Toward Hierarchical Tensor Representations of Endurance Training Stimulus: Integrating External Load, Physiological State, and Temporal Structure

Part I: Pure-Load Tensor (PLT)

Gabriel Della Mattia
AGMT2 Team

June 27, 2024

Abstract

We present two complementary tensorizations of endurance training data. (i) A *Pure-Load Tensor* (PLT) constructed only from training-derived scalars: TSS, energy per body mass (kJ/kg), within-activity mean heart rate computed over quartiles of session duration, pre- and post-session HRV, and an exponential return-to-baseline rate (λ_{HRV}). The PLT includes an order-aware inheritance operator to propagate the consequences of the first stimulus to a second same-day stimulus. (ii) A *Full Multimodal Tensor* (FMT) that preserves intra-session sequences, day-level context, and multi-day dependencies as in the main hierarchy. Within both tensorizations, TSS becomes one feature among many. We derive operators for anaerobic work balance, sequence-aware aggregation, and multi-task learning for short-term recovery events and medium-term performance changes. We provide theoretical arguments explaining why scalar reductions cannot identify key nonlinear responses and present a practical model consisting of a session encoder, an attention-based day aggregator, and a temporal Transformer.

1 Introduction

Scalar training-load metrics such as TSS or TRIMP Banister1975,Allen2019 remain popular due to simplicity and approximate comparability across sessions. However, endurance adaptation depends on the *distribution* of intensities, *ordering* of stimuli and recoveries, and *physiological state* (sleep, HRV, stress). Two sessions with identical TSS may differ substantially in their anaerobic work expenditure, oxygen kinetics, and neuromuscular stress. This paper formalizes two routes to sequence-aware modeling: a compact Pure-Load Tensor (PLT) when only minimal metrics exist, and a Full Multimodal Tensor (FMT) when rich intra-session waveforms and context are available.

Contributions. (i) A formal tensorization across three levels (session, day, window) in two flavors (PLT and FMT); (ii) operators modeling physiologically relevant dynamics. (iii) a sequence model yielding short-term risk and medium-term performance deltas via multi-task losses; (iv) theoretical results highlighting the non-identifiability of scalar reductions and the benefits of sequence-aware representations.

2 Notation and tensor representations

We denote by C the number of within-session channels (e.g., power, heart rate, cadence, grade, altitude, skin temperature), by T_s the length (in samples) of session s, and by S_d the number of sessions on day d.

2.1 Intra-session tensor

A session is a variable-length time series

$$\boldsymbol{X}_{d,s} \in \mathbb{R}^{T_s \times C}, \qquad \boldsymbol{M}_{d,s} \in \{0,1\}^{T_s \times C}, \tag{1}$$

where $\boldsymbol{M}_{d,s}$ is a binary mask (1 if observed). Typical channels include instantaneous power P, heart rate H, cadence K, grade G, altitude A, and derived variables (e.g., W′ balance, zone indicators). For batching, we pad to $T_{\max}$ and retain $\boldsymbol{M}_{d,s}$. Downsampling to $\Delta t \in [1,5]\,\mathrm{s}$ is allowed. Let $\mathcal{N}$ be a per-athlete normalizer (robust z-score or percentile scaling) applied channel-wise.

2.2 Day-level representation

Each session has a summary vector $\boldsymbol{v}_{d,s} \in \mathbb{R}^{C_{\text{sess}}}$ collecting non-sequential descriptors (e.g., duration, NP, IF, variance index, percent time in zones, interval counts, W′ spent/recovered, decoupling, TSS). Day-level physiological context is $\boldsymbol{z}_d \in \mathbb{R}^{C_{\text{day}}}$ (... morning HRV as rMSSD TaskForce1996, resting HR, sleep duration and architecture, subjective stress, body mass, symptoms). Stack session summaries with padding to $S_{\max}$ to obtain $\boldsymbol{D}_d \in \mathbb{R}^{S_{\max} \times C_{\text{sess}}}$.

