Supplementary Materials

Assessing the Complementary Contributions of Weather and Temporal Information to Wildfire Occurrence Prediction: Evidence from South Korea, Spain, and Canada. Boksoo Choi and Gye-Young Kim.

Table S1. Key differences between Choi and Kim (2026) [11] and the present study.

Aspect	Choi and Kim (2026) [11]	Present study
Research focus	Predictive value of time proxy variables and their complementarity with FWI	Conditional contributions of F/W/T and cross-country replication; country-specific reference contrasts
Data	South Korea; 100 ASOS stations, 2011–2025	Same Korean source data plus independent Spanish (2011–2023) and Canadian (2011–2024) panels
Sample processing	Nearest-station matching (planar distance); 6639 positive station-days; 1:2 train undersampling	Haversine matching in Korea (6643 positives); country-specific matching and preprocessing; 1:2 train undersampling (within each year in Canada)
Predictors	Time (5), Weather (4), FWI (6) and their combinations	F (6), W (6), T (2); Korean K and D; country-specific references
Model training	Eight models (logistic regression, six tree ensembles and MLP); separate deep-learning comparison (MLP and LSTM variants); validation-based tuning; SHAP	Harmonized RF (500 trees, depth 10, min. leaf 20, seed 42); no feature-set tuning
Temporal partitioning	Korea: train 2011–2018; validation 2019–2021; test 2022–2025	Same Korean split; country-specific chronological splits for Spain and Canada
Evaluation and inference	ROC-AUC, Recall, Precision, F1; bootstrap and DeLong tests	Prespecified paired ΔROC-/PR-AUC; clustered bootstrap, Holm correction, calibration; LightGBM check

Note: Both studies use the same Korean source data (7140 fires; 517,448 station-days). Haversine matching reassigned 358 fires (5.01%) relative to the earlier planar method, resulting in 6643 versus 6639 positive station-days. At the level of station-day keys, 6319 positive station-days (95.18%) were shared; 320 were unique to the earlier study and 324 to the present study. Re-scoring the locked test predictions against the earlier labels changed the planned ΔROC-AUC values by at most 0.0005. Country-specific sample processing, matching rules, missing-value handling and temporal splits are given in Table 1 of the main text.

Table S2. Calibration of test probabilities from the harmonized random forest.

Country	Structure	Version	Brier score	Brier skill score	Mean predicted	Observed rate
South Korea	FWT	Raw	0.0866	−5.3370	0.2012	0.0139
	FWT	Prior-corrected	0.0132	0.0322	0.0099	0.0139
	FWT	Validation isotonic	0.0132	0.0329	0.0147	0.0139
	KD	Raw	0.0955	−5.9917	0.2184	0.0139
	KD	Prior-corrected	0.0133	0.0245	0.0113	0.0139
	KD	Validation isotonic	0.0133	0.0278	0.0143	0.0139
Spain	FWT	Raw	0.1230	−2.1067	0.2991	0.0413
	FWT	Prior-corrected	0.0387	0.0228	0.0596	0.0413
	FWT	Validation isotonic	0.0385	0.0285	0.0505	0.0413
	S	Raw	0.1327	−2.3523	0.3092	0.0413
	S	Prior-corrected	0.0396	−0.0008	0.0628	0.0413
	S	Validation isotonic	0.0390	0.0145	0.0521	0.0413
Canada	FWT	Raw	0.1208	−7.5015	0.2823	0.0144
	FWT	Prior-corrected	0.0139	0.0220	0.0145	0.0144
	FWT	Validation isotonic	0.0139	0.0217	0.0136	0.0144
	F_{off}	Raw	0.1235	−7.6938	0.2949	0.0144
	F_{off}	Prior-corrected	0.0140	0.0126	0.0145	0.0144

Country	Structure	Version	Brier score	Brier skill score	Mean predicted	Observed rate
	F_{off}	Validation isotonic	0.0140	0.0124	0.0129	0.0144

Note: Prior correction uses the train sampling ratio β of negative station-days (South Korea 0.02366; Spain 0.12643; Canada 0.02719). For Canada, where negatives were sampled within each year, β is the whole-period ratio, so the correction is approximate. Isotonic regression was fitted on validation data only. The Brier skill score is relative to a constant forecast equal to the validation occurrence rate (South Korea 0.0145; Spain 0.0472; Canada 0.0117). Recalibration does not change ROC-AUC.

