FrostTR: A calibrated and spatially transferable machine-learning framework for 24-hour agricultural frost forecasting across Türkiye

Electronic Supplementary Material - Online Resource 1

Mehmet Tevfik AĞDAŞ | Munzur University | mtagdas@munzur.edu.tr

Supplementary tables and figure

Online Resource 1 contains station-level, fold-level, seed-level, endpoint-sensitivity, model-configuration, and runtime details that support direct audit of the aggregate results reported in the manuscript.

Table S1. Station coordinates, elevation, observation coverage, and strict-QC status.

Station ID	Station	Lat.	Lon.	Elev. (m)	All-hour cov.	Cold cov.	Frost days	Strict QC
170920-99999	Erzincan	39.710	39.527	1153.1	0.959	0.955	1328	Yes
172020-99999	Elazig	38.607	39.291	892.1	0.959	0.955	967	Yes
171700-99999	Van	38.468	43.332	1670.3	0.959	0.955	1296	Yes
172440-99999	Konya	37.979	32.562	1030.5	0.959	0.955	974	Yes
172800-99999	Diyarbakir	37.894	40.201	686.1	0.959	0.955	640	Yes
172400-99999	Isparta	37.785	30.582	990.6	0.944	0.937	644	Yes
171240-99999	Eskisehir	39.784	30.582	786.7	0.941	0.937	1093	Yes
170950-99999	Erzurum Bölge	39.900	41.283	1869.0	0.917	0.932	1588	Yes
171300-99999	Ankara Central	39.950	32.883	891.0	0.940	0.932	607	Yes
170900-99999	Sivas	39.750	37.017	1285.0	0.920	0.912	1526	Yes
171900-99999	Afyon	38.726	30.601	1008.9	0.829	0.831	779	Yes
172370-99999	Denizli	37.783	29.083	425.0	0.789	0.795	146	Yes
171750-99999	Ayvalik	39.300	26.700	4.0	0.719	0.729	64	No
172460-99999	Karaman	37.200	33.217	1023.0	0.680	0.706	683	No
171400-99999	Yozgat	39.817	34.800	1298.0	0.677	0.698	976	No
170840-99999	Corum	40.550	34.950	776.0	0.669	0.690	839	No
171930-99999	Nevsehir	38.617	34.700	1260.0	0.668	0.690	681	No
171650-99999	Tunceli	39.117	39.550	981.0	0.680	0.687	720	No
170500-99999	Edirne	41.667	26.567	51.0	0.645	0.679	380	No
172380-99999	Burdur	37.667	30.333	967.0	0.661	0.670	470	No
172030-99999	Bingol	38.867	40.500	1177.0	0.658	0.662	871	No
171550-99999	Kutahya	39.417	29.967	969.0	0.639	0.659	753	No
172500-99999	Nigde	37.967	34.683	1210.0	0.617	0.644	755	No
172620-99999	Kilis	36.717	37.117	638.0	0.569	0.591	45	No
170700-99999	Bolu	40.733	31.600	743.0	0.550	0.563	759	No
170880-99999	Gumushane	40.467	39.467	1219.0	0.507	0.515	842	No
170240-99999	Inebolu	41.983	33.783	64.0	0.439	0.450	66	No
172850-99999	Hakkari	37.567	43.767	1728.0	0.436	0.437	785	No
170590-99999	Kumkoy	41.250	29.033	30.0	0.439	0.436	63	No
170460-99999	Ardahan	41.117	42.717	1829.0	0.427	0.432	1399	No

Table S1 provides the station-level foundation of the entire study by listing coordinates, elevation, coverage metrics, and strict-QC status for all 30 assembled stations. This level of detail allows the reader to audit the spatial and climatological diversity of the network directly and to understand why only a subset satisfied every pre-specified quality threshold. The table also supports the interpretation of later LOSO and strict-QC analyses, because those results depend fundamentally on the heterogeneity summarized here. In that sense, Table S1 is not merely descriptive; it is part of the evidence base showing that the study’s strong results were obtained across a genuinely diverse and challenging station network.

