Task 2: Feature Extraction & Interpretation for Mobile Sensors

Understanding Mean/Variance Windows, 3D Magnitudes, and What the Plots Tell Us

Luisa Faust

2025

Abstract

The goal of this task is to extract informative features from tri-axial mobile sensors and to *understand* what these features express physically. I explain in plain language the meaning of windowed **mean** and **variance**, engineered features such as **3D magnitudes**, **jerk** and **angular acceleration**, and I interpret each visualization. Screenshots from my Jupyter workflow and all resulting plots are included. The code supports both live download from EdgeML and a local CSV fallback; this report analyzes the fallback example for clarity.

1 Signals and Their Physical Meaning

Accelerometer (X_a, Y_a, Z_a) measures *linear acceleration (lineare Beschleunigung)*. A device at rest on a table still measures $\approx 9.81\,\mathrm{m/s^2}$ along "down" because of gravity.

Gyroscope (X_g, Y_g, Z_g) measures *angular velocity (Winkelgeschwindigkeit)* in °/s. It reacts to rotations, not translations.

Orientation (α, β, γ) are angles (yaw–pitch–roll) derived from sensor fusion. They describe *pose* in space.

2 Windowing: Why Mean and Variance?

We summarize the streaming data in short, overlapping time windows of W samples. For each axis $x_t^{(\cdot)}$ in a window W_k:

$$\textbf{Mean } \mu_k^{(\cdot)} = \frac{1}{|W_k|} \sum_{t \in W_k} x_t^{(\cdot)} \qquad \text{(average level / pose)},$$

$$\textbf{Variance } \sigma_k^{2(\cdot)} = \frac{1}{|W_k| - 1} \sum_{t \in W_k} (x_t^{(\cdot)} - \mu_k^{(\cdot)})^2 \qquad \text{(activity intensity)}.$$

Intuition. The mean moves slowly and reflects the typical orientation or bias (e.g., phone upright vs. flat). The variance grows with movement: sitting ⇒ low; walking ⇒ medium, periodic; running/shaking ⇒ high. Windowing reduces noise and produces fixed-length feature vectors suitable for ML **pandas_rolling**.

3 Engineered Features: What They Capture

From the fused table (all sensors aligned by timestamp) I compute:

$$
\begin{aligned}
\text{AccMag} &= \sqrt{X_a^2 + Y_a^2 + Z_a^2} && \text{overall movement incl. gravity,} \\
\text{GyroMag} &= \sqrt{X_g^2 + Y_g^2 + Z_g^2} && \text{how fast we rotate,} \\
\text{OrientMag} &= \sqrt{\alpha^2 + \beta^2 + \gamma^2} && \text{pose change amplitude,} \\
\text{Jerk}_{X,Y,Z} &= \Delta\{X_a, Y_a, Z_a\} && \text{abrupt changes of linear motion,} \\
\text{AngAcc}_{X,Y,Z} &= \Delta\{X_g, Y_g, Z_g\} && \text{abrupt changes of rotation,} \\
\text{TotalEnergy} &= \text{AccMag} + \text{GyroMag} && \text{combined motion intensity.}
\end{aligned}
$$

Rules of Thumb

- **Pose vs. Activity:** Orientation means → pose; accelerometer/gyroscope variances and magnitudes → activity intensity.
- **Sudden events:** Spikes in *Jerk* or *AngAcc* indicate steps, taps, turns, or impacts.
- **Device placement:** Pocket vs. hand changes axis alignment; magnitudes are more robust than raw axes.

