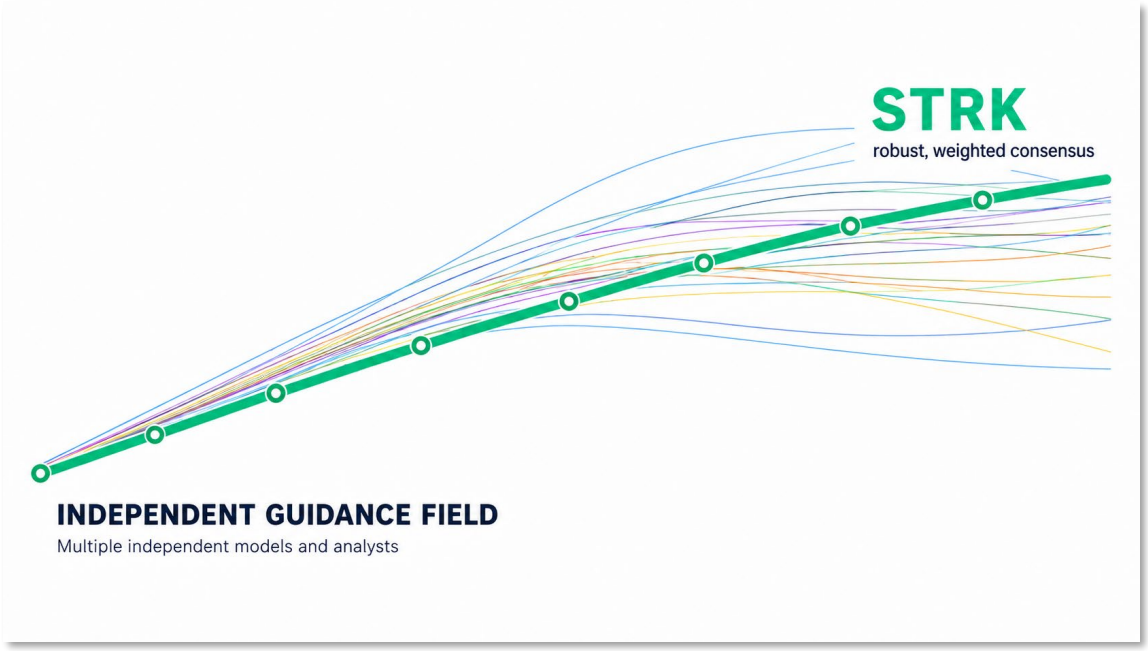
Statistically-Tuned Real-time Integrated Kinematic Ensemble

STRIKE SuperEnsemble

Operational formulation, leakage-resistant walk-forward verification, and a 2026 prospective test plan for robust multimodel tropical cyclone track and intensity consensus guidance.

Presented by Adam Hazell



METHOD	MODEL MODULE	TRAINER
2.1.0	1.6.4	1.11.0
IMPACT CORRIDOR	**REPORT DATE**	**STATUS**
1.7.1	4 August 2026	Experimental / Prospective

Scientific and operational documentation for external review

strikesuperensemble.org

Report Status and Scope

Submission position

- This report documents the current implementable STRIKE SuperEnsemble methodology and presents finalized 2019–2025 walk-forward evidence as a basis for scientific review, independent verification, and prospective evaluation.
- It does not claim operational adoption, endorsement by NOAA/NHC, or superiority to the official forecast. OFCL is treated as a separate human-integrated benchmark, not as a member of the STRIKE blend.
- The 2026 season is the prospective phase. Real-time forecasts will be preserved as cumulative, publicly accessible STRIKE-only ATCF A-decks so that favorable and unfavorable cases can be evaluated after the season.
- The public tracker and Impact Corridor are experimental decision-support interfaces. Official forecasts, watches, warnings, and local emergency-management instructions remain authoritative.

Abstract

STRIKE SuperEnsemble is a deterministic, auditable multimodel tropical cyclone guidance system that constructs track and intensity forecasts from independent numerical, statistical-dynamical, ensemble-mean, and machine-learning aids available to a forecast cycle. Method 2.1.0 publishes three four-character ATCF techniques: STRK, the primary basin- and horizon-dependent consensus; STRR, a raw performance-weighted control; and STRE, a fixed equal-weight control. A proposed AI-only STRA technique is a separate 2027 development concept and is not part of the method or verification reported here. The live builder applies motion-relative member bias correction, robust median/MAD screening, availability-driven renormalization, a 45% member dominance cap, spherical position averaging, and immutable issuance manifests. Existing official and consensus aids are excluded from the member blend and retained only for contextual display and verification.

The finalized evidence archive contains 13,365,524 verification rows from 275 Atlantic and eastern North Pacific storms during 2019–2025. A cumulative season-blocked replay begins from a fixed equal-weight seed and advances the training boundary only after each completed season, preventing direct information leakage from future storms. In 112 season–basin–forecast-hour track ranking cells constructed with NHC-style homogeneous rosters, STRK placed in the top five automated aids in 99 cells and ranked first in 6; STRR placed in the top five in 93 cells and ranked first in 17; STRE placed in the top five in 100 cells and ranked first in 11. These results show a consistently competitive consensus family but also show that more complex weighting does not automatically dominate the equal-weight control.

On exact pairwise cases, STRK track errors are lower than OFCI by 12.9%, 10.3%, and 7.6% in the short-, medium-, and long-range bands, respectively, while differences from OFCL are approximately neutral (−0.4%, −0.4%, and +0.2%). STRK is slightly worse than HCCA at short and medium range and slightly better at long range; it is modestly better than TVCA/TVCN across the three bands. Intensity results are mixed against corrected consensus guidance and close to OFCL. Because the historical results remain developmental and because candidate weights and bias fields are review artifacts rather than auto-promoted operational parameters, the decisive evidence must come from frozen, pre-issued 2026 forecasts.

During the 2026 season, cumulative STRIKE-only ATCF decks will be publicly available without an access request, together with a public cyclone tracker. The tracker will also expose the proprietary, experimental STRIKE Impact Corridor: a location-query tool that triangulates an NHC source score, a STRK source score, and a Grand Ensemble source score into a 0–10 guidance-interaction index and confidence estimate. The Impact Corridor is intended to help users interrogate where converging guidance suggests potential real-world effects well before landfall; it is not a calibrated damage, loss, surge, rainfall, or outage prediction and will be revised as 2026 evidence accumulates.

Key findings and 2026 commitments

Table 1. Executive evidence summary and prospective commitments.

Area	Current evidence	2026 commitment
Historical archive	275 storms; 13,365,524 verification rows; 2019–2025	Freeze archive and retain reproducible manifests
Track ranking	STRK top five in 99/112 automated-aid cells	Verify only pre-issued real-time records
Operational robustness	Minimum 5 contributors; 2.75×MAD screen; 45% cap	Archive both issued and failed/non-issued cycles
Public ATCF	Cumulative STRK/STRR/STRE storm decks implemented	Publish plain-text, gzip, and manifest files
Public tracker	No access request required	Expose cycle-pure guidance and provenance
Impact Corridor	Triangulated v1; 256×256 scored grid	Test calibration, usability, and failure modes throughout 2026

Report map

Section	Purpose
1–3	Scientific rationale, system definition, and exact operational mathematics
4–6	Training design, verification contract, and 2019–2025 results
7	Frozen 2026 prospective forecast and public ATCF plan
8	Impact Corridor formulation, intended use, and validation questions
9–10	Limitations and a concrete research-to-operations evaluation pathway
Appendices	Constants, member catalog, benchmark tables, reproducibility record, and references

1. Scientific Rationale and Method Lineage

A trained consensus between simple averaging and unconstrained optimization.

1.1 Tropical cyclone guidance as a multimodel problem

Tropical cyclone track and intensity forecasting is a changing multimodel decision problem. Global deterministic models, regional hurricane models, ensemble means, statistical-dynamical intensity techniques, and machine-learning systems differ in initialization, resolution, physics, latency, forecast horizon, and failure mode. Their errors are neither stationary nor perfectly independent. A simple mean can suppress random error, but it gives the same influence to members with different historical skill and availability. An unconstrained statistical fit can produce a lower retrospective error while becoming fragile when the sample is small, the member pool changes, or correlated systems are over-weighted.

STRIKE is designed for the operational space between those extremes. Its target is a transparent consensus that: uses only information available to the current cycle; differentiates members by basin, component, and lead-time band; corrects persistent, historically supported tendencies; resists individual departures from the current guidance cluster; degrades cleanly when a member is missing; and leaves an immutable record that can be independently scored.

1.2 Relationship to the Florida State Superensemble

The term “superensemble” as used in this work is rooted in the Florida State University work of Krishnamurti and collaborators. The classical formulation estimates multiple-regression coefficients during a training period and applies those coefficients to member departures from their own climatological means in a forecast period (Krishnamurti et al. 1999, 2000; Williford et al. 2003). Its enduring contribution is methodological: separate training from forecasting, estimate member-specific coefficients that minimize collective error over the training sample, and verify the combined system against the same truth data used to evaluate its members.

$$O'_j = \Sigma_i\, a_i(F_{ij} - \bar{F}_i) + \varepsilon_j \tag{1}$$

$$S = \bar{O} + \Sigma_i\, a_i(F_i - \bar{F}_i) \tag{2}$$

O is the verifying observation, F is a member forecast, a is a trained coefficient, and S is the superensemble forecast.

STRIKE is not a direct implementation of this regression. The live system uses externally versioned basin/horizon weights, recency-weighted and shrunk motion-relative bias estimates, robust member screening, capped normalization, and spherical averaging. The word “superensemble” is therefore used in the broader operational sense of a trained, performance-differentiated multimodel ensemble. This distinction is essential for reproducibility and for fair comparison with classical FSSE methods.

1.3 Relationship to HCCA and HFIT

The HFIP Corrected Consensus Approach (HCCA) and the Hurricane Track Fit (HFIT) model are closer contemporary comparators. HCCA corrects member tendencies and combines track and intensity guidance using a trained consensus framework. HFIT fits forecast-hour-specific latitude and longitude changes with interpretable multilinear coefficients and evaluates the result with homogeneous samples and cross-validation (Simon et al. 2018; Ginis and Marchok 2022). STRIKE shares the goals of bias correction, interpretability, and timely member selection, but differs in four important respects:

- STRIKE uses constrained nonnegative performance weights rather than unconstrained direct-fit latitude/longitude regression coefficients.
- Track and intensity use separate member pools, weights, bias fields, and robust screening.
- Member availability is handled explicitly through issuance gates, renormalization, and a post-renormalization dominance cap.
- The primary retrospective experiment is a cumulative season-blocked replay using the production builder, rather than a single fitted cross-validation model.

> **Design objective**
>
> STRIKE is intended to be an independent reference aid in the tropical cyclone forecasting toolbox—not a replacement for the official forecast and not a recursively blended derivative of existing consensus aids.