Table S3. Operational coverage of official CWFIS FWI information in the full Canadian panel (2011–2024).

Status of official CWFIS FWI	Station-days	Positive station-days	Share of positives
Official information provided	2,406,077	32,079	89.10%
Planned non-production (seasonal operating switch)	1,137,248	3887	10.80%
Missing because of operational gaps	5349	37	0.10%
Full panel	3,548,674	36,003	100.00%

Note: Planned non-production refers to days outside the operational fire season defined by the CWFIS seasonal switch. Shares of positive station-days without official FWI information were 23.85% in Atlantic Canada (New Brunswick, Nova Scotia, Newfoundland and Labrador), 14.73% in the Prairies (Alberta), 8.96% in Quebec and 4.66% in British Columbia.

Table S4. Detailed results of the Spanish sensitivity analyses (harmonized random forest, ΔROC-AUC).

Scenario	Split	Positives	FWT–F	FWT–FT	FWT–FW	SFWT–S
Primary (locked models, original labels)	Validation	18,450	0.0428	0.0216	0.0103	0.0791
	Test	16,137	0.0405	0.0227	0.0114	0.0796
Evaluation excludes fires > 50 km	Validation	17,768	0.0431	0.0215	0.0104	0.0790
	Test	15,599	0.0404	0.0222	0.0116	0.0795
Evaluation excludes fires > 30 km	Validation	15,728	0.0419	0.0209	0.0100	0.0764
	Test	13,801	0.0391	0.0207	0.0118	0.0768
50 km labels: locked models	Validation	17,768	0.0430	0.0215	0.0104	0.0789
	Test	15,599	0.0404	0.0222	0.0116	0.0794
50 km labels: re-fitted models	Validation	17,768	0.0434	0.0220	0.0106	0.0786
	Test	15,599	0.0404	0.0226	0.0118	0.0790
Year-stratified 1:2: re-fitted models	Validation	18,450	0.0433	0.0221	0.0104	0.0788
	Test	16,137	0.0414	0.0228	0.0119	0.0796

Note: Evaluation-exclusion scenarios keep the locked models and remove from evaluation the positive station-days whose matched fires all lie beyond the stated distance. The 50 km label scenarios redefine the outcome: fires farther than 50 km from the nearest station are not assigned, so station-days linked only to such fires become negative, whereas those with at least one fire within 50 km remain positive. These are sensitivity analyses, not corrections of the primary labels; "locked models" evaluates the original models on these labels and "re-fitted models" retrains all models with the same specification and 1:2 undersampling. The year-stratified scenario re-fits all models with 1:2 undersampling within each year, as in Canada, and keeps the original labels. Test confidence intervals are given in Table 4 of the main text; robustness analyses used 2000 date-cluster bootstrap resamples.

Table S5. Exploratory top-capacity capture of positive test station-days in Spain (post hoc).

Top-capacity share	S	F	SF	SF – S
5%	1881 (11.66%)	2171 (13.45%)	3194 (19.79%)	+1313

Top-capacity share	S	F	SF	SF − S
10%	3424 (21.22%)	3831 (23.74%)	5105 (31.64%)	+1681
20%	6103 (37.82%)	6554 (40.62%)	7892 (48.91%)	+1789

Note: Station-days were ranked by the scores of the country-specific LightGBM models for S, F and SF, and positive station-days captured within the top 5%, 10% and 20% of the test sample (391,044 station-days; 16,137 positives) were counted. This descriptive comparison was not prespecified and was not used to select thresholds or models.

Table S6. Test ΔROC-AUC under the harmonized random forest and the LightGBM models, evaluated with the same procedure.

