SWANSTANT

WORKING PAPER — LIVE DOCUMENT, UPDATED AS VALIDATION PROGRESSES

Climate-Driven Cotton Yield Forecasting: A Cross-Geography Validation

François Souchet · Swanstant

April 2026 · Version 1.0

This is a live working paper. Metrics are updated as new growing seasons complete and are validated against realised outcomes.

Abstract. We present a climate-yield forecasting system validated across three structurally different cotton-producing geographies: US Georgia (18 counties, 29 years), US Texas (15 counties, 29 years), and India (116 districts, 23 years post-Bt) — together covering 149 administrative regions and approximately 25% of global cotton production.[0] The system uses phenologically-phased climate features derived from reanalysis data to forecast seasonal yield anomalies at sub-national resolution, with calibrated uncertainty intervals from a proprietary multi-stage quantification framework.

Walk-forward validation on held-out years demonstrates persistent cross-sectional signal: mean Information Coefficient (IC) of 0.29 with ICIR of 1.76 for India, IC of 0.49 with ICIR of 1.53 (1.24 HAC-corrected) for Texas, and aggregate directional accuracy of 100% (14/14 years) for Georgia. The system independently discovers different physical mechanisms per geography: ENSO-mediated mechanisms in Georgia and India, direct heat stress in Texas. The signal is robust across all ENSO regimes. Calibrated prediction intervals achieve near-nominal coverage via a proprietary multi-stage calibration framework. Yield-signal-implied Sharpe ratios (measuring forecast skill on the fundamental variable, not realised commodity trading returns) range from 1.20 to 1.37 across geographies. This paper demonstrates validated yield forecasting; Section 10 addresses the remaining steps to a fully tradeable signal, including factor neutralisation, a futures-conditioned backtest, and a live timestamped track record.

1. Motivation

Climate intelligence for commodity markets occupies a gap between two established disciplines: meteorological forecasting (which produces weather predictions without systematic, validated yield translation) and agronomic modelling (which simulates crop physiology without the probabilistic calibration and out-of-sample statistical validation that institutional decision-makers require). The result is that procurement teams and agricultural lenders routinely make seasonal-horizon decisions with either no climate input or with ad hoc qualitative assessments; even on commodity trading desks with dedicated meteorological capability, the yield and cost translation step remains largely manual and unvalidated.

This paper presents a system designed to bridge that gap. Our forecasting model produces calibrated, probabilistic yield forecasts at sub-national resolution with 1–6 month lead times. The architecture is universal: the same approach to feature engineering, validation, and uncertainty quantification applies to every crop and geography, while crop-specific components handle local phenology, climate variables, and data sources.

We focus on cotton as the primary validation case because it spans both irrigated and rainfed production systems, multiple climate transmission mechanisms, and geographies with fundamentally different climate-yield coupling. The three case studies presented — US Georgia, US Texas, and India — together cover 149 administrative regions (counties and districts) representing approximately 25% of global cotton production.[0] An additional US state (California, 7 counties) is in the pipeline but does not yet meet our minimum validation threshold for publication.[1] If the same methodology produces validated forecasts across these structurally different settings, that is evidence of a transferable signal extraction framework, not a geography-specific curve fit.

---

2. Methodology Framework

This paper describes the validation framework, uncertainty quantification, and signal evaluation methodology. The specific climate features, model architectures, and proprietary pipeline details are not disclosed.

2.1 Phenological decomposition

Each growing season is decomposed into biologically meaningful phases (pre-planting, planting/sowing, vegetative growth, flowering, boll development, maturation). Climate variables are computed separately for each phase, creating a feature space that respects the biological reality that the same climate variable can have opposite effects at different growth stages. This decomposition is the primary mechanism for generating candidate features, which are then subjected to automated selection.

2.2 Feature selection and model competition

From up to 690 candidate climate features per geography, automated selection identifies 3–8 features per model. Our machine learning algorithm evaluates candidates independently per geography and per regional cluster — features that win in Georgia are not assumed to win in Texas or India, and features that win in one Indian climate cluster are not assumed to win in another.

2.3 Walk-forward validation

All reported metrics use strict walk-forward (expanding window) validation with no future data leakage. For a panel of N entities (counties or districts) over T years, each fold trains on years 1…t and tests on year t+1 across all entities simultaneously. Models are cloned per fold to prevent state contamination. This protocol is the honest-test standard: it simulates the real operational setting where the model must forecast the next season using only data available up to the current point.

