

KPACE: A Course- and Conditions-Adjusted Measure of Cross Country Performance

A Public Overview

Peter Dunphy, Tommy Mulc, Gabe Cupino, & Derick Lawrence
Kailo, San Francisco, C.A.

September 30, 2026

Abstract

A coach preparing an athlete for a meet needs to know what that athlete should run on a given course, in given conditions. A finishing time conflates four quantities — fitness, the course, the weather, and the distance — and the tools available to coaches separate few of these. Pace calculators return equivalences under flat, ideal conditions; published venue ratings are high-school products that transfer poorly to collegiate fields; and the one independent collegiate system fuses a course’s durable character with the conditions of a single day into one per-race score. We present the Kailo Performance-Adjusted Course Equivalence (KPACE) framework, a projection assembled from separable components: a fitness base composing prior-season ability, a career-stage progression coefficient, the intervening track record, and the current season’s races as they arrive; a course multiplier estimated from every athlete who raced the venue and weighted toward its recent editions; a conditions term priced as the day’s departure from the course’s own typical weather; and a distance standardization. We evaluate KPACE by forward validation across the five most recent collegiate seasons (2021–2025; 207,774 NCAA Division I–III races), refitting every component on seasons strictly prior to the one predicted. KPACE reaches a pooled mean absolute error of 2.86% of race time against 3.95% for a last-season-mean baseline, with 48% of races inside 2% and a bias of +0.04%. It improves mean absolute error in every subgroup examined (gender, division, career season, and ability decile), and it transfers outside its headline population to NAIA (3.23%) and NJCAA (4.49%) fields. The course component agrees with two independently built systems at rank correlations of 0.92 and 0.74, and predicts races better than either on the venues they share.

1 Introduction

Hundreds of cross-country coaches spend hours preparing meet strategies. The core question this coach returns to is: given what this athlete has shown, what should they run here, on this course, in these conditions? These four components of success — the athlete’s fitness, the course, the weather, and the distance — are all intertwined in a raw finishing time. Yet existing infrastructure fails to adequately combine these components in a way that allows for accurate prediction for cross-country athletes.

This paper provides a solution through the Kailo Performance-Adjusted Course Equivalence (KPACE) model. The framework is based on a product of adjustments that are the core of

race-time variance,

$$\text{predicted time} = \text{fitness} \times \text{course} \times \text{weather} \times \text{distance}. \quad (1)$$

Figure 1 applies Equation 1 to one race from the 2026 season, using only what was known before the season began.

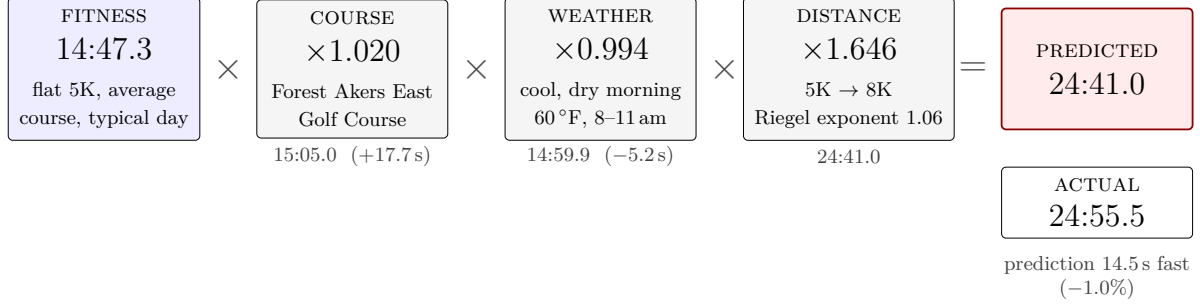


Figure 1: One projection worked through Equation 1: a Michigan State senior at the 2026 Spartan Invitational (8K, September 11), projected from pre-season information only — the athlete’s 2025 cross-country fitness, a career-stage progression, and spring 2026 track marks. The course rates 2.0% slower than a typical course; the cool, dry morning ran 0.57% faster than the course’s typical day. Running totals below each factor are 5K equivalents until distance is applied. The projection missed by 1.0%; across 14,022 returning athletes’ first 2026 races, the median pre-season miss was 2.7%.

While the model provides a structured way to separate these components, the ultimate goal is to empower coaches and athletes with actionable insights rather than just numerical predictions. A projection is therefore a starting point for a coach’s judgment about what a model may not capture, such as an athlete’s current form and role in a team race.

KPACE extends existing component tools. Fitness proxies such as VDOT¹ provide a useful basis for pace predictions in flat, ideal conditions. Kailo’s weather models provide adjustments for weather conditions during races. Existing course-adjustment systems are either focused on high-school competitions (MileSplit’s venue ratings, TullyRunners speed ratings, joshm.dev’s CIF-CCS pace factors)² or mix course effects with other factors (LACCTiC³). We compare KPACE against these existing tools. KPACE offers a lower bias and mean absolute error to these systems.

This overview describes the data the model is built from (§1.1), what each component of the model measures and how it behaves (§2; the flow from data to projection is drawn in Figure 3), the accuracy the assembled model achieves under forward validation (§3), and the limits it inherits from its inputs (§4). The estimators themselves, their fitted constants, and the feature definitions of the conditions model are currently internal to Kailo and are not reproduced here.

