

UNIVERSITY OF CRETE, Spring 2025

A Comparative Analysis of Robust Penalized Estimators for Periodic Time Series

Panos Evangelidakis

Tuesday 25th February, 2025

Professor: Dr. Yiannis Kamarianakis, Principal Researcher at IACM/FORTH



I extend my deepest gratitude to my esteemed professor for their invaluable guidance, to my friends for their unwavering support, and especially to my fiancée for their boundless patience and encouragement throughout this thesis journey. This endeavor would not have been possible without the invaluable advice and companionship offered by all who supported me.

“

The most important questions of life are, for the most part, really only problems of probability.

~Pierre Simon Marquis de Laplace

”

1 Motivation

This study conducts a rigorous comparative analysis of two state-of-the-art statistical approximation methods for robust variable selection: the **Least Squares Approximation (LSA)** [1] and **Maximum Tangent Likelihood Estimation (MTE)** [2], implemented through Least Absolute Deviations (LAD) and Least Squares (LS) regression frameworks under the **Maximum Entropy Bootstrap (MEB)** paradigm [3]. The investigation is motivated by the critical need for robust and interpretable traffic volume forecasting along a major arterial roadway in Athens (April 2000, TRANSFOR dataset). Our methodological approach employs variable selection to identify the optimal subset of low harmonic frequencies in traffic volume data collected from seven loop detector locations. This approach builds upon the foundational work of F.Rodrigues [4], who demonstrated that extracting harmonic seasonal components (representing historical averages) from non-stationary time series yields stationary residuals amenable to conventional time series analysis. This decomposition permits the application of classical methods (e.g., ARMA models) while maintaining forecasting accuracy comparable to computationally intensive deep learning approaches. While conventional LS and LAD regression remain widely employed, they exhibit well-documented limitations: LS estimators demonstrate sensitivity to multicollinearity and deviations from Gaussian error assumptions, while LAD estimators, though robust to outliers, lack inherent variable selection capabilities. To address these limitations, we implement a comprehensive evaluation framework utilizing **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)** minimization to optimize the trade-off between model fidelity and parsimony. Recognizing the demonstrated importance of low harmonic frequencies in traffic pattern analysis, we quantify their relative contributions through **Variable Inclusion Probabilities (VIPs)**. The culmination of this analysis is an innovative **adaptive LAD-Lasso** implementation, where the VIP vector informs the weighting scheme, thereby prioritizing the most statistically significant harmonic components. This integrated methodology advances the field by providing a principled, robust framework for traffic modeling that simultaneously addresses noise robustness, variable selection, and non-stationarity concerns.

2 Data

The dataset used in this study comes from the 2013 TRANSPORTATION DATA FORECASTING Competition (TRANSFOR), whose findings were revealed at the 92nd annual conference of the Transportation Research Board (TRB) in Washington D.C. [5]. The data was given by the National Technical University of Athens and consists of traffic occupancy measures taken every 90 seconds in April 2000 from seven loop detectors on traffic lights, indicated as $L \in \{A, B, C, D, E, F, G\}$ (see appendix section for more info). These detectors are located on Alexandras Avenue, a key arterial artery in Athens, Greece. Figure 1 shows that detectors A, B, C, and D monitor westbound traffic, whereas detectors E, F, and G assess eastbound traffic flow. The time and space where the detectors are positioned are critical because in April 2000, certain key events occurred that might explain causally some data anomalies. For example, the elections of April 9th, 2000 (where the central-left party, named "PASOK", won), the president of that party had various election campaigns at party's office is located on Charilaou Trikoupi (which crosses Alexandras Avenue on F detector), and on April 20, 2000, Panathinaikos, a Greek basketball club, won the Final Four of the Euroleague for the second time, and the team's association celebrated on Alexandras Avenue (between F and G

loop detectors). Collecting raw traffic occupancy data at 90-second intervals is challenging owing to the large volume of data and low signal-to-noise ratio [6]. Temporal aggregation is a typical strategy to solve these challenges, which smoothes the data and mitigates nonlinear and heteroskedastic effects ([7]). Traditionally, traffic occupancy data has been aggregated into 15-minute intervals before modeling, which improves data stability but makes extremely short-term forecasts less useful for practical applications such as real-time traffic management. Rather than using a multi-scale modeling technique, this study combines traffic occupancy data at 3-minute intervals, yielding 480 observations per day for each detector position. This technique strikes a compromise between data smoothing and the necessity for high-resolution forecasting, resulting in useful insights for current traffic control centers.



Figure 1: Loop detector locations on Alexandra's Avenue in Athens, Greece, with arrows indicating traffic flow direction. The image is taken from publication [8]

3 Methods

3.1 Harmonic Seasonal Model

By plotting the data we can clearly see that the data of traffic occupancy and volume are ruled by a seasonal effect as it was expected. Seasonal effects often exhibit smooth variations over time, making it more efficient to model them using continuous functions rather than discrete indices. A common method for incorporating smooth variation in a seasonal model is to use sine and cosine functions. Periodicity is a fundamental concept found in nature, such as in seasonal variations and oscillatory motion. A function is considered periodic if it repeats at regular intervals, known as the period. Sinusoidal functions, like sine and cosine, are prime examples of periodic functions. Given a constant $f > 0$, the function $\sin(2\pi ft)$ has a period $T = 1/f$, as it satisfies the condition:

$$\sin(2\pi f(t + T)) = \sin(2\pi ft) \quad \forall t \in \mathbb{R}.$$

Here, f denotes the frequency (cycles per unit time), while $\omega = 2\pi f$ represents the angular frequency (radians per unit time). According to Fourier theory, any piecewise continuous function $x(t)$ with period T can be decomposed into a sum of sinusoids with frequencies $f_k = k/T$ for $k = 1, 2, \dots$:

$$x(t) = A_0 + \sum_{k=1}^{\infty} \{A_k \cos(2\pi f_k t) + B_k \sin(2\pi f_k t)\}.$$

The coefficients A_k and B_k are computed using integrals over a full period, ensuring that the function is optimally expressed as a sum of orthogonal sinusoidal components.

In real-world applications, like ours, this infinite sum is typically approximated by a finite number of terms, leading to a truncation error:

$$\int_0^T |x(t) - \tilde{x}(t)|^2 dt = \sum_{k=K+1}^{\infty} \frac{1}{2} T (A_k^2 + B_k^2).$$

where $\tilde{x}(t)$ is defined as

$$\tilde{x}(t) := A_0 + \sum_{k=1}^K \{A_k \cos(2\pi f_k t) + B_k \sin(2\pi f_k t)\}.$$

A significant advantage of using sinusoidal approximations is their time-invariance. If the function undergoes a time shift by τ , the approximation preserves the same frequencies and error, though the coefficients adjust as follows:

$$A'_k := A_k \cos(2\pi f_k \tau) + B_k \sin(2\pi f_k \tau), \quad B'_k := -A_k \sin(2\pi f_k \tau) + B_k \cos(2\pi f_k \tau).$$

Since the sum of squared coefficients remains unchanged, the total error remains unaffected by the shift. This property underscores the essential role of sinusoidal functions in accurately modeling periodic signals.

It's worth noting that the coefficients formula are calculated based on the squared error loss. Changing the error function into a non-smooth one, is going to affect the resultant coefficients estimation. Since the original equation is nonlinear due to the presence of ϕ within the sine function, the reformulated version is preferred, as it allows for estimation using ordinary least squares (OLS) [9], least absolute deviations (LAD) or any penalized version of those two. For a time series $\{x_t\}$ with s seasons, there are $\lfloor s/2 \rfloor$ possible cycles, and the harmonic seasonal model is defined as:

$$x_t = m_t + \sum_{i=1}^{\lfloor s/2 \rfloor} s_i \sin\left(\frac{2\pi i t}{s}\right) + c_i \cos\left(\frac{2\pi i t}{s}\right) + z_t,$$

where m_t represents the trend component, and s_i and c_i are unknown parameters.

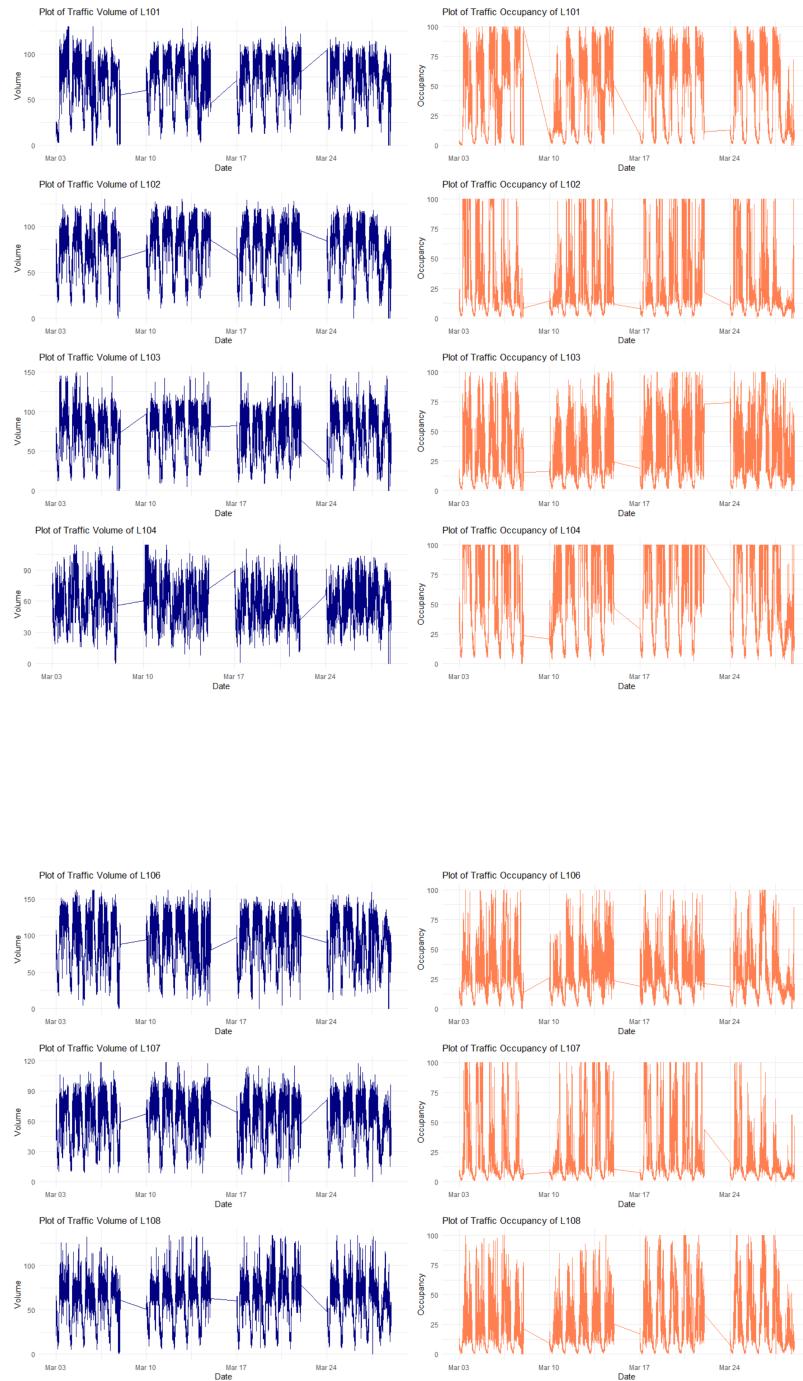


Figure 2: *Traffic Volume and Occupancy on Alexandras Avenue, Athens Greece, 2000*

The trend can be modeled using a polynomial form. When s is an even number, the sine term at frequency $1/2$ (i.e., when $i = s/2$) is always zero for all $t > 0$, meaning it can be excluded from the model. Consequently, the maximum number of parameters in the harmonic model matches that of a seasonal indicator model, ensuring identical fits. Including cycles with frequencies greater than the basic seasonal frequency $1/s$ may appear paradoxical. However, incorporating additional harmonics perturbs the standard sine wave, making the seasonal pattern more flexible and realistic. For instance, if monthly data is analyzed, introducing harmonics with frequencies of $2/12$ and $4/12$ can develop a more sophisticated seasonal pattern compared to a simple sine wave of period s . The afore described model is also known as **Harmonic Seasonal Model** (HMS) since it uses the harmonics to capture the seasonal pattern.

It's also noteworthy that low harmonic frequencies are significant for assessing traffic volume because they indicate the dominating, long-period cyclical patterns inherent in human activity, which fundamentally drive traffic flow [10]. These low frequencies correspond to significant, predictable variations such as the daily commute cycle (peaking twice, resulting in a primary 1 cycle/day component and its immediate harmonics), the weekly differences between weekday and weekend traffic (1 cycle/week), and even larger seasonal or annual trends (1 cycle/year). These long-period cycles often account for the biggest amplitudes or changes in traffic data, and so have the most "energy" or predictive power in the signal. Capturing these fundamental rhythms is crucial for accurate traffic modeling and forecasting, as higher frequencies tend to represent more localized, smaller-amplitude, or random fluctuations rather than the overarching structural patterns. This is the reason for restricting our analysis on the first 500 harmonic frequencies out of 2400 of the weekly 3-min distanced traffic volume data from 7 locations of Alexandras' Avenue in Athens on April 2000.

3.2 Adaptive LAD Lasso

The flexibility of harmonic modeling allows for a wide range of seasonal variations, making it a powerful alternative to simple seasonal indicator models. In practice, the order of harmonics and the polynomial trend are typically unknown. However, harmonic coefficients are statistically independent, allowing for the removal of non-significant terms. The selection of significant variables is often subjective, but a common approach is to let statistical variable selection procedures to filter-out the features with negligible statistical importance.

In this kind of problem, we usually penalize the estimation using the ℓ_1 -penalty. By leveraging the penalized ℓ_1 procedure, which is the method of the LASSO (Least Absolute Deviation – Least Absolute Shrinkage and Selection Operator), we can select the most significant features. In fact, Knight and Fu ([11]) proved that if $\lambda_n/n \rightarrow \lambda_0 \geq 0$, then the lasso estimate $\hat{\beta}^{(n)} = \arg \min_{\beta} \left\| \mathbf{y} - \sum_{j=1}^p x_j \beta_j \right\|^2 + \lambda_n \sum_{j=1}^p |\beta_j|$, converges in probability to the minimizer of the function $\mathcal{L}_1(\beta) = (\beta - \beta^*)^T \left(\frac{1}{n} X^\top X \right) (u - \beta^*) + \lambda_0 \sum_{j=1}^p |\beta_j|$, where λ_n is the parameter of LASSO estimation. In the same research publication, we can find the proof of the \sqrt{n} -consistency of the LASSO operator. Based on the last result, Hui Zou ([12]) stated that the convergence rate of $\hat{\beta}^{(n)}$ is slower than \sqrt{n} , and the limiting quantity is nonrandom. While the optimal estimation rate is achieved when $\lambda_n = O(\sqrt{n})$, it leads to inconsistent variable selection.

Hence, the crucial issue has been indirectly indicated questioning the existence of a way which ensures oracle properties for the LASSO-type estimators. This question answered by a proposed two-step procedure, where we first perform a simple regression (mean or median estimation) and,

afterwards, we apply LASSO where we divide the λ parameter of the procedure with the non-zero coordinated of the coefficient from the previous step [12], where the two-step procedure is named **Adaptive LASSO**. More specific, suppose that we set the convex minimization problem with ℓ_1 constraints dividing the λ parameter with the non-zero coordinated of a regression and

$$\begin{cases} \lambda_n/\sqrt{n} \rightarrow 0, \\ \lambda_n n^{(\gamma-1)/2} \rightarrow \infty, \gamma > 0, \\ \hat{\Sigma} = \frac{1}{n} X^\top X = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}, \text{ where } \hat{\Sigma}_{11} \text{ is a } p \times p \text{ matrix,} \end{cases}$$

Suppose also that \mathcal{A}_n^* the indices of the non-zero coefficients that LASSO indicates, and \mathcal{A} the indices of the real coefficients. Hence, under mild considerations, it is proven that $\lim_n P(\mathcal{A}_n^* = \mathcal{A}) = 1$ (consistency in variable selection), and $\sqrt{n}(\hat{\beta}_A^{(n)} - \beta_A^*) \rightarrow_d \mathcal{N}(0, \sigma^2 \times \hat{\Sigma}_{11}^{-1})$ (asymptotic normality). Alternatively, the initial coefficients of the adaptive form of the adaptive lasso are not constrained to have \sqrt{n} consistency for targeting the oracle properties of the adaptive scheme of the lasso, and the weights for predictors with zero coefficients become arbitrarily large, while the weights for nonzero coefficients converge to finite constants, as the sample size increases. This allows for simultaneous asymptotic unbiased estimation of both large coefficients and small threshold estimates. It's also noteworthy that the oracle properties are closely tied to the super-efficiency phenomenon ¹[14]. Last but not least, we can conclude that the adaptive lasso solution is continuous by definition which assures that the procedure is optimal. It's also On his study, Hui Zou shows also that the Breiman's Nonnegative garrote [15] can be considered as an adaptive lasso with ℓ_2 minimization and initial coefficients taken by minimizing the squared residuals with som constraints. The side effect of this proof is that the nonnegative garrote enjoys the oracle property which hasn't been reported since the adaptive LASSO's scheme proposal.²

In our investigation, the application of robust variable selection with the LASSO methodology appears to be a computationally costly strategy. To address this issue, we will compare the findings of most recent statistical approximations such as LSA and MTE, regarding the AIC and BIC minimization and the initial method of estimating coefficients (by minimizing squared or absolute deviations).

3.3 Least Square Approximation of Unified LASSO

The first, in chronological order, is the **Least Squares Approximation (LSA) of the Unified Lasso Estimation** [1]. Hansheng Wang and Chenlei Leng in 2007 introduced this methodology to integrate various LASSO estimation techniques. This innovative approach simplifies the application of LASSO across different regression models, including ordinary least squares (OLS) and quantile regression, by reformulating various LASSO objective functions into asymptotically equivalent least

¹The super-efficiency phenomenon ensures the boosting scheme of the efficiency on a two (or more) step procedure while the asymptotic variance of the first-step estimator is equal to the Cramer-Rao bound and the asymptotic variance of last estimation is strictly decreased. The phenomenon of superefficiency implies that asymptotically there exists no uniformly minimum-variance parameter estimator. [13]

²Let $\{\hat{\beta}_k\}$ represent the original ordinary least squares (OLS) estimates. Define $\{c_k\}$ as the set of values that minimize the objective function $\sum_k (y_n - \sum_k c_k \hat{\beta}_k x_{kn})^2$, subject to the constraints $c_k \geq 0$ and $\sum_k c_k \leq s$. The resulting predictor coefficients, $\tilde{\beta}_k(s) = c_k \hat{\beta}_k$, are the new estimates. As the parameter s is reduced, tightening the garrote, more of the coefficients c_k approach zero, while the remaining nonzero $\tilde{\beta}_k(s)$ are further shrunk. This process is referred to as the nonnegative garrote.

squares problems. A significant aspect of this methodology is its ability to ensure that the resulting LSA estimator achieves efficiency comparable to that of the oracle estimator, particularly when utilizing an adaptive LASSO penalty in conjunction with a Bayes/Schwarz information criterion-type tuning parameter. Notably, it is shown that regarding the LAD estimator consistency and asymptotic normality under specific conditions, thereby the LSA addresses the challenges posed by the non-smooth loss function effectively, and through the study of Wang-Leng [1], it's shown that the LSA estimator often outperforms traditional estimation methods, highlighting its robustness and effectiveness in real-data applications, particularly in scenarios demanding resilience against outlier effects. The case of the minimization of the absolute deviations (LAD) exemplifies the versatility of the LSA approach, as it adeptly manages the complexities associated with non-smooth loss functions while achieving the fundamental goals of variable selection and parameter estimation akin to conventional LASSO methodologies. Ultimately, the study suggests that the LSA method provides substantial benefits, especially within regression frameworks such as LAD.

The LSA method takes advantage on the Taylor Series expansion of the Least Squares (LS) or Absolute Deviations (LAD) loss function over the number of measurements, $\frac{\mathcal{L}_n(\beta)}{n}$, assuming that it has a continuous second-order derivative with respect to its vector variable at the unpenalized initial vector coefficient $\tilde{\beta} = \arg \min_{\beta \in \mathbb{R}^d} (\mathcal{L}_n(\beta))$ and the nullification of the first derivative of the

aforementioned loss function. Taking into account that $\mathbb{E} \left[\frac{\mathcal{L}_n(\beta)}{n} \right] \approx \Sigma^{-1}$ where Σ^{-1} is the inverse of the variance covariance matrix. Therefore, using the second order Taylor expansion explained above, we get the following squared Mahalanobis Distance based on the inverse of the covariance matrix:

$$\begin{aligned} \arg \min_{\beta \in \mathbb{R}^d} \left(\frac{\mathcal{L}_n(\beta)}{n} \right) &= \arg \min_{\beta \in \mathbb{R}^d} \left(\frac{\mathcal{L}_n(\tilde{\beta})}{n} + (\beta - \tilde{\beta})^\top \frac{\mathcal{L}'_n(\tilde{\beta})}{n} (\beta - \tilde{\beta}) \right) \\ &\approx \arg \min_{\beta \in \mathbb{R}^d} \left((\beta - \tilde{\beta})^\top \hat{\Sigma}^{-1} (\beta - \tilde{\beta}) \right) \end{aligned}$$

This is extremely intriguing in case of LAD-LASSO minimization, where the loss function is non-smooth, because we can apply the Least Squares Approximation if the sample covariance matrix $\hat{\Sigma}$ is not singular (namely, $\hat{\Sigma}^{-1}$ exists). This estimation can be done by simply leveraging the estimation of the Pollards asymptotic covariance matrix, $\hat{\Sigma}^{-1} \approx 4f_e^2(0)\text{cov}(x^S)$ where $S = \{j_1, \dots, j_d\}$, denote an arbitrary candidate model, which contains the predictors j_i, \dots, j_d th ($d > 1$) and $f_e(0)$ is the probability density function of the noise at zero value [16]. The LSA implementation[1] performs both LARS [17] and LASSO approaches, where we focus on the latter part. The method is not asymptotically efficient compared to the oracle when we use $\tilde{\beta}$, but it is shown that it is asymptotically efficient as the oracle if we properly fine-tune the parameters and the resultant coefficient vector is \sqrt{n} -consistent, as stated by Knight and Fu [11]. Wang and Leng recalled the \sqrt{n} -consistency of any vector of coefficients from LAD or LS LASSO (that contains the oracle estimator's coefficients). Moreover, they also showed that those two share the same semi-positive definite asymptotic covariance matrix, Ω_S . Particularly, the covariance matrix of the LS-LASSO problem is calculated to be $\Omega_S = \hat{\sigma}^{-2}\text{Cov}(\mathbf{x}_S)$, where $\hat{\sigma}^{-2}$ is the estimated variance of the errors, and the LAD-LASSO uses the implementation based on the Pollards' asymptotic estimation, which is $\Omega_S = 4f_e^2(0)\text{Cov}(\mathbf{x}_S)$. In addition, they prove that if the regularization applied to the relevant predictors diminishes to zero at a rate faster than \sqrt{n} , then the resulting LSA estimator remains \sqrt{n} -consistent. Afterwards, with probability approaching 1, all coefficients that are truly zero must be estimated as zero. the combination of the last two proven properties ensures that the proposed

LSA estimator can consistently recover the true model. Furthermore, the LSA estimator for the LASSO, denoted as $\beta_{\lambda,\alpha}$, where $\alpha = \max\{\lambda_j \mid j \leq d_0\}$ and $S_T = \{N_{\leq d_0}\}$ (the index set of true nonzero coefficients), follows the same asymptotic distribution as the oracle estimator. This implies that $\beta_{\lambda,\alpha}$ can achieve oracle efficiency.

For computational reasons, avoiding the generalized cross validation, on the adaptive scheme of the LSA, the method leverages the estimator for the degrees of freedom of Lasso estimators [18] and minimizes the BIC criterion, using the rule of thumb given by Zou [12] in the introduction publication of the adaptive lasso technique, $\lambda_j = \frac{\lambda_0}{|\tilde{\beta}_j|^\gamma}$, where γ is a prespecified positive real-valued number, and $\tilde{\beta}$ the unpenalized initial estimation, avoid paying the demanding computational cost of the exhaustive searching strategy in the Euclidean Space R^d . The question that is naturally arose brings up the heteroskedasticity existence on the scene. In case of the violation of homoskedasticity assumption, one could employ the Heteroscedasticity Consistent and Heteroscedasticity Autocorrelation Consistent estimators of the covariance matrix for a more estimation of the covariance matrix, based on the work of H.White[19], J.G.MacKinnon[20], F.Cribar- Neto [21], and W.B.D. Da Silva [22], keeping in mind that the last two estimators are useful in the small-sampled datasets, and Newey-West (NW-HAC, [23]) in case of autocorrelation persistence.

On the side of the code, the `lars.lsa` function uses a path-following approach to solve linear regression problems using either the least-angle regression (LARS) or the Least Absolute Shrinkage and Selection Operator (LASSO) technique. It is designed to deal with precomputed covariance matrices and coefficient estimations, and it can handle models with or without an intercept term. The function accepts a covariance matrix Σ_0 , a coefficient vector \mathbf{b}_0 , a Boolean flag `intercept` indicating whether to adjust for an intercept, the sample size n , a choice of algorithm type (either "lasso" or "lars"), a numerical precision threshold `eps` (defaulting to machine epsilon), and a user-specified maximum number of steps `max.steps`. If an intercept is supplied, the function applies a Schur complement-based modification to the covariance matrix to eliminate its contribution. The response vector \mathbf{b}_0 is also updated, and the appropriate scalar intercept estimate β_0 is initialized. If no intercept is specified, the function continues with the original inputs. The covariance matrix is symmetrically scaled using the absolute values of \mathbf{b} , and the sign vector of \mathbf{b} is saved for future calculations. The function starts an iterative loop, which progressively constructs the solution route. At each cycle, the current correlations between predictors and residuals are calculated. Variables are added to the active set when their highest absolute correlation exceeds a modest threshold. The Cholesky factor of the Gram matrix is updated via a rank-one update mechanism to efficiently incorporate new active predictors. If the final design matrix becomes rank-deficient, the related variables are temporarily removed from consideration. Otherwise, new active variables are added to the model with their corresponding signs recorded. The technique then solves a triangular problem to identify a viable step direction and step size, as well as the precise amount to advance along the piecewise-linear solution path while retaining the LAR or LASSO path's geometric features. In `lasso` mode, extra logic is used to check for coefficients that should be reduced exactly to zero. If such conditions occur, the related variables are removed from the active set, and the Cholesky factor is adjusted correspondingly. This ensures the sparsity of the solution at each level of the algorithm. At the end of iteration, the function reconstructs the entire coefficient pathways and adjusts them for the original scale and sign of \mathbf{b}_0 . If the model has an intercept, the bias factor β_0 is recalculated to match the final estimates. The function also calculates model selection criteria such as the residual sum of squares (RSS), the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC), which use the number of nonzero coefficients to estimate the model's degrees of freedom. The function provides a list that includes the matrix of regression coefficients

along the solution route, the intercept term (if any), and the AIC/BIC values for each step. This allows for detailed model selection and diagnostic assessment of the LAR or LASSO solution paths.

3.4 Penalized Maximum Tangent Likelihood Estimation

The second alternative method we employ is the **Maximum Tangent Likelihood estimation** (MTE) and its penalized version proposed by Y. Qin, Sh. Li, Y. Li, and Y. Yu [2]. This approach is presented 6 years later as a more reliable alternative to LSA, which frequently performs poorly in the presence of outliers. The penalized MTE seeks to combine the desirable features of robustness with the precision of high-dimensional statistical methods. The penalized MTE is \sqrt{n} -consistent and has the oracle property, indicating theoretical dependability. The framework includes several well-known methods as special instances, demonstrating its adaptability.

In particular, this family of estimators is shown to have an ideal rate of convergence ($\frac{\sqrt{\ln(d)}}{n}$), especially in circumstances where the number of variables, d , increases exponentially with sample size, n , and outperforms existing variable selection methods, such as Lasso and its variants on real-world data. In particular, the maximum tangent likelihood estimation is based on $\tilde{\beta} = \arg \max_{\beta \in \mathbb{R}^d} \left(\sum_{i=1}^n \ln_t(f(y_i - x_i^T \beta)) - n \sum_{j=1}^d p_{\lambda_j}(|\beta_j|) \right)$ where $\ln_t(f)$ is the p-th order Taylor expansion of, $\ln(u)\ln_t(u) = \mathbb{1}_{[u>t]}\ln(u) + \mathbb{1}_{[0 \leq u \leq t]}\ln(t) + \sum_{k=1}^p \frac{\partial^k \ln(v)}{\partial v^k} \Big|_{v=t} \frac{(u-t)^k}{k!}$ with $t > 0$ is the tuning parameter ³. Therefore, we get

$$\sum_{i=1}^n \frac{\partial}{\partial \beta} \ln_t(f(y_i - x_i^T \beta)) = \sum_{i=1}^n w_i \frac{\partial}{\partial \beta} \ln(f(y_i - x_i^T \beta))$$

where $w_i = \left(1 - \left(1 - \frac{f(y_i - x_i^T \beta)}{t}\right)^p\right)^{\delta_{[i \in \mathcal{A}^c]}}$ ⁴. Then, the simple MTE problem can be induced to the following minimization problem

$$\tilde{\beta} = \arg \max_{\beta \in \mathbb{R}^d} \left(\underbrace{\sum_{i \in \mathcal{A}} \ln(f(y_i - x_i^T \beta))}_{\text{minimizing KL divergence}} - \underbrace{\frac{1}{t} \sum_{i \in \mathcal{A}} f(y_i - x_i^T \beta) + n \sum_{j=1}^d \frac{\partial}{\partial \beta} p_{\lambda_j}(|\beta_j|)}_{\text{minimizing } \ell_2 \text{ metric}} \right)$$

where $\mathcal{A} = \{i \in \mathbb{N}_{\leq n} : f(y_i - x_i^T \beta) \geq t\}$ leveraging trimmed mean regression advantages on high-dimensional data ([24],[25]). The asymptotic properties of the MTE are the following:

- ☞ Probabilistic convergence of the coefficients to those values that maximize the expected mean value tangent of the natural logarithm of the Likelihood⁵ with respect to the true parameters β_0

$$\tilde{\beta} \xrightarrow[n \rightarrow \infty]{p} \beta_t^*, \text{ with } \beta_t^* = \arg \max_{\beta \in \mathcal{B}} \mathbb{E}_{\beta_0}(\ln_t(f(y - X^T \beta_t^*)))$$

³Note that $\lim_{t \rightarrow 0^+} \ln_t(u) = \ln(u)$ and for $t = 0$ we get the MLE

⁴It's easily notable that $w_i \xrightarrow{t \rightarrow 0} 0$

⁵and hence the Likelihood, since $\ln(\cdot)$ is a monotonic increase function

- The Central Limit Theorem under the distributional convergence occurs with covariance matrix to be decomposed as $\Omega = A^{-1}BA^{-1}$ where $A = \frac{\partial^2}{\partial \beta^2} \mathbb{E}_{\beta_0}[\ln_t(f(y - X^\top \beta_t^\star))]$ and $B = \mathbb{E}_{\beta_0} \left[\left(\frac{\partial}{\partial \beta} \ln_t(f(y - X^\top \beta_t^\star)) \right) \left(\frac{\partial}{\partial \beta} \ln_t(f(y - X^\top \beta_t^\star)) \right)^\top \right]$. The covariance matrix can be an approximation of the inverse Information Fisher as $t \rightarrow 0^+$ and then we get $\beta_t^\star \rightarrow \beta_0$. Namely,

$$\sqrt{n}\Omega^{-1/2}(\tilde{\beta} - \beta_t^\star) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, I_{d \times d})$$

- It worth to notice that in the previous property if $\beta_t^\star = \beta_0$ then the MTE is a consistent estimator with asymptotic normality, $\forall t > 0$. This occurs under specific regularity conditions for a linear dataset case (f.i. $y = X^\top \beta + e$ where e follows a symmetric distribution with zero mean).

Moving a step further, combining the MTE with a penalization scheme we can develop a technique that can also be used as a variable selector and has nice features. For sake of simplicity, let's define some notations and important assumptions:

- Let's note the true coefficients of the harmonic seasonal model as $\beta_0 = (\beta_{0i})_{i=1}^d$
- We set a partition of any coefficient vector $\beta = (\beta_S^\top, \beta_{S^c}^\top), \beta \in \mathbb{R}^d$, where we assume that $S = \{i \in \mathbb{N}_{\leq d} : \beta_{0i} \neq 0\} = \{i\}_{i=1}^s, |S| = s$ is the set of the indices of the nonzero coordinates of the coefficient vector. Note that $\beta_S \in \mathbb{R}^s$ whereas $\beta_{S^c} \in \mathbb{R}^{d-s}$ indicates the zero coordinates of the coefficient vector.
- We assume that d is fixed. This is given in our scenario since we compute the harmonics of the model based on the time seasonal pattern.
- We denote the maximization of the tangent of penalty term and the second derivative of the penalty term on the non-zero coordinates of the true coefficients as a_n and b_n respectively. Namely, $a_n = \max_{\beta_{0j} \neq 0} (p'_{\lambda_{nj}}(|\beta_{0j}|))$ and $b_n = \max_{\beta_{0j} \neq 0} (p''_{\lambda_{nj}}(|\beta_{0j}|))$

Then, the oracle property and the \sqrt{n} -consistency occur. In particular, for consistency, we get that if $a_n = O_p(1/\sqrt{n}), b_n = o_p(1) : \forall t > 0, \exists \hat{\beta} \in \mathbb{R}^d \|\hat{\beta} - \beta_0\|_2 = O_p(1/\sqrt{n})$ and for the oracle property, we know that if $\liminf_{n \rightarrow \infty} \liminf_{\theta \rightarrow 0+} \left\{ \min_{s+1 \leq j \leq d} p'_{\lambda_{nj}}(\theta)/\lambda_{nj} \right\} > 0$, and

$$\begin{cases} \max_{1 \leq j \leq s} (\sqrt{n}\lambda_{nj}) = o_p(1) \text{ and} \\ \frac{1}{\min_{s+1 \leq j \leq d} (\sqrt{n}\lambda_{nj})} = o_p(1). \end{cases}$$

Then sparsity of $\hat{\beta}_{S^c}$ and asymptotic normality $\hat{\beta}_S$ achieved. Namely, $\forall t > 0 : \hat{\beta}_{S^c} = 0$ with probability 1 (sparsity) and $\sqrt{n}(J_S + \Sigma_1) \left\{ \hat{\beta}_S - \beta_{0S} + (J_S + \Sigma_1)^{-1}\mathbf{b} \right\} \xrightarrow{d} N(0, \Sigma_2)$, where

$$\begin{cases} \Sigma_1 = \text{diag} \left((p''_{\lambda_{nj}}(|\beta_{0i}|))_{i \in \mathbb{N}_{\leq s}} \right) \\ \Sigma_2 = \text{Cov}[\partial \ln t(f(z; \beta_0))/\partial \beta_S] \\ J_S = \mathbb{E}[\partial^2 \ln t(f(z; \beta_0))/\partial \beta_S \partial \beta_S^\top] \\ \mathbf{b} = (p'_{\lambda_{ni}}(|\beta_{0i}|) \text{sgn}(\beta_{0i}))_{i=1}^s \end{cases}$$

Therefore, the asymptotic covariance matrix for $\hat{\beta}_S$ is calculated to be

$$\text{Var}(\hat{\beta}_S) = \frac{1}{n}(J_S + \Sigma_1)^{-1}\Sigma_2(J_S + \Sigma_1)^{-1}$$

The consistency of the penalized MTE scheme is achieved based on the restricted eigenvalue condition (REC)⁶ of the distance of the candidate coefficients, $\hat{\beta}$ from the true values, β_0 , $\Delta = \hat{\beta} - \beta_0$, over the constrain set $C(S) = \{\Delta \in \mathbb{R}^d : 3\|\Delta_S\|_1 \leq \|\Delta\|_{S^c}\}$, where Δ_S and Δ_{S^c} are the projections of Δ onto the coordinate sets S and S^c respectively. Given that our regressors are bounded (the regressors are sines and cosines of $2\pi \cdot \text{Time}/\text{frequency}$) and the design matrix, X , satisfies the REC, $\frac{\|X\Delta\|_2^2}{n} \leq \kappa_{REC}\|\Delta\|_2^2$ ($\kappa_{REC} > 0$), $\forall \Delta \in C(S)$ [25], then the loss function \mathcal{L}_λ satisfies restricted strongly convexity (SRC) in the neighborhood of the true coefficients and its gradient is bounded on the true coefficients, β_0 . The SRC assumption ensures that the loss function, \mathcal{L}_λ , contains enough curvature in the direction of $\Delta \in C(S)$ such that small value of $\mathcal{L}_\lambda(\tilde{\beta}) - \mathcal{L}_\lambda(\beta_0) = \mathcal{L}_\lambda(\tilde{\beta})$ translates to a small ℓ_2 norm $\|\Delta\|_2^2$. Moreover, it is proven ([27]) that we can force $\hat{\Delta} \in C(S)$ by choosing the regularization parameter of the ℓ_1 penalization to be $\lambda_n = 2\xi\sqrt{\frac{\ln(d)}{n}}$. Then, a sequence of publications ([28], [29], [30]) presented that for $\lambda = \left(\frac{\ln(d)}{n|\beta_j|}\right)_{j=1}^d$ can minimize BIC for lasso implementation in case of scaled design-matrix to be $\|X\|_2 = n$. Unfortunately, we need to avoid scaling because this will lead us to unstable rationale against the time dependency of the data. Therefore, we end up tuning the hyperparameter greedily in the domain defined by the following partition $\mathcal{D}_\lambda \subset (0, 0.1] : 50\mathcal{D}_\lambda = \mathbb{N}_{\leq 5}$. The explanation for this approach is that larger-valued lambda will easily unintentionally cancel the harmonic frequency coefficients since the regressors' values are between $[-1, 1]$ and hence close to zero. Moreover, we can understand that the low values of the regressors can cause huge values of the coefficients, since the target values can reach upto in 0-162 vehicles. Therefore, by

The aforementioned properties and problem reductions coupled with the coordinate descent optimization algorithm accelerates the speed of process regarding the conventional implementation of ℓ_1 -Lasso and the Least Square Approximation of the unified Lasso that we mentioned before. For the model-wise comparison we are based on the minimization of the Mean Absolute Error (MAE) or Square root of Mean Squared Error (RMSE). When evaluating predictive accuracy, MAE and RMSE are preferred over AIC because they measure the difference between predicted and actual values in the same units as the dependent variable. While MAE treats all errors equally, RMSE "penalizes" larger errors more severely, making it appropriate when outliers are an issue. In contrast, AIC focus on model selection by balancing goodness of fit with complexity, penalizing excessive parameters to prevent overfitting. However, a model with the lowest AIC does not necessarily yield the best predictions. Therefore, when comparing models for predictive performance, MAE or RMSE should be used, whereas AIC is more appropriate for selecting a statistically optimal model with a balance between fit and complexity. This choice is made regarding the suggestion of

⁶The Restricted Eigenvalue (RE) condition assures that the squared ℓ_2 -error of a Lasso estimate reaches the ideal minimax rate. The design matrix X must meet specified parameters for sustained and accurate sparse recovery in high-dimensional regression. Traditionally, this condition exists when X meets the Restricted Isometry Property (RIP). This is frequently assured when X contains i.i.d. entries or unitary row structure. This study demonstrates that the RE condition applies with high probability even for Gaussian matrices with dependent predictors [26], expanding the applicability of ℓ_1 -based approaches beyond the conventional assumptions. Recall that if we assume that A is an $m \times p$ matrix and let $1 < s < p$ be an integer. Suppose that there exists a constant $\delta_s \in (0, 1)$ such that, for every $m \times s$ submatrix A_s of A and for every s -dimensional vector y , $(1 - \delta_s)\|y\|_2^2 \leq \|A_s y\|_2^2 \leq (1 + \delta_s)\|y\|_2^2$. Then, the matrix A is said to satisfy the s -restricted isometry property with restricted isometry constant δ_s .

Burham and Anderson on 2006 [31]. On the side of code, the `MTElasso` function uses *Maximum Tangent Likelihood Estimation* (MTE) to create a robust and sparse linear regression framework. This method combines the resilience of heavy-tailed error distributions with ℓ_1 -norm penalization to promote sparsity in the predicted regression coefficients. The MTE approach substitutes the traditional log-likelihood function with a tangent approximation, which reduces the impact of tiny likelihood contributions, which are usually outliers or heavy-tailed residuals. The function starts with two inputs: a predictor matrix $X \in \mathbb{R}^{n \times d}$ and a response vector $y \in \mathbb{R}^n$. If an initial coefficient vector is not supplied, it is calculated via penalized quantile regression. When the intercept flag is set, the design matrix is expanded with a column of ones, and the penalty factor is changed to exclude the intercept term from penalization. The `penalty.factor` vector supports varied penalization across variables and defaults to uniform weights. The approach has a tuning parameter $t > 0$ that sets the threshold below which the log-likelihood is substituted with a piecewise Taylor (polynomial) approximation. The approximation order is determined by a discrete parameter $p \in \mathbb{N}$, which determines the smoothness and accuracy of the likelihood surrogate in the tails. The approach is based on an iterative coordinate descent scheme. At each iteration, the current residuals are utilized to evaluate the standard normal density, which serves as the foundation for likelihood contributions. The log-likelihood does not change for residuals with densities higher than the threshold t . However, for lesser values, the tangent approximation replaces the log-likelihood with a lower-degree Taylor polynomial, retaining continuity and reducing the impact of extreme values. As a result, the objective function is smoothed, allowing for efficient differentiation and optimization. Each coefficient is updated by solving a univariate optimization problem while keeping the other coefficients fixed. The objective function reduced for each coordinate consists of a tangent likelihood component and a ℓ_1 penalty term. When the adaptive penalty is activated, the penalty weight for each coefficient is inversely proportional to its size, decreasing bias on big coefficients and enhancing variable selection consistency. If the adaptive option is disabled, the user must provide a fixed penalty parameter manually. The optimization process continues until the largest change in the coefficient vector is less than a preset threshold (defaulting to 10^{-4}) or a maximum of 200 iterations are achieved. Each iteration prunes coefficients with absolute values less than 10^{-5} and excludes the relevant variables from subsequent updates, quickening convergence and dynamically ensuring sparsity. Once completed, the function reconstructs the whole coefficient vector, restoring zeros for removed variables. If an intercept was included, it is labeled appropriately, and the remaining coefficients are given names based on the predictors. The function returns both the final estimated coefficients and the fitted values, which are generated via matrix multiplication with the design matrix.

