Holm p = 0.011). Among the three configurations that use geostationary data, PSwt-OK was significantly higher than PS-IDW and PS-OK in SC and than PS-OK in r. On this basis, PSwt-OK was selected as the default configuration that uses geostationary data.

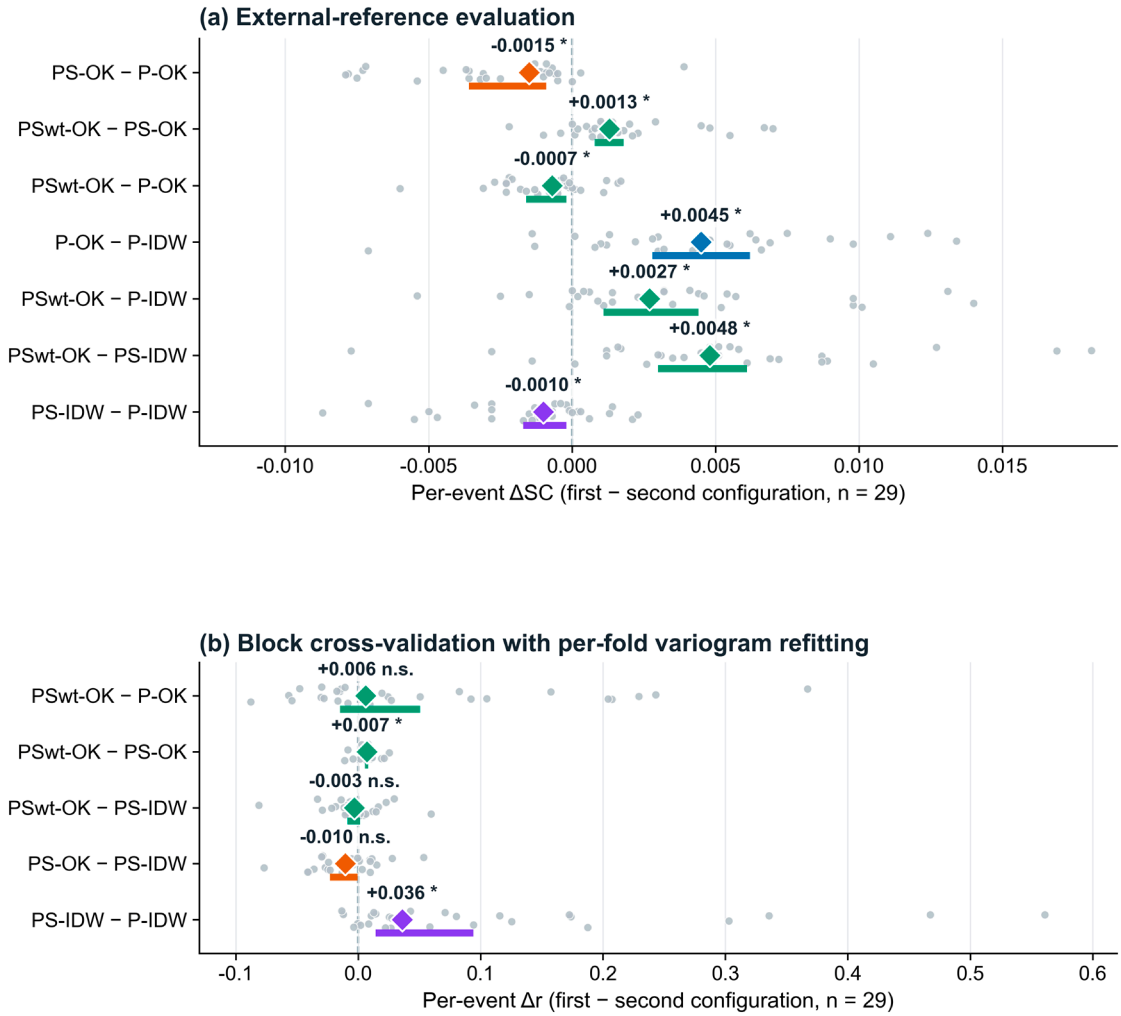


Figure 4. Distribution of per-event paired differences for the two tests. (**a**) ΔSC of the external-reference evaluation. (**b**) Δr of the spatial block cross-validation. The points in each row are the per-event differences over 29 events (first method − second method). The diamond is the median difference. The thick bar is the 95% percentile confidence interval from 10,000 event-resampling bootstrap iterations. The *p*-values of the two-sided exact sign test are given in Tables S4 and S7. An asterisk denotes $p < 0.05$ in the two-sided exact sign test after Holm correction, and n.s. denotes that no difference was detected.

Table 5. Comparison of the five configurations in the spatial block cross-validation. Values are the medians over 29 events of the per-event Pearson r pooled across all valid folds. The distribution of per-event paired differences and the median difference are shown in Figure 4b.

Configuration	Block Cross-Validation Training Input Data	Median r
P-IDW	VIIRS	0.813
PS-IDW	VIIRS + GOES	0.879
P-OK	VIIRS	0.845
PS-OK	VIIRS + GOES, equal weighting	0.879
PSwt-OK	VIIRS + GOES, empirical variance weighting	0.880

Events with higher block cross-validation r also had higher SC evaluated against the NIROPS reference (Spearman $\rho = 0.743$, $p < 0.001$; Figure 5).

