

## 1. Spatial and temporal variability of traffic data

I examined volume, occupancy, and speed data from two Sunnyside detectors, identified as 1345 and 1858. Detector 1345 monitors northbound (NB) traffic, while Detector 1858 captures southbound (SB) traffic. The time-series observations for each detector are illustrated in Figures 1 and 4.

Daily patterns were analyzed by computing the mean of five-minute observations for a randomly selected day. As shown in Figure 2, the NB data exhibit a typical Thursday pattern, with higher volumes occurring early in the morning and late in the afternoon. In contrast, Figure 5, capturing a Wednesday for SB, reveals lower volumes. A possible conclusion is that NB traffic may be heavier than SB traffic, with nearly double the volume when comparing the sampled days. However, because these observations come from different weekdays, direct comparisons should be made cautiously.

Finally, Figures 3 and 6 demonstrate the presence of weekly patterns in the data, suggesting recurring traffic trends over time.

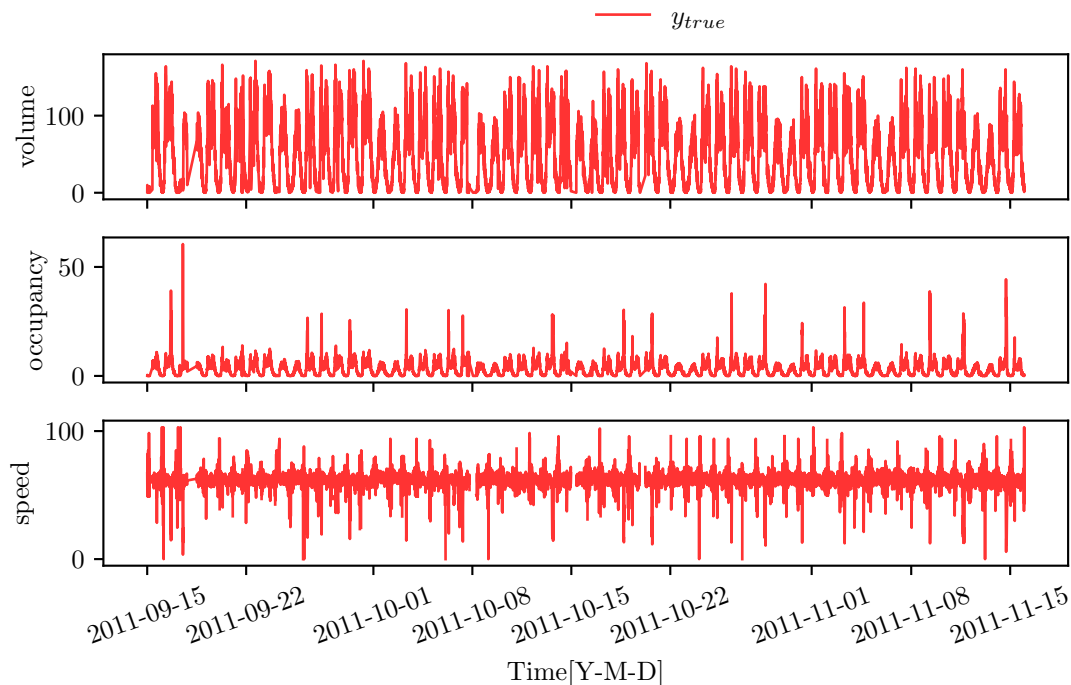


Figure 1: Time Series for Detector 1345.