3.5 Maximum Entropy Bootstrap

For sake of completeness, in order to make the comparison of the aforementioned methods to be proper, we need to focus on the sampling distribution of the time series data. Bootstrap techniques are key statistical tools for estimating a statistic's sampling distribution when the theoretical distribution is unknown or difficult to calculate, particularly with small samples or complicated data structures. The **Maximum Entropy Bootstrap (MEBoot)** applies this utility to time-series data, where existing bootstrap approaches frequently fail owing to temporal dependency and non-stationarity. MEBoot provides a principled, nonparametric resampling strategy that produces smooth, continuous-valued samples while preserving the time-order and dependency structure of the original data. This is accomplished by using maximum entropy densities, which generate the

least biased distribution that is compatible with observable restrictions such as the empirical mean and order statistics. As a result, ME Bootstrap is very well suited to our objective.

Vinod ([32], [33], and [3]) invented the MEBoot, based on the Wiener-Khinchine-Khinchine (WKK) process which is a pipeline arose from their authors analysis ([34], [35], and [36])⁷. Beside that WKK is based on zero memory $I(0)$ white noise type processes with circular events, the ME-Boot method, offers considerable flexibility for time series analysis. There are some limitations though. One notable limitation is its sensitivity to non-stationarity. Although the method does not strictly require stationarity in the traditional sense, it performs optimally when the underlying data-generating process exhibits stable statistical properties over time. Applying the ME bootstrap to non-stationary series may yield misleading results and distort meaningful temporal relationships, particularly if transformations to induce stationarity compromise interpretability. Another important constraint lies in the method's reliance on accurate model specification. If the model employed does not adequately reflect the data structure, especially in the presence of temporal dependence, the resulting inference can be significantly flawed. Moreover, the ME bootstrap may encounter difficulties in handling time series characterized by structural breaks or abrupt regime shifts, which can obscure the method's flexibility and complicate the interpretation of outcomes. On the other hand, the ME bootstrap assumes that the underlying process does not possess infinite memory, thereby enabling the satisfaction of both the ergodic theorem and the central limit theorem, which are fundamental for valid inference. A key procedural requirement involves the use of the empirical cumulative distribution function (CDF), from which the method constructs its sampling ensemble. Consequently, the observed data must be sufficiently rich and representative to ensure that the empirical CDF captures the essential features of the true distribution. Unlike standard bootstrap approaches, which may overlook time dependence structures, the MEBoot retains such temporal dynamics by employing a distribution with maximum entropy, a concept that provides the least biased estimate consistent with existing data:

1. (Ranking) Sorting the observed time series data x_t in ascending order yields the order statistics $x_{(t)}$. Create an index vector with the original time locations of these sorted data. This vector represents the temporal structure of the original series.
2. (Interval Boundaries) Compute the midpoint of each pair of consecutive order statistics ($z_t = \frac{x_{(t)} + x_{(t+1)}}{2}$, $t = 1, \dots, T - 1$). These midpoints mark the bounds of intervals that resemble the empirical distribution.
3. (Extending Tails) To determine the outermost intervals (beyond the smallest and largest data points), compute the trimmed mean m_{trm} of the sequential differences $x_t - x_{t-1}$. To expand the domain, define: $z_0 = x_{(1)} - m_{\text{trm}}$, $z_T = x_{(T)} + m_{\text{trm}}$. These represent the left and right limits for the lowest and highest intervals, respectively.
4. (Assigning Interval Means - Maximum Entropy, ME) Within each period $[z_{t-1}, z_t]$, assign a number m_t indicating the mean of the highest entropy density confined to that interval. This assignment guarantees that the predicted value of the created sample corresponds to the mean of the observed data. Because of their uneven support, the first and last intervals' means are determined using simpler formulae.

⁷The WKK method is based on the Wiener-Khinchine (it is also known as Wiener-Kolmogorov-Einstein and Wiener-Kolmogorov) which states that the autocovariance is the Fourier transformation of the absolute value of the square of the probability density function. Its proof is presented in the Appendix.

5. (Sampling from ME Distribution) Create uniform random variables $u_i \sim \mathcal{U}(0, 1)$ and turn them into sample quantiles using the maximum entropy distribution. The quantiles are then sorted in increasing order.
6. (Restoring Order) Using the index vector from Step 1, rearrange the sorted quantiles in the same temporal order as the original data. This step assures that the created sample has the same time-dependent structure as the original series.
7. Repeat.

The sorting stage gives the dataset a systematic order, which helps to preserve the links between data points. This structure is critical for appropriately representing the underlying data distribution in later analysis. Defining intervals for the bootstrap technique allows a more exact estimation of the distribution. Afterwards, creating these bounds, the algorithm accommodates for extreme observations while also ensuring that the resampled data are realistic and typical of the original dataset, particularly in terms of central tendency and dispersion. Then, we enforce adherence to the maximum entropy principle, which is required to ensure the statistical integrity of bootstrap samples. This aids in producing samples that correctly reflect the variability inherent in the original data. Then, randomization is introduced via quantile-based sampling. This is crucial for creating bootstrap samples that maintain the original dataset's statistical features while accounting for variability. The use of quantiles preserves distributional properties like skewness and kurtosis. Next, reorder the collected data to restore any temporal dependencies that existed in the original time series. This is critical in time series analysis since maintaining the structural integrity of time-dependent interactions is required. Finally, repeating the bootstrap technique increases the robustness of the inference. Multiple resamples allow for the assessment of variability and uncertainty in estimates, resulting in a more thorough and solid basis for statistical findings.

Respecting the requirements of the method and paying attention of its limitations, we used the Maximum Entropy Bootstrap via `mboot` R-library performing 20 repetitions based on the whole month of the whole dataset. This method provide us the statistics of the important harmonic frequencies of the traffic volume of our dataset. The vector of inclusion probability of each harmonic in the predicting model constructs the weight vector of the final Adaptive Lasso model.

3.6 Hyperparameter tuning

In the conventional lasso estimation, we need to tune the hyperparameter λ , where this issue also appears on the maximum tangent likelihood estimate. For the hyperparameter tuning, we are based on the **Akaike Information Criterion (AIC)** and **Schwarz/Bayesian Information Criterion (BIC)** minimization in both scenarios of least squares and least absolute deviations minimization methods. The application of those two criteria aims finding the best model that describes the data generating mechanism and the data generating mechanism itself ⁸ as it is proved by Burham and Anderson [37]. This occurs because the formula of AIC and BIC is $2k - \log(L)$ and $k\log(n) - \log(L)$ in correspondence, where k is the degrees of freedom, n is the number of samples of the dataset and L is the likelihood of the model. As we can easily spot, the BIC has heavier penalty than AIC. especially in large datasets. The BIC frequently produces smoother predictions than the AIC due to a fundamental difference in how each criteria balances goodness-of-fit versus model complexity. BIC is based on a Bayesian framework and posits that one of the candidate models represents the

⁸assuming that the set of candidate models includes the data generating mechanism

actual data-generating process. As a result, it imposes a higher penalty for each new parameter, particularly as the sample size grows. This harsher penalization discourages the use of too complicated models, which may fit the data well in-sample but perform badly out-of-sample. AIC, on the other hand, is based on information theory and operates on a more pragmatic assumption: no candidate model is completely right, but some may better resemble the unknown data-generating process. AIC is more tolerant of complexity if it increases the model's capacity to capture the data's underlying structure. This frequently prompts AIC to prefer more flexible models, which might result in predictions that are more variable or "wiggly," especially in the presence of noise. BIC's smoother forecasts reflect its preference for simpler, more parsimonious models, which are usually more stable and less sensitive to data volatility. This conservative tendency is frequently useful when interpretability and generalization are important, but it can lead to underfitting if the real process is genuinely complicated. In practice, neither criteria is clearly superior. AIC may be selected when the objective is to resemble a complicated system as closely as feasible, but BIC is frequently preferred when the danger of overfitting is more significant. Many academics believe that contacting both is beneficial, and that any disagreement should be seen as a signal to further analyze the models. Of course, we need to pay attention avoiding the misusage of AIC. According to Burham and Anderson [31].⁹

3.7 Metrics Used and Model Comparison

In this study, a broad range of out-of-sample measures, including Huber Loss, RMSE, MAE, MAPE, and MdAPE, is leveraged to evaluate and compare the approaches used, resulting in a more thorough and rigorous performance assessment. Each of these measures identifies various areas of model inaccuracy, providing complementing information. RMSE is sensitive to big variances and emphasizes punishing considerable prediction mistakes, which is important when such errors have a high cost. MAE, on the other hand, is a simple measure of average error that does not overemphasize outliers, making it more interpretable and resilient. Huber Loss provides a hybrid technique that combines the sensitivity of RMSE in the center with the robustness of MAE for greater deviations, which is particularly useful in the presence of outliers. MAPE and MdAPE quantify relative error, allowing model performance to be evaluated in percentage terms, which is useful for comparing variables or datasets of varying dimensions. MdAPE, in particular, offers resilience to skewed distributions and excessive percentage errors. Using this collection of measures allows practitioners to get a more comprehensive knowledge of model behavior, minimize overreliance on a single measure, and make fairer, more generalizable comparisons between models and datasets. Let's take a more clear look into these metrics, so we can clarify the significance of their impact to our study.

The **Huber loss** is a loss function commonly used in robust regression to limit the influence of outliers. It combines the properties of the squared error loss and the absolute error loss, behaving quadratically near the origin and linearly for large residuals. This makes it less sensitive to outliers than the pure squared loss, while still retaining sensitivity to small errors. Originally introduced

⁹This means that we don't use it to compare nested models, the formula is modified in case we want to get compatible criterion with robust models, the maximization of the likelihood is assumed. We also know that AIC is not used for estimating the forecast accuracy, its minimization is asymptotically equivalent to minimizing the leave-one-out cross-validation MSE for cross-sectional data, and equivalent to minimizing the out-of-sample one-step forecast MSE for time series models and its constant term is eliminated. Of course, as it has already been mentioned above, the AIC isn't used to catch the true generating mechanism.

by Huber ([38]), the loss is defined piecewise as:

$$L_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - f(x))^2 = \frac{1}{2}\varepsilon^2 & \text{if } |y - f(x)| \leq \delta, \\ \delta(|y - f(x)| - \frac{1}{2}\delta) = \delta(|\varepsilon| - \frac{1}{2}\delta) & \text{otherwise,} \end{cases}$$

where $\varepsilon = y - \hat{y}$ the residuals. The function smoothly transitions between the quadratic and linear regimes at the threshold $|\varepsilon| = \delta$, maintaining continuity and a matching derivative at the join. Geometrically, the Huber loss can be interpreted as a smoothed version of the absolute error loss. Specifically, it is the convolution of the absolute value function with a rectangular (uniform) kernel, scaled and shifted appropriately. This smoothing effect eliminates the sharp corner at the origin present in the absolute loss function, which improves the behavior of gradient-based optimization algorithms. Two widely used loss functions in statistical estimation are the squared loss and the absolute loss, defined respectively as:

$$L(\varepsilon) = \varepsilon^2 \quad \text{and} \quad L(\varepsilon) = |\varepsilon|.$$

The squared loss leads to estimators that are mean-unbiased, meaning the arithmetic mean minimizes the expected loss. On the other hand, the absolute loss yields median-unbiased estimators in one dimension, and geometric median-unbiased estimators in higher dimensions. However, a key drawback of the squared loss is its high sensitivity to outliers. When aggregating over a sample,

$$\sum_{i=1}^n L(\varepsilon_i) = \sum_{i=1}^n \varepsilon_i^2,$$

a few large values of ε_i can disproportionately affect the result—especially when the underlying distribution has heavy tails. In such settings, the asymptotic relative efficiency of the mean becomes low, making the squared loss less suitable for robust estimation. These will be discussed further below.

The Huber loss function, as previously defined, offers a compromise between these two extremes. It is strongly convex in a uniform neighborhood around the origin $\varepsilon = 0$, which provides desirable stability properties for optimization. At the boundary points $\varepsilon = -\delta$ and $\varepsilon = \delta$, the loss transitions smoothly into a linear function, preserving differentiability. These structural features allow the Huber loss to retain the efficiency and sensitivity of the mean-based estimator under normal conditions, while gaining the robustness characteristics of the median-based estimator in the presence of outliers. In this way, the Huber loss function effectively balances efficiency and robustness in statistical modeling.

The **Mean Absolute Error (MAE)** is a statistical metric that calculates the average size of errors between paired observations of the same phenomena. Pairings can arise in circumstances like projected vs. actual values, measurements across time, or various measurement procedures. The formal definition is as follows:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t| = \frac{1}{T} \sum_{t=1}^T |\varepsilon_t|.$$

In certain forms, weights based on relative frequency may be used. Because MAE is assessed in the same units as the data, it is a scale-dependent accuracy metric and hence inappropriate for

comparing predictions across datasets of varying sizes. MAE is extensively used in time series forecasting and is occasionally mistaken with Mean Absolute Deviation (MAD); while similar in form, the two are fundamentally distinct. Although MAE is sometimes compared with or mistaken for the *Root Mean Square Error (RMSE)*, the two differ in interpretation and sensitivity. While MAE is a linear average of absolute errors, RMSE squares the differences, amplifying the influence of larger errors. MAE is conceptually simpler and represents the average absolute distance from the identity line $y = \hat{y}$ in a scatter plot. MAE is sometimes compared to or confused for the *Root Mean Square Error (RMSE)*, however the two differ in interpretation and sensitivity. While MAE is a linear average of absolute mistakes, RMSE squares the differences to increase the impact of greater errors. MAE indicates the average absolute distance from the identity line $y = \hat{y}$ in a scatter plot.

The **Root Mean Square Error (RMSE)** (also known as the Root Mean Square Deviation, RMSD), is a commonly used statistical metric that measures the discrepancies between values predicted by a model or estimator and actual values observed. These values might relate to either genuine observations or reference estimators. RMSE is defined as the square root of the mean of squared deviations, where the deviations might be scalar differences or vector magnitudes (e.g., the root mean square deviation of atomic locations in bioinformatics). It generates a single aggregated measure of predictive accuracy by integrating the magnitudes of individual prediction mistakes.

The RMSE is calculated as $\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2}$, using observed values y_t and expected values \hat{y}_t over T time points for cross-sectional data (the subscript t is typically replaced by i , and T by n). In practice, the RMSE is non-negative and only achieves zero in the ideal (and uncommon) scenario in which forecasts exactly match observations. A lower RMSE suggests improved prediction ability. However, because it is dependent on the magnitude of the data, RMSE should not be used to evaluate performance across datasets measured in various units. When the RMSE is done on the same dataset that was used to fit the model, the results are referred to as residuals. When computed with fresh or external data, the variances are referred to as errors or prediction mistakes. Despite this discrepancy, the mathematical definition is the same. Because RMSE entails squaring each variation before averaging, higher mistakes have a greater impact on the final result. This renders RMSE more vulnerable to outliers. If \hat{y}_t is an unbiased estimator of y_t , then the RMSE simplifies to the standard deviation of the estimator.

The **Mean Absolute Percentage Error (MAPE)** (also known as Mean Absolute Percentage Deviation, MAPD), is a commonly used metric for evaluating forecasting model performance in statistical and machine learning applications. It measures the prediction error as a percentage and is defined by the following formula: $\text{MAPE} = 100 \cdot \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$, where n is the number of prediction occurrences. The formula calculates the absolute percentage error for each observation, averages the results, and presents the result as a percentage. MAPE is especially recognized for its easy interpretation, as it displays the average relative error in percentage terms. This makes it a common choice for assessing predicting accuracy and acting as an effective loss function in regression problems. The purpose of the MAPE framework is to reduce the predicted absolute percentage error. $g_{\text{MAPE}}(x) = \arg \min_{g \in \mathcal{G}} \mathbb{E} \left[\left| \frac{g(X) - Y}{Y} \right| \mid X = x \right]$, Empirical risk minimization can be used to estimate the optimum model g_{MAPE} . The relevant empirical optimization issue is: $\hat{g}_{\text{MAPE}}(x) = \arg \min_{g \in \mathcal{G}} \sum_{i=1}^n \left| \frac{g(X_i) - Y_i}{Y_i} \right|$. This may be restated as a weighted mean absolute error (MAE) regression problem, with each sample's weight being inversely proportional to the magnitude

of the response variable:

$$\hat{g}_{MAPE}(x) = \arg \min_{g \in \mathcal{G}} \sum_{i=1}^n \omega(Y_i) |g(X_i) - Y_i|; \quad \text{with } \omega(Y_i) = \left| \frac{1}{Y_i} \right|.$$

Thus, MAPE regression may be efficiently applied with tools designed for quantile regression with sample-dependent weights. From a theoretical and practical viewpoint, the usage of MAPE as a loss function in regression analysis makes sense. Under the right conditions, one may prove the existence of an optimum prediction function and demonstrate the coherence of the empirical risk reduction strategy. This theoretical soundness, together with its practical applicability, encourages the usage of MAPE in a variety of modeling settings.

The same rationale is used in **Median Absolute Percentage Error (MdAPE)** where the mean calculation is replaced by the estimation of the median, $MdAPE = Median\left(100 \cdot \left| \frac{y_t - \hat{y}_t}{y_t} \right| \right)$, where y_t is the actual value at time t , \hat{y}_t is the forecast value, and n is the number of observations. The MdAPE mitigates the impact of outliers by substituting the mean with the median. The MdAPE, which uses the median, provides a more robust measure of usual prediction accuracy since it is less impacted by skewed error distributions against outliers and noise presence. It is worth noting that both measures assume non-zero real values.

The MAPE, despite its persistent popularity in practice, possesses several well-documented deficiencies that undermine its reliability as an error metric. Its intended purpose is to produce a scale-free, easily interpretable indicator of forecasting accuracy. However, this formulation introduces significant issues. Firstly, MAPE is sensitive to the scale of the data: in high-volume contexts, the large denominator tends to suppress error magnitude, potentially disguising poor model performance. Conversely, when actual values are small, even minor deviations inflate the error disproportionately, potentially exceeding several hundred percent. Secondly, MAPE exhibits an inherent asymmetry—it tends to favor underforecasting over overforecasting, a bias previously highlighted by Fildes [39]. It's worthy to note also that MAPE is criticized for its asymmetry ([40], [41], [42]), particularly because it penalizes overestimation ($\hat{y}_t > y_t$) more heavily than underestimation by the same amount. This bias arises from the presence of the actual value y_t in the denominator, making errors disproportionately large when actuals are small. In response to this issue, Armstrong [40] proposed the **adjusted MAPE**, $\overline{MAPE} = Mean\left(100 \cdot \frac{2|y_t - \hat{y}_t|}{(y_t + \hat{y}_t)}\right)$ as a asymmetric alternatives of MAPE and seven years later, Makridakis [42] introduced **Symmetric Mean Absolute Percentage Error (sMAPE)**, $sMAPE = Mean\left(100 \cdot \frac{2|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}\right)$. ¹⁰ However, these formulations suffer from inconsistency, ambiguous definitions, and incorrect assumptions about their range. Both symmetric variations of adjusted and symmetric MAPE sometimes claim a bounded range (e.g., 0–200), which is not always accurate due to the possibility of the denominator approaching zero. Then there was an improvement by Goodwin and Lawton ([43]) and the final version proposed by Chen and Yang ([44]), $sMAPE = 100 \cdot \frac{1}{n} \sum_{t=1}^n \frac{2|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}$, is considered the most robust, as it handles zero and negative values more appropriately and is bounded between 0 and 200. Nonetheless, the symmetry debate remains contentious. Some authors argue that MAPE is already symmetric with respect to relative percentage deviations, while others note that many sMAPE formulations lack this property. Importantly, in practical applications like the M3 forecasting competition, the precise definition of sMAPE becomes crucial, especially given the presence

¹⁰These often modify the denominator to include both actual and forecast values, for example $y_t + \hat{y}_t$ or $|y_t| + |\hat{y}_t|$, leading to several variants known as adjusted MAPE or sMAPE

of negative forecasts despite positive actuals. Replicability issues in M3 suggest a lack of clarity regarding the specific metric definition employed. Ultimately, the use of MAPE or sMAPE should be contingent on the data characteristics and application context. Thirdly, MAPE cannot be reliably computed for intermittent demand where actual values may be zero, resulting in undefined or infinite errors. While its simplicity makes it intuitively appealing, MAPE does not provide a meaningful reflection of model quality due to these distortions and lacks a well-defined optimization criterion. The root of these problems lies in the division of absolute forecast errors by actual values. To address this, Hyndman and Koehler ([45]) introduced the **Mean Absolute Scaled Error (MASE)**, which replaces the denominator with the in-sample mean absolute differences, thereby removing scale dependency and asymmetry. Although MASE resolves many limitations of MAPE, it is less interpretable and suboptimal for intermittent demand, as it is minimized by the median. A more suitable alternative for such cases is the **Root Mean Squared Scaled Error (RMSSE)**, as proposed by Makridakis et al. ([46]), which follows a similar structure to MASE but utilizes the RMSE and in-sample squared differences. Although still not easily interpretable, RMSSE mitigates several key shortcomings. For those prioritizing interpretability, Petropoulos and Kourentzes ([47]) suggest dividing MAE or RMSE by the in-sample mean, which offers partial mitigation of the scale issue and improves upon other deficiencies. Ultimately, while percentage-based error measures like MAPE remain widely used, their inherent flaws necessitate careful consideration and, where appropriate, substitution with more robust alternatives. For the reasons explained above, we are going to focus our comparative analysis on the metrics of RMSE, MAE and Huber Loss (as an balanced metric between the robustness of MAE and sensitivity of RMSE), and then we are going to mention the MAPE, MdAPE and the pair of metrics sMAPE and sMdAPE for sake of completeness.

The process starts with the original dataset, which is then resampled using the Maximum Entropy Bootstrap (MEB) technique. This technique creates 20 synthetic datasets $MEB\ SD_1, SD_2, \dots, SD_K$ that retain the dependence structure and temporal properties of the original series, allowing for robust inference under uncertainty. Each synthetic dataset is then modeled using a *Harmonic Seasonal Model* (HSM), which captures periodic variations inherent in traffic data by estimating coefficients of harmonic basis functions using either ℓ_1 -norm (LAD) or ℓ_2 -norm (LS) coefficients estimation. These models extract the historical seasonal trends found in each resampled dataset. Next, *Least Squares Approximation* (LSA) is used to refine the estimated coefficients, utilizing ℓ_1 and ℓ_2 norms. Model selection is guided by minimizing the *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC), ensuring a balance between model fit and complexity. In parallel, each bootstrapped dataset undergoes *Penalized Maximum Tangent Likelihood Estimation* (MTE), a robust penalized approach intended to handle outliers and heavy-tailed distributions. To improve model performance, a greedy Hyperparameter Tuning (HPT) is performed across the domain $\mathcal{D} = (0, 0.1]$ with a step size of 0.02 to tweak the penalty parameters. The LSA and MTE outputs are used to construct variable inclusion probabilities (VIPs), which are aggregated into important weights representing the stability and significance of each variable throughout the ensemble of models. These weights are then included into a *Adaptive ℓ_1 -Lasso* regression framework, which implements a 10-fold *Bidaily Leave-One-Block-Out Cross-Validation* (LOBOCV) for each model performed separately. This validation structure preserves temporal coherence by separating two-day blocks, which improves the model's generalizability to previously unexplored time segments. Within this Lasso phase, a supplementary HPO step is applied across $\mathcal{D} = [0, 0.1]$ with a step size of 0.02, to fine-tune the regularization strength. This pipeline produces a traffic forecasting model that is both statistically and operationally robust. It combines harmonic decomposition, regularized estimates, resampling-based uncertainty quantification, and rigorous model selection to provide a

complete solution for time series modeling in complicated urban contexts. The whole pipeline is depicted in Figure 3 below.

The aforementioned analysis is written in R statistical software. It utilizes a comprehensive suite of libraries, each serving specific roles in data manipulation, statistical modeling, time series analysis, parallel computation, and web-based visualization. The `collapse`, `dplyr`, `reshape2`, and `magrittr` libraries are primarily employed for efficient data manipulation, transformation, and pipeline syntax, with `collapse` offering particularly fast grouped statistics for large datasets. For time series handling and visualization, `tseries`, `xts`, `lubridate`, and `ggfortify` provide functions for testing stationarity, managing time-indexed data, simplifying date-time operations, and plotting time series with `ggplot2`, respectively. Statistical modeling and diagnostics are supported by `lmtest` and `sandwich`, which facilitate hypothesis testing and robust standard error estimation, while `jtools` enhances the interpretation and visualization of regression results. Additionally, `quantreg` and `rqPen` is used for fitting penalized and unpenalized quantile regression models, useful for robust predictive modeling, `MTE` is used for the penalized maximum tangent likelihood estimation, and `meboot` library provides us the Maximum Entropy Bootstrap technique. The inclusion of `foreach`, `doParallel`, `future`, `furrr`, and `purrr` enables scalable and parallelized computation across multiple cores or asynchronous workers, streamlining iterative tasks or simulation procedures. The script also incorporates interactive data visualization and web application development using `shiny`, `shinydashboard`, and `shinyWidgets`, which together allow the creation of responsive dashboards. `plotly` and `leaflet`, along with `leaflet.extras`, further enhance interactivity by enabling dynamic plotting and geospatial visualizations. The `DT` package renders interactive tables, often essential in user-facing dashboards. Aesthetic customization is facilitated by `Polychrome`, which ensures the use of visually distinguishable color palettes for complex plots. Lastly, the Least Square approximation of the adaptive LASSO is integrated externally the R-script provided by the authors of the corresponding publication ([1]).

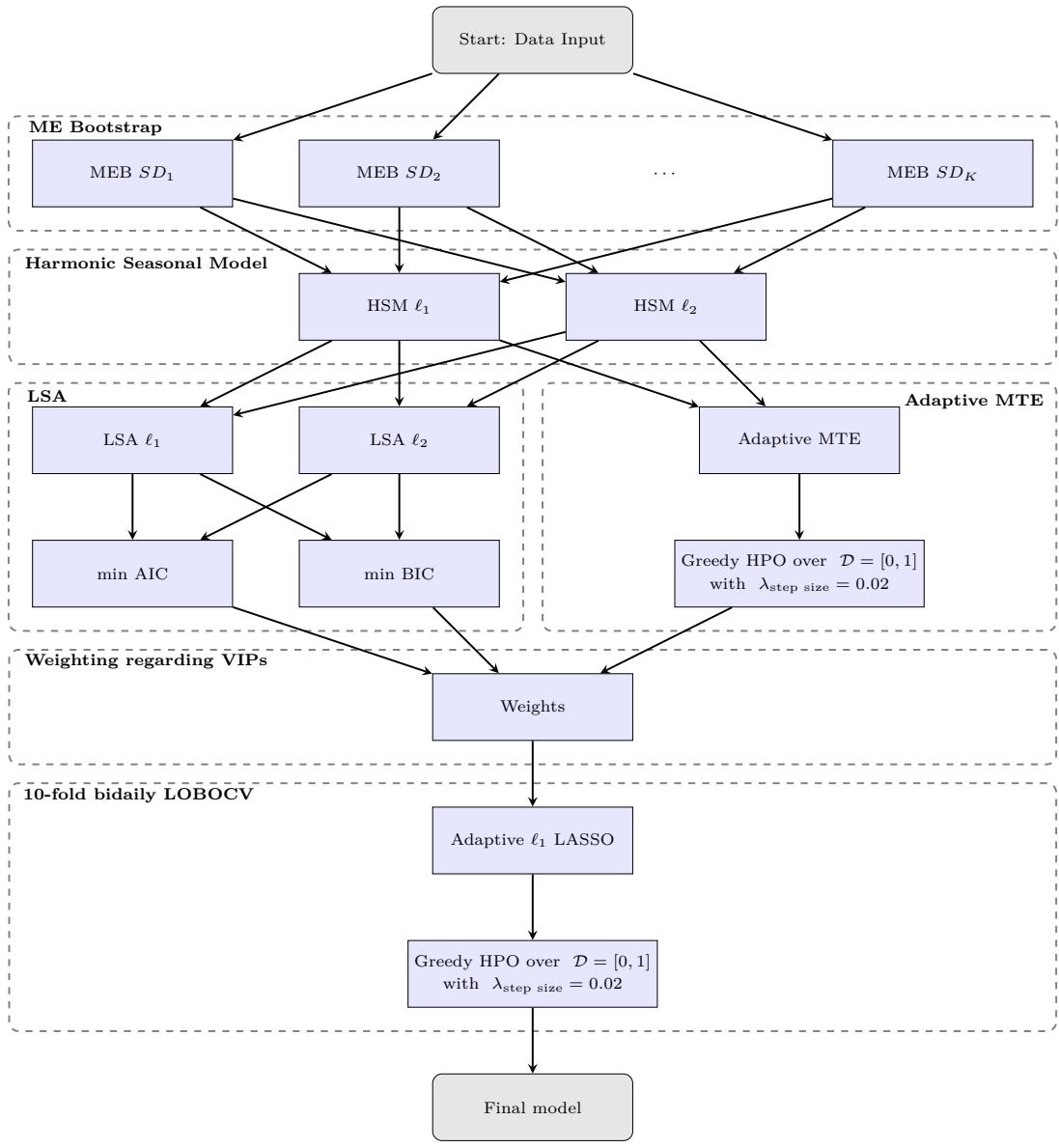


Figure 3: The pipeline we are following in our comparative study.

4 Results

This section presents a comparative analysis of models within the family $\{\ell_x M_i | x \in \{1, 2\}, M \in \{\text{MTE, LSA}\}, i \in \{\text{AIC, BIC}\}\}$. Here, ℓ_x specifies the norm used for the loss function and/or penalty (e.g., ℓ_1 for taxicab norm minimization, ℓ_2 for euclidean norm minimization). M indicates the robust penalized estimators (Maximum Tangent likelihood Estimation or Least Squares Approximation of the unified LASSO) to which penalization is applied for variable selection. Finally, i denotes the information criterion (Akaike Information Criterion or Bayesian Information Criterion) employed for optimizing model complexity, typically by selecting the regularization parameter. Feature importance is quantified using Variable Inclusion Probabilities (VIPs), derived via the Maximum Entropy Bootstrap (MEBoot) methodology, reflecting the probability of a regressor's inclusion in the model.

Data for this study were acquired from loop detectors positioned at seven signalized intersections along Alexandras Street. To clearly illustrate core model behaviors, the analysis focuses on location L102 (B), situated at the intersection of Alexandras Avenue and Varvakis Street. This site was selected due to its anticipated "regular" traffic flow, contrasting with other locations potentially affected by externalities. Specifically, Varvakis Street is not a major thoroughfare, and the traffic direction at L102 (B) is towards the Campus Martius, where access from Charilaou Trikoupi Street (a unidirectional street feeding into Alexandras Avenue from the opposite direction) is prohibited. This configuration mitigates traffic singularities that might otherwise arise from events such as those observed during election campaign periods (e.g., 3rd-7th and 10th-14th of April) on Charilaou Trikoupi Street, thus providing a dataset with fewer irregularities for model evaluation.

An examination of the Variable Inclusion Probabilities, visually summarized in the heatmap of Figure 4 and detailed in Table 2, reveals a consistent pattern of identified significant harmonic frequencies (including the intercept) across most penalized variable selection methods, irrespective of the minimization criterion (AIC or BIC). The primary exception to this trend is the ℓ_1 -norm Least Squares Approximation ($\ell_1\text{LSA}$), which yields markedly sparser models. Consequently, this analysis will first provide an explicit exposition of the robust MTE method's results. Findings from other methods that align with MTE, such as $\ell_2\text{LSA}$, will be presented more compactly. This will be followed by a dedicated discussion of the robust $\ell_1\text{LSA}$ method to highlight distinctions, particularly those arising from smooth versus non-smooth criteria minimization processes.

A key aspect of the MTE model implementation throughout this study is the utilization of a second-order Taylor series expansion of the log-probability function. This approximation is employed for computational tractability, ensuring a balance between numerical stability and estimation precision within the MTE framework.

Robust Maximum Tangent Likelihood Estimation ($\ell_1\text{MTE}$)

The evaluation of the $\ell_1\text{MTE}$ approach commences with the Akaike Information Criterion ($\ell_1\text{MTE}_{\text{AIC}}$). Employing this model, features with a VIP of 100% (denoting consistent selection across all 20 ME-Boot samples) include the Intercept and the harmonic terms $\sin(2\pi \cdot x \frac{t}{T})$ for $x \in \{5, 10, 20\}$, and $\cos(2\pi \cdot x \frac{t}{T})$ for $x \in \{5, 15\}$, where $T = 2400$ represents the period. Other regressors exhibit lower inclusion probabilities: $\sin(2\pi \cdot 15 \frac{t}{T})$ and $\cos(2\pi \cdot 25 \frac{t}{T})$ were selected with a probability of 35%; $\sin(2\pi \cdot 3 \frac{t}{T})$ with 15%; and both $\sin(2\pi \cdot 45 \frac{t}{T})$ and $\cos(2\pi \cdot 10 \frac{t}{T})$ with 5%. These VIPs are visually depicted in Figure 5 and correspond numerically to the $\text{L1MTE}_{\text{AIC}}$ column in Table 2.

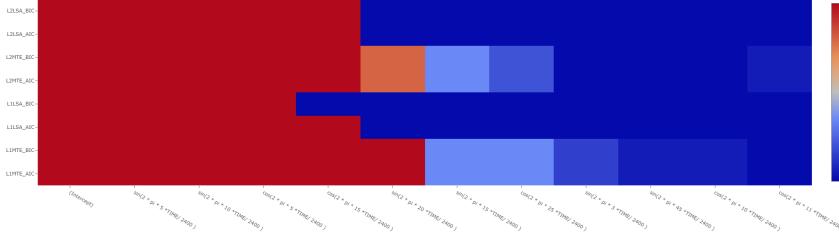


Figure 4: Heatmap of Variable Inclusion Probabilities across all the models performed against 30 MEBoot samples regarding L102 location's data dynamics.

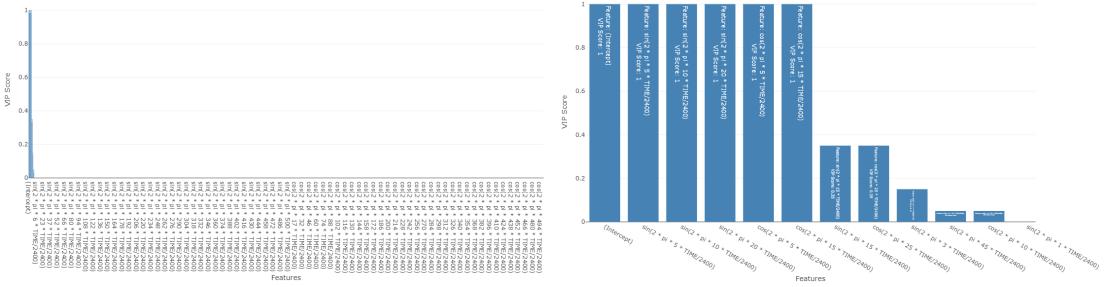


Figure 5: Variable Importance Plot for the ℓ_1 MTEAIC model applied to location L102 data. The right panel provides a magnified view of lower-ranked regressors.

This variable selection process leads to highly sparse coefficient vectors, as illustrated in Figure 6 and in table 10. The coefficient estimates demonstrate notable consistency in magnitude across the MEBoot samples, which translates to stable patterns in both fitted and predicted values. The ℓ_1 MTEAIC model predominantly selects low harmonic frequency regressors. Interpreting traffic volume within a signal processing paradigm, these low frequencies, corresponding to high wavelengths, represent the persistent, dominant characteristics of the traffic flow. The penalized regression framework effectively isolates these fundamental attributes. This outcome suggests that the selected low harmonic frequencies encapsulate the primary characteristics of the traffic volume signal, a hypothesis validated by the robust MTE model's performance on MEBoot samples.

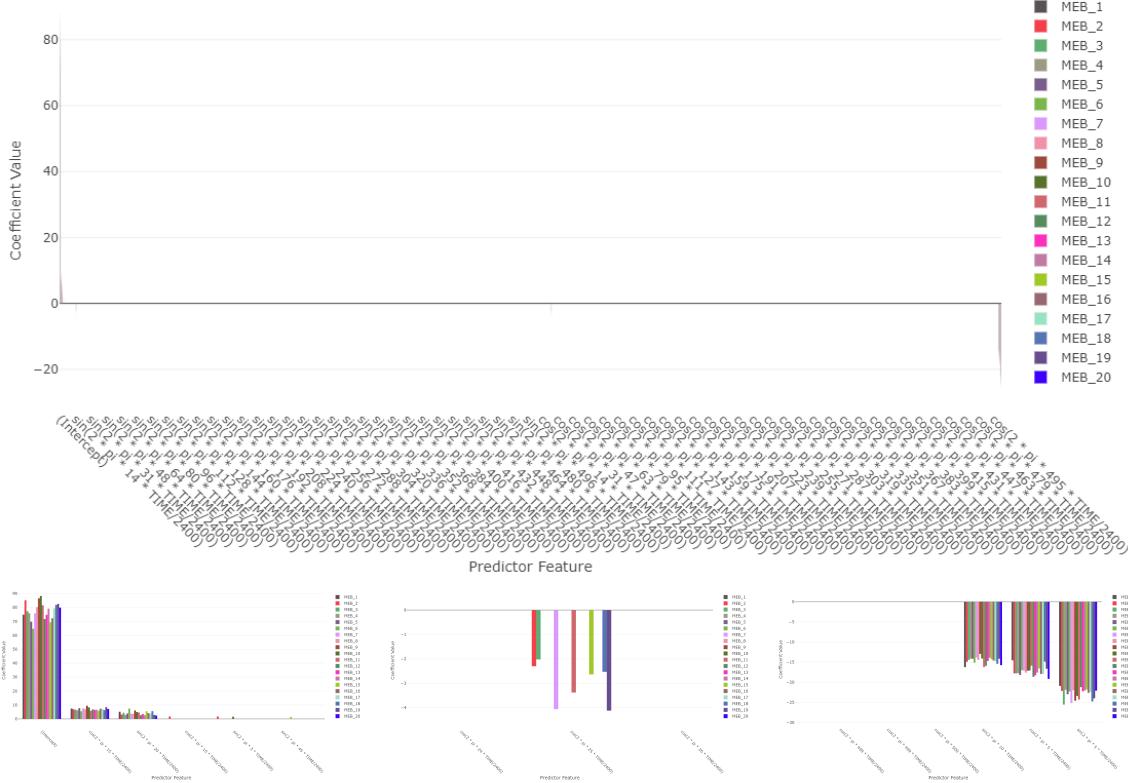


Figure 6: Barplot of coefficient estimates for each MEBoot dataset using the ℓ_1 MTEAIC model at location L102. Bottom panels offer magnified views of specific coefficient groups.

Figure 7 illustrates the performance of the ℓ_1 MTEAIC estimator at location L102, providing in-sample (training period: 3rd-14th April) and out-of-sample (test period: 18th-22nd April) evaluations. During the training phase, fitted values (top-left panel) demonstrate strong concordance with observed data across all MEBoot samples, signifying accurate capture of primary temporal characteristics, particularly underlying periodicity. The corresponding training residuals (top-right panel) fluctuate around zero without discernible bias or heteroscedasticity, indicative of a well-calibrated fit. In the test phase, the model sustains its predictive structure (bottom-left panel), maintaining the fundamental waveform and exhibiting robust generalization. Test period residuals (bottom-right panel), while manifesting slightly increased variability, remain centered and controlled. The robustness inherent in minimizing absolute deviations is evident during specific periods with known externalities: the late hours of April 5th and 7th (coinciding with political campaigns¹¹) and April 20th (following a significant local sporting event¹²). The model yields approximately stationary residual time series, and the absence of residual seasonality confirms effective capture of seasonal patterns.

¹¹Political campaigns occurred in central Athens during this week.

¹²Panathinaikos basketball club, with a fan club near Alexandras Street, won a European title on April 20th-21st.