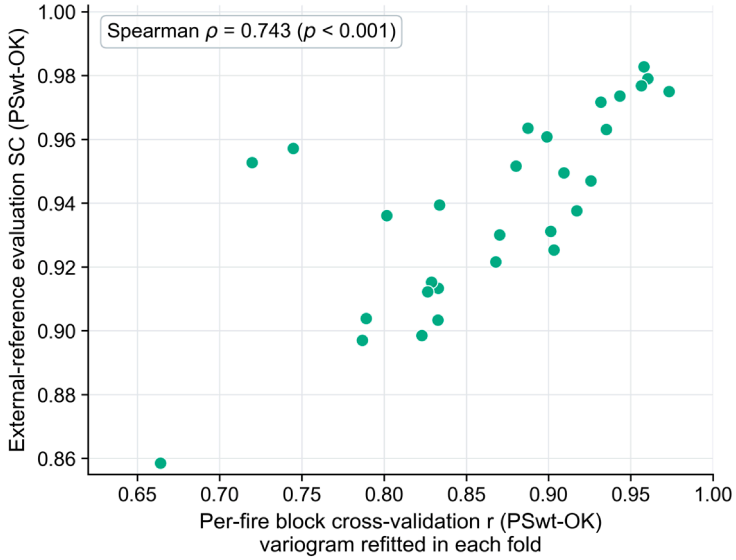


Figure 5. Correspondence between block cross-validation r and the external-reference evaluation. This is the relationship between the per-event block cross-validation r of PSwt-OK and the external NIROPS SC.

For the 29 U.S. events, the median ME of PSwt-OK was −15.7 h and the median MAE was 91.9 h. The per-event r of the five configurations and the ME and MAE of PSwt-OK are listed in Table S6.

For PSwt-OK, the median per-event difference between the r obtained with the variogram refitted in each fold and the r obtained with the variogram of the whole event data held fixed was 0.0000. This difference was not significant in the sign test ($p = 1.000$). Sensitivity analyses that exclude folds with a variogram range below 0.5 km, or the corresponding events, are listed in Table S7.

3.3. Korean Quality-Screened Dataset

Of the 21 Korean events, PSwt-OK ToA and kriging standard-deviation rasters were produced for 15 (Table 6; Figure 6). Six events had fewer than 5 VIIRS summary cells inside the perimeter or failed variogram fitting and were excluded from production. The ToA and standard-deviation rasters of the 15 produced events had no missing pixels inside the perimeter. The Korean results are not an independent regional validation but an application and demonstration of the method of this study.

The seven events for which a valid block cross-validation was defined were classified as evaluated, whereas the remaining eight produced events were classified as evaluation-withheld (Table 6).

The representative events show that r is not a quantity such as fire size or kriging standard deviation. Uljin used 1117 VIIRS cells and 93 geostationary summary points. Its block cross-validation r was 0.942, ME was −6.0 h, and MAE was 10.4 h. The ToA range produced was 192.2 h. Uiseong, the largest Korean event, had 3108 VIIRS cells and 419 geostationary summary points, but its block cross-validation r was lower than Uljin's (r = 0.771, ME = −0.2 h, MAE = 10.9 h). Gangneung had the smallest MAE among the seven evaluated events (r = 0.739, ME = +0.2 h, MAE = 5.8 h). Its r was nevertheless lower than that of Uljin. Even when the MAE is small, r can be low if the ToA range is short, so the two statistics are not substitutes for each other. The full map of the 15 Korean products is provided in Figure S2. The per-event blocks, evaluation pairs, r, ME, MAE, and variogram basis are listed in Tables S8 and S9.

Table 6. Production status and spatial block cross-validation statistics of the 21 Korean events. r, ME, and MAE are presented only for events with a valid block cross-validation. ME is the mean of the predicted ToA minus the observed ToA of the held-out cells. ME and MAE are in units of h. "Evaluation withheld" means that the raster exists but the conditions for defining valid statistics were not met. A dash is an explicit status of not produced or no valid statistics.

ID	Fire	Raster	Block CV r	ME (h)	MAE (h)	Quality Screening	Basis for Exclusion/Evaluation Withheld
KFS_01	Uiseong	Produced	0.771	−0.2	10.9	Evaluated	Meets evaluation conditions
KFS_02	Uljin	Produced	0.942	−6.0	10.4	Evaluated	Meets evaluation conditions
KFS_03	Gangneung	Produced	0.739	+0.2	5.8	Evaluated	Meets evaluation conditions
KFS_04	Sancheong	Produced	0.837	+7.5	20.4	Evaluated	Meets evaluation conditions
KFS_05	Hongseong	Produced	0.714	+2.5	7.7	Evaluated	Meets evaluation conditions
KFS_06	Ulju	Produced	0.699	−6.2	18.8	Evaluated	Meets evaluation conditions
KFS_07	Geumsan	Produced	—	—	—	Evaluation withheld	Insufficient spatial blocks
KFS_08	Hapcheon	Excluded from production	—	—	—	Excluded from production	Fewer than 5 VIIRS summary cells inside perimeter
KFS_09	Yanggu	Produced	—	—	—	Evaluation withheld	Variogram degeneracy
KFS_10	Hampyeong	Excluded from production	—	—	—	Excluded from production	Fewer than 5 VIIRS summary cells inside perimeter
KFS_11	Miryang	Produced	0.532	+1.4	12.2	Evaluated	Meets evaluation conditions
KFS_12	Yeongdeok	Produced	—	—	—	Evaluation withheld	Fewer than 10 validation pairs
KFS_13	Daegu Buk-gu	Produced	—	—	—	Evaluation withheld	Insufficient spatial blocks
KFS_14	Uljin Geunnam	Produced	—	—	—	Evaluation withheld	Variogram degeneracy
KFS_15	Gunwi	Produced	—	—	—	Evaluation withheld	Insufficient spatial blocks
KFS_16	Suncheon	Produced	—	—	—	Evaluation withheld	Variogram degeneracy
KFS_17	Hapcheon Yongju	Produced	—	—	—	Evaluation withheld	Insufficient spatial blocks
KFS_18	Seosan	Excluded from production	—	—	—	Excluded from production	Fewer than 5 VIIRS summary cells inside perimeter
KFS_19	Bonghwa	Excluded from production	—	—	—	Excluded from production	Fewer than 5 VIIRS summary cells inside perimeter
KFS_20	Hadong	Excluded from production	—	—	—	Excluded from production	Fewer than 5 VIIRS summary cells inside perimeter
KFS_21	Gangneung Nangok	Excluded from production	—	—	—	Excluded from production	Variogram fitting failure