2.3 Window tensor (multi-day sequence)

For a window length L (e.g., 28 days), we define the sequence

$$\boldsymbol{W}_t = \left[(\boldsymbol{D}_{t-L+1},\, \boldsymbol{z}_{t-L+1}), \ldots, (\boldsymbol{D}_t,\, \boldsymbol{z}_t)\right]. \tag{2}$$

For multi-sport settings with K sports (bike/run/swim), we can allocate a sport axis, yielding $\boldsymbol{W}_t \in \mathbb{R}^{L \times K \times C_{\text{day}}}$ by segregating day features per sport. TSS appears as one coordinate within $\boldsymbol{v}_{d,s}$ and possibly aggregated in $\boldsymbol{z}_d$.

3 Complementary tensorizations: Pure-Load vs Full Multimodal

3.1 Pure-Load Tensor (PLT)

Let body mass be m and session duration be T_s. Define total mechanical work $W = \sum_{n=1}^{N} P_n \, \Delta t$ and specific energy $E_{\text{kg}} = \frac{W/1000}{m}$ in kJ/kg. Partition the session into four equal-duration quartiles $Q_j = [t_{j-1}, t_j]$, $j \in \{1, 2, 3, 4\}$ with $t_j = jT_s/4$. Define within-quartile mean heart rate

$$\bar{H}_j^{(q)} \;=\; \frac{4}{T_s} \int_{Q_j} H(t)\, dt \quad (\text{discrete: } \bar{H}_j^{(q)} = \tfrac{1}{|Q_j|} \sum_{t \in Q_j} H_t). \tag{3}$$

Let $H_0 = \text{HRV}^{\text{pre}}$ be pre-session rMSSD (baseline over a short window before the session), $H_1 = \text{HRV}^{\text{post}}$ be the first post-session measurement, and suppose a later measurement H_2 is available after Δt_{12}. The exponential return-to-baseline rate is estimated by

$$\lambda_{\text{HRV}} \;=\; \frac{1}{\Delta t_{12}} \ln\left(\frac{|H_1 - H_0|}{|H_2 - H_0|}\right), \tag{4}$$

with a single-sample surrogate $v_{\text{HRV}} = \frac{H_1 - H_0}{\Delta t_{01}}$ when only H_1 exists. We define the PLT session vector

$$\boldsymbol{v}_{d,s}^{\text{PLT}} \;=\; \left[\, \text{TSS}_{d,s},\; E_{\text{kg},d,s},\; \bar{H}_{1:4}^{(q)},\; \text{HRV}_d^{\text{pre}},\; \text{HRV}_{d,s}^{\text{post}},\; \lambda_{\text{HRV},d,s} \,\right] \in \mathbb{R}^{C_{\text{PLT}}}, \quad C_{\text{PLT}} = 9. \tag{5}$$

Stacking yields $\boldsymbol{D}_d^{\text{PLT}} \in \mathbb{R}^{S_{\max} \times C_{\text{PLT}}}$. No intra-session waveform is needed; the encoder for PLT can be a small MLP.

Inheritance for two stimuli in a day. Let the same day have two sessions $s = 1, 2$ separated by a gap τ_{gap}. The predicted pre-session HRV of the second session, inheriting the effect of the first, is

$$\widehat{\text{HRV}}_{d,2}^{\text{pre}} \;=\; \text{HRV}_d^{\text{pre}} + \left(\text{HRV}_{d,1}^{\text{post}} - \text{HRV}_d^{\text{pre}}\right) e^{-\lambda_{\text{HRV},d,1}\, \tau_{\text{gap}}}, \tag{6}$$

which is used in place of (or alongside) a measured value when unavailable. More generally, define a carry-over state $\boldsymbol{s}_{i+1} = \mathcal{H}(\boldsymbol{s}_i, \boldsymbol{v}_{d,i}^{\text{PLT}})$ with

$$\boldsymbol{s}_{i+1} \;=\; \sigma\!\left(\boldsymbol{U}\, \boldsymbol{s}_i + \boldsymbol{G}\, \boldsymbol{v}_{d,i}^{\text{PLT}}\right), \qquad \sigma = \tanh, \tag{7}$$

so the effective representation of the second stimulus becomes $\tilde{\boldsymbol{v}}_{d,2}^{\text{PLT}} = \Pi(\boldsymbol{v}_{d,2}^{\text{PLT}} \| \boldsymbol{s}_2)$, making aggregation explicitly order-sensitive.