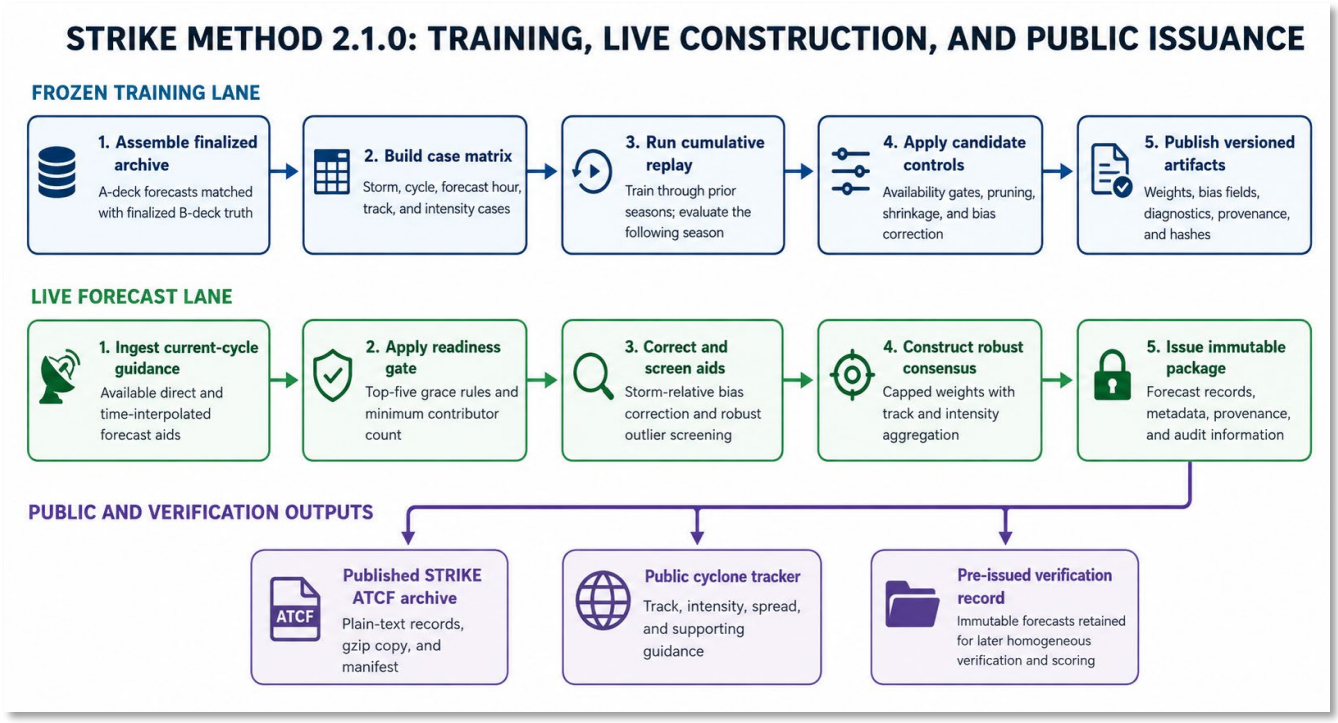


Figure 1. End-to-end architecture of Method 2.1.0. The training lane produces versioned candidate artifacts; the live lane constructs cycle-pure forecasts; the public lane preserves ATCF and verification records.

2. Current Operational System

Techniques, member eligibility, readiness controls, and ATCF publication.

2.1 Published techniques

Table 2. Current Method 2.1.0 ATCF techniques.

Technique	Operational role	Coefficient mode	Interpretive purpose
STRK	STRIKE Primary	Adopted basin/horizon weights	Primary experimental forecast; constrained performance weighting with walk-forward adoption and baseline retention.
STRR	STRIKE Raw	Raw cumulative performance weights	Control that isolates the direct historical weighting signal while using the same correction and production machinery.
STRE	STRIKE Equal	Fixed equal weights	Control that isolates the value of multimodel averaging from differential weighting.

Technique naming boundary

The current source code and public ATCF builder publish STRK, STRR, and STRE. The website concept STRA ("STRIKE AI") is planned for 2027 development using 2026 AI-guidance performance and is not part of this report's method, member pool, or verification.

2.2 Independent member pool

The configured member universe is deliberately restricted to independent forecast aids. Existing official and consensus techniques—including OFCL/OFCI, HCCA, TVCN, IVCN, FSSE, and related products—are retained for display and verification but are not allowed to enter the STRIKE blend. Individual GEFS members are also excluded so that a large correlated family cannot dominate merely by member count.

Table 3. Operational member universe by forecast component.

Component	Eligible families in current configuration	Excluded from blend
Track	EMNI; UKXI; HFAI; HFBI; GDMI; AVNI; CMCI; HMNI; HWFI; CTCI; NVGI	OFCL/OFCI; HCCA; TVCN/TVCA; FSSE; individual GEFS members
Intensity	Track-capable families where intensity is present, plus DSHP, LGEM, SHIP, and ICON	OFCL/OFCI; HCCA; IVCN; FSSE and other consensus products

EMNI is the operational ECMWF ensemble-mean product of record. The historical deterministic ECMWF aid EMXI remains a comparator.

2.3 Core operational constants

Table 4. Implemented constants in the current live builder.

Control	Implemented value	Operational reason
Forecast grid	0–120 h every 12 h	Produces 11 forecast points and aligns the three STRIKE variants.
Horizon cells	SHORT 0–48; MEDIUM 60–96; LONG 108–120 h	Allows member skill and bias behavior to change with lead time.
Minimum retained contributors	5	Suppresses low-diversity consensus points.
Top required contributors	5	Defines the highest-priority readiness set.
Top-five grace period	2 polling passes	Waits longer for high-value missing guidance.

Control	Implemented value	Operational reason
Additional-member grace	1 polling pass	Allows lower-ranked members a limited arrival window.
Maximum normalized weight	0.45	Prevents one surviving member from becoming the forecast.
Track outlier rule	2.75× distance MAD plus tau-dependent floor	Rejects extreme spatial departures without over-screening tight clusters.
Intensity outlier rule	2.75× intensity MAD with 25-kt floor	Rejects extreme VMAX departures while preserving plausible spread.
Issuance gate	Track component	A valid track can issue when intensity is absent.

2.4 Cycle purity, immutability, and public ATCF decks

The NHC public A-deck is cumulative by storm. The collector conditionally retrieves the upstream deck, normalizes its content, calculates source and cycle SHA-256 identities, and writes only changed cycle slices. The production builder then creates separate issued packages for STRK, STRR, and STRE. Guidance from a later initialization is never inserted into an earlier cycle.

For each storm, the public publisher rebuilds a cumulative STRIKE-only A-deck from issued packages of record. Records are de-duplicated by cycle, technique, and forecast hour; output is written as an ASCII .dat file, a deterministic gzip copy, and a JSON manifest containing record count, cycle count, source package identities, byte count, and SHA-256. This structure is intentionally familiar to users of NHC ATCF archives and provides a stable object for independent verification.

2026 public-access commitment

- Cumulative STRK/STRR/STRE ATCF decks will be accessible publicly without an access request.
- Late-arriving upstream guidance may improve future cycles but will not rewrite the scientific meaning of an already issued forecast.
- The archive will preserve missing and failed cycles, not only successful forecasts, so issuance reliability can be scored alongside forecast error.

3. Mathematical Formulation

Bias correction, robust screening, capped weighting, track geometry, and intensity construction.

3.1 Basin and horizon selection

$$H(\tau) = \text{SHORT}, 0 \le \tau \le 48;\ \text{MEDIUM}, 60 \le \tau \le 96;\ \text{LONG}, 108 \le \tau \le 120,\ \tau \in \{0, 12, \ldots, 120\} \quad (3)$$

Atlantic cases use the AL profile; eastern and central North Pacific cases use the EP profile. The component-specific weight vector and bias field are selected independently for track and intensity at every forecast hour. The horizon boundaries were selected both to maintain adequate training samples and to limit abrupt coefficient-regime transitions. Because forecasts are issued on a 12-hour grid, the 48–60 h and 96–108 h transition intervals reduce visible "horizon snap" while preserving distinct short-, medium-, and long-range weighting regimes.

3.2 Historical motion-relative correction

Track bias is represented in a verifying-motion coordinate frame. Positive along-track error is ahead of the verifying motion; positive cross-track error is to the right. Intensity bias is signed forecast VMAX minus verifying VMAX. Bias coefficients are recency-weighted, trimmed in both tails, and shrunk by sample count before application. Operational correction subtracts the supported bias from the member forecast.

$$e_{\parallel} = d\cos(\beta_{error} - \beta_{motion});\quad e_{\perp} = d\sin(\beta_{error} - \beta_{motion}) \quad (4)$$

$$\hat{b}^{shrunk}_{mch} = \hat{b}^{trimmed}_{mch} \times N_{mch}/(N_{mch} + K_c) \quad (5)$$

3.3 Robust track screening

At a forecast hour, the member cluster center is the component-wise median latitude and longitude. Each member's great-circle distance from that center is calculated. The screening threshold is the larger of 2.75 times the distance median absolute deviation and a forecast-hour-dependent floor, preventing the algorithm from rejecting realistic members merely because the central cluster is unusually tight.

$$\tilde{c}_\tau = [\text{median}_{i\in A}(\phi_i), \text{median}_{i\in A}(\lambda_i)] \quad (6)$$

$$\tilde{d} = \text{median}_{i\in A}(d_i);\quad MAD_d = \text{median}_{i\in A}\ |d_i - \tilde{d}| \quad (7)$$

$$T_d(\tau) = \max\{2.75\ MAD_d, F(\tau)\};\quad \text{retain } i \text{ when } d_i \le T_d(\tau) \quad (8)$$

A track point is withheld when fewer than five contributors survive. The reason is recorded as a tau-level failure rather than filled by persistence or a future cycle.

3.4 Availability renormalization and dominance cap

Configured weights are restricted to the members available and retained at a forecast hour, then normalized. If any normalized member exceeds 0.45, the excess is iteratively redistributed among the remaining members in proportion to their uncapped weights. The result remains a weighted consensus even in degraded-member conditions.

$$u_i = w_i / \Sigma_{j\in A}\ w_j \quad (9)$$

$$0 \le \hat{w}_i \le 0.45;\quad \Sigma_{i\in A}\ \hat{w}_i = 1 \quad (10)$$

3.5 Spherical track mean

A latitude/longitude arithmetic mean is inappropriate near longitude discontinuities and introduces geometry error at high latitudes. STRIKE converts each retained member position to a unit Cartesian vector, computes the weighted vector mean, and converts the result back to latitude and longitude.

$$x = \Sigma_{i\in A}\ \hat{w}_i \cos\phi_i \cos\lambda_i;\quad y = \Sigma_{i\in A}\ \hat{w}_i \cos\phi_i \sin\lambda_i;\quad z = \Sigma_{i\in A}\ \hat{w}_i \sin\phi_i \quad (11)$$

$$\hat{\phi} = \text{atan2}(z, \sqrt{(x^2 + y^2)});\quad \hat{\lambda} = \text{atan2}(y, x) \quad (12)$$

3.6 Intensity construction

Intensity is screened separately from track. Members farther than max(2.75×MAD, 25 kt) from the median VMAX are rejected. The retained weighted mean is rounded to the nearest 5 kt. Minimum sea-level pressure is averaged when sufficient pressure values are available from retained intensity contributors.

$$T_V = max\{2.75\ MAD_V,\ 25\ kt\};\quad retain\ i\ when\ |V_i - \tilde{V}| \le T_V \tag{13}$$

$$\hat{V} = 5 \times round[(\Sigma_{i \in A}\ \hat{w}_i\ V_i)/5] \tag{14}$$

4. Training, Coefficient Governance, and Leakage Control

The model must earn complexity against a stable baseline before a candidate is adopted.