4 Interpreting My Plots

(a) Window Means & Variances

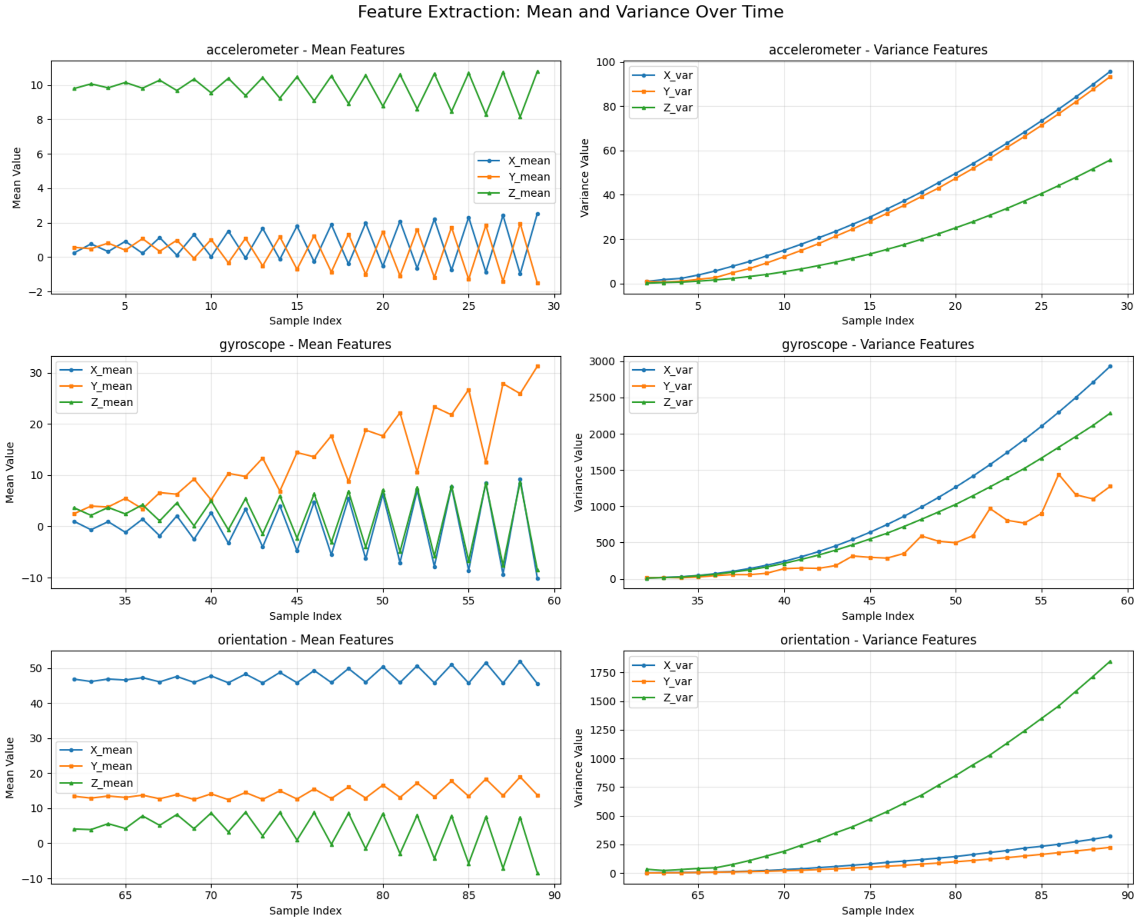

Figure 1: Window statistics (mean and variance) across sensors.

Reading the figure.

- *Accelerometer means* hover around gravity-dominated values (Z near ≈ 9.8). Slow drifts indicate tilt changes.
- *Variances* increase steadily in later windows ⇒ motion gets stronger (transition from idle to walking-like pattern).
- *Gyroscope variance* rises faster than accelerometer variance when rotations dominate (turning the phone).
- *Orientation means* show staircase-like changes ⇒ pose adjustments; their variance remains smaller during steady holds.

Conclusion: What is happening in the data?

Short answer: The person starts *stationary/idle*, then transitions into *walking* with increasing intensity; towards the end there are *pronounced turns/re-orientations* of the device.

Where is this visible in Fig. 1?

- **Phase 1 – Idle (early windows):** Very low variance in *accelerometer*, *gyroscope*, and *orientation*. Orientation means are nearly flat. ⇒ phone is held or lying still.
- **Phase 2 – Walking (middle windows):** *Accelerometer variance* in the top-right panel rises steadily; the accelerometer *means* (top-left) show a subtle sawtooth pattern. ⇒ periodic linear accelerations from steps.
- **Phase 3 – Turns/Rotations (late windows):** *Gyroscope variance* (middle-right) increases *faster* than accelerometer variance; *orientation means* (bottom-left) show staircase-like jumps and *orientation variance* (bottom-right) grows. ⇒ clear device rotations/turns.

Interpretation.

- Rising *accelerometer variance* ⇒ growing **movement intensity** (idle → walking).
- Strongly rising *gyroscope variance* ⇒ **rotation-dominant motion** (turns, quick heading changes).
- Stair-step *orientation means* ⇒ distinct **pose changes**; low orientation variance during steady holds.

Overall picture. The sequence is consistent with **starting from standstill**, then **walking with increasing vigor**, and finally **executing noticeable turns** (or actively rotating the phone).