2.4 Uncertainty quantification

Prediction intervals are constructed through a proprietary multi-stage calibration framework that jointly propagates climate scenario uncertainty (51 ECMWF SEAS5 ensemble members) and model parameter uncertainty (200 parameter draws per scenario, yielding 10,200 Monte Carlo samples per forecast). The framework produces intervals with three properties that distinguish them from naïve model-based confidence intervals:

Forecast-responsive width. Intervals are wider when the model is less certain about a specific prediction and narrower when model components agree — they are not fixed-width bands. This means the uncertainty estimate is itself informative: a wide interval signals genuine ambiguity, while a narrow interval signals convergent evidence.

Calibrated coverage. Raw Monte Carlo intervals are adjusted to achieve near-nominal coverage rates (80% and 90%) as measured against held-out walk-forward observations. The calibration procedure is applied out-of-sample, ensuring the coverage guarantee is honest.

Robustness floor. The framework prevents overconfident intervals in low-disagreement regimes, ensuring that even when model components agree strongly, the interval respects the irreducible uncertainty in seasonal climate-yield prediction.

The specific combination and sequencing of calibration techniques is proprietary. The observable result — near-nominal coverage with forecast-responsive widths — is validated in the case studies below.

2.5 Regional clustering

For large, climatically heterogeneous geographies (India), the national panel is segmented into regional clusters using unsupervised learning on climate-response features. Each cluster receives its own feature selection and model training, with cluster-level forecasts aggregated to national estimates using production weights. The key empirical finding is that **zero features overlap across clusters**, confirming that district-level climate-yield coupling is genuinely heterogeneous and that a single national model would miss important regional variation.

3. Case Study: US Cotton (Georgia)

PANEL: 18 COUNTIES · 1996-2024 · 511 OBSERVATIONS

Georgia cotton is predominantly irrigated across the humid subtropical US Southeast with established USDA data infrastructure. The 18-county panel provides moderate cross-sectional breadth with deep time-series coverage (29 years). Walk-forward validation is conducted over 14 held-out years (2011–2024).

METRIC	VALUE	INTERPRETATION
Aggregate directional accuracy	100% (14/14)	Every out-of-sample year correctly classified as above or below trend.[6] For context, USDA's August crop production forecast — issued mid-season with observed data — achieves cotton yields within 10% of the final estimate in only 74% of years (14 of 19 years historically), with an average absolute error of 20.9 lb/acre.[4]
Aggregate Spearman IC	0.924	Near-perfect rank correlation between predicted and realised anomalies at state level
Aggregate Pearson IC	0.909	Magnitude fidelity, not just rank ordering
Signal Sharpe (state)	1.20	Year-aggregate directional betting Sharpe
Cross-sectional IC (mean)	0.088	Limited by breadth: 18 counties per cross-section
ICIR	0.33	Low power per year due to small N
Positive IC years	9/14 (64%)	Cross-sectional signal is real but noisy
CRPSS vs climatology	+0.675	Strongest probabilistic skill score across all regions
Calibrated coverage (80/90)	89% / 89%	Near-nominal coverage after proprietary calibration
Backtest RMSE	4.5%	Absolute forecast error on yield anomaly scale
All actuals within CI90	14/14	Zero interval exceedances over backtest period

GEORGIA SIGNAL PROFILE

Georgia's strength is at the aggregate level: 14/14 directional accuracy over 14 years, IC of 0.924, and the highest CRPSS of any crop (+0.675). The cross-sectional signal is weaker because 18 counties provide limited statistical power for within-year rank correlation. For institutional purposes, Georgia demonstrates that our forecasting system **calls the direction of a specific commodity market with near-perfect accuracy at seasonal frequency**.

---

4. Case Study: US Cotton (Texas)

PANEL: 15 COUNTIES · 1996–2024 · 414 OBSERVATIONS · 3 FEATURES

Texas is the largest US cotton-producing state (~40% of US output) and operates under a fundamentally different climate regime from Georgia. Across the High Plains and Rolling Plains—which together comprise approximately 4.17 million planted acres (3.83M High Plains + 0.34M Rolling Plains in 2024) of predominantly dryland cotton production in semi-arid conditions[5]—the dominant climate risk is direct heat stress during flowering and boll development, a different transmission mechanism from the ENSO-mediated and moisture-driven signals that dominate in Georgia and India.