Country	Contrast	Random forest Δ [95% CI]	LightGBM Δ [95% CI]	LightGBM Holm p	Agreement
South Korea	FWT−F	+0.0448 [0.0384, 0.0511]	+0.0445 [0.0373, 0.0519]	0.0010	Yes
	FWT−FT	+0.0167 [0.0134, 0.0202]	+0.0153 [0.0111, 0.0195]	0.0010	Yes
	FWT−FW	+0.0096 [0.0061, 0.0129]	+0.0098 [0.0051, 0.0143]	0.0010	Yes
	FWT−KD	+0.0243 [0.0192, 0.0295]	+0.0249 [0.0189, 0.0313]	0.0010	Yes
	FWT−FWD	+0.0023 [0.0007, 0.0039]	+0.0024 [−0.0005, 0.0053]	0.1032	No
Spain	FWT−F	+0.0405 [0.0365, 0.0447]	+0.0511 [0.0471, 0.0550]	0.0008	Yes
	FWT−FT	+0.0227 [0.0196, 0.0258]	+0.0315 [0.0284, 0.0347]	0.0008	Yes
	FWT−FW	+0.0114 [0.0096, 0.0132]	+0.0172 [0.0148, 0.0195]	0.0008	Yes
	SFWT−S	+0.0796 [0.0745, 0.0847]	+0.0833 [0.0774, 0.0889]	0.0008	Yes
Canada	FWT−F	+0.0337 [0.0304, 0.0371]	+0.0400 [0.0363, 0.0436]	0.0014	Yes
	FWT−FT	+0.0167 [0.0145, 0.0190]	+0.0241 [0.0209, 0.0271]	0.0014	Yes
	FWT−FW	+0.0054 [0.0041, 0.0069]	+0.0077 [0.0056, 0.0097]	0.0014	Yes
	F_{off}WT−F_{off}	+0.0371 [0.0331, 0.0412]	+0.0415 [0.0369, 0.0457]	0.0014	Yes
	F^0T−F^0	+0.0242 [0.0208, 0.0276]	+0.0248 [0.0208, 0.0288]	0.0014	Yes
	F^0W−FW	+0.0001 [−0.0006, 0.0009]	−0.0008 [−0.0021, 0.0004]	0.2192	Yes (neither significant)
	F^0WT−FWT	+0.0020 [0.0015, 0.0025]	+0.0008 [−0.0002, 0.0019]	0.2192	No

Note: Both models were evaluated on the same test station-days with the same procedure: Δ = ROC-AUC(A) − ROC-AUC(B), with 95% intervals from 5000 date-cluster paired bootstrap resamples. Holm-adjusted bootstrap p-values are shown for LightGBM, adjusted across the contrasts listed for each country; random forest values reproduce Table A1. LightGBM was the primary model in the original Spanish analysis and a supplementary model in South Korea and Canada; the Korean LightGBM models were refitted with the locked specification and train sample and reproduced the locked test ROC-AUC of all previously fitted feature sets. Agreement = same direction and same significance decision (95% interval excluding or including zero). The two disagreements (Korean FWT−FWD; Canadian F^0WT−FWT) concern small secondary contrasts on seasonal representation.

