XC-SPI Technical Specification

Version 2.7: roster-adjusted preseason index, dynamic athlete model, and exact NCAA championship simulation

Colin C. Caso

Oklahoma State University

August 28, 2026

Abstract

This document is the formal statistical specification for XC-SPI after the August 2026 model-design audit. The revision preserves the frozen Week 0 rankings while tightening what the equations are allowed to mean. Preseason Score is defined as a roster-adjusted scoring index rather than an expected NCAA score; preseason evidence confidence is separated conceptually from modeled availability; the depth term is explicitly a roster-fragility penalty; state uncertainty grows with elapsed time; observation uncertainty includes uncertainty in context adjustment; the production model adds a team-by-race shared shock and allows difficult conditions to alter residual variance; the optional Bradley–Terry layer is removed from the production specification unless future validation demonstrates unique information; lineup availability and selection are separated; Monte Carlo stopping is adaptive; and championship-probability calibration requires a longer multi-season record than a single held-out NCAA final. The former Preseason Score-to-lognormal preseason odds bridge is deprecated and is no longer treated as a public probability generator.

1. Objects and Notation

Symbol	Definition
T_{ir}	Official time for athlete i in race r.
$s(i)$	Sex/model stratum for athlete i; men's and women's performance models are fit separately or with explicit sex-specific parameters.
$\alpha_{i,t}$	Latent current cross-country ability on the log-time scale; smaller is faster.
Δt_{it}	Elapsed time since the athlete's previous informative state update, on the model's chosen time unit.
d_{ir}	Race distance.
γ_c	Persistent course effect, centered on the reference scale.
w_r	Race-window weather vector.
q_r	Antecedent moisture/ground vector.
u_r	Race-wide residual effect used primarily to diagnose unmeasured common race context.
v_{jr}	Team-by-race shared performance shock for team j in race r.
c_i	Week 0 Confidence weight used only in Preseason Score; not a probability.
p_i^{avail}	Modeled probability that athlete i is available to compete, only when historically estimable.
p_i^{select}	Modeled or rule-based probability/decision that an available athlete is selected to the championship group.
$P_i^{(m)}$	Raw overall championship finish place in simulation m.
$Q_i^{(m)}$	NCAA team-scoring place after the official scoring filter.

Symbol	Definition
$S_j^{(m)}$	Team j NCAA score in simulation m.

2. Week 0 XC-SPI Preseason Score

The Week 0 system is a transparent index, not a fitted probability model. Define the uncertainty-adjusted contribution

$$C_i = c_i X_i + (1 - c_i) R_0, \tag{1}$$

where X_i is the athlete's comparable historical scoring value, c_i is a declared Confidence weight, and R_0 is the replacement baseline. The team index is

$$F_j = \lambda[(C_{j(6)} - C_{j(5)})_+ + (C_{j(7)} - C_{j(5)})_+], \tag{2}$$

$$PS_j = \sum_{k=1}^{5} C_{j(k)} + F_j. \tag{3}$$

Here F_j is the **Depth Risk**. It is a preseason roster heuristic: it does not attempt to reproduce the actual NCAA displacement points created by runners six and seven. Those displacement effects are handled only in the full championship simulator.

The launch constants remain $R_0 = 200$, $X_{\text{new}} = 150$, and $\lambda = .05$. They are declared design choices and sensitivity-tested. Because c_i is not a participation probability, C_i is not a mathematical expectation and Preseason Score must not be described as predicted or expected NCAA points.

3. Preseason Athlete Distribution

Before a current-season cross-country result, athlete ability follows a sex-specific or sex-stratified prior such as

$$\alpha_{i,0} \sim N\left(\theta_{0,s(i)} + \theta_{1,s(i)} XC_{i,2025} + \theta_{2,s(i)} TRK_{i,2026},\ V_{i,0}\right). \tag{4}$$

Track is standardized recency evidence rather than a direct XC-time conversion. Prior variance increases with sparse, old, weakly comparable, or uncertain evidence.