The system selected only 3 features from 690+ candidates, the most parsimonious model across all geographies. All three are heat-stress variables concentrated in the reproductive phases (flowering and boll development).

METRIC	VALUE	INTERPRETATION
Cross-sectional IC (mean)	0.485	Strongest cross-sectional signal of any geography
ICIR (IC / Std)	1.53	Stable signal: positive IC in 13 of 14 test years
ICIR — HAC-corrected	1.24	Newey-West correction for serial correlation (lag-1 $\rho = 0.59$); see footnote 7
Overall Spearman IC	0.537	Pooled rank correlation, $p < 10^{-15}$
Panel directional accuracy	69.6%	133/191 county-year predictions correct
Aggregate directional accuracy	67% (6/9)	State-level aggregate backtest
Signal Sharpe (year-aggregate)	1.27	Year-aggregate directional betting Sharpe
CRPSS vs climatology	+0.627	Strong probabilistic skill, comparable to Georgia
Calibrated coverage (80/90)	89% / 100%	Excellent coverage after proprietary calibration
Features selected	3 (of 690+)	Extreme parsimony, all heat stress during reproductive phases
Positive IC years	13/14 (93%)	Near-universal cross-sectional signal presence

TEXAS SIGNAL PROFILE

The extreme parsimony (3 features) and high ICIR (1.53; 1.24 HAC-corrected, see footnote 7) of the Texas cotton model suggest a clean, robust signal that is unlikely to be an artefact of overfitting. Texas also illustrates the system's readiness gate's intended behaviour: the 2026 forecast was correctly classified as SKIPPED ("0/3 features from completed phases") because flowering and boll development have not yet occurred and the system refuses to forecast when its features are not yet observable.

5. Case Study: Indian Cotton

PANEL: 116 DISTRICTS · 2002–2024 (POST-BT ONLY) · 2,406 OBSERVATIONS · 5 REGIONAL CLUSTERS

Indian cotton represents a predominantly monsoon-dependent system across a climatically heterogeneous subcontinent. The panel is restricted to the post-Bt era (2002 onwards) because the adoption of Bt cotton fundamentally altered the climate-yield response function — pre-Bt data represents a different crop regime. Extending to 2024 includes DES (Directorate of Economics and Statistics) yield data integrated via source-specific anomaly rebasing to handle the systematic measurement discontinuity between ICRISAT and DES reporting.

The 116-district panel is segmented into 5 regional clusters via unsupervised climate-response clustering. Cluster production weights range from 5.2% to 55.4% of national output.

METRIC	VALUE	INTERPRETATION
Cross-sectional IC (mean)	0.292	Consistently ranks which districts will outperform across 15 test years
ICIR (IC / Std)	1.76	Signal is nearly twice as stable as it is strong
Positive IC years	14/15 (93%)	Cross-sectional signal is present in nearly every year
Overall Spearman IC	0.359	Pooled rank correlation, $p < 10^{-47}$
Panel directional accuracy	65.4%	1,009/1,542 district-year predictions correct
Signal Sharpe (year-aggregate)	1.37	Aggregate annual signal return / variability
CRPSS vs climatology	+0.569	Substantial probabilistic skill above base rate
Backtest directional accuracy (national)	62.5% (5/8)	National aggregation loses cross-sectional richness
Max drawdown (panel signal)	-5.2%	Shallow; 13/15 years produce positive signal returns
IC persistence (lag-1 autocorrelation)	-0.12	Near-zero: each year's IC is independent of the last
Zero feature overlap across clusters	Confirmed	Each cluster selects entirely different climate drivers

INDIA SIGNAL PROFILE

India's strength is in the cross-section: IC of 0.29, ICIR of 1.76, positive in 14 of 15 years. The 116-district breadth means the model consistently identifies **where within the cotton belt the climate-yield impact will be strongest**.

6. Regime Robustness

Georgia and India depend on ENSO teleconnections as primary forecast drivers. Texas uses a different mechanism entirely (direct heat stress). This provides a natural test of whether our signal identification is structurally dependent on ENSO.