¹Jack Daniels and Jimmy Gilbert, *Oxygen Power: Performance Tables for Distance Runners*, published by the authors, 1979. The VDOT performance-equivalence system.

²MileSplit venue ratings, <https://www.milesplit.com/>; Bill Tully, TullyRunners Speed Ratings, <http://www.tullyrunners.com/>; joshm.dev, Crystal Springs Pace Factors for CIF-CCS Cross Country, <https://joshm.dev/>. Accessed 2026-08-25.

³Bijan Mazaheri, LACCTiC: Live Analysis of College Cross Country (TiC), <https://www.lacctic.com/>. Accessed 2026-08-26.

1.1 Data

Race results. Results come from the Track & Field Results Reporting System (TFRRS), operated by DirectAthletics. TFRRS holds 1.99 million collegiate cross-country results over the 2012–2025 seasons, of which 1.42 million are attributable to NCAA Division I–III programs; the remainder are NAIA, NJCAA, and other association fields. A further 3.3 million track running results (2020–2026 seasons, field events excluded) supply the track record that supplements fitness estimates (Table 1, Figure 2).

	Men	Women	Unmarked	Total	Unique Meets
All Seasons (2012–2025)	912,024	903,301	173,619	1,988,944	10,058
2017–2025 Seasons	635,120	611,681	165,747	1,412,548	7,268
2012–2016 Seasons	276,904	291,620	7,872	576,396	2,790

Table 1: The cross-country corpus by sex, all captured associations included. Per-season detail is in Figure 2; 2020 is genuinely small (COVID); the unmarked column is concentrated in association fields whose rosters carry no sex label. Counts as of 2026-09-14.

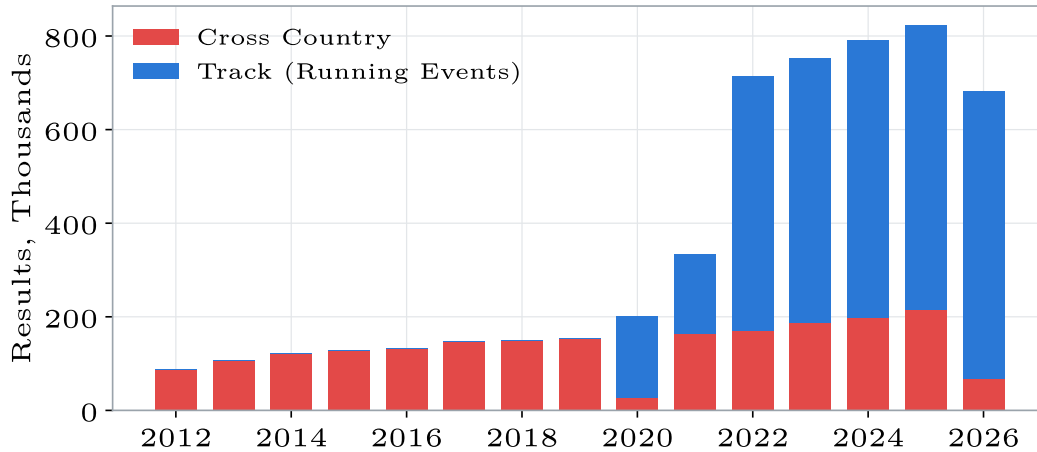


Figure 2: Running results per season by surface (field events excluded; track deduplicated). Cross-country coverage is effectively complete from 2017, substantial from 2012; track capture spans the 2020–2026 seasons.

Venues, dates, and course identity. Meet names and dates are parsed from TFRRS, which records a venue as free text. Each venue string is resolved to a course name, city, and state, and venue strings are then matched into canonical courses, so that a course’s editions pool across the several spellings under which it is recorded: over 2012–2025, 3,730 distinct venue strings resolve to 3,001 canonical courses. The resulting locations are geocoded, and those coordinates are what attach weather to a race.

Race start times. TFRRS publishes no race start (gun) times, which matters because the weather a field races in is the weather of a particular few hours. Start times are supplemented

from various auxiliary sources. Roughly half of recent championship-season meets carry a start time from these sources; coverage across all seasons is about a tenth of meets, since older seasons predate the timing archives. A meet with no start time for a given year but a recorded time in another year inherits that time; the remainder are assigned a default window of 9:00am–1:00pm local time.