(b) Pairplot of Magnitudes

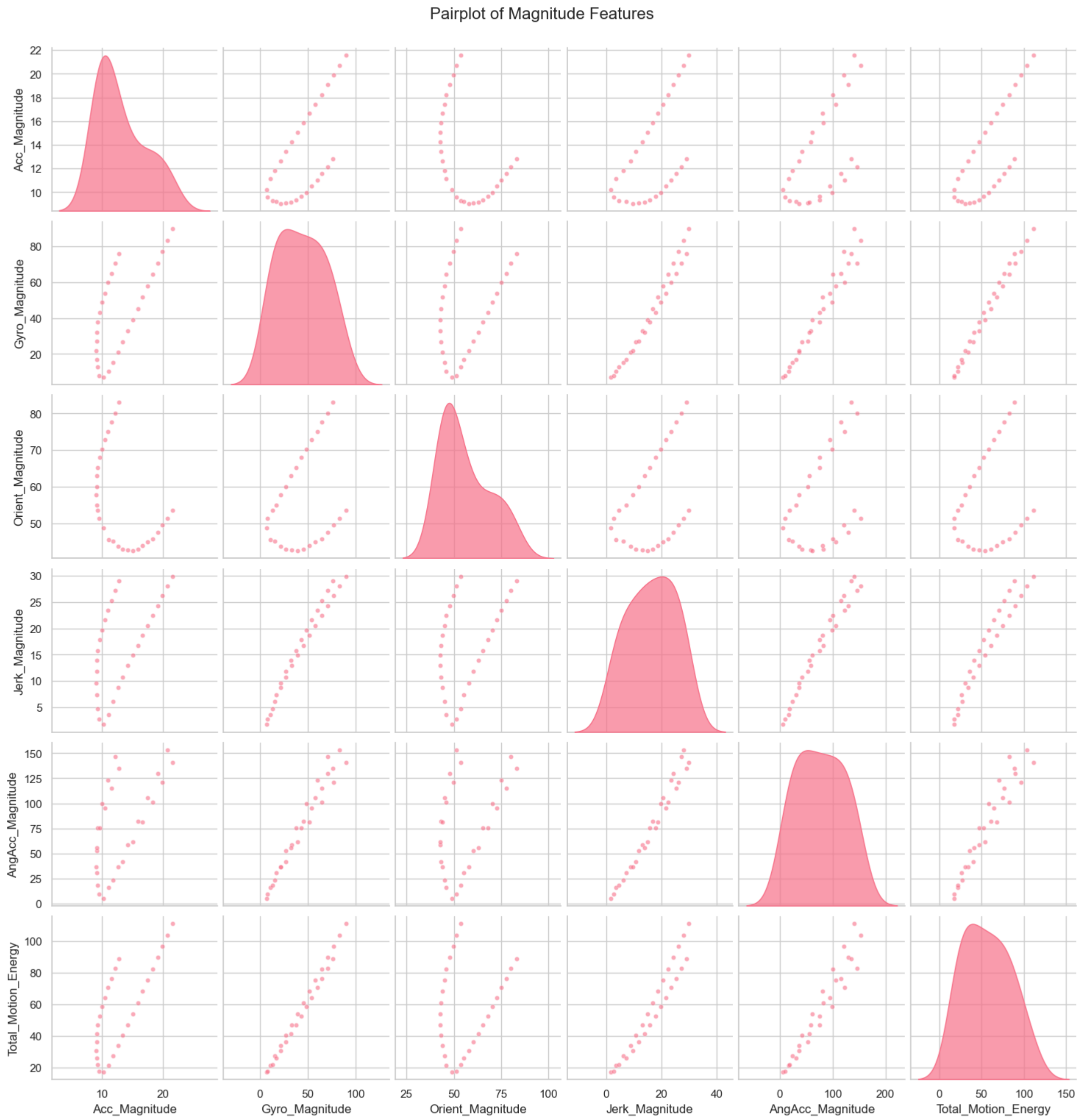

Figure 2: Pairwise relations among Acc/Gyro/Orient/Jerk/AngAcc/TotalEnergy magnitudes.

What it says. We see clear monotonic relations: higher **GyroMag** correlates with higher **AngAcc_Magnitude** (rotations and their change co-occur). **TotalEnergy** increases with both **AccMag** and **GyroMag**. The clouds are well-separated along diagonals ⇒ good candidates for simple classifiers.

