XC-SPI Model Limitations & Responsible Interpretation

Version 2.7 boundaries on what the rankings, simulations, and odds can mean

Colin C. Caso
Oklahoma State University

August 28, 2026

Abstract

XC-SPI is designed to make uncertainty visible rather than to remove it rhetorically. Version 2.7 follows a model-design audit that preserved the frozen Week 0 ranking but narrowed several interpretations and retired the former public preseason probability bridge. This statement records what the current system can support and what remains provisional.

1. Preseason Score Is an Index, Not Expected NCAA Points

Preseason Score combines historical athlete scoring evidence, evidence-confidence shrinkage toward a replacement baseline, five-runner ordering, and a small Depth Risk. Its constants are declared and sensitivity-tested. Because the Confidence weight is not a literal participation probability, the adjusted contribution is not a mathematical expectation. A Preseason Score value is therefore a lower-is-better Week 0 index, not a predicted NCAA score.

2. The Depth Term Is Roster Fragility, Not Displacement

The Week 0 depth term measures how far runners six and seven sit behind scorer five in the index. Actual NCAA displacement depends on where those runners finish relative to the entire national field. Only the full championship simulator models displacement directly.

3. Roster Confidence and Availability Are Different

A runner can be clearly listed on the roster while current performance evidence is weak, or have excellent historical evidence while actual championship availability remains uncertain. Version 2.7 therefore separates evidence confidence from future modeled availability/selection. The Week 0 0.99/0.85/0.70/0 values must not be quoted as participation probabilities.

4. Course and Weather Effects Are Not Magic Reordering Terms

Additive course/weather terms help normalize historical races. In a future championship, common additive terms apply to everyone and therefore do not by themselves change relative order. XC-SPI allows difficult conditions to change residual variance and will add athlete-specific interactions only if repeated data support them.

5. Team Performances Are Correlated

Runners from the same team may share an unusually good or bad day. Version 2.7 includes a team-by-race shared shock in the planned model rather than treating all athlete residuals as independent after context adjustment.

6. The Production Model Does Not Double-Count Head-to-Head

A separate Bradley–Terry layer is not used in v2.7 because finishing order is already encoded in the same-race clock. It can return only if preregistered rolling validation shows incremental information.

7. Expected Score Is Not Title Probability

The primary power ranking is designed around expected simulated NCAA score. That is not mathematically identical to highest title probability. XC-SPI will therefore publish expected score, median score, expected place, and calibrated event probabilities as distinct quantities rather than using them interchangeably.

8. Monte Carlo Precision Is Not Model Accuracy

A tiny MCSE only means the simulation has precisely estimated the output of the assumed model. It does not prove athlete distributions, course effects, availability, or championship probabilities are correct. Longshot probabilities also require relative precision/minimum-event checks, not merely a fixed total draw count.

9. Title Calibration Needs More Than One NCAA Final

Weekly athlete/race prediction can be evaluated across many races in recent seasons. National-title probability calibration is much rarer: there is only one men's and one women's champion per season. A single held-out NCAA championship cannot establish that 10%, 30%, or 60% probabilities are calibrated. A longer historical championship archive is required.

10. The Former Week 0 Odds Board Is Deprecated

The v2.5 scenario board mapped Preseason Score into equal-dispersion lognormal team distributions. The model was transparent, but it implicitly treated Preseason Score ratios as meaningful. Because the Preseason Score origin/scale depend partly on design constants, that assumption is too strong for public "fair price" language. Those values are retained internally for provenance but withdrawn from the current public probability presentation.

11. Betting Arithmetic Does Not Validate a Probability

Fair odds, implied probability, vig removal, expected value, and Kelly are algebraically valid once a defensible p exists. They do not make an uncalibrated probability correct. XC-SPI does not provide staking recommendations.

12. Preferred Public Language

Appropriate: “XC-SPI ranks Team A first under the frozen Week 0 Preseason Score assumptions”; “Team A has the strongest expected scoring profile in the validated in-season simulation”; “the title probability is historically calibrated over the declared archive” only after that test exists.

Avoid: “Preseason Score predicts 167 NCAA points”; “three million simulations make this probability certain”; “the model proves Team A will win”; or quoting deprecated v2.5 scenario percentages as current calibrated XC-SPI probabilities.

References

[1] Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev.* 1950;78(1):1–3.
[2] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc.* 2007;102(477):359–378.
[3] Kelly JL Jr. A new interpretation of information rate. *Bell Syst Tech J.* 1956;35(4):917–926.
[4] 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.