3.2 Full Multimodal Tensor (FMT)

The FMT follows the hierarchy in Sections 2.1–2.3. It preserves intra-session sequences $\boldsymbol{X}_{d,s} \in \mathbb{R}^{T_s \times C}$ with masks, session summaries $\boldsymbol{v}_{d,s}$, and day context $\boldsymbol{z}_d$, assembled into a window tensor $\boldsymbol{W}_t$. TSS is a coordinate within $\boldsymbol{v}_{d,s}/\boldsymbol{z}_d$, not the decision variable.

4 Physiological operators as tensor channels

4.1 Anaerobic work balance (W')

(see Monod1965,Vanhatalo2011,Jones2017,Skiba2012). Let instantaneous power be $P(t)$ and critical power be CP. Define the positive-part operator $(x)_+ = \max(x, 0)$. A general $\mathrm{W}' dynamicscanbewrittenas \frac{d\,\mathrm{W}'(t)}{dt}$ $-\,(P(t) - \mathrm{CP})_+ + r\big(P(t), \mathrm{W}'(t);\, \theta_r\big), (8)$ where $r(\cdot)$ models sub-CP recovery (e.g., mono-exponential with time constant depending on intensity or an empirically learned function). The state $\mathrm{W}'(t)$ enters as a channel in $\boldsymbol{X}_{d,s}$, allowing the model to distinguish sessions with identical TSS but different supra-CP demands.

4.2 Cardio-metabolic stress

Define a convex surrogate of metabolic strain using a power nonlinearity on normalized power (e.g., NP Allen2019). $\tilde{P}(t)$:

$$\mathcal{S}(\boldsymbol{X}_{d,s}) \;=\; \int \tilde{P}(t)^{\,p}\, dt \quad (p > 1), \tag{9}$$

with $p \approx 3$–4 related to moving-average constructs (e.g., NP). By Jensen's inequality, aggregating before exponentiation underestimates peaks: $\left(\frac{1}{T}\int \tilde{P}\, dt\right)^{p} \le \frac{1}{T}\int \tilde{P}^{\,p}\, dt$.

4.3 Environmental and contextual modulation

Let $E(t)$ collect environmental covariates (temperature, humidity, altitude). Introduce a modulation $\gamma(E(t))$ multiplying channel intensities or recovery $r(\cdot)$ in (8). Similarly, morning state $\boldsymbol{z}_d$ shifts day-level priors via conditioning in the sequence model.

5 Sequence model on hierarchical tensors

Input flexibility. The session encoder accepts either (i) intra-session sequences $\boldsymbol{X}_{d,s}$ (FMT) or (ii) PLT vectors $\boldsymbol{v}_{d,s}^{\mathrm{PLT}}$ (mapped through an MLP). The same day aggregator and temporal model apply to both, with the carry-over state in Eq. (7) activated when sequential same-day stimuli are present.

5.1 Session encoder

Given $\boldsymbol{X}_{d,s} \in \mathbb{R}^{T_s \times C}$ and mask $\boldsymbol{M}_{d,s}$, define an encoder f_θ producing an embedding $\boldsymbol{e}_{d,s} \in \mathbb{R}^d$; for PLT, use a projection Π_{PLT} of $\boldsymbol{v}_{d,s}^{\mathrm{PLT}}$ instead:

$$\boldsymbol{e}_{d,s} = \begin{cases} f_\theta(\mathcal{N}(\boldsymbol{X}_{d,s}),\, \boldsymbol{M}_{d,s}), & \text{FMT}, \\ \Pi_{\mathrm{PLT}}\, \boldsymbol{v}_{d,s}^{\mathrm{PLT}}, & \text{PLT}. \end{cases} \tag{10}$$