REGIME	INDIA IC	INDIA HIT	INDIA SHARPE	GEORGIA IC	GEORGIA HIT	GEORGIA SHARPE	TEXAS IC	TEXAS HIT
El Niño	0.410	76.5%	1.84	0.715	74.3%	0.81	0.578	70.9%
La Niña	0.313	63.7%	1.36	0.444	68.2%	1.36	0.397	58.0%
Neutral	0.183	58.7%	1.66	0.428	64.4%	0.63	0.612	80.6%

For India, the signal is strongest during ENSO-active phases, weakest in neutral years — consistent with monsoon-mediated coupling. For Georgia, El Niño years produce the strongest signal (IC 0.715), while La Niña and neutral years are comparable (0.444 vs 0.428). For Texas, the signal is strongest in neutral years (IC 0.612, hit rate 80.6%) — consistent with a heat-stress mechanism that is independent of ENSO and operates under normal climatological conditions rather than teleconnection extremes. The Texas pattern is consistent with reduced yield variance from precipitation in neutral years: ENSO-active phases introduce additional moisture-driven variability that competes with the heat signal, lowering its precision rather than its physical relevance. All three geographies show positive IC across every regime. The combined result demonstrates that the system extracts robust signals regardless of whether the underlying mechanism is ENSO-mediated or direct thermal stress.

7. Cross-Geography Consistency

The three case studies present complementary signal profiles that together make a stronger case than any single geography:

DIMENSION	INDIA	GEORGIA	TEXAS	WHAT IT DEMONSTRATES
Primary signal strength	Cross-section	Aggregate	Cross-section	Complementary profiles
Cross-sectional IC	0.292	0.088	0.485	TX strongest; GA limited by 18-county N
CRPSS	+0.569	+0.675	+0.627	All three strongly beat climatology
Panel depth	2,406	511	414	3,331 total county/district-year obs
Features selected	7	6	3	Different counts, zero overlap
Dominant mechanism	ENSO × monsoon	ENSO + dewpoint	Direct heat stress	System discovers locally dominant driver
Irrigation regime	Rainfed	Irrigated	Dryland	Signal works across all water regimes

The critical finding is that different physical mechanisms emerge independently through automated selection in each geography. Georgia and India both select ENSO-related features through different transmission pathways — the system independently identifies ENSO as a relevant signal in both geographies without being told to look for it. Texas selects no ENSO features, instead finding direct heat stress during reproductive development. The divergent findings (different features, different mechanisms, different model parsimony) confirm that the system adapts to local conditions rather than applying a rigid template.

THE BREADTH ARGUMENT

The Fundamental Law of Active Management (Grinold 1989; Grinold & Kahn 1999) states $IR \approx IC \times \sqrt{N}$. India's cross-sectional breadth ($N \approx 103$ per year) translates IC = 0.29 into IR ≈ 2.96. Texas adds a further 15 counties with IC = 0.49 (IR ≈ 1.90). As our coverage expands to additional geographies and crops (the same approach already validates on California almonds and Ethiopian coffee), N grows further. Total validated administrative regions: 149 across three geographies, covering approximately 25% of global cotton production.[0] The scalability is a function of the architecture, not of any single geography's characteristics.

8. Signal Metrics Summary

METRIC	INDIA	GEORGIA	TEXAS	BENCHMARK
Cross-sectional IC	0.292	0.088	0.485	>0.05 exploitable; >0.15 strong
ICIR	1.76	0.33	1.53	>1.0 is stable signal
ICIR — HAC-corrected	–	–	1.24	Texas serial correlation correction (footnote 7)
Signal Sharpe	1.37	1.20	1.27	>1.0 strong at annual frequency[3]
CRPSS	+0.569	+0.675	+0.627	>0 beats climatology
Hit rate (panel)	65.4%	68.6%	69.6%	>60% exploitable
Calibration (80/90)	Multi-stage	89%/89%	89%/100%	Nominal 80/90 coverage
Regime robustness	Positive IC across all 3 ENSO regimes in all geographies			Not regime-conditional
Max drawdown	-5.2%	-4.5%	-6.9%	Shallow across all three
IC persistence (lag-1)	-0.12	-0.27	+0.59	TX shows persistence[2]

On yield vs. price. These metrics measure signal quality on the fundamental variable (yield anomalies), not on tradeable commodity returns. The mapping from yield forecast to position P&L introduces basis risk, timing mismatch, and the fact that any single geography's yield is one input among many to global cotton pricing. We report signal-implied Sharpe as an upper bound on the fundamental signal's information content, not as a claim of realised trading performance. A forthcoming extension (Section 10) addresses the yield-to-price translation directly.

