How XC-SPI Works

A readable explanation of the v2.7 ranking model

Colin C. Caso
Oklahoma State University
August 28, 2026

Abstract

XC-SPI is designed around the fact that cross-country times are contextual and NCAA team scoring is nonlinear. The preseason release uses a transparent roster-adjusted index. The in-season system is designed to estimate athlete ability from repeated races, normalize race context, and simulate the actual NCAA scoring procedure. Version 2.7 tightens the interpretation of the Week 0 index and the future probability model without changing the frozen preseason ranking.

1. Start With Athletes, Not School Names

A team's 2025 NCAA score belongs to the athletes who produced it. A transfer takes the athlete's evidence to the new school; a professional departure no longer contributes to the old team. Week 0 therefore begins by reconstructing who owns each historical performance in 2026.

2. Week 0 Is an Index

For athlete i,

$$C_i = c_i X_i + (1 - c_i) R_0.$$

X_i is the historical scoring value, c_i is a rule-based confidence/shrinkage weight, and R_0 is the conservative replacement baseline. Because c_i is not a calibrated probability, C_i is not an expected finish. The team adds its five strongest C_i values and a small Depth Risk based on the gap from scorer five to runners six and seven. Lower Preseason Score means a stronger roster-adjusted preseason profile.

3. Why the Depth Term Is Only a Heuristic

Sixth and seventh runners matter in real NCAA scoring because they can displace opponents. The Week 0 fragility term does not try to calculate those displacement points. Actual displacement depends on every athlete in the national field, so the in-season championship simulator handles it directly.

4. Add a Historical Success Lens

XC-SPI calls 2020 forward the **Postmodern Era (Super-Shoe Era onward)**.[8] The Historical Success Model compares four simple profile features—Team Score, Front Runner, Fifth Runner, and Depth Gap—with next-season NCAA outcomes. It reports Historical Title, Podium, Top-10, and Top-20 rates. Historical Success Rank uses podium rate as its primary ordering because podium is less sparse than title.

This layer is deliberately separate from the Power Ranking. A team's Preseason Score can be excellent while its profile resembles historically volatile, top-heavy teams; another team can have a slightly worse current score but a more historically durable shape. The historical comparison does not rewrite the Preseason Score.

5. Athletes Change Over Time

A runner is not assumed to have one permanent rating. The athlete state is allowed to drift, with uncertainty increasing as time passes:

$$\eta_{it} \sim N(0, \sigma_\eta^2 \Delta t).$$

A long period without useful evidence therefore creates more uncertainty than a short gap. This is a standard state-space idea adapted to athlete performance.[7]

6. One Time Can Mean Different Things

A 24-minute race on a flat, dry course and a 24-minute race on a muddy, hilly course are not interchangeable. The observation model separates athlete ability from distance, course, weather/ground, race-wide effects, team-day effects, and individual noise. Course effects are partially pooled so a venue cannot acquire a dramatic adjustment from one strange race. Penalized smooth functions are used where a rigid linear assumption is not justified.[5;6]

7. Weather Has Two Jobs

Weather and ground conditions help explain why a historical race was faster or slower than expected. But if every athlete in a future championship receives the same additive weather term, that common term cancels when runners are compared. Version 2.7 therefore allows difficult conditions to change the **spread** of performances as well as their common mean. Mud, heat, wind, or technical footing may make outcomes more variable and create more upset risk. Athlete-specific "mud ability" is not introduced unless repeated data support it.

8. Teams Can Have Shared Good or Bad Days

Independent athlete errors would imply five runners from the same team have no shared race-day fluctuation after measured context is removed. That is too strong. XC-SPI therefore allows a small team-by-race random effect, which can represent shared execution, travel, illness, strategy, or other team-level variation without manually awarding a team bonus.

9. Normalize, Then Update

A new result is first adjusted for estimated context. The amount that it moves the athlete depends on observation uncertainty. A race with uncertain course, weather, ground, or timing information receives less effective weight than a well-measured race. The explanatory Kalman-style gain is

$$K = \frac{V^-}{V^- + R},$$

where R includes individual residual noise and uncertainty in the context correction.

10. No Extra Head-to-Head Layer by Default

Earlier drafts allowed an optional Bradley–Terry likelihood. Version 2.7 removes it from production because same-race times already contain the finishing order; adding a second pairwise likelihood risks double-counting the same information. It can return only if rolling validation shows unique predictive value.

11. Availability Is Not Selection

An athlete can be available but not selected to the championship seven. The future model therefore distinguishes availability from lineup selection. Until historical data support calibrated probabilities for those decisions, XC-SPI uses transparent roster gates and rule-based selection rather than converting uncertainty labels into fake probabilities.

12. From Athlete Distributions to NCAA Scores

Each simulation produces athlete performances in the same championship environment. The model retains raw overall finish for the individual race and then applies a separate NCAA team-scoring filter. Individual qualifiers, runners from incomplete teams, and runners beyond a complete team's seventh are removed from the team-scoring sequence. Five score; six and seven displace; ties use the NCAA scorer-by-scorer rule.[1]

For simulation m,

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

The primary power-ranking target is expected simulated NCAA score. Median score, expected team place, and title/podium probabilities are separate outputs because the team with the best mean score need not have the highest probability of winning in every distributional shape.

13. How Many Simulations?

There is no magic fixed number. The model starts with a large pilot batch and continues until the public quantities meet precision rules. Probability simulation error is

$$MCSE(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/M}.$$

For very small probabilities, relative precision and the number of simulated wins matter too. Millions of draws can make MCSE tiny while leaving model assumptions wrong, so simulation precision is never treated as model validation.[2]

14. What the Model Must Prove

Race and team components are evaluated chronologically against simpler baselines. Metrics include next-race MAE/RMSE, rank and team-place error, NCAA-score error, interval coverage, Brier/log scores, calibration, and ablation tests. National-title probability calibration requires many historical championships rather than one held-out NCAA final.[3;4]

15. Why the Old Preseason Odds Board Was Withdrawn

The former Week 0 odds bridge treated Preseason Score as the mean of equal-dispersion lognormal team distributions. That shortcut remains retired. Version 2.7 instead publishes separate Postmodern Era Historical Title Odds: raw historical title rates are normalized across the displayed Top 25 to form a coherent conditional historical market. Full current-model fair prices still wait for the validated athlete-level championship simulator.

References

[1] 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/

[2] Metropolis N, Ulam S. The Monte Carlo method. *J Am Stat Assoc*. 1949;44(247):335–341.

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

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

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

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

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

[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.