4.1 Case matrix and recency weighting

Each finalized historical case is indexed by storm, initialization cycle, and forecast hour. Eligible member forecasts are matched to best-track truth, producing great-circle track error, signed intensity error, and motion-relative along- and cross-track errors. Duplicate aid/case records are resolved before training. Only configured independent aids are eligible in a basin/component/horizon cell.

For candidate performance weights, a member quality score is proportional to inverse recency-weighted mean absolute error raised to α=1.2, multiplied by availability raised to 0.25. A three-year half-life is used by the supplied trainer configuration. Candidate weights are blended 50/50 with the baseline, weights below 2.5% are pruned, GDMI is capped at 10% (to prevent overfitting due to recent aid success), and at least three families must remain in the training score matrix.

$$qi = (1/MAEi)^{1.2} \times availabilityi^{0.25} \quad (15)$$

$$wcandidate = normalize[0.50\ wperformance + 0.50\ wbaseline] \quad (16)$$

4.2 Season-blocked adoption gate

The optimizer is tested one completed season at a time. For test year *y*, candidate weights are derived only from seasons before *y* and scored against the fixed baseline on year *y*. A candidate cell is adopted only when aggregate walk-forward skill is at least +0.25% and at least 50% of tested seasons are positive. Otherwise the baseline is retained. This is intentionally conservative because very small historical differences are not stable evidence of a superior operational configuration.

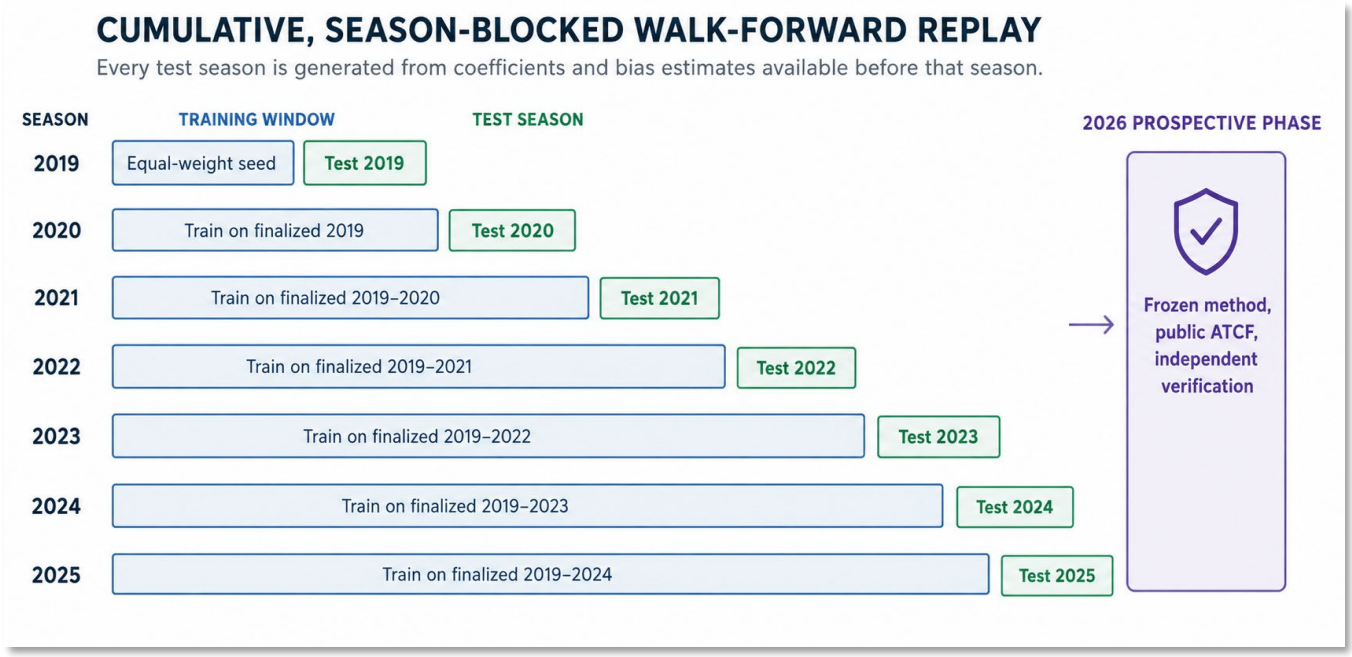


Figure 2. Cumulative season-blocked replay. The training boundary moves only after a season is finalized; 2026 is reserved for prospective evidence.

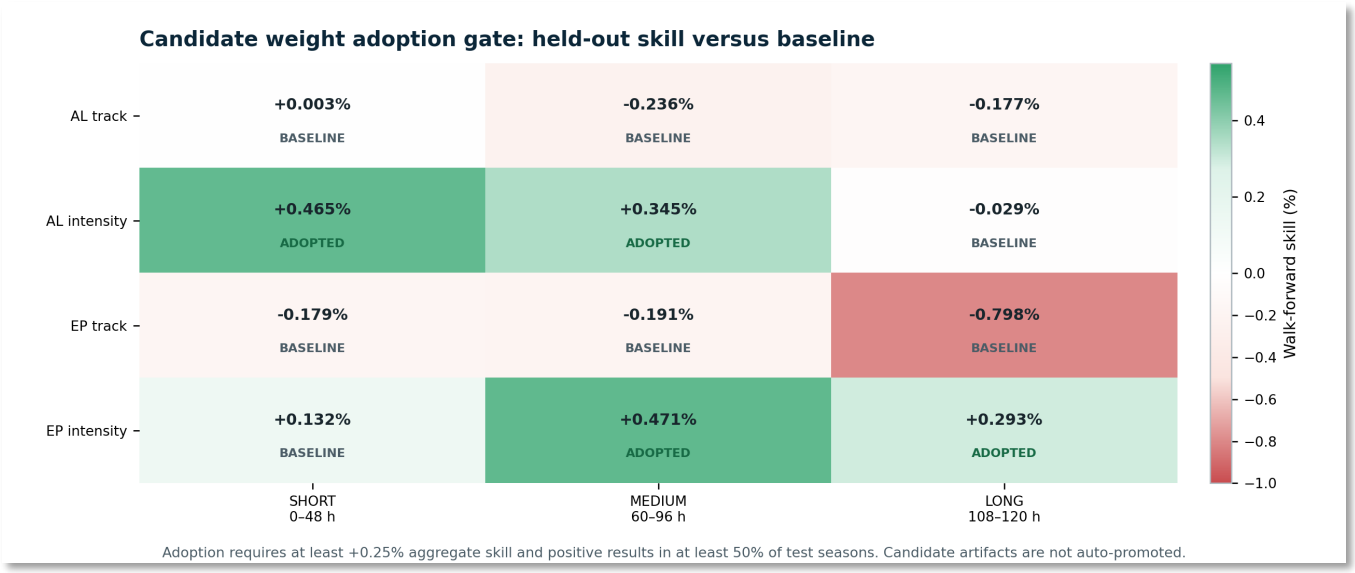


Figure 3. Current candidate weight adoption results. All track cells retained the baseline; selected intensity cells cleared the minimum gate. Values are small by design and the candidate file is not auto-promoted.

The current diagnostics are scientifically informative. They show that the trainer did not manufacture an apparent track improvement by forcing every optimized cell into production. Atlantic short- and medium-range intensity and eastern Pacific medium- and long-range intensity cleared the gate; every track cell and the remaining intensity cells retained the baseline.

4.3 Bias governance

Bias coefficients use a weighted trimmed mean, minimum sample requirements, and N/(N+K) shrinkage. The source artifact records sign convention, sample count, raw and shrunk values, retained recency weight, and application strength. Track and intensity corrections are therefore inspectable and can be disabled or attenuated by horizon without retraining the entire system.

> **No silent promotion**
>
> Both the current weight candidate and bias candidate are marked *candidate_not_auto_promoted*. Production changes require an explicit review and release decision; generating a better retrospective artifact is not sufficient.

5. Verification Framework and Evidence Archive

NHC-style homogeneous rankings, exact pairwise cases, and issuance reliability are separate evidence layers.

5.1 Source archive

Table 5. Evidence archive used in this report.

Attribute	Value
Finalized seasons	2019–2025
Basins	Atlantic (AL) and eastern/central North Pacific profile (EP)
Completed storm pairs	275
Verification rows	13,365,524
Historical archive status	275 reused; 0 rebuilt in the supplied run
Current season exclusion	2026 excluded from historical training
Trainer run ID	20260804T071828Z
Historical signature	4fa288017a79c71db6470e984cdf24f2efdf275be7d9a5286f324eb78b0a18b8

The large verification-row count is not the number of independent forecasts. It is the row-level archive from which storm–cycle–tau cases, candidate members, and error components are constructed. Reported forecast comparisons use explicit case matching and must not be interpreted as millions of independent forecast events.

5.2 NHC-style homogeneous ranking

The annual ranking layer follows the supplied NHC scoring contract: standard forecast hours 12, 24, 36, 48, 60, 72, 96, and 120 h; tropical or subtropical verifying status; an availability screen; and a common forecast roster within each season, basin, metric, and forecast hour. OFCL is shown as the official benchmark. PeerRank excludes OFCL and ranks the automated aids.

The availability screen follows the rule encoded for each verification year: through 2024, at least two-thirds of OFCL opportunities at both 48 and 120 h; from 2025 forward, at least two-thirds at every standard forecast hour. This prevents sparse aids from entering only on favorable cases.

5.3 Exact pairwise comparison

The pairwise layer compares STRK with one comparator at a time on identical storm–cycle–tau cases. It answers a different question than the annual roster: when both aids existed and truth was valid, which had lower mean error? Percent improvement is defined as:

$$SkillB = 100 \times (EB - ESTRK) / EB \tag{17}$$

Pairwise N therefore varies by comparator and horizon. The values should not be combined across comparators as though they came from one common universal sample. Positive skill favors STRK; a near-zero value means practical parity in the pooled historical sample and is not evidence of statistical equivalence.

5.4 Issuance reliability

The replay also evaluates whether the operational builder could issue a valid track at all. Cycle failures and tau failures are preserved. Across 2019–2025, STRK issued 84.9% of attempted cycles and produced 50,613 verified replay points. STRE issued slightly more cycles because its fixed equal-weight member set can remain viable in some availability configurations where differential-weight controls do not.