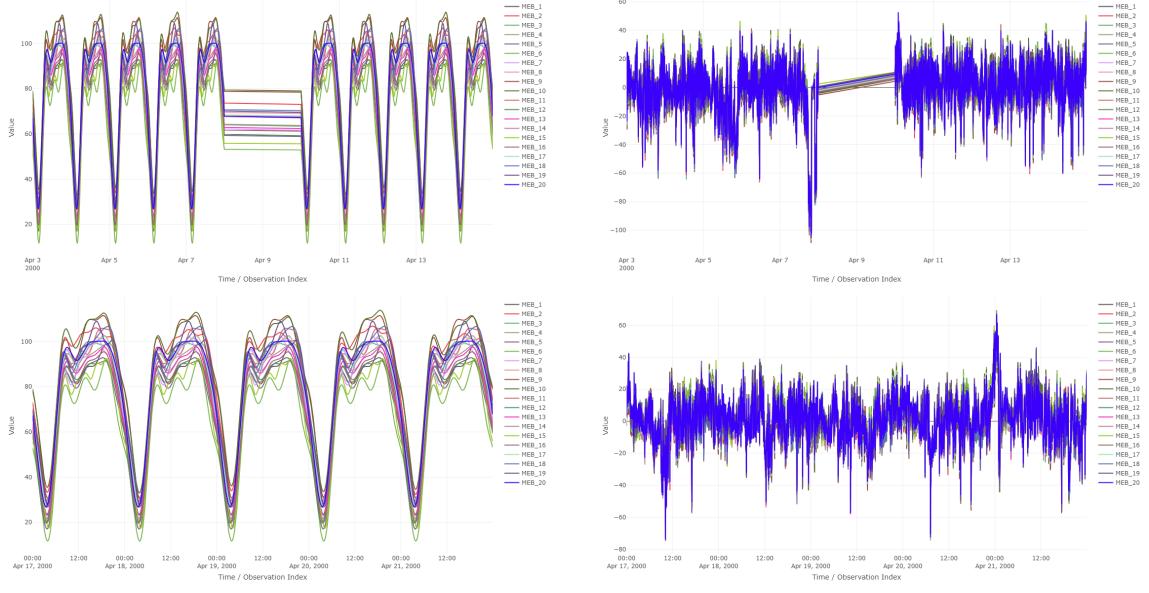


Figure 7: Performance of the ℓ_1 MTE_{AIC} model at location L102. Top-left: fitted values for training period (3rd–14th April). Top-right: training residuals. Bottom-left: forecasted values for test period (18th–22nd April). Bottom-right: forecast errors.

A comparison of coefficient estimates from the ℓ_1 MTE_{AIC} model (averaged over MEBoot samples and applied to the original sample) with those from a post-hoc adaptive ℓ_1 -Lasso (**ADL1Lasso**) model (penalized using weights derived from VIPs) is presented in Figure 8. The ℓ_1 MTE_{AIC} yields sparser estimations compared to the **ADL1Lasso**. The comprehensive view (top panel) illustrates the corrective influence of the adaptive Lasso, particularly relative to the MEBoot and original sample estimates. Magnified views (bottom panels) highlight key predictors, underscoring their consistent sign and relative magnitude across these estimation approaches for influential terms.

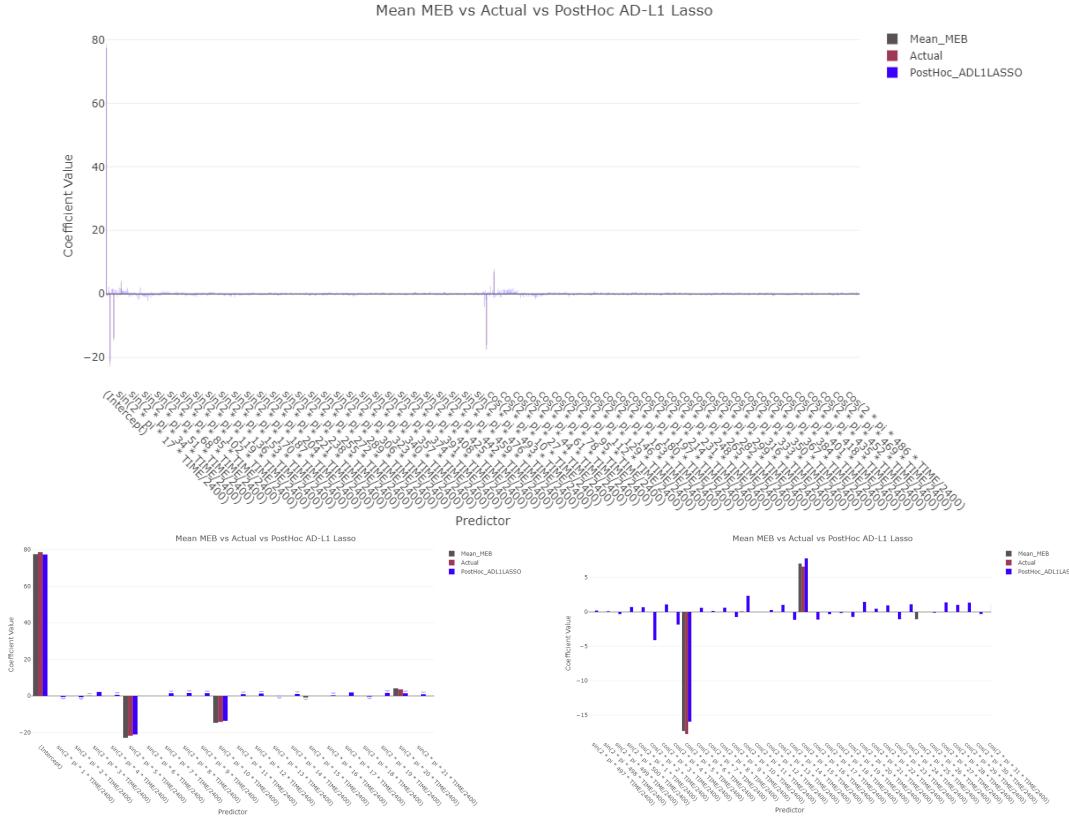


Figure 8: Comparative barplot of average coefficients from $\ell_1\text{MTE}_{\text{AIC}}$ on MEBoot samples, $\ell_1\text{MTE}_{\text{AIC}}$ on the actual sample, and post-hoc adaptive ℓ_1 -Lasso (weights from VIPs) for location L102.

Figure 9 displays violin plots for various error metrics, Huber Loss (HL), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Huber Loss, Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and Symmetric Median Absolute Percentage Error (SMdAPE), for both training and testing datasets under the $\ell_1\text{MTE}_{\text{AIC}}$ framework. A marked consistency between training and testing performance is evident. MAE, RMSE, and Huber Loss exhibit proximate central tendencies and variability, with MAE and RMSE values typically between 11 and 13 units. It is worth emphasizing that the model averaging of the $\ell_1\text{MTE}_{\text{AIC}}$ across 20 Maximum Entropy Bootstrap samples consistently yields the second-lowest Huber loss across all three evaluation metrics (HL, MAE, and RMSE) in both the training and testing phases, deviating by merely 0.10-0.12 units from the absolute minimum. Moreover, the percentage-based error metrics further underscore its competitiveness: the method achieves the lowest MdAPE during training, and both the lowest MdAPE and SMdAPE during testing. For the remaining percentage error measures, the deviations from the minimum lie between 0.06-0.49 units in training and 0.25-0.89 units in testing. In contrast, when the $\ell_1\text{MTE}_{\text{AIC}}$ is applied directly to the original traffic dataset at location L102 (B), it attains the

lowest RMSE, despite its foundation in minimizing absolute deviations. Additionally, it records the lowest MAPE and MdAPE in both the training and testing datasets. The values for the remaining performance metrics differ only slightly, ranging from 0.01 to 0.04 units relative to the optimal results in both training and testing contexts. Conversely, the model incorporating post-hoc variable selection exhibits a more pronounced deviation from the optimal metrics, with discrepancies of 0.35-0.50 units in HL, MAE, and RMSE, and 0.25-0.38 units in percentage errors for the training data. In the testing phase, the variations range from 0.04-0.50 units for HL, MAE, and RMSE, and from 0.15-0.50 units in the percentage error metrics. Percentage-based metrics (MAPE, MdAPE) show wider distributions, centered around 100 (as they are percentages), reflecting sensitivity to low traffic volumes. Conversely, symmetric percentage metrics (SMAPE, SMdAPE) present narrower, more stable distributions, with medians generally between 15 and 18. The performance of the post-hoc adaptive ℓ_1 -Lasso (navy blue diamond) and the ℓ_1 MTE_{AIC} on the original data (black diamond) are depicted for comparison. The adaptive ℓ_1 -Lasso exhibits somewhat larger test errors compared to its training performance and relative to MEBoot results for several metrics, though it demonstrates competitive RMSE values.

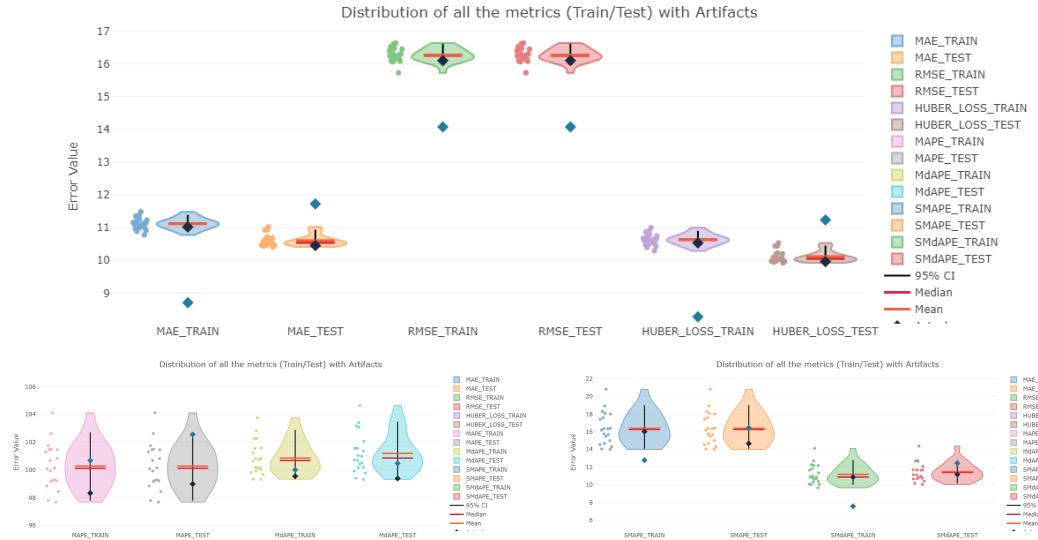


Figure 9: Violin plots of performance metrics (Huber Loss, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for location L102 with ℓ_1 MTE_{AIC}. Navy blue diamond: adaptive ℓ_1 -Lasso. Black diamond: ℓ_1 MTE_{AIC} on initial data. Other distributions: ℓ_1 MTE_{AIC} on MEBoot samples.

Transitioning to the Bayesian Information Criterion, the ℓ_1 MTE_{BIC} model yields remarkably similar results to its AIC counterpart regarding regressor significance and inclusion order across MEBoot samples (Figure 10, Table 2 [L1MTE_{BIC} column], and Figure 4). Crucially, coefficient estimates from ℓ_1 MTE_{BIC} (Figure 11) are virtually indistinguishable from those obtained with ℓ_1 MTE_{AIC}. This congruence is significant as it suggests that the non-smooth optimization potentially associated with AIC-based penalty selection can be supplanted by the smoother BIC-based minimization within the ℓ_1 MTE framework without altering model structure or performance, thereby offering

computational benefits.

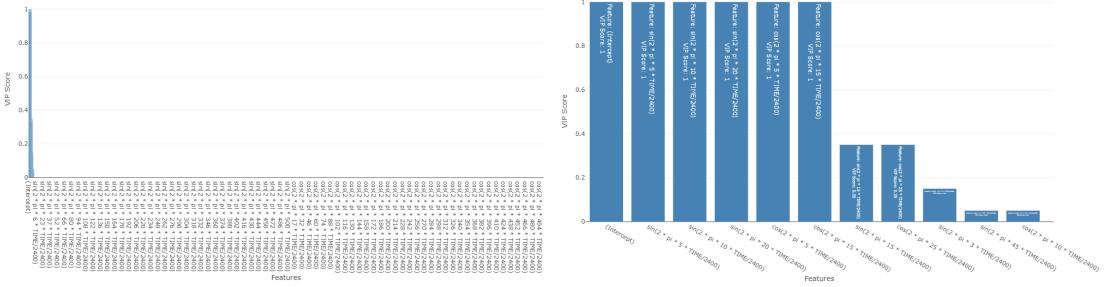


Figure 10: Variable Importance Plot for the ℓ_1 MTE_{BIC} model applied to location L102 data. The right panel provides a magnified view.

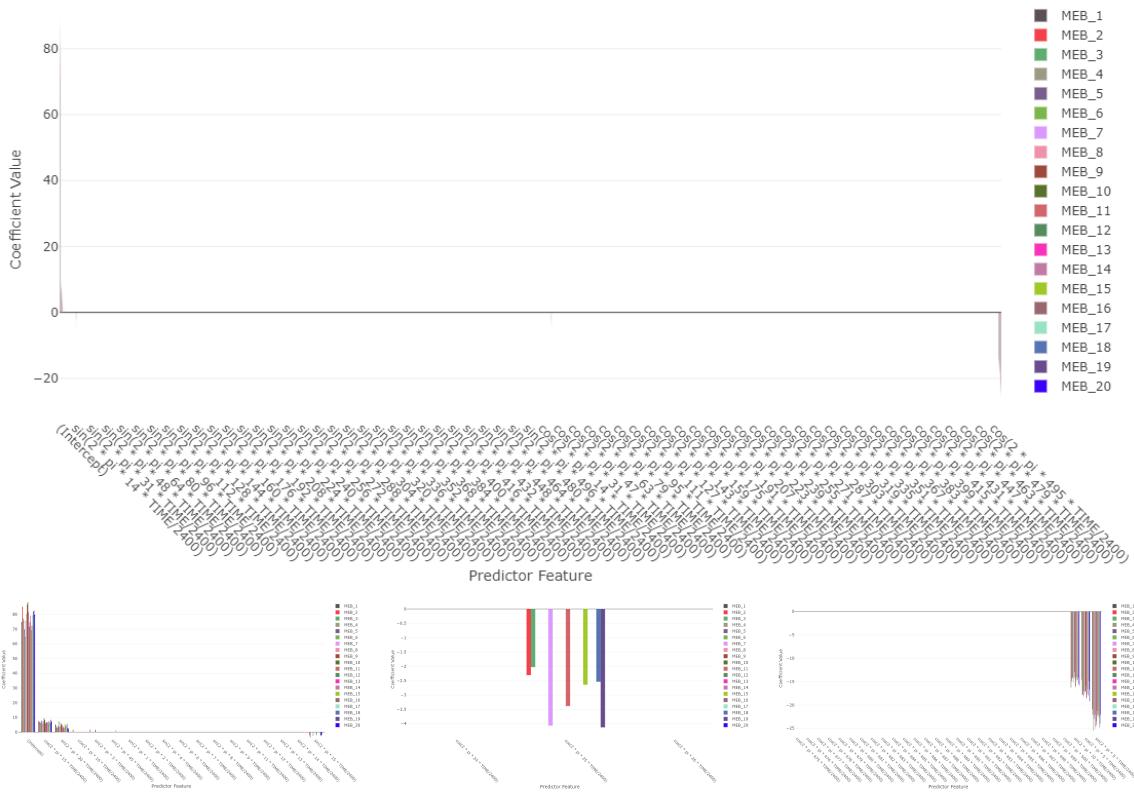


Figure 11: Barplot of coefficient estimates for each MEBoot dataset using the ℓ_1 MTE_{BIC} model at location L102. Bottom panels offer magnified views.

Consistent with coefficient similarity, the in-sample fitted values, residuals, out-of-sample predictions, and forecast errors for $\ell_1\text{MTE}_{\text{BIC}}$ (Figure 12) closely mirror those for $\ell_1\text{MTE}_{\text{AIC}}$. This parallelism further attests to the model's robustness and generalization capacity. Consequently, the comparative analysis involving the post-hoc adaptive ℓ_1 -Lasso (Figure 13) for $\ell_1\text{MTE}_{\text{BIC}}$ reiterates the findings from the AIC-based analysis.

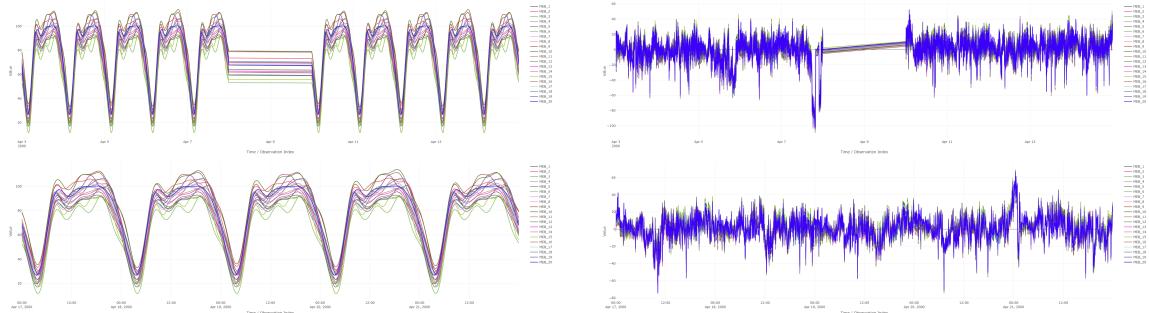


Figure 12: *Performance of the $\ell_1\text{MTE}_{\text{BIC}}$ model at location L102, analogous to Figure 7.*

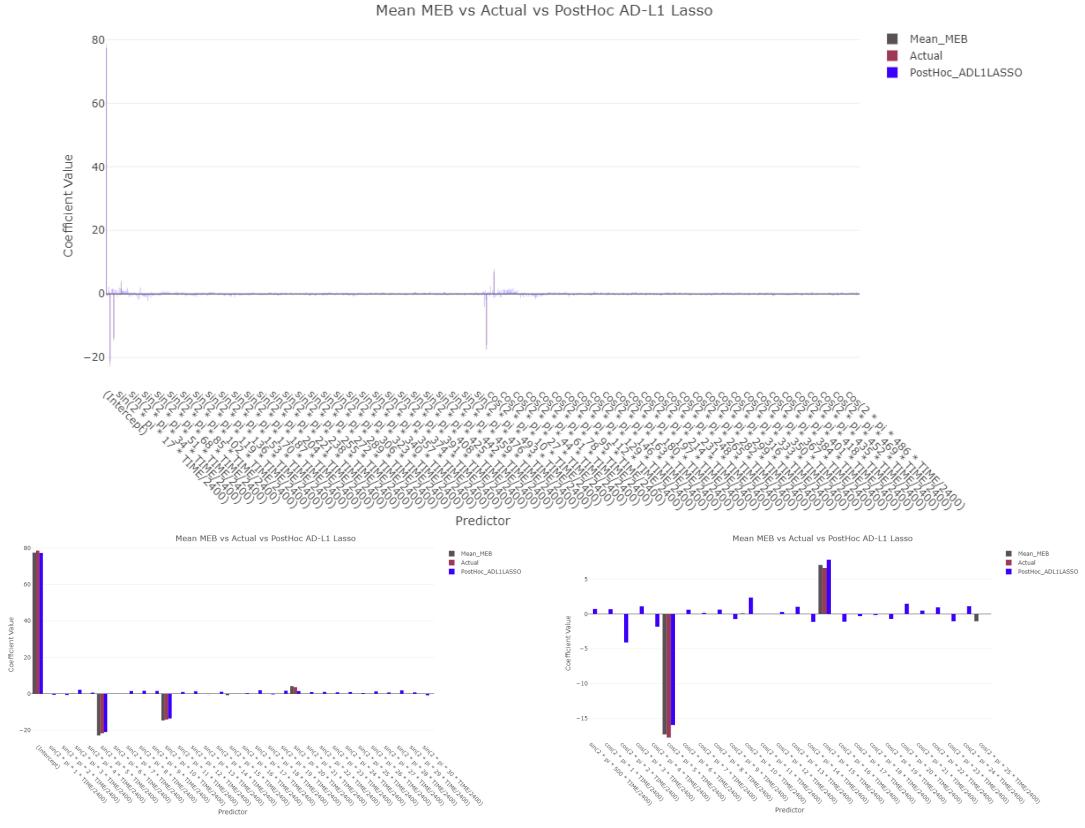


Figure 13: Comparative barplot of average coefficients for $\ell_1\text{MTE}_{\text{BIC}}$ analogous to Figure 8.

The error metric distributions for $\ell_1\text{MTE}_{\text{BIC}}$ (Figure 14) are also virtually identical to those of $\ell_1\text{MTE}_{\text{AIC}}$. A consistent performance profile is observed across training and testing phases for all metrics. Minor variations might suggest slight advantages for $\ell_1\text{MTE}_{\text{BIC}}$ in terms of median MAE, Huber Loss, and median-based percentage errors, but the overall performance landscape remains highly comparable.

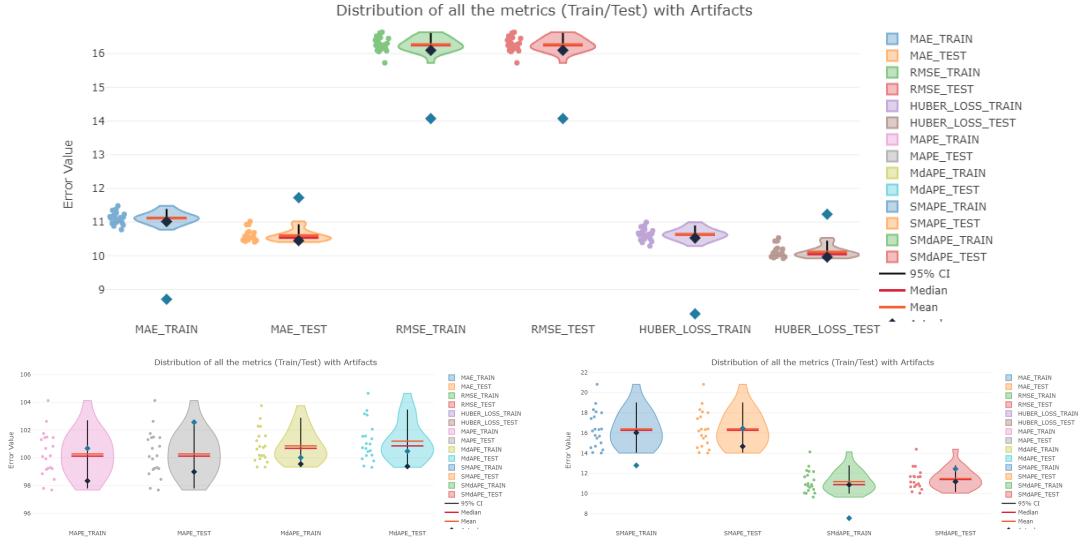


Figure 14: Violin plots of performance metrics for location L102 with $\ell_1\text{MTE}_{\text{BIC}}$, analogous to Figure 9.

Conventional Maximum Tangent Likelihood Estimation ($\ell_2\text{MTE}$)

This subsection details the performance of the conventional Maximum Tangent Likelihood Estimation, characterized by an ℓ_2 -norm in its objective function ($\ell_2\text{MTE}$), again employing the second-order Taylor approximation for the log-likelihood. A salient observation is that both AIC and BIC minimization procedures for selecting the regularization parameter yield identical sets of significant harmonic regressors. Specifically, the intercept term and the regressors within the set $\{\sin(2\pi \cdot x \frac{t}{T}) | x \in \{5, 10\}\} \cup \{\cos(2\pi \cdot x \frac{t}{T}) | x \in \{5, 15\}\}$, where $T = 2400$, are unanimously identified as important predictors with a Variable Inclusion Probability (VIP) of 100%. Beyond these, $\sin(2\pi \cdot 20 \frac{t}{T})$ is included with a VIP of 0.80 (16 out of 20 MEBoot samples). Less frequently selected features include $\sin(2\pi \cdot 15 \frac{t}{T})$ with a VIP of 35% (7/20 samples), $\cos(2\pi \cdot 25 \frac{t}{T})$ with 20% (4/20 samples), and $\cos(2\pi \cdot 11 \frac{t}{T})$ with 5% (1/20 samples). These probabilities are documented in Table 2 (columns $\mathbf{L2MTE}_{\text{AIC}}$ and $\mathbf{L2MTE}_{\text{BIC}}$) and visualized in Figure 15.

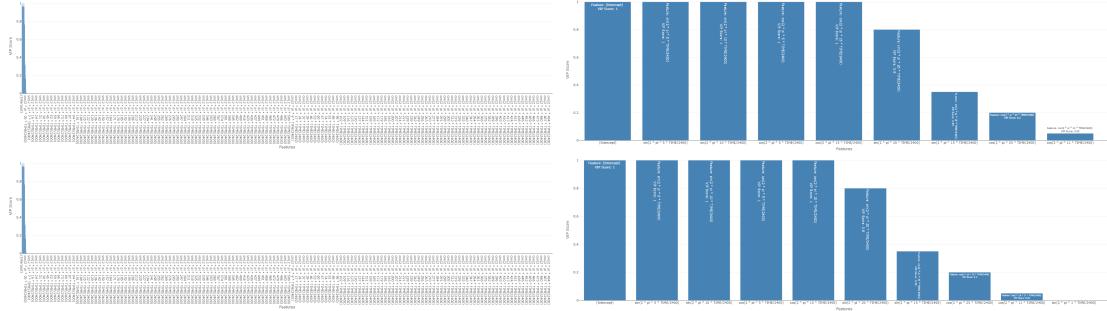
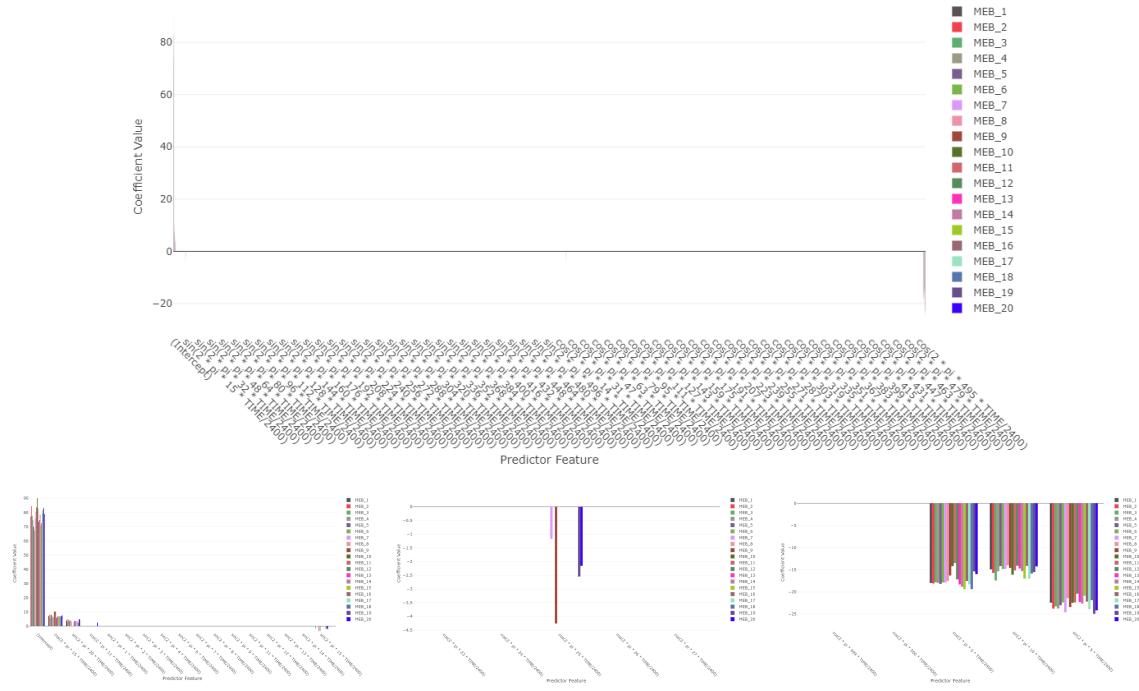


Figure 15: *Variable Importance Plots for ℓ_2 MTE_{AIC} (top row) and ℓ_2 MTE_{BIC} (bottom row) applied to location L102 data.* Right panels offer magnified views.

The coefficient estimates derived from the ℓ_2 MTE approach (Figure 16 and Table 10) exhibit general similarity. Compared to the ℓ_1 MTE models, a primary distinction lies in the VIP for $\sin(2\pi \cdot 20\frac{t}{T})$, which is 80% under ℓ_2 MTE rather than 100%. Furthermore, $\sin(2\pi \cdot 15\frac{t}{T})$ maintains a VIP of 35%, while $\cos(2\pi \cdot 25\frac{t}{T})$ has a reduced VIP of 20% (down from 35% in ℓ_1 MTE_{AIC/BIC}), and $\cos(2\pi \cdot 11\frac{t}{T})$ is selected with a VIP of 0.05. These differences in VIPs contribute to some variability in coefficient magnitudes across MEBoot samples, particularly for less consistently selected regressors, and may result in sparser representations in some zoomed views of the coefficient barplots if certain terms are frequently zero.



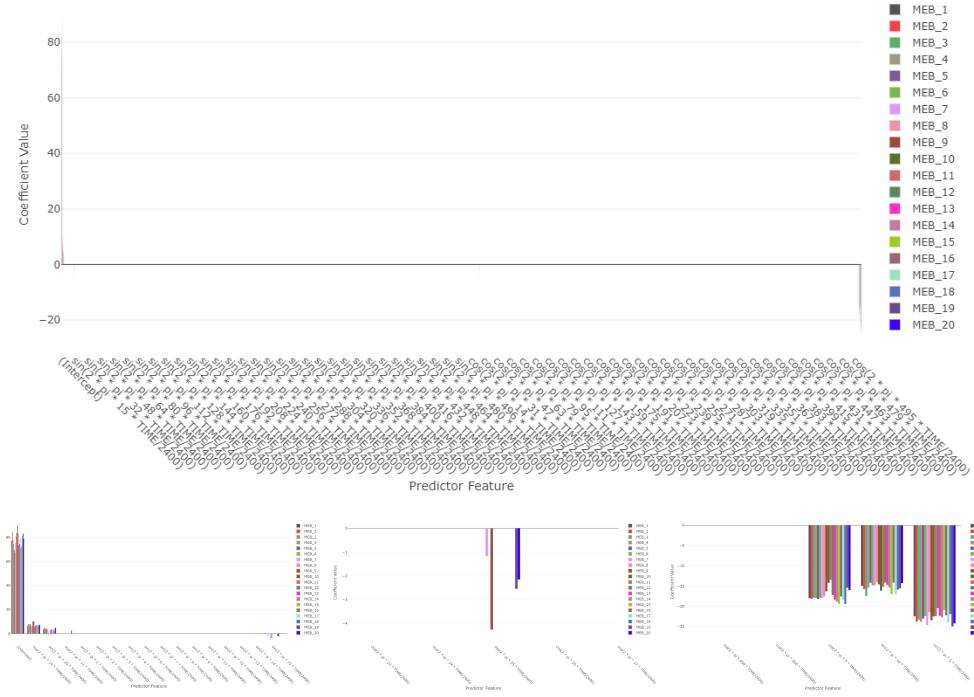


Figure 16: *Coefficients estimation barplot for each MEBoot dataset while performing ℓ_2 MTE_{AIC} (the first top foursome plots) and ℓ_2 MTE_{BIC} on L102.*

Given that the AIC and BIC minimization procedures for the regularization parameter (λ) lead to the same set of selected regressors within the ℓ_2 MTE framework, the resulting fitted values, forecasted values, and associated residuals are virtually identical. Figure 17 presents these outcomes for both ℓ_2 MTE_{AIC} (first quartet of plots) and ℓ_2 MTE_{BIC} (second quartet). In each group, the panels display fitted values and residuals for the training period (3rd–14th April), alongside predicted values and forecast errors for the test period (18th–22nd April). Both estimators effectively capture the underlying traffic flow periodicities and demonstrate robust predictive capability. While largely similar, any minimal variations in residual distributions might reflect subtle differences in how each criterion fine-tunes the model’s sensitivity to data volatility.

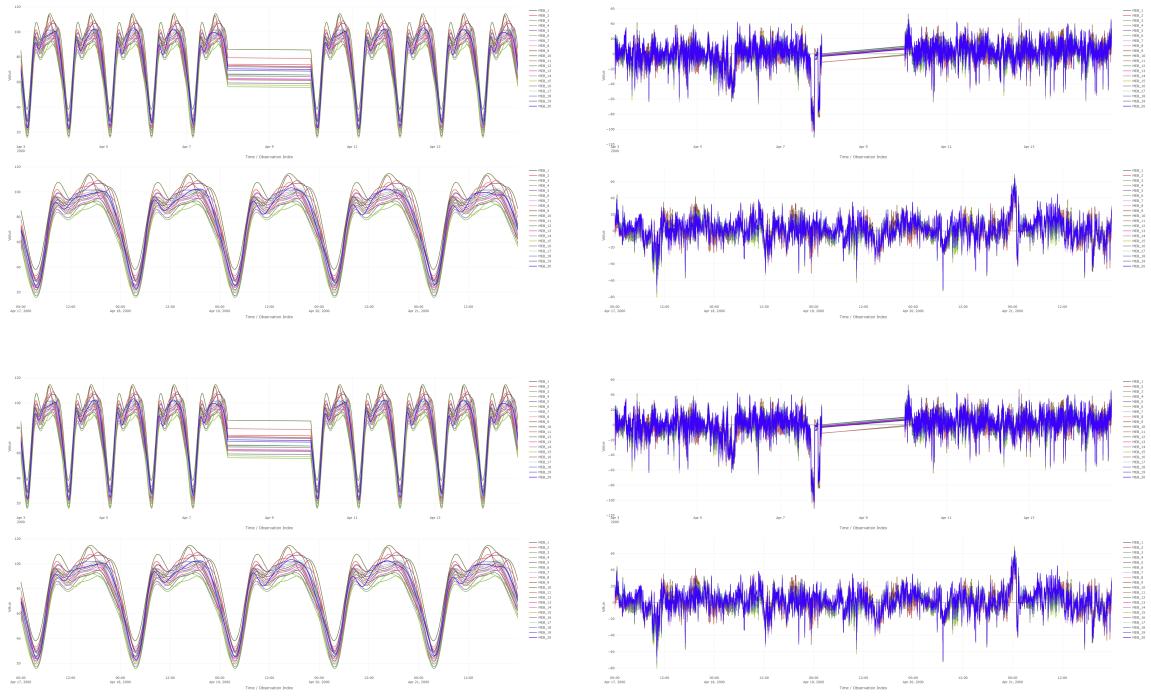


Figure 17: Performance of $\ell_2\text{MTE}_{\text{AIC}}$ (top quartet) and $\ell_2\text{MTE}_{\text{BIC}}$ (bottom quartet) at location L102. Each quartet shows training fitted values (top-left), training residuals (top-right), test predictions (bottom-left), and test forecast errors (bottom-right).

A post-hoc analysis was undertaken using an adaptive ℓ_1 -Lasso (Adaptive LAD LASSO), with penalty weights obtained from the VIPs of the $\ell_2\text{MTE}$ models, as we did before. Figure 18 compares average coefficient estimates from: (i) $\ell_2\text{MTE}$ applied to MEBoot samples, (ii) $\ell_2\text{MTE}$ applied to the original sample, and (iii) the aforementioned adaptive ℓ_2 -Lasso. The top set of three panels corresponds to $\ell_2\text{MTE}_{\text{AIC}}$ and the bottom set to $\ell_2\text{MTE}_{\text{BIC}}$. A general alignment is observed between coefficients averaged from MEBoot samples and those from the original sample for both criteria. However, variations in magnitude and occasionally sign for certain coefficients indicate sensitivity to sample-specific characteristics or the choice of information criterion influencing the regularization.

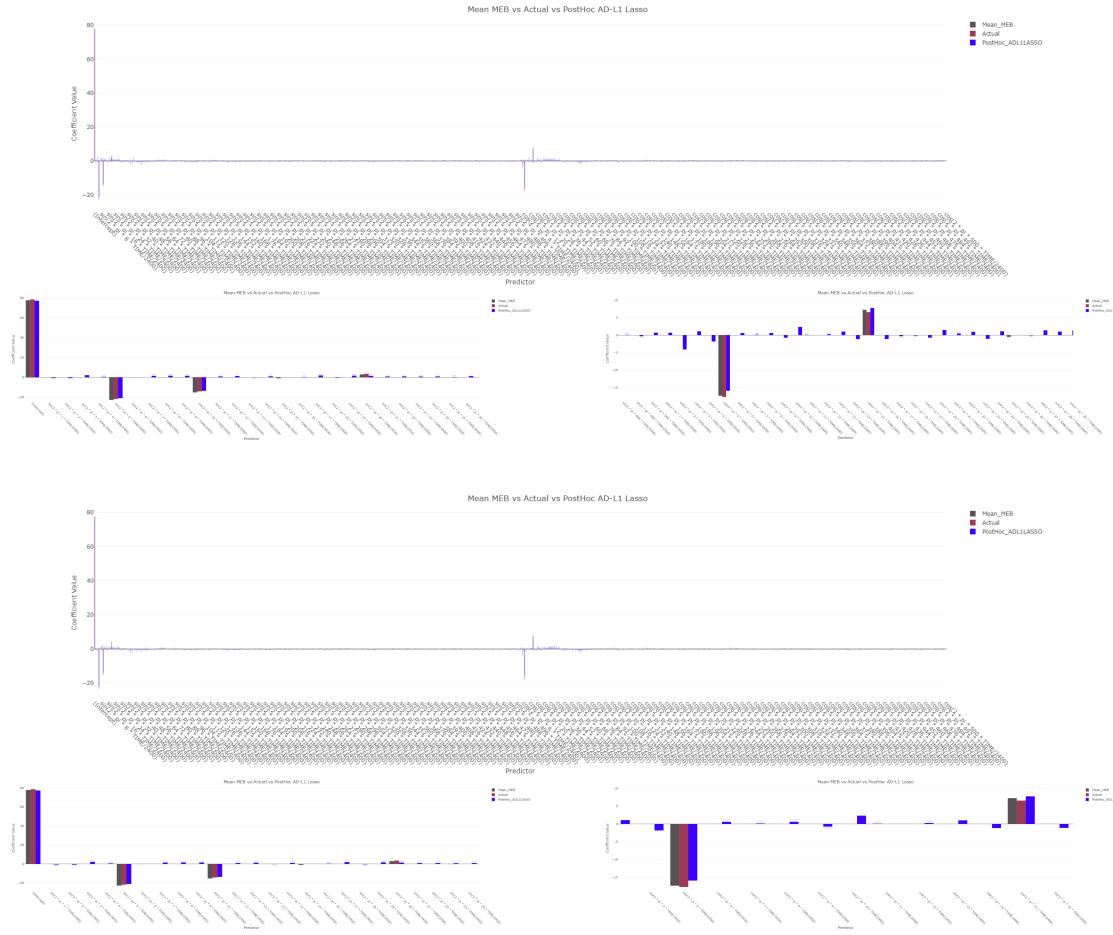


Figure 18: Comparative barplots of average coefficients for location L102. Top three panels: ℓ_2 MTE_{AIC}. Bottom three panels: ℓ_2 MTE_{BIC}. Comparisons are between MEBoot sample averages, actual sample estimates, and adaptive ℓ_2 -Lasso (weights from VIPs).

The distributions of key performance metrics (Huber Loss, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for ℓ_2 MTE_{AIC} (top trinity of plots) and ℓ_2 MTE_{BIC} (bottom threesome plots) at location L102 are illustrated in Figure 19. Each plot shows the distribution across MEBoot samples, with the **navy blue** diamond indicating the adaptive ℓ_2 -Lasso performance on the original data, and the black diamond representing the ℓ_2 MTE model's performance on the original data. The concordance between distributions from MEBoot samples and performance on the actual data supports the MEBoot methodology's utility for assessing model stability. These results collectively affirm the robustness of the ℓ_2 MTE framework and the adaptive estimation schemes in characterizing traffic flow dynamics.

Since the performance of the ℓ_2 MTE estimator remains identical under both AIC and BIC minimization criteria, we present unified results for these cases. When applying model averaging over 20 Maximum Entropy Bootstrap (MEB) samples, the ℓ_2 MTE_{AIC/BIC} achieves the lowest values for

HL, MAE, and RMSE across both training and testing phases. Despite being grounded in minimizing squared errors, its performance in terms of MAE and HL surpasses even that of its robust ℓ_1 -based counterparts. Regarding percentage error metrics, the method exhibits deviations of only 0.89 (MAPE) and 0.34 (MdAPE) units from the minimum in training, and 0.89 and 0.41 units, respectively, in testing. Interestingly, its symmetric error counterparts (SMAPE and SMdAPE) are minimized by the conventional ℓ_1 MTE_{AIC} under model averaging, with the sole exception of SMdAPE in the test set, where the ℓ_2 MTE result is merely 0.03 units above the minimum. When applied directly to the original traffic dataset, the ℓ_2 MTE estimator continues to perform robustly, yielding minimum HL and MAE in both training and testing. Furthermore, it attains the lowest SMAPE on training data, and both SMAPE and SMdAPE on the testing data. However, it slightly underperforms in RMSE, exceeding the minimum by just 0.03 units in both train and test sets. The MAPE and MdAPE deviate by 2.69 and 2.76 units, respectively, in training, and by 2.62 and 1.74 in testing—values that remain within an acceptable margin. The post-hoc adaptive ℓ_1 LASSO model, constructed using VIPs from the MEBoot samples, displays commendable generalization capabilities. On training data, HL, MAE, and RMSE are recorded as 8.48, 8.91, and 14.60, respectively, deviating by 0.55, 0.53, and 1.03 units from the corresponding minima. Notably, it achieves the minimum values for MAPE and MdAPE, while SMAPE and SMdAPE trail the best by 0.71 and 0.43 units, respectively. In the testing scenario, this LASSO-based model remarkably attains minimum values for HL, MAE, MAPE, MdAPE, and SMdAPE, with respective scores of 11.19, 11.68, 102.25, 99.97, and 12.29. Its RMSE and SMAPE are marginally higher than their best-performing counterparts by 1.03 and 0.13 units, respectively. Overall, it is noteworthy that the direct application of ℓ_2 MTE and its model-averaged variant demonstrate slightly superior generalization performance on the testing set relative to the training set, underscoring their stability and predictive consistency.