5.2 Attention-based day aggregation

We aggregate a set of session embeddings $\{\boldsymbol{e}_{d,s}\}_{s=1}^{S_d}$ with content-based attention and integrate physiological state $\boldsymbol{z}_d$ (if available):

$$\alpha_{d,s} = \frac{\exp\left(\boldsymbol{q}^\top \tanh(\boldsymbol{W}\,\boldsymbol{e}_{d,s})\right)}{\sum_{j=1}^{S_d} \exp\left(\boldsymbol{q}^\top \tanh(\boldsymbol{W}\,\boldsymbol{e}_{d,j})\right)}, \tag{11}$$

$$\boldsymbol{h}_d = \phi\Big(\sum_{s=1}^{S_d} \alpha_{d,s}\,\boldsymbol{e}_{d,s} \parallel \Pi_{\text{day}}\,\boldsymbol{z}_d\Big) \in \mathbb{R}^d. \tag{12}$$

5.3 Temporal Transformer

Consider the window $\{\boldsymbol{h}_{t-L+1}, \ldots, \boldsymbol{h}_t\}$. A Transformer encoder $\mathrm{T}_\psi with positional encodings models long-range dependencies~Vaswani2017:$

$$\boldsymbol{H}_t = \mathcal{T}_\psi(\boldsymbol{h}_{t-L+1:t}) \in \mathbb{R}^{L\times d}, \qquad \boldsymbol{g}_t = \boldsymbol{H}_t[L,\, :]. \tag{13}$$

Outputs are produced by task-specific heads: a classifier for short-term recovery risk and a regressor for medium-term performance deltas.

5.4 Targets and multi-task loss

Let $y_{t+1}^{(\text{rec})} \in \{0,1\}$ indicate a recovery event (e.g., next-day HRV drop $> \delta$), and $y_{t+\tau}^{(\Delta\text{CP})} \in \mathbb{R}$ the change in critical power at horizon τ (e.g., 28 days). Define

$$\hat{p}_{t+1} = \sigma(\boldsymbol{w}_{\text{rec}}^\top \boldsymbol{g}_t + b_{\text{rec}}), \tag{14}$$

$$\widehat{\Delta\text{CP}}_{t+\tau} = \boldsymbol{w}_\Delta^\top \boldsymbol{g}_t + b_\Delta. \tag{15}$$

The joint loss is

$$\mathcal{L} = \lambda_{\text{rec}}\,\text{BCE}(y_{t+1}^{(\text{rec})}, \hat{p}_{t+1}) \;+\; \lambda_\Delta\,\text{Huber}_\delta(y_{t+\tau}^{(\Delta\text{CP})}, \widehat{\Delta\text{CP}}_{t+\tau}). \tag{16}$$

6 Why scalars are insufficient: theory

6.1 Scalar reductions lose convex, order-sensitive information

Let $u : \mathbb{R}_+ \to \mathbb{R}_+$ be convex (e.g., $u(x) = x^p$ with $p > 1$). Consider a scalar reduction $\text{TSS}(P) = \int u(P(t))\,dt$. Then by Jensen,

$$u\Big(\frac{1}{T}\int P(t)\,dt\Big) \;\le\; \frac{1}{T}\int u(P(t))\,dt, \tag{17}$$

so compressing P before applying u underestimates the contribution of peaks. Further, any scalar map $\mathcal{S} : \mathbb{R}^T \to \mathbb{R}$ is many-to-one; there exist $P_1 \neq P_2$ with $\mathcal{S}(P_1) = \mathcal{S}(P_2)$ but different supra-CP content $\int (P_i - \mathrm{CP})_+ \, dt$.

6.2 Non-commutativity with W' dynamics

Let Φ be the flow induced by (8). In general $\mathrm{TSS}(P_1) = \mathrm{TSS}(P_2)$ does not imply $\Phi(P_1) = \Phi(P_2)$. Because Φ is nonlinear and state-dependent, averaging P or equating any integral functional does not preserve the trajectory of $\mathrm{W}'(t)$.