Why this pairplot matters. A pairplot shows, for every feature pair, how their values co-vary (off-diagonal) and how each feature is distributed on its own. Because I plotted *orientation-invariant magnitudes* (Acc, Gyro, Orient, Jerk, AngAcc, TotalEnergy), this view tells us (i) which signals carry similar information (redundancy), (ii) which combinations separate behaviours (useful for simple rules or classifiers), and (iii) whether relations are linear or curved (model choice).

What the figure says.

- **GyroMag vs. AngAcc_Magnitude:** strong, almost monotonic positive relation. Rotations and their rate-of-change co-occur. Actionable: one of the two may be dropped to reduce redundancy; or keep both if noise characteristics differ.

- **TotalEnergy vs. (AccMag, GyroMag):** clear positive trends—TotalEnergy grows when either translation (Acc) or rotation (Gyro) increases. This confirms TotalEnergy as a compact activity-intensity feature.
- **AccMag vs. Jerk_Magnitude:** generally increasing but often *curved/J-shaped*. Jerk reacts to *changes* in acceleration: high jerk spikes appear at transitions (steps, taps) rather than during steady motion—hence correlation is positive but not strictly linear.
- **Acc/Gyro vs. OrientMag:** weaker or curved trends. Pose magnitude rises when the device is tilted/turned, but not proportionally to linear acceleration. This indicates OrientMag contributes complementary pose information rather than pure intensity.
- **Diagonals (KDEs = Kernel Density Estimate (deutsch: Kerndichteschätzung)):** right-skewed distributions for *AccMag*, *TotalEnergy*, and often *Jerk/AngAcc*: many low-activity windows with a tail of high activity—typical for sequences that move from idle to walking/turning.

Why it is useful for modelling.

- **Feature selection:** High redundancy (e.g. GyroMag–AngAcc) suggests selecting one or combining them to avoid multicollinearity.
- **Simple rules possible:** The diagonal "bands" indicate that thresholding AccMag/GyroMag (and optionally Jerk/AngAcc) can already separate idle vs. walking vs. turning.
- **Model choice:** Mostly monotonic but partly *curved* relations imply linear models can work with interaction terms, while trees/boosting capture the curvature without manual engineering.
- **Diagnostics:** Outliers or heteroscedastic spreads would flag sensor glitches or placement changes; none dominate here.

Takeaway. The pairplot confirms a coherent activity progression: as overall movement increases, both *AccMag* and *GyroMag* rise, *AngAcc* rises particularly during turns (rotation onsets), and *TotalEnergy* aggregates both effects. It validates that a small, robust set {*AccMag*, *GyroMag or AngAcc*, *JerkMag*, *TotalEnergy*} captures most discriminative structure for downstream tasks.