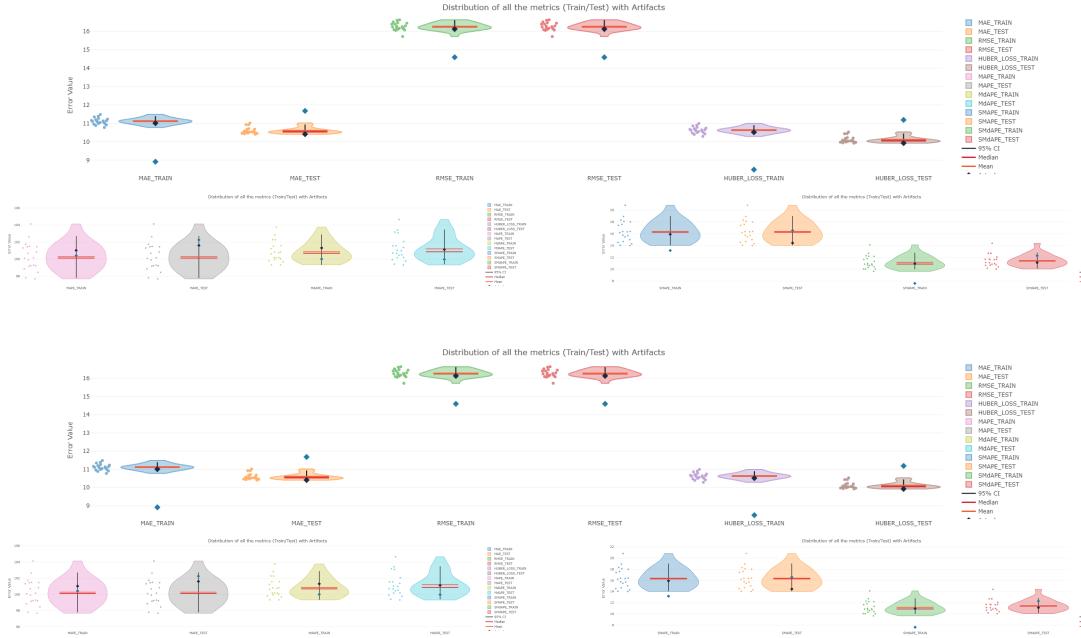


Figure 19: *Violin plots of performance metrics for location L102.* Top three plots: ℓ_2 MTE_{AIC}. Bottom three plots: ℓ_2 MTE_{BIC}. The navy blue diamond presents the adaptive ℓ_2 Lasso and the black one shows the corresponding method’s performance on the initial data of L102 location. The rest came from the application of the corresponding method on the MEBoot samples of L102 location data.

Diagnostic assessment of model adequacy through residual analysis is provided by ACF and PACF plots. Figure 20 presents these plots for the residuals of the ℓ_2 MTE model under AIC minimization (left panel) and BIC minimization (right panel). For both specifications, the ACF and PACF coefficients predominantly reside within the 95% confidence intervals, suggesting an absence of significant autocorrelation in the residuals. This indicates that both models effectively capture the temporal dependencies in the data. The BIC-based specification may exhibit a slightly more constrained autocorrelation structure, potentially reflecting a more parsimonious yet adequately specified model. Comparatively, the residual behavior of these ℓ_2 MTE models contrasts with that observed for various ℓ_1 MTE-based approaches (including averages over MEBoot samples, application to actual data, and post-hoc adaptive ℓ_1 -Lasso). While the ℓ_1 MTE specifications generally produce well-behaved residuals, particularly when applied to the original dataset, certain configurations (e.g., model averaging over MEBoot samples for ℓ_1 MTE) might show more pronounced PACF spikes. The current ACF/PACF analysis for ℓ_2 MTE (Figure 20) confirms the suitability of these models, and this broader comparison underscores the importance of both the chosen norm (ℓ_1 vs. ℓ_2) and the specific model configuration in achieving uncorrelated residuals, a hallmark of good model fit.

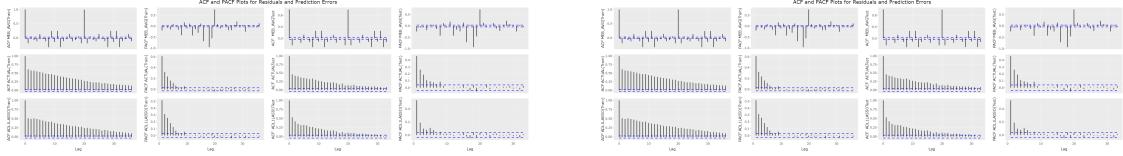


Figure 20: *Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) of residuals from $\ell_2\text{MTE}_{\text{AIC}}$ (left panel) and $\ell_2\text{MTE}_{\text{BIC}}$ (right panel) models for location L102.*

Robust Least Square Approximation of Unified LASSO ($\ell_1\text{LSA}$)

We now analyze the results obtained from the robust implementation of the least-squares approximation within the unified LASSO framework ($\ell_1\text{LSA}$). A detailed examination is crucial at this juncture to fully comprehend the implications of the regularization process and the variable selection strategy. Our focus begins with the approach that minimizes the Akaike Information Criterion ($\ell_1\text{LSA}_{\text{AIC}}$). When applied to 20 Maximum Entropy Bootstrap (MEB) samples, this technique consistently identified the following features as significant, demonstrating a Variable Inclusion Probability (VIP) of 100%:

$$\{\text{Intercept}\} \cup \left\{ \sin\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5, 10\} \right\} \cup \left\{ \cos\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5, 15\} \right\}.$$

This unanimous selection is visually presented in Figure 21 and numerically substantiated in Table 2. It is immediately apparent that $\ell_1\text{LSA}_{\text{AIC}}$ imposes a more pronounced shrinkage compared to the MTE-based methods, ultimately retaining only 5 out of the initial 1001 regressors.

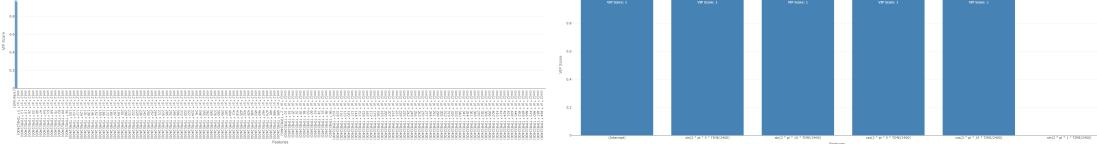


Figure 21: *Variable Importance Plot for $\ell_1\text{MTE}_{\text{AIC}}$ on L102*

Examination of the coefficients reveals that the $\ell_1\text{LSA}_{\text{AIC}}$ model selects nearly identical values for the chosen coefficients as the MTE models (Figure 22). A notable distinction, however, is the absence of certain regressors, such as $\sin\left(2\pi \cdot 20 \frac{t}{2400}\right)$, which were frequently identified as significant by the MTE-based approaches. The influence of these omitted regressors appears to be implicitly distributed among the selected features within the $\ell_1\text{LSA}$ framework, rendering them unnecessary for inference purposes. Furthermore, the coefficient estimates exhibit remarkable consistency across MEB samples, demonstrating minimal variance (Figure 22 and Table 10). Consistent with previous findings, the $\ell_1\text{LSA}_{\text{AIC}}$ model continues to prioritize low-frequency characteristics, highlighting their persistent importance in capturing the underlying signal.

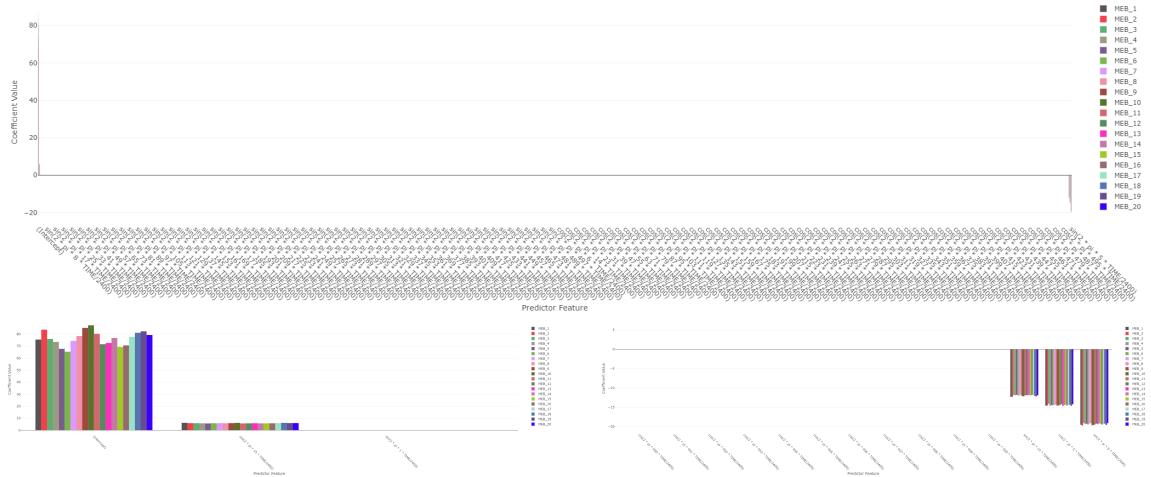


Figure 22: Coefficients estimation barplot for each MEBoot dataset while performing ℓ_1 LSA_{AIC} on L102

For computational feasibility, we employed a naive estimation of the variance-covariance matrix for the absolute deviation minimization within the ℓ_1 LSA method ¹³. This approach, based on Pollard’s method [16], involves generating a random sample from a Gaussian normal distribution and using the quadruple of the location multiplied by the covariance of the data as the inverse of robust covariance estimation. This choice offers a more robust scheme for the ℓ_1 LSA, as illustrated in Figure 23.

The fitted data exhibit a strong and consistent periodicity, with distinct cyclical patterns indicating well-defined daily or subdaily seasonality (Figure 23, top-left panel). The presence of multiple approximately parallel lines moving synchronously suggests a strong, unified model that effectively captures the underlying deterministic structure of the data. Constant vertical offsets and phase coherence further underscore the model’s ability to identify the fundamental patterns.

Residuals, depicted in the top-right panel of Figure 23, fluctuate around zero, indicating minimal average bias and successful extraction of the primary deterministic components. While some periods of increased variability suggest the presence of minor unmodeled stochasticity, the absence of strong, systematic residual patterns confirms the model’s efficacy in capturing the key dynamics of the series. Notably, singularities observed on April 7th and 10th persist, suggesting these data points may represent distinct regimes within our dataset. Similar irregularities are evident in the forecasting errors (bottom-right panel) around 11:30 AM on April 17th and 00:00 AM on April 21st. Forecast errors are also centered around zero, indicating unbiased predictions with variability comparable to training residuals. Occasional transitory spikes indicate greater variances but do not reveal long-term trends or patterns. This pattern suggests that the model correctly differentiated the key deterministic signal from the residual noise, demonstrating its generalizability and strong

¹³We attempted to account for autocorrelation and heteroskedasticity of the samples using Barrett kernel filtering in Andrews’s [48] implementations (Newey-West[23]). We attempted to use Cribari-Neto [21], Cribari-Neto-Da Silva [?], and Galvao-Yoon [49] implementations for a more robust estimation of the covariance matrix, however the computing cost was tremendous. After almost a month, we took it down and focused our investigation on Pollard’s [16] naïve but robust estimation of the covariance matrix.

predictive performance.

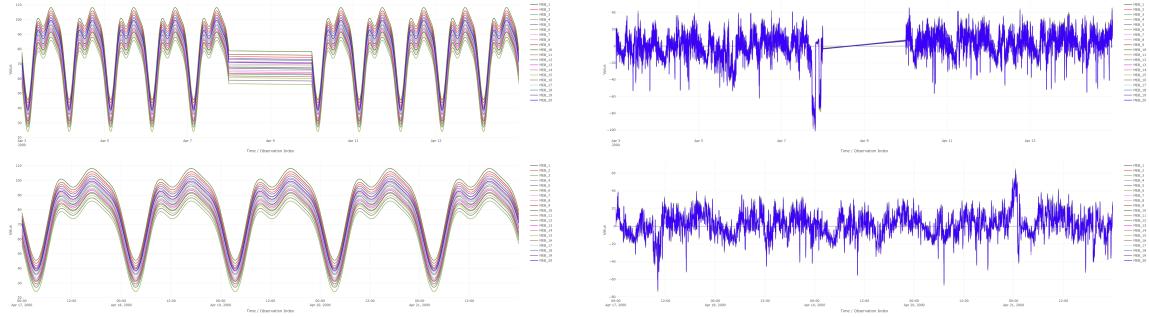


Figure 23: The first couple of panels (in top-left and bottom-right direction) depict the fitted values and residuals, respectively, for the training period (3rd–14th April). The remaining pair illustrates the forecasted values and corresponding errors during the test period (18th–22nd April), based on the application of ℓ_1 LSA_{AIC} on location L102.

As in previous analyses, we applied ℓ_1 LSA_{AIC} to the original sample, calculated the VIPs, and subsequently applied the adaptive robust (ℓ_1) LASSO to the original dataset. The comparative bar graph in Figure 24 indicates that the robust LSA-based model yields sparser estimates than the post-hoc adaptive LASSO. A noteworthy observation is the striking similarity in the coefficient estimates between the actual sample, the model averaging of the models applied to MEBoot samples, and the post-hoc adaptive LASSO for commonly selected features. The only significant deviation is observed for $\cos(2\pi \cdot 15 \frac{t}{2400})$, where the post-hoc LASSO estimate is approximately double that of the other two models.

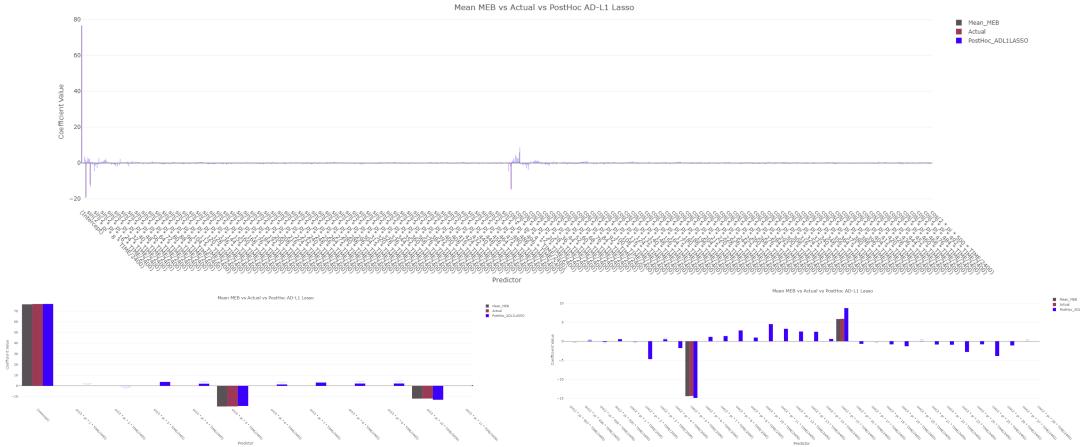


Figure 24: Comparative barplot between average coefficients from maximum entropy bootstrap samples, from actual sample and actual samples performing adaptive ℓ_1 -Lasso with weights constructed by VIPs for L102 while applying ℓ_1 LSA_{AIC}

Figure 25 provides a comprehensive depiction of various error metrics through violin plots, illustrating their distributional properties and central tendencies for ℓ_1 LSA_{AIC}. Each violin plot effectively portrays the probability density of the data, while embedded boxplots offer a concise summary of the median, mean, and interquartile ranges. This dual representation facilitates a thorough understanding of the measures' characteristics, including their dispersion and propensity to cluster around specific values.

For the MAE, the violin plots reveal a tight concentration around a central value for both training and test data, with minimal dispersion, indicating consistent model performance. Similar observations hold for RMSE and Huber Loss, which exhibit symmetric and slender violin shapes, signifying consistent accuracy and resilience. Conversely, MAPE and MdAPE plots display slightly broader outlines, suggesting greater variability. Nevertheless, the close alignment of median and mean values confirms that the central tendency of these measurements is accurately reflected. According to table 17, when applied through model averaging over 20 Maximum Entropy Bootstrap (MEB) samples, the ℓ_1 LSA method exhibits stable yet non-optimal performance. In both the training and testing scenarios, it does not secure the lowest values in the primary error metrics, HL, MAE, and RMSE, but remains within a reasonable proximity. Specifically, Huber loss (HL), MAE, and RMSE values are observed at 11.43, 11.92, and 16.38 in both datasets under MEB sampling. In terms of percentage-based error metrics, this model performs relatively weakly. For the training data, the MAPE and MdAPE reach values of 99.41 and 102.31, both of which are higher than those achieved by the ℓ_2 MTE or adaptive LASSO alternatives. SMAPE and SMdAPE stand at 18.16 and 12.55, respectively, which also fall short of the best-performing benchmarks. When applied directly to the original dataset (i.e., without MEB-based model averaging), ℓ_1 LSA offers modest improvements. HL, MAE, and RMSE slightly decrease to 11.28, 11.77, and 16.17, respectively. However, these gains remain insufficient to surpass the leading models. MAPE and MdAPE also remain high, at 99.46 and 102.27, while SMAPE improves marginally to 17.61, and SMdAPE to 12.29. Overall, the ℓ_1 LSA model—though grounded in robust regularization and well-suited for sparse feature

selection, does not achieve competitive results across the majority of error metrics under either AIC or BIC minimization. Its consistent performance across bootstrap samples and the original dataset highlights its robustness, yet the relatively high deviation from the optimal values suggests limited efficacy in this traffic forecasting context compared to alternatives like ℓ_2 MTE and adaptive ℓ_1 LASSO. The comparison of training and test errors across all measures is particularly insightful. The marginal increase in variability from the training to the testing phase, indicated by slightly broader violin plots for test data, is a natural and expected outcome of predictive modeling. This increase reflects the inherent generalization challenge in predictive models, where predictions on unseen data are inherently more uncertain than fits to historical data. However, the consistent alignment of the median and mean lines across training and testing phases underscores the model's capacity to deliver unbiased predictions without systematic deviations. The visualization in Figure 25 effectively highlights the efficiency of the ℓ_1 LSA_{AIC} approach, both in its baseline form and when applied to MEBoot samples. The clear and consistent portrayal of metric distributions provides a valuable tool for assessing model quality and discerning subtle changes induced by the bootstrapped sampling strategy. The effective use of violin plots, boxplots, and diamond markers offers a thorough yet accessible representation of the model's error characteristics and its robust generalization capabilities across the observed dataset.

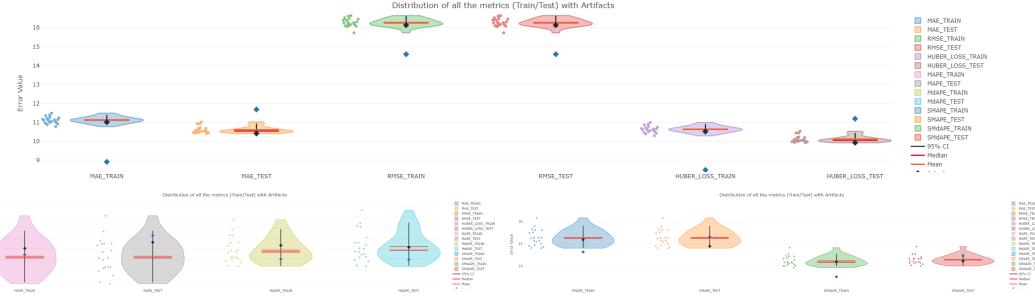


Figure 25: *Violin plots of performance metrics of Huber Loss, MAE, RMSE, MAPE, MdAPE, and their symmetric invariants For location L102 while applying ℓ_1 LSA_{AIC}. The navy blue diamond presents the adaptive ℓ_1 Lasso and the black one shows the ℓ_1 LSA_{AIC} on the initial data of L102 location. The rest came from the application of ℓ_1 LSA_{AIC} on the MEBoot samples of L102 location data.*

Analysis of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots for residuals and prediction errors (Figure 26) reveals distinct patterns. For the training data, the ACF plots for the ACTUAL and ADL1LASSO residuals exhibit a gradually decaying pattern, indicating some remaining autocorrelation structure not entirely captured by the models. This suggests that, while the fundamental dynamic components have been well described, minor dependencies persist in the residuals, pointing to potential areas for model refinement or integration of higher-order effects. In contrast, the PACF plots for these residuals show a rapid attenuation after the initial lags, implying that dependencies are primarily captured by a limited number of lags. The test data plots present a similar, albeit slightly more pronounced, pattern of autocorrelation, as expected due to the inherent uncertainty in prediction. Notably, the ACF for the MEB_AVG technique demonstrates lower autocorrelation than the other approaches, suggesting that model

averaging may provide a more parsimonious description of the temporal structure in out-of-sample predictions. This highlights the benefit of model averaging in mitigating overfitting and fostering more reliable forecasting trends.

Overall, the systematic presentation of ACF and PACF plots in this figure offers a nuanced understanding of model performance. The limited and rapidly declining partial autocorrelations across all approaches signify that primary temporal dependencies have been effectively removed from the residuals, leading to largely white noise behavior. However, the presence of subtle autocorrelation in both training and testing residuals points to potential avenues for improvement, such as incorporating additional lag terms or exploring more flexible model architectures to fully represent the complex dynamics of traffic data. This comprehensive illustration underscores the rigorous examination of model adequacy through residual diagnostics, emphasizing the interplay between model definition, error structure, and predictive fidelity. It demonstrates the diligent application of statistical principles to ensure that predictive models not only fit historical data well but also maintain their validity and robustness when extrapolated to unobserved data.

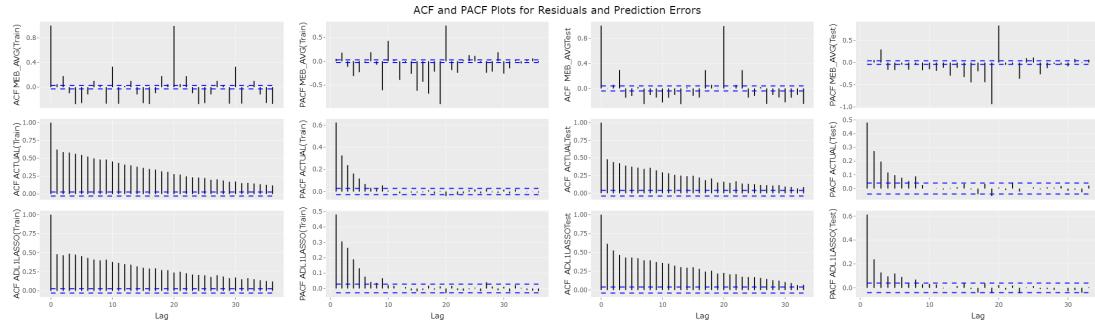


Figure 26: Autocorrelation and Partial Autocorrelation Functions of residuals and forecasting errors. The *MEB_AVG*, *ACTUAL* and *ADL1LASSO* correspond to the model averaging of the ℓ_1 LSA_{AIC} instance on 20 MEBoot synthetic dataset, on the original one and the post-hoc adaptive LAD Lasso based on the VIPs of the MEBoot experiments respectively.

Focusing on the BIC minimization for the ℓ_1 LSA method (ℓ_1 LSA_{BIC}), we observe that the same features were selected as for ℓ_1 LSA_{AIC}, with the notable exception of the harmonic cosine frequency on $15\frac{t}{2400}$ (Table 2). Intriguingly, this omitted feature was not selected in any of the MEBoot sample scenarios, whereas the remaining features were consistently chosen across all scenarios (Figure 27). This outcome is particularly surprising given that BIC minimization, due to its heavier penalization, is generally expected to yield a more parsimonious model. Hence, the regressors ultimately selected are:

$$\{\text{Intercept}\} \cup \left\{ \sin\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5, 10\} \right\} \cup \left\{ \cos\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5\} \right\},$$

where it is evident that the dominant features consistently correspond to the lower harmonic frequencies. This finding is clearly depicted in Figure 27 and detailed in Table 2. For further visual confirmation, refer also to Figure 4.

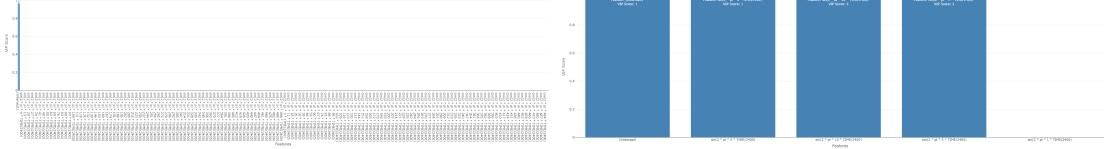


Figure 27: *Variable Importance Plot for ℓ_1 LSA_{BIC} on L102*

The coefficient estimates for ℓ_1 LSA_{BIC} are nearly identical to those obtained in the AIC minimization case, differing only by the nullification of the excluded coefficient. This consistency underscores the stability of coefficient estimation using this ℓ_1 LSA invariant method. As shown in Figure 28 and Table 10, the dispersion of value estimations is not statistically significant, indicating that, even with a limited set of 20 maximum entropy bootstrap samples, the ℓ_1 LSA invariant models converge to specific values for each regressor.

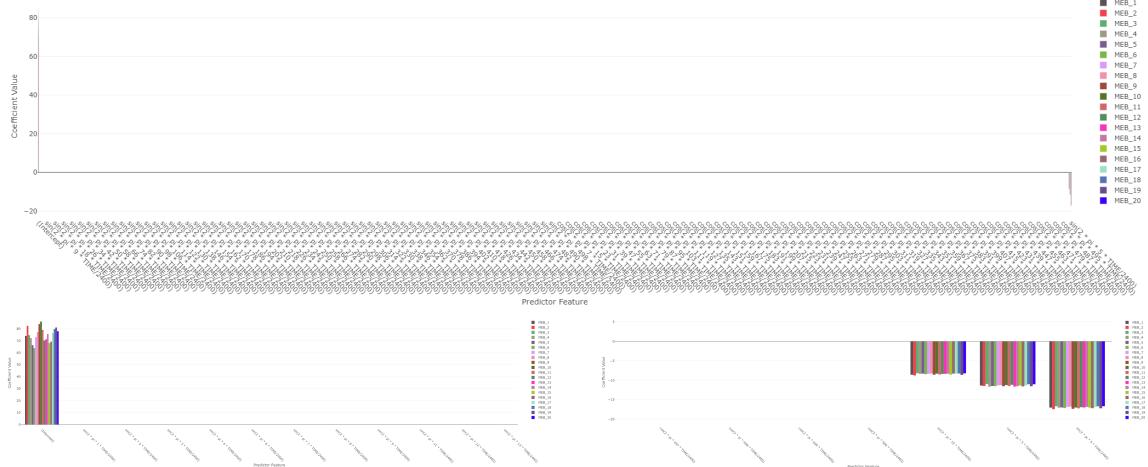


Figure 28: *Coefficients estimation barplot for each MEBoot dataset while performing ℓ_1 LSA_{BIC} on L102*

Excluding the $\cos(2\pi \cdot 15 \frac{t}{2400})$ regressor results in fitted values exhibiting less steep and wide fluctuations. However, the residual plots reveal more irregularities (Figure 29). Specifically, the ℓ_1 LSA_{BIC} estimation, again utilizing Pollard's implementation for covariance matrix estimation, lags behind the previously discussed techniques in capturing the periodicity of the traffic volume data at location L102 (B). This is evident in Figure 29, where the right-hand plots, representing residuals on training and testing data, appear to exhibit a seasonal pattern.

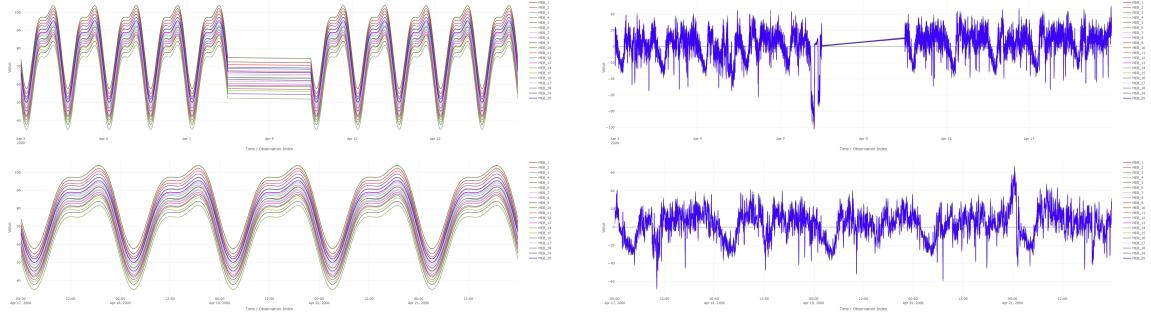


Figure 29: The first couple of panels (in top-left and bottom-right direction) depict the fitted values and residuals, respectively, for the training period (3rd–14th April). The remaining pair illustrates the forecasted values and corresponding errors during the test period (18th–22nd April), based on the application of $\ell_1\text{LSA}_{\text{BIC}}$ on location L102.

As anticipated, Figure 30 indicates that the post-hoc adaptive ℓ_1 -LASSO, employing VIPs derived from the ℓ_1 -LSA invariant minimizing the AIC, yields coefficient estimates identical to those of the previously presented models. Furthermore, the model averaging from the application of $\ell_1\text{LSA}_{\text{BIC}}$ against 20 MEBoot samples produces almost equal coefficient estimates to the model applied solely on the original dataset, with both sets of values closely aligning with those of the post-hoc adaptive LASSO for the selected predictors.

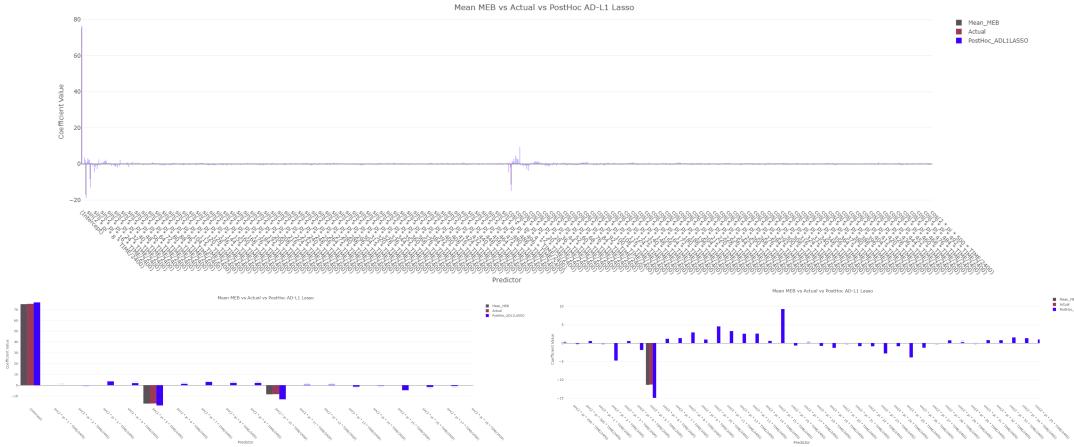


Figure 30: Comparative barplot between average coefficients from maximum entropy bootstrap samples, from actual sample and actual samples performing adaptive ℓ_1 -Lasso with weights constructed by VIPs for L102 while applying $\ell_1\text{LSA}_{\text{BIC}}$

Figure 31 demonstrates that the $\ell_1\text{LSA}_{\text{BIC}}$ model, when applied to location L102, yields well-balanced error metrics with consistently low variance between training and testing sets. As we have

seen in the case of the AIC minimization, whose results are identical, the modest variability observed under bootstrapping suggests that the model's parameter estimates remain robust even when the underlying stochastic process is resampled. Furthermore, the post-hoc adaptive ℓ_1 -Lasso provides a straightforward enhancement, achieving a modest reduction of approximately 2–3% in error values without introducing additional dispersion, mirroring the behavior observed with previous models. This makes it an appealing option for practitioners seeking incremental performance gains without altering the model architecture.

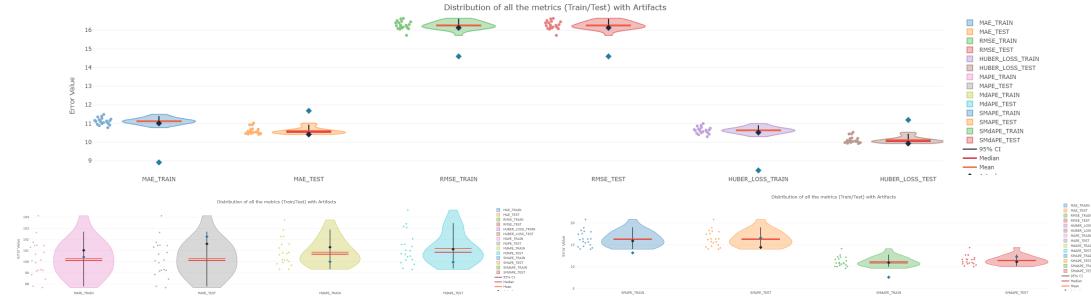


Figure 31: *Violin plots of performance metrics of Huber Loss, MAE, RMSE, MAPE, MdAPE, and their symmetric invariants For location L102 while applying ℓ_1 LSA_{BIC}. The navy blue diamond presents the adaptive ℓ_1 Lasso and the black one shows the ℓ_1 LSA_{BIC} on the initial data of L102 location. The rest came from the application of ℓ_1 LSA_{BIC} on the MEBoot samples of L102 location data.*

Figure 32 depicts the ACF and PACF of residuals and forecasting errors from three distinct modeling techniques within the ℓ_1 LSA_{BIC} framework. The **MEB_AVG** model, which represents the averaged performance across 20 MEBoot samples, demonstrates minimal autocorrelation in both ACF and PACF plots for both training and test sets. This indicates stable and generalizable performance. In contrast, the **ACTUAL** model, fitted directly on the original traffic volume data, exhibits pronounced autocorrelation patterns, particularly in the training residuals. This suggests unmet time dependencies and potential model underfitting. The post-hoc adaptive LAD Lasso, utilizing VIPs derived from the MEBoot procedure, significantly reduces autocorrelation across both phases, with residuals closely resembling white noise. This underscores the effectiveness of the adaptive re-estimation technique in producing more accurate and well-calibrated estimates. Overall, the figure illustrates that bootstrapped VIP-driven adaptive modeling yields statistically decorrelated residuals and robust, reliable forecast errors.

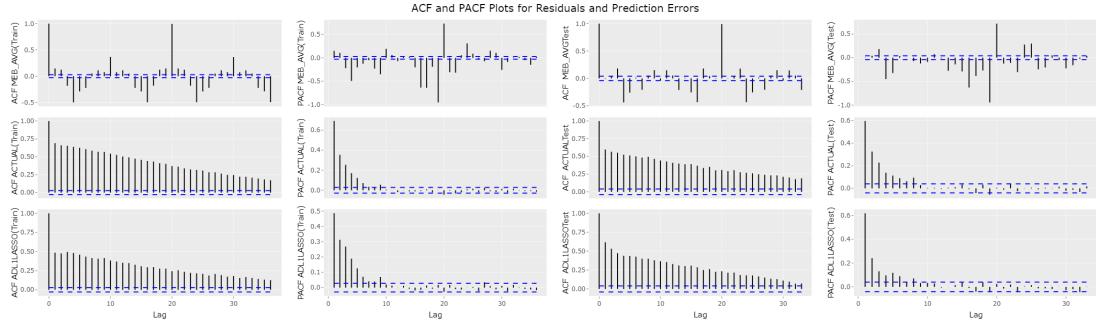


Figure 32: Autocorrelation and Partial Autocorrelation Functions of residuals and forecasting errors. The MEB_AVG , $ACTUAL$ and $ADL1LASSO$ correspond to the model averaging of the ℓ_1 LSA_{AIC} instance on 20 MEBoot synthetic dataset, on the original one and the post-hoc adaptive LAD Lasso based on the VIPs of the MEBoot experiments respectively.

Conventional Least Square Approximation of Unified LASSO (ℓ_2 LSA)

We now turn our attention to examining the performance of the ℓ_2 LSA method under both AIC and BIC minimization frameworks. Notably, both the ℓ_2 LSA_{AIC} and ℓ_2 LSA_{BIC} approaches consistently identified the same set of regressors as significant across all 20 MEBoot-sampled scenarios. This unanimously selected set of predictors is comprised of:

$$\{\text{Intercept}\} \cup \left\{ \sin\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5, 10\} \right\} \cup \left\{ \cos\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5, 15\} \right\}.$$

This selection precisely mirrors the subset of predictors identified as significant by the ℓ_1 LSA_{AIC} procedure. It is crucial to emphasize that both ℓ_2 LSA procedures, similar to ℓ_1 LSA_{AIC}, uniformly selected these features across all MEBoot-sampled scenarios, underscoring their consistent role as crucial predictors for inferring and forecasting the underlying periodic behavior. Visual confirmation of this consistent selection is provided in Figure 33.

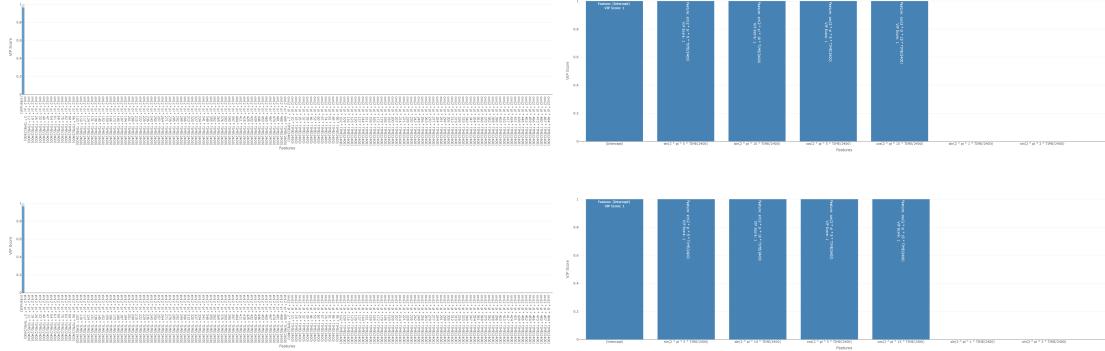
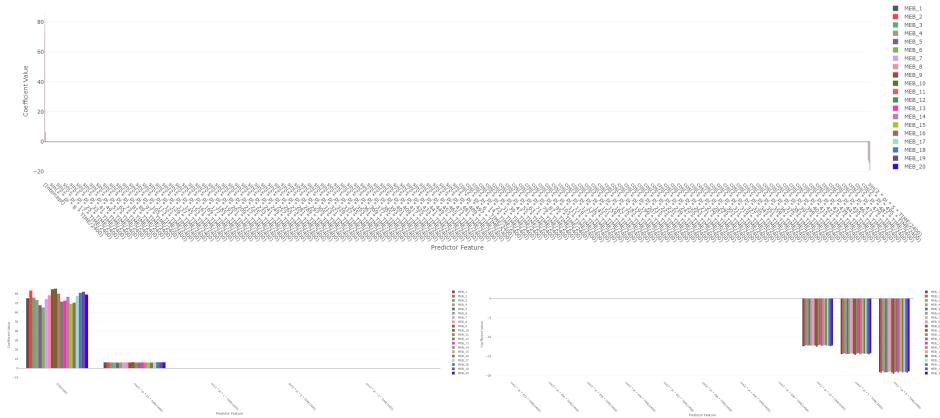


Figure 33: Variable Importance Plot for ℓ_2 LSA_{AIC} (top left and right plots) and ℓ_2 LSA_{BIC} (bottom left and right plots) on L102

Given the consistent identification of the same predictor set by the ℓ_1 LSA_{AIC} procedure, it is entirely expected that the resulting coefficient estimates from ℓ_2 LSA_{AIC} and ℓ_2 LSA_{BIC} are nearly indistinguishable. This observation is clearly evidenced in Figure 34 and in Table 10, which illustrates that the estimated coefficients across the various MEBoot samples exhibit only negligible variations. This phenomenon is anticipated because both AIC and BIC criteria, when implementing the LSA invariant model, employ the same vector of coefficients and an identical covariance matrix estimate, thereby maintaining an invariant model structure.

It should also be noted that the same set of λ values, consisting of five equally spaced points within the interval [0.02, 0.1], was utilized across all examined models. This upper bound of 0.1 was deliberately chosen, as empirical investigations demonstrated that the application of a λ value equal to or exceeding 0.1 consistently resulted in pronounced over-shrinkage of the models across all evaluated locations. Consequently, the minimization procedure must select one of these five possible λ values. Thus, it is unsurprising that both the AIC and BIC minimization processes converge on the same optimal λ value, subsequently yielding identical coefficient estimations.