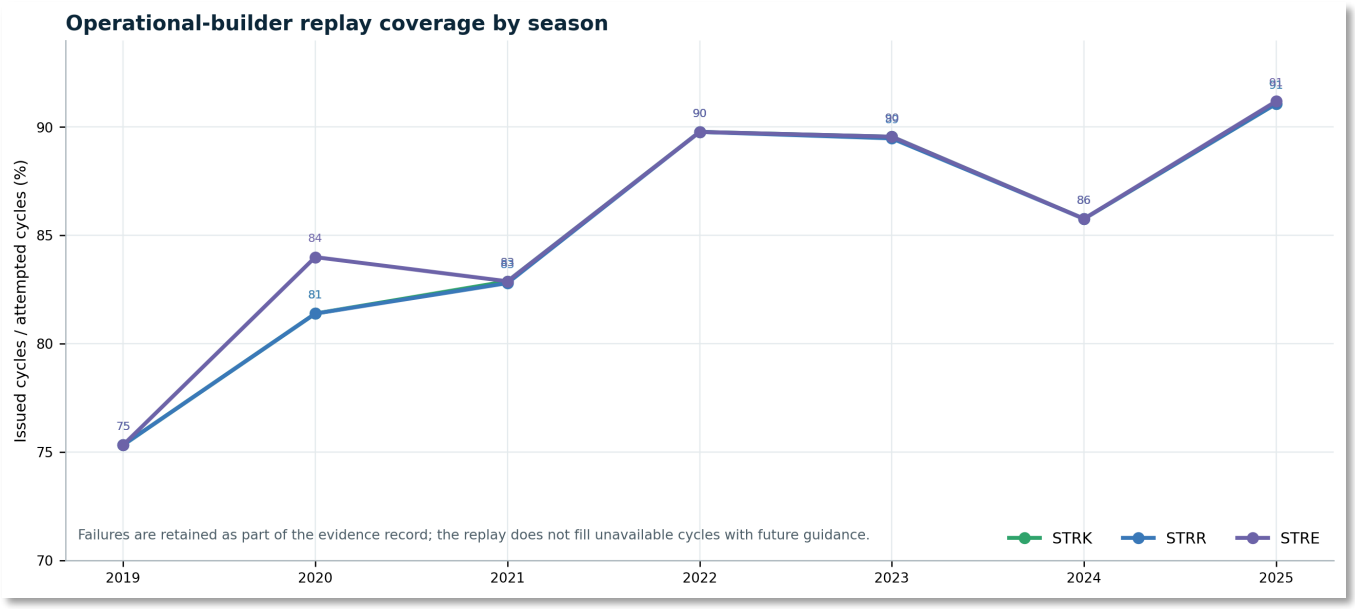


Figure 4. Historical operational-builder replay coverage. Issuance rate is an independent reliability metric and is not replaced by forecast-error skill.

6. Developmental Results, 2019–2025

A competitive consensus family with clear strengths, neutral comparisons, and control cases.

6.1 Automated-aid ranking cells

Table 6. Track ranking summary across 112 NHC-style homogeneous cells.

Technique	Rank 1	Top 3	Top 5	Median rank	Interpretation
STRK	6	52	99	4	Primary weighting is consistently competitive but not dominant.
STRR	17	52	93	4	Raw weighting produces more first-place cells but somewhat fewer top-five cells.
STRE	11	59	100	3	Equal weighting is a strong control and leads several seasons/cells.

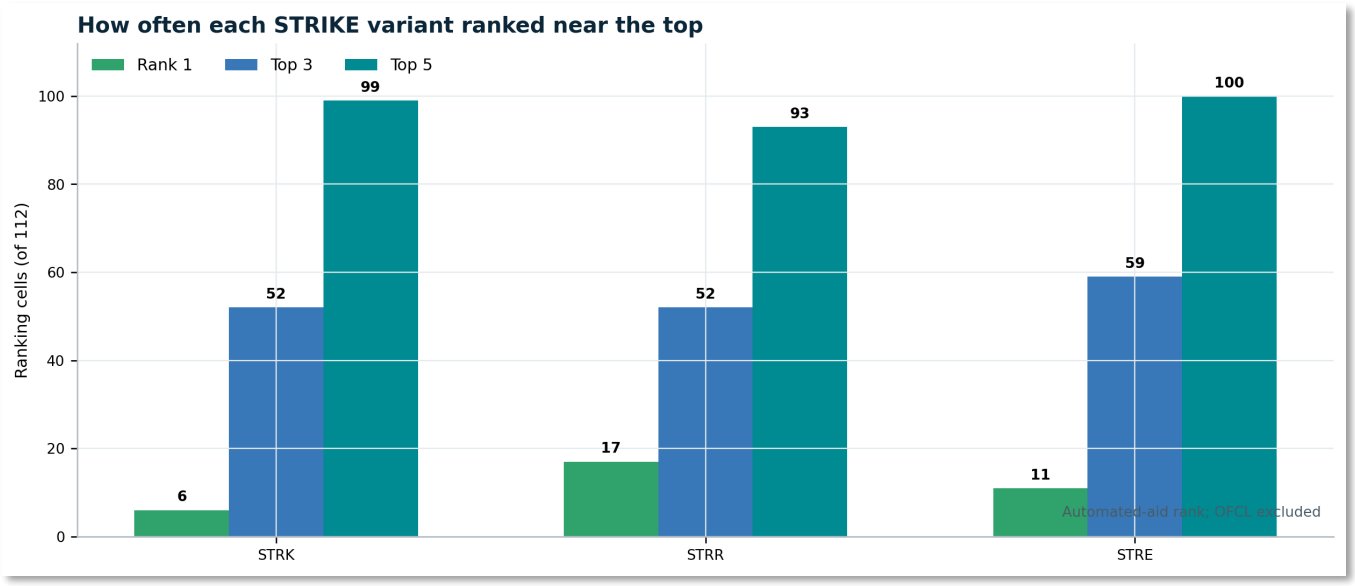


Figure 5. Frequency of first-place, top-three, and top-five automated-aid ranks. The control results are retained because they directly test whether added weighting complexity creates value.

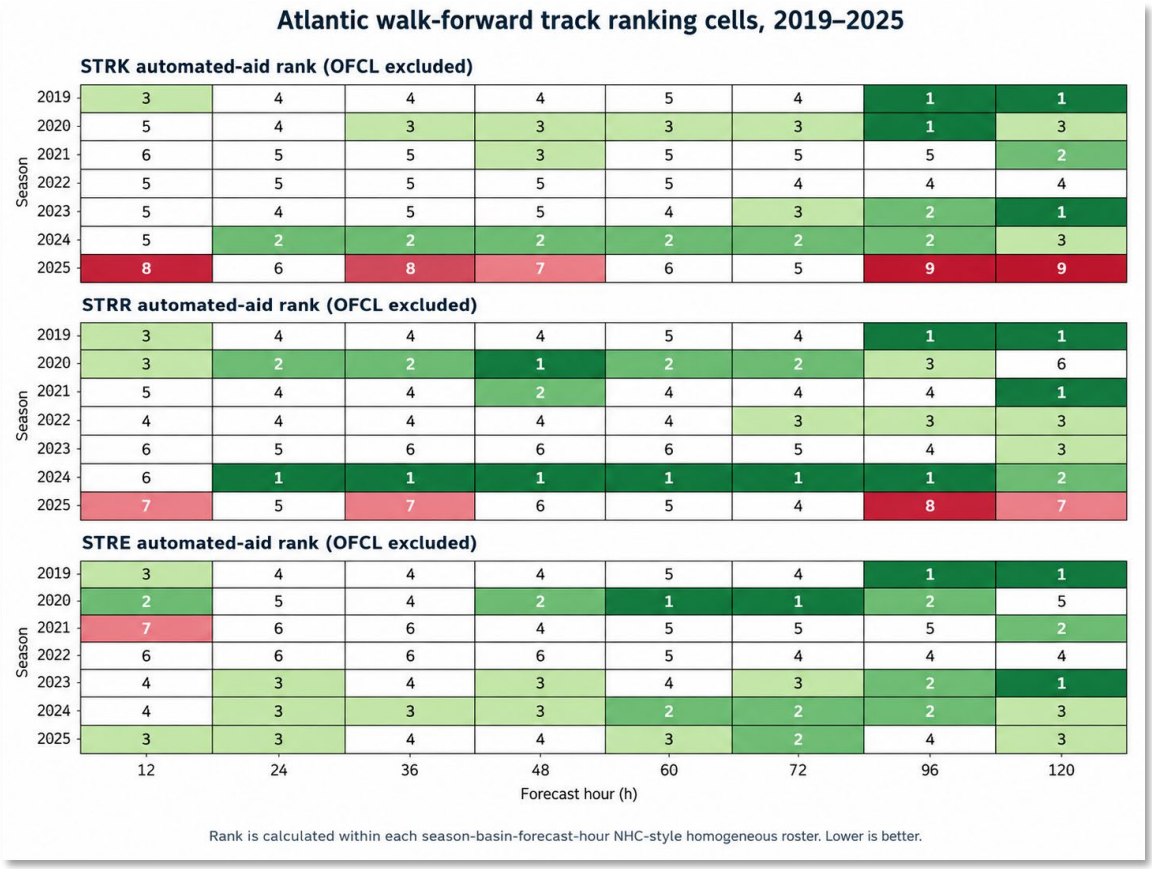


Figure 6. Atlantic track ranks by season and standard forecast hour. Ties can appear when variants are identical under the available-member configuration.