(c) Correlation Heatmap

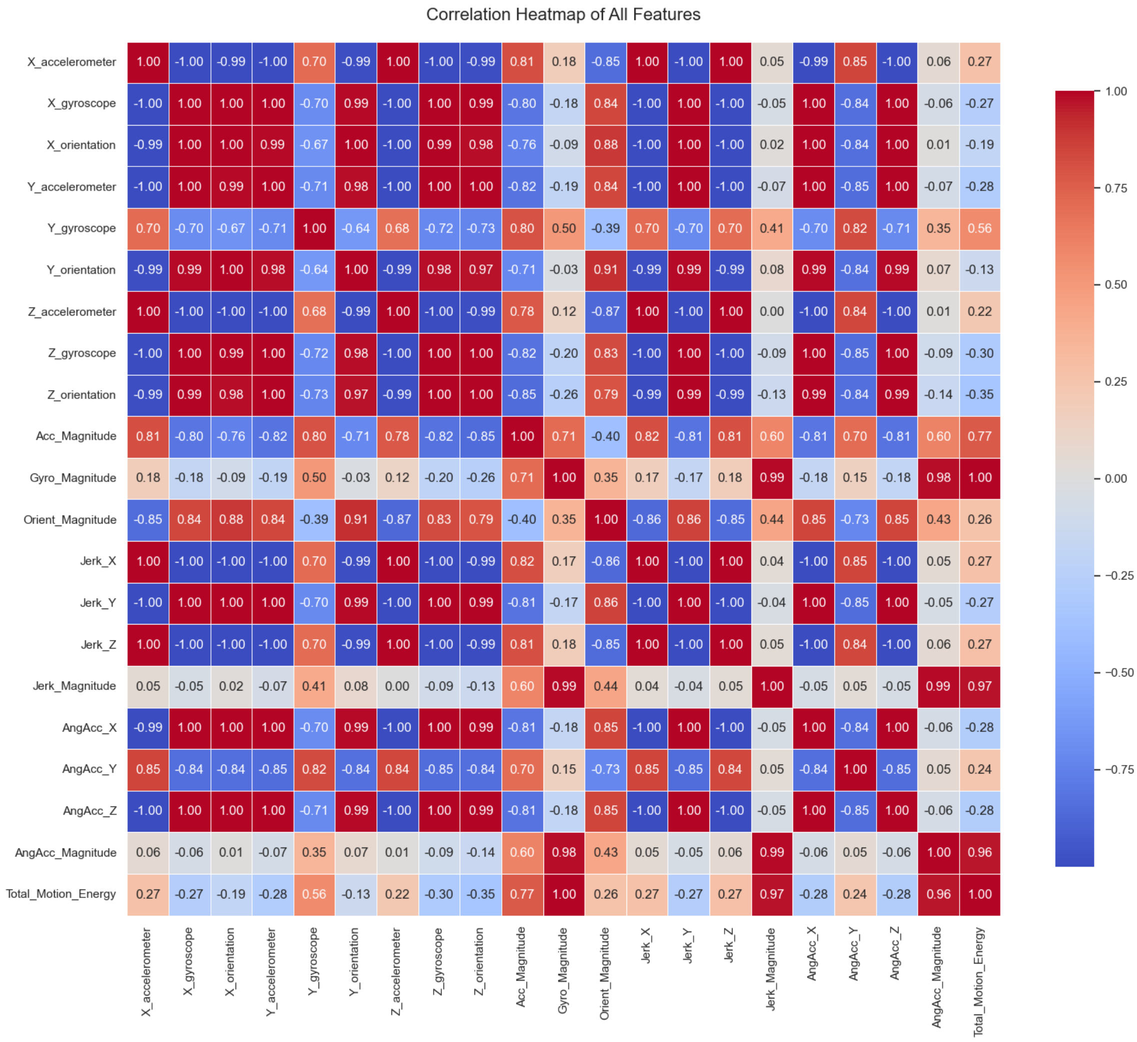

Figure 3: Feature correlation. Red/blue → strong ± correlation.

Interpretation. Axis pairs (e.g., X_a vs. X_o) show strong ± 1 correlations because device pose maps axes deterministically. Magnitudes are *less collinear* and thus carry complementary information. For modeling, we can keep magnitudes and drop many raw axes or use PCA to reduce redundancy.

(d) Distributions of Magnitudes

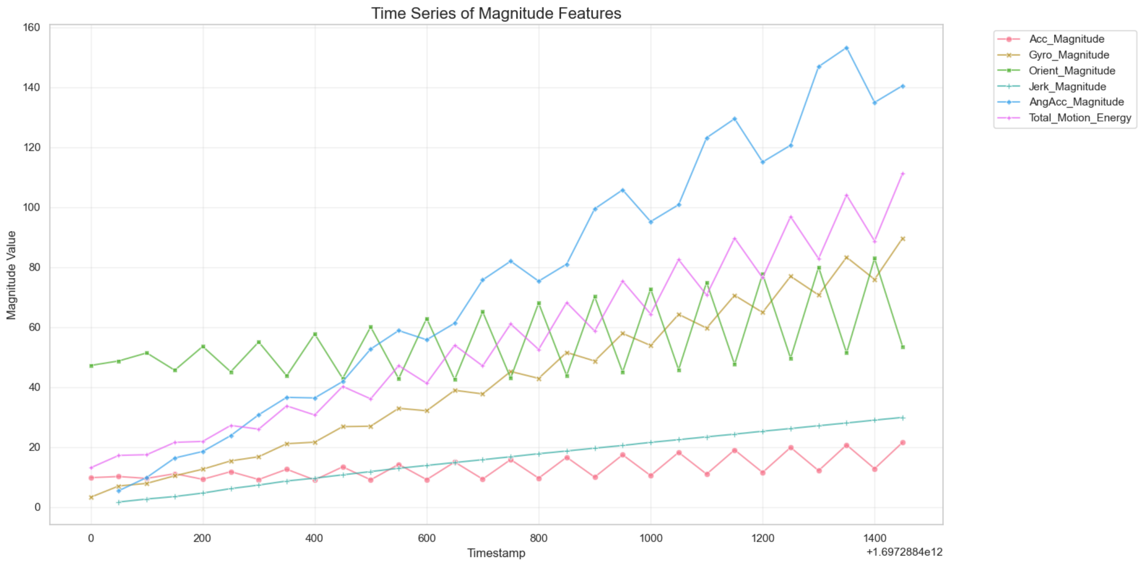

Figure 4: Histograms with KDE overlays for magnitude features.