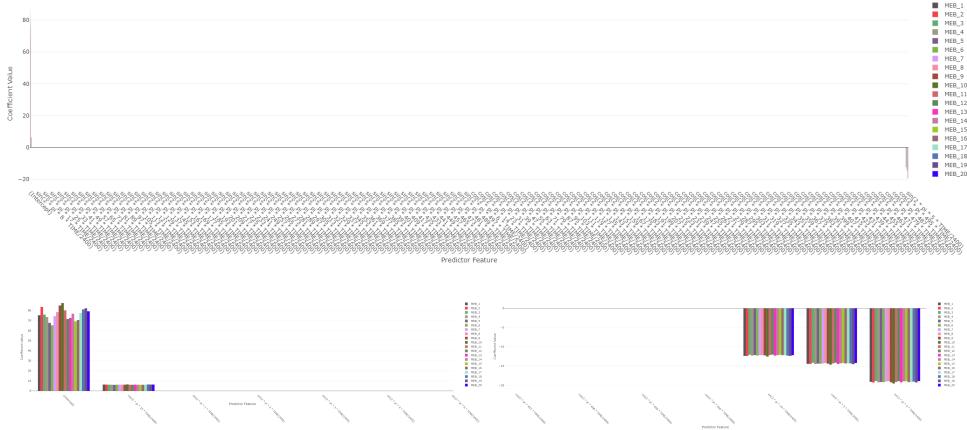


Figure 34: *Coefficient estimates for each MEBoot dataset while performing ℓ_2 LSAAC (top row) and ℓ_2 LSABIC (bottom row) on L102.*

As a direct consequence of the consistent parameter estimation and model structure, we observe a persistent issue of over-shrinkage across all MEBoot sample applications. The fitted and predicted values are depicted as distinct, equally spaced lines along the time axis, as seen in Figure 35. This over-shrinkage results in similar abnormalities being identified in both the residuals and forecasting errors. These include pronounced deviations around April 7th and 10th, corresponding to the weekend of the elections, and April 21st, which marks the victory of the local Greek basketball team. The clear seasonal pattern evident in the right-hand plots of Figure 35 for the residuals unequivocally indicates that the coefficients were overly suppressed towards zero, leading to the omission of substantial periodic components that should have been captured by the model.

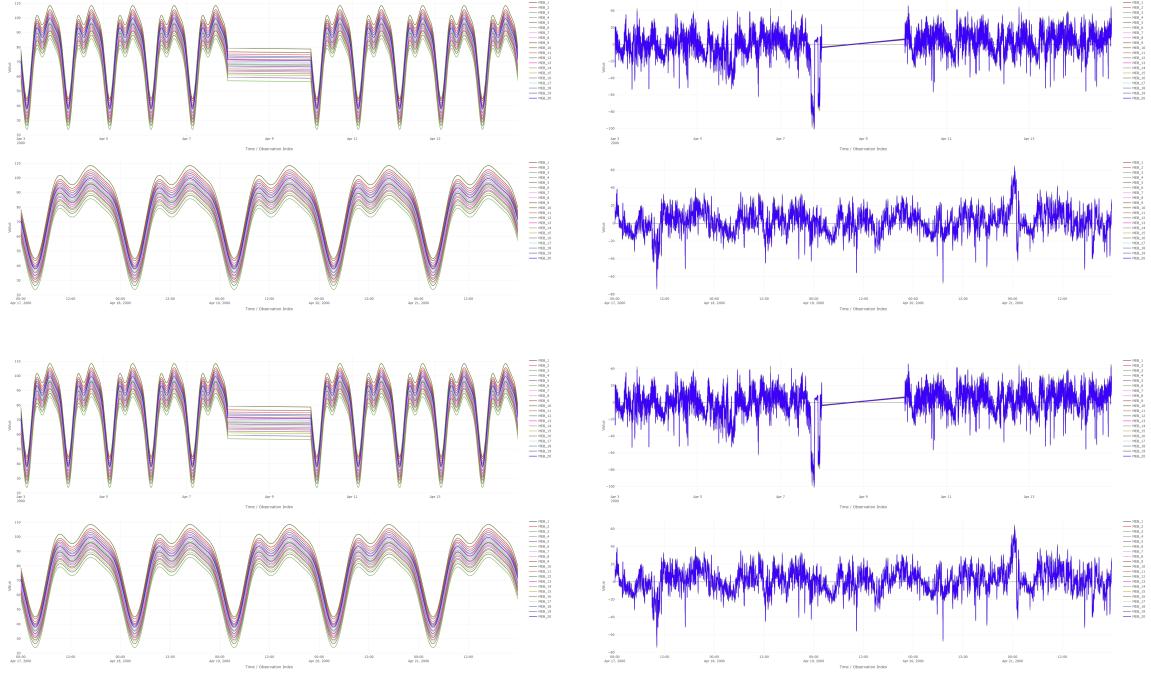


Figure 35: The top quartet of panels refers to $\ell_2\text{LSA}_{\text{AIC}}$, and the bottom quartet to $\ell_2\text{LSA}_{\text{BIC}}$. In each group of plots, the first two panels (top-left and top-right) depict the fitted values and residuals for the training period (April 3rd–14th), respectively. The subsequent two panels (bottom-left and bottom-right) illustrate the forecasted values and corresponding errors during the test period (April 18th–22nd), based on the application of the respective method on location L102.

The post-hoc analysis, which employed the VIPs from the $\ell_2\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{BIC}}$ as initial weights for the adaptive least absolute deviations (LAD) LASSO, yielded identical results in both cases. This consistency holds when comparing the performance against the application of the corresponding model to the original traffic volume data and the model averaging produced by its application with the 20 MEBoot samples. Figure 36 demonstrates that both the model averaging and the model applied directly to the original traffic volume dataset effectively identify the coefficients with the highest values using the post-hoc variable selection technique. Furthermore, the values predicted by model averaging and the model applied to the original dataset are highly congruent with those estimated by the post-hoc model for the specified predictors.

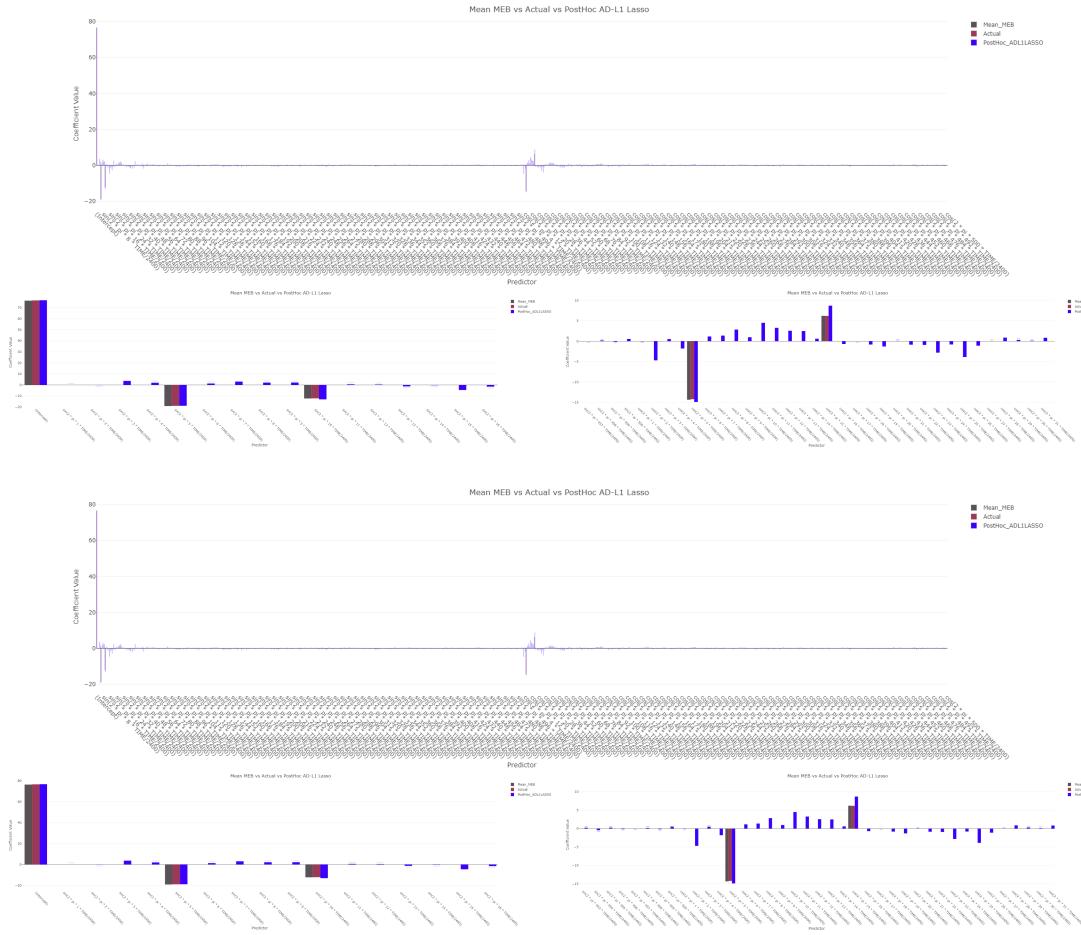


Figure 36: The top set of three plots corresponds to $\ell_2\text{LSA}_{\text{AIC}}$, and the bottom set to $\ell_2\text{LSA}_{\text{BIC}}$. These bar plots provide a comparative analysis of average coefficients obtained from maximum entropy bootstrap samples, the actual sample, and the actual samples performing adaptive ℓ_2 -Lasso with weights constructed by VIPs for L102, under both criterion minimization processes.

The performance of the $\ell_2\text{LSA}$ estimator under both AIC and BIC model selection criteria demonstrates remarkable consistency, with negligible differences observed between the two formulations across all key metrics. As such, we present a unified evaluation encompassing both approaches. Under model averaging over 20 MEBoot samples, the $\ell_2\text{LSA}$ achieves solid, though not leading, performance. In the training scenario, the method yields HL, MAE, and RMSE values of 11.45, 11.94, and 16.41, respectively, which are slightly higher than those of the best-performing $\ell_2\text{MTE}$ models. On the testing set, the same metrics remain stable at 10.64, 11.13, and 16.41, again indicating reliable but non-optimal error reduction. In terms of percentage-based errors, the $\ell_2\text{LSA}$ records MAPE and MdAPE values of 99.39 and 102.29 in training, and 99.39 and 102.04 in testing, figures that suggest modest deviation from the minimum and signal lower efficiency compared to

the MTE variants. SMAPE and SMdAPE values are also moderate: 18.22 and 12.53 in training, and 18.22 and 12.20 in testing. These results affirm the model's robustness, albeit at the cost of predictive sharpness. When directly applied to the original dataset, the ℓ_2 LSA produces marginal improvements. For instance, HL and MAE in the test set are slightly reduced to 10.49 and 11.00, respectively, and RMSE remains at 16.17. Percentage metrics such as MAPE and MdAPE improve subtly but remain suboptimal at 100.12 and 101.53. These findings suggest that while the model is inherently stable, it does not exploit the full potential of the data structure in minimizing predictive errors. Figure 37, mirroring the analysis conducted for ℓ_1 LSA_{AIC}, offers a detailed and statistically informative visualization of the error characteristics associated with the ℓ_2 LSA_{AIC} and ℓ_2 LSA_{BIC} approaches. The violin plots, augmented by embedded boxplots and diamond markers, encapsulate both distributional variability and central tendencies across a range of error metrics for both training and testing phases. Specifically, the plots for MAE, RMSE, and Huber Loss exhibit narrow, symmetric shapes with minimal dispersion, indicative of consistent and robust performance across MEBoot samples. In contrast, MAPE and MdAPE demonstrate moderately broader distributions, reflecting an increased sensitivity to percentage-based errors; nevertheless, the alignment of their mean and median values confirms the reliability of their central tendency estimates. A comparison of training versus testing metrics reveals the expected slight increase in variability for test data, without substantial shifts in central location, suggesting effective generalization and the absence of overfitting. Furthermore, the placement of the navy blue and black diamonds—representing model performance on the original data—within the dense central regions of the violins confirms that the initial results are statistically representative and not anomalous. Collectively, these findings underscore the ℓ_2 LSA method's capacity for stable, unbiased predictive performance and robust generalization across bootstrapped samples, affirming its methodological soundness and practical applicability.

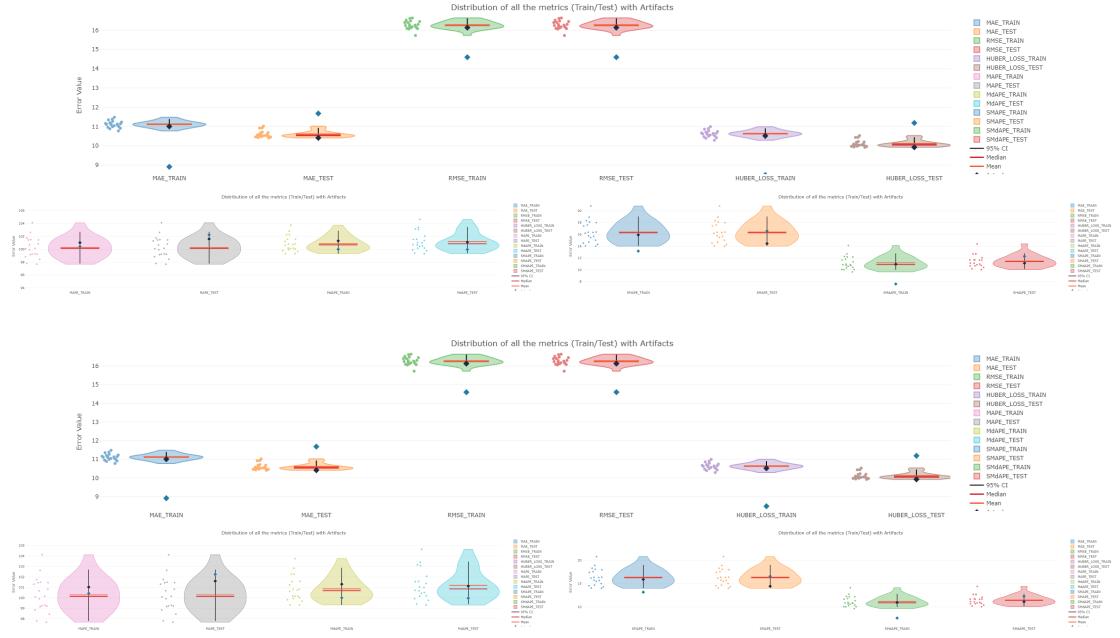


Figure 37: These two sets of violin plots illustrate the performance metrics—Huber Loss, MAE, RMSE, MAPE, MdAPE, and their symmetric invariants—for location L102. The upper triad corresponds to ℓ_2 LSA_{AIC} and the lower triad to ℓ_2 LSA_{BIC}. The navy blue diamond denotes the adaptive ℓ_2 Lasso, while the black diamond represents the corresponding method's performance on the initial L102 location data. All other data points originate from the application of the corresponding method on the MEBoot samples of L102 location data.

Consequently, and as previously observed in the residual plots depicted in Figure 35, we can corroborate the persistence of periodic structures within the residuals of the model averaging across the 20 MEBoot samples. The sinusoidal patterns identified in the ℓ_1 LSA_{AIC} scenario similarly suggest that the model was unable to fully account for an underlying seasonal component. However, the test data, while not strictly resembling white noise, appears to have captured and managed the seasonality to a satisfactory extent. The ℓ_2 LSA model, when applied to the original dataset, along with the post-hoc variable selection procedure based on the corresponding VIPs, appears to have effectively addressed the seasonality. This treatment has, in turn, rendered the residuals more amenable to a refined statistical model capable of further capturing the regressive serial dependencies present in the data, as shown in Figure 38.

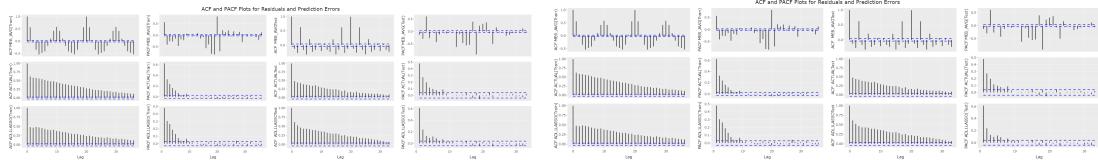


Figure 38: *Partial and Conventional Autocorrelation Function of ℓ_2 LSA_{AIC} (left plot) and ℓ_2 LSA_{BIC} (right plot)*

LSA-invariants instability on other locations

This section presents an in-depth analysis of model behavior at two distinct loop detector sites: a site with a relatively low incidence of singularities during the observation period and a more complex site, **L106 (E)**, located at the intersection of Alexandras Avenue and Charilaou Trikoupi Street, an area historically prone to political events in April. At location **L106 (E)**, the LSA-invariant models exhibit notable instability, a deviation from the more stable variable selection observed at sites with fewer data singularities. As illustrated in the heatmap (Figure ??) and detailed in Table ??, most methods demonstrate a consistent variable selection process irrespective of the AIC/BIC minimization criteria, with the notable exception of the LSA-invariant approaches. Our focus will be on the ℓ_2 LSA models, as their behavior reveals more pronounced differences across the various information criterion minimization processes compared to the ℓ_1 LSA models.

Specifically, the ℓ_2 LSA method, when minimizing the AIC, consistently selects 100% of the predictors from the set:

$$\{(\text{Intercept})\} \cup \left\{ \sin\left(2\pi \cdot x \frac{t}{2400}\right) \right\}_{x \in \{2, 4, 5, 8, 9, 10, 15, 16, 20\}} \cup \left\{ \cos\left(2\pi \cdot x \frac{t}{2400}\right) \right\}_{x \in \{4, 5, 10, 15, 17, 20, 25\}}$$

Conversely, minimizing the BIC for the same model results in the selection of less than half of these predictors. As evidenced in Table ??, Figure ??, and Figure 39, the selected regressors under BIC minimization are:

$$\{(\text{Intercept})\} \cup \left\{ \sin\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5, 10\} \right\} \cup \left\{ \cos\left(2\pi \cdot x \frac{t}{2400}\right) \mid x \in \{5, 15\} \right\}.$$

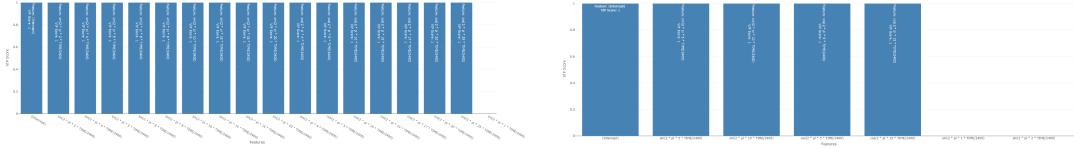


Figure 39: *Bar plots of Variable Inclusion Probabilities for ℓ_2 LSA_{AIC} (left panel) and ℓ_2 LSA_{BIC} (right panel) applied to 20 MEBoot samples based on L106 location.*

Naturally, this divergence in selected predictors leads to substantial differences in the coefficient estimates between the ℓ_2 LSA_{AIC} and ℓ_2 LSA_{BIC} models, primarily through the elimination of numerous

coefficients under BIC minimization, as explicitly depicted in Figure 40. While the intercept and the coefficient for $\cos(2\pi \cdot 15 \frac{t}{2400})$ remain highly similar across both models, $\ell_2\text{LSA}_{\text{BIC}}$ estimates the remaining common coefficients with considerably reduced amplitudes.

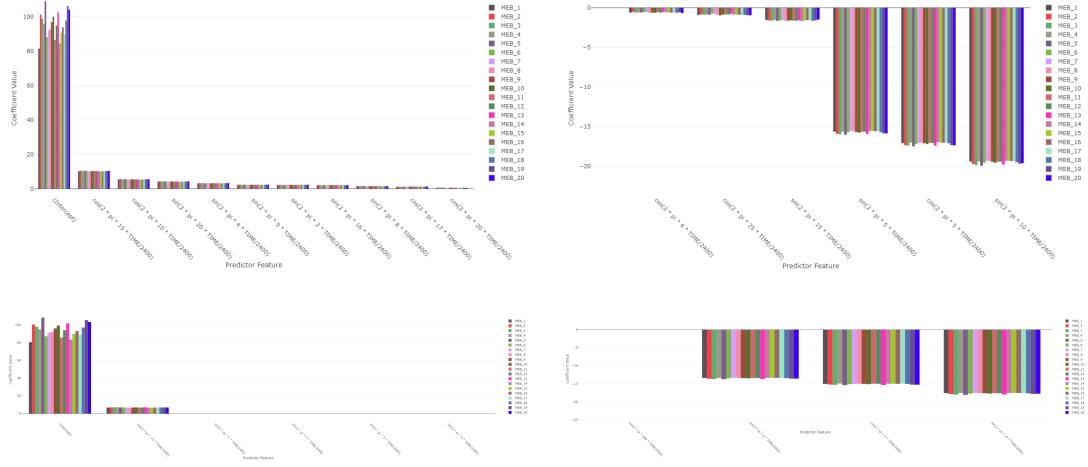


Figure 40: *Coefficient estimates for $\ell_2\text{LSA}_{\text{AIC}}$ (left pair of plots) and $\ell_2\text{LSA}_{\text{BIC}}$ (right pair of plots) applied to 20 MEBoot samples based on L106 location.*

The $\ell_2\text{LSA}_{\text{BIC}}$ model, being significantly simpler due to its sparser predictor set, consequently demonstrates a limited capacity to capture intricate seasonal patterns. Interestingly, even the more complex $\ell_2\text{LSA}_{\text{AIC}}$ model does not fully eliminate periodicity, as evidenced by the residual plots in Figure 41. This suggests that despite its parsimony, the $\ell_2\text{LSA}_{\text{BIC}}$ model achieves a comparable level of accuracy to that of the more complex $\ell_2\text{LSA}_{\text{AIC}}$ model.

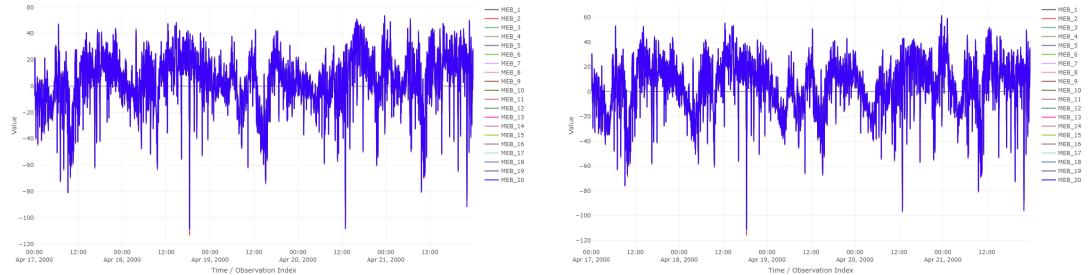


Figure 41: *Residual plots on the test data for $\ell_2\text{LSA}_{\text{AIC}}$ (left plot) and $\ell_2\text{LSA}_{\text{BIC}}$ (right plot) applied to 20 MEBoot samples based on L106 location.*

Following the application of the models to 20 MEBoot samples, their testing on the original data, and the subsequent incorporation of the resulting VIPs as weights in the post-hoc variable selection

procedure, a clear pattern emerges. The AIC minimization consistently yields models with similar coefficient values for the commonly selected predictors, with the sole exception of $\cos(2\pi \cdot 20 \frac{t}{2400})$. Conversely, BIC minimization leads to models that also share similar coefficients, albeit with the exception of the predictor $\cos(2\pi \cdot 15 \frac{t}{2400})$, as depicted in Figure 42.

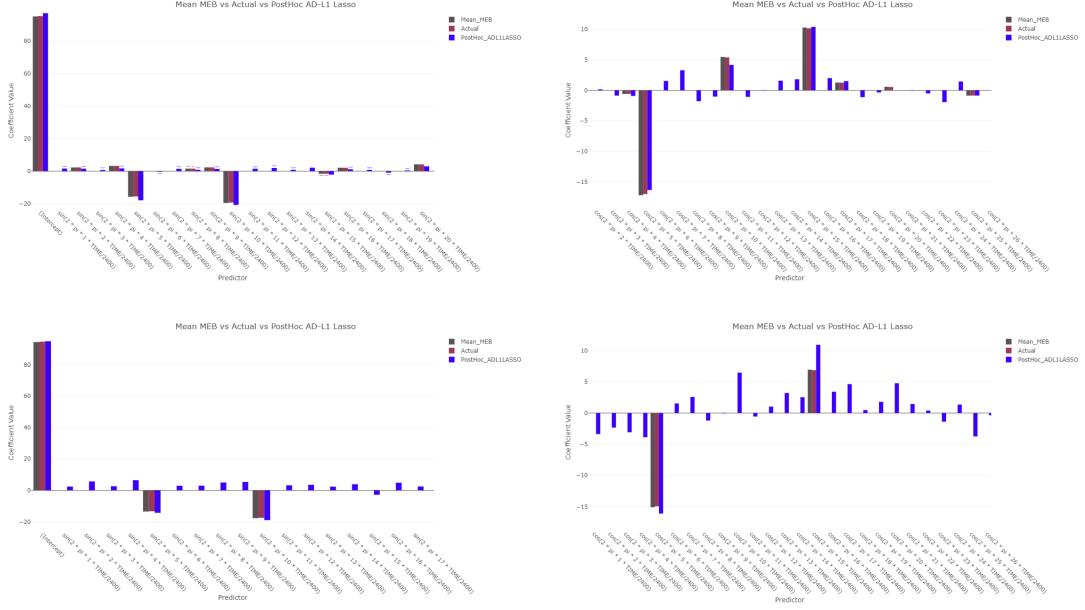


Figure 42: Coefficient estimates for ℓ_2 LSA_{AIC} (upper pair of plots) and ℓ_2 LSA_{BIC} (bottom pair of plots) applied to 20 MEBoot samples, the original traffic volume data at L106, and the corresponding post-hoc variable selection technique.

Significant variations across competing model classes may be seen when looking at error measures like as MAE, RMSE, and Huber loss. While no one model dominates across all measures, the ℓ_2 LSA strategy under AIC minimization, which picks a bigger range of relevant predictors, generally outperforms its BIC counterpart. This improved efficacy is most clearly reflected in the lowering of important error metrics by around 4 to 9 units, which is graphically verified in Figure 43 and numerically in table 20. Using MEBoot sampling, ℓ_2 LSA_{AIC} achieves HL, MAE, and RMSE values of 18.02, 18.51, and 24.62 in training and 15.66, 16.15, and 24.62 in testing. These scores are relatively high, indicating consistent but not superior prediction accuracy. The adaptive LASSO refinement using VIPs achieves the best performance, with HL, MAE, and RMSE dropping to 12.44, 12.89, and 20.06 in training and 13.80, 13.92, and 20.06 in testing, respectively. This results in significant gains of 6-9 units across all metrics compared to the raw MEB output. In comparison, the ℓ_2 LSA_{BIC} model performs similarly at the MEB level, with HL, MAE, and RMSE of 19.11, 19.60, and 25.52 in training and 16.64, 17.13, and 25.52 in testing. Post-hoc adaptive LASSO improves performance, achieving the lowest RMSE (19.93) in training and testing. Nonetheless, its improvements are somewhat fewer than those of AIC, highlighting the value of more comprehensive model selection criteria. Percentage error measurements provide a detailed picture. MAPE and

MdAPE reach 98.74 and 102.98 in MEB testing, whereas SMAPE and SMdAPE are 22.17 and 14.33, respectively. Adaptive LASSO improves symmetric errors to 21.64 and 14.75. Under BIC reduction, MAPE and MdAPE stand at 99.00 and 103.99, whereas SMAPE and SMdAPE exhibit somewhat worse performance at 23.17 and 15.14, respectively. Although percentage-based measures reveal no difference between AIC and BIC formulations, the general trend favors the AIC-driven strategy. Figure 43 compares the error distributions for ℓ_2 LSA_{AIC} and ℓ_2 LSA_{BIC} across training and testing sets. The violin plots clearly illustrate that the AIC-minimized model has narrower distributions and more concentrated central tendency, especially for HL and RMSE. Notably, the diamond markers indicating performance on the original dataset are centered inside the majority of the distributions, confirming that those values are consistent and statistically representative. These findings highlight the methodological benefit of adopting AIC as a selection criterion in ℓ_2 LSA estimation, resulting in increased predictive performance and greater generalization as confirmed across both sampling-based and original data assessments.

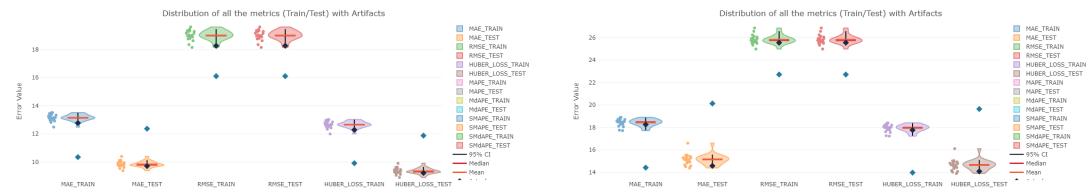


Figure 43: *Violin Plots of MAE, RMSE, and Huber loss on both test and train data for ℓ_2 LSA_{AIC} (left plot) and ℓ_2 LSA_{BIC} (right plot) applied to 20 MEBoot samples based on L106 location.*

A similar irregularity in the behavior of the LSA models is observed at location **L104 (D)**, where Panormou Avenue, a major thoroughfare, intersects with our primary avenue of interest. This location experienced more pronounced irregularities in the testing data compared to L106 (likely due to its proximity to the local basketball club, which secured a championship on April 21st, 2025). Unlike L106, which exhibits singularities in both training and testing data, the irregularities at L104 were predominantly observed in the testing data.

Comparison across locations

Let us make a comparison of those methods across locations. Applying a meticulous analysis on the model averaging over 20 the maximum entropy bootstrap samples, reveals several critical observations regarding the models' performance. Firstly, concerning overall sparsity and coefficient magnitudes, the plots below (49, 50, 51, 52, 45, 46, 47, and 48) uniformly demonstrate that the vast majority of estimated coefficients converge to values remarkably close to zero, irrespective of the specific model type or the information criterion employed. This pervasive sparsity strongly suggests that only a select subset of the potential predictors significantly contributes to explaining the observed variations in traffic volume across the analyzed sites. Amidst this widespread sparsity, a few key predictors consistently exhibit discernible non-zero coefficient values, occasionally reaching substantial magnitudes for certain locations. Notably, the intercept term and specific low-frequency components, such as those associated with terms like $\cos(2\pi \cdot 5 \frac{t}{2400})$, $\sin(2\pi \cdot 5 \frac{t}{2400})$, $\cos(2\pi \cdot 15 \frac{t}{2400})$, and $\sin(2\pi \cdot 10 \frac{t}{2400})$, appear to be the most frequently selected and profoundly

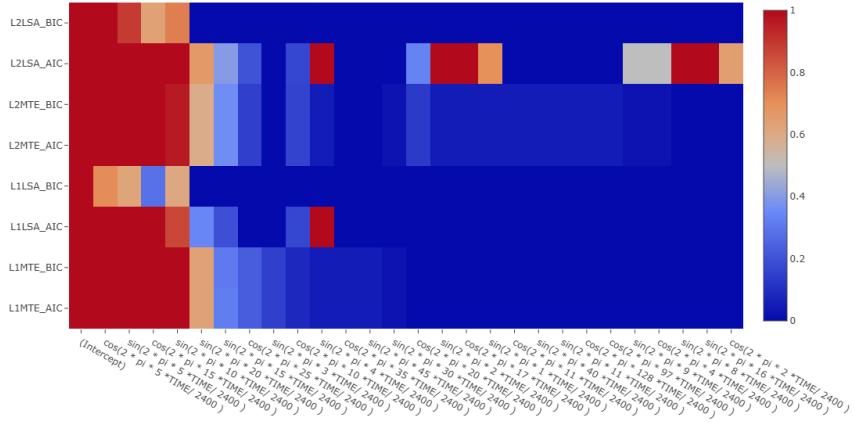


Figure 44: Enter Caption

influential regressors. Furthermore, for those limited predictors that do manifest non-zero coefficients, their estimated values generally display relative consistency across the different locations (L101-L108) within each distinct model and criterion combination. This consistency is visually corroborated by the largely uniform, non-zero value estimation across the different loop detectors Secondly, the impact of AIC versus BIC minimization is starkly evident. The 44 heatmap unequivocally illustrates that BIC minimization consistently engenders a higher degree of sparsity, characterized by a predominance of cooler colors (e.g., blue cells) signifying lower Variable Inclusion Probabilities across the ℓ_1 LSA_{BIC}, ℓ_2 LSA_{BIC}, ℓ_1 MTE_{BIC}, and ℓ_2 MTE_{BIC} models, when compared to their AIC-minimized counterparts. This behavior is entirely anticipated, given BIC's more stringent penalty on model complexity (i.e., the number of predictors), which inherently promotes the selection of more parsimonious models. Conversely, AIC minimization tends to retain a significantly larger ensemble of predictors, manifested by a higher frequency of warmer colors (e.g., red/orange cells) indicating elevated VIPs. This suggests that AIC-based models aim to capture more intricate patterns or endeavor to account for a greater proportion of the observed variance, even if some of the retained predictors might be considered less critical under BIC's stricter regularization. Thirdly, a distinct comparison emerges between the ℓ_1/ℓ_2 LSA and ℓ_1/ℓ_2 MTE model families. The 44 heatmap highlights that ℓ_1 LSA_{AIC} and ℓ_2 LSA_{AIC} models exhibit instances where certain predictors show high VIPs, while their corresponding BIC-minimized versions within the same LSA family display drastically reduced VIPs. This observation underscores a significant "instability" in the variable selection of LSA-invariant models, particularly concerning their sensitivity to the choice of information criterion. A compelling example is observed with terms such as $\sin(2\pi \cdot 20 \frac{t}{2400})$ and $\cos(2\pi \cdot 25 \frac{t}{2400})$, which register very high VIPs for ℓ_2 LSA_{AIC} but approach near-zero VIPs for ℓ_2 LSA_{BIC}. In contrast, the ℓ_1 MTE and ℓ_2 MTE models demonstrate a notably more consistent variable selection process between AIC and BIC minimization. This is inferred from the visually similar color patterns across the AIC and BIC rows for MTE models in 44, implying that the MTE regularization technique may confer a more robust variable selection, rendering it less susceptible to the specific information criterion employed. Finally, concerning specific predictor importance derived from 44, the intercept term, along with $\sin(2\pi \cdot 5 \frac{t}{2400})$, and $\cos(2\pi \cdot 5 \frac{t}{2400})$, consistently exhibit exceptionally high VIPs (approaching 1, depicted by dark red) across nearly all

model configurations and information criteria. This finding strongly posits that a constant baseline offset and a fundamental time dependent cycle are the most critical and universally recognized components in effectively modeling traffic volume. Other harmonic terms, such as $\sin(2\pi \cdot 10 \frac{t}{2400})$ and $\cos(2\pi \cdot 15 \frac{t}{2400})$, also demonstrate high VIPs for numerous models, signifying the presence of additional significant periodic components within the traffic patterns. However, the VIPs for these terms generally display a greater susceptibility to the particular information criterion and model type utilized. Furthermore, higher frequency terms typically exhibit lower VIPs, especially under BIC minimization. This suggests that their individual contributions to explaining traffic volume are often less pronounced, or that more parsimonious models preferentially exclude them to mitigate the risk of overfitting. In summation, the comprehensive analysis of these estimations reveals that while a fundamental set of predictors, including the intercept and a specific low-frequency harmonics features, are universally deemed important, the choice between AIC and BIC minimization significantly influences the ultimate composition of selected predictors and, consequently, the overall complexity of the model, particularly for the LSA family. Conversely, MTE models appear to offer a more stable and less criterion-dependent approach to variable selection. The pervasive sparsity observed in the coefficient estimates collectively suggests that a parsimonious model structure is likely sufficient for accurately capturing the essential dynamics of traffic volume across the diverse locations. To achieve more precise conclusions, a more granular examination of the coefficients estimation across the different locations, specifically focusing on predictors with non-zero coefficients, or a direct numerical tabulation of these values, would prove highly advantageous.

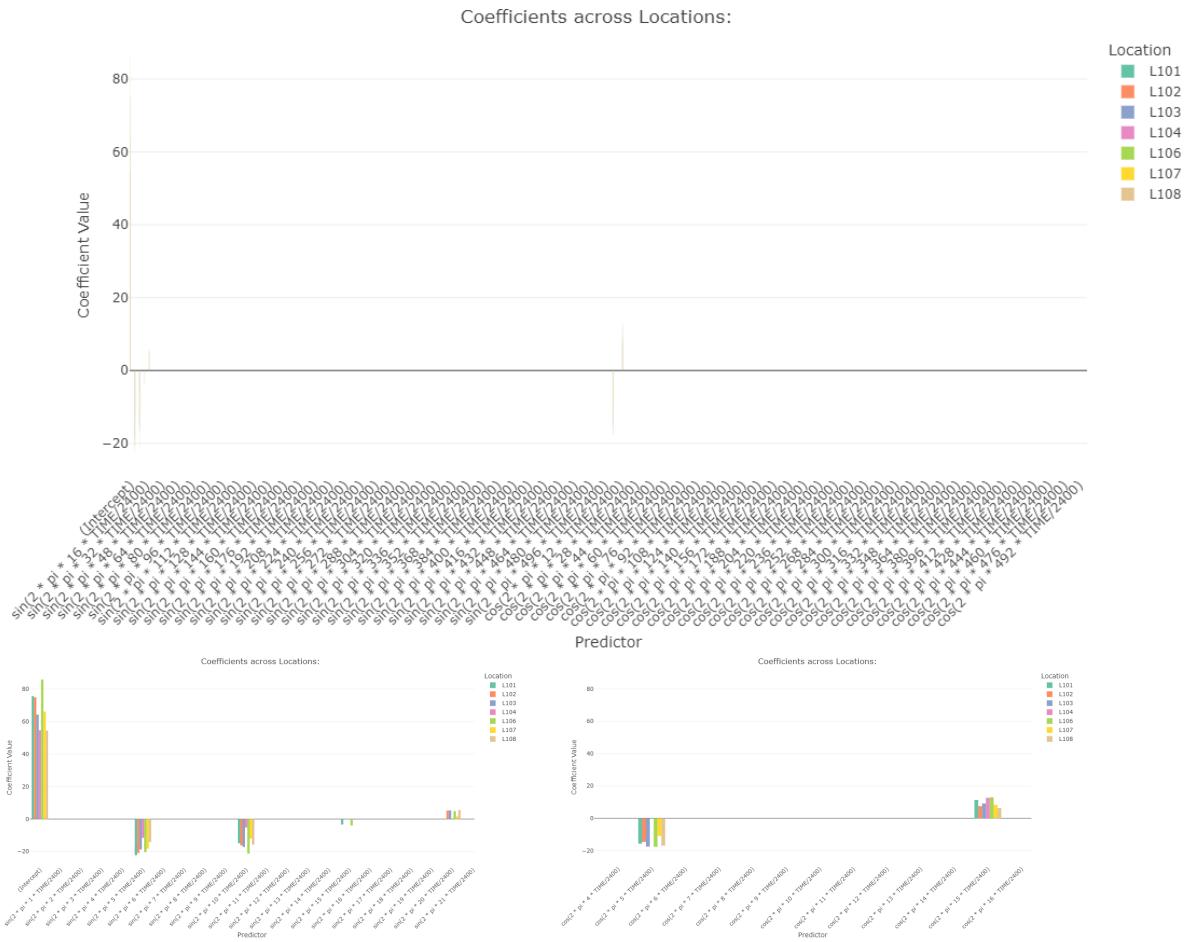


Figure 45: *Coefficients estimations of ℓ_1 MTE_{AIC} across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

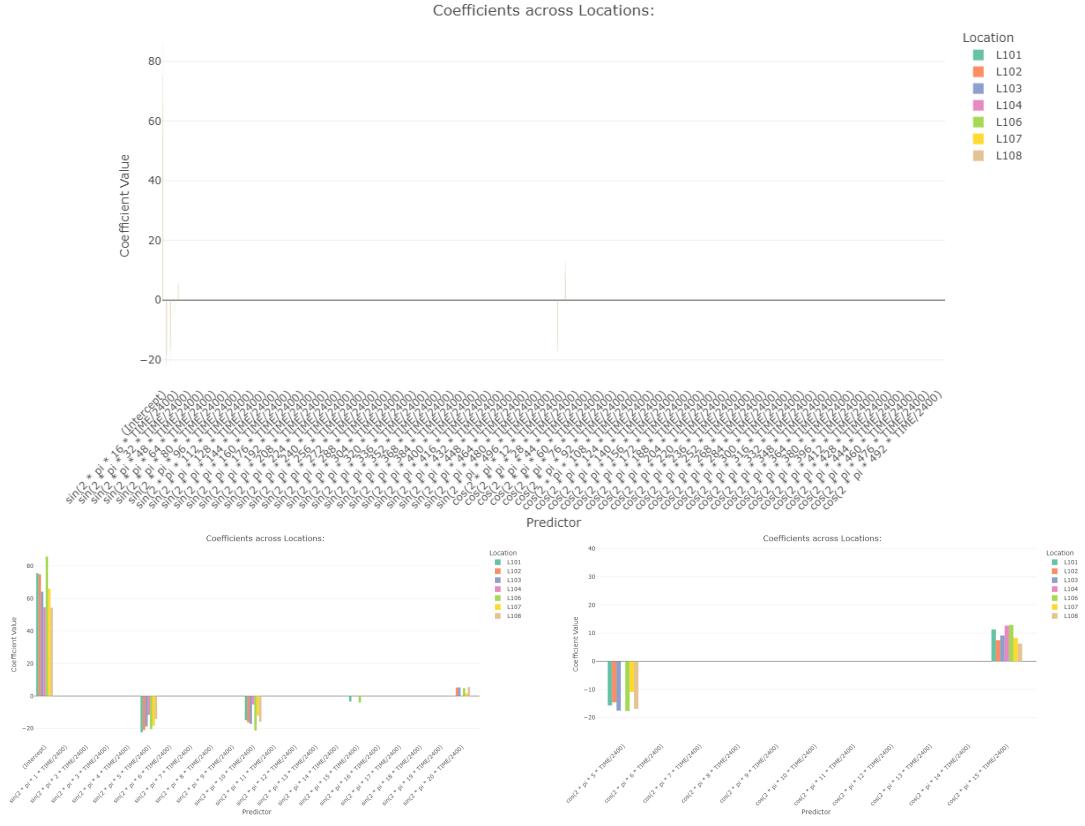


Figure 46: *Coefficients estimations of ℓ_1 MTEBIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

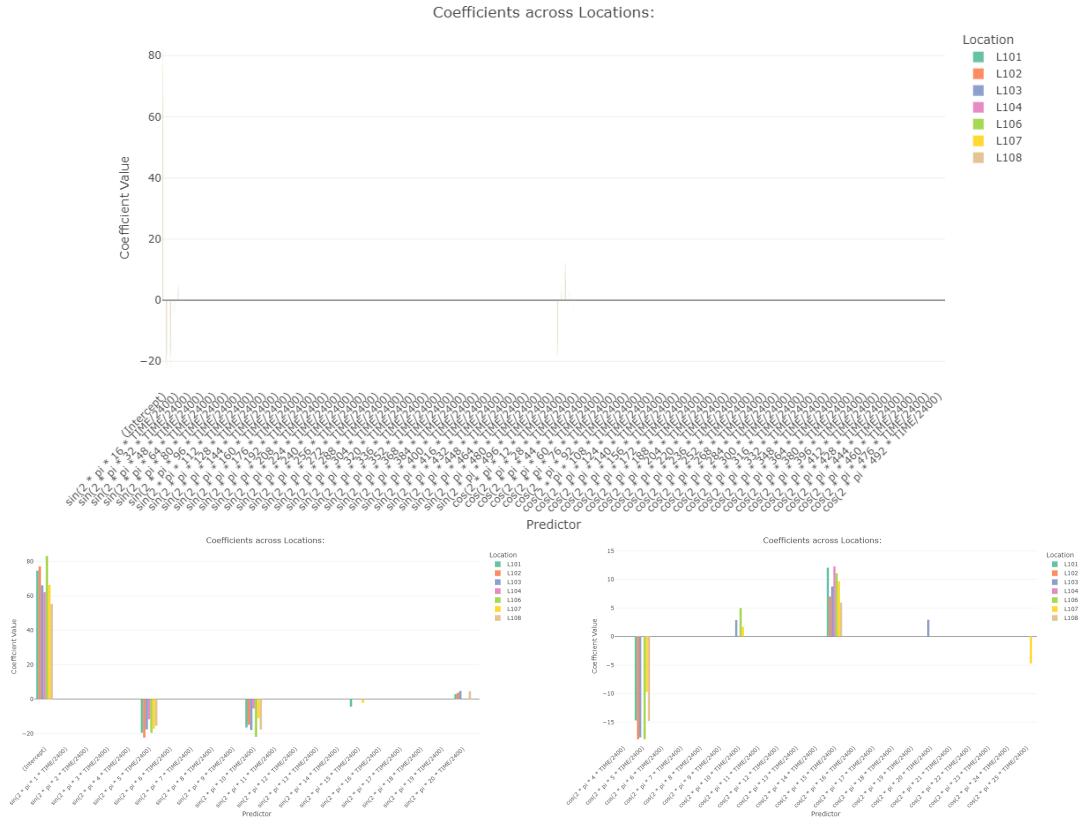


Figure 47: Coefficients estimations of ℓ_2 MTE_{AIC} across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.