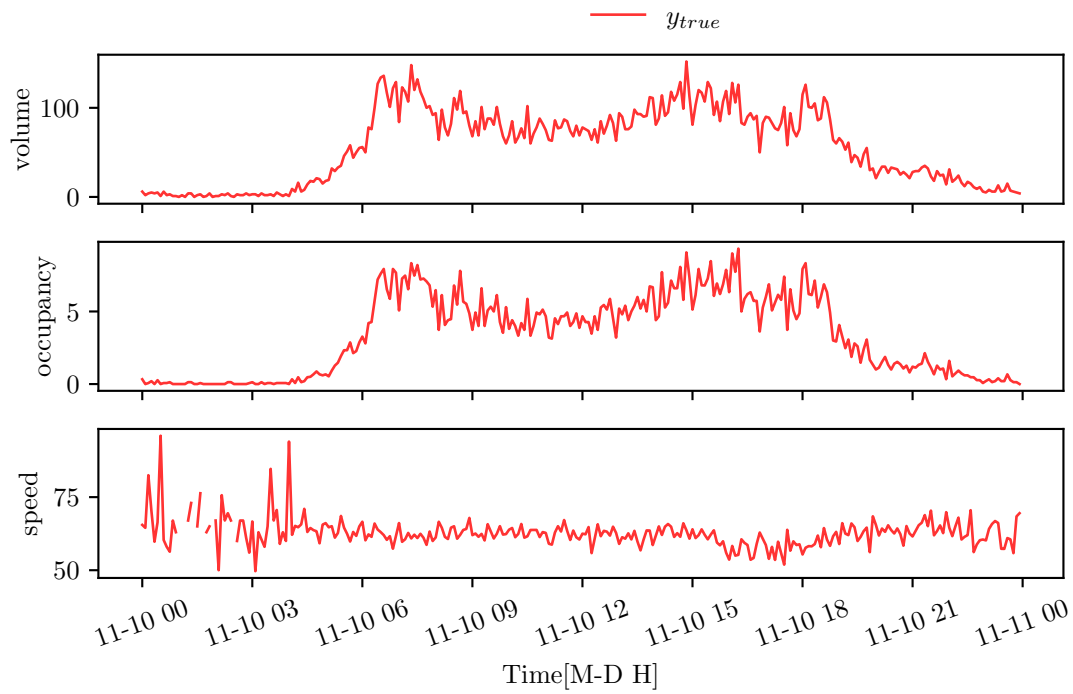


Figure 2: Daily Patterns for Detector 1345.

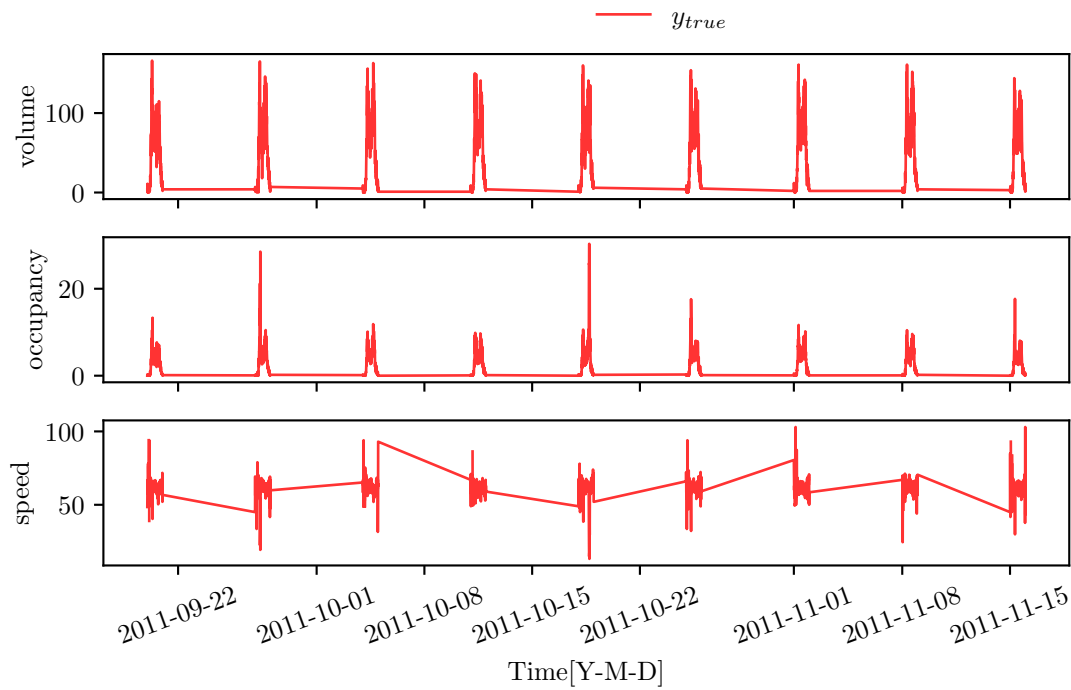


Figure 3: Weekly Patterns for Detector 1345.