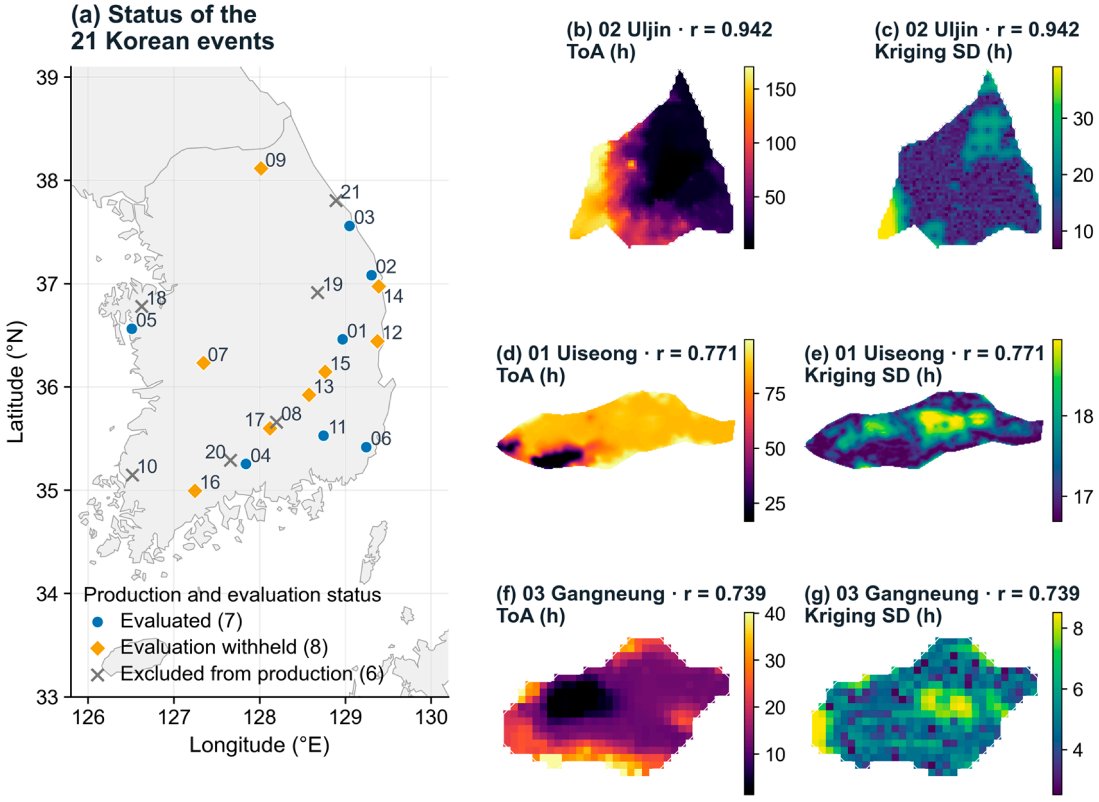


Figure 6. Status and representative rasters of the Korean products. (**a**) The evaluated, evaluation-withheld, and excluded-from-production status of the 21 events. (**b**,**c**) ToA and kriging standard deviation of Uljin, (**d**,**e**) Uiseong, (**f**,**g**) Gangneung. Only the interior of the final perimeter is shown. Because the color range is the per-event 2nd–98th percentile, colors are not directly compared as absolute times across events.

4. Discussion

4.1. Comparison of External-Overlap Accuracy with Prior Satellite Products

Across the 29 U.S. fires, the between-event median obtained by aggregating the per-survey Sørensen coefficient (SC) as a per-event median was about 0.94. Per-fire overlap of VIIRS-only tracking was F1 = 0.85 [12]. Machine-learning fire-arrival-time estimation from polar-orbiting active fires had a mean SC of 0.82 [17]. Generative models that combine satellite data with physics-based simulation had a mean SC of 0.81 [18,19]. The final overlap of hourly geostationary-only perimeters was IoU = 0.77 [11].