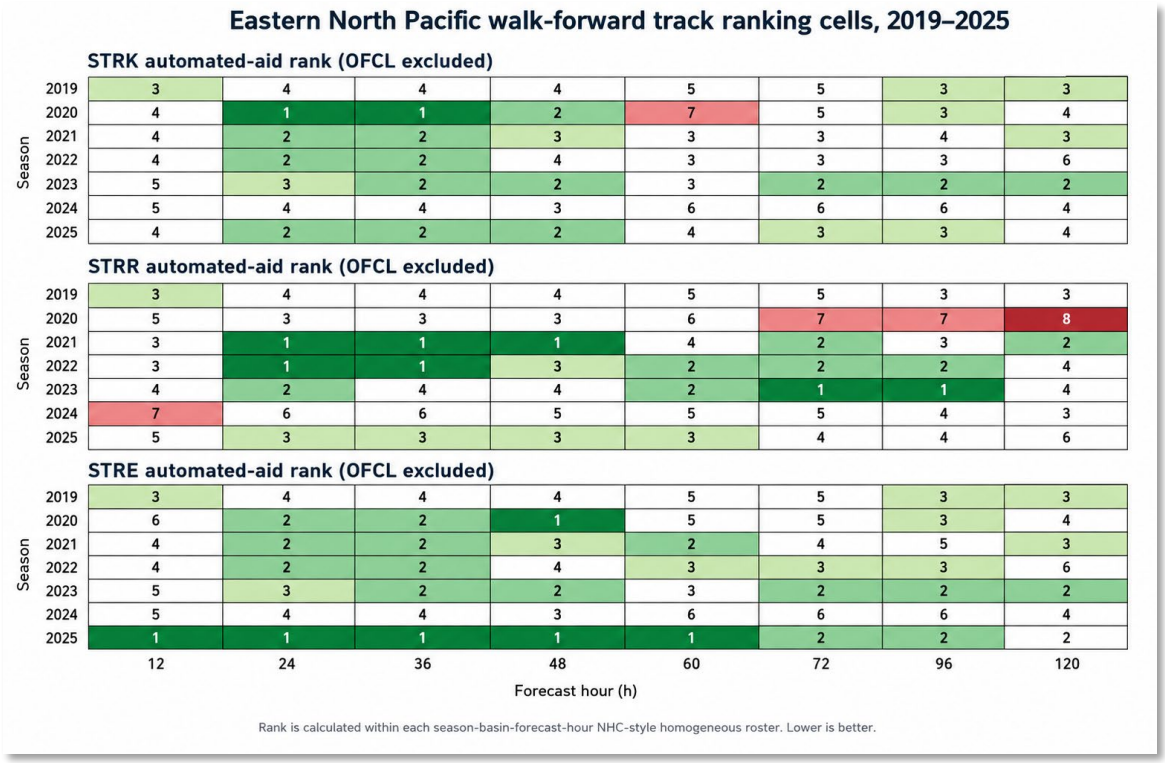


Figure 7. Eastern North Pacific track ranks by season and standard forecast hour. Performance changes by season and horizon, reinforcing the need for prospective evidence.

6.2 Pairwise track comparisons

Table 7. STRK track improvement on exact pairwise cases.

Comparator	0–48 h	60–96 h	108–120 h
OFCI	+12.9% (N 16,938)	+10.3% (N 11,031)	+7.6% (N 3,712)
OFCL	-0.4% (N 17,899)	-0.4% (N 8,976)	+0.2% (N 1,861)
HCCA	-2.8% (N 21,884)	-1.6% (N 16,292)	+2.0% (N 5,868)
TVCA	+1.3% (N 22,362)	+2.1% (N 16,673)	+1.7% (N 6,042)
FSSE	-0.5% (N 14,213)	+2.9% (N 9,556)	+5.9% (N 3,149)
EMNI	+13.3% (N 18,894)	+12.9% (N 14,124)	+11.9% (N 5,067)
EMXI	+7.9% (N 17,589)	+8.2% (N 13,063)	+9.0% (N 4,609)

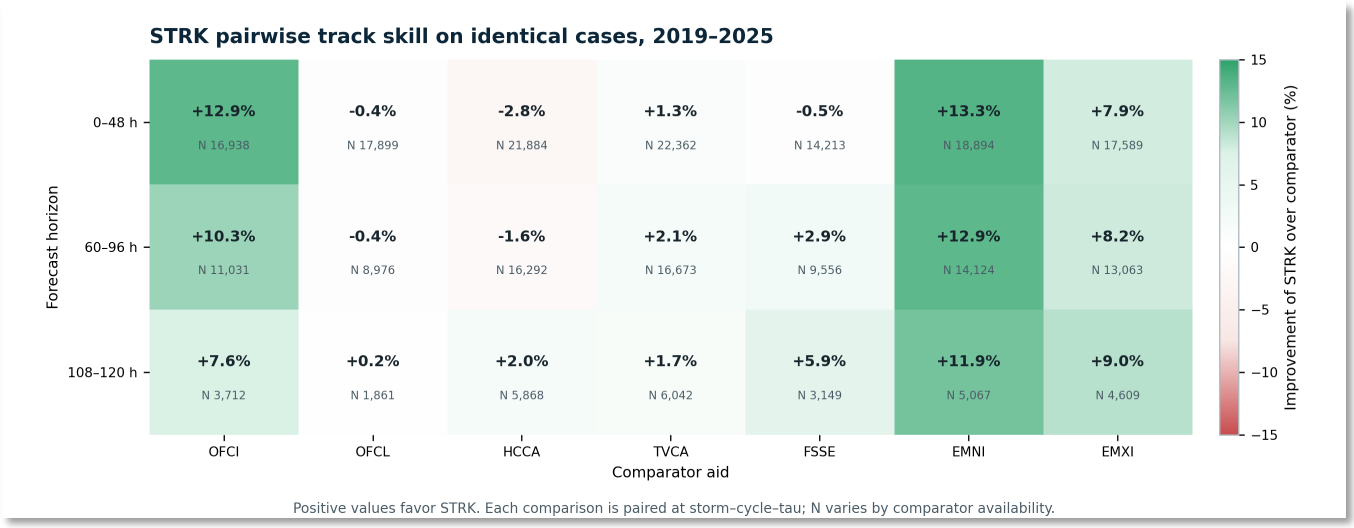


Figure 8. Exact pairwise STRK track skill. The near-zero OFCL comparisons are a key boundary on the interpretation of the stronger OFCI and single-model results.

The track evidence supports three conclusions. First, STRK provides a meaningful reduction relative to OFCI and several independent model/ensemble-mean aids. Second, it is close to TVCA/TVCN and HCCA, with no blanket dominance over corrected consensus guidance. Third, pooled errors are essentially neutral to OFCL. The report therefore does not use OFCI skill as a proxy for superiority to the contemporaneous official forecast.

6.3 Pairwise intensity comparisons

Table 8. STRK intensity improvement on exact pairwise cases.

Comparator	0–48 h	60–96 h	108–120 h
OFCI	+6.0% (N 16,876)	+6.3% (N 10,927)	+7.3% (N 3,609)
OFCL	-4.6% (N 17,837)	-2.2% (N 8,897)	-0.1% (N 1,806)
HCCA	+1.0% (N 21,728)	-1.6% (N 15,695)	-0.3% (N 5,483)
FSSE	-2.5% (N 14,206)	+6.7% (N 9,479)	+17.7% (N 3,084)
EMNI	+39.1% (N 18,840)	+45.0% (N 13,934)	+41.2% (N 4,833)
EMXI	+31.3% (N 17,542)	+29.5% (N 12,897)	+18.2% (N 4,426)

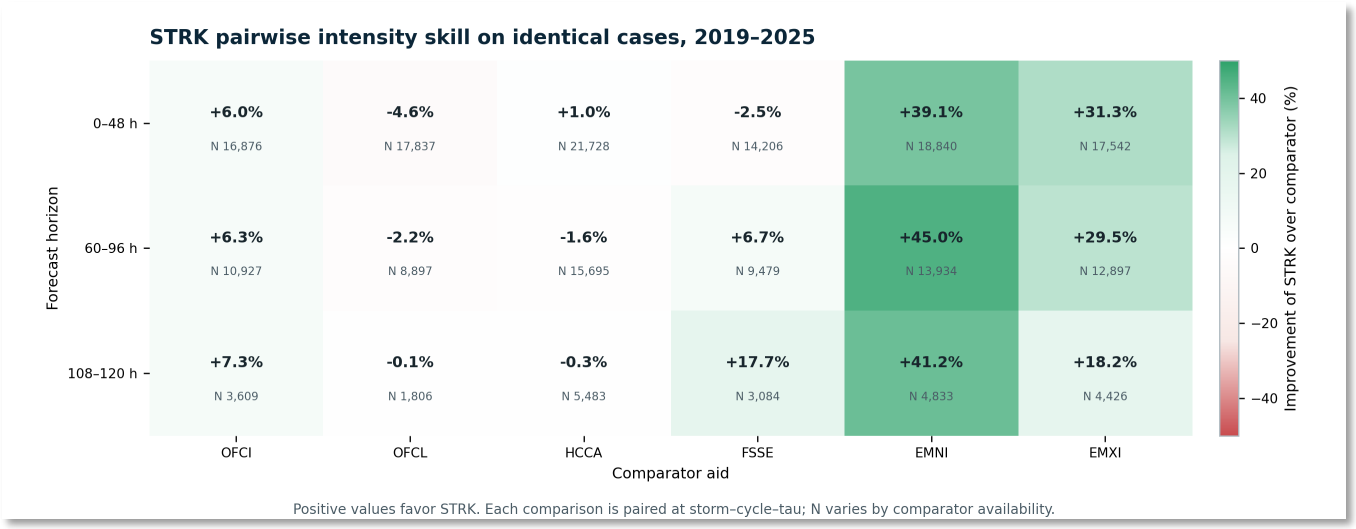


Figure 9. Exact pairwise STRK intensity skill. Large improvements over individual ECMWF products do not imply an advantage over OFCL or HCCA.

Intensity results are strongest against individual model aids and OFCI, but are mixed against HCCA, FSSE, and OFCL. At short range, STRK is 4.6% worse than OFCL in the pooled exact pairwise sample; at medium range it is 2.2% worse; at long range it is effectively neutral. These findings support use as experimental supplemental guidance, not a claim that the system has solved tropical cyclone intensity forecasting.

6.4 What the controls say

STRR and STRE are not cosmetic variants. Their strong performance exposes the central scientific question: how much of the result comes from ensemble error cancellation, and how much from differential weighting? STRE's 2025 eastern Pacific performance and STRR's first-place frequency show that weighting complexity must continually justify itself. The current adoption gate, which retained every track baseline cell, is consistent with that evidence.

Interpretation boundary

The historical replay demonstrates constructible, leakage-resistant competitiveness. It is not independent prospective validation because development decisions were informed by the same broad archive and because upstream real-time availability cannot be perfectly reconstructed from finalized records.

7. 2026 Prospective Forecast and Verification Program

Frozen methods, public ATCF records, failed-cycle accounting, and a deliberately limited early sample.

7.1 Prospective objective

The 2026 season is the first season intended to create an externally inspectable real-time performance record for Method 2.1.0. The scientific objective is not to accumulate only successful examples. It is to determine whether the system remains stable when guidance is late, missing, truncated, or internally inconsistent—and whether any historical skill survives under the actual decision-time information set.

Frozen-season rules

- Freeze the published method version, production coefficient identity, bias identity, member catalog, and scoring contract before evaluating performance.
- Issue only from guidance available to the forecast cycle and record the source cycle and package SHA-256.
- Do not retroactively fill a non-issued cycle or replace an issued point with late guidance for verification.
- Preserve all public ATCF decks and manifests by storm, including unfavorable forecasts.
- Separate changes required for software integrity from changes that alter meteorological behavior; document either type before it affects a forecast.

7.2 2026 data collection period

The 2026 season serves as the first live data collection period for the STRIKE SuperEnsemble. Real-time storm forecasts will begin in August 2026 and will be published at strikesuperensemble.org. Forecast records collected during the season will be preserved for later independent verification and used to evaluate availability, operational behavior, and performance across a broader set of storms, basins, and forecast conditions.