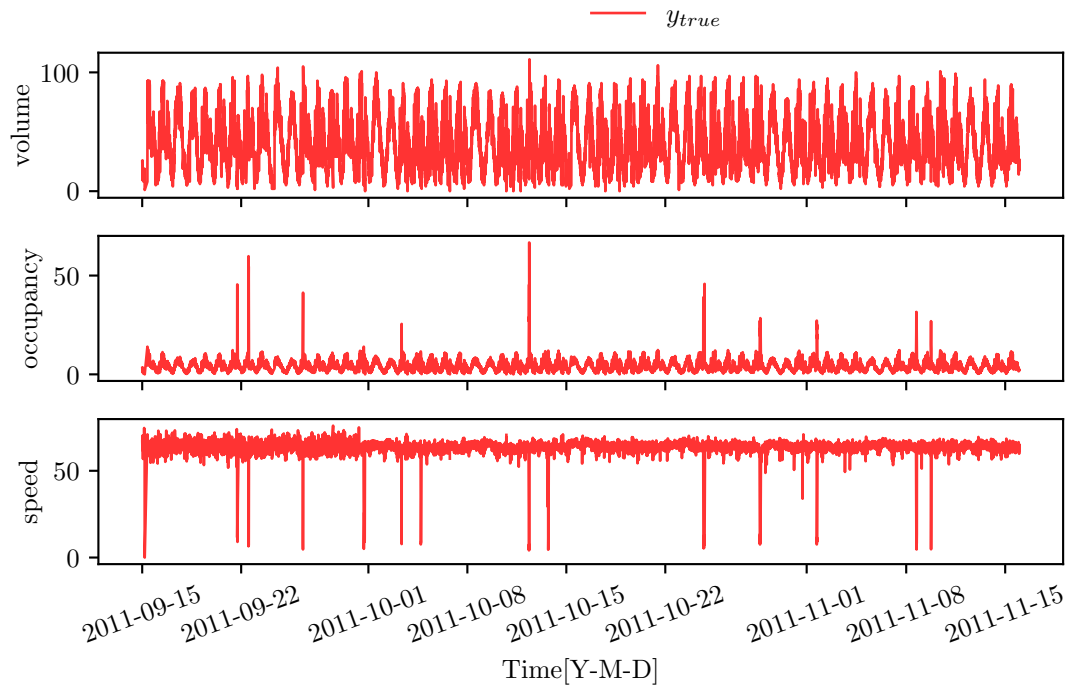


Figure 4: Time Series for Detector 1858.