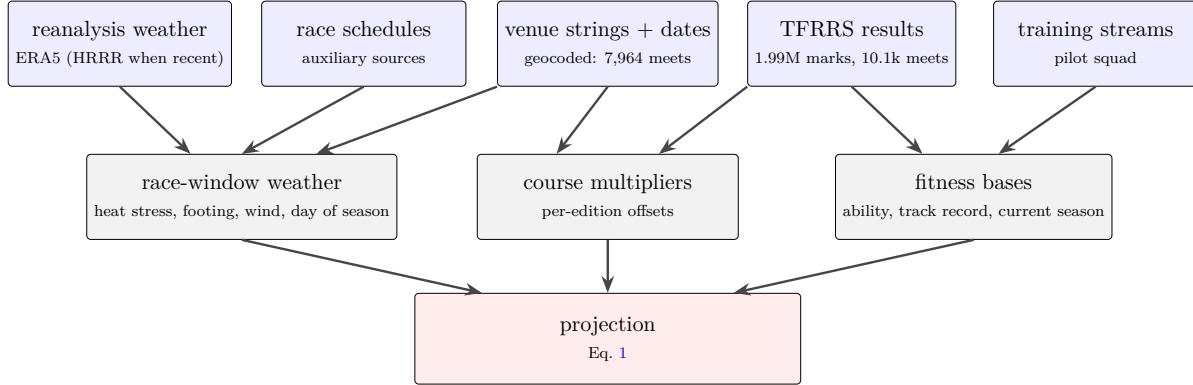


Figure 3: Data collection and how it feeds a projection. Weather attaches through the geocoded venue and the scheduled (or inherited) race window; the day-of-season term enters the conditions model from the meet date, while the race schedule sets the clock window within that day. Venue strings both locate the weather and name the courses the multipliers attach to. Weather is the ERA5 reanalysis, held and queried locally rather than through any third-party weather service; meets too recent for ERA5 to have been issued are enriched from NCEP HRRR analysis in the interim, and re-enriched from ERA5 once it arrives; the conditions model uses only variables both products carry, so a meet is priced identically from either.

2 How KPACE works

A KPACE projection is assembled from three estimated components and a distance standardization. This section states what each component measures, what it is estimated from, and how it behaves.

2.1 Separating fitness from course difficulty

The quantities the model requires are not directly observed. A finishing time is a single number, and the fitness it reflects is interpretable only after adjusting for the difficulty of the meets the athlete ran, while a meet’s difficulty is interpretable only after adjusting for the ability of the field that ran it. Estimating either in sequence therefore presupposes the other, and any procedure that settles one first — rating courses against a reference field, or rating athletes on raw times — propagates that choice into every quantity downstream.

KPACE estimates the two jointly, so that no course rating is conditioned on an assumed field and no fitness estimate is conditioned on an assumed course (Figure 4). Two of its properties govern how its output should be read and are stated here.

Overlap is the first, which makes decomposition identifiable. Athletes run in several courses within a season, and every course hosts teams that also race elsewhere. Each corpus, therefore,

forms a dense connected network of comparisons: each athlete who races two courses is a direct comparison.

The second is the scale. The system is anchored so that the average rated meet has no effect on the projections. A course rating of +3% is accordingly three percent slower than a typical course under its typical conditions — not just slower than an idealized flat course, nor slower relative to any single venue.

Two further separations follow from the fitted meet terms. We define a course’s persistent character as the average of its editions, with recent seasons weighted more heavily to account for courses being rerouted or remeasured; what an individual edition retains beyond that average is the signature of its own race day, and those residuals are the quantity the conditions model is fit against. The result is one consistent decomposition of every result into athlete, course, and day, estimated from the complete field of every race rather than from a hand-set table of adjustments.

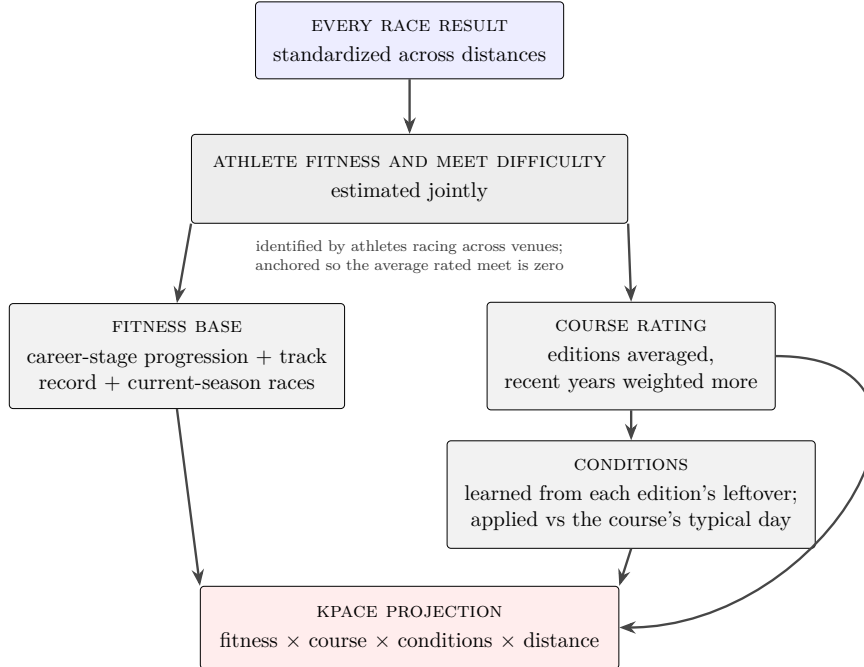


Figure 4: The estimation flow. Fitness and meet difficulty are estimated jointly rather than in sequence; the course rating and the conditions model then separate what is persistent about a venue from what a particular race day contributed.