These values should not be read as a simple ranking. The reference perimeter and the aggregation unit differ among studies. Our value is the per-survey overlap during the growth phase summarized as a per-event median and then aggregated as a between-event median, whereas the studies above report per-fire means or use the final perimeter. Including early, small phases generally lowers the overlap value.

Farguell et al. [17] and Shaddy et al. [18,19] computed the same Sørensen coefficient against the NIROPS aerial perimeters as we did. Those values were 0.81–0.82. However, the fact that the same metric and the same series of reference perimeters were used does not mean that the evaluation conditions were the same. We therefore do not conclude that our procedure is superior.

4.2. Role Decomposition of Geostationary Inclusion and Variance Weighting

The two tests gave different answers. In the U.S. external-reference evaluation, the differences in SC among the five configurations were small ($|\text{median } \Delta SC| \leq 0.0058$ between configurations; Table S4). In spatial block cross-validation, the differences among configurations were larger (median Δr of +0.057 for PSwt-OK versus P-IDW). This difference arises from the test conditions. Where observations are dense, any configuration produces a similar surface. In U.S. fires, a high share of the pixels inside the perimeter contain a VIIRS detection, with a median of 91.9% (Table S6). The difference between configurations was therefore diluted in SC, which is a full-coverage overlap. Spatial block cross-validation empties whole observation blocks and thereby exposes that difference.

We separated the two effects using controls. The configuration that added geostationary data had a per-fire median cross-validation r that was +0.036 higher. When geostationary data were added with equal weighting only, no additional gain beyond that was detected. The configuration that also added variance weighting was significantly higher than equal-weighted kriging. In short, geostationary inclusion fills empty locations, and variance weighting suppresses the distortion of coarse pixels and thereby makes that inclusion safe.

Studies that estimated wildfire arrival time by kriging existed before. Veraverbeke et al. kriged single-sensor MODIS active fires to produce daily progression and per-pixel kriging standard error [2]. Scaduto et al. combined MODIS and VIIRS with several interpolation methods but did not address geostationary data, measurement-error weighting, or per-pixel uncertainty [14]. Both studies left geostationary integration as future work. We combined polar-orbiting and geostationary data through measurement-error variance-weighted kriging and produced per-pixel uncertainty as well. This measurement-error kriging framework is an established method in geostatistics [41], and this study is the first to apply it to wildfire arrival time and satellite sensor fusion.

4.3. The Need for Geostationary Data in the Korean Application

Small Korean fires are easily missed by polar-orbiting sensors alone. Jang et al. reported that the detection probability of MODIS active fires in Korea is only 15.2% [23]. In the same study, the detection probability of the Himawari geostationary sensor was much higher, at 93%. The early-detection advantage of geostationary data has also been confirmed in Korea. The first-detection delay of the geostationary sensor was about 13 min, whereas that of polar-orbiting sensors exceeded 210 min [24]. These results support our reason for always including geostationary data in the Korean products.

Prior Korean studies and our study answer different questions. Prior geostationary studies addressed whether a fire is present and how quickly it is detected [23,24]. Monitoring of spread extent that combines GK2A and VIIRS has also been attempted [25]. We address when each pixel burned and how reliable that value is. We are the first to produce per-pixel arrival-time rasters of Korean fires together with uncertainty.

The SC observed in the United States cannot be transferred directly to the expected accuracy of Korean fires. The two regions differ in fire size, sensors used, and reference data. Korean values therefore indicate reliability only through internal diagnostics, not through U.S. accuracy.

4.4. Providing Reliability Where External Validation Is Impossible

Korea has no aerial or field data with which to validate arrival time. We therefore used internal diagnostics as a reliability metric. In the United States, fires with a high spatial block cross-validation r also had a high external SC ($\rho = 0.743$). After confirming this

relationship in the United States, where validation is possible, we applied this diagnostic, rather than accuracy, to Korea, where validation is impossible.

The character of a quality-screened dataset is present in prior work as well. Benali et al.'s database of large Portuguese fires was distributed with a five-level per-fire reliability flag [4]. That flag is a subjective grade based on the type of input data and has no external validation. Our reliability metric follows the same idea. However, we confirmed the relationship between this metric and the external-reference evaluation in the United States.

The difficulty of small fires also supports the character of this dataset. Geostationary-only mapping has markedly lower overlap in the small, early phase after ignition [11]. Coarse satellite-based arrival-time estimation has been reported to be inappropriate below 500 ha [13]. Korean fires of the 100–1000 ha class have few valid blocks and readily remain evaluation-withheld or excluded from production. We did not force such fires to be filled, and we distributed them together with their status and reason.

We make clear what this metric measures. This metric is the spatial block cross-validation r. This r indicates how well the arrival time at a location is recovered from the remaining observations after part of the observations is masked. A high r means that the observations are mutually consistent and that the arrival-time map is well defined. The r is not, however, an absolute time error. Because r is a correlation coefficient, it comes out low even for a small error when the fire's arrival-time range is narrow. Gangneung, noted earlier, was such a case. A low r should therefore not be read as inaccurate.

Users make two decisions with these data. First, which fires to use is chosen by the evaluation status and r. Reading ME and MAE together with r here also reveals the direction and magnitude of the time error. Second, which locations within a fire have a large kriging uncertainty is judged from the per-pixel standard-deviation raster. Even when r is high, there can be pixels with a large standard deviation. The two axes therefore do not replace each other.

