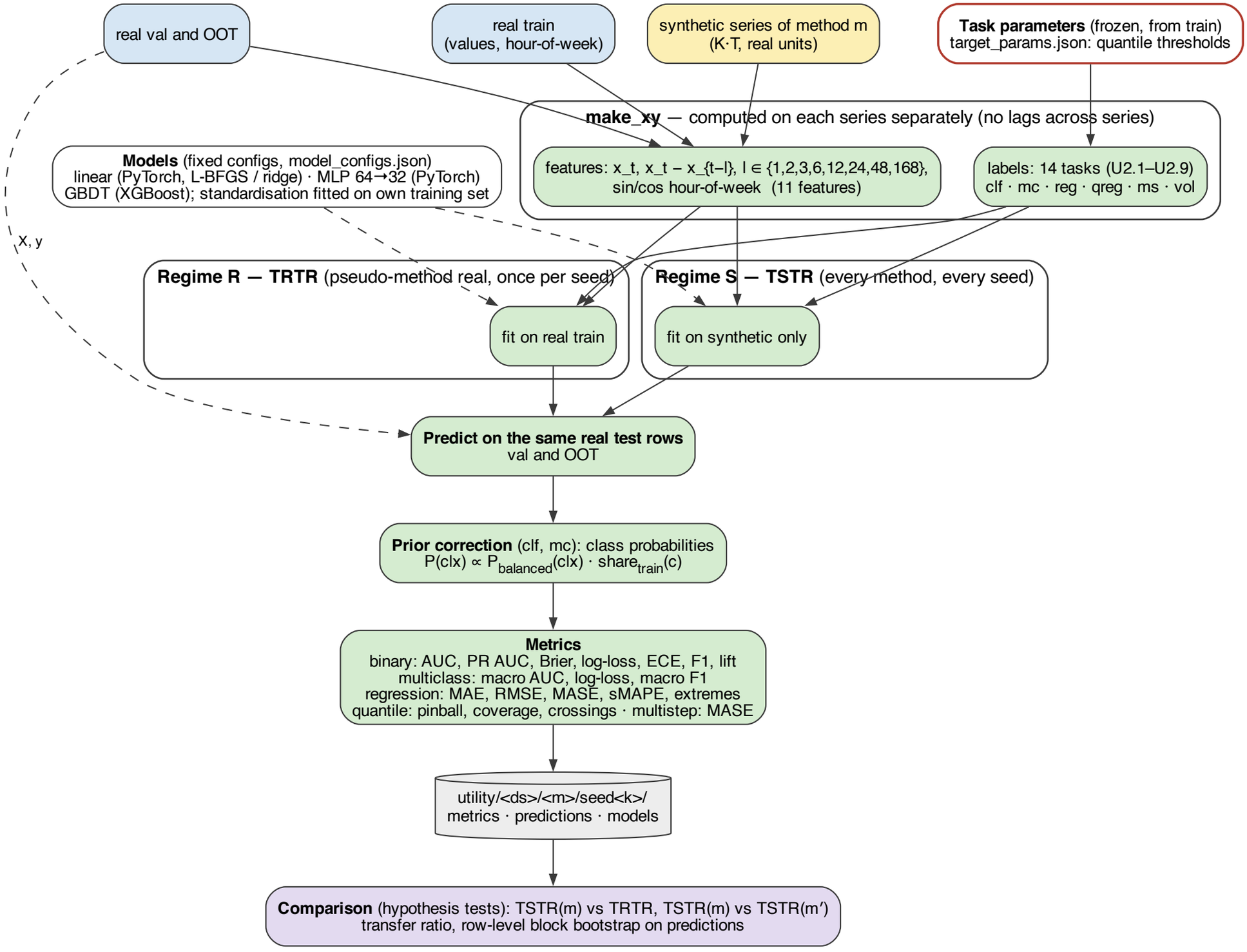

real val and OOT
real train
(values, hour-of-week)
synthetic series of method m
(K·T, real units)
Task parameters (frozen, from train)
target_params.json: quantile thresholds
make_xy — computed on each series separately (no lags across series)
features: x_t, x_t − x_{t−l}, l ∈ {1,2,3,6,12,24,48,168},
sin/cos hour-of-week (11 features)
labels: 14 tasks (U2.1–U2.9)
clf · mc · reg · qreg · ms · vol
Models (fixed configs, model_configs.json)
linear (PyTorch, L-BFGS / ridge) · MLP 64→32 (PyTorch)
GBDT (XGBoost); standardisation fitted on own training set
X, y
Regime R — TRTR (pseudo-method real, once per seed)
fit on real train
Regime S — TSTR (every method, every seed)
fit on synthetic only
Predict on the same real test rows
val and OOT
Prior correction (clf, mc): class probabilities
P(c|x) ∝ P_balanced(c|x) · share_train(c)
Metrics
binary: AUC, PR AUC, Brier, log-loss, ECE, F1, lift
multiclass: macro AUC, log-loss, macro F1
regression: MAE, RMSE, MASE, sMAPE, extremes
quantile: pinball, coverage, crossings · multistep: MASE
utility/<ds>/<m>/seed<k>/
metrics · predictions · models
Comparison (hypothesis tests): TSTR(m) vs TRTR, TSTR(m) vs TSTR(m′)
transfer ratio, row-level block bootstrap on predictions