2.2 Fitness

The fitness base is an estimate of what an athlete would run on a neutral course in neutral conditions.⁴ Each athlete’s prior race records (both track and cross-country), adjusted for the course it was run on and the distance it covered, contributes to a season-level ability estimate. Three further components enter. First, athletes improve with experience, and expected year-over-year improvement depends on how many collegiate seasons an athlete has completed

⁴Again, neutral course does not mean a flat course, nor do neutral conditions mean ideal conditions. Each is the average across the corpus.

(first-year athletes improve fastest). Pre-season estimates, therefore, carry a career-stage progression coefficient estimated from tens of thousands of athlete-season pairs. Second, the track season between cross-country campaigns is informative: an athlete’s middle- and long-distance track marks are composed into the pre-season estimate, with weights estimated separately for men and women. Third, once the current season begins, the estimate updates race by race, completed races drawing weight away from the pre-season estimate under a safeguard that prevents a single anomalous race from dominating. By mid-season the athlete’s current races carry most of the estimate.

The construction above presumes an athlete competed in the prior cross-country season. Two populations have none, and each is treated explicitly rather than omitted from the rankings.

2.3 First-year athletes

A first-year athlete has no prior collegiate season, so neither the progression coefficient nor the cross-country or track composite has anything to act on. Once such an athlete appears on a posted team roster and completes a collegiate race, their estimate is built from their collegiate races alone.

A single race taken at face value over-rates a first-year who raced well and under-rates one who raced badly, so with one to three races the estimate is regressed toward the first-year average of the athlete’s division: past first-years’ next race, fitted against their first, shows they keep about nine tenths of their distance from that average. The adjustment uses no information from the field of the race or from any other athlete. From the fourth race the athlete’s own races stand.

As of September 14, 2026, current KPACE rankings include 5,742 first-year athletes on this basis of which 4,643 have run solely one race. This method increases the reach of KPACE while sacrificing some precision. For first-season athletes, a projection built on one race predicts the next race to 4.2% of race time, on two races to 3.3%, and on three to 3.0%. A single race is therefore approximately as informative as the naive last-season-average baseline is for a returning athlete, and appreciably less precise than a returning athlete’s own pre-season estimate. A first-year athlete’s fitness estimate based on one race should be treated as provisional.

2.4 Returning after a missed season

An athlete returning from a missed season presents the converse problem: an estimate exists but has gone stale. Athletes who miss exactly one season return, at the median, 1.2% faster than their last season’s estimate, against 1.4% for a consecutive season. At the median, in other words, a returning athlete resumes on their development trend.

However, outcomes for returning athletes vary considerably more than for athletes who raced the previous season, and more again after two missed seasons. Their estimates carry correspondingly more uncertainty. As with first-year athletes, the rule applies only once the athlete has raced in the current season; the rankings make no claim about a return athlete who has not raced yet. The current rankings admit 651 athletes on this basis, 590 of them after a single missed season.

2.5 Course difficulty

Each course carries a difficulty rating: the percentage a typical athlete loses, or gains, in pace relative to a neutral course. Ratings are estimated from the complete field of every meet, so that thousands of athletes crossing between venues tie the courses to a common scale, and each course’s rating averages over its recent editions, with newer editions weighted more heavily because courses are rerouted and remeasured. A rating is less certain the fewer rated editions a course has, and a rating from a single edition is the least certain.

The ratings agree with the systems available for comparison. Against LACCTiC’s published collegiate scores the rank correlation is 0.92; against an independently constructed ridge-regression system it is 0.74 on shared venues. On the races both systems rate, KPACE predicts more accurately than either: a mean absolute error of 3.26% against 3.38% for LACCTiC’s scores after recalibration to our scale (289,410 races), and 3.29% against 3.51% for the ridge-regression system (27,243 races on ten shared courses). The ratings are also physically coherent: championship venues built for fast racing rate near neutral, and the courses with documented severe terrain, such as Van Cortlandt Park and Crystal Springs, occupy the slow extreme.

Altitude is isolated most cleanly indoors, where a track carries neither terrain nor weather (Figure 5). Indoor facility multipliers rise about 1.8% per additional 1,000 m ($r = 0.71$), from near zero at sea level to more than 5% at Gunnison and Alamosa. Cross-country courses at altitude follow the same direction less tightly, because terrain differs far more between courses than between tracks: Buffalo Park, a hilly mesa in Flagstaff, rates more than 7% slow, while the flat Cattails Golf Course in Alamosa rates under 2%, against the 5% of the indoor track in the same town.