How the data are used differs by study type. For an analysis of multiple fires together, it is safer to view results separated by r. Here, one should also check whether the conclusion changes when evaluation-withheld products are included. When used as machine-learning training data, training and validation should be split by fire or by spatial block, not by pixel. Because neighboring pixels are similar to each other, splitting by pixel inflates performance beyond its true level.

4.5. Limitations

First, the NIROPS reference is the cumulative perimeter at the survey time, so its temporal resolution is limited to the survey interval, which was about one day. The evaluation is also an areal overlap matched to this interval-censored structure, so the absolute time accuracy between surveys and the prediction of spread beyond the perimeter are outside the scope of the evaluation. The positional error of the survey perimeter and the uncertainty of some survey times also remain. Because the polar-orbiting observations were treated as error-free and only the geostationary observations had their variance inflated, the temporal response of the product where polar-orbiting observations exist is effectively governed by the polar-orbiting revisit interval. Responses at time scales shorter than one day remain a task for a follow-up evaluation that holds out the geostationary observations from the input and uses them as the reference.

Second, the geostationary variance factor of 27.4 is a fixed empirical approximation based on the support-area ratio. The results of the external-reference evaluation were nearly the same under a sensitivity sweep of 1/4 to 4 times the factor. This sensitivity does not, however, establish 27.4 as the true value of the measurement-error variance. This diagonal addition assumes that the observation errors are mutually independent and have zero mean.

Systematic time bias and correlated errors between pixels are therefore not modeled directly. It also does not model the actual heteroscedasticity due to view angle, cloud and smoke, or day–night conditions, and other geostationary products require separate confirmation. The kriging standard deviation is a model-based uncertainty, and the coverage was not calibrated against an external reference. The uncertainty of the variogram estimate and the perimeter uncertainty are not included in this standard deviation. The standard-deviation raster is an auxiliary layer for finding locations with a large kriging uncertainty inside a fire. The median cell summary reduces early noise but can delay the first arrival of a long-lived active cell.

Third, spatial block cross-validation is a test of operational conditions in which geostationary observations are available. Large-scale polar-orbiting data loss, sensor outages, and common cloud obscuration [47] are outside the scope of this paper and are subjects for follow-up study. The variogram was refitted using only the training polar-orbiting cells of each fold. The block size and assignment were nevertheless determined from the variogram fitted to the whole polar-orbiting data of the event. For this reason, the values of the validation block contribute indirectly to the block configuration.

Fourth, Korea has no external reference, and the character of the final damage perimeter also differs from NIROPS. If the input perimeter is inaccurate, the output surface is inaccurate to the same degree. Fires of the 100–1000 ha class have few valid blocks and readily remain evaluation-withheld or excluded from production.

4.6. Data Release and Reuse

The reproduction package was organized to include the manifest of the 29 U.S. fires; the per-survey and per-fire SC; the spatial block cross-validation results of the five configurations; the PSwt-OK r, ME, and MAE; the status table of the 21 Korean fires; the Korean final ToA and standard-deviation rasters; and the analysis code. The Magstadt raw data are not redistributed but are linked through the original DOI and license [26], and the derivatives preserve the fire event ID and the raw-data version.

Users should check, together with the Korean rasters, the status in Table 6, the variogram, block, and evaluation-pair basis in Tables S8 and S9, the ME and MAE, and the standard-deviation raster. When evaluating a fire-spread model, distinguishing the progression accuracy inside the final perimeter from the prediction of spread beyond the perimeter as separate metrics is consistent with the evaluation target of this study.

The time axis of a reuse analysis must also be stated per fire. ToA is not the absolute UTC time itself but the elapsed time after a per-fire origin. Combining the raster values of multiple fires as they are mixes different calendar times and weather conditions. For comparison across fires, one must restore the absolute time using the origin metadata, or realign to the common definition of elapsed time after ignition. The colors in Figures 6 and S2 are visualizations normalized to the internal range of each fire, so the color tones of the figures should not be read as numeric data. Quantitative analysis should be performed on the GeoTIFF values and the perimeter mask. The figures should be used only for inspecting spatial patterns.

5. Conclusions

This study presented the PSwt-OK procedure, which combines the spatial detail of 375 m VIIRS with the temporal frequency of GOES and Himawari within ordinary kriging and lowers the influence of large geostationary pixels through empirical variance inflation. In the external-reference evaluation of 29 NIROPS fires in the western United States, the SC of PSwt-OK was lower than that of polar-orbiting-only P-OK by a per-event median of 0.0007, and this difference remained detectable after Holm correction. In spatial block cross-

validation, which held out polar-occupied blocks and predicted with operational input data, PSwt-OK had the highest median r among the five configurations. The differences from P-OK and the control PS-IDW were not detected, however. Among the three configurations that use geostationary data, PSwt-OK was significantly higher than PS-IDW and PS-OK in SC and than PS-OK in r.

Applying the two tests together, we selected PSwt-OK as the default method. We unified the second test for method selection and the per-fire reliability diagnostic under the same block cross-validation. The higher this r for a fire, the higher its SC in the United States ($\rho = 0.743$). This result does not mean that r is a calibrated value of true accuracy; it is a basis for indicating, in fires without an external reference, the spatial self-predictability of the operational input data. We also reported ME and MAE, which show the direction and time scale of the error.