Table S2. 2025 analytical-calendar, observed-temperature support, model-side imputation, and endpoint availability by station.

Station ID	Rows	Full calendar	Observed temperature ratio	Mean model-side imputation	Observed-target ratio	Completed-target ratio
170240-99999	8760	True	0.626	0.573	0.626	0.997
170460-99999	8760	True	0.623	0.589	0.616	0.997
170500-99999	8760	True	0.626	0.561	0.627	0.997
170590-99999	8760	True	0.000	1.000	0.000	0.997
170700-99999	8760	True	0.629	0.557	0.633	0.997
170840-99999	8760	True	0.628	0.555	0.634	0.997
170880-99999	8760	True	0.611	0.593	0.604	0.997
170900-99999	8760	True	0.639	0.628	0.640	0.997
170920-99999	8760	True	0.639	0.642	0.640	0.997
170950-99999	8760	True	0.628	0.581	0.631	0.997
171240-99999	8760	True	0.629	0.620	0.626	0.997
171300-99999	8760	True	0.633	0.534	0.636	0.997
171400-99999	8760	True	0.627	0.590	0.622	0.997
171550-99999	8760	True	0.632	0.541	0.634	0.997
171650-99999	8760	True	0.633	0.537	0.635	0.997
171700-99999	8760	True	0.639	0.636	0.640	0.997
171750-99999	8760	True	0.627	0.559	0.623	0.997
171900-99999	8760	True	0.634	0.597	0.637	0.997
171930-99999	8760	True	0.629	0.589	0.629	0.997
172020-99999	8760	True	0.639	0.615	0.640	0.997
172030-99999	8760	True	0.633	0.586	0.638	0.997
172370-99999	8760	True	0.631	0.538	0.634	0.997
172380-99999	8760	True	0.632	0.532	0.635	0.997
172400-99999	8760	True	0.634	0.531	0.637	0.997
172440-99999	8760	True	0.639	0.632	0.640	0.997
172460-99999	8760	True	0.632	0.601	0.635	0.997
172500-99999	8760	True	0.624	0.587	0.624	0.997
172620-99999	8760	True	0.609	0.582	0.597	0.997
172800-99999	8760	True	0.639	0.618	0.641	0.997
172850-99999	8760	True	0.623	0.588	0.620	0.997

Table S2 reports the 2025 analytical calendar using metrics that remain interpretable at station level. The previously reported simultaneous all-core raw completeness indicator was removed because it was zero for every station and therefore provided no discriminatory information; it reflected the overly strict requirement that all heterogeneous raw channels be present at the same timestamp, not the absence of all inputs. The revised table reports observed-temperature support, mean model-side imputation, observation-supported target availability, and completed-target availability separately. Observed-temperature ratios of approximately 0.61–0.64 at most contributing stations indicate substantial reporting gaps and justify the primary endpoint restriction. The table does not establish that those gaps are random, and possible hour-, season-, station-, or temperature-dependent missingness remains a limitation. Kumköy is retained for source auditing but did not contribute observation-supported primary endpoints.

Table S3. Full station-by-station leave-one-station-out results with rare-event support audit.

Station	Test n	Positive n	Prevalence (95% Wilson CI)	F1 (descriptive)	AUPRC	Brier	MAE (°C)	Support flag
Inebolu	1,168	55	0.0471 (0.0364–0.0608)	0.5248	0.8738	0.0481	2.005	Moderate (50–99)
Edirne	1,162	195	0.1678 (0.1474–0.1904)	0.6977	0.8450	0.0624	1.666	Adequate (≥100)