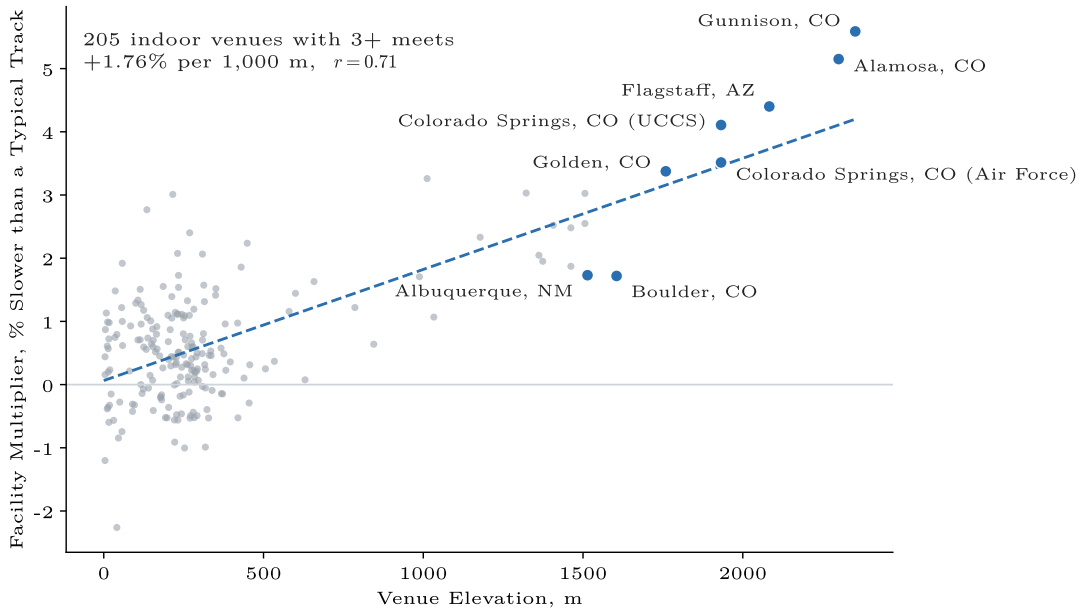


Figure 5: Indoor track facility multipliers against venue elevation, for the 205 indoor venues with at least three meets. With no terrain or weather, altitude is isolated: multipliers rise about 1.8% per additional 1,000 m ($r = 0.71$). Albuquerque’s banked track and Boulder’s oversized (300 m) track run below the line, consistent with banked and oversized tracks running faster than standard flat 200 m tracks.

2.6 Conditions

Weather moves race times. We used LASSO regression to determine which weather measurements within each meet’s race window were predictive of race outcomes, and the final model retains only a handful of them (for example heat stress, maximum air temperature, wind gusts, and rain in the preceding day). The candidates are restricted to variables carried by both weather products in use, so that a meet enriched from either is priced by the same model. Two results characterize the fit. Heat dominates, carried jointly by air temperature and humidity-adjusted measures such as wet-bulb temperature. Antecedent conditions matter greatly: rain in the days before a meet slows races more than rain falling during the race itself.

The adjustment is constructed as the day’s departure from the course’s own typical conditions, because a course’s difficulty rating already contains its usual climate. The rating of a Texas course whose meets run in September already reflects Texas heat, so only an unusually hot day moves the projection further. For a future race the term is evaluated on forecast weather.

3 Validation

All accuracy figures are forward validation: for each season from 2021 through 2025, every component is refit using only seasons strictly prior to it, and that season’s races are then predicted. Each of the 207,774 NCAA Division I–III races is held out exactly once. NAIA and NJCAA races, captured at shared meets, are evaluated separately as a test of transfer outside

the training population and show the same character (NAIA 3.23%, NJCAA 4.49%, each well below its naive baseline). Errors are reported in percent of race time; for scale, 1% of a 26:00 8K is approximately 16 seconds.

The full KPACE model produces an error of 2.86% of race time, against 3.95% for the last-season-mean baseline (a naive baseline). Nearly half of races (48%) fall within 2% of their projection, against a third for the baseline, and the mean signed error is +0.04% (projections are unbiased in aggregate; Table 2, Figure 6). Pre-season information alone reaches 3.38% (Figure 7), and the balance of the gain accrues as the season’s races arrive (Figure 9): error falls from 3.32% at an athlete’s season opener to approximately 2.1% by their fifth start, and from 3.03% in September to 2.30% in November. The improvement is not concentrated by subpopulation. It holds for both sexes, in every division, at every career stage, and in every ability decile (Figure 11), and it survives transfer to NAIA and NJCAA fields the model was never fit on. Against the public alternatives (Figures 7 and 8), a VDOT-style best-mark equivalence granted KPACE’s own course and conditions adjustments trails at 4.66%, and the course component agrees with two independently constructed systems at rank correlations of 0.92 and 0.74 while predicting races more accurately than either on the venues they share. The remainder of this section documents each of these results in turn.

model	bias	MAE	within 2%
Naive baseline (project last season’s mean)	+1.47	3.95	33%
KPACE, pre-season information only	+0.44	3.38	41%
KPACE, full model (updating in season)	+0.04	2.86	48%

Table 2: Pooled five-fold forward validation, $n = 207,774$ NCAA races. The pre-season row uses only information available before the season starts; the full model updates as the season’s races arrive.