Proposition 1 (Non-identifiability of TSS)**.** *There exist power profiles P_A, P_B such that $\mathrm{TSS}(P_A) = \mathrm{TSS}(P_B)$ while for the same CP and initial W'_0, the terminal anaerobic reserve satisfies $\mathrm{W}'_A(T) \neq \mathrm{W}'_B(T)$ and the time-above-CP satisfies $\int (P_A - \mathrm{CP})_+ \, dt \neq \int (P_B - \mathrm{CP})_+ \, dt$.*

Sketch. Construct P_A as steady Z2 with short spikes to match $\int u(P)\, dt$ of an interval session P_B with long supra-CP bouts and long recoveries. Equality of $\int u(P)\, dt$ holds by design; however, the piecewise-constant supra-CP segments in P_B generate distinct W' depletion trajectories due to the $(\cdot)_+$ term and state-dependent recovery $r(\cdot)$, which do not commute with time-averaging. □

6.3 Fading-memory functionals and sequence models

Let $\mathcal{F}$ be the class of causal, time-invariant, fading-memory functionals mapping windows of inputs to outputs. Under mild conditions, sequence models with sufficient capacity (e.g., CNN/TCN/Transformers with bounded kernels or attention) approximate $\mathcal{F}$ uniformly on compacts. The tensor representation supplies the requisite inputs (multi-channel, context, masks); the scalar TSS does not.

7 Preprocessing, normalization, and missingness

Per-athlete scaling reduces inter-individual variance: for channel c, $\tilde{x}_c = \frac{x_c - \mathrm{median}_c}{\mathrm{IQR}_c}$ or percentile scaling. Heavy-tailed variables (e.g., $\mathrm{W}' events) benefit from \log(1+x)$. Missing values are represented by explicit masks appended as channels. Morning state variables (HRV, sleep) are aligned to day d, while performance targets are aligned to $d + \tau$.

8 Evaluation protocol

We recommend leave-one-athlete-out or grouped cross-validation to assess generalization. Metrics include AUROC/PR for recovery events, MAE/R^2 for performance deltas, and calibration curves. Utility is reported as decision curves for prescriptive thresholds (e.g., probability of next-day HRV drop).

9 Practical implications

Planning shifts from hitting a scalar load to prescribing *structures* whose expected benefit and risk are quantified for the specific athlete and context. Attention weights in (12) provide interpretability, highlighting which session structures and state variables drove predictions.

10 Limitations and extensions

Data quality (device changes, artifacts) must be handled via robust preprocessing and masking. Sparse labels for medium-term outcomes can be mitigated with self-supervised pretraining (masked time-series prediction) at the session level and multi-task learning. Extensions include multi-modal sensing (core temperature, HRV during exercise), Bayesian personalization layers, and reinforcement learning policies that optimize sequences of sessions under risk constraints.

11 Conclusion

The PLT and FMT formulations elevate training data from a scalar summary to physiologically faithful states. In both, TSS remains a useful coordinate but loses primacy. Sequence-aware models operating on these tensors predict recovery and adaptation more accurately and enable individualized, explainable prescription.

12 Use case: Mapping PLT/FMT to an impulse–response (Banister) model

Goal. Produce a daily scalar "training impulse" u_d from either the PLT or FMT representation and feed it to a fitness–fatigue impulse–response (IR) model to obtain a latent performance trajectory $\hat{P}_d$ suitable for visualization and decision-making.

12.1 Daily impulse from PLT (minimal-data route)

For day d with sessions $s = 1, \ldots, S_d$, compute PLT vectors $\boldsymbol{v}_{d,s}^{\mathrm{PLT}} \in \mathbb{R}^9$ (Sec. 3.1), apply the same-day carry-over for multiple stimuli (Eq. (6)) to obtain order-aware $\tilde{\boldsymbol{v}}_{d,s}^{\mathrm{PLT}}$, and aggregate to a day vector

$$\boldsymbol{v}_d^{\mathrm{PLT}} \;=\; \sum_{s=1}^{S_d} \omega_{d,s}\, \tilde{\boldsymbol{v}}_{d,s}^{\mathrm{PLT}}, \qquad \omega_{d,s} \ge 0,\ \sum_s \omega_{d,s} = 1. \tag{18}$$

Apply per-athlete robust scaling to get $\tilde{\boldsymbol{v}}_d^{\mathrm{PLT}}$. Define the daily impulse as a nonnegative projection

$$u_d \;=\; \mathrm{softplus}\big(\boldsymbol{\beta}^\top \tilde{\boldsymbol{v}}_d^{\mathrm{PLT}}\big) \;=\; \log\big(1 + e^{\boldsymbol{\beta}^\top \tilde{\boldsymbol{v}}_d^{\mathrm{PLT}}}\big), \tag{19}$$

where $\boldsymbol{\beta} \in \mathbb{R}^9$ are athlete-specific or population-shared weights. *Fallback:* set $u_d = \mathrm{TSS}_d$ for backward compatibility.