Station	Test n	Positive n	Prevalence (95% Wilson CI)	F1 (descriptive)	AUPRC	Brier	MAE (°C)	Support flag
Ayvalik*	1,152	26	0.0226 (0.0154–0.0329)	0.7222	0.8871	0.0149	1.532	Low (<50 positives)
Diyarbakir	1,199	431	0.3595 (0.3328–0.3870)	0.7631	0.9583	0.0989	1.799	Adequate (≥100)
Denizli	1,186	82	0.0691 (0.0560–0.0850)	0.7639	0.9007	0.0203	1.360	Moderate (50–99)
Kilis*	1,079	40	0.0371 (0.0273–0.0501)	0.7742	0.9206	0.0137	2.160	Low (<50 positives)
Isparta	1,188	280	0.2357 (0.2124–0.2607)	0.7963	0.8919	0.0672	1.350	Adequate (≥100)
Nevsehir	1,168	389	0.3330 (0.3066–0.3606)	0.8067	0.9237	0.0920	2.337	Adequate (≥100)
Burdur	1,186	240	0.2024 (0.1805–0.2262)	0.8161	0.9157	0.0512	1.309	Adequate (≥100)
Ankara Central	1,186	355	0.2993 (0.2739–0.3260)	0.8276	0.9171	0.0740	1.663	Adequate (≥100)
Eskisehir	1,170	521	0.4453 (0.4170–0.4739)	0.8309	0.9483	0.0985	1.996	Adequate (≥100)
Afyon	1,188	441	0.3712 (0.3442–0.3991)	0.8354	0.9449	0.0774	1.494	Adequate (≥100)
Corum	1,185	466	0.3932 (0.3658–0.4214)	0.8388	0.9454	0.0814	1.614	Adequate (≥100)
Kutahya	1,181	456	0.3861 (0.3588–0.4142)	0.8483	0.9550	0.0784	1.631	Adequate (≥100)
Karaman	1,186	439	0.3702 (0.3431–0.3980)	0.8542	0.9342	0.0795	1.667	Adequate (≥100)
Bolu	1,181	435	0.3683 (0.3413–0.3962)	0.8636	0.9503	0.0694	1.470	Adequate (≥100)
Elazig	1,195	527	0.4410 (0.4131–0.4693)	0.8688	0.9642	0.0765	1.961	Adequate (≥100)
Tunceli	1,187	445	0.3749 (0.3478–0.4028)	0.8799	0.9545	0.0641	1.429	Adequate (≥100)
Nigde	1,156	454	0.3927 (0.3650–0.4212)	0.8825	0.9633	0.0632	1.641	Adequate (≥100)
Konya	1,195	521	0.4360 (0.4081–0.4643)	0.8842	0.9713	0.0647	1.639	Adequate (≥100)
Yozgat	1,180	526	0.4458 (0.4176–0.4743)	0.9086	0.9720	0.0591	1.567	Adequate (≥100)
Gumushane	1,099	495	0.4504 (0.4212–0.4799)	0.9163	0.9814	0.0507	1.598	Adequate (≥100)

Station	Test n	Positive n	Prevalence (95% Wilson CI)	F1 (descriptive)	AUPRC	Brier	MAE (°C)	Support flag
Bingol	1,190	521	0.4378 (0.4099–0.4662)	0.9381	0.9880	0.0352	1.280	Adequate (≥100)
Van	1,195	633	0.5297 (0.5014–0.5579)	0.9428	0.9896	0.0462	1.709	Adequate (≥100)
Erzincan	1,195	666	0.5573 (0.5290–0.5853)	0.9452	0.9927	0.0402	1.928	Adequate (≥100)
Ardahan	1,136	714	0.6285 (0.6000–0.6561)	0.9457	0.9916	0.0479	2.911	Adequate (≥100)
Erzurum Bölge	1,178	718	0.6095 (0.5813–0.6370)	0.9620	0.9954	0.0347	1.808	Adequate (≥100)
Hakkari	1,145	512	0.4472 (0.4186–0.4761)	0.9631	0.9956	0.0252	1.436	Adequate (≥100)
Sivas	1,195	679	0.5682 (0.5399–0.5960)	0.9732	0.9941	0.0313	1.875	Adequate (≥100)

Table S3 presents the complete LOSO results together with the number of observed positive windows and Wilson 95% intervals for prevalence. The support audit makes clear that Ayvalık (26 positives) and Kilis (40 positives) are low-support folds, while Inebolu contains 55 positives and remains relatively sparse. Because F1 is nonlinear and valid block-bootstrap intervals require prediction-level fold records, Wilson intervals are shown for prevalence rather than incorrectly applied to F1. The F1 values for stations marked with an asterisk are therefore descriptive and should not be interpreted with the same precision as estimates from stations containing hundreds of positive windows. This revision makes the spatial-transfer heterogeneity more explicit and prevents low-prevalence point estimates from implying unwarranted certainty.