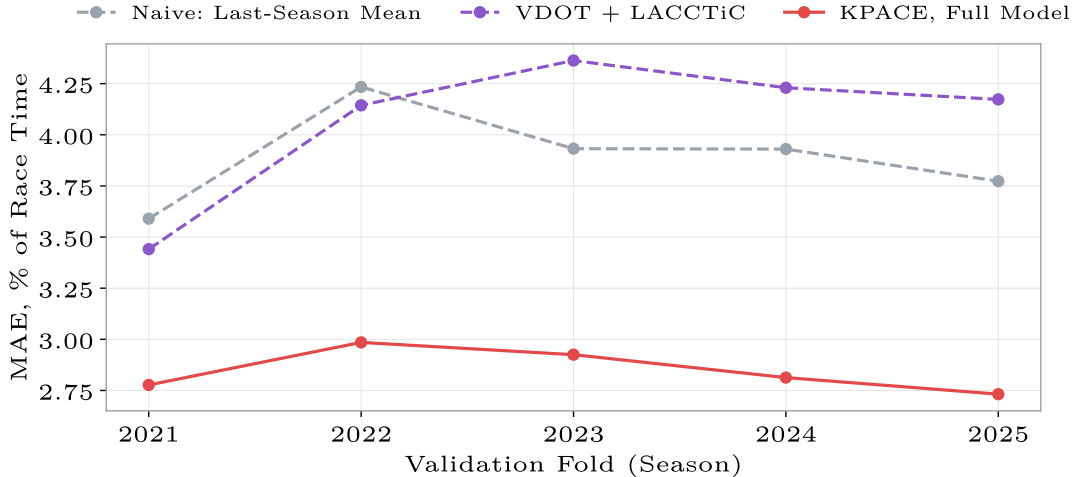


Figure 6: Accuracy by held-out season. The VDOT + LACCTiC comparator uses LACCTiC’s published race score, which postdates the race.

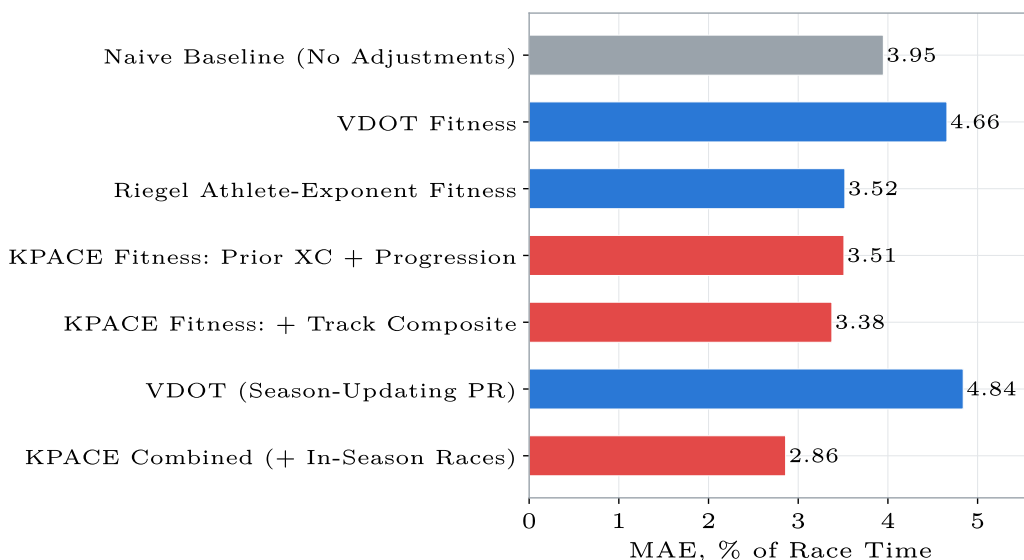


Figure 7: Fitness estimates compared, each given the identical KPACE course and conditions adjustments. A best-mark equivalence (VDOT) is systematically optimistic; KPACE’s season-based estimate improves further as the current season’s races arrive.

Figure 9 tracks the within-season gain on the most recent held-out season, by calendar month and by the number of races the athlete has already completed; the naive baseline is flat on both axes, since it has nothing to learn from the season in progress. Two forms of history contribute independently. Error declines with each additional race the athlete runs through the fifth. It rises again for athletes with six or more prior races, which reflects who those athletes are rather than a loss of accuracy: they are 1% of the season’s races and three quarters or more of them are Division III athletes, against two fifths at a season opener, and the naive baseline rises for the same group. Followed across their own season, the same athletes’ error falls from 3.5% at their opener to 2.35% and does not rise. The course adjustment is worth the most at venues with several rated past editions, being worth little at a venue rated from a single edition (Figure 10). The subgroup improvement is largest in absolute terms among developing athletes, for whom a last-season projection errs most (Figure 11).

