3. Predictive Modeling

We used a cleaned dataset, specifically, the dataset collected hourly electricity usage data for 734 buildings over the past two years (2016-2017), generating approximately 17,500 timestamps per series. Our goal was to build a parsimonious forecasting framework that (1) uncovers the common drivers affecting many buildings, (2) allows for building-specific dynamics, and (3) produces reliable multi-step forecasts at daily, weekly, and monthly horizons.

We began by randomly subsampling ten of our original hourly building series (N = 10) to ease the computational burden of fitting a high-dimensional state-space model, while still retaining a representative cross-section of usage patterns. We then explored three aggregation levels—daily (731 observations), weekly (105), and monthly (24)—to balance noise reduction against the loss of short-term dynamics. Fitting a first-order Dynamic Factor Model (DFM) with three common factors to the daily averages yielded a log-likelihood of –31 245.7 (AIC = 62 589.4, BIC = 62 814.5). Despite some instability in the standard errors of the factor loadings (e.g. loading .f1 = 92.999 ±6 162 for “Wolf_education_Clarissa”), the daily model produced richly dynamic seven-day forecasts: “Wolf_education_Clarissa” plunges from 40.43 on January 1 to 7.09 by January 5, and “Rat_assembly_Ezequiel” dips to –113.5 before rebounding. By contrast, the weekly two-factor fit (LL = –3 916.4, AIC = 7 900.7) forecasts essentially flat four-week trajectories—“Robin_office_Victor” hovers around 36—while the monthly fit (LL = –841.2, AIC = 1 750.3) projects only gentle declines (e.g. “Wolf_education_Clarissa” drifts from 47.15 in January to 44.05 by May). We therefore favor the daily aggregation: it retains meaningful intra-week swings that are crucial for short-term planning, yet still captures cross-series co-movements via a parsimonious three-factor structure.

To capture the shared dynamics across buildings, we opted for a Dynamic Factor Model (DFM), whose core idea is that a small number of latent factors drive the co-movement of all series. By modeling ten series jointly through, say, three factors instead of ten separate ARIMA processes, we achieve massive dimensionality reduction, more stable estimates of common trends, and improved multivariate forecasts. They summarize common movements via a small number (k) of latent factors, each following an autoregressive process of order p. At the same time, each building retains its own “idiosyncratic” AR term, capturing local inertia. And by including the day-of-week dummies as exogenous inputs, we explicitly model seasonality.

Our first DFM fit on the raw daily data used three factors with AR(1) dynamics (no idiosyncratic autocorrelation). Although this “3 × AR(1)” specification ran quickly, it delivered a log-likelihood of –31 245.7 (AIC 62 589.4) and unstable loading standard errors, and its seven-day forecasts (e.g. “Wolf_education_Clarissa” plunging from ~40.4 to ~7.1, “Rat_assembly_Ezequiel” dipping below –113) were arguably over-reactive. We also tested the same first-order model on weekly (105 observations, 2 factors; LL = –3 916.4, AIC 7 900.7) and monthly (24 obs., 2 factors; LL = –841.2, AIC 1 750.3) aggregates. While the coarser series gave steadily lower AICs—monthly smoothing produced only gentle downward slopes (e.g. ~47.1→44.1 for “Wolf_education_Clarissa”)—they washed out the intra-week swings we needed for short-term planning. That trade-off led us to stick with daily data.

Next, we standardized each series (zero mean, unit variance) and allowed the factors themselves to follow AR(2) processes. The “3 × AR(2)” model improved log-likelihood to –7 337.2 (AIC 14 790.3) but still left highly significant residual autocorrelation (all Ljung–Box $p \ll 0.001$). Introducing an AR(1) idiosyncratic term for each series—that is, asking each building’s residual to follow its own AR(1)—brought the log-likelihood up to –5 291.3 (AIC 10 718.5) and most Ljung–Box p-values into comfortable nonsignificance (e.g. “Wolf…” $p \approx 1.3\times10^{-9}$), indicating that we had finally tamed the unmodeled serial dependence. The full set of AR(2) factor coefficients and nearly all factor loadings became highly significant ($|z| > 4.3$, $p < 0.001$), and forecasts showed realistic momentum and mean-reversion.

We briefly explored adding a second AR term to each idiosyncratic error (error_order = 2), but the AIC crept up slightly (to 10 729.4) and some series again exhibited low Ljung–Box p-values. Likewise, a seasonal extension with day-of-week dummy regressors failed to improve fit (LL = –7 774.7, AIC 15 805.4) and produced tiny, insignificant weekday coefficients.

Finally, in pursuit of parsimony, we reduced the number of common factors from three to two—our “strong” latent trends—while retaining AR(2) factor dynamics and AR(2) idiosyncratic errors. This “2-factor, AR(2) idio” specification yielded log-likelihood –6 086.1 (AIC 12 288.1), with all sixteen loadings significant (for example, loading f1 = 2.1109, z = 10.08 for “Wolf_education_Clarissa”) and most residual autocorrelations quashed. Its seven-day forecasts in original units remain richly dynamic—“Rat_assembly_Ezequiel” falls from ~107.9 on Jan 1 to ~7.1 by Jan 7—while our two-factor structure keeps the model lean and interpretable.

Recognizing that DFMs excel at summarizing shared structure but may miss nonlinear, building-specific quirks, we also sketched a gradient-boosted pipeline. We engineered lag features (1 h, 24 h, 168 h), rolling means (24 h, 168 h), and sine/cosine encodings of hour, day-of-week and month to capture cyclicality. By unstacking into a long panel and merging building metadata (primary use, building ID), we prepared an XGBoost with time-series cross-validation and sample weights inversely proportional to use-type frequency. Although a full randomized search crashed on our laptop, preliminary runs confirmed the value of each feature set—particularly 24-h lags and cyclical encodings—in reducing one-step RMSE.

In summary, our final “2-factor, AR(2) idiosyncratic” DFM on daily data strikes the sweet spot between parsimony and flexibility: it collapses ten noisy series into two interpretable drivers, captures local inertia via idiosyncratic ARs, and delivers nuanced seven-day forecasts that honor both common and building-specific dynamics. Future work will explore ensemble blends of latent-factor and tree-based models, deeper stratified subsampling, and automated pipelines to extend the framework back to all 734 buildings.