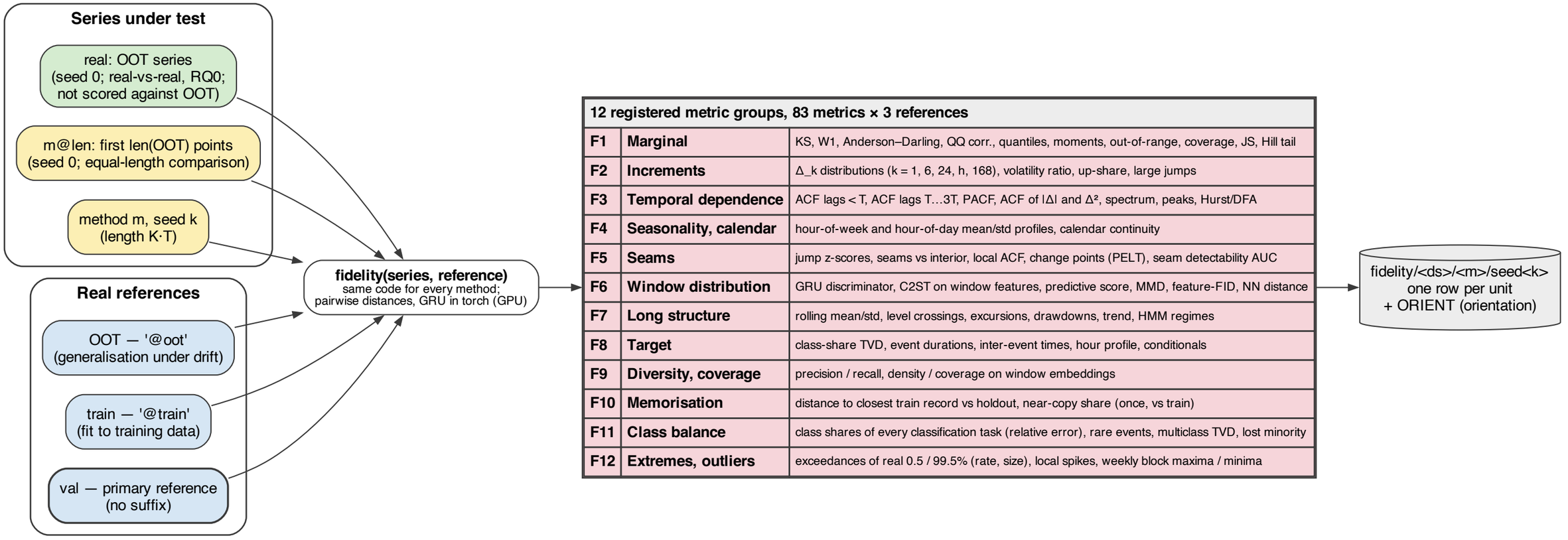

Series under test
real: OOT series
(seed 0; real-vs-real, RQ0;
not scored against OOT)
m@len: first len(OOT) points
(seed 0; equal-length comparison)
method m, seed k
(length K·T)
Real references
OOT — '@oot'
(generalisation under drift)
train — '@train'
(fit to training data)
val — primary reference
(no suffix)
fidelity(series, reference)
same code for every method;
pairwise distances, GRU in torch (GPU)
12 registered metric groups, 83 metrics × 3 references
F1 Marginal KS, W1, Anderson–Darling, QQ corr., quantiles, moments, out-of-range, coverage, JS, Hill tail
F2 Increments Δ_k distributions (k = 1, 6, 24, h, 168), volatility ratio, up-share, large jumps
F3 Temporal dependence ACF lags < T, ACF lags T…3T, PACF, ACF of |Δ| and Δ², spectrum, peaks, Hurst/DFA
F4 Seasonality, calendar hour-of-week and hour-of-day mean/std profiles, calendar continuity
F5 Seams jump z-scores, seams vs interior, local ACF, change points (PELT), seam detectability AUC
F6 Window distribution GRU discriminator, C2ST on window features, predictive score, MMD, feature-FID, NN distance
F7 Long structure rolling mean/std, level crossings, excursions, drawdowns, trend, HMM regimes
F8 Target class-share TVD, event durations, inter-event times, hour profile, conditionals
F9 Diversity, coverage precision / recall, density / coverage on window embeddings
F10 Memorisation distance to closest train record vs holdout, near-copy share (once, vs train)
F11 Class balance class shares of every classification task (relative error), rare events, multiclass TVD, lost minority
F12 Extremes, outliers exceedances of real 0.5 / 99.5% (rate, size), local spikes, weekly block maxima / minima
fidelity/<ds>/<m>/seed<k>
one row per unit
+ ORIENT (orientation)