4. State Evolution with Irregular Time

Athlete fitness should become more uncertain as time passes without informative evidence. Use

$$\alpha_{i,t}^- = \alpha_{i,t-1} + \eta_{i,t}, \tag{5}$$

$$\eta_{i,t} \sim N(0, \sigma^2_{\eta,s(i)} \Delta t_{it}), \tag{6}$$

$$V_{i,t}^- = V_{i,t-1} + \sigma^2_{\eta,s(i)} \Delta t_{it}. \tag{7}$$

This is a continuous-time-scaled random-walk approximation: five weeks without evidence create more state uncertainty than five days. The time scale and process variance are estimated chronologically.[5]

5. Race Observation Model

For athlete i from team $j(i,r)$,

$$\log T_{ir} = \mu_{s(i)} + \alpha^{-}_{i,t(r)} + \delta_{s(i)}(d_{ir}) + \gamma_{c(r),s(i)} + h_{s(i)}(w_r) + g_{s(i)}(q_r) + u_r + v_{j(i,r),r} + \epsilon_{ir}. \tag{8}$$

The model separates current athlete ability from distance, persistent course context, weather/ground context, common race-wide misspecification, a shared team-race performance component, and athlete-specific residual noise. Robust/heavy-tailed residuals are permitted if held-out diagnostics justify them.

6. Distance

Use a centered penalized spline

$$\delta_s(d) = \sum_{b=1}^{B} \kappa_{b,s} B_b(\log d), \qquad \sum_r \omega_r \delta_s(d_r) = 0, \tag{9}$$

with sex-specific functions retained only if chronological validation supports them. Centering prevents the distance basis from absorbing the intercept.[4]

7. Course Hierarchy and Identifiability

Persistent course effects are partially pooled:

$$\gamma_{k,s} \sim N(x_k^\top \beta_{c,s}, \tau_{c,s}^2), \qquad \sum_k \pi_k \gamma_{k,s} = 0. \tag{10}$$

A connected network of athletes and repeated races is required to distinguish athlete ability from course effects. Sparse courses are strongly shrunk toward the structural course prediction. Large stand-alone course penalties are not allowed from one unusual race.[3]

8. Weather, Ground, and Condition-Dependent Variance

Let w_r contain temperature, dew point, wind, and race-window precipitation. Let

$$q_r = (P_{2h:24h}, P_{24:72h}, P_{72h:7d}, \text{surface}, \text{drainage}, \text{ground evidence}).$$

The mean functions $h_s(w_r)$ and $g_s(q_r)$ help normalize historical performances. Because all athletes in a future championship share the same course and common weather, purely additive common-condition terms cancel from pairwise ordering. Therefore difficult conditions may also affect race-to-race unpredictability through a residual-scale model such as

$$\log \sigma_{\epsilon,r,s} = \xi_{0,s} + \xi_s^\top z_r, \tag{11}$$

where z_r may include heat stress, mud/ground severity, wind, and technicality. This allows hard conditions to widen the performance distribution without claiming unsupported athlete-specific "mud ability." Athlete-by-condition random slopes should be added only if repeated historical evidence supports them.

9. Race-Wide and Team-Race Effects

A centered common residual may be retained as

$$u_r \sim N(0, \sigma_u^2), \tag{12}$$

principally as a diagnostic for systematic race-wide misspecification. Team-correlated execution is represented by

$$v_{jr} \sim N(0, \sigma_{\text{team}}^2), \tag{13}$$

so runners from the same team can move modestly together on an unusually good or bad team day. This correlation matters for the tails of championship team-score distributions.

10. Normalized Observation and Measurement Uncertainty

Let the context-adjusted observation be

$$z_{ir} = y_{ir} - \hat{\mu}_s - \hat{\delta}_s(d_{ir}) - \hat{\gamma}_{c,s} - \hat{h}_s(w_r) - \hat{g}_s(q_r) - \hat{u}_r - \hat{v}_{jr}. \tag{14}$$

For an explanatory sequential update, define observation variance

$$R_{ir} = \sigma_{\epsilon,r,s}^2 + V_{\text{course,ir}} + V_{\text{weather,ir}} + V_{\text{ground,ir}} + V_{\text{measurement,ir}}. \tag{15}$$

Then

$$K_{ir} = \frac{V_{i,t}^-}{V_{i,t}^- + R_{ir}}, \tag{16}$$

$$\hat{\alpha}_{i,t} = \hat{\alpha}_{i,t}^- + K_{ir}(z_{ir} - \hat{\alpha}_{i,t}^-), \tag{17}$$

$$V_{i,t} = (1 - K_{ir})V_{i,t}^-. \tag{18}$$

The production fit may use the equivalent joint hierarchical posterior rather than literal closed-form updates. The equations expose the intended information flow: a poorly measured or poorly contextualized race moves the athlete less.

11. Head-to-Head Information

The production v2.7 model does **not** add a Bradley–Terry likelihood on top of the time likelihood. Same-race ordering is already encoded in observed times, so an additional pairwise likelihood risks double-counting the same race information and introducing strong within-race dependence. A head-to-head component may be reconsidered only if a preregistered rolling-origin comparison demonstrates out-of-sample information beyond normalized time.