12.2 Daily impulse from FMT (rich-data route)

Let $\boldsymbol{h}_d$ be the day embedding from Eq. (12). Map it to a scalar impulse

$$u_d \;=\; \mathrm{softplus}(\boldsymbol{\gamma}^\top \boldsymbol{h}_d), \qquad \boldsymbol{\gamma} \in \mathbb{R}^d. \tag{20}$$

This choice lets the IR layer sit on top of the same encoder used for other tasks.

12.3 Fitness–fatigue IR dynamics and output to plot

Use either a single-impulse or dual-impulse formulation. The dual-impulse variant often fits better:

$$u_d^{(f)} = \mathrm{softplus}(\boldsymbol{\beta}_f^\top \tilde{\boldsymbol{v}}_d^{\mathrm{PLT}}) \quad \text{or} \quad \mathrm{softplus}(\boldsymbol{\gamma}_f^\top \boldsymbol{h}_d), \tag{21}$$

$$u_d^{(g)} = \mathrm{softplus}(\boldsymbol{\beta}_g^\top \tilde{\boldsymbol{v}}_d^{\mathrm{PLT}}) \quad \text{or} \quad \mathrm{softplus}(\boldsymbol{\gamma}_g^\top \boldsymbol{h}_d), \tag{22}$$

and the recursions (daily indexing, $d = 1, 2, \ldots$):

$$F_{d+1} = \rho_f\, F_d \;+\; u_d^{(f)}, \qquad \rho_f = e^{-1/\tau_f},\ \tau_f > 0, \tag{23}$$

$$G_{d+1} = \rho_g\, G_d \;+\; u_d^{(g)}, \qquad \rho_g = e^{-1/\tau_g},\ \tau_g > 0, \tag{24}$$

$$\hat{P}_d = P_0 \;+\; k_f\, F_d \;-\; k_g\, G_d. \tag{25}$$

Here F (fitness) is slow-decaying and G (fatigue) is fast-decaying; $k_f, k_g > 0$ scale their effects; P_0 is a baseline. **Value to plot:** $\hat{P}_d$ versus d (line), with u_d as bars and optionally F_d, G_d as auxiliary lines.

Single-impulse simplification. Set $u_d^{(f)} = u_d^{(g)} = u_d$ from Eq. (19) or (20).

12.4 Parameter estimation

Given observed targets $\{y_d\}$ (e.g., CP/FTP every few weeks or periodic TT power), estimate

$$\Theta = \{P_0, k_f, k_g, \tau_f, \tau_g, \boldsymbol{\beta}\ (\text{or}\ \boldsymbol{\gamma})\}$$

by minimizing $\sum_d w_d\,(y_d - \hat{P}_d)^2$ with constraints $k_f, k_g, \tau_f, \tau_g > 0$ and optional monotonicity priors $\tau_f > \tau_g$. For binary recovery events, replace the last line of Eq. (25) by $\mathrm{logit}\, p_{d+1} = \theta_0 + \theta_f F_d - \theta_g G_d$ and use Bernoulli likelihood.

12.5 Minimal recipe for practitioners

1. Compute $\boldsymbol{v}_{d,s}^{\mathrm{PLT}}$ per session; apply same-day inheritance (Eq. (6)); aggregate to $\boldsymbol{v}_d^{\mathrm{PLT}}$ and scale per athlete.
2. Build u_d with Eq. (19) (or use TSS_d initially).

3. Choose τ_f, τ_g initial guesses (e.g., $\tau_f \in [20, 60]$ days, $\tau_g \in [3, 15]$ days); fit Θ on your historical (u_d, y_d).