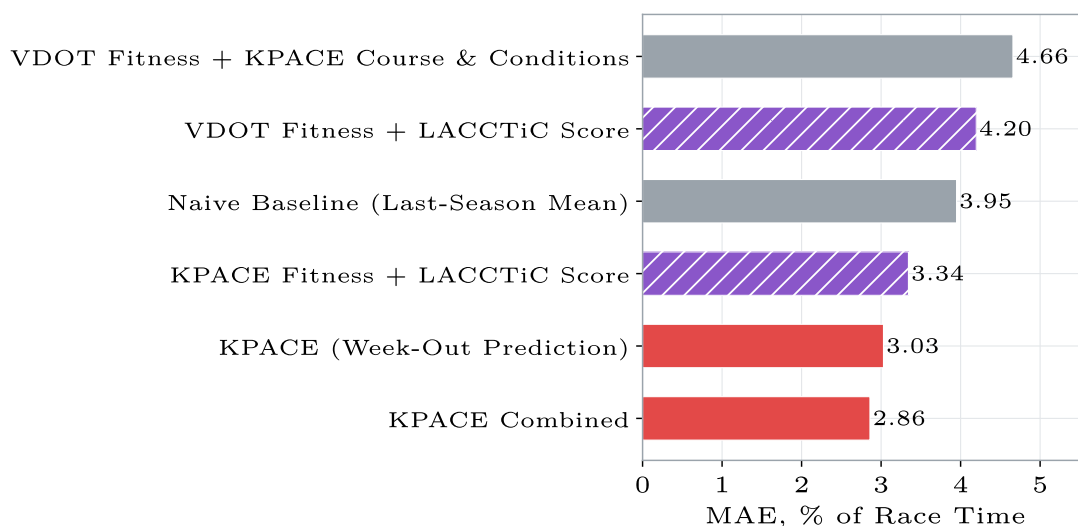


Figure 8: Assembled systems compared, pooled across the five held-out seasons. The week-out bar restricts KPACE to information available seven days before the meet. Hatched bars pair a fitness base with LACCTiC’s published race score as the day correction; that score postdates the race, so those bars are retrospective diagnostics, not forward errors.

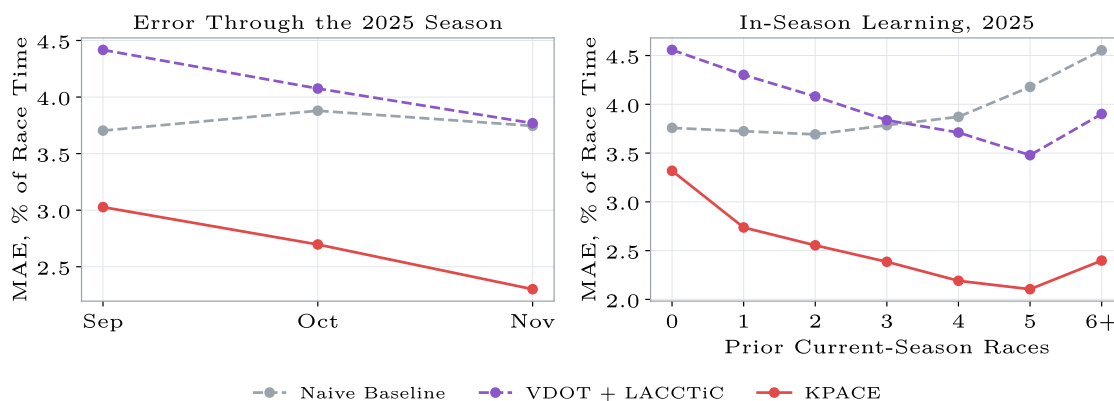


Figure 9: The 2025 season tracked forward. Left: monthly mean absolute error. Right: error by the number of the athlete’s completed races in the season. The naive baseline is flat; KPACE sharpens as evidence accumulates.

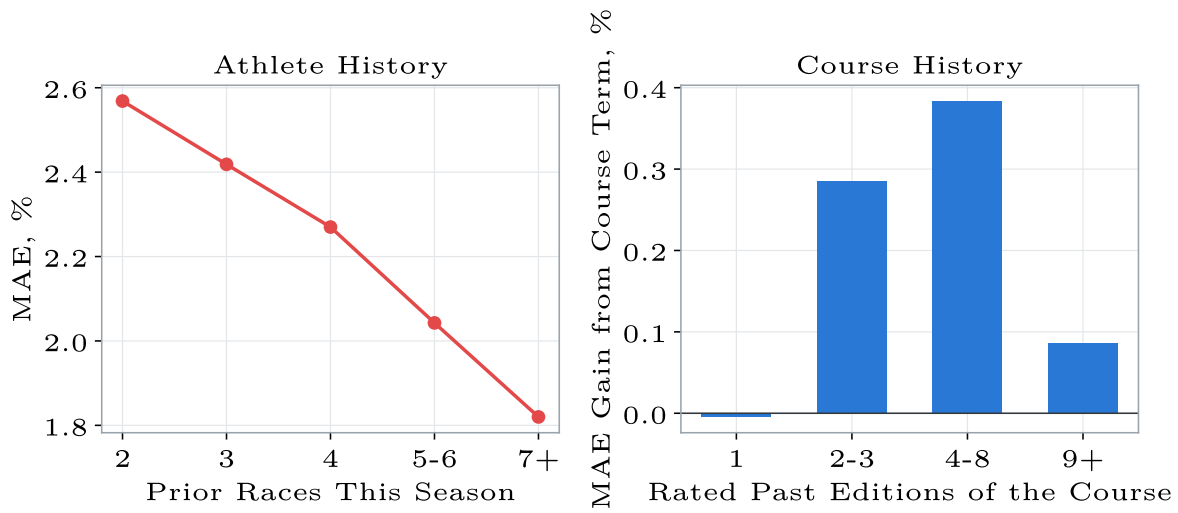


Figure 10: The two depth axes. Left: in-season error falls with each additional current-season race. Right: the course adjustment’s value by how many past editions inform it.

