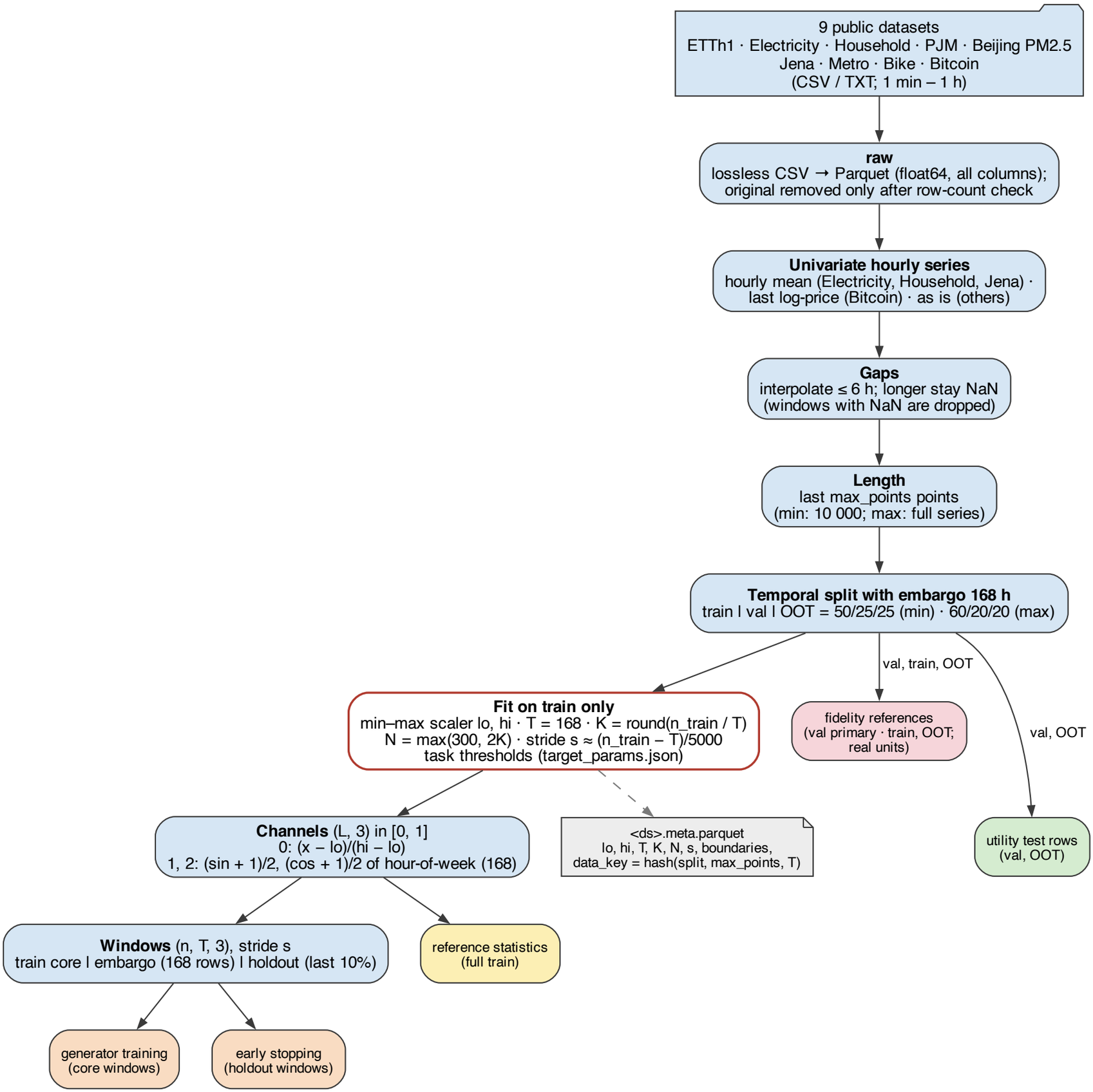

9 public datasets
ETTh1 · Electricity · Household · PJM · Beijing PM2.5
Jena · Metro · Bike · Bitcoin
(CSV / TXT; 1 min – 1 h)
raw
lossless CSV → Parquet (float64, all columns);
original removed only after row-count check
Univariate hourly series
hourly mean (Electricity, Household, Jena) ·
last log-price (Bitcoin) · as is (others)
Gaps
interpolate ≤ 6 h; longer stay NaN
(windows with NaN are dropped)
Length
last max_points points
(min: 10 000; max: full series)
Temporal split with embargo 168 h
train | val | OOT = 50/25/25 (min) · 60/20/20 (max)
val, train, OOT
val, OOT
Fit on train only
min–max scaler lo, hi · T = 168 · K = round(n_train / T)
N = max(300, 2K) · stride s ≈ (n_train – T)/5000
task thresholds (target_params.json)
fidelity references
(val primary · train, OOT;
real units)
Channels (L, 3) in [0, 1]
0: (x – lo)/(hi – lo)
1, 2: (sin + 1)/2, (cos + 1)/2 of hour-of-week (168)
<ds>.meta.parquet
lo, hi, T, K, N, s, boundaries,
data_key = hash(split, max_points, T)
utility test rows
(val, OOT)
Windows (n, T, 3), stride s
train core | embargo (168 rows) | holdout (last 10%)
reference statistics
(full train)
generator training
(core windows)
early stopping
(holdout windows)