4. Plot u_d (bars), F_d and G_d (lines), and $\hat{P}_d$ (main line).

Appendix A: Practical data schema for PLT

Daily context vector ($\boldsymbol{z}_d$). From the daily table, build $\boldsymbol{z}_d$ by stacking robustly scaled (per-athlete) versions of: `HRV (rMSSD)` [ms], `minHeartRate` [bpm], `sleepRate` [a.u.], `sleepDuracion` [h], `deepSleepHoras` [h], `lightSleepHoras` [h], `remSleepHoras` [h], `awakeSleepHoras` [h], plus optional `bodyMass` [kg], `stressSubjectivo` [a.u.], `symptoms` [binary].

Session summary for PLT. For each session s on day d, compute

$$\boldsymbol{v}_{d,s}^{\mathrm{PLT}} = [\mathrm{TSS}_{d,s},\ E_{\mathrm{kg},d,s},\ \bar{H}_{1:4,d,s}^{(q)},\ \mathrm{HRV}_d^{\mathrm{pre}},\ \mathrm{HRV}_{d,s}^{\mathrm{post}},\ \lambda_{\mathrm{HRV},d,s}] \in \mathbb{R}^9.$$

Stack with padding to obtain the daily matrix $\boldsymbol{D}_d^{\mathrm{PLT}} \in \mathbb{R}^{S_{\max} \times 9}$.

Field list (recommended).

Column (source)	Symbol	Units
HRV (rMSSD)	$\mathrm{HRV}_d^{\mathrm{pre}}$	ms
minHeartRate	resting HR	bpm
sleepRate	sleep quality score	a.u.
sleepDuracion, deep/light/rem/awake	sleep architecture	h
bodyMass	m	kg
Session duration	T_s	s
Specific energy	$E_{\mathrm{kg},d,s}$	kJ/kg
TSS	$\mathrm{TSS}_{d,s}$	a.u.
Quartile HR	$\bar{H}_{1:4,d,s}^{(q)}$	bpm
Post-session HRV	$\mathrm{HRV}_{d,s}^{\mathrm{post}}$	ms
Return rate	$\lambda_{\mathrm{HRV},d,s}$	h^{-1}

Preprocessing. Per-athlete robust scaling (median/IQR or percentiles); explicit missingness masks; time alignment: morning HRV and sleep $\rightarrow d$, performance targets $\rightarrow d + \tau$. Escape literal percent signs as `\%` in LaTeX when documenting "% time in zones".

Appendix B: Estimating λ_{HRV} and return-to-baseline metrics

Acquisition protocol. (i) Record $\mathrm{HRV}^{\mathrm{pre}}$ (rMSSD) during 5–10 min of seated rest immediately before the session; (ii) record $\mathrm{HRV}^{\mathrm{post}}$ within 10–20 min after finishing, same posture; (iii) optionally record a late sample H_2 at 2–6 h.

Estimator. Given baseline $H_0 = \mathrm{HRV}^{\mathrm{pre}}$, early post $H_1 = \mathrm{HRV}^{\mathrm{post}}$ and late post H_2 with lag Δt_{12} (hours),

$$\lambda_{\mathrm{HRV}} = \frac{1}{\Delta t_{12}} \ln\!\left(\frac{|H_1 - H_0|}{|H_2 - H_0|}\right) \quad [\mathrm{h}^{-1}].$$

If only H_0, H_1 exist, use the velocity proxy $v_{\mathrm{HRV}} = \dfrac{H_1 - H_0}{\Delta t_{01}}$ (ms/h) and apply shrinkage toward the athlete prior: $\tilde{v}_{\mathrm{HRV}} = \alpha\, v_{\mathrm{HRV}} + (1-\alpha)\, \mu_{\mathrm{athlete}}$ with $\alpha \in [0,1]$.

Time-to-threshold. For a tolerance $\varepsilon > 0$, the time t_ε such that $|H(t) - H_0| \le \varepsilon$ under the exponential model is

$$t_\varepsilon = \frac{1}{\lambda_{\mathrm{HRV}}} \ln\!\left(\frac{|H_1 - H_0|}{\varepsilon}\right).$$

Quality control. Discard segments with artifacts (ectopy, motion); use winsorization or Hampel filters on short HRV windows; log-transform HRV deltas if heavy-tailed; report units explicitly (ms, h^{-1}).