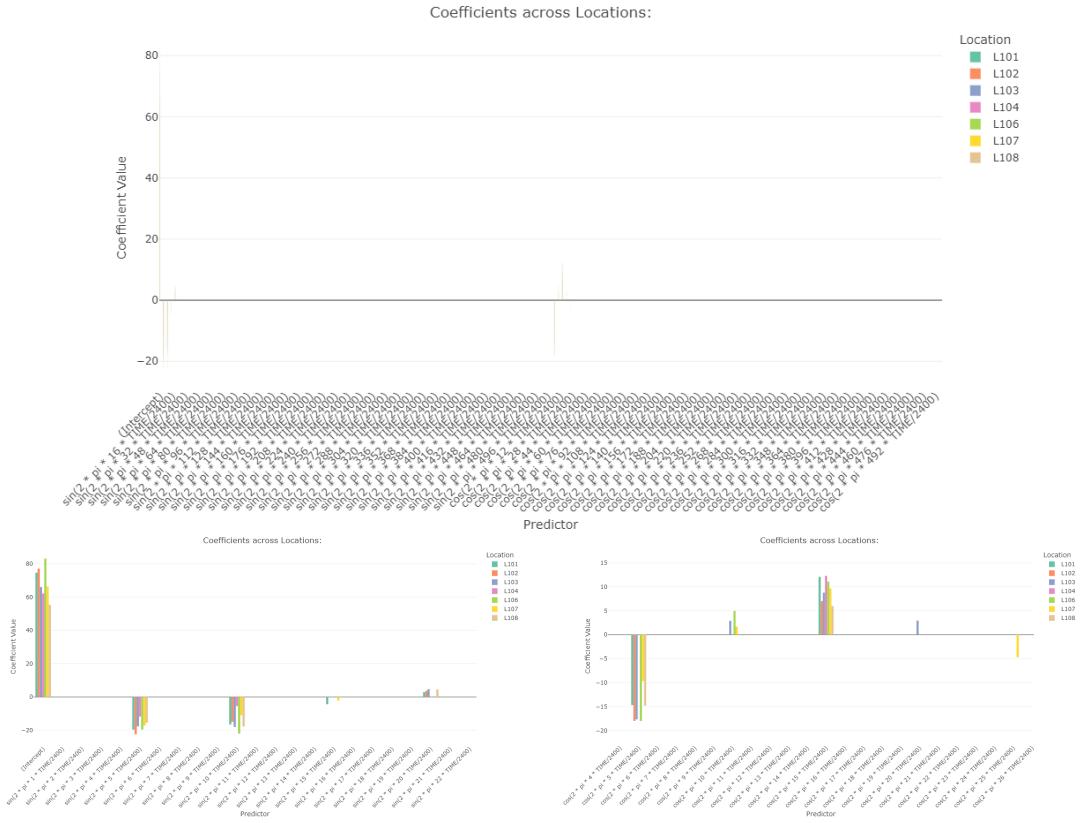


Figure 48: *Coefficients estimations of ℓ_2 MTEBIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

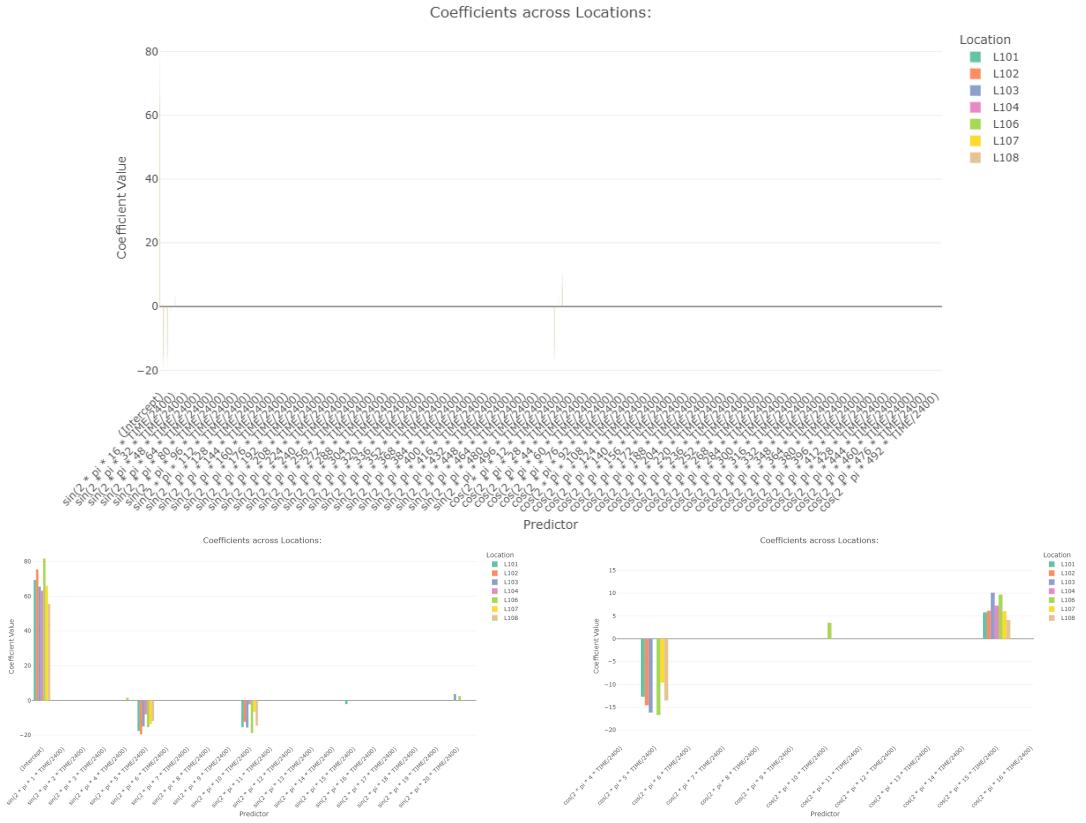


Figure 49: *Coefficients estimations of ℓ_1 LSAAIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

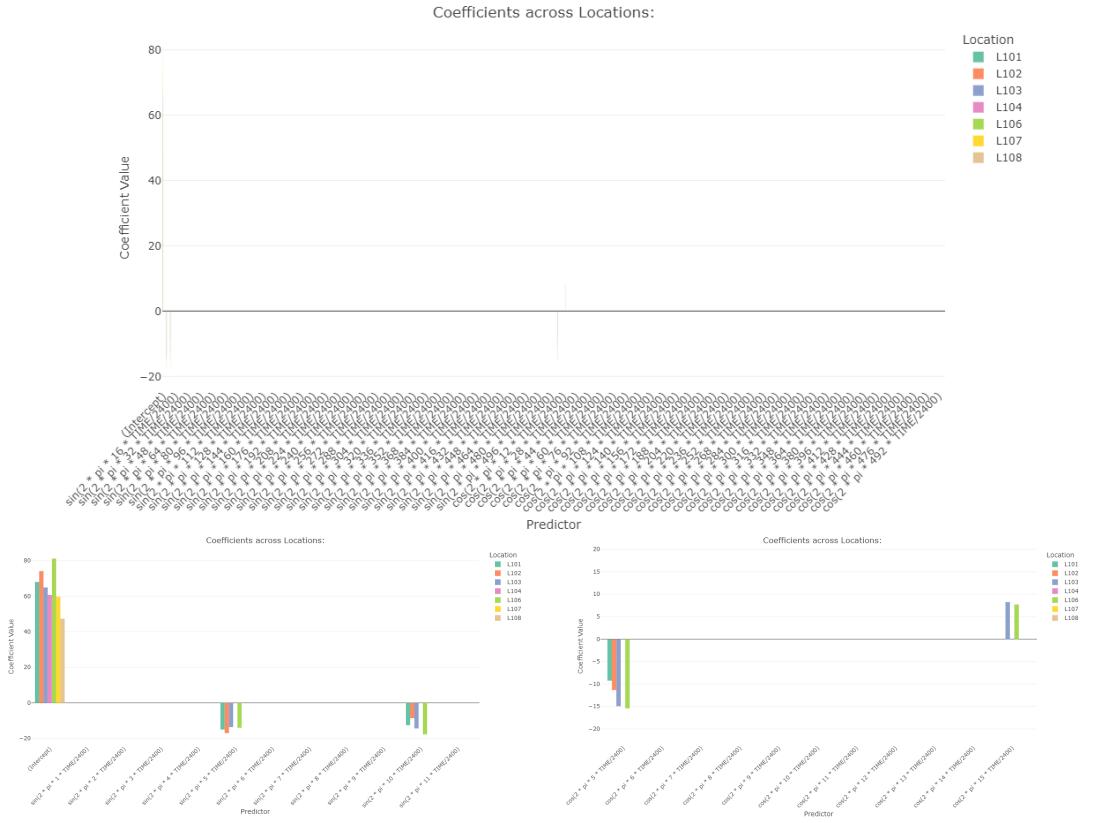


Figure 50: *Coefficients estimations of ℓ_1 LSA_{BIC} across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

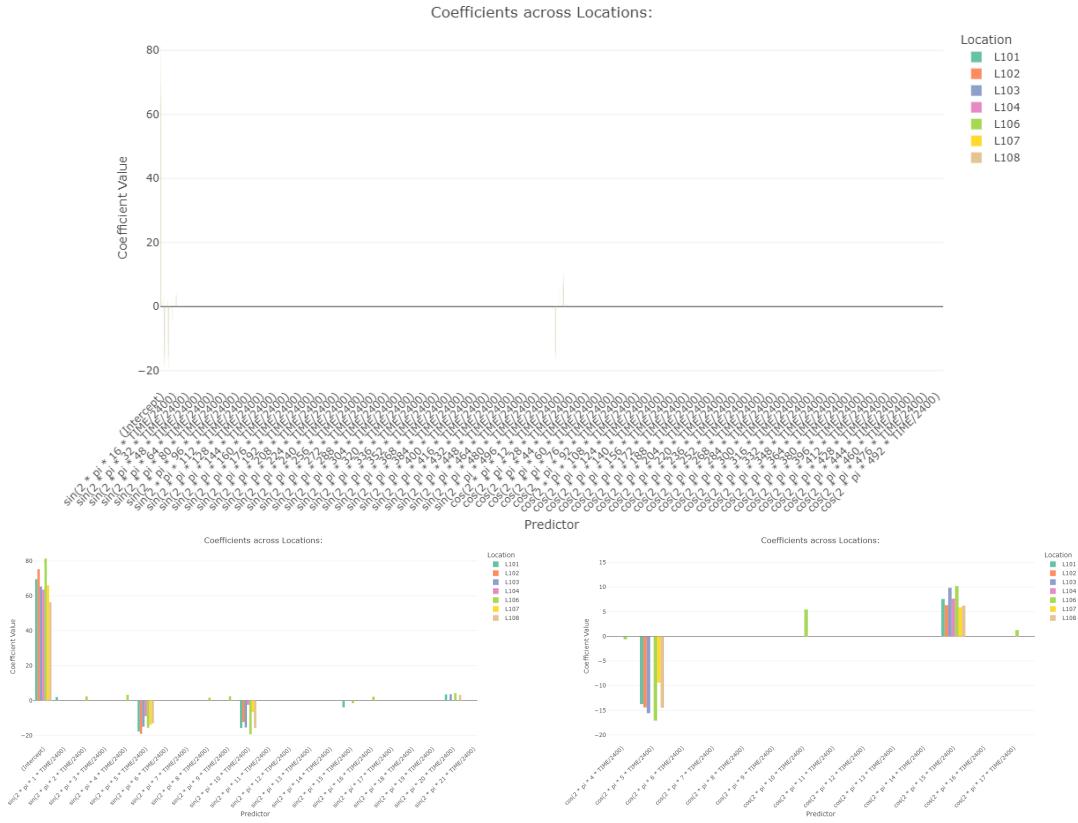


Figure 51: *Coefficients estimations of ℓ_2 LSAAIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

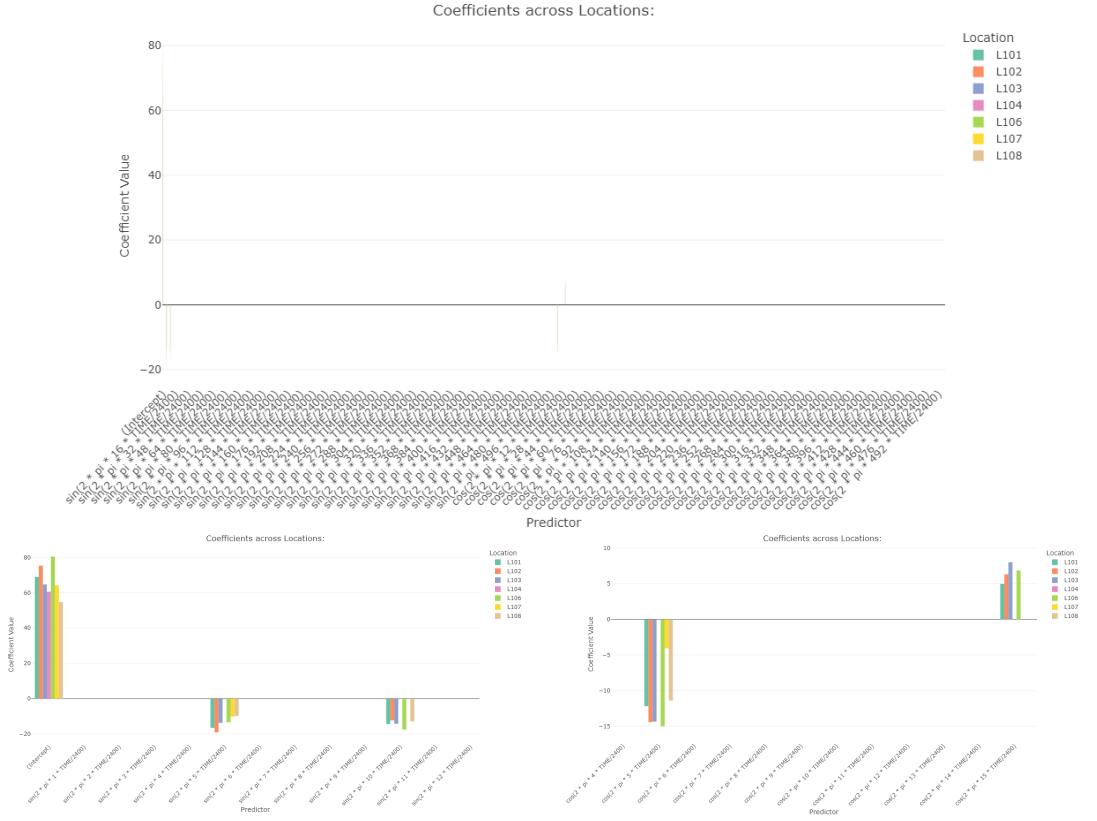


Figure 52: *Coefficients estimations of ℓ_2 LSA BIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

The comparative analysis of the models applied to the original dataset yielded nearly identical outcomes, with only minor discrepancies observed in certain parameter estimates. Consequently, we will focus our comparison on the post-hoc adaptive ℓ_1 LASSO model, which is implemented on the original dataset and leverages the VIPs derived from the application of the models across 20 MEBoot samples for each location. The 44 heatmap provides critical insights into the harmonic frequencies deemed significant by the adaptive ℓ_1 Lasso regularizer across various model configurations and information criteria. Notably, as we can verify in the plots below (57, 58, 59, ??, 53, 54, 55, and 56) the intercept term, along with the low-frequency harmonic components $\sin(2\pi \cdot 5 \frac{t}{2400})$ and $\cos(2\pi \cdot 5 \frac{t}{2400})$, consistently exhibit uniformly exceptionally high VIPs, depicted by deep red in the heatmap. This emphatically underscores their foundational role in accurately capturing the constant offset and primary periodic patterns inherent in traffic volume data. Beyond these ubiquitous terms, other harmonic components, such as $\sin(2\pi \cdot 10 \frac{t}{2400})$ and $\cos(2\pi \cdot 15 \frac{t}{2400})$, also frequently display elevated VIPs for numerous models, indicating the presence of additional significant periodicities contributing to the temporal dynamics. However, the inclusion probabilities for these secondary terms demonstrate greater variability and susceptibility to the specific information cri-

terion (AIC versus BIC) and the chosen model type. Conversely, higher frequency terms, generally exhibit substantially lower VIPs, particularly under BIC minimization. This indicates that their contribution to explaining traffic volume variance is often marginal, leading to their preferential exclusion by more parsimonious models to prevent overfitting and enhance generalization capabilities. Thus, the heatmap rigorously identifies a core set of low-frequency harmonics as universally significant, while providing a nuanced view of the adaptive ℓ_1 Lasso's variable selection across different complexities and model stabilities.

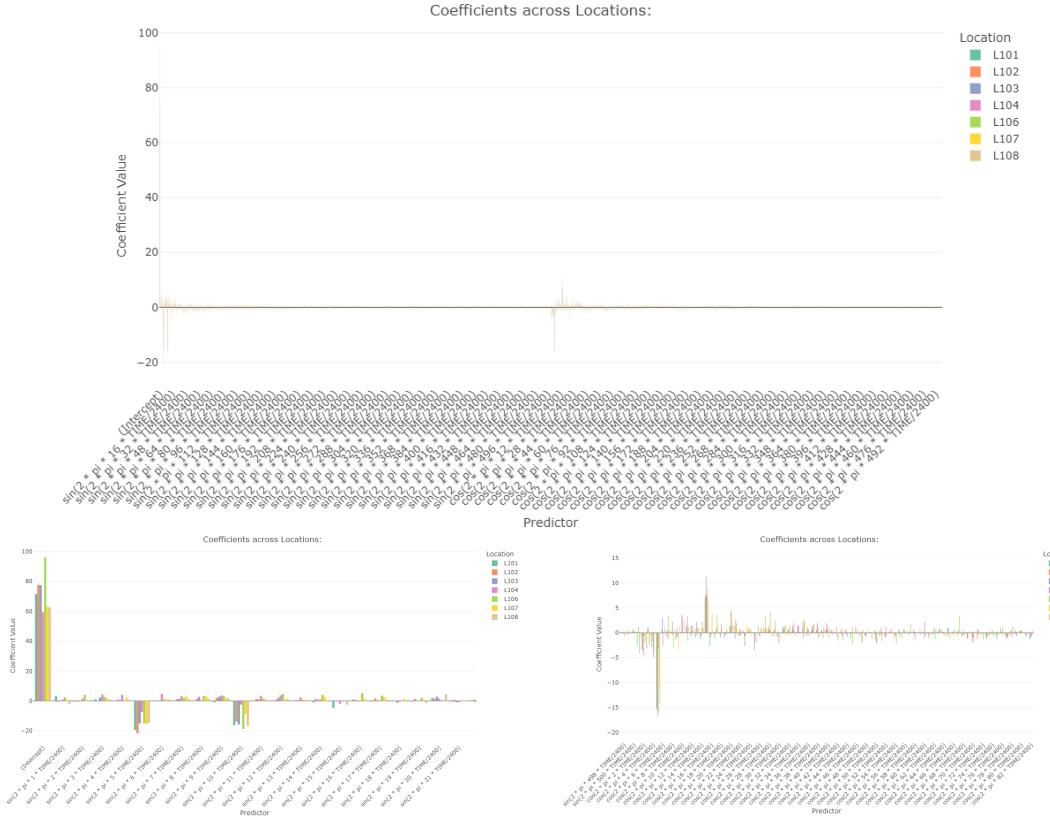


Figure 53: *Coefficients estimations of ℓ_1 MTE_{AIC} across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

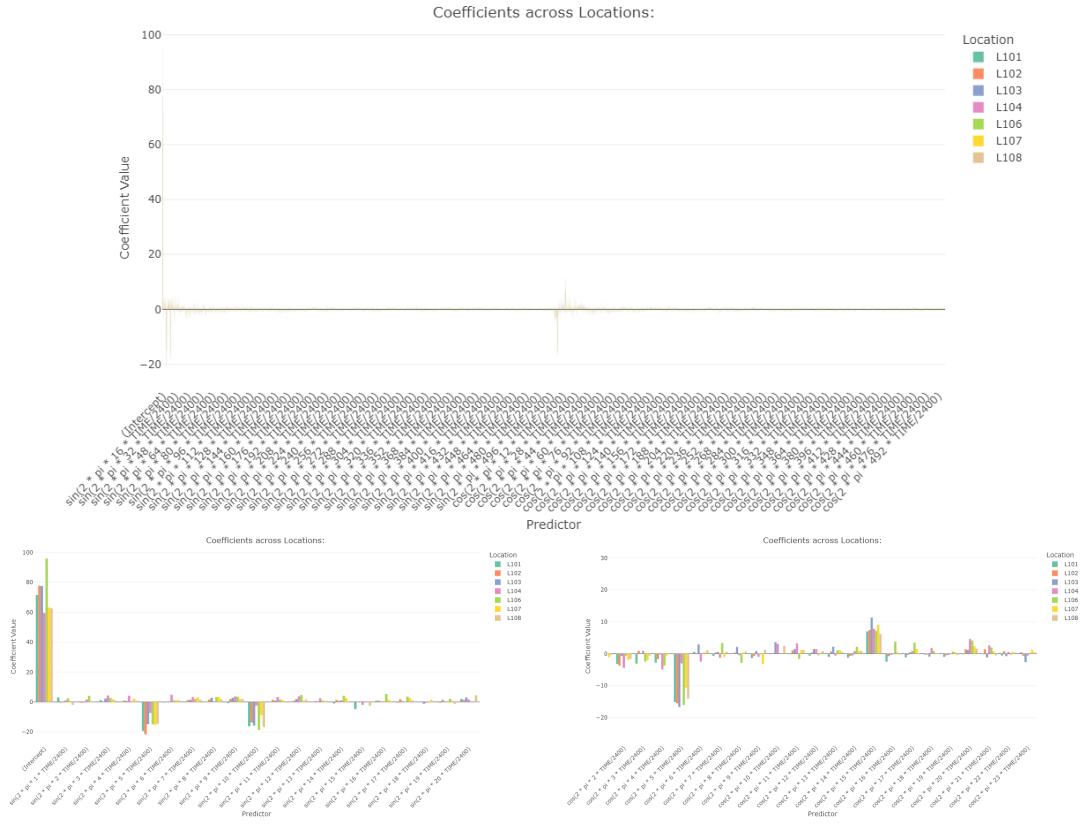


Figure 54: *Coefficients estimations of ℓ_1 MTEBIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

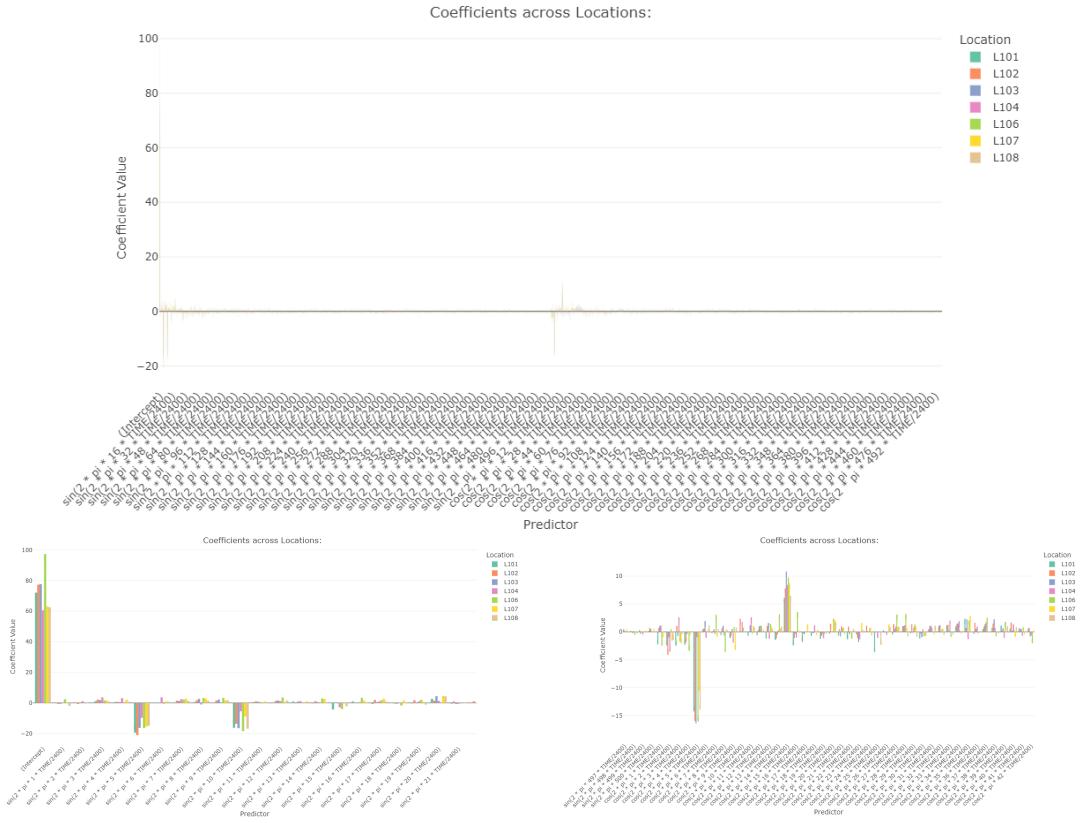


Figure 55: *Coefficients estimations of ℓ_2 MTE_{AIC} across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

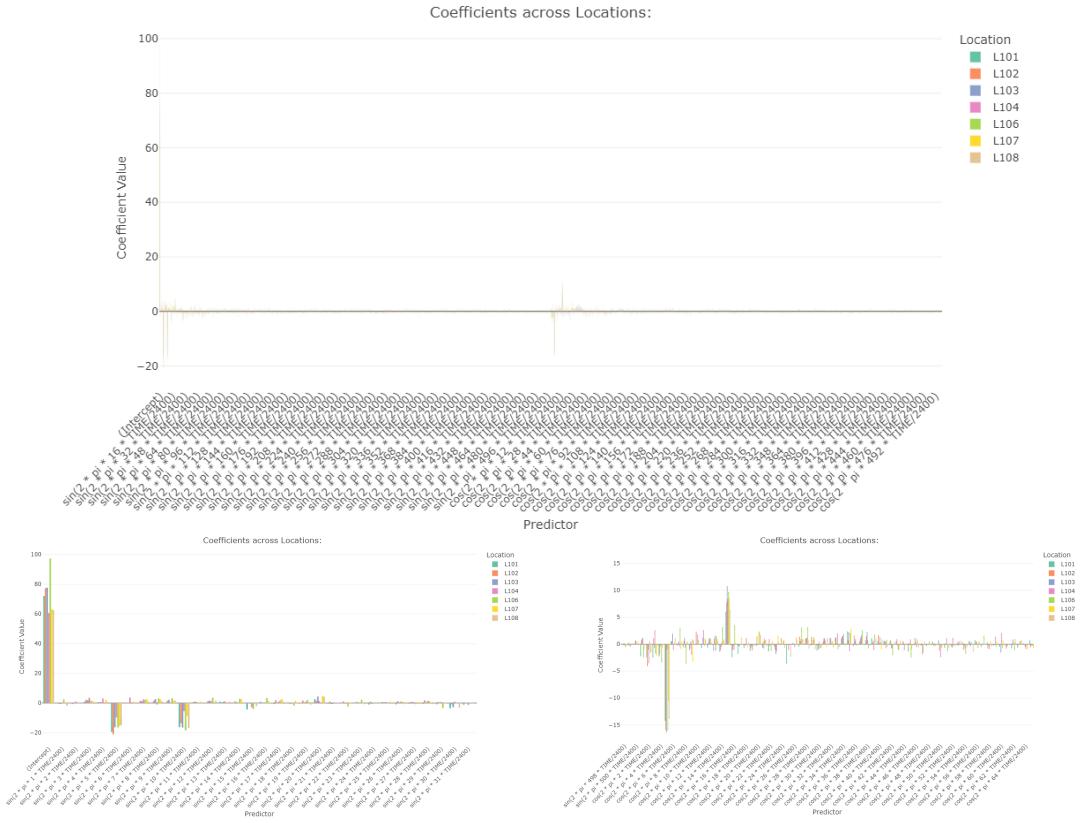


Figure 56: Coefficients estimations of ℓ_2 MTEBIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.

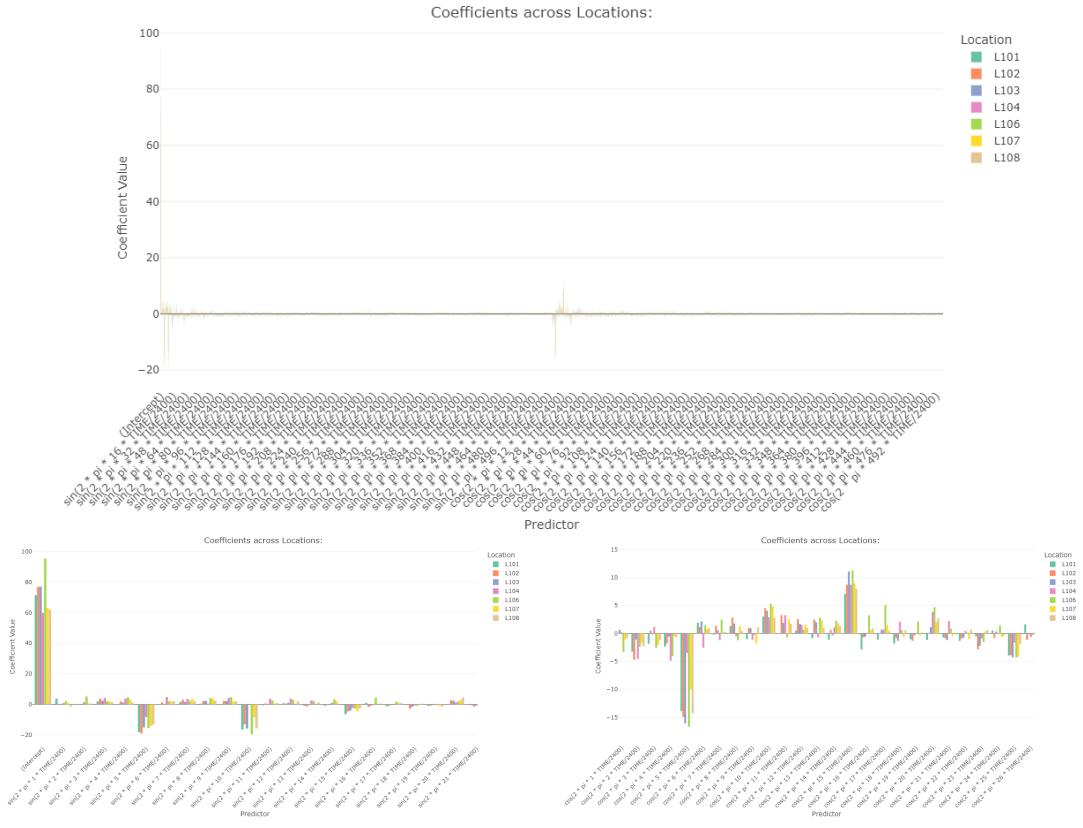


Figure 57: *Coefficients estimations of ℓ_1 LSAAIC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

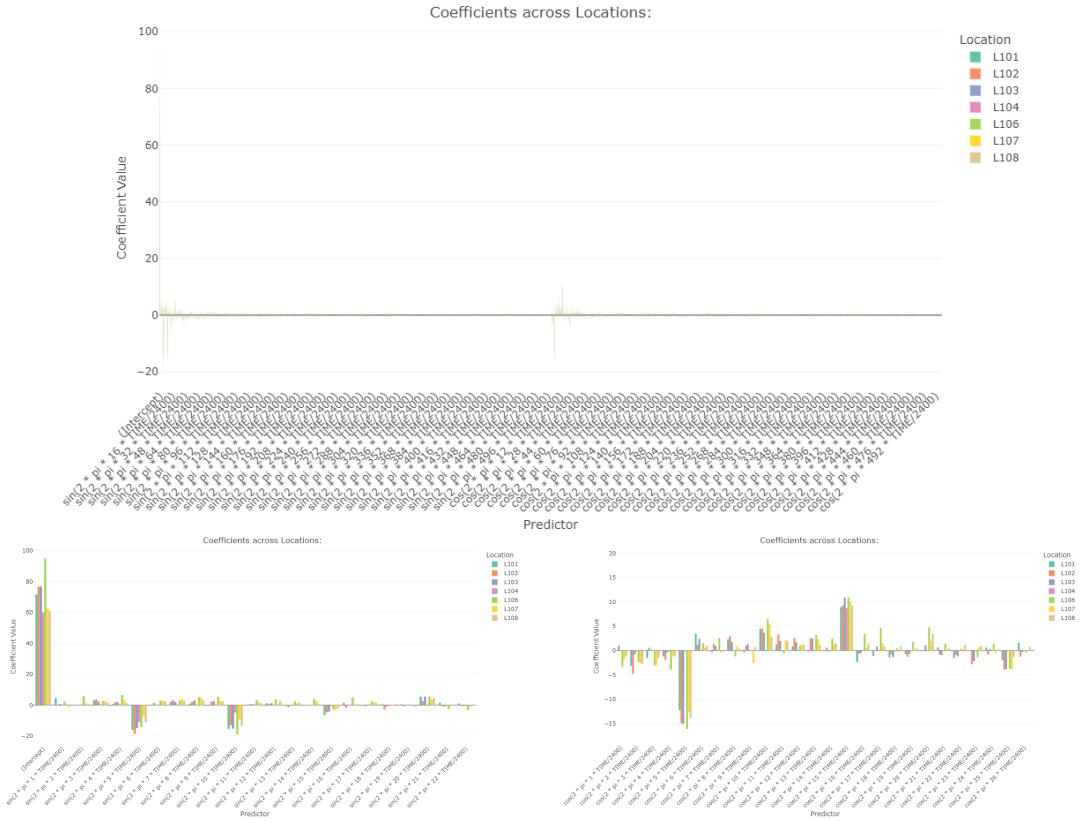


Figure 58: *Coefficients estimations of ℓ_1 LSA_{BIC} across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

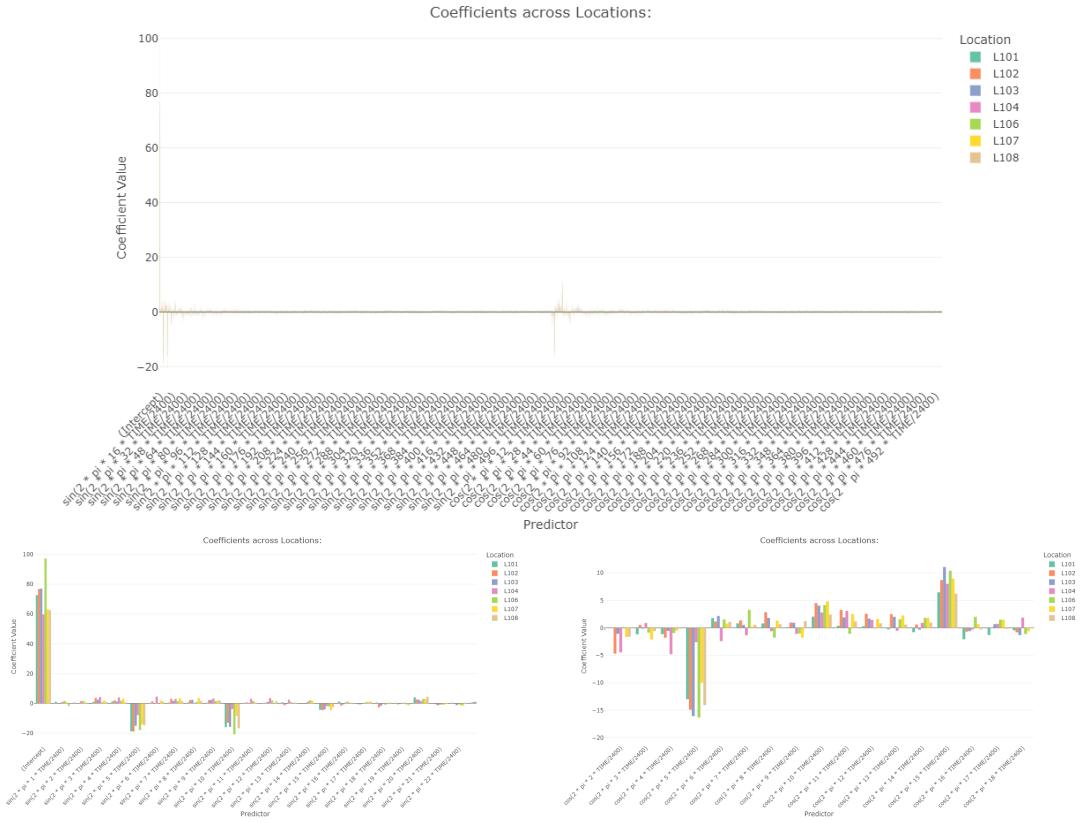


Figure 59: *Coefficients estimations of ℓ_2 LSAAC across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

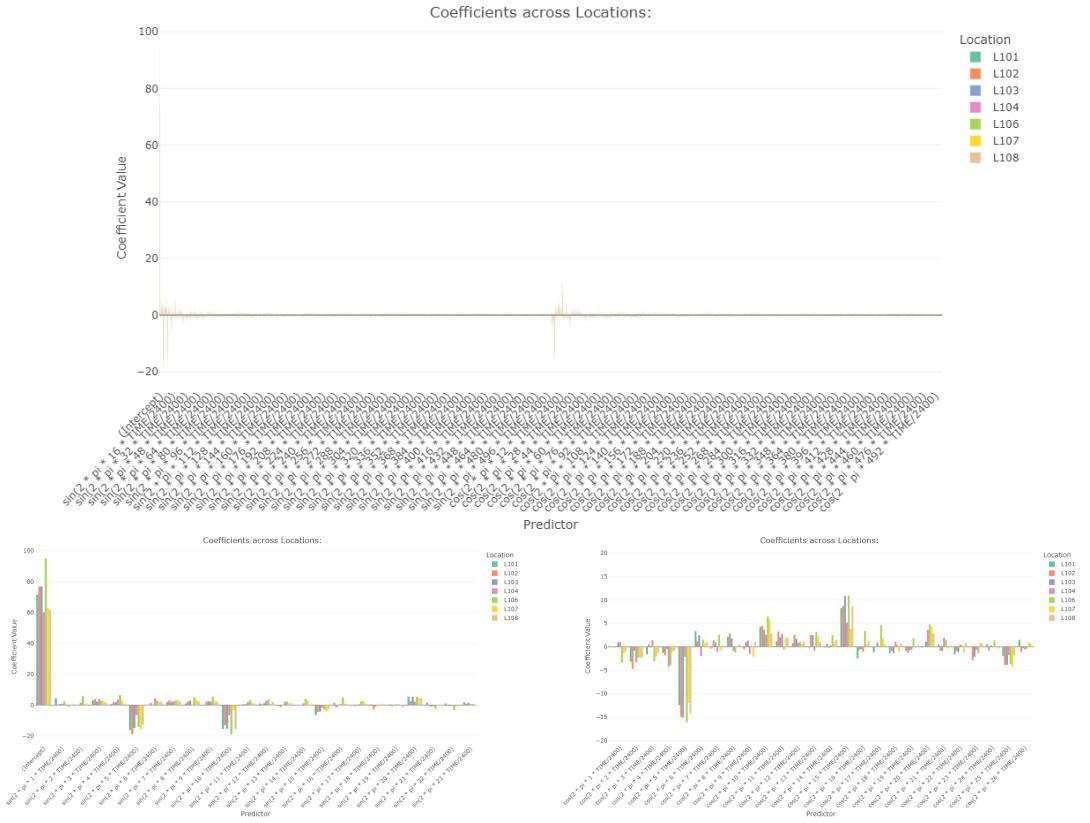


Figure 60: *Coefficients estimations of ℓ_2 LSA_{BIC} across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

Taking into account the performance metrics for various model configurations applied to traffic volume prediction, we can perform a systematic examination of the provided performance metrics for the training and testing sets which reveals distinct patterns concerning model efficacy and generalization capabilities. For the training set, post-hoc adaptive ℓ_1 LASSO models generally achieve superior performance, as indicated by their consistently lower values across metrics such as HL, MAE, and RMSE. Specifically, on the L107 (F) tend to reside at the apex of training performance within each data category. The consistently high values observed for MAPE and MdAPE (around 100%) across all models, when contrasted with the more tractable SMAPE and SMdAPE figures, strongly suggest the pervasive presence of numerous zero or near-zero actual values in the dataset. This inherent characteristic of the data often inflates traditional percentage errors due to division by, or proximity to, zero, thus rendering MAE and RMSE as more robust primary indicators of absolute deviation in this context. For metric scores derived from the robust adaptive LAD LASSO framework, a remarkable degree of parity is observed between the ℓ_1 LSA, ℓ_2 LSA models and their ℓ_1 MTE, ℓ_2 MTE counterparts on the training set. In numerous instances, particularly on location L107 (F), the performance metrics are virtually identical. This suggests that when the adaptive Lasso metric scores, both LSA and MTE formulations are highly effective at minimizing training

errors, exhibiting strikingly similar fitting capabilities. On the testing set, post-hoc LASSO models, as a class, generally demonstrate lower RMSE values (ranging from approximately 12 to 16) compared to models applied to models applied on original data and model averaging. This indicates superior generalization when the data structure inherently aligns with the adaptive Lasso framework. Within the post-hoc variable selection, both LSA and MTE models maintain competitive performance on the testing set, with no single type consistently outperforming the other across all locations under both minimization of information criterion techniques.