---

9. Validation Gating Framework

Not every model that runs through our pipeline qualifies for inclusion in this paper. We apply a three-tier gating framework that determines whether a geography is reported as *validated*, reported with disclosed caveats, or excluded entirely from validated publication. This framework is part of our methodological commitment: by stating the gates explicitly, we make our publication standard auditable and repeatable across geographies, crops, and future model iterations.

Tier 1 — Hard Gates (failure of any single gate disqualifies the geography)

Tier 1 gates are non-negotiable. A geography that fails any one of them cannot be reported as validated under any circumstance. They establish the minimum conditions for the model to have economic value at all.

GATE	THRESHOLD	WHAT IT TESTS
CRPSS	`> 0`	Probabilistic forecast must beat climatology baseline
Walk-forward R^2	`> 0`	Out-of-sample variance explanation must be positive
Walk-forward directional accuracy	`> 0.55`	Above noise floor for binary directional outcomes
Sign-agreement gate	`All features pass`	Each selected feature must show consistent directional sign across >55–65% of entities (counties or districts) — threshold scaled to panel size

Tier 2 — Quality Gates (all must pass for "validated" status)

A geography that passes Tier 1 but fails any Tier 2 gate is reportable but flagged as preliminary rather than validated. Tier 2 establishes the standard for institutional-grade reliability.

GATE	THRESHOLD	WHAT IT TESTS
LOO directional accuracy	> 0.60	Robustness to alternative resampling scheme
LOO R^2	> 0	Cross-validates walk-forward variance explanation
Magnitude flags	= 0	No coefficients implausibly large vs. literature
Feature selection ratio	< 0.05	Selected features < 5% of candidates — filtering discipline
Residual stationarity	$p < 0.05$	Residuals stationary over time (no drift)

Tier 3 — Diagnostic Flags (do not block publication; require explicit disclosure)

Tier 3 diagnostics describe known imperfections that do not invalidate the forecast but matter for how institutional users interpret and apply it. These are flagged in the relevant case studies and explained in terms of operational implications.

DIAGNOSTIC	TRIGGER	OPERATIONAL IMPLICATION IF FLAGGED
Residual autocorrelation	Detected	Year-over-year errors not fully independent; reduces effective sample size for time-series significance tests
Extreme-year stress test	Below threshold or not run	Directional accuracy on the K=3 most extreme observed years (|z| > 2 on year-mean residuals, with top-K fallback). Performance below threshold means the signal degrades in tail conditions; users in risk applications should size accordingly

A note on temporal stability. Across all three validated geographies, the temporal-stability diagnostic (which tests whether signal strength is constant across multi-year sub-windows of the historical period) returns False. This is not a defect specific to any single geography — it is an expected property of climate-yield forecasting in a non-stationary climate. The underlying physical relationships shift gradually as crop varieties evolve, irrigation infrastructure changes, climate variability intensifies, and structural shifts (such as the Bt cotton transition in India or aquifer depletion in semi-arid US production) reshape the response function. A model that claimed perfect temporal stability would either be overfitting to the training period or smoothing away regime change. We treat temporal instability as a systemic feature of the problem rather than a per-geography flag, and we address it through annual re-validation against newly realised data — the live monitoring programme described in Section 10 is designed precisely to detect any degradation early. Users should plan for annual re-calibration rather than treating historical backtest performance as a permanent guarantee.

Application to the geographies in this paper

GEOGRAPHY	TIER 1	TIER 2	TIER 3 DISCLOSURES	STATUS
Georgia (US)	Pass	Pass	Extreme-year stress test: 96.2% — signal holds robustly in tail years	Validated
Texas (US)	Pass	Pass	Residual autocorrelation present (HAC-corrected ICIR reported, footnote 7); extreme-year stress test: 80.0% — signal holds in tail years	Validated
India	Pass	Pass	Residual autocorrelation present; extreme-year stress test: 76.0% on national aggregate — signal holds in tail years, with reliability varying across districts[8]	Validated
California	Fail (CRPSS -0.22)	N/A	N/A — Tier 1 failure stops evaluation	Excluded

Tier 2 LOO metrics are reported using the highest-weighted base model in each geography's ensemble. The Super Learner meta-ensemble (used for production forecasts) is fitted on walk-forward out-of-fold predictions of base models and does not have a separate LOO block by design; LOO validation is run on each underlying base model individually.

10. Current Limitations and Roadmap

Transparency about what has and has not been demonstrated:

Demonstrated

Walk-forward validated yield forecasting with persistent cross-sectional and aggregate signal across three structurally different geographies. Multi-stage uncertainty quantification with near-nominal calibration. Regime robustness across ENSO states. Universal architecture producing validated forecasts on additional crops (almonds, coffee) using the same methodology.