Appendix C: Same-day non-commutative carry-over operator

Two sessions in a day. For sessions $s = 1, 2$ on day d separated by a gap τ_{gap} (h), predict the pre-session HRV of the second session by

$$\widehat{\mathrm{HRV}}^{\mathrm{pre}}_{d,2} = \mathrm{HRV}^{\mathrm{pre}}_{d} + \left(\mathrm{HRV}^{\mathrm{post}}_{d,1} - \mathrm{HRV}^{\mathrm{pre}}_{d}\right) e^{-\lambda_{\mathrm{HRV},d,1}\, \tau_{\mathrm{gap}}}.$$

Operator definition. Define $\odot$ acting on an ordered pair $(\boldsymbol{v}^{\mathrm{PLT}}_{d,1}, \boldsymbol{v}^{\mathrm{PLT}}_{d,2})$ by updating the second vector with $\widehat{\mathrm{HRV}}^{\mathrm{pre}}_{d,2}$ and concatenating contextual residues:

$$(\boldsymbol{v}^{\mathrm{PLT}}_{d,1}, \boldsymbol{v}^{\mathrm{PLT}}_{d,2}) \;\mapsto\; \tilde{\boldsymbol{v}}^{\mathrm{PLT}}_{d,2} = \Pi\big(\boldsymbol{v}^{\mathrm{PLT}}_{d,2} \,\|\, \tau_{\mathrm{gap}} \,\|\, E_{\mathrm{kg},d,1} \,\|\, \lambda_{\mathrm{HRV},d,1}\big).$$

In general $\boldsymbol{v}^{\mathrm{PLT}}_{d,2} \odot \boldsymbol{v}^{\mathrm{PLT}}_{d,1} \neq \boldsymbol{v}^{\mathrm{PLT}}_{d,1} \odot \boldsymbol{v}^{\mathrm{PLT}}_{d,2}$ whenever τ_{gap} or λ_{HRV} differ.

Extension to $S_d > 2$. Recursively apply the update left-to-right: $\tilde{\boldsymbol{v}}^{\mathrm{PLT}}_{d,i+1} = \Pi\big(\boldsymbol{v}^{\mathrm{PLT}}_{d,i+1} \,\|\, \tau_{\mathrm{gap},i\to i+1} \,\|\, E_{\mathrm{kg},d,i} \,\|\, \lambda_{\mathrm{HRV},d,$ This yields an ordered, gap-aware daily stack before aggregation.

References

[1] E. W. Banister, T. W. Calvert, M. V. Savage, and T. Bach, "A systems model of training for athletic performance," *Australian Journal of Sports Medicine*, 7(3):57–61, 1975.

[2] T. W. Calvert, E. W. Banister, M. V. Savage, and T. Bach, “A systems model of the effects of training on physical performance,” *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(2):94–102, 1976.

[3] H. Monod and J. Scherrer, “The work capacity of a synergic muscular group,” *Ergonomics*, 8(3):329–338, 1965.

[4] A. Vanhatalo, A. M. Jones, and M. Burnley, “Application of critical power in sport,” *Int. J. Sports Physiol. Perform.*, 6(1):128–136, 2011.

[5] A. M. Jones, A. Vanhatalo, and M. Burnley, “The ‘Critical Power’ Concept: Applications to Sports Performance,” *Sports Medicine*, 47:65–78, 2017.

[6] P. F. Skiba, W. Chidnok, J. D. Clarke, and A. M. Jones, “Modeling the expenditure and reconstitution of work capacity above critical power,” *Med. Sci. Sports Exerc.*, 44(8):1526–1532, 2012.

[7] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, “Heart rate variability: standards of measurement, physiological interpretation and clinical use,” *Circulation*, 93(5):1043–1065, 1996.

[8] H. Allen, A. R. Coggan, and S. McGregor, *Training and Racing with a Power Meter*, 3rd ed. Boulder, CO: VeloPress, 2019.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.