7.3 Public access

The public release will expose cumulative storm A-decks in standard ATCF-style text and gzip formats, plus a manifest. The public-facing cyclone tracker will not require an access request. It will provide cycle identification, technique selection, track and intensity points through 120 h, and explicit experimental labeling. Public accessibility is part of the verification design: a forecast that can be downloaded before the outcome is inherently easier to audit than a retrospective plot.

The project website will serve as the stable entry point for the method description, release notes, public deck discovery, verification summaries, and tracker access. Archived deck files should remain addressable after storms become post-tropical so independent reviewers can reproduce annual results.

7.4 End-of-season verification package

- Storm-level and annual cumulative ATCF decks, including STRK, STRR, and STRE.
- Issuance availability table listing every attempted and issued cycle and explicit failure reason.
- NHC-style homogeneous track and intensity tables at standard forecast hours.
- Exact pairwise tables against OFCL/OFCI, recognized consensus aids, and independent member aids.
- Storm-block uncertainty analysis and sensitivity to roster/availability rules.
- A frozen reproducibility manifest containing source hashes, method version, coefficients, and code identities.

8. STRIKE Impact Corridor

An experimental, proprietary location-query technique for interrogating convergent guidance well before landfall.

8.1 Product concept

The public cyclone tracker will include the STRIKE Impact Corridor, available for testing by any viewer without an access request. The viewer can pinpoint a location on the map. The system evaluates that point against synchronized NHC guidance, the STRK primary consensus, and the Grand Ensemble, then publishes a 0–10 score, classification, source breakdown, confidence estimate, closest approach, and relevant geometry/evidence fields.

The technique is designed to answer a narrow operational question: how strongly do three complementary guidance systems interact with this location at the current forecast cycle? It is intended to surface emerging location-specific concern earlier than a landfall-only narrative, particularly when track proximity, forecast intensity, official probability/radii information, and ensemble clustering reinforce one another.

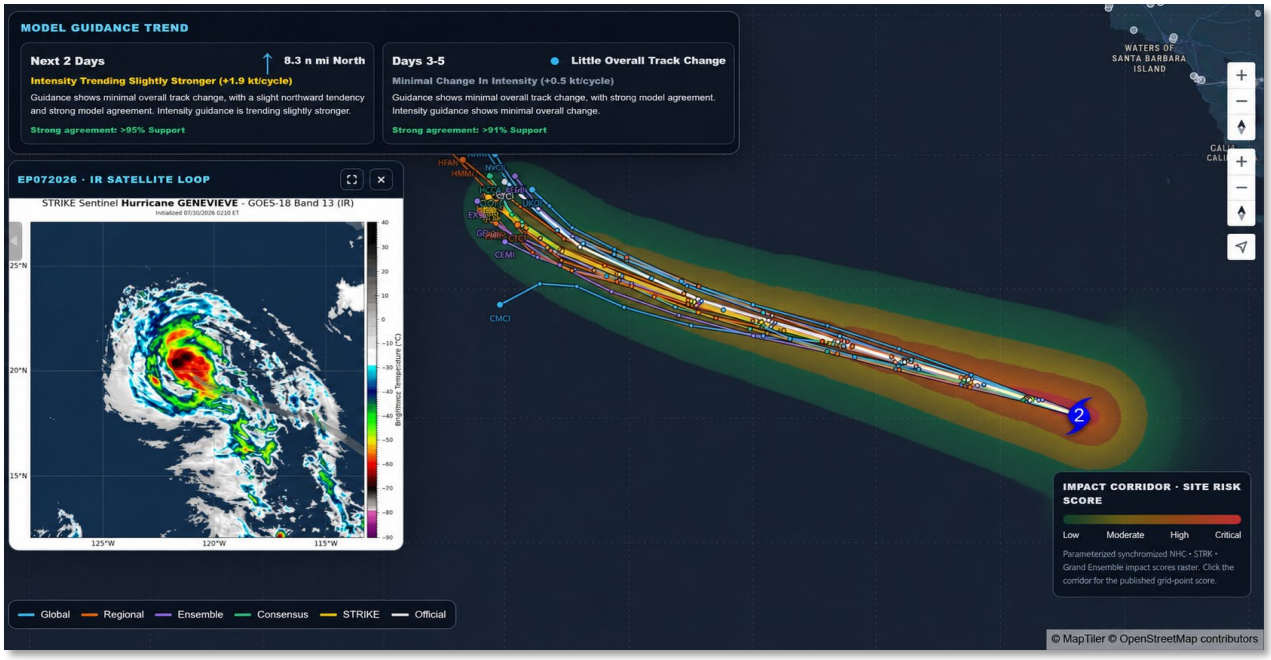


Figure 10. Public tracker implementation of the STRIKE Impact Corridor for Hurricane Genevieve (EP072026), showing the model-guidance field, location-specific corridor shading, guidance-trend panel, site-risk score, and infrared satellite context.

Product note

- The STRIKE Impact Corridor is a deterministic guidance-interaction index generated from current forecast geometry, timing, intensity, probabilities, and member clustering.
- It is not a probability that a location will be damaged, flooded, lose power, or experience a specific wind speed.
- It is not calibrated to economic loss, casualties, storm-surge depth, rainfall accumulation, wave height, tornado occurrence, or outage duration.
- The risk-language catalog is an experimental interpretive layer and must remain subordinate to official products and hazard-specific forecasts.
- The formula, bands, explanatory language, and visual design may be updated as 2026 evidence accumulates. Every behavior-changing update should receive a new formulaVersion, effective timestamp, release note, and a preserved record of the prior formula so retrospective validation remains possible.

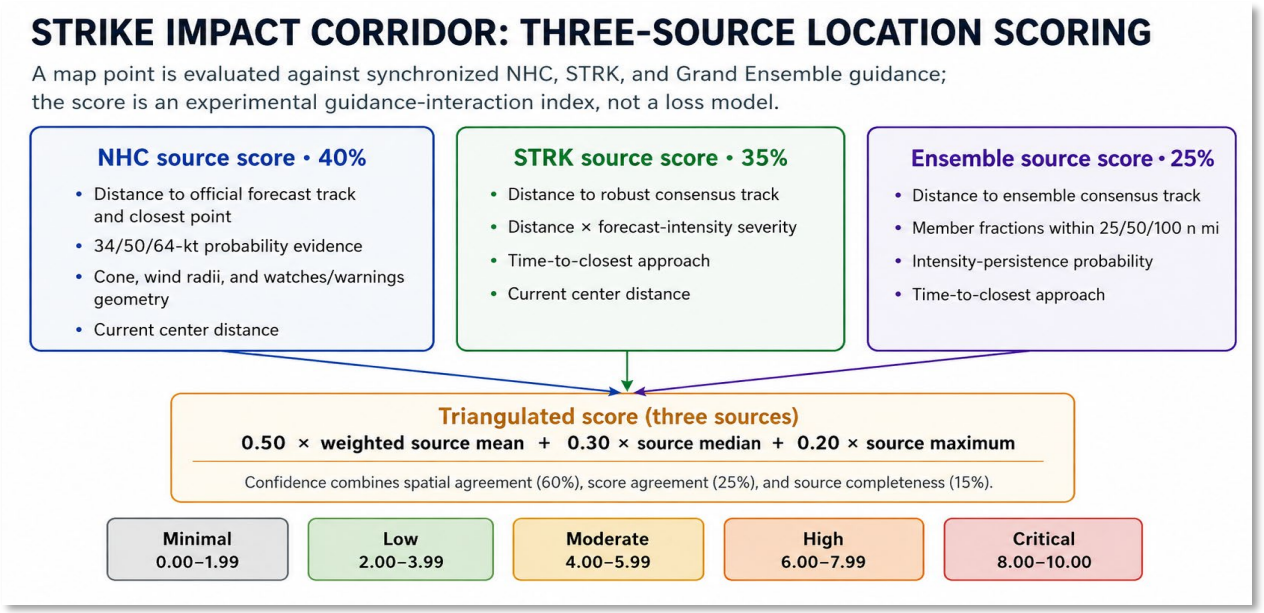


Figure 11. Implemented Impact Corridor architecture. Source weights affect the weighted mean; the median and maximum terms preserve agreement and high-end evidence without allowing a single source to define the entire score.

8.2 Source-score equations

All source scores are clipped to 0–10. Distances are measured to the closest point on a synchronized forecast track; τ is the closest forecast hour. Center-distance terms are deliberately weak and decay over 750 n mi. Eligibility can extend beyond the 2.00 display threshold when a location intersects official geometry or has nonzero probability evidence.

$$NHC = 3P200 + U120 + 3.5W + R + 0.7C + 0.3Q + 0.5D750 \tag{18}$$

P200 is normalized proximity within 200 n mi; U120 is urgency through 120 h; W is the maximum normalized 34/50/64-kt probability evidence; R, C, and Q indicate wind-radii, cone, and watch/warning geometry; D750 is current-center proximity.

$$STRK = 4P200 + 4P200S(V) + 1.5U120 + 0.5D750 \tag{19}$$

S(V) is the implemented forecast-intensity severity transform. The interaction P200S(V) prevents a distant intense forecast from receiving the same score as a similarly intense track near the queried location.

$$ENS = 3P200 + p25 + 1.8p50 + 1.2p100 + U120 + 1.5I + 0.5D750 \tag{20}$$

p25, p50, and p100 are the fractions of eligible ensemble members passing within 25, 50, and 100 n mi; I is the maximum ensemble intensity-persistence probability near the closest tau.

8.3 Triangulation and confidence

$$Weighted = 0.40\,NHC + 0.35\,STRK + 0.25\,ENS \tag{21}$$

$$Impact = 0.50\,Weighted + 0.30\,Median(NHC, STRK, ENS) + 0.20\,Maximum(NHC, STRK, ENS) \tag{22}$$

When only two sources are available, the formula changes to 0.65 weighted mean + 0.20 median + 0.15 maximum. With one source, the source score is returned directly. Confidence is based on 60% spatial agreement, 25% score agreement, and 15% source completeness; confidence is capped at 75 when fewer than three sources are available.

$$Confidence = 100[0.60Aspatial + 0.25Ascore + 0.15Ccomplete] \tag{23}$$

Table 9. Current Impact Corridor score bands.

Band	Score	Operational display meaning
Minimal	0.00–1.99	Below corridor display threshold unless other eligibility evidence is present.
Low	2.00–3.99	Weak but nontrivial interaction with current guidance.
Moderate	4.00–5.99	Multiple terms indicate a meaningful location interaction.
High	6.00–7.99	Strong proximity/urgency/intensity or multi-source clustering.
Critical	8.00–10.00	Highest modeled guidance interaction; requires careful official-context review.