Not yet demonstrated

Factor-neutralised residual signal. The IC has not yet been regressed against standard commodity factors (momentum, term structure, basis). The structural argument for signal orthogonality to price-based factors is strong — the input data (satellite-era climate reanalysis) is entirely independent of price, flow, and sentiment signals — but formal demonstration requires the completed backtest. This is in development using ICE cotton futures.

Tradeable backtest with transaction costs. Converting yield signal to futures P&L requires bid-ask spread estimates, market impact modelling, and position sizing rules. A long/short ICE Cotton No. 2 futures backtest conditioned on our directional signal is in development, with initial results expected Q3 2026.

Live timestamped track record. All reported metrics are from walk-forward backtesting on historical data. A live forecasting programme with timestamped, immutable monthly signals is now underway. Six to twelve months of live predictions against realised outcomes will constitute the first out-of-sample live track record.

Signal decay analysis. We expect the signal to sharpen as each growing season progresses (from preliminary ENSO-derived forecasts to actionable observed-climate forecasts), and we expect slower decay than price-based signals due to the physical (rather than behavioural) transmission mechanism. Formal decay analysis requires multiple years of monthly signal vintages.

11. Data Sources

All data inputs to our model are drawn from publicly available, institutionally maintained sources. No alternative data, proprietary sensors, or MNPI sources are used.

SOURCE	PROVIDER	COVERAGE	ACCESS
ICRISAT District-Level Data	ICRISAT (CGIAR)	India district yields, 1966–2017	Public, open access
DES Crop Statistics	Directorate of Economics & Statistics, India	India district yields, 2018–2025	Public, government publication
USDA NASS	US Dept. of Agriculture	US county yields, 1960–present	Public, open access API
ERA5 Reanalysis	ECMWF / Copernicus CDS	Global climate, 1940–present, hourly	Public, free registration
NASA POWER (MERRA-2)	NASA Langley	Global climate, 1981–present, daily	Public, open access API
SEAS5 Seasonal Forecast	ECMWF / Copernicus CDS	Global, 51-member ensemble, monthly	Public, free registration
NOAA CPC ENSO Diagnostics	NOAA Climate Prediction Center	ONI, ENSO outlooks	Public, open access

This public-data-only approach eliminates alternative data compliance concerns and ensures full reproducibility of the input pipeline. The proprietary value resides in the feature engineering, model architecture, calibration framework, and regional clustering methodology — not in the data.

12. Engagement

Swanstant engages through three models, matched to the depth of integration required:

Co-development partnership. Joint exploration of specific crops, geographies, or transmission pathways (yield-to-price, yield-to-quality) with shared validation. Suitable for institutional buyers seeking to internalise or extend the signal for their own coverage requirements. The universal pipeline architecture means new geographies can be validated in weeks, not months.

Signal subscription. Monthly probabilistic forecasts with calibrated intervals, delivered as structured data. Suitable for integration into existing commodity research or risk frameworks without requiring internal model development.

Research collaboration. Partnership on methodology extension — factor neutralisation, signal decay analysis, cross-commodity portfolio construction. Structured for research-oriented institutions or academic partners.

Contact: contact@swanstant.com

[0] Production coverage calculation. Global cotton production: ~25 million metric tons (USDA WASDE, MY 2024/25). India represents ~23% of global production; the United States ~13%, of which Texas accounts for ~40% of US output (~5.2% of global) and Georgia ~10% (~1.3% of global). Within each geography, our validated panels capture the bulk of national or state-level production: India 116 districts (~85% of national cotton production), Texas 15 counties (~75% of state production), Georgia 18 counties (~85% of state production). Combined coverage: 23% × 0.85 + 5.2% × 0.75 + 1.3% × 0.85 ≈ 24.6%, rounded to ~25% of global cotton production.

[1] California cotton (7 counties, 25 years, 127 observations) has been run through the full validation pipeline including feature selection, model competition, and walk-forward validation, but is excluded from validated reporting because it fails our Tier 1 gating criteria (CRPSS = -0.22, indicating the model does not yet beat climatology). See Section 9 for the full validation gating framework. We include California in our internal coverage tracking (bringing the total to 156 administrative regions across 4 geographies) but do not report its metrics as validated results.