Table S4. Individual five-seed results.

Seed	Threshold	F1	AUPRC	Brier	ECE	MAE (°C)
13	0.473205	0.892918	0.968383	0.053578	0.005791	1.586836
29	0.481487	0.892428	0.968537	0.053344	0.006389	1.582612
42	0.494153	0.890828	0.968092	0.053811	0.006564	1.586362
71	0.473721	0.891223	0.968316	0.053786	0.006507	1.587099
101	0.464527	0.893266	0.968297	0.053555	0.005677	1.587882

Table S4 reports the individual five-seed fits used to construct the mean, standard deviation, and t intervals in Table 8. The close agreement among thresholds and performance measures confirms that the principal results are reproducible across the tested random states.

Table S5. Secondary completed-target sensitivity analysis.

Measure	Result
Test windows	57,360
Threshold	0.4902
Accuracy	0.9441
Precision	0.9146
Recall	0.9048
F1	0.9097
MAE	1.258 °C
RMSE	1.761 °C
Mean target imputed fraction	0.4061
Fully observed target share	0.4652

Table S5 reports the completed-target sensitivity analysis separately from the observation-supported primary endpoint. Its higher apparent skill is expected to be optimistic because 40.6% of the future target interval was completed on average, reducing missingness and smoothing part of the evaluation target. These values quantify sensitivity to endpoint definition only; they must not be used as a performance comparison, as evidence of superior real-world forecasting, or as a substitute for the observation-supported results.

Table S6. Final XGBoost configuration and test-isolation rules.

Component	Final specification	Role/test-isolation rule
Training period	2015–2023	All model fitting and training-only sampling
Validation period	2024	Early stopping and binary threshold selection only
Primary test period	2025	No fitting, tuning, balancing, or threshold optimization
Main training cap	180,000 windows	Applied only to the pre-2024 training pool
Fold training cap	60,000 windows	Applied to LOSO, rolling-origin, and overlap repetitions
Estimators	450 main; 220 repeated folds	Early stopping limited effective tree count
Maximum depth	6	Held constant across reported experiments
Learning rate	0.04	Held constant across reported experiments
Minimum child weight	2	Held constant across reported experiments
Row / column subsampling	0.85 / 0.85	Training-only regularization
L1 / L2 regularization	0.05 / 1.0	Held constant across reported experiments
Tree construction	Histogram; CUDA when available, CPU fallback	Computational implementation
Binary class weighting	None	No scale_pos_weight multiplier
Severity weighting	Inverse-frequency sample weights	Applied only to four-class severity
Decision threshold	0.4942	Selected by 2024 validation F1; fixed for 2025

Table S6 makes the model specification and the separation between development and evaluation explicit. The 2025 endpoint was not used for hyperparameter choice, early stopping, class balancing, or threshold selection. Repeated validation experiments retained the same substantive configuration and reduced only the training and estimator caps for computational feasibility, improving comparability across the comprehensive analyses.

Table S7. Computational environment used for the strict-QC and extended-network sensitivity analyses.

Component	Reported environment	Interpretation
Backend	Google Compute Engine GPU runtime	Cloud notebook environment used for the reported sensitivity analyses
Python environment	Python 3	Executable notebook environment
System memory	51.0 GB total; 37.7 GB active at capture	Snapshot during the memory-intensive analysis
GPU memory	15.0 GB total; 0.2 GB active at capture	GPU allocation at the recorded moment
Disk	235.7 GB total; 47.4 GB used at capture	Workspace storage snapshot

Table S7 documents the computational environment recorded during the memory-intensive sensitivity analysis. The values are a runtime snapshot rather than a standardized peak-memory benchmark and should not be interpreted as the maximum allocation over the full workflow.