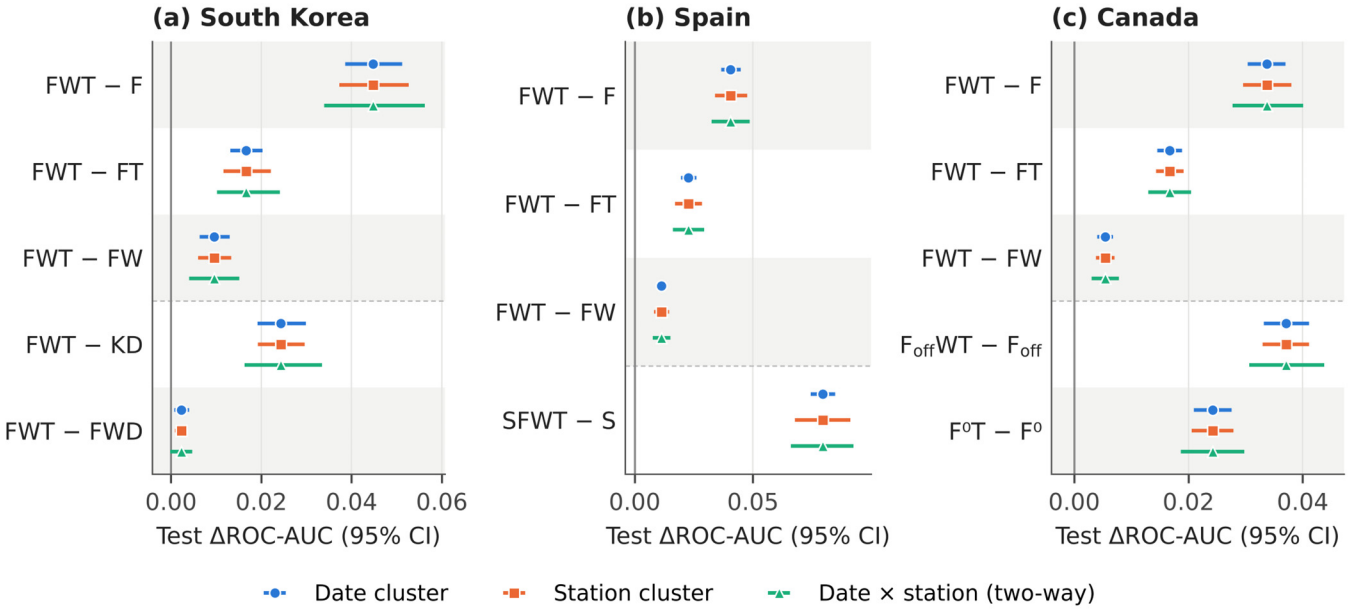


Figure S1. Robustness of the planned test contrasts to the bootstrap cluster unit. Points show ΔROC-AUC under the harmonized random forest; bars show 95% percentile intervals from 2000 paired bootstrap resamples clustered by calendar date, by weather station or by both (two-way pigeonhole bootstrap). Contrasts above the dashed line are common to the three countries; those below are country-specific reference or secondary contrasts.

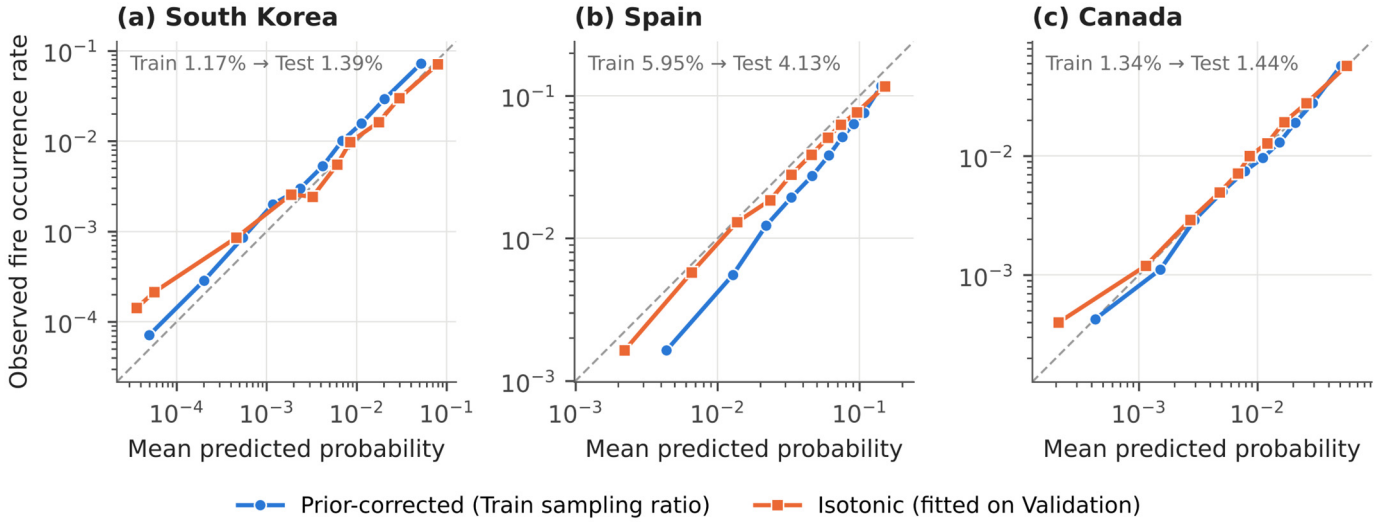


Figure S2. Reliability of FWT probabilities in the test period after prior correction for train undersampling and after isotonic recalibration fitted on validation data. Points are decile means on log–log axes; the dashed line marks perfect calibration. Annotated rates show the change in fire occurrence rate between the train and test periods.