[2] Texas's positive IC persistence (lag-1 autocorrelation = 0.59) contrasts with near-zero persistence in India and Georgia. This may reflect the underlying climatological structure: multi-year heat regimes in semi-arid Texas create autocorrelated climate conditions, whereas ENSO-mediated signals alternate more rapidly. The institutional implication is that the Texas signal may exhibit higher serial correlation in trading applications, requiring appropriate position management. See also footnote 7 for the implications of this autocorrelation on statistical inference.

[3] Sharpe ratio benchmarks are frequency-dependent. Medium-frequency tradeable strategies (daily or weekly turnover) typically target Sharpe ratios of 1.5–2.5 because they compound across many independent bets per year. Annual-frequency fundamental signals operate on a different basis: a single observation per geography per year limits the achievable Sharpe regardless of signal quality. At annual frequency, Sharpe > 1.0 is strong and Sharpe > 1.5 is exceptional. The reported Sharpes (1.20–1.37) all clear the strong threshold. **Strategy specification:** Signal Sharpe is computed as the mean annual return of an equal-weight ±1 position (long the geography when the aggregate signal is positive, short when negative) divided by the standard deviation of annual returns. No transaction costs, market impact, leverage, or signal-strength-proportional position sizing are applied. The live tradeable backtest (in development, see Section 10) will report Sharpe at the appropriate trading frequency with realistic costs.

[4] Source: Southern Ag Today, "How Is the Cotton Crop Looking in 2023?" (August 2023), based on historical USDA NASS Crop Production reports. Available at southernagtoday.org.

[5] Sources: Plains Cotton Growers, 2024 PCG Production Report; Texas A&M Soil & Crop Sciences, 2024 Rolling Plains RACE Report. Acreage figures represent total planted cotton acreage across the Texas High Plains and Rolling Plains regions for the 2024 crop year (irrigated and dryland combined).

[6] Statistical significance of Georgia 14/14 directional accuracy. Under a null hypothesis of random directional guessing ($p = 0.5$), the probability of 14 consecutive correct calls is $(0.5)^{14} \approx 0.006\%$ ($p < 10^{-4}$). A more stringent benchmark is naïve persistence (predicting that this year's direction matches last year's): on the same 13-year overlapping window, persistence achieves 7/13 = 53.8%, while our model achieves 13/13 = 100% over the same window (the persistence benchmark requires one prior-year observation, reducing the comparable window by one year from the full 14-year backtest). The signal therefore adds substantial information beyond both random guessing and the most basic time-series benchmark.

[7] HAC correction for Texas ICIR. The naive ICIR computation assumes year-over-year IC values are independent draws. Texas's lag-1 autocorrelation of 0.59 (footnote 2) violates this assumption, inflating the apparent stability of the signal. Applying a Newey-West (HAC) correction with one lag yields an HAC-corrected ICIR of 1.24, compared to the naive 1.53. Equivalently, the effective sample size shrinks from 14 to approximately 3.6 independent observations. The corrected ICIR remains above 1.0 (the strong-signal threshold for annual-frequency signals) but the underlying time-series evidence for Texas is necessarily limited by the short 14-year backtest window. The cross-sectional evidence is not affected by this autocorrelation: the pooled overall Spearman IC of 0.537 across 191 county-year observations ($p < 10^{-15}$) pools across districts and years simultaneously, and remains highly significant. Future updates will report HAC-corrected estimates as additional years of out-of-sample data accumulate.

[8] Extreme-year stress test methodology and India result. The stress test evaluates directional accuracy on the K=3 most extreme observed years per geography, identified by $|z|$-score > 2 on year-mean residuals (with a top-K fallback when no observations exceed the strict z-threshold). For India, the test selects 2002, 2023, and 2024 — capturing both the post-Bt era's earliest year and two of the most volatile recent kharif seasons. The model achieves 76.0% directional accuracy on these years at the production-weighted national aggregate. Underlying this aggregate, reliability varies across the regional clusters that compose the Indian forecast: the dominant cluster (representing the majority of national production) shows high directional accuracy on extreme years, while smaller clusters with lower production weights are more variable. Because the published forecast is a production-weighted national aggregate, the user-facing signal benefits from the dominant cluster's reliability. Users with regionally-concentrated exposure should request cluster-level diagnostics directly. We continue to refine cluster-level stress testing as additional years of out-of-sample data accumulate.

This document describes methodology and validation outputs. Proprietary feature selection, model internals, and pipeline architecture are not disclosed.