Applying the fixed procedure to the 21 Korean fires produced ToA and standard-deviation rasters for 15 fires. Seven were classified as evaluated and eight as evaluation withheld. The remaining six were designated as excluded from production. The Korean results are not an independent regional validation but an application and demonstration of the fixed procedure. The constructed data provide a candidate reference layer for Korean fire-spread research, but until external aerial or field data become available they should be used not as a validated accuracy product but as a self-screened dataset selected by explicit quality criteria.

These data are a product that reconstructs arrival time inside the known final perimeter of fires that have already occurred. They are not an independent prediction product for spread beyond the perimeter or for real-time field command. The Korea Forest Service, the National Institute of Forest Science, local governments, and fire-spread modelers can use these data for post-suppression fire analysis, for studies of weather, terrain, and fuel factors, and for the evaluation of spread models.

A subject for future research is to obtain aerial and field data for Korean fires and to validate the arrival-time reconstruction independently. Modeling the effects of view angle, cloud, and smoke is also a subject for follow-up study. Follow-up studies may apply the same procedure to fires in other regions and to other geostationary satellites, and use the produced arrival times for spread-model validation and analysis of spread drivers.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18193321/s1, Table S1: Sample selection chain; Table S2: Specification of the final 29 events; Table S3: Sensitivity to the evaluation cutoff thresholds; Table S4: Fire-level exact sign test by cutoff; Table S5: Stability of the re-test excluding one fire at a time; Table S6: Per-fire design characteristics and block cross-validation r of the five configurations; Table S7: Method comparison and SC correlation; Table S8: Status of all 21 fires and spatial block cross-validation statistics; Table S9: Variogram and raster diagnostics for the 15 produced fires; Table S10: Sensitivity to the empirical variance-inflation coefficient; Figure S1: Map of NIROPS reference and VIIRS detection distribution for the 29 U.S. fires; Figure S2: Maps of the 15 produced Korean PSwt-OK ToA rasters; Supplementary File (S1–S6): Coordinates, time origin, grid alignment, NoData, block cross-validation reproducibility conventions, and software environment.

Author Contributions: Conceptualization, C.C. and J.C.; methodology, C.C. and J.C.; software, C.C.; validation, C.C. and J.C.; formal analysis, C.C.; investigation, C.C. and H.S.; data curation, C.C. and H.S.; writing—original draft preparation, C.C.; writing—review and editing, C.C., H.S. and J.C.; visualization, C.C.; supervision, J.C.; project administration, J.C.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the National Institute of Ecology (NIE), funded by the Ministry of Climate, Energy and Environment (MCEE) of the Republic of Korea (NIE-B-2026-44).

Data Availability Statement: The original NIROPS data are provided at Mendeley Data DOI 10.17632/95rj5d379g.1. The derived manifest, the per-fire SC, the spatial block cross-validation results of the five configurations, the U.S. and Korean r, ME, and MAE, the status table of the 21 Korean fires, the final rasters, and the code implementing the method (preprocessing, kriging, cross-validation, and evaluation) are openly available in Zenodo at https://doi.org/10.5281/zenodo.21788344. The redistribution conditions of the raw satellite data follow the terms of the original providers NASA FIRMS, NOAA, and JAXA.