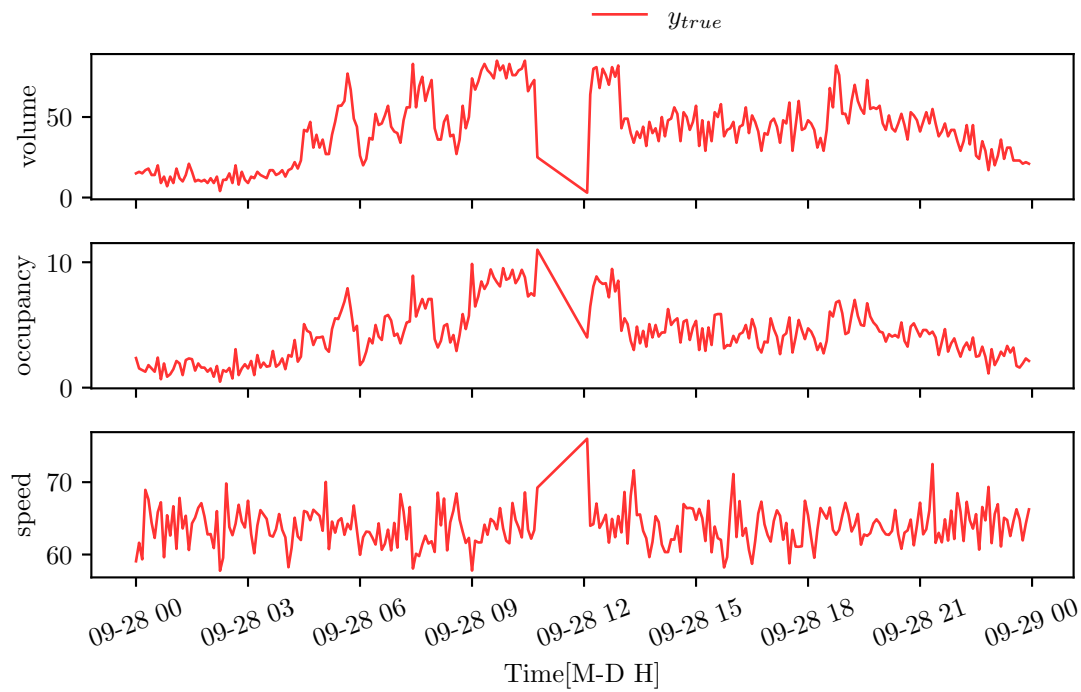


Figure 5: Daily Patterns for Detector 1858.

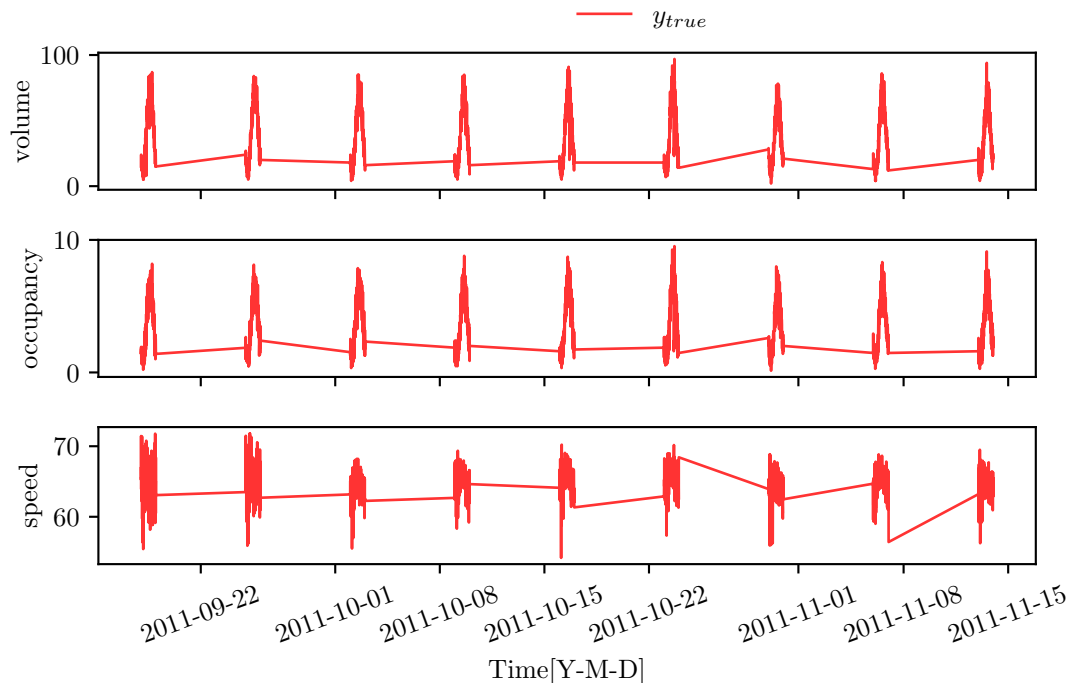


Figure 6: Weekly Patterns for Detector 1858.

## 2. Correlation between sensors

The correlation matrix shown in Figure 7 reveals that many sensor locations exhibit moderate to high correlations in their volume measurements (indicated by warmer/red colors). This structure typically reflects both physical proximity and shared traffic patterns, such as morning and afternoon rush hours. Sensors along the same corridor or in close proximity to each other could often show stronger correlations, suggesting that if data go missing at one location, readings from correlated sensors can help reconstruct the lost observations.

Sophisticated methods such as matrix completion, graph-based approaches, or low-rank factorization can exploit these spatio-temporal relationships to impute missing values. Moreover, such strong correlations are highly effective in global modelling approaches where one model is trained on a set of time series. Allowing the model to learn and identify these correlations will lead to better forecasting capabilities. This is something that I have tried to explore in my research proposal where I have looked at a global modelling with a probabilistic LSTM model that can also learn embeddings to represent time series in an unsupervised approach. These embeddings are designed to allow the model to build better a better local representation of the time series in a global setup. One thing that I have not explored yet is how to extract valuable information from these embeddings where in the context of this exercise to test whether these embedding are able to model the correlation observed between the time series.