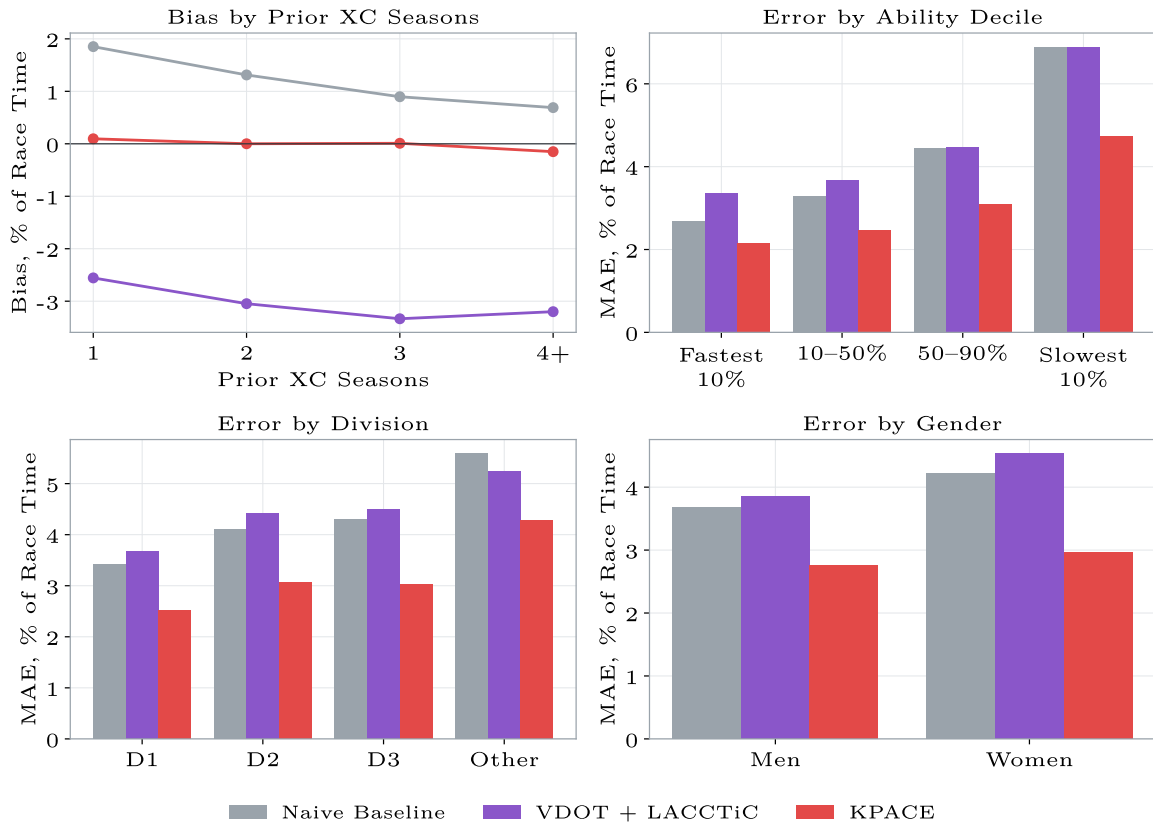


Figure 11: Performance by subgroup: bias by career season, and error by ability decile, division, and gender. KPACE improves accuracy in every group.

Against public alternatives. A VDOT-style personal-best equivalence, granted KPACE’s own course and conditions adjustments, returns a 4.66% mean absolute error with a large systematic bias (Figure 7): a best mark is not an expectation. Pairing a fitness estimate with LACCTiC’s published scores — a retrospective comparison, since the score postdates the race — still trails the KPACE day-of adjustment on identical races (Figure 8).

4 Limitations

KPACE is built from race results and inherits their limits. An athlete with no results — a first-year or returning athlete who has not yet raced this season, a transfer appearing under a new identity — cannot be projected until a race arrives, and the projections of those who have raced once are correspondingly uncertain (§2.3, §2.4). A first-year athlete’s rank still rests on very few races: the regression toward the first-year average (§2.3) tempers a strong first performance, and is in turn cautious about the first-years who really are among the best, whom it under-rates until further races arrive. In contrast, a fit athlete who ran a race as a tempo effort, or simply raced badly, may be weighed down by that single relatively poor performance. We are currently investigating ways to mitigate this, such as ranking on the best normalized performance of the season. The system is also occasionally delayed by manual intervention when a meet changes its name and is not correctly mapped back to its existing identifier. Meets whose results are never posted, and the great majority of meets carrying no published start time — which fall back to a default race window that the harvested schedules show mis-times about half of races — introduce noise the model cannot remove. Course ratings at single-edition venues are weakly determined, and a course rerouted between seasons retains its former rating until new editions accumulate; both cases carry more uncertainty than a rating built on several recent editions. The residual error — approximately three percent of race time at the start of a season and under two percent by its end — is dominated by genuine day-to-day athlete variation, which no results-based system observes in advance.