Acknowledgments: During the preparation of this manuscript, the authors used Anthropic Claude Opus 5 for the purposes of English translation and for debugging and refining the analysis scripts that implement the authors' analytical framework. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Durlević, U.; Ilić, V.; Radovanović, M.M.; Milanović Pešić, A.; Petrović, M.D.; Milenković, M.; Jovanović, J.M.; Atasoy, E. Wildfire Susceptibility Mapping in China Combining Machine Learning, Deep Learning, and Transformer-Based Models. *Earth* **2026**, *7*, 119. [CrossRef]
2. Veraverbeke, S.; Sedano, F.; Hook, S.J.; Randerson, J.T.; Jin, Y.; Rogers, B.M. Mapping the Daily Progression of Large Wildland Fires Using MODIS Active Fire Data. *Int. J. Wildland Fire* **2014**, *23*, 655–667. [CrossRef]
3. Cardil, A.; Tapia, V.M.; Monedero, S.; Quiñones, T.; Little, K.; Stoof, C.R.; Ramirez, J.; de-Miguel, S. Characterizing the Rate of Spread of Large Wildfires in Emerging Fire Environments of Northwestern Europe Using Visible Infrared Imaging Radiometer Suite Active Fire Data. *Nat. Hazards Earth Syst. Sci.* **2023**, *23*, 361–373. [CrossRef]
4. Benali, A.; Guiomar, N.; Gonçalves, H.; Mota, B.; Silva, F.; Fernandes, P.M.; Mota, C.; Penha, A.; Santos, J.; Pereira, J.M.C.; et al. The Portuguese Large Wildfire Spread Database (PT-FireSprd). *Earth Syst. Sci. Data* **2023**, *15*, 3791–3818. [CrossRef]
5. Sá, A.C.L.; Benali, A.; Fernandes, P.M.; Pinto, R.M.S.; Trigo, R.M.; Salis, M.; Russo, A.; Jerez, S.; Soares, P.M.M.; Schroeder, W.; et al. Evaluating Fire Growth Simulations Using Satellite Active Fire Data. *Remote Sens. Environ.* **2017**, *190*, 302–317. [CrossRef]
6. Chang, D.Y.; Jeong, S.; Park, C.-E.; Park, H.; Shin, J.; Bae, Y.; Park, H.; Park, C.R. Unprecedented Wildfires in Korea: Historical Evidence of Increasing Wildfire Activity Due to Climate Change. *Agric. For. Meteorol.* **2024**, *348*, 109920. [CrossRef]
7. Kim, J.; Kim, T.; Lee, Y.-E.; Im, S. Spatial and Temporal Variability of Forest Fires in the Republic of Korea over 1991–2020. *Nat. Hazards* **2025**, *121*, 9801–9821. [CrossRef]
8. Park, J.; Suh, J.; Baek, M. Climatic and Forest Drivers of Wildfires in South Korea (1980–2024): Trends, Predictions, and the Role of the Wildland–Urban Interface. *Forests* **2025**, *16*, 1476. [CrossRef]
9. Mahood, A.L.; Lindrooth, E.J.; Cook, M.C.; Balch, J.K. Country-Level Fire Perimeter Datasets (2001–2021). *Sci. Data* **2022**, *9*, 458. [CrossRef]
10. Andela, N.; Morton, D.C.; Giglio, L.; Paugam, R.; Chen, Y.; Hantson, S.; van der Werf, G.R.; Randerson, J.T. The Global Fire Atlas of Individual Fire Size, Duration, Speed and Direction. *Earth Syst. Sci. Data* **2019**, *11*, 529–552. [CrossRef]
11. Liu, T.; Randerson, J.T.; Chen, Y.; Morton, D.C.; Wiggins, E.B.; Smyth, P.; Foufoula-Georgiou, E.; Nadler, R.; Nevo, O. Systematically Tracking the Hourly Progression of Large Wildfires Using GOES Satellite Observations. *Earth Syst. Sci. Data* **2024**, *16*, 1395–1424. [CrossRef]
12. Chen, Y.; Hantson, S.; Andela, N.; Coffield, S.R.; Graff, C.A.; Morton, D.C.; Ott, L.E.; Foufoula-Georgiou, E.; Smyth, P.; Goulden, M.L.; et al. California Wildfire Spread Derived Using VIIRS Satellite Observations and an Object-Based Tracking System. *Sci. Data* **2022**, *9*, 249. [CrossRef]
13. Parks, S.A. Mapping Day-of-Burning with Coarse-Resolution Satellite Fire-Detection Data. *Int. J. Wildland Fire* **2014**, *23*, 215–223. [CrossRef]
14. Scaduto, E.; Chen, B.; Jin, Y. Satellite-Based Fire Progression Mapping: A Comprehensive Assessment for Large Fires in Northern California. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5102–5114. [CrossRef]
15. Barber, Q.E.; Jain, P.; Whitman, E.; Thompson, D.K.; Guindon, L.; Parks, S.A.; Wang, X.; Hethcoat, M.G.; Parisien, M.-A. The Canadian Fire Spread Dataset. *Sci. Data* **2024**, *11*, 764. [CrossRef]
16. Schiks, T.J.; Wotton, B.M.; Martell, D.L. Remote Sensing Active Fire Detection Tools Support Growth Reconstruction for Large Boreal Wildfires. *Fire* **2024**, *7*, 26. [CrossRef]
17. Farguell, A.; Mandel, J.; Haley, J.; Mallia, D.V.; Kochanski, A.; Hilburn, K. Machine Learning Estimation of Fire Arrival Time from Level-2 Active Fires Satellite Data. *Remote Sens.* **2021**, *13*, 2203. [CrossRef]