Reading the figure. AccMag is right-skewed: most windows are low-movement with a tail of higher activity. **GyroMag** shows a broader spread when rotations occur. Distinct distribution shapes hint that a threshold or quantile-based rule could already separate low/high activity.

(e) Time Series of Magnitudes

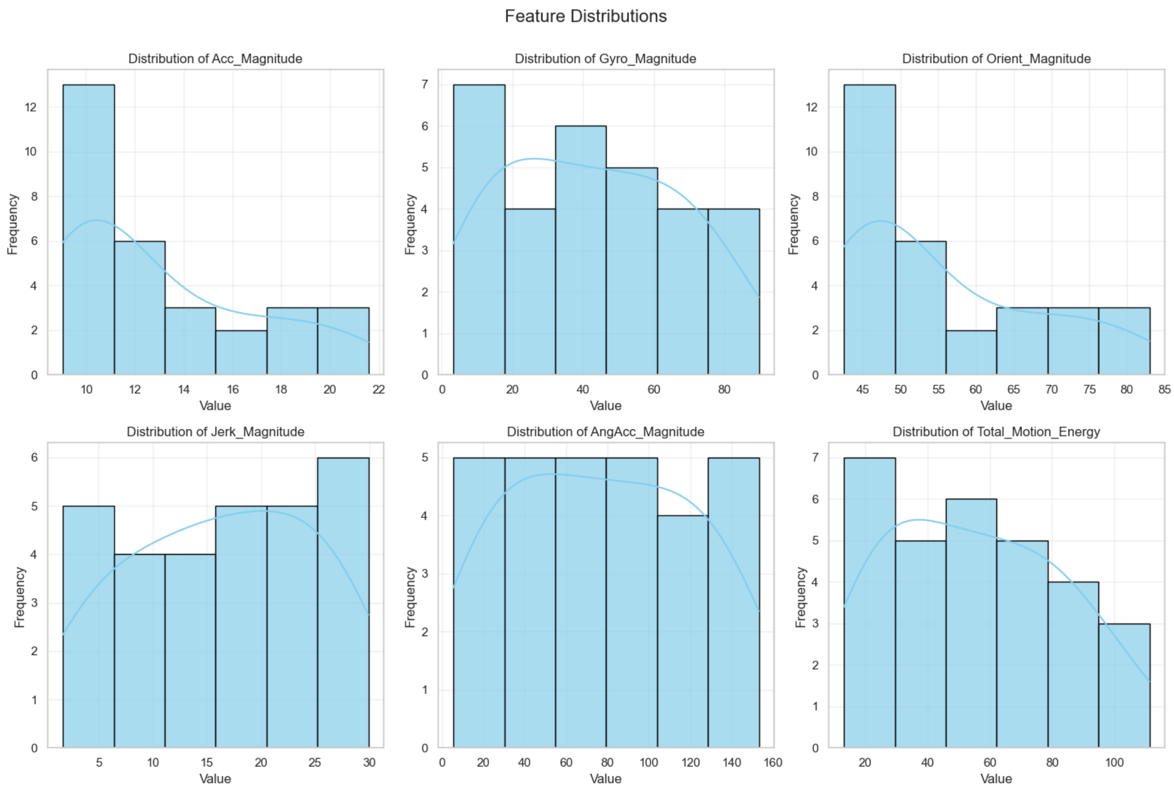

Figure 5: Temporal evolution of magnitude features.

Interpretation. All magnitudes trend upward during the motion sequence: **AngAcc** grows fastest (frequent reorientation), and **TotalEnergy** accumulates both translation and rotation. Peaks likely align with steps/turns.

Why this plot is important.

- **Threshold design:** The shapes and value ranges indicate practical cutoffs (e.g., idle vs. walking vs. turning).
- **Feature scaling:** Skewed or heavy–tailed features suggest using robust scaling or log transforms (e.g., for *AngAcc* and *TotalEnergy*) before modeling.
- **Class separability:** Distinct distribution shapes imply complementary signal—combining *AccMag* (translation), *Gyro/AngAcc* (rotation), and *Jerk* (transitions) covers different motion aspects.
- **Quality check:** Outliers, multi–modality, or very narrow peaks can reveal sensor issues, mixed behaviors, or overly aggressive windowing.