A more pronounced distinction in model efficacy emerges when evaluating performance on model averaging over the models applied on the original dataset and models applied against 20 MEBoot samples. For both the training and, critically, the testing sets, MTE models (e.g., on the first location L101, A) consistently present marginally to significantly superior performance. This superiority is evinced by lower values across primary error metrics such as HL, MAE, and RMSE when compared to their LSA equivalents. For instance, on the testing set, the leading performers in the models applied on the original and the model averaging categories are predominantly MTE variants, such as the ℓ_1 MTE_{AIC} on the original vehicular data one of the L101 (A) location (RMSE: 18.662) and corresponding model averaging (RMSE: 18.846). These MTE models generally surpass LSA models (e.g., ℓ_1 MTE_{AIC} on the original transportation data, RMSE: 17.766, which paradoxically has a lower RMSE but higher MAE and HL than some MTE models on testing, suggesting a complex interplay of error distribution, and model averaging of ℓ_1 MTE_{AIC}, RMSE: 18.050). This observed superior generalization of MTE models for model averaging and the models applied on the original real-world dataset robustly suggests that the minimum test error framework, or its inherent regularization properties, confers a greater advantage in predictive accuracy for unseen data, particularly when the underlying data structure may deviate from a purely adaptive Lasso assumption. This quantitative finding further corroborates previous qualitative observations from the VIPs heatmap 44, which indicated that MTE models exhibit a more stable and consistent variable selection profile across different information criteria, thereby hinting at a more intrinsically robust model specification.

In conclusion, while adaptive ℓ_1 Lasso models generally demonstrate commendable performance across diverse scenarios, the optimal selection between LSA and MTE frameworks appears to be contingent upon the specific characteristics of the data-generating process. For datasets explicitly constructed under post-hoc adaptive Lasso principles, both LSA and MTE models achieve comparable, high-fidelity results. However, for models applied on the original traffic volume data and model averaging, the MTE models consistently manifest as the more robust and reliable choice, offering superior generalization capabilities. This underscores their pronounced efficacy in contexts where the true underlying model may not perfectly conform to a simplified Lasso structure, highlighting their practical utility for real-world traffic prediction.

4.1 Conclusion

The comprehensive analysis of various penalized regression models, including Maximum Tangent likelihood Estimation (MTE) and Least Squares Approximation (LSA) under ℓ_1 and ℓ_2 norms, optimized with Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), offers profound insights into traffic volume prediction across diverse locations. This comparative study encompasses the family of models $\{\ell_x M_i | x \in \{1, 2\}, M \in \{\text{MTE}, \text{LSA}\}, i \in \{\text{AIC}, \text{BIC}\}\}$, where ℓ_x defines the norm, M denotes the robust penalized estimator, and i specifies the information criterion for regularization parameter selection. Feature importance is quantified using Variable

Inclusion Probabilities (VIPs), obtained via the Maximum Entropy Bootstrap (MEBoot) methodology . The application of variable selection methodologies is employed to ascertain the fundamental periodic characteristics embedded within the data. These 'fundamental characteristics' are defined as the time-dependent features that wield the most substantial influence on the dataset. Upon conceptualizing traffic volume data as a signal, these impactful characteristics would be those that demonstrate enduring persistence even after prolonged transmission of the signal. Such persistent features inherently possess long wavelengths, which, owing to the inverse relationship between wavelength and frequency, implies that the most influential data characteristics are expected to correspond to low-frequency components. This has been verified through our analysis, since after performing a fundamental observation across all model configurations is the pervasive sparsity of coefficient estimates, indicating that only a select subset of predictors significantly influences traffic dynamics. Foremost among these are the intercept term and specific low-frequency harmonic components, such as $\sin(2\pi \cdot 5 \frac{t}{2400})$ and $\cos(2\pi \cdot 5 \frac{t}{2400})$, which consistently exhibit exceptionally high Variable Inclusion Probabilities (VIPs) across nearly all models and criteria. This affirms their foundational role in capturing the constant baseline and primary periodic patterns. Other key periodicities, like $\sin(2\pi \cdot 10 \frac{t}{2400})$ and $\cos(2\pi \cdot 15 \frac{t}{2400})$, also frequently show high VIPs, while higher frequency terms generally contribute less and are often excluded by more parsimonious models, particularly under BIC minimization . The analyses were conducted on data from loop detectors at seven signalized intersections along Alexandras Street, with a primary focus on location L102 (B) for detailed illustration due to its anticipated "regular" traffic flow.

The choice of information criterion profoundly impacts model complexity: BIC consistently fosters greater sparsity due to its more stringent penalty, leading to the selection of fewer predictors compared to AIC, which tends to retain a larger ensemble . This distinction is particularly pronounced and reveals a notable instability in the variable selection of LSA-invariant models . For instance, at locations like L106 (E), ℓ_2 LSA_{AIC} selects significantly more predictors than ℓ_2 LSA_{BIC}, demonstrating high sensitivity to the information criterion. In stark contrast, MTE models exhibit a remarkably more consistent and robust variable selection process between AIC and BIC minimization , suggesting that MTE regularization confers greater stability irrespective of the specific criterion employed. The ℓ_1 MTE_{BIC} model, for example, yielded results virtually indistinguishable from ℓ_1 MTE_{AIC} in terms of regressor significance, inclusion order, coefficient estimates, and performance metrics. Overall, a consistent increase in error metrics is observed when transitioning from MTE to LSA models. Specifically, the Huber Loss (HL) demonstrated an average percentage increase of 10.28%, with its growth ranging from a minimum of 1.71% to a maximum of 22.65%. The Mean Absolute Error (MAE) similarly registered an average increase of 9.97%, fluctuating between a minimum of 1.65% and a maximum of 22.05%. For the Root Mean Squared Error (RMSE), the average increase was 4.13%; notably, while its maximum ascent reached 13.58%, a decrease of 2% was observed in certain instances.

Regarding predictive performance, the post-hoc adaptive ℓ_1 -Lasso models, using VIPs derived from the bootstrap procedure, generally achieve superior error metrics (Huber Loss, MAE, RMSE) in both training and testing phases, especially in terms of lower RMSE values. These results are presented in Tables 23, 24, 25, 26, 27, 28 and 29 where we refined the presentation style of the error metrics to show with respect to locations. This underscores their strong generalization capabilities when the data structure aligns with an adaptive Lasso framework. However, a more critical distinction emerges when models are averaged over maximum entropy bootstrap (MEBoot) samples or applied directly to the original dataset. In these scenarios, MTE models consistently demonstrate marginally to significantly superior performance, yielding lower values across primary error metrics

compared to their LSA equivalents. This enhanced generalization for unseen data by MTE models, particularly when the underlying data might not perfectly conform to a simplified Lasso structure, highlights their intrinsic robustness and predictive accuracy for real-world traffic prediction. For example, the ℓ_2 MTE_{AIC/BIC} achieved the lowest values for HL, MAE, and RMSE when model averaged over MEBoot samples . Conversely, while LSA models are grounded in robust regularization and are suited for sparse feature selection, they often do not achieve competitive results across the majority of error metrics, exhibiting relatively high deviations from optimal values and a tendency towards over-shrinkage, which can leave persistent seasonal patterns in residuals . The ℓ_1 LSA_{BIC} model, for instance, showed seasonal patterns in its residuals , and the ℓ_2 LSA models consistently demonstrated over-shrinkage, leading to omitted periodic components and abnormalities in residuals. Transitioning from MTE to LSA models results in a continuous increase in error measures. The HL increased by an average of 11.84%, ranging from a low of 4.19% to a maximum of 23.94%. MAE increased by an average of 11.47%, with values ranging from 4.02% to 23.29%. The average rise in RMSE was 4.13%, with a maximum climb of 13.59% and the minimum value reached at 2%. The selection of the model family and the information criterion significantly influence the parsimony of the model and the predictive effectiveness. MTE models offer a more stable and robust approach to variable selection, exhibiting superior generalization capabilities, especially when applied to original data or averaged over bootstrap samples. This renders MTE models a more reliable choice for accurate and robust traffic prediction in practical applications where the underlying vehicular data dynamics may be complex and susceptible to external influences. The pervasive sparsity across all models suggests that a parsimonious structure is often sufficient for capturing essential traffic dynamics. It is particularly noteworthy how effectively the Least Square Approximation (LSA) performed, even when confronted with data exhibiting both heteroskedasticity and autocorrelation. This commendable performance was achieved despite the covariance matrix being estimated through Pollard's [16] methodology, an approach that inherently disregards the temporal dependence and serial correlation present in the data points, thereby presenting a challenging condition for model compatibility.

To facilitate a more rigorous comparative analysis between these models, a sophisticated estimation of the covariance matrix could be carried out, ideally using cutting-edge methodologies. For instance, the approach proposed by Galvao-Yoon [49], by specifically accounting for heteroskedasticity and serially correlated data, would confer greater robustness upon the Mahalanobis distance estimation utilized by LSA. Concurrently, the application of Newey-West [23] estimation could further refine the least squares components of these models. Furthermore, an intriguing avenue for exploration involves investigating the impact of employing a higher-order Taylor expansion for the Maximum Tangent Estimations (MTE), a refinement that could potentially illuminate which model framework ultimately demonstrates superior predictive prowess. Additionally, augmenting the comprehensiveness of the hyperparameter tuning process represents another valuable modification, entailing not only enriching the search domain but also substantially expanding the number of maximum entropy bootstrap samples from the current twenty to nine hundred ninety-nine, a quantity recommended by Vinod [32], [33] to fully leverage the ergodic theorem's properties inherent in the MEBoot procedure.

5 Appendix

5.1 Tables of Results for all locations

We are going to present here the results of the experiments we did on the rest of locations separated by their content.

5.1.1 Variable Inclusion Probabilities (VIPs)

<i>Features</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	0.95	0.00	0.90	0.90	1.00	0.00
$\sin(2\pi \cdot 20 \frac{t}{2400})$	0.40	0.40	0.05	0.00	0.60	0.60	1.00	0.00
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.25	0.25	0.00	0.00	0.15	0.15	0.00	0.00
$\sin(2\pi \cdot 30 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$\cos(2\pi \cdot 25 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$\cos(2\pi \cdot 35 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$\sin(2\pi \cdot 1 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.70	0.00
$\cos(2\pi \cdot 2 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	0.65	0.00

Table 1: Variable Inclusion Probabilities for location L101

<i>Features</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 20 \frac{t}{2400})$	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 15 \frac{t}{2400})$	0.35	0.35	0.00	0.00	0.35	0.35	0.00	0.00
$\cos(2\pi \cdot 25 \frac{t}{2400})$	0.35	0.35	0.00	0.00	0.20	0.20	0.00	0.00
$\sin(2\pi \cdot 3 \frac{t}{2400})$	0.15	0.15	0.00	0.00	0.00	0.00	0.00	0.00
$\sin(2\pi \cdot 45 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$\cos(2\pi \cdot 11 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 2: Variable Inclusion Probabilities for location L102

<i>Methods</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 20 \frac{t}{2400})$	1.00	1.00	0.00	1.00	1.00	1.00	0.00	0.00
$\sin(2\pi \cdot 15 \frac{t}{2400})$	0.10	0.05	0.00	0.20	0.20	0.00	0.00	0.00
$\sin(2\pi \cdot 40 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00
$\cos(2\pi \cdot 20 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00
$\cos(2\pi \cdot 25 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00

Table 3: Variable Inclusion Probabilities for location L103

<i>Methods</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.25
$\cos(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.25
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	0.05	0.00	0.75	0.75	1.00	0.00
$\cos(2\pi \cdot 20 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.35	0.35	0.00	0.00
$\sin(2\pi \cdot 9 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$\sin(2\pi \cdot 11 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$\cos(2\pi \cdot 4 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$\cos(2\pi \cdot 128 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 4: Variable Inclusion Probabilities for location L104

<i>Methods</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 20 \frac{t}{2400})$	0.20	0.20	1.00	0.00	0.15	0.15	1.00	0.00
$\cos(2\pi \cdot 25 \frac{t}{2400})$	0.10	0.10	0.00	0.00	0.10	0.10	1.00	0.00
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.05	0.05	1.00	0.00	0.40	0.40	1.00	0.00
$\sin(2\pi \cdot 4 \frac{t}{2400})$	0.05	0.05	1.00	0.00	0.05	0.05	1.00	0.00
$\sin(2\pi \cdot 15 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	1.00	0.00
$\sin(2\pi \cdot 24 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	1.00	0.00
$\cos(2\pi \cdot 17 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	1.00	0.00
$\sin(2\pi \cdot 8 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$\sin(2\pi \cdot 9 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$\sin(2\pi \cdot 16 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$\cos(2\pi \cdot 4 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$\cos(2\pi \cdot 20 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00

Table 5: Variable Inclusion Probabilities for location L106

<i>Methods</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	0.10	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.20
$\cos(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.20
$\cos(2\pi \cdot 15 \frac{t}{2400})$	0.65	0.65	0.00	0.00	0.40	0.40	0.00	0.00
$\sin(2\pi \cdot 20 \frac{t}{2400})$	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.15	0.15	0.00	0.00	0.30	0.30	0.00	0.00
$\sin(2\pi \cdot 15 \frac{t}{2400})$	0.10	0.10	0.00	0.00	0.35	0.35	0.00	0.00
$\sin(2\pi \cdot 30 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 6: Variable Inclusion Probabilities for location L107

<i>Methods</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	0.25	1.00	1.00	1.00	1.00
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	1.00	0.25	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	0.25	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00
$\sin(2\pi \cdot 20 \frac{t}{2400})$	1.00	1.00	0.00	0.00	1.00	1.00	1.00	0.00
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$\cos(2\pi \cdot 97 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 7: Variable Inclusion Probabilities for location L108

<i>Features</i>	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}	Overall
(Intercept)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\cos(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	0.71	1.00	1.00	1.00	1.00	0.96
$\sin(2\pi \cdot 5 \frac{t}{2400})$	1.00	1.00	1.00	0.62	1.00	1.00	1.00	0.89	0.94
$\cos(2\pi \cdot 15 \frac{t}{2400})$	1.00	1.00	1.00	0.29	1.00	1.00	1.00	0.64	0.87
$\sin(2\pi \cdot 10 \frac{t}{2400})$	1.00	1.00	0.86	0.61	0.96	0.96	1.00	0.74	0.89
$\sin(2\pi \cdot 20 \frac{t}{2400})$	0.63	0.63	0.34	0.00	0.59	0.59	0.67	0.00	0.43
$\sin(2\pi \cdot 15 \frac{t}{2400})$	0.32	0.31	0.19	0.00	0.36	0.36	0.40	0.00	0.24
$\cos(2\pi \cdot 25 \frac{t}{2400})$	0.23	0.23	0.00	0.00	0.15	0.15	0.20	0.00	0.12
$\sin(2\pi \cdot 3 \frac{t}{2400})$	0.15	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.04
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.08	0.08	0.17	0.00	0.16	0.16	0.17	0.00	0.10
$\sin(2\pi \cdot 4 \frac{t}{2400})$	0.05	0.05	1.00	0.00	0.05	0.05	1.00	0.00	0.28
$\cos(2\pi \cdot 35 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.01
$\sin(2\pi \cdot 45 \frac{t}{2400})$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.01
$\sin(2\pi \cdot 30 \frac{t}{2400})$	0.02	0.02	0.00	0.00	0.02	0.02	0.00	0.00	0.01
$\cos(2\pi \cdot 20 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.13	0.13	0.33	0.00	0.07
$\sin(2\pi \cdot 2 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	1.00	0.00	0.14
$\cos(2\pi \cdot 17 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	1.00	0.00	0.14
$\sin(2\pi \cdot 1 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.05	0.05	0.70	0.00	0.10
$\cos(2\pi \cdot 11 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.05	0.00	0.00	0.01
$\sin(2\pi \cdot 40 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.05	0.00	0.00	0.01
$\sin(2\pi \cdot 11 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.05	0.00	0.00	0.01
$\cos(2\pi \cdot 128 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.05	0.00	0.00	0.01
$\cos(2\pi \cdot 97 \frac{t}{2400})$	0.00	0.00	0.00	0.05	0.05	0.05	0.00	0.00	0.01
$\sin(2\pi \cdot 9 \frac{t}{2400})$	0.00	0.00	0.00	0.02	0.02	0.02	0.50	0.00	0.07
$\cos(2\pi \cdot 4 \frac{t}{2400})$	0.00	0.00	0.00	0.02	0.02	0.02	0.50	0.00	0.07
$\sin(2\pi \cdot 8 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.12
$\sin(2\pi \cdot 16 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.12
$\cos(2\pi \cdot 2 \frac{t}{2400})$	0.00	0.00	0.00	0.00	0.00	0.00	0.65	0.00	0.08

Table 8: Variable Inclusion Probabilities for according to all locations

5.1.2 Coefficients Estimation regarding models across locations

Frequencies	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	75.05 (65.55, 75.10 , 85.02)	75.05 (65.55, 85.02)	70.65 (60.90, 70.58 , 80.50)	69.20 (59.33, 69.12 , 79.03)	74.63 (65.10, 74.28 , 85.09)	74.63 (65.10, 74.28 , 85.09)	70.70 (60.87, 70.72 , 80.52)	70.14 (60.27, 70.00 , 79.95)
$\cos(2\pi \cdot 15 \frac{t}{2400})$	10.20 (7.01, 10.04 , 12.71)	5.93 (7.01, 10.04 , 12.71)	0.00 (5.09, 5.96 , 6.62)	0.00 (0.00, 0.00 , 0.00)	10.16 (7.45, 10.34 , 11.97)	10.16 (7.45, 10.34 , 11.97)	7.26 (6.40, 7.48 , 7.88)	4.95 (4.63, 4.95 , 5.33)
$\sin(2\pi \cdot 20 \frac{t}{2400})$	1.22 (0.00, 1.00 , 4.52)	0.10 (0.00, 0.00 , 4.52)	0.10 (0.00, 0.00 , 1.04)	0.00 (0.00, 0.00 , 0.00)	2.23 (0.00, 2.02 , 5.08)	2.23 (1.72, 3.38 , 3.59)	2.93 (0.00, 2.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 2 \frac{t}{2400})$	0.74 (0.00, 0.00 , 4.18)	0.74 (0.00, 0.00 , 4.18)	0.00 (0.00, 0.00 , 0.00)	0.21 (0.00, 0.00 , 0.00)	0.21 (0.00, 0.00 , 0.00)	0.21 (0.00, 0.00 , 0.00)	0.00 (0.00, 1.90 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 35 \frac{t}{2400})$	0.08 (0.00, 0.00 , 0.81)	0.08 (0.00, 0.00 , 0.81)	0.00 (0.00, 0.00 , 0.00)					
$\cos(2\pi \cdot 1 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	1.32 (0.00, 0.00 , 2.03)	0.00 (0.00, 0.00 , 0.00)					
$\sin(2\pi \cdot 2 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	-0.69 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)					
$\cos(2\pi \cdot 25 \frac{t}{2400})$	-0.11 (-1.12, 0.00 , 0.00)	-0.11 (-1.12, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 30 \frac{t}{2400})$	-0.15 (-1.54, 0.00 , 0.00)	-0.15 (-1.54, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, -1.03 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 25 \frac{t}{2400})$	-4.95 (-7.25, -4.95 , 2.64)	-4.95 (-7.25, -4.95 , 2.64)	-2.24 (-3.11, -2.24 , 0.92)	0.00 (0.00, 0.00 , 0.00)	-3.61 (-6.55, -3.61 , 0.00)	-3.61 (-6.55, -3.61 , 0.00)	-3.53 (-4.07, -3.53 , 2.55)	0.00 (0.00, 0.00 , 7.00)
$\cos(2\pi \cdot 5 \frac{t}{2400})$	-14.45 (-15.81, -14.56 , 13.33)	-14.45 (-15.81, -14.56 , 13.33)	-12.64 (-13.28, -12.57 , 12.01)	-9.04 (-9.34, -9.00 , -8.76)	-13.96 (-15.41, -14.20 , 11.34)	-13.96 (-15.41, -14.20 , 11.34)	-12.57 (-14.11, -13.71 , 12.05)	-12.18 (-12.54, -12.17 , 11.93)
$\sin(2\pi \cdot 10 \frac{t}{2400})$	-15.66 (-17.65, -16.81 , -14.08)	-15.66 (-17.65, -16.81 , -14.08)	-15.53 (-16.30, -16.45 , -14.79)	-12.49 (-12.92, -12.50 , -12.08)	-17.18 (-18.83, -16.94 , -15.61)	-17.18 (-18.83, -16.94 , -15.61)	-15.64 (-16.22, -16.74 , -15.00)	-14.42 (-14.84, -14.37 , -14.10)
$\sin(2\pi \cdot 5 \frac{t}{2400})$	-21.51 (-24.33, -21.66 , -18.59)	-21.51 (-24.33, -21.66 , -18.59)	-17.67 (-18.35, -17.62 , -16.99)	-14.94 (-15.35, -14.94 , -14.64)	-20.15 (-22.10, -20.18 , -17.50)	-20.15 (-22.10, -20.18 , -17.50)	-17.74 (-18.34, -17.78 , -17.11)	-16.65 (-17.11, -16.97 , -16.32)

Table 9: Location L101 location. The coral tiny numbers indicate the 2.5th and 97.5th quantiles, the navy-blue tiny numbers represent the median values and normal-sized numbers refer to the mean value of the estimated coefficients model-wise

Frequencies	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	77.53 (66.89, 78.21, 87.45)	77.53 (66.89, 78.21, 87.45)	76.42 (66.48, 76.37, 86.29)	75.16 (65.07, 75.14, 85.12)	77.82 (68.47, 77.90, 87.35)	77.82 (68.47, 77.90, 87.35)	76.40 (66.47, 76.40, 86.30)	76.40 (66.47, 76.40, 86.30)
$\cos(2\pi \cdot 15 \frac{t}{2400})$	7.03 (5.65, 6.35, 8.04)	7.03 (5.65, 6.35, 8.04)	5.84 (5.65, 6.35, 6.12)	0.00 (0.00, 0.00, 0.00)	7.28 (5.49, 6.21, 10.31)	7.28 (5.49, 6.21, 10.31)	6.22 (6.07, 6.22, 6.45)	6.22 (6.07, 6.22, 6.45)
$\sin(2\pi \cdot 20 \frac{t}{2400})$	4.16 (2.52, 3.84, 6.77)	4.16 (2.52, 3.84, 6.77)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	2.95 (0.00, 3.52, 4.80)	2.95 (0.00, 3.52, 4.80)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\sin(2\pi \cdot 3 \frac{t}{2400})$	0.17 (0.00, 0.00, 1.68)	0.17 (0.00, 0.00, 1.68)	0.00 (0.00, 0.00, 0.00)					
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.09 (0.00, 0.00, 0.93)	0.09 (0.00, 0.00, 0.93)	0.00 (0.00, 0.00, 0.00)					
$\sin(2\pi \cdot 45 \frac{t}{2400})$	0.07 (0.00, 0.00, 0.72)	0.07 (0.00, 0.00, 0.72)	0.00 (0.00, 0.00, 0.00)					
$\cos(2\pi \cdot 11 \frac{t}{2400})$	0.00 (-0.00, 0.00, 0.09)	0.00 (-0.00, 0.00, 0.09)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.12 (0.00, 0.00, 1.28)	0.12 (0.00, 0.00, 1.28)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\sin(2\pi \cdot 15 \frac{t}{2400})$	-0.87 (-3.56, 0.00, 0.00)	-0.87 (-3.56, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	-0.91 (-4.36, 0.00, 0.00)	-0.91 (-4.36, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\cos(2\pi \cdot 25 \frac{t}{2400})$	-1.05 (-4.10, 0.00, 0.00)	-1.05 (-4.10, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	-0.51 (-3.44, 0.00, 0.00)	-0.51 (-3.44, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\sin(2\pi \cdot 12 \frac{t}{2400})$	-14.71 (-16.25, -14.52, -13.37)	-14.71 (-16.25, -14.52, -13.37)	-11.97 (-12.30, -11.93, -11.78)	-8.39 (-8.69, -8.34, -8.20)	-15.24 (-17.21, -15.03, -14.02)	-15.24 (-17.21, -15.03, -14.02)	-12.23 (-12.46, -12.21, -12.04)	-12.23 (-12.46, -12.21, -12.04)
$\cos(2\pi \cdot 5 \frac{t}{2400})$	-17.29 (-18.94, -17.47, -14.71)	-17.29 (-18.94, -17.47, -14.71)	-14.44 (-14.44, -13.84, -14.19)	-11.36 (-11.36, -10.94, -10.98)	-17.37 (-19.92, -17.84, -18.80)	-17.37 (-19.92, -17.84, -18.80)	-14.35 (-14.54, -14.34, -14.14)	-14.35 (-14.54, -14.34, -14.14)
$\sin(2\pi \cdot 5 \frac{t}{2400})$	-22.93 (-25.38, -22.35, -21.06)	-22.93 (-25.38, -22.35, -21.06)	-19.36 (-19.71, -19.32, -19.10)	-16.99 (-17.38, -16.99, -16.66)	-22.80 (-24.76, -22.54, -20.62)	-22.80 (-24.76, -22.54, -20.62)	-19.10 (-19.42, -19.13, -18.88)	-19.10 (-19.42, -19.13, -18.88)

Table 10: Location L102 location. The coral tiny numbers indicate the 2.5th and 97.5th quantiles, the navy-blue tiny numbers represent the median values and normal-sized numbers refer to the mean value of the estimated coefficients model-wise

Frequencies	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	76.19 (65.41, 76.76, 87.25)	76.21 (65.41, 76.76, 87.25)	77.17 (66.70, 77.00, 87.75)	76.62 (66.08, 76.52, 87.25)	75.88 (66.43, 75.50, 86.71)	75.88 (66.43, 75.50, 86.71)	77.00 (66.51, 76.50, 87.55)	76.45 (65.90, 76.40, 87.05)
$\cos(2\pi \cdot 15 \frac{t}{2400})$	3.95 (8.70, 9.67, 11.47)	10.19 (8.62, 9.67, 11.47)	10.19 (10.05, 10.17, 10.37)	2.28 (8.08, 8.30, 8.51)	4.48 (8.48, 9.42, 11.00)	9.48 (8.48, 9.42, 11.00)	8.93 (9.75, 9.91, 10.17)	8.03 (7.88, 8.02, 8.22)
$\sin(2\pi \cdot 20 \frac{t}{2400})$	5.05 (4.30, 5.12, 5.73)	5.08 (4.30, 5.12, 5.73)	3.73 (3.49, 3.70, 3.98)	0.00 (0.00, 0.00, 0.00)	5.32 (3.81, 5.10, 7.59)	5.32 (3.81, 5.10, 7.59)	0.00 (3.46, 3.61, 3.78)	0.00 (0.00, 0.00, 0.00)
$\cos(2\pi \cdot 20 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.15 (0.00, 0.15, 1.55)	0.15 (0.00, 0.15, 1.55)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.14 (0.00, 0.00, 1.52)	0.14 (0.00, 0.00, 1.52)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\sin(2\pi \cdot 40 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.13 (0.00, 0.00, 1.37)	0.13 (0.00, 0.00, 1.37)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\cos(2\pi \cdot 25 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	-0.17 (-1.77, 0.00, 0.00)	-0.17 (-1.77, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\sin(2\pi \cdot 15 \frac{t}{2400})$	-0.20 (-1.99, 0.00, 0.00)	-0.08 (-0.83, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	-0.37 (-2.36, 0.00, 0.00)	-0.37 (-2.36, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)
$\sin(2\pi \cdot 10 \frac{t}{2400})$	-17.88 (-19.27, -17.39, -16.24)	-17.90 (-19.27, -17.39, -16.24)	-15.74 (-16.81, -15.72, -15.52)	-14.44 (-15.04, -14.45, -14.22)	-17.34 (-19.56, -17.22, -16.03)	-17.34 (-19.56, -17.22, -16.03)	-15.65 (-16.13, -15.59, -15.45)	-14.37 (-14.84, -14.32, -14.18)
$\cos(2\pi \cdot 5 \frac{t}{2400})$	-17.91 (-19.54, -17.84, -16.47)	-17.95 (-19.54, -17.84, -16.47)	-16.39 (-15.51, -16.33, -14.90)	-15.14 (-14.14, -15.06, -13.92)	-17.23 (-19.36, -17.29, -16.50)	-17.23 (-19.36, -17.29, -16.50)	-15.80 (-15.50, -15.78, -14.07)	-14.52 (-14.16, -14.49, -13.63)
$\sin(2\pi \cdot 5 \frac{t}{2400})$	-18.15 (-19.65, -18.57, -16.60)	-18.24 (-19.65, -18.57, -16.60)	-15.09 (-16.13, -15.02, -16.14)	-13.74 (-14.83, -13.68, -14.81)	-18.33 (-19.40, -18.55, -14.59)	-18.33 (-19.40, -18.55, -14.59)	-15.16 (-15.97, -15.14, -15.60)	-13.84 (-14.60, -13.79, -14.30)

Table 11: Location L103 location. The coral tiny numbers indicate the 2.5th and 97.5th quantiles, the navy-blue tiny numbers represent the median values and normal-sized numbers refer to the mean value of the estimated coefficients model-wise

Frequencies	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}	
(Intercept)	54.56 (54.56, 54.56, 54.56)	54.56 (54.56, 54.56, 54.56)	59.43 (51.96, 59.38, 66.77)	57.10 (49.30, 57.13, 64.58)	58.64 (51.20, 58.50, 66.36)	58.64 (51.20, 58.50, 66.36)	60.03 (52.61, 60.01, 67.37)	57.63 (49.20, 57.08, 66.98)	
$\cos(2\pi \cdot 15 \frac{t}{2400})$	12.66 (12.66, 12.66, 12.66)	12.66 (12.66, 12.66, 12.66)	6.18 (8.95, 6.35, 6.87)	0.00 (0.00, 0.00, 0.00)	10.85 (8.15, 10.57, 12.91)	10.85 (7.50, 7.67, 7.85)	7.62 (7.00, 7.57, 8.44)	1.60 (0.00, 0.00, 0.00)	
$\cos(2\pi \cdot 20 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	1.32 (0.00, 0.00, 1.00)	1.32 (0.00, 0.00, 1.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	
$\sin(2\pi \cdot 11 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.15 (0.00, 0.00, 4.00)	0.15 (0.00, 0.00, 4.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	
$\sin(2\pi \cdot 9 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.11 (0.00, 0.00, 1.18)	0.11 (0.00, 0.00, 1.18)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	
$\cos(2\pi \cdot 128 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	-0.14 (-1.43, 0.00, 0.00)	-0.14 (-1.43, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	
$\cos(2\pi \cdot 4 \frac{t}{2400})$	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	0.00 (0.00, 0.00, 0.00)	-3.56 (-5.63, -4.74, 0.00)	-3.56 (-5.63, -4.74, 0.00)	-2.41 (-2.50, -2.40, -2.32)	0.00 (0.00, 0.00, 0.00)	
$\sin(2\pi \cdot 10 \frac{t}{2400})$	-5.23 (-11.61, -5.23)	-5.23 (-11.61, -5.23)	-4.11 (-4.20, -4.09, 0.00)	-7.02 (-7.07, -6.90, -6.81)	0.00 (0.00, 0.00, 0.00)	-13.97 (-18.19, -13.81, -11.32)	-13.97 (-18.19, -13.81, -11.32)	-8.85 (-9.08, -8.81, -8.70)	-1.94 (-7.82, 0.00, 0.00)
$\sin(2\pi \cdot 5 \frac{t}{2400})$	-11.61 (-11.61, -11.61)	-11.61 (-11.61, -11.61)	-11.61 (-11.61, -11.61)	-7.02 (-7.07, -6.90, -6.81)	0.00 (0.00, 0.00, 0.00)	-11.32 (-18.19, -13.81, -11.32)	-11.32 (-18.19, -13.81, -11.32)	-8.85 (-9.08, -8.81, -8.70)	-1.94 (-7.82, 0.00, 0.00)

Table 12: Location L104 location. The coral tiny numbers indicate the 2.5th and 97.5th quantiles, the navy-blue tiny numbers represent the median values and normal-sized numbers refer to the mean value of the estimated coefficients model-wise

Frequencies	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	98.57 (87.54, 97.58 , 112.90)	98.57 (87.54, 97.58 , 112.90)	95.59 (83.11, 95.58 , 107.94)	95.13 (82.60, 95.09 , 107.54)	96.79 (83.43, 96.94 , 110.47)	96.79 (83.43, 96.94 , 110.47)	95.23 (82.82, 95.17 , 107.56)	94.45 (81.05, 94.40 , 106.85)
$\cos(2\pi \cdot 15 \frac{t}{2400})$	12.36 (0.28, 12.36 , 14.71)	12.36 (0.28, 12.36 , 14.71)	9.67 (0.38, 9.67 , 9.90)	7.73 (7.43, 7.73 , 8.00)	13.04 (10.27, 13.04 , 14.57)	13.04 (10.27, 13.04 , 14.57)	10.28 (10.03, 10.28 , 10.49)	6.95 (6.75, 6.95 , 7.12)
$\sin(2\pi \cdot 20 \frac{t}{2400})$	1.00 (0.00, 0.00 , 5.11)	1.00 (0.00, 0.00 , 5.11)	3.28 (3.23, 3.42 , 2.79)	0.00 (0.00, 0.00 , 0.00)	0.37 (0.00, 0.00 , 3.03)	0.37 (0.00, 0.00 , 3.03)	4.23 (5.34, 5.46 , 4.33)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.29 (0.00, 0.29 , 3.04)	0.29 (0.00, 0.29 , 3.04)	3.44 (2.25, 2.47 , 3.68)	0.00 (0.00, 0.00 , 0.00)	1.91 (0.00, 1.91 , 5.96)	1.91 (0.00, 1.91 , 5.96)	5.46 (4.06, 4.23 , 5.58)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 4 \frac{t}{2400})$	0.21 (0.00, 0.21 , 2.16)	0.21 (0.00, 0.21 , 2.16)	0.00 (1.37, 1.50 , 1.64)	0.00 (0.00, 0.00 , 0.00)	0.19 (0.00, 0.19 , 1.77)	0.19 (0.00, 0.19 , 1.77)	3.01 (3.17, 3.23 , 3.31)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 2 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	0.17 (0.00, 0.17 , 1.77)	0.17 (0.00, 0.17 , 1.77)	2.29 (2.31, 2.35 , 2.36)	0.00 (0.00, 0.00 , 0.00)			
$\cos(2\pi \cdot 17 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	0.15 (0.00, 0.15 , 1.58)	0.15 (0.00, 0.15 , 1.58)	2.24 (2.24, 2.30 , 1.93)	0.00 (0.00, 0.00 , 0.00)			
$\sin(2\pi \cdot 9 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	2.35 (2.09, 2.13 , 2.42)	0.00 (0.00, 0.00 , 0.00)					
$\sin(2\pi \cdot 16 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	2.13 (1.52, 2.13 , 2.18)	0.00 (0.00, 0.00 , 0.00)					
$\sin(2\pi \cdot 8 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	1.57 (1.20, 1.28 , 1.61)	0.00 (0.00, 0.00 , 0.00)					
$\cos(2\pi \cdot 20 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	0.58 (0.35, 0.58 , 0.80)	0.00 (0.00, 0.00 , 0.00)					
$\cos(2\pi \cdot 4 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	-0.60 (-0.65, -0.60 , -0.52)	0.00 (0.00, 0.00 , 0.00)					
$\sin(2\pi \cdot 15 \frac{t}{2400})$	-0.48 (-2.09, -0.48 , 0.00)	-0.48 (-2.09, -0.48 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	-0.50 (-5.05, -0.50 , 0.00)	-0.50 (-5.05, -0.50 , 0.00)	-1.60 (-1.67, -1.60 , 0.75)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 25 \frac{t}{2400})$	-0.48 (-4.86, -0.48 , 0.00)	-0.48 (-4.86, -0.48 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	-0.50 (-5.05, -0.50 , 0.00)	-0.50 (-5.05, -0.50 , 0.00)	-0.87 (-0.96, -1.00 , -1.52)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 5 \frac{t}{2400})$	-17.84 (-21.17, -17.82 , -13.11)	-17.84 (-21.17, -17.82 , -13.11)	-16.74 (-16.99, -16.70 , -16.56)	-15.50 (-15.79, -15.47 , -15.28)	-16.79 (-20.82, -16.56 , -13.10)	-16.79 (-20.82, -16.56 , -13.10)	-17.18 (-17.46, -17.18 , -16.99)	-15.00 (-15.35, -15.01 , -14.94)
$\sin(2\pi \cdot 5 \frac{t}{2400})$	-18.46 (-23.04, -19.50 , -16.55)	-18.46 (-23.04, -19.50 , -16.55)	-18.38 (-15.09, -18.38 , -15.22)	-18.93 (-14.40, -14.02 , -13.94)	-17.82 (-21.73, -19.25 , -16.53)	-22.63 (-21.73, -19.25 , -16.53)	-19.53 (-16.02, -19.68 , -15.54)	-17.69 (-13.71, -19.41 , -13.32)
$\sin(2\pi \cdot 10 \frac{t}{2400})$	-23.71 (-26.10, -23.71 , -20.87)	-23.71 (-26.10, -23.71 , -20.87)	-18.93 (-19.35, -18.86 , -18.72)	-17.82 (-18.27, -22.58 , -19.36)	-22.63 (-24.77, -22.58 , -19.36)	-22.63 (-19.92, -19.45 , -19.33)	-19.53 (-18.04, -17.61 , -19.33)	-17.69 (-14.30, -13.96 , -13.84)

Table 13: Location L106 location. The **coral** tiny numbers indicate the 2.5th and 97.5th quantiles, the **navy-blue** tiny numbers represent the median values and normal-sized numbers refer to the mean value of the estimated coefficients model-wise