12. Availability and Lineup Selection

Availability and selection are separate concepts. If sufficient historical evidence exists,

$$p_i^{\text{avail}} = P(A_i = 1 \mid \mathcal{E}_t), \qquad A_i^{(m)} \sim \text{Bernoulli}(p_i^{\text{avail}}), \tag{19}$$

$$p_i^{\text{select}} = P(L_i = 1 \mid A_i = 1, \mathcal{E}_t). \tag{20}$$

Until these quantities can be estimated credibly, availability remains a transparent roster gate and championship-seven selection is rule-based from the current athlete distributions. Early-season absence is not an injury indicator without public evidence.

13. Championship Projection

For a selected and available athlete,

$$\log \tilde{T}_i^{(m)} = \mu_s + \tilde{\alpha}_{i,t}^{(m)} + \delta_s(d^*) + \gamma_{c^*,s} + h_s(w^{*(m)}) + g_s(q^{*(m)}) + u^{*(m)} + v_{j(i),*}^{(m)} + \tilde{\epsilon}_i^{(m)}, \tag{21}$$

with

$$\tilde{\epsilon}_i^{(m)} \sim F_s\left(0, \sigma_{\epsilon,*,s}^{(m)}\right). \tag{22}$$

Uncertainty in athlete state, course/environmental parameters, team-day effects, and residual scale is propagated. Common additive conditions affect absolute time; condition-dependent variance and any historically supported interactions are what can alter relative ordering.

14. NCAA Team-Scoring Algorithm

The simulator retains raw overall finish $P_i^{(m)}$ for individual results and constructs a separate NCAA team-scoring order $Q_i^{(m)}$ consistent with NCAA cross-country rules.[1;2]

1. Sort all finishers by simulated overall performance.
2. Determine whether each institution has at least five finishers from its declared championship group; incomplete teams are DNF for team scoring.
3. Among complete teams, only their first seven finishers are team-scoring/displacement eligible.
4. Remove individual qualifiers, athletes from incomplete teams, and runners beyond a complete team's seventh from the team-scoring sequence.
5. Re-number the retained sequence $1, 2, \ldots$ to obtain $Q_i^{(m)}$.
6. Sum the first five Q values for each complete team. Sixth and seventh retain Q values and displace opponents but do not add to their own subtotal.
7. Resolve ties using the NCAA scorer-by-scorer comparison procedure.

15. Team and Individual Outputs

$$S_j^{(m)} = \sum_{k=1}^{5} Q_{jk}^{(m)}, \tag{23}$$

$$\bar{S}_j = M^{-1} \sum_m S_j^{(m)}, \tag{24}$$

$$\widetilde{S}_j = \text{median}_m \, S_j^{(m)}, \tag{25}$$

$$\overline{RANK}_j = M^{-1} \sum_m \text{teamplace}_j^{(m)}, \tag{26}$$

$$\widehat{P}_j(\text{title}) = M^{-1} \sum_m I(\text{teamplace}_j^{(m)} = 1), \tag{27}$$

$$\widehat{P}_i(\text{individual title}) = M^{-1} \sum_m I(P_i^{(m)} = 1). \tag{28}$$

The weekly **power ranking** is based primarily on expected simulated NCAA score, with median score, expected team place, and calibrated event probabilities reported separately. "Best expected scoring profile" is not treated as mathematically identical to "highest title probability."

16. Monte Carlo Diagnostics and Adaptive Stopping

Simulation count is determined by precision, not by a single universal minimum. Start with a substantial pilot batch (for example 50,000 draws) and continue until the published quantities meet preregistered tolerances. For a probability,

$$\mathrm{MCSE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{M}}. \tag{29}$$

For common probabilities, require a small absolute MCSE; for longshots, also require a relative-MCSE or minimum-event-count criterion so a probability based on only a handful of simulated wins is not displayed with false precision. Expected-score MCSE is $s(S_j)/\sqrt{M}$. Report M, seed, MCSE, stopping criterion, and any convergence warning.

17. Missing Data and Source Quality

All inputs carry provenance and source grade. Missing course/ground data widen context or observation uncertainty rather than triggering an undocumented correction. Contradictory sources are resolved by primary-source priority or retained as unresolved. No undocumented manual seconds/points correction is permitted.

