Part 4: Performance Monitoring Flag System: Threshold Justification

4.1 Overview

The performance monitoring flag system was designed to highlight meaningful changes in athlete performance, training load, and testing consistency. It combines individualized baselines, statistical benchmarks, and team-level norms to create a layered approach to athlete monitoring. Each threshold was chosen based on physiological reasoning, established best practices in performance analysis, and statistical reliability.

4.2 Data Cleaning and Validity Thresholds

Before applying any flags, the dataset was cleaned using linear interpolation and carry-forward/backfill techniques.. Implausible values were removed using gender-specific physiological cut-offs to ensure that only realistic data contributed to baseline calculations and risk identification.

Invalid Value Thresholds

Jump Height (Women): <0.30 m
Peak Propulsive Power (Women): <3200 W
Peak Velocity (Women): <2.4 m/s
Speed Max (Women): <5.0 m/s
Jump Height (Men): <0.40 m
Peak Propulsive Power (Men): <4500 W
Peak Velocity (Men): <2.8 m/s
Speed Max (Men): <6.0 m/s
Distance Total (Men & Women): <80% of baseline

Justification:

These thresholds represent the lower limits of what would be expected from trained collegiate athletes. Values below these ranges typically indicate measurement errors, sensor issues, or incomplete data capture, and underperformance. Because metrics like speed, power, velocity, and jump height change gradually over time, linear interpolation helps maintain natural performance trends, while forward/backfill methods ensure complete data at the start or end of a sequence.

4.3 Flag System Justification

4.3A Flag 1: Decline Relative to Individual Baseline

Baselines were established using the oldest valid data point for each metric within the past year.

Thresholds:

Speed Max: <90% of baseline
Jump Height: <90% of baseline
Peak Velocity: <95% of baseline
Peak Propulsive Power: <95% of baseline
Distance Total: >80% above baseline

Justification:
A drop of 5–10% in high-velocity or power-based metrics is generally considered meaningful in trained athletes.[1] These thresholds help identify signs of neuromuscular fatigue, under-recovery, or performance decline.[2] Distance Total uses an upper threshold because unusually high volume is more strongly associated with overload risk.[3]

4.3B Flag 2: Below Published Risk Thresholds (Statistical)
This flag uses statistical thresholds to identify athletes performing below their team's typical range.
Thresholds:
Standard Deviation Method: Value < (mean − 1 SD)
Percentile Method: Value < 10th percentile
Justification:
These thresholds help identify statistical outliers and low performers. Using both standard deviation and percentile methods increases sensitivity to different distribution patterns, allowing the system to capture both consistently low performers and sudden performance drops.

4.3C Flag 3: Testing Frequency (>30 Days Since Last Test)

Threshold:
More than 30 days since the last test.

Justification:
A 30-day threshold aligns with common monitoring cycles.[4] When athletes go untested for long periods, meaningful changes in performance may go unnoticed, reducing the reliability of trend analysis and baseline comparisons.

4.3D Flag 4: Deviation from Team Norms (Z Score + Player Mean Decline)
This flag combines team-level statistics with individual performance trends.

Thresholds:
Z Score: |z| > 2
Player Mean Decline: value/player_mean is < threshold
Speed Max & Jump Height: <90% of player mean
Peak Velocity & Power: <95% of player mean
Distance Total: <80% of player mean

Justification:
A z-score greater than 2 standard deviations indicates a substantial deviation from team norms.[5] The player-mean thresholds add an individualized layer, capturing meaningful declines even when an athlete still falls within the broader team range. Together, these measures create a **strong dual-indicator system** for identifying unusual or concerning performance patterns.

4.4 Integrated Rationale

The four flags each capture a different aspect of performance risk:

- Baseline Decline: Detects individual performance drops.
- Statistical Risk: Identifies underperformance relative to team norms.
- Testing Frequency: Ensures consistent and reliable data collection.
- Team Deviation: Highlights unusual values and meaningful individual declines.

This multi-layered structure improves sensitivity, reduces false positives, and provides a more complete picture of athlete performance and readiness.

Assumptions and Limitations

The code assumes the following:

- Valid data points accurately reflect performance
- Missing values can be interpolated without distorting trends, and
- Baselines are defined by the oldest valid measurement for each metric within the past year.
- Player means are treated as stable indicators of typical performance
- Thresholds are applied as meaningful markers of decline, and
- A 30 day testing window is considered sufficient for reliable monitoring.

The flagging system has the following notable limitations.

- The athlete dataset had limited data because testing and data collection is not robust and consistent, lacking relevant longitudinal data for study and predictive power, limiting the efficiency of the flagging system.
- Irregular or limited athlete testing can undermine the stability of baselines and player means
- Using uniform thresholds may fail to capture individual differences.
- Because of the heavy dependence on data quality, the approach is susceptible to measurement errors or GPS noise.
- Decline detection relies solely on percentage reductions, without statistical safeguards such as confidence intervals or practically significant change thresholds.
- Excludes important contextual factors like training load, fatigue, injury status, and competition demands, which restrict how deeply flagged changes can be interpreted.

References

1) Alba-Jiménez C, Moreno-Doutres D, Peña J. Trends Assessing Neuromuscular Fatigue in Team Sports: A Narrative Review. Sports. 2022; 10(3):33. https://doi.org/10.3390/sports10030033

2) Muñoz-Gracia JL, Alentorn-Geli E, Casals M, Hewett TE, Baiget E. Assessment Methods of Sport-Induced Neuromuscular Fatigue: A Scoping Review. Int J Sports Phys Ther. 2025;20(7):943-956. Published 2025 Jul 2. doi:10.26603/001c.141230

3) Coultman L. What neuromuscular fatigue really means for sprint performance. The Speed Project. Updated April 12, 2025. Accessed December 16, 2025. https://www.thespeedproject.com/sprinting/.

4) Gabbett TJ, Nassis GP, Oetter E, et al. The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data. Br J Sports Med. 2017;51(20):1451-1452. doi:10.1136/bjsports-2016-097298

5) Clubb J. Mastering Z-score and T-score statistics in sports science. Global Performance Insights. May 15, 2024. Updated April 4, 2025. Accessed December 16, 2025. https://www.globalperformanceinsights.com/blog/categories/data.