Frequencies	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	63.84 (55.23, 63.84 , 72.52)	63.84 (55.23, 63.84 , 72.52)	62.66 (54.24, 62.72 , 70.07)	56.40 (46.80, 56.29 , 67.78)	63.22 (54.85, 63.22 , 71.28)	63.22 (54.85, 63.22 , 71.28)	62.78 (54.39, 62.78 , 71.12)	61.24 (52.25, 61.24 , 71.12)
$\cos(2\pi \cdot 15 \frac{t}{2400})$	8.58 (6.39, 8.58 , 10.66)	8.58 (6.39, 8.58 , 10.66)	6.02 (5.88, 6.02 , 6.16)	0.00 (0.00, 0.00 , 0.00)	7.67 (6.04, 7.69 , 10.04)	7.67 (6.04, 7.69 , 10.04)	5.86 (5.77, 5.83 , 6.03)	1.19 (0.00, 0.00 , 0.03)
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.50 (0.00, 0.50 , 4.26)	0.50 (0.00, 0.50 , 4.26)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.70 (0.00, 0.70 , 3.66)	0.70 (0.00, 0.70 , 3.66)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 20 \frac{t}{2400})$	0.37 (0.00, 0.37 , 2.35)	0.37 (0.00, 0.37 , 2.35)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 30 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 15 \frac{t}{2400})$	-0.27 (-2.72, -0.27 , 0.00)	-0.27 (-2.72, -0.27 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	-0.95 (-3.23, -0.95 , 0.00)	-0.95 (-3.23, -0.95 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 25 \frac{t}{2400})$	-2.14 (-4.65, -2.85 , 0.00)	-2.14 (-4.65, -2.85 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	-1.73 (-3.37, -0.00 , 0.00)	-1.73 (-3.37, -0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 5 \frac{t}{2400})$	-10.57 (-12.37, -10.57 , -9.62)	-10.57 (-12.37, -10.57 , -9.62)	-9.53 (-9.77, -9.53 , -9.30)	0.00 (0.00, 0.00 , 0.00)	-10.19 (-10.98, -10.19 , -9.20)	-10.19 (-10.98, -10.19 , -9.20)	-9.36 (-9.54, -9.35 , -9.26)	-5.10 (-9.54, -9.35 , -9.26)
$\sin(2\pi \cdot 10 \frac{t}{2400})$	-11.68 (-13.72, -11.68 , -9.99)	-11.68 (-13.72, -11.68 , -9.99)	-10.57 (-10.71, -10.56 , -9.43)	0.00 (0.00, 0.00 , 0.00)	-11.08 (-11.95, -11.18 , -9.23)	-11.08 (-11.95, -11.18 , -9.23)	-10.87 (-10.80, -10.66 , -6.58)	-1.35 (-3.80, -1.00 , 0.00)
$\sin(2\pi \cdot 5 \frac{t}{2400})$	-19.14 (-21.55, -19.33 , -17.18)	-19.14 (-21.53, -19.33 , -17.18)	-13.64 (-13.95, -13.60 , -13.46)	-0.66 (-6.58, -0.00 , 0.00)	-17.54 (-20.46, -17.21 , -16.28)	-17.54 (-20.46, -17.21 , -16.28)	-13.98 (-14.30, -13.96 , -13.84)	-10.87 (-14.30, -10.66 , -10.08)

Table 14: Location L107 location. The **coral** tiny numbers indicate the 2.5th and 97.5th quantiles, the **navy-blue** tiny numbers represent the median values and normal-sized numbers refer to the mean value of the estimated coefficients model-wise

Frequencies	L1MTE _{AIC}	L1MTE _{BIC}	L1LSA _{AIC}	L1LSA _{BIC}	L2MTE _{AIC}	L2MTE _{BIC}	L2LSA _{AIC}	L2LSA _{BIC}
(Intercept)	59.99 (50.50, 59.81 , 69.32)	59.99 (50.50, 59.81 , 69.32)	61.16 (51.72, 61.03 , 70.61)	54.92 (42.64, 53.47 , 69.57)	60.21 (50.75, 60.33 , 69.31)	60.21 (50.75, 60.33 , 69.31)	61.79 (52.29, 61.79 , 71.24)	60.23 (50.57, 60.23 , 69.77)
$\cos(2\pi \cdot 15 \frac{t}{2400})$	6.58 (5.64, 6.50 , 7.94)	6.58 (5.64, 6.50 , 7.94)	4.38 (4.13, 4.39 , 4.56)	0.00 (0.00, 0.00 , 0.00)	6.56 (5.69, 6.25 , 8.38)	6.56 (5.69, 6.25 , 8.38)	6.35 (6.10, 6.34 , 6.65)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 20 \frac{t}{2400})$	4.82 (3.98, 4.82 , 6.15)	4.82 (3.98, 4.82 , 6.15)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	4.52 (3.63, 4.33 , 6.06)	4.52 (3.63, 4.33 , 6.06)	3.28 (3.08, 3.28 , 3.52)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 10 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.08 (0.00, 0.08 , 0.70)	0.08 (0.00, 0.08 , 0.70)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\cos(2\pi \cdot 97 \frac{t}{2400})$	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	-0.00 (-0.06, -0.00 , 0.00)	-0.00 (-0.06, -0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$\sin(2\pi \cdot 5 \frac{t}{2400})$	-10.57 (-15.22, -14.50 , -14.02)	-10.57 (-15.22, -14.50 , -14.02)	-12.48 (-12.48, -12.19 , -11.85)	-2.46 (-5.94, -2.46 , 0.00)	-14.88 (-15.74, -14.84 , -14.23)	-14.88 (-15.74, -14.84 , -14.23)	-13.75 (-13.75, -13.25 , -13.14)	-9.91 (-10.13, -9.90 , -9.67)
$\cos(2\pi \cdot 5 \frac{t}{2400})$	-16.08 (-17.44, -16.83 , -14.52)	-16.08 (-17.44, -16.83 , -14.52)	-13.60 (-15.07, -15.52 , -13.31)	-2.90 (-12.86, -2.90 , 0.00)	-15.44 (-17.93, -15.41 , -13.56)	-15.44 (-17.93, -15.41 , -13.56)	-14.69 (-16.49, -14.63 , -14.46)	-11.53 (-13.40, -11.53 , -11.25)
$\sin(2\pi \cdot 10 \frac{t}{2400})$	-16.40 (-17.85, -16.54 , -15.20)	-16.40 (-17.85, -16.54 , -15.20)	-14.71 (-13.99, -14.70 , -14.40)	-3.20 (-11.67, -3.20 , 0.00)	-16.94 (-17.04, -16.94 , -15.36)	-16.94 (-17.04, -16.94 , -15.36)	-16.00 (-15.11, -15.89 , -15.77)	-13.09 (-11.79, -13.07 , -12.83)

Table 15: Location L108 location. The **coral** tiny numbers indicate the 2.5th and 97.5th quantiles, the **navy-blue** tiny numbers represent the median values and normal-sized numbers refer to the mean value of the estimated coefficients model-wise

5.1.3 Error Metrics

Location: L101															
Methods	TRAIN							TEST							
	HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	
L1LSA	MEB	12.63	13.12	18.05	99.31	103.40	23.44	14.07	10.24	10.73	18.05	99.31	106.98	23.44	13.09
	Actual	12.40	12.89	17.77	99.27	103.22	22.28	13.66	9.95	10.44	17.77	104.83	106.45	15.63	12.68
	ADL1LASSO	9.53	9.96	15.98	97.97	100.04	17.68	8.35	10.76	11.24	15.98	105.15	103.11	16.79	12.11
AIC	MEB	13.22	13.71	18.71	99.03	103.30	24.47	14.86	10.59	11.08	18.71	99.03	107.29	24.47	13.19
	Actual	12.88	13.37	18.30	99.03	102.95	23.18	14.29	10.24	10.73	18.30	104.47	106.66	16.40	12.57
	ADL1LASSO	9.59	10.02	15.96	103.08	100.02	18.04	8.67	11.98	12.47	15.96	113.89	103.65	19.34	13.27
L2MTE	MEB	12.59	13.08	18.85	95.15	98.30	23.05	12.78	9.28	9.77	18.85	95.15	101.70	23.05	10.81
	Actual	12.37	12.85	18.66	93.32	96.62	21.86	12.07	9.04	9.53	18.66	98.47	99.28	14.14	10.33
	ADL1LASSO	9.64	10.07	16.17	99.81	100.03	17.89	8.51	11.02	11.51	16.17	107.43	103.23	17.68	12.12
L2LSA	MEB	12.66	13.15	19.00	94.70	97.71	23.20	12.70	9.35	9.84	19.00	94.70	100.98	23.20	10.92
	Actual	12.29	12.77	18.26	95.80	98.95	22.16	12.77	9.22	9.71	18.26	101.11	102.04	14.98	10.75
	ADL1LASSO	9.91	10.35	16.10	102.25	100.02	18.75	8.88	11.88	12.37	16.10	110.03	104.06	19.57	13.31
BIC	MEB	13.85	14.34	19.18	99.21	104.01	25.53	16.16	11.23	11.72	19.18	99.21	107.87	25.53	14.52
	Actual	13.52	14.02	18.77	99.27	103.85	24.25	15.60	10.91	11.40	18.77	104.67	107.23	17.41	13.94
	ADL1LASSO	9.43	9.87	15.41	102.94	99.99	18.04	8.66	11.99	12.48	15.41	114.75	103.44	18.78	12.76
L1MTE	MEB	15.45	15.95	20.56	100.24	105.91	28.43	18.53	13.20	13.69	20.56	100.24	109.62	28.43	17.02
	Actual	15.06	15.55	20.10	100.15	105.48	27.04	17.84	12.77	13.26	20.10	105.51	108.76	20.82	16.29
	ADL1LASSO	9.42	9.87	15.52	103.21	100.00	18.04	8.74	11.99	12.48	15.52	115.48	103.20	18.84	12.58
L2MTE	MEB	12.59	13.08	18.85	95.15	98.30	23.05	12.78	9.28	9.77	18.85	95.15	101.70	23.05	10.81
	Actual	12.37	12.85	18.66	93.32	96.62	21.86	12.07	9.04	9.53	18.66	98.47	99.28	14.14	10.33
	ADL1LASSO	9.64	10.07	16.17	99.81	100.03	17.89	8.51	11.02	11.51	16.17	107.43	103.23	17.68	12.12
L2LSA	MEB	12.66	13.15	19.00	94.70	97.71	23.20	12.70	9.35	9.84	19.00	94.70	100.98	23.20	10.92
	Actual	12.29	12.77	18.26	95.80	98.95	22.16	12.77	9.22	9.71	18.26	101.11	102.04	14.98	10.75
	ADL1LASSO	9.91	10.35	16.10	102.25	100.02	18.75	8.88	11.88	12.37	16.10	110.03	104.06	19.57	13.31

Table 16: Comparison of seven performance metrics (HL, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for L1LSA, L2LSA, L1MTE, and L2MTE methods, evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L101

Methods		Location: L102														
		TRAIN							TEST							
		HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	
AIC	L1LSA	MEB	11.43	11.92	16.38	99.41	102.31	18.16	12.55	10.62	11.11	16.38	99.41	102.07	18.16	12.23
	Actual	11.28	11.77	16.17	99.46	102.27	17.61	12.29	10.51	11.00	16.17	100.20	101.53	16.05	11.98	
	ADL1LASSO	7.93	8.38	13.57	101.65	100.02	12.47	7.23	11.69	12.18	13.57	104.70	100.57	17.16	12.62	
L2MTE	L2LSA	MEB	11.45	11.94	16.41	99.39	102.29	18.22	12.53	10.64	11.13	16.41	99.39	102.04	18.22	12.20
	Actual	11.27	11.76	16.19	99.39	102.17	17.62	12.23	10.49	10.98	16.19	100.12	101.53	16.07	11.86	
	ADL1LASSO	7.93	8.38	13.57	101.65	100.02	12.47	7.23	11.69	12.18	13.57	104.70	100.57	17.16	12.62	
L2MTE	L1MTE	MEB	10.74	11.22	16.38	99.88	100.52	16.65	11.24	10.12	10.61	16.38	99.88	100.79	16.65	11.46
	Actual	10.53	11.02	16.10	98.34	99.55	16.03	10.87	9.96	10.45	16.10	98.99	99.38	14.67	11.18	
	ADL1LASSO	8.28	8.71	14.07	100.68	100.01	12.79	7.57	11.23	11.72	14.07	102.56	100.47	16.46	12.44	
L2MTE	L2LSA	MEB	10.63	11.12	16.27	100.28	100.86	16.40	11.18	10.11	10.60	16.27	100.28	101.20	16.40	11.49
	Actual	10.52	11.01	16.13	101.03	101.31	15.91	10.94	9.93	10.41	16.13	101.61	101.12	14.46	11.12	
	ADL1LASSO	8.48	8.91	14.60	100.43	100.01	13.18	7.62	11.19	11.68	14.60	102.25	99.97	16.59	12.29	
BIC	L1LSA	MEB	11.43	11.92	16.38	99.41	102.31	18.16	12.55	10.62	11.11	16.38	99.41	102.07	18.16	12.23
	Actual	11.28	11.77	16.17	99.46	102.27	17.61	12.29	10.51	11.00	16.17	100.20	101.53	16.05	11.98	
	ADL1LASSO	7.93	8.38	13.57	101.65	100.02	12.47	7.23	11.69	12.18	13.57	104.70	100.57	17.16	12.62	
L2MTE	L2LSA	MEB	13.45	13.94	18.16	100.46	104.95	21.74	15.08	12.76	13.25	18.16	100.46	105.18	21.74	14.62
	Actual	13.27	13.77	17.94	100.42	104.93	21.11	14.71	12.58	13.07	17.94	101.25	104.57	19.80	14.34	
	ADL1LASSO	7.93	8.38	13.62	101.94	100.02	12.48	7.19	11.71	12.20	13.62	105.12	100.59	17.24	12.55	
L2MTE	L1MTE	MEB	10.74	11.22	16.38	99.88	100.52	16.65	11.24	10.12	10.61	16.38	99.88	100.79	16.65	11.46
	Actual	10.53	11.02	16.10	98.34	99.55	16.03	10.87	9.96	10.45	16.10	98.99	99.38	14.67	11.18	
	ADL1LASSO	8.28	8.71	14.07	100.68	100.01	12.79	7.57	11.23	11.72	14.07	102.56	100.47	16.46	12.44	
L2MTE	L2LSA	MEB	10.63	11.12	16.27	100.28	100.86	16.40	11.18	10.11	10.60	16.27	100.28	101.20	16.40	11.49
	Actual	10.52	11.01	16.13	101.03	101.31	15.91	10.94	9.93	10.41	16.13	101.61	101.12	14.46	11.12	
	ADL1LASSO	8.48	8.91	14.60	100.43	100.01	13.18	7.62	11.19	11.68	14.60	102.25	99.97	16.59	12.29	

Table 17: Comparison of seven performance metrics (HL, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for L1LSA, L2LSA, L1MTE, and L2MTE methods, evaluated using model averaging from MEBoot samples (**MEB**), the corresponding models applied on the original traffic data (**Actual**), and post-hoc variable selection algorithm **ADL1LASSO** models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L102

Methods		Location: L103														
		TRAIN							TEST							
		HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	
AIC	L1LSA	MEB	11.58	12.07	16.95	99.22	100.30	17.89	11.81	13.98	14.47	16.95	99.22	96.49	17.89	14.59
	Actual	11.39	11.88	16.71	99.28	100.34	17.28	11.49	13.82	14.31	16.71	94.22	96.12	20.82	14.21	
	ADL1LASSO	8.61	9.04	14.35	101.84	99.99	12.94	7.51	15.99	16.49	14.35	100.04	96.28	22.99	16.66	
L2MTE	L2LSA	MEB	11.54	12.03	16.93	99.12	100.07	17.79	11.73	14.03	14.53	16.93	99.12	96.21	17.79	14.65
	Actual	11.34	11.83	16.68	99.19	100.08	17.16	11.37	13.87	14.36	16.68	94.16	95.87	20.79	14.32	
	ADL1LASSO	8.61	9.04	14.35	101.84	99.99	12.94	7.51	15.99	16.49	14.35	100.04	96.28	22.99	16.66	
L2MTE	L1MTE	MEB	11.42	11.91	17.07	102.78	101.89	17.15	11.74	13.76	14.25	17.07	102.78	98.86	17.15	14.38
	Actual	11.27	11.75	16.86	103.07	102.10	16.58	11.40	13.51	14.00	16.86	97.82	98.69	19.76	13.96	
	ADL1LASSO	8.85	9.28	14.76	103.79	100.00	13.21	7.61	15.67	16.16	14.76	103.35	95.84	22.40	16.26	
L2MTE	L1LSA	MEB	11.42	11.91	17.08	102.61	101.66	17.14	11.65	13.83	14.33	17.08	102.61	98.69	17.14	14.39
	Actual	11.22	11.71	16.85	101.98	100.99	16.52	11.16	13.86	14.36	16.85	96.82	97.66	20.13	13.95	
	ADL1LASSO	8.69	9.12	14.44	100.96	100.00	13.03	7.49	16.36	16.86	14.44	98.82	95.85	23.36	17.17	
BIC	L1LSA	MEB	12.45	12.94	17.66	99.33	101.30	19.73	12.93	14.57	15.07	17.66	99.33	97.03	19.73	15.41
	Actual	12.27	12.76	17.43	99.40	101.20	19.07	12.58	14.41	14.90	17.43	94.31	96.64	22.37	14.96	
	ADL1LASSO	8.57	9.02	13.97	101.18	100.03	12.94	7.80	16.05	16.54	13.97	98.70	96.44	23.11	16.72	
L2MTE	L2LSA	MEB	12.39	12.88	17.61	99.19	101.03	19.60	12.83	14.61	15.10	17.61	99.19	96.80	19.60	15.48
	Actual	12.21	12.70	17.38	99.27	101.00	18.97	12.55	14.44	14.94	17.38	94.21	96.46	22.37	15.12	
	ADL1LASSO	8.57	9.02	13.97	101.18	100.03	12.94	7.80	16.05	16.54	13.97	98.70	96.44	23.11	16.72	
L2MTE	L1MTE	MEB	11.42	11.91	17.07	102.78	101.89	17.15	11.74	13.76	14.25	17.07	102.78	98.86	17.15	14.38
	Actual	11.27	11.75	16.86	103.07	102.10	16.58	11.40	13.51	14.00	16.86	97.82	98.69	19.76	13.96	
	ADL1LASSO	8.85	9.28	14.76	103.79	100.00	13.21	7.61	15.67	16.16	14.76	103.35	95.84	22.40	16.26	
L2MTE	L1LSA	MEB	11.42	11.91	17.10	102.63	101.64	17.13	11.65	13.84	14.33	17.10	102.63	98.65	17.13	14.40
	Actual	11.22	11.71	16.85	101.98	100.99	16.52	11.16	13.86	14.36	16.85	96.82	97.66	20.13	13.95	
	ADL1LASSO	8.69	9.12	14.43	100.94	100.00	13.03	7.52	16.36	16.85	14.43	98.78	95.80	23.35	17.21	

Table 18: Comparison of seven performance metrics (HL, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for L1LSA, L2LSA, L1MTE, and L2MTE methods, evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L103

Methods		Location: L104														
		TRAIN							TEST							
			HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE		HL	MAE	RMSE	MAPE	MdAPE	SMAPE
AIC	L1LSA	MEB	13.70	14.19	18.02	100.27	98.93	25.20	20.33	13.63	14.12	18.02	100.27	92.41	25.20	21.30
	Actual	MEB	13.46	13.95	17.71	100.24	99.07	24.50	19.84	13.47	13.96	17.71	94.36	92.09	25.39	20.88
	ADL1LASSO	MEB	8.99	9.44	14.17	104.76	99.98	17.10	10.74	13.81	14.30	14.17	102.04	96.25	25.86	19.07
L2MTE	L2LSA	MEB	14.34	14.84	18.51	101.20	100.16	26.41	21.80	13.97	14.46	18.51	101.20	93.35	26.41	22.32
	Actual	MEB	14.11	14.61	18.21	101.18	100.20	25.72	21.34	13.81	14.30	18.21	95.21	93.10	26.14	22.01
	ADL1LASSO	MEB	9.20	9.63	14.68	113.53	99.99	17.29	10.85	14.39	14.88	14.68	113.18	95.53	27.29	19.61
L2MTE	L1MTE	MEB	13.40	13.89	18.30	104.55	102.31	24.36	18.96	13.68	14.17	18.30	104.55	96.89	24.36	20.45
	Actual	MEB	12.98	13.47	17.91	104.84	102.50	23.32	18.11	13.48	13.97	17.91	98.90	97.39	24.79	20.02
	ADL1LASSO	MEB	9.24	9.67	14.76	101.01	100.03	17.03	10.60	14.22	14.71	14.76	97.35	95.28	25.82	19.81
L2MTE	L2LSA	MEB	14.41	14.90	19.40	112.70	110.19	26.32	21.31	13.91	14.40	19.40	112.70	105.21	26.32	21.86
	Actual	MEB	13.01	13.50	17.84	104.35	102.30	23.41	18.07	13.21	13.70	17.84	98.41	96.92	24.45	19.28
	ADL1LASSO	MEB	9.07	9.51	14.48	108.68	99.98	17.15	10.57	13.96	14.45	14.48	107.53	96.38	26.44	19.37
BIC	L1LSA	MEB	16.00	16.50	20.34	105.10	103.86	29.31	25.18	14.77	15.26	20.34	105.10	97.17	29.31	23.95
	Actual	MEB	16.29	16.78	20.59	105.85	104.58	29.38	25.31	14.81	15.31	20.59	99.23	97.61	27.98	23.97
	ADL1LASSO	MEB	9.21	9.64	14.66	102.00	100.00	17.16	10.61	13.73	14.22	14.66	98.08	95.18	25.06	19.07
L2MTE	L2LSA	MEB	16.58	17.07	20.95	106.07	104.76	30.20	25.96	15.03	15.52	20.95	106.07	98.25	30.20	24.43
	Actual	MEB	16.28	16.78	20.58	105.73	104.46	29.37	25.20	14.81	15.31	20.58	99.12	97.50	27.98	23.86
	ADL1LASSO	MEB	12.99	13.48	17.56	101.19	99.88	23.47	18.50	13.18	13.67	17.56	95.31	93.52	24.60	19.87
L2MTE	L1MTE	MEB	13.40	13.89	18.30	104.55	102.31	24.36	18.96	13.68	14.17	18.30	104.55	96.89	24.36	20.45
	Actual	MEB	12.98	13.47	17.91	104.84	102.50	23.32	18.11	13.48	13.97	17.91	98.90	97.39	24.79	20.02
	ADL1LASSO	MEB	9.24	9.67	14.76	101.01	100.03	17.03	10.60	14.22	14.71	14.76	97.35	95.28	25.82	19.81
L2MTE	L2LSA	MEB	14.41	14.90	19.40	112.70	110.19	26.32	21.31	13.91	14.40	19.40	112.70	105.21	26.32	21.86
	Actual	MEB	13.01	13.50	17.84	104.35	102.30	23.41	18.07	13.21	13.70	17.84	98.41	96.92	24.45	19.28
	ADL1LASSO	MEB	9.07	9.51	14.48	108.68	99.98	17.15	10.57	13.96	14.45	14.48	107.53	96.38	26.44	19.37

Table 19: Comparison of seven performance metrics (HL, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for L1LSA, L2LSA, L1MTE, and L2MTE methods, evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L104

Methods		Location: L106														
		TRAIN							TEST							
		HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	
AIC	L1LSA	MEB	17.04	17.53	23.46	99.36	103.21	20.88	14.80	15.65	16.15	23.46	99.36	105.79	20.88	14.17
	Actual	MEB	16.86	17.35	23.20	99.39	103.28	20.34	14.52	15.43	15.93	23.20	102.58	105.35	17.55	13.90
	ADL1LASSO	MEB	13.02	13.46	21.30	98.69	99.96	16.34	8.70	17.61	18.10	21.30	103.88	101.42	19.97	14.19
L2MTE	L1LSA	MEB	18.02	18.51	24.62	98.74	102.98	22.17	15.72	15.66	16.15	24.62	98.74	105.58	22.17	14.33
	Actual	MEB	17.79	18.28	24.34	98.77	102.85	21.53	15.40	15.38	15.88	24.34	101.59	105.13	17.79	13.89
	ADL1LASSO	MEB	12.44	12.89	20.06	102.35	100.01	15.93	8.55	18.93	19.43	20.06	111.42	102.35	21.64	15.18
L2MTE	L1MTE	MEB	17.90	18.40	25.40	99.05	101.43	21.68	14.59	14.97	15.46	25.40	99.05	104.03	21.68	12.83
	Actual	MEB	17.86	18.36	25.60	96.70	99.42	21.27	13.66	14.86	15.35	25.60	99.27	101.48	17.05	12.26
	ADL1LASSO	MEB	13.83	14.27	22.54	98.79	100.00	17.39	9.14	18.45	18.94	22.54	104.30	101.67	20.97	14.75
L2MTE	L2LSA	MEB	17.94	18.43	25.82	97.50	100.09	21.65	14.11	14.65	15.14	25.82	97.50	102.12	21.65	12.18
	Actual	MEB	17.77	18.26	25.55	97.25	100.12	21.14	13.75	14.10	14.59	25.55	99.71	101.49	16.37	11.54
	ADL1LASSO	MEB	13.98	14.42	22.71	102.10	99.99	17.83	9.41	19.64	20.14	22.71	110.19	101.54	22.77	15.16
BIC	L1LSA	MEB	19.38	19.88	25.65	99.63	104.85	24.12	17.31	17.06	17.55	25.65	99.63	107.29	24.12	15.74
	Actual	MEB	19.18	19.67	25.38	99.66	104.74	23.50	17.02	16.80	17.30	25.38	102.48	106.84	19.83	15.39
	ADL1LASSO	MEB	12.38	12.84	19.93	102.73	99.99	15.97	8.72	19.60	20.09	19.93	111.66	102.45	22.56	15.73
L2MTE	L2LSA	MEB	19.11	19.60	25.52	99.00	103.99	23.77	16.79	16.64	17.13	25.52	99.00	106.52	23.77	15.14
	Actual	MEB	18.91	19.40	25.25	99.01	103.95	23.17	16.52	16.39	16.88	25.25	101.79	106.13	19.37	14.73
	ADL1LASSO	MEB	12.38	12.84	19.93	102.73	99.99	15.97	8.72	19.60	20.09	19.93	111.66	102.45	22.56	15.73
L2MTE	L1MTE	MEB	17.90	18.40	25.40	99.05	101.43	21.68	14.59	14.97	15.46	25.40	99.05	104.03	21.68	12.83
	Actual	MEB	17.86	18.36	25.60	96.70	99.42	21.27	13.66	14.86	15.35	25.60	99.27	101.48	17.05	12.26
	ADL1LASSO	MEB	13.83	14.27	22.54	98.79	100.00	17.39	9.14	18.45	18.94	22.54	104.30	101.67	20.97	14.75
L2MTE	L2LSA	MEB	17.94	18.43	25.82	97.50	100.09	21.65	14.11	14.65	15.14	25.82	97.50	102.12	21.65	12.18
	Actual	MEB	17.77	18.26	25.55	97.25	100.12	21.14	13.75	14.10	14.59	25.55	99.71	101.49	16.37	11.54
	ADL1LASSO	MEB	13.98	14.42	22.71	102.10	99.99	17.83	9.41	19.64	20.14	22.71	110.19	101.54	22.77	15.16

Table 20: Comparison of seven performance metrics (HL, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for L1LSA, L2LSA, L1MTE, and L2MTE methods, evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L106

Methods		Location: L107														
		TRAIN							TEST							
		HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	
AIC	L1LSA	MEB	12.00	12.49	16.12	99.46	102.39	22.69	16.49	11.58	12.07	16.12	99.46	101.66	22.69	16.36
	Actual	11.87	12.36	15.96	99.53	102.27	22.04	16.17	11.48	11.97	15.96	98.20	101.16	21.82	16.01	
	ADL1LASSO	8.10	8.54	12.96	101.66	99.99	14.74	9.07	11.25	11.74	12.96	101.76	100.02	20.24	14.86	
L2MTE	L2LSA	MEB	12.04	12.53	16.13	99.63	102.55	22.76	16.64	11.66	12.15	16.13	99.63	101.93	22.76	16.57
	Actual	11.91	12.40	15.96	99.67	102.40	22.11	16.24	11.55	12.04	15.96	98.36	101.43	21.92	16.25	
	ADL1LASSO	8.10	8.54	12.96	101.66	99.99	14.74	9.07	11.25	11.74	12.96	101.76	100.02	20.24	14.86	
L2MTE	L1MTE	MEB	11.17	11.66	15.86	100.67	101.51	20.43	14.55	10.54	11.03	15.86	100.67	100.61	20.43	14.03
	Actual	11.26	11.75	15.83	99.51	101.15	20.51	14.69	10.67	11.16	15.83	98.00	99.37	19.87	13.96	
	ADL1LASSO	8.61	9.04	14.02	101.73	100.00	15.51	9.35	11.77	12.26	14.02	101.78	99.99	21.15	15.41	
L2MTE	L2LSA	MEB	11.16	11.65	16.06	100.59	100.87	20.07	14.29	10.51	11.00	16.06	100.59	99.86	20.07	13.74
	Actual	11.03	11.52	15.93	99.74	100.09	19.41	13.95	10.33	10.82	15.93	98.09	98.74	18.50	13.22	
	ADL1LASSO	8.84	9.27	14.21	101.65	99.99	16.06	9.85	12.10	12.59	14.21	101.87	99.79	21.90	15.74	
BIC	L1LSA	MEB	14.70	15.19	18.65	102.45	107.70	27.84	20.61	14.10	14.59	18.65	102.45	106.77	27.84	20.03
	Actual	15.17	15.66	19.07	102.87	108.45	28.15	20.96	14.52	15.01	19.07	101.56	107.14	27.61	20.13	
	ADL1LASSO	8.86	9.29	14.17	102.26	99.98	16.11	9.71	12.48	12.97	14.17	102.73	99.58	22.48	16.48	
BIC	L2LSA	MEB	19.72	20.22	23.42	112.70	122.03	36.04	31.97	19.28	19.78	23.42	112.70	121.10	36.04	31.19
	Actual	19.70	20.20	23.37	112.58	122.09	35.38	31.82	19.24	19.74	23.37	111.33	120.32	35.13	30.69	
	ADL1LASSO	8.98	9.41	14.60	103.19	100.02	16.10	9.99	12.97	13.46	14.60	104.14	99.38	23.21	16.54	
L2MTE	L1MTE	MEB	11.17	11.66	15.86	100.67	101.51	20.43	14.55	10.54	11.03	15.86	100.67	100.61	20.43	14.03
	Actual	11.26	11.75	15.83	99.51	101.15	20.51	14.69	10.67	11.16	15.83	98.00	99.37	19.87	13.96	
	ADL1LASSO	8.61	9.04	14.02	101.73	100.00	15.51	9.35	11.77	12.26	14.02	101.78	99.99	21.15	15.41	
L2MTE	L2LSA	MEB	11.16	11.65	16.06	100.59	100.87	20.07	14.29	10.51	11.00	16.06	100.59	99.86	20.07	13.74
	Actual	11.03	11.52	15.93	99.74	100.09	19.41	13.95	10.33	10.82	15.93	98.09	98.74	18.50	13.22	
	ADL1LASSO	8.84	9.27	14.21	101.65	99.99	16.06	9.85	12.10	12.59	14.21	101.87	99.79	21.90	15.74	

Table 21: Comparison of seven performance metrics (HL, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for L1LSA, L2LSA, L1MTE, and L2MTE methods, evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L107

Methods	Location: L108															
	TRAIN								TEST							
	HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE		HL	MAE	RMSE	MAPE	MdAPE	SMAPE	SMdAPE	
AIC	MEB	10.03	10.51	15.26	98.99	99.39	19.91	12.96	11.59	12.08	15.26	98.99	101.66	19.91	15.00	
	Actual	9.85	10.34	15.02	99.11	99.44	18.93	12.57	11.38	11.87	15.02	103.22	101.11	19.64	14.62	
	ADL1LASSO	7.36	7.79	12.51	99.78	100.00	14.03	8.35	12.04	12.53	12.51	105.11	102.28	19.96	15.56	
L2MTE	MEB	10.85	11.34	15.94	99.23	100.91	22.18	14.03	12.65	13.14	15.94	99.23	103.34	22.18	16.94	
	Actual	10.68	11.17	15.70	99.35	100.92	21.19	13.59	12.45	12.94	15.70	103.68	102.72	22.33	16.59	
	ADL1LASSO	7.35	7.79	12.45	101.94	100.02	14.23	8.34	12.34	12.83	12.45	110.25	102.24	20.72	15.48	
L2MTA	MEB	10.00	10.49	15.36	103.96	102.33	19.42	13.07	11.51	12.00	15.36	103.96	104.78	19.42	14.93	
	Actual	9.79	10.28	15.08	103.89	102.64	18.43	12.60	11.31	11.80	15.08	107.93	104.40	18.86	14.52	
	ADL1LASSO	7.43	7.86	12.72	99.62	100.00	14.19	8.23	11.97	12.46	12.72	104.91	101.72	19.86	15.43	
L2LSA	MEB	10.01	10.50	15.37	104.37	102.64	19.48	13.10	11.61	12.09	15.37	104.37	105.18	19.48	15.09	
	Actual	9.80	10.29	15.08	103.78	102.74	18.45	12.51	11.31	11.80	15.08	107.82	104.49	18.87	14.38	
	ADL1LASSO	7.36	7.79	12.51	99.78	100.00	14.03	8.35	12.04	12.53	12.51	105.11	102.28	19.96	15.56	
L1LSA	MEB	12.25	12.74	17.24	100.38	103.60	25.36	15.86	14.23	14.72	17.24	100.38	106.37	25.36	19.45	
	Actual	12.05	12.54	16.98	100.42	103.48	24.28	15.37	13.97	14.46	16.98	104.97	105.76	25.77	18.96	
	ADL1LASSO	7.36	7.80	12.53	102.86	100.02	14.25	8.36	12.47	12.96	12.53	112.04	102.58	21.02	15.48	
BIC	MEB	19.64	20.14	24.23	113.52	122.67	38.70	30.34	22.15	22.64	24.23	113.52	126.68	38.70	34.11	
	Actual	21.57	22.07	26.00	115.47	125.98	40.73	32.60	24.07	24.57	26.00	121.33	129.68	42.87	36.40	
	ADL1LASSO	7.57	8.00	13.03	107.38	99.99	14.75	8.37	13.04	13.53	13.03	122.61	102.16	22.34	15.89	
L2MTE	MEB	10.00	10.49	15.36	103.96	102.33	19.42	13.07	11.51	12.00	15.36	103.96	104.78	19.42	14.93	
	Actual	9.79	10.28	15.08	103.89	102.64	18.43	12.60	11.31	11.80	15.08	107.93	104.40	18.86	14.52	
	ADL1LASSO	7.43	7.86	12.72	99.62	100.00	14.19	8.23	11.97	12.46	12.72	104.91	101.72	19.86	15.43	
L2LSA	MEB	10.01	10.50	15.37	104.37	102.64	19.48	13.10	11.61	12.09	15.37	104.37	105.18	19.48	15.09	
	Actual	9.80	10.29	15.08	103.78	102.74	18.45	12.51	11.31	11.80	15.08	107.82	104.49	18.87	14.38	
	ADL1LASSO	7.36	7.79	12.51	99.78	100.00	14.03	8.35	12.04	12.53	12.51	105.11	102.28	19.96	15.56	

Table 22: Comparison of seven performance metrics (HL, MAE, RMSE, MAPE, MdAPE, SMAPE, SMdAPE) for L1LSA, L2LSA, L1MTE, and L2MTE methods, evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L108

5.1.4 Location-wise Errors regarding Metrics Among Models

Methods	Metric: Huber Loss																														
	ℓ_1 LSA _{AIC}				ℓ_2 LSA _{AIC}				ℓ_1 MTE _{AIC}				ℓ_2 MTE _{AIC}				ℓ_1 LSA _{BIC}				ℓ_2 LSA _{BIC}				ℓ_1 MTE _{BIC}						
	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO				
L101	12.63	12.40	9.53	13.22	12.88	9.59	12.59	12.37	9.64	12.66	12.29	9.91	13.85	13.52	9.43	15.45	15.06	9.42	12.59	12.37	9.64	12.66	12.29	9.91	12.66	12.29	9.91				
L101	11.43	11.28	7.93	11.45	11.27	7.93	10.74	10.53	8.28	10.63	10.52	8.48	11.43	11.28	7.93	13.45	13.27	7.93	10.74	10.53	8.28	10.63	10.52	8.48	10.63	10.52	8.48				
L101	11.58	11.39	8.61	11.54	11.34	8.61	11.42	11.27	8.85	11.42	11.22	8.69	12.45	12.27	8.57	12.39	12.21	11.42	11.27	8.85	11.42	11.22	8.69	11.42	11.22	8.69	11.42	11.22	8.69		
L101	12.70	12.50	9.99	12.69	12.41	9.11	12.50	12.38	9.28	12.51	12.30	9.07	13.09	13.09	9.21	16.53	16.28	11.99	12.88	12.54	9.07	12.70	12.50	9.99	12.70	12.50	9.99	12.70	12.50	9.99	
L106	17.04	16.86	13.02	18.02	17.79	12.44	17.90	17.86	13.83	17.94	17.77	13.98	19.18	19.18	12.38	19.11	18.91	12.38	17.90	17.86	13.83	17.94	17.77	13.98	17.94	17.77	13.98	17.94	17.77	13.98	
L106	12.00	11.87	8.10	12.04	11.91	8.10	11.47	11.26	8.61	11.16	11.04	8.84	14.70	15.17	8.86	19.72	19.70	8.95	11.17	11.26	8.61	11.16	11.04	8.84	11.16	11.04	8.84	11.16	11.04	8.84	
L108	10.03	9.85	7.36	10.85	10.68	7.35	10.60	9.79	7.43	10.01	9.80	7.36	12.25	12.05	7.36	19.64	19.64	7.43	21.57	19.64	7.43	10.01	9.80	7.36	10.01	9.80	7.36	10.01	9.80	7.36	
L102	10.62	10.51	11.69	10.64	10.49	11.69	10.12	9.96	11.23	10.11	9.93	11.19	10.62	10.51	11.69	12.76	12.58	11.71	10.12	9.96	11.23	10.11	9.93	11.19	10.11	9.93	11.19	10.11	9.93	11.19	
L102	13.98	13.82	15.99	14.03	13.87	15.99	13.76	13.51	15.67	13.83	13.36	14.36	14.57	14.41	16.05	14.61	14.44	13.73	15.03	14.81	13.18	13.68	13.48	14.22	13.91	13.21	13.91	13.68	13.48	14.22	
L104	13.63	13.47	13.81	13.97	13.58	14.39	13.68	13.48	14.22	13.91	13.21	13.96	14.77	14.14	14.01	13.73	15.03	14.81	13.18	13.68	13.48	14.22	13.91	13.21	13.91	13.68	13.48	14.22	13.91	13.21	13.91
L105	12.55	12.47	13.75	12.55	12.55	13.83	12.51	12.38	13.84	12.41	12.10	12.50	13.74	13.64	13.59	14.60	14.60	14.59	13.54	13.20	14.60	13.74	13.64	14.59	13.54	13.20	14.60	13.74	13.64	14.59	
L107	11.58	11.48	11.25	11.66	11.55	11.25	10.54	10.67	11.37	10.51	10.33	12.10	14.10	14.52	12.48	19.28	19.24	12.97	10.54	10.67	11.77	10.51	10.33	12.10	10.54	10.67	11.77	10.51	10.33	12.10	
L107	11.59	11.48	12.04	12.65	12.45	12.34	11.51	11.31	11.97	11.61	11.31	12.04	14.23	13.97	12.47	22.15	24.07	13.04	11.51	11.31	11.97	11.61	11.31	12.04	11.61	11.31	12.04	11.61	11.31	12.04	

Table 23: Comparison of all the methods (ℓ_1 MTE, ℓ_2 MTE, ℓ_1 LSA, ℓ_2 LSA under AIC and BIC minimization) under Huber Loss metric, evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model.

		Metric: Mean Absolute Error (MAE)																															
		ℓ_1 LSA _{AIC}				ℓ_2 LSA _{AIC}				ℓ_1 MTE _{AIC}				ℓ_2 MTE _{AIC}				ℓ_1 LSA _{BIC}				ℓ_2 LSA _{BIC}				ℓ_1 MTE _{BIC}				ℓ_2 MTE _{BIC}			
Method	Execution	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO		
		L101	13.12	12.89	9.96	13.71	13.37	10.02	13.08	12.85	10.07	13.15	12.77	10.35	14.34	14.02	9.87	15.95	15.55	9.87	13.08	12.85	10.07	13.15	12.77	10.35	13.12	12.89	9.96	13.12	12.89	10.35	
L102	11.92	11.77	8.38	11.94	11.76	8.38	11.22	11.02	8.71	11.12	11.01	8.91	11.91	11.71	9.12	12.94	12.76	9.02	12.88	12.70	9.06	11.91	11.75	9.28	11.91	11.77	9.12	11.91	11.77	9.12	11.91	11.77	9.12
L103	12.07	11.88	9.04	12.03	11.83	9.04	11.96	11.75	9.26	11.45	11.34	9.03	13.48	13.47	10.31	14.45	14.33	9.04	17.01	17.08	13.48	14.45	14.33	9.04	14.45	14.33	9.04	14.45	14.33	9.04	14.45	14.33	9.04
L106	17.53	17.35	13.46	18.51	18.28	12.89	18.40	18.36	18.26	18.43	18.36	18.26	14.42	19.88	19.67	12.84	19.60	19.40	12.84	18.40	18.36	14.27	18.43	18.26	14.42	18.43	18.26	14.42	18.43	18.26	14.42		
L107	12.49	12.36	8.54	12.53	12.40	8.54	11.66	11.75	9.04	11.65	11.52	9.27	15.19	15.66	9.29	20.22	20.20	9.41	11.66	11.75	9.04	11.65	11.52	9.27	11.65	11.52	9.27	11.65	11.52	9.27	11.65	11.52	9.27
L108	10.51	10.34	7.79	11.34	11.17	7.79	10.49	10.28	7.86	10.50	10.29	7.86	12.74	12.54	7.80	20.14	22.07	8.06	11.49	10.28	7.86	10.50	10.29	7.79	10.50	10.29	7.79	10.50	10.29	7.79	10.50	10.29	7.79

Table 24: Comparison of all the methods (ℓ_1 MTE, ℓ_2 MTE, ℓ_1 LSA, ℓ_2 LSA under AIC and BIC minimization) under Mean Absolute Error (MAE) metric , evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model.

		Metric: Root Mean Square Error (RMSE)																																		
		ℓ_1 LSA _{AIC}				ℓ_2 LSA _{AIC}				ℓ_1 MTE _{AIC}				ℓ_2 MTE _{AIC}				ℓ_1 LSA _{BIC}				ℓ_2 LSA _{BIC}				ℓ_1 MTE _{BIC}				