18. Shaddy, B.; Ray, D.; Farguell, A.; Calaza, V.; Mandel, J.; Haley, J.; Hilburn, K.; Mallia, D.V.; Kochanski, A.; Oberai, A.A. Generative Algorithms for Fusion of Physics-Based Wildfire Spread Models with Satellite Data for Initializing Wildfire Forecasts. *Artif. Intell. Earth Syst.* **2024**, *3*, e230087. [CrossRef]
19. Shaddy, B.; Binder, B.; Dasgupta, A.; Qin, H.; Haley, J.; Farguell, A.; Hilburn, K.; Mallia, D.V.; Kochanski, A.; Mandel, J.; et al. Generative Algorithms for Wildfire Progression Reconstruction from Multi-Modal Satellite Active Fire Measurements and Terrain Height. *Remote Sens.* **2026**, *18*, 227. [CrossRef]
20. McCarthy, N.F.; Tohidi, A.; Aziz, Y.; Dennie, M.; Valero, M.M.; Hu, N. A Deep Learning Approach to Downscale Geostationary Satellite Imagery for Decision Support in High Impact Wildfires. *Forests* **2021**, *12*, 294. [CrossRef]
21. Badhan, M.; Shamsaei, K.; Ebrahimian, H.; Bebis, G.; Lareau, N.P.; Rowell, E. Deep Learning Approach to Improve Spatial Resolution of GOES-17 Wildfire Boundaries Using VIIRS Satellite Data. *Remote Sens.* **2024**, *16*, 715. [CrossRef]
22. Ramsey, S.; Reinke, K.; Edwards, A.; Jones, S. Reconstructing Savannah Wildfire Events Using a Deep Learning Framework. *Remote Sens. Appl. Soc. Environ.* **2026**, *41*, 101837. [CrossRef]
23. Jang, E.; Kang, Y.; Im, J.; Lee, D.-W.; Yoon, J.; Kim, S.-K. Detection and Monitoring of Forest Fires Using Himawari-8 Geostationary Satellite Data in South Korea. *Remote Sens.* **2019**, *11*, 271. [CrossRef]
24. Sung, T.; Lee, G.; Kim, D.; Kim, W.; Yang, S.; Im, J. Real-Time Wildfire Monitoring via Geostationary Satellite and Artificial Intelligence: Insights from the March 2025 South Korea Wildfires. *Korean J. Remote Sens.* **2025**, *41*, 565–580. [CrossRef]
25. Lee, D.; Kim, S.-I.; Ahn, D.-S.; Kim, S.-C.; Seo, D.; Choi, M.; Lee, Y.; Kim, J. Machine Learning-Based Near-Real-Time Monitoring of Wildfire Spread Extent Using GK2A and VIIRS. *Korean J. Remote Sens.* **2025**, *41*, 869–881. [CrossRef]
26. Magstadt, S.; Wadhwani, R.; Whitley, M.A.; Martin, T.J.; Sitter, E.; Wei, Y.; O'Connor, C.D. A High Spatial Resolution Daily Fire Perimeter Progression Dataset for Wildfires in the Western United States: 2020–2024 [Dataset]. *Mendeley Data* **2026**, Version 1. [CrossRef]
27. Magstadt, S.; Wadhwani, R.; Whitley, M.A.; Martin, T.J.; Sitter, E.; Wei, Y.; O'Connor, C.D. A High Spatial Resolution Daily Fire Perimeter Progression Dataset for Wildfires in the Western United States: 2020–2024. *Sci. Data* **2026**. *published online*. [CrossRef]
28. National Institute of Forest Science. *Sanbul Jedaero Algi 2025 [Understanding Forest Fires 2025]*; Research Data No. 1156; National Institute of Forest Science: Seoul, Republic of Korea, 2025; p. 30. Available online: https://www.forest.go.kr/kfs/images/data/down/sb_all.pdf (accessed on 5 September 2026). (In Korean)
29. Davies, D.K.; Ilavajhala, S.; Wong, M.M.; Justice, C.O. Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 72–79. [CrossRef]
30. 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.* **2014**, *143*, 85–96. [CrossRef]
31. Giglio, L. *VIIRS/NPP Active Fires 6-Min L2 Swath 375 m V002 (VNP14IMG) [Dataset]*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2024. [CrossRef]
32. Giglio, L. *VIIRS/JPSS1 Active Fires 6-Min L2 Swath 375 m V002 (VJ114IMG) [Dataset]*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2024. [CrossRef]
33. Giglio, L. *VIIRS/JPSS2 Thermal Anomalies/Fire 6-Min L2 Swath 375 m V002 (VJ214IMG) [Dataset]*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2026. [CrossRef]
34. Schmidt, C.C.; Hoffman, J.; Prins, E.; Lindstrom, S. *GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document for Fire/Hot Spot Characterization, Version 2.7*; NOAA: Washington, DC, USA; NESDIS: Silver Spring, MD, USA; STAR: College Park, MD, USA, 2020. Available online: https://www.star.nesdis.noaa.gov/goesr/documents/ATBDs/Enterprise/ATBD_Enterprise_Fire_Hot_Spot_v2.7_2020-10-31.pdf (accessed on 15 July 2026).
35. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T.; et al. An Introduction to Himawari-8/9—Japan's New-Generation Geostationary Meteorological Satellites. *J. Meteorol. Soc. Jpn.* **2016**, *94*, 151–183. [CrossRef]
36. Japan Aerospace Exploration Agency Earth Observation Research Center (JAXA EORC). *Himawari-8 Wild Fire Product (L2WLF), Version 1.00—Product Description*; JAXA EORC: Tsukuba, Japan, 2020. Available online: https://www.eorc.jaxa.jp/ptree/documents/README_H08_L2WLF.txt (accessed on 15 July 2026).
37. Korea Forest Service. National Forest Fire Information System. Available online: https://fd.forest.go.kr/ (accessed on 15 July 2026). (In Korean)
38. Cressie, N.A.C. *Statistics for Spatial Data*, revised ed.; John Wiley & Sons: New York, NY, USA, 1993. [CrossRef]
39. Chilès, J.-P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012. [CrossRef]
40. Oliver, M.A.; Webster, R. A Tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging. *Catena* **2014**, *113*, 56–69. [CrossRef]
41. Silva, V.M.; Coimbra Costa Leite, J.F.; Deutsch, C.V. Kriging Data with Measurement Error: A Review and a Generalized Approach. *Appl. Earth Sci.* **2025**, *134*, 5–17. [CrossRef]
42. Kyriakidis, P.C. A Geostatistical Framework for Area-to-Point Spatial Interpolation. *Geogr. Anal.* **2004**, *36*, 259–289. [CrossRef]

43. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
44. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
45. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
46. Pohjankukka, J.; Pahikkala, T.; Nevalainen, P.; Heikkonen, J. Estimating the Prediction Performance of Spatial Models via Spatial k-Fold Cross Validation. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2001–2019. [CrossRef]
47. Schroeder, W.; Csiszar, I.; Morisette, J. Quantifying the Impact of Cloud Obscuration on Remote Sensing of Active Fires in the Brazilian Amazon. *Remote Sens. Environ.* **2008**, *112*, 456–470. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.