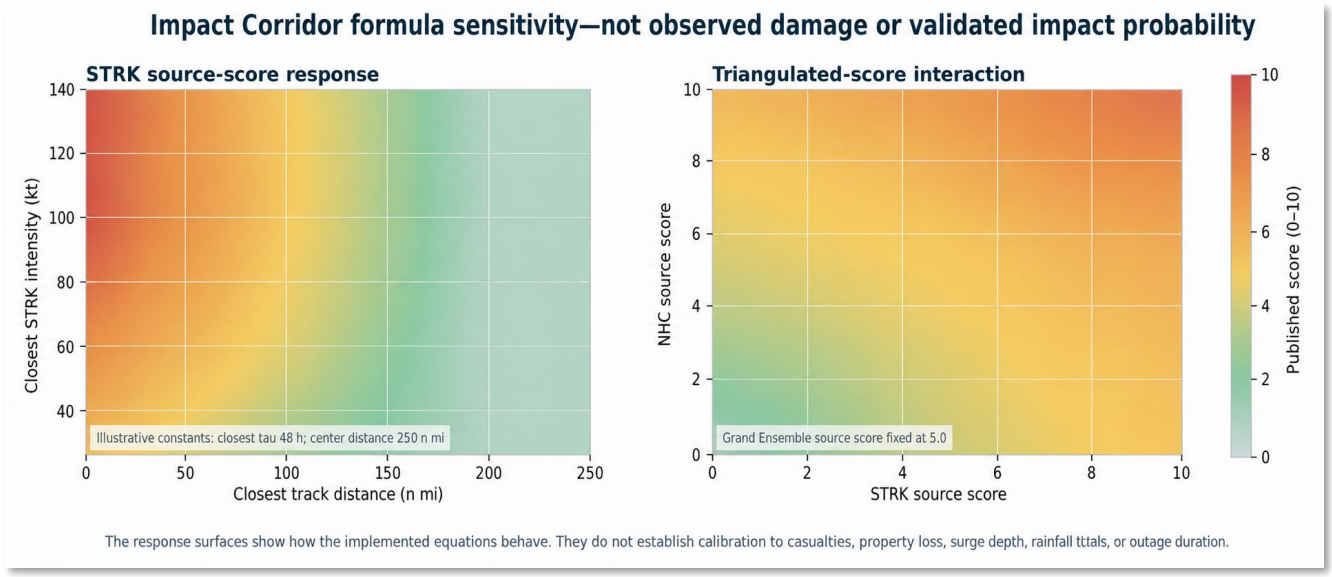


Figure 12. Formula-response surfaces. These visuals are generated from the implemented equations to make model behavior inspectable; they are not observed impact calibration.

8.4 Surface-aware interpretation

The backend classifies the clicked point as over water, coastal, or inland using a geospatial surface mask and a configurable coastal band. The interpretive catalog then emphasizes different plausible hazard families: marine wind/wave conditions over water, surf/erosion/coastal inundation near the coast, and wind/rain/flooding/transportation disruption inland. Inland locations are not assigned storm-surge language. This taxonomy is more physically defensible than applying one generic impact paragraph everywhere, but it still requires prospective validation against hazard observations and user interpretation.

8.5 2026 research questions

1. Calibration: Does each score band correspond to increasing observed hazard frequency or severity when evaluated by surface class and forecast lead time?
2. Discrimination: Does the score distinguish affected from unaffected locations better than track distance alone or a simple maximum-source score?
3. Reliability: Are high scores systematically over- or under-issued, and how does that behavior differ between land, coast, and water?
4. Lead time: Does triangulation provide useful advance signal before conventional location-specific thresholds are reached?
5. Incremental value: Does the STRK source add measurable information beyond official and ensemble terms?
6. Communication: Do users interpret a 0–10 score and confidence correctly without confusing it with an official warning or an impact probability?
7. Stability: How sensitive are scores to cycle-to-cycle track shifts, ensemble membership changes, and partial source availability?

9. Limitations and Responsible Interpretation

Known boundaries that must remain visible in research, publication, and public use.

9.1 Historical-development limitations

- The broad 2019–2025 archive informed model development. Season-blocked replay reduces direct leakage but does not convert the entire archive into a fully independent validation set.
- Finalized A-decks do not reproduce every real-time latency, outage, data correction, and polling-state condition. The production builder is replayed, but upstream availability is still reconstructed.
- Member systems changed during 2019–2025. A technique identifier can span upgrades in model physics, resolution, initialization, or post-processing.
- Annual/basin samples are uneven, especially at 108–120 h. Long-range first-place ranks can arise from small common rosters and should be read with N.
- Many forecast cases from the same storm are correlated. Row counts are not independent-sample counts.
- The exact production coefficient JSON was not included in the supplied evidence bundle. This report therefore documents the operational contract, current source path, and candidate diagnostics rather than reproducing unverified “exact current weights.” Issued audit manifests remain the coefficient source of record.

9.2 Forecast interpretation limitations

- A consensus can fail when several members share the same error or when the true solution lies outside the main guidance cluster.
- Robust outlier screening can reject a physically correct minority member when the majority is wrong.
- A dominance cap protects diversity but can reduce the benefit of a genuinely superior member in a particular regime.
- Bias estimates describe historical average tendencies, not storm-specific dynamical causes.
- Track and intensity skill do not directly establish skill for rainfall, storm surge, wave, tornado, or infrastructure effects.

9.3 Public communication limitations

The public tracker can make complex guidance accessible, but accessibility can create false precision. A plotted line is not a deterministic path of the center, and an Impact Corridor score is not an official local forecast. Public displays should retain source dates, cycles, tau, units, uncertainty context, and a persistent link to NHC products. The interface should avoid language that implies evacuation or protective-action authority.

> **Required standing disclaimer**
>
> STRIKE is independent experimental guidance. **National Hurricane Center official forecasts, watches, warnings, and local emergency-management instructions remain authoritative**.

10. Path to Independent Verification and Operational Consideration

A staged proposal aligned with established tropical cyclone forecast-development practice.

10.1 Evidence sequence

A new guidance technique should not seek operational visibility from a favorable retrospective table alone. The appropriate sequence is a frozen real-time record, independent verification, technical review, forecaster evaluation, and only then consideration for a formal experimental pathway. NHC verification practice emphasizes consistent case definitions, homogeneous comparisons, and post-analyzed best track; HFIP research-to-operations activities emphasize real-time experiments, post-season evaluation, and transition readiness.

Table 10. Proposed research-to-operations evidence ladder.

Stage	Required evidence or action	Decision gate
1. Freeze	Method, member pool, production weights, bias field, code identity, ATCF format, and verification contract fixed for a defined period.	Can another party identify exactly what was forecast?
2. Prospective issuance	Public, timestamped ATCF records and failed-cycle accounting generated before outcomes are known.	Is the record genuinely real-time and complete?
3. Independent verification	Annual NHC-style homogeneous and pairwise evaluation by an external meteorologist or institution.	Does skill persist under external case construction?
4. Scientific review	Methods manuscript, source-level audit, leakage review, sensitivity tests, and reproducibility package.	Are the assumptions and safeguards technically defensible?
5. Forecast experiment	Structured real-time display to forecasters or researchers, ideally through an HFIP/HREx-style experiment or partner testbed.	Does it add usable information in the forecast cycle?
6. Transition assessment	Compatibility, latency, reliability, maintainability, support ownership, and forecaster training assessed.	Can it be sustained without disrupting operations?
7. Operational consideration	Formal agency decision only after stable value is demonstrated.	Is operational benefit greater than implementation and maintenance cost?

10.2 Concrete external requests

8. Data access: research access to major aids at or near native availability, with highest priority on deterministic ECMWF tropical cyclone guidance (EMXI) and other AI-derived aids that are not consistently available in public real-time A-decks.
9. Annual external verification: independent scoring of public STRIKE ATCF records using NHC procedures and case definitions, explicitly separated from internal hindcast products.
10. Methodological review: scrutiny of member independence, training boundaries, weight adoption, bias correction, roster construction, outlier handling, and information-leakage safeguards.
11. Real-time research partnership: a university, NOAA laboratory, HFIP participant, or other tropical cyclone research group willing to observe the 2026 forecast stream and conduct post-season analysis.
12. Forecaster usability assessment: evaluation of whether STRK, STRR, and source/audit information can be interpreted within the forecast cycle without adding confusion or excessive workload.
13. Impact Corridor validation partnership: access to hazard observations and expertise needed to test calibration by surface class, lead time, and hazard family.

10.3 Candidate submission package

- This technical report and a concise method note suitable for journal or conference review.
- Current and archived source code with tagged release and environment specification.
- Production coefficients and bias field with machine-readable schema and hashes.
- Public 2026 ATCF storm decks, manifests, and issuance availability records.
- Historical and prospective verification tables with exact case keys or a privacy-neutral reproducible derivative.

- A one-page forecaster quick reference explaining member composition, failure modes, and interpretation.
- Impact Corridor algorithm specification, formula-version history, sensitivity tests, and observed-impact validation plan.

10.4 Operational criteria

Established testbed transition criteria provide a useful final screen even when the exact administrative pathway changes: forecast benefit, computational and human efficiency, ease of use, compatibility with operational systems, and long-term sustainability. A statistically positive technique can still be unsuitable if it arrives too late, fails silently, requires excessive support, or creates a confusing signal. Conversely, a modest-error improvement may be valuable if it is reliable, transparent, and exposes complementary information at a difficult forecast horizon.

> **Standard for consideration**
>
> External access, review, or operational consideration should be based on reproducible methodology, pre-issued ATCF records, homogeneous case matching, issuance reliability, and transparent reporting of both successful and unsuccessful forecasts.

11. Conclusions

What is established, what remains experimental, and what 2026 must decide.

Method 2.1.0 is a complete, auditable tropical cyclone consensus architecture rather than a retrospective averaging exercise. It separates track and intensity, constrains the member pool to independent aids, organizes coefficients by basin and horizon, applies historically supported motion-relative correction, screens extreme departures robustly, caps member dominance, computes track positions on the sphere, and preserves issued packages and cumulative ATCF records.

The 2019–2025 cumulative replay shows that the STRIKE SuperEnsemble family is consistently competitive among automated aids. STRK appears in the top five in 99 of 112 track ranking cells and shows useful exact-pairwise gains over OFCI and several individual aids. At the same time, differences from OFCL are approximately neutral, HCCA comparisons are mixed, and the raw/equal-weight controls frequently match or beat the primary variant. These are scientifically valuable constraints, not weaknesses to hide: they identify the specific value proposition that must be tested prospectively.

The 2026 season provides the correct next experiment. Public, pre-issued ATCF decks and a no-request public tracker will make the record independently inspectable. Only two storms in the supplied evidence currently contain live STRK issuance, so no early performance conclusion is justified. The project should preserve a frozen method, failed cycles, source manifests, and all unfavorable cases through the season, then invite external NHC-style verification.

The Impact Corridor extends the project from forecast consensus to location interrogation. Its three-source triangulation is technically transparent and potentially useful, but it remains an experimental guidance-interaction score. Its labels and effects language must be calibrated against observed hazards and tested for communication risk before any operational implication is considered. Formula versions should be frozen and archived whenever behavior changes.

Bottom line

STRIKE is ready for prospective scientific evaluation as an independent experimental technique. It is not yet validated for operational adoption. The pathway forward is public real-time issuance, independent verification, methodological review, and structured forecast-user testing.

Appendix A. Operational Constants and Member Catalog

A1. Method and module identities

Table A1. Current software and artifact identities.