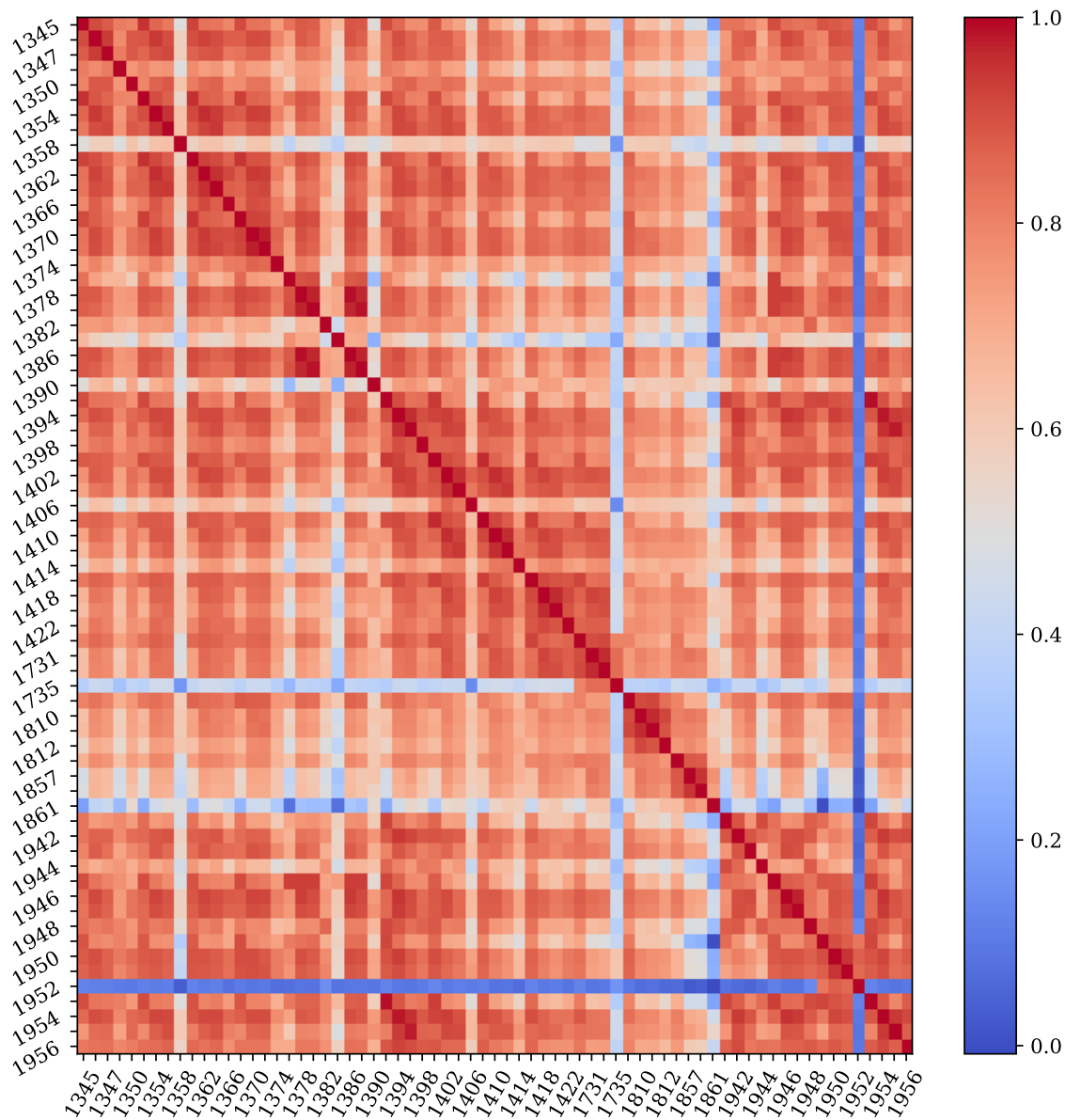


Figure 7: Correlation matrix between detectors for the volume observations.