18. Postmodern Era Historical Success Model

The historical benchmark is separate from both the Preseason Score and the athlete-level race simulator. XC-SPI defines the **Postmodern Era (Super-Shoe Era onward)** as 2020 forward.[8] For source season $s \in \{2020, \dots, 2024\}$, let $\mathbf{z}_{js}$ contain standardized Team Score, Front Runner, Fifth Runner, and Depth Gap for team j. The next-season NCAA outcome is coded into five categories: champion; 2nd–4th; 5th–10th; 11th–20th; other. A regularized multinomial model estimates

$$P(Y_{j,s+1} = k \mid \mathbf{z}_{js}) = \frac{\exp(\alpha_k + \boldsymbol{\beta}_k^T \mathbf{z}_{js})}{\sum_{\ell=1}^{5} \exp(\alpha_\ell + \boldsymbol{\beta}_\ell^T \mathbf{z}_{js})}. \tag{30}$$

Candidate feature subsets and regularization strengths are compared by held-out-season cross-validation. The selected 2026 historical benchmark uses the four-variable profile and moderate L_2 regularization. The disrupted 2020 source season remains in the main Postmodern Era fit and is removed in a declared sensitivity rerun.

For the 2026 historical benchmark, the current audited adjusted-runner profile is standardized within sex and mapped into the same four profile dimensions. The model returns direct Historical Title, Podium, Top-10, and Top-20 rates. Historical Success Rank is ordered primarily by podium rate because title is sparse. The Historical Success Model is a profile-persistence model: it does not reconstruct every historical transfer or injury and must not be described as the final current-athlete probability model.

For a coherent title-only betting-style display among the published Top 25, raw historical title rates r_j are normalized:

$$p_j^H = \frac{r_j}{\sum_{\ell=1}^{25} r_\ell}. \tag{31}$$

This is explicitly a **Top-25-conditional Historical Title Market**; it is not full-field current XC-SPI title probability. Podium and top-10 rates remain direct binary-event probabilities and are not normalized across teams because multiple teams can satisfy those outcomes.

19. Calibration and Validation

The Historical Success Model uses held-out-season cross-validation across its strict 2020-forward Postmodern Era sample and a separate exclusion test for the disrupted 2020 source profile. The athlete-level weekly model still requires rolling-origin evaluation with no future leakage. Full current-model championship calibration should use as many reconstructable seasons as feasible and should not be inferred from the Historical Success benchmark alone; the Postmodern Era sample contains only five next-season outcome years and ten champions across men and women. Report next-race MAE/RMSE, rank correlation, team-place and score error, interval coverage, Brier/log score, reliability, and calibration with sample-size context.[6;7]

20. Ablation and Simplification

Compare the full model with variants removing course, weather/ground, track prior, dynamic state, team-race shock, condition-dependent residual variance, and any future optional interaction. A component survives only if it improves held-out performance, calibration, or uncertainty quality enough to justify the added complexity.

21. Probability and Betting Release Gate

The former Week 0 Preseason Score-to-lognormal odds bridge is deprecated because it imposed ratio-scale meaning on an index whose zero point and scaling constants are design choices. Historical Title Odds may be published now when they are clearly labeled as Top-25-conditional outputs from the Postmodern Era Historical Success Model. Full current-model Championship Odds should enter the primary public presentation only when they come from the athlete-level championship simulation and have passed the calibration gate. Fair-odds, implied-probability, vig, EV, and Kelly equations remain valid translations once a defensible probability p exists; they do not create probability validity themselves.

22. Versioning

Each release stores data cutoff, code version, parameter version, source manifest, random seed, simulation count, validation split, and change log. A published historical ranking is not silently rebuilt with later information. The frozen Week 0 Preseason Score order is unchanged by the v2.7 conceptual revision; the revision changes interpretation and the future probability architecture.

References

[1] National Collegiate Athletic Association. *2025 and 2026 NCAA Men's and Women's Track and Field and Cross Country Rules Book*. Accessed August 28, 2026. https://ncaaorg.s3.amazonaws.com/championships/sports/crosstrack/rules/PRXTF_RulesBook.pdf

[2] National Collegiate Athletic Association. Cross Country and Track and Field Playing Rules. Accessed August 28, 2026. https://www.ncaa.org/championships/playing-rules/cross-country-track-and-field-playing-rules/

[3] Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press; 2007.

[4] Wood SN. *Generalized Additive Models: An Introduction With R*. 2nd ed. Chapman & Hall/CRC; 2017.

[5] Durbin J, Koopman SJ. *Time Series Analysis by State Space Methods*. 2nd ed. Oxford University Press; 2012.

[6] Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev.* 1950;78(1):1–3.

[7] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc.* 2007;102(477):359–378.

[8] Caso CC. *XC-SPI Postmodern Era Historical Success Model: Super-Shoe Era onward, 2020–2025 team-profile persistence, 2026 success rates, and historical title odds.* Oklahoma State University; 2026.