Artifact	Version / status
STRIKE method	2.1.0
tc_models module	1.6.4
STRK trainer	1.11.0
NHC scoring contract	1.0.0
EMNI method	1.0.0
Impact Corridor module	1.7.1
Impact Corridor formula	triangulated_v1
Impact grid	256 × 256
Candidate weight status	candidate_not_auto_promoted
Bias status	candidate_not_auto_promoted

A2. ATCF technique and family identifiers

Table A2. Principal four-character aid identifiers used in the report.

Identifier	Description / current role
STRK	STRIKE Primary weighted consensus
STRR	STRIKE Raw performance-weighted control
STRE	STRIKE Equal-weight control
EMNI	STRIKE-derived ECMWF ensemble-mean product of record
EMXI	Historical deterministic ECMWF comparator; not synonymous with EMNI
AVNI	Interpolated GFS deterministic
AEMI	Interpolated GEFS mean
GDMI	Interpolated Google DeepMind ensemble mean
HFAI / HFBI	Interpolated HAFS-A / HAFS-B
HWFI / HMNI	Interpolated HWRF / HMON
UKXI	Interpolated UKMET
CMCI	Interpolated Canadian deterministic
CTCI	Interpolated COAMPS-TC
NVGI	Interpolated NAVGEM
DSHP / LGEM / SHIP / ICON	Statistical or dynamical intensity members where configured
OFCL / OFCI	Official forecast and interpolated official comparator; excluded from member blend
HCCA / TVCN / IVCN / FSSE	Consensus comparators; excluded from member blend

A3. Impact Corridor output fields

Table A3. Principal Impact Corridor point-query fields.

Field	Meaning
score	Published 0–10 triangulated guidance-interaction index
classification	Minimal / Low / Moderate / High / Critical
confidence	0–100 agreement/completeness indicator; capped at 75 with fewer than three sources
nhcScore / strkScore / ensembleScore	Source-level contributions before triangulation
closestUtc / closestTauH	Median source closest-approach time and forecast hour
trackDistanceMi	Source-specific distance from point to forecast track
prob34/50/64Pct	NHC probability evidence at the point
insideCone / insideRadii / insideWarning	Official geometry intersection flags
membersWithin25/50/100Pct	Ensemble member clustering around the point
triangulationStatus	synchronized / partial / unsynchronized / unavailable
surface	Over Water / Coastal / Inland classification

Appendix B.

Selected Benchmark Values and Seasonal Context

B1. Exact pairwise track errors

Table B1. Pooled exact-pairwise track comparison values.

Comparator	Horizon	STRK mean n mi	Comparator mean n mi	Skill	N
OFCI	SHORT	40.6	46.6	+12.9%	16,938
OFCI	MEDIUM	98.6	109.9	+10.3%	11,031
OFCI	LONG	170.1	184.1	+7.6%	3,712
OFCL	SHORT	41.1	40.9	-0.4%	17,899
OFCL	MEDIUM	95.6	95.2	-0.4%	8,976
OFCL	LONG	183.3	183.7	+0.2%	1,861
HCCA	SHORT	45.2	44.0	-2.8%	21,884
HCCA	MEDIUM	104.9	103.3	-1.6%	16,292
HCCA	LONG	169.5	173.0	+2.0%	5,868
TVCA	SHORT	45.3	45.8	+1.3%	22,362
TVCA	MEDIUM	104.6	106.8	+2.1%	16,673
TVCA	LONG	168.5	171.4	+1.7%	6,042
TVCN	SHORT	45.2	45.9	+1.6%	22,344
TVCN	MEDIUM	104.4	107.5	+2.9%	16,629
TVCN	LONG	168.5	172.3	+2.2%	6,016
FSSE	SHORT	39.5	39.3	-0.5%	14,213
FSSE	MEDIUM	101.2	104.2	+2.9%	9,556
FSSE	LONG	178.7	189.8	+5.9%	3,149
EMNI	SHORT	43.9	50.7	+13.3%	18,894
EMNI	MEDIUM	102.3	117.4	+12.9%	14,124
EMNI	LONG	168.6	191.4	+11.9%	5,067
EMXI	SHORT	43.6	47.3	+7.9%	17,589
EMXI	MEDIUM	101.0	110.0	+8.2%	13,063
EMXI	LONG	166.4	182.9	+9.0%	4,609

Note: Each basin-season-horizon row is paired on exact storm–cycle–tau cases before pooling. N is not constant across comparators.

B2. Exact pairwise intensity errors

Table B2. Pooled exact-pairwise intensity comparison values.

Comparator	Horizon	STRK mean kt	Comparator mean kt	Skill	N
OFCI	SHORT	7.72	8.22	+6.0%	16,876
OFCI	MEDIUM	10.97	11.71	+6.3%	10,927
OFCI	LONG	12.74	13.74	+7.3%	3,609
OFCL	SHORT	7.70	7.36	-4.6%	17,837
OFCL	MEDIUM	11.11	10.86	-2.2%	8,897
OFCL	LONG	13.19	13.18	-0.1%	1,806
HCCA	SHORT	7.36	7.44	+1.0%	21,728
HCCA	MEDIUM	11.65	11.47	-1.6%	15,695
HCCA	LONG	13.96	13.93	-0.3%	5,483
FSSE	SHORT	8.29	8.09	-2.5%	14,206
FSSE	MEDIUM	11.39	12.20	+6.7%	9,479
FSSE	LONG	13.07	15.89	+17.7%	3,084
EMNI	SHORT	7.30	12.00	+39.1%	18,840
EMNI	MEDIUM	11.23	20.43	+45.0%	13,934
EMNI	LONG	13.32	22.65	+41.2%	4,833
EMXI	SHORT	7.36	10.71	+31.3%	17,542
EMXI	MEDIUM	11.24	15.94	+29.5%	12,897
EMXI	LONG	13.32	16.29	+18.2%	4,426

B3. Replay issuance totals

Table B3. Historical operational-builder replay coverage, 2019–2025.

Technique	Cycles attempted	Cycles issued	Issuance rate	Verified replay points
STRK	7,548	6,407	84.9%	50,613
STRR	7,548	6,405	84.9%	50,586
STRE	7,548	6,444	85.4%	50,964

Appendix C.

Reproducibility Record and Evidence Provenance

C1. Source identities

Table C1. Primary source hashes used to build this report.

Source artifact	SHA-256
strk_trainer 1.11.0	a94e2ab16bfc95949620527bde06ba0a38a94007909ae05e4e57c329cf2bedcc
tc_models 1.6.4	03d93e37c7b3152420f4aea3d5298ca23abbd4f943bb5f037937c4b567db3579
impact_corridor 1.7.1	df482d99a090178a962546116c6a467fb86484f7fb83a2b145d68dbfcc111b90
public website source	8e92b0e7c1aecc6da9dcf40a5c6cbdecca8f08c2866a65d362bae321a9feb034
analysis archive	3152622d49b12d41059a8852579d30af15cbef67661b5aa4365271c4754e495d
superseded report source	45eb6aa705795e4068d4d0d8c586887938a1e58e7fdd930037b4fb7407381ef9

C2. Historical archive manifest

Table C2. Historical source manifest.

Field	Value
schemaVersion	1
trainerVersion	1.11.0
generatedUtc	2026-08-04T07:18:38Z
pairs	275
built	0
reused	275
verificationRows	13365524
years	2019|2020|2021|2022|2023|2024|2025
archiveRoot	C:\zzEnvironment\TC\STRK_trainer\archive
currentSeasonExcluded	2026

C3. Reconstruction and scoring sequence

14. Read finalized historical ATCF verification caches and normalize legacy identifiers without collapsing EMNI and EMXI.
15. Build independent storm–cycle–tau member cases for track and intensity; retain current-season data outside historical training.
16. Advance the cumulative training boundary one completed season at a time and replay STRK, STRR, and STRE with the production builder.
17. Record attempted cycles, failed cycles, issued cycles, tau-level contributor failures, and published points.
18. Construct NHC-style homogeneous annual rosters and exact pairwise case sets at standard forecast hours.
19. Aggregate only after case matching; preserve N, basin, season, horizon, comparator, and metric.
20. Generate this report from the supplied source code, data tables, manifests, and website specification; do not infer exact production coefficients from candidate files.

References.

Scientific, Operational, and Project Sources

Ginis, N., and T. Marchok, 2022: The Hurricane Track Fit Consensus Model for Improving Hurricane Forecasting. arXiv:2210.16382. https://doi.org/10.48550/arXiv.2210.16382.

Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. Monthly Weather Review, 128, 1187–1193.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. Science, 285, 1548–1550. https://doi.org/10.1126/science.285.5433.1548.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. Journal of Climate, 13, 4196–4216.

National Hurricane Center, 2025: 2024 Hurricane Season Forecast Verification Report. NOAA/NWS/NHC.

National Hurricane Center, 2026: Forecast verification procedures, official forecast error database, and track/intensity model descriptions. NOAA/NWS/NHC, accessed 4 August 2026.

NOAA Hurricane Forecast Improvement Program, 2026: Hurricane Research Experiment and research-to-operations program information, accessed 4 August 2026.

Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). Bulletin of the American Meteorological Society, 81, 1231–1240.

Simon, A., A. Penny, M. DeMaria, J. L. Franklin, R. J. Pasch, E. N. Rappaport, and D. A. Zelinsky, 2018: A description of the real-time HFIP Corrected Consensus Approach (HCCA) for tropical cyclone track and intensity guidance. Weather and Forecasting, 33, 37–57.

STRIKE Tropical Cyclone Guidance Project, 2026a: STRIKE TC Models, module version 1.6.4 and method version 2.1.0. Source and immutable product schema.

STRIKE Tropical Cyclone Guidance Project, 2026b: STRK Trainer, version 1.11.0. Historical walk-forward replay, NHC-style scoring, candidate weights, and bias artifacts.

STRIKE Tropical Cyclone Guidance Project, 2026c: STRIKE Impact Corridor, module version 1.7.1, formula version triangulated_v1.

STRIKE Tropical Cyclone Guidance Project, 2026d: Public project website and research collaboration statement. strikesuperensemble.org.

Vijaya Kumar, T. S. V., T. N. Krishnamurti, M. Fiorino, and M. Nagata, 2003: Multimodel superensemble forecasting of tropical cyclones in the Pacific. Monthly Weather Review, 131, 574–583.

Williford, C. E., 2002: Real-time Superensemble Tropical Cyclone Prediction. Ph.D. dissertation, Florida State University, 144 pp.

Williford, C. E., T. N. Krishnamurti, R. C. Torres, S. Cocke, Z. Christidis, and T. S. V. Vijaya Kumar, 2003: Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. Monthly Weather Review, 131, 1878–1894. https://doi.org/10.1175/2571.1.

Public links

STRIKE SuperEnsemble project website

NHC forecast verification procedures

NHC official forecast error database

2024 NHC Forecast Verification Report

HFIP Hurricane Research Experiment

Hurricane Track Fit paper