Predicting the Age-at-Peak of Elite Athletes with Functional Survival Models

1. Motivation and Research Questions

Athletic careers bend, peak, and fade. We treat "peak" as a time-to-event:

- Can functional predictors (e.g., season-long training load curves, intra-year performance trajectories, weekly GPS intensity profiles) improve prediction of age-at-peak compared with scalar-only models?
- How do functional Cox models compare to oblique random survival forests (ORSF) used as a baseline (as in PeaksNPrimes)?

2. Data & Outcome Definition

2.1 Outcome (time-to-peak).

- Event: first season in which an athlete achieves their lifetime best or a smoothed "career apex".
- Time scale: age; time origin = first appearance in sanctioned results; censoring for athletes who have not yet peaked when data end. This is right-censoring in a standard survival frame.

2.2 Predictors.

- **Scalar**: gender, nationality, event category, "Olympic year" indicator, training age—matching PeaksNPrimes for fair comparison.
- **Functional**: curves that vary within seasons/years: weekly workload; within-season performance trajectory; micro-cycle intensity; or "time-of-day / day-of-week" usage where appropriate (modeled as functions over a domain).

2.3 Data structure & preprocessing.

- Store functional predictors as matrices or lists of curves inside a single data frame column (wide format for fast access), following the survival+functional pattern.
- If needed, bin or average raw high-frequency data like converting minute/weekly signals to coarser grids to lighten computation before fitting.

3. Exploratory Analysis

- Describe cohort, censoring, median peak age via Kaplan–Meier. (PeaksNPrimes reported a median ~27 with tight uncertainty, compare with it).

- Visualize functional predictors (mean curves, variability ribbons), and scalar distributions by event and gender.
- Sanity-check proportional hazards assumptions for scalar-only Cox models before stepping up to functional modeling.

4. Methods

4.1 Baselines (scalar predictors).

- (i) KM stratified summaries to show trend.
- (ii) Standard Cox with scalar covariates; report HRs for key features (training age, event family, Olympic year).
- (iii) ORSF on the scalar set to mirror PeaksNPrimes; use variable importance to compare feature influence.

4.2 Functional Cox models.

- FLCM:

  $$\log \lambda_i(t \mid Z_i, W_i(\cdot)) = \log \lambda_0(t) + Z_i^\top \gamma + \int W_i(s)\, \beta(s)\, ds.$$

  Implement with penalized splines for $\beta(s)$ using mgcv/refund, as in chapter7.
- AFCM:

  $$\log \lambda_i(t) = \log \lambda_0(t) + Z_i^\top \gamma \; + \; \int F\{s, W_i(s)\}\, ds,$$

  where F is a smooth surface—use mgcv::gam tensor terms; this relaxes linearity and can capture interactions like “time-of-season × intensity”.
- Note:
  - Use refund::pfr/pffr or mgcv::gam/bam with survival family options; the chapter outlines estimation, inference, and prediction workflows.
  - For heavy designs, pre-aggregate (e.g., to hourly/weekly bins) to avoid memory blow-ups.

4.3 Assumptions, diagnostics, inference.

- Interpret $\beta(s)$ as a weight function: positive regions increase hazard (earlier peaks), negative regions decrease hazard (later peaks). Provide pointwise and multiplicity-adjusted inference on $\beta(s)$
- For AFCM, visualize the $s\times$intensity surface and interpret regions linked to earlier vs later peaks

5. Prediction & Evaluation

5.1 Per-athlete survival curves & peak-age distributions.

- Get $\widehat{S}_i(t)$; summarize predicted peak age as the time when the event probability crosses a threshold. Output the prediction curve for each athlete.

5.2 Comparative performance.

- Train/test splits by cohort years or nested CV.
- Metrics: time-dependent concordance (C-index), integrated Brier score, calibration of predicted peak-age quantiles, and ranking utility.
- Compare: Scalar Cox vs ORSF (scalar) vs FLCM vs AFCM. For ORSF.
- Show functional cox has predictive advantage and explanatory advantage.

6. Case Studies & Visuals

- Attach two athletes' functional curves with their predicted survival/peak-age curves to illustrate interpretability.
- maybe compare the predicted difficulty of different events. e.g., sprinters vs throwers—PeaksNPrimes found throwers' peaks harder to predict; test whether functional models close that gap.

7. Results

- Main: Does functional modelling provide better performance for age-at-peak than scalar baselines?
- Secondary: Which regions of the curve matter? E.g., pre-Olympic season load shape associated with earlier peaks; late-season volatility pushing peaks later.
- Variable influence: report $\beta(s)$ (FLCM) or $s\times$ intensity effects (AFCM) with uncertainty; contrast to ORSF variable importances.

8. Discussion

- **Interpretability:** Functional Cox models yield transparent mechanisms (where on the curve hazard shifts), which is often richer than black-box feature splits.
- **PeaksNPrimes comparison:** Training age remained influential in scalar forests; we test whether **functional shape** adds incremental value and whether it reduces difficulty for events like throws.
- **Physiology meets statistics:** Align curve regions with plausible training theory (base vs taper vs peaking micro-cycles). Frame as **working hypotheses**, not causal claims.

9. Limitations

- peak definition
- noise
- ...

10. Expected Contribution

A systematic survival-FDA blueprint for age-at-peak that is interpretable at the curve level, competitive to scalar-only forests, and easily extended to multilevel functional structures in future work.