### 3. Forecasting

Three broad categories of models are commonly used for short- or medium-term forecasting and for understanding relationships among time series variables: (1) Exponential Smoothing Models (ESM), which are straightforward and effective for data with level, trend, or seasonal components; (2) ARIMA-type models, including SARIMA for handling seasonal patterns and ARIMAX or VAR for incorporating explanatory variables or multiple interdependent series; and (3) Unobserved Component Models (UCM), which break down a series into trend, seasonal, and irregular components, and can readily include covariates. For more complex or highly nonlinear data, deep learning approaches (RNNs, LSTMs, GRUs, and Transformers) can capture intricate temporal dependencies and relationships.

For the purpose of this question I have used the local TAGI-LSTM model also presented in my research proposal. I did not perform any hyper parameter fine-tuning but rather used an

out of the box approach where I have trained the parameters of the model on a training set, and evaluated it's forecasting capacity on a test set that is highlighted in 8. Since, I have worked with 5min periodic data my training set included a great amount of data to train on, and due to the strong seasonality pattern the deployed model is able to accurately forecast on the test set as shown in Figure 9.

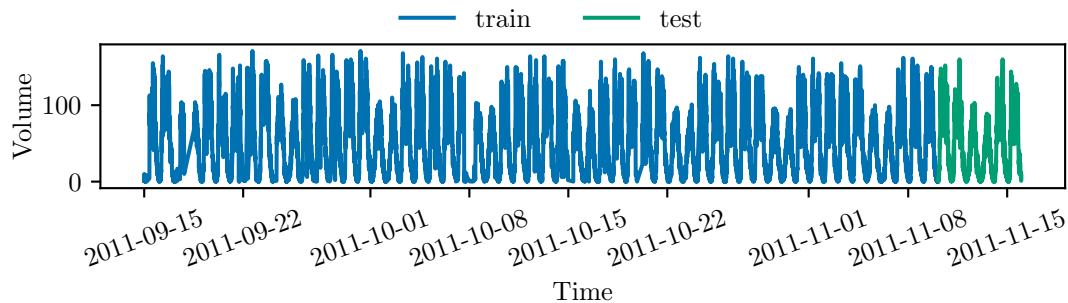


Figure 8: Target time series showing the training set and the test set.

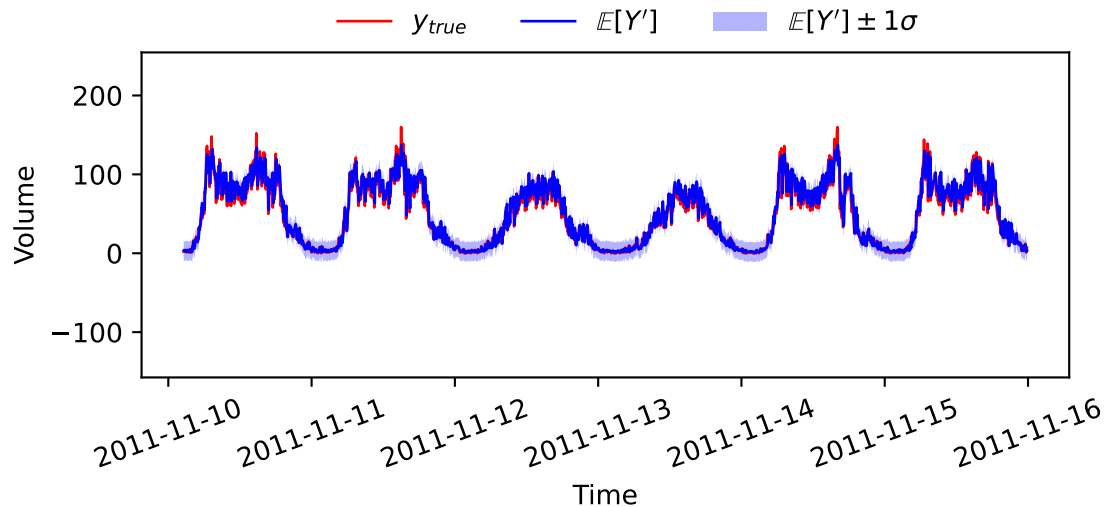


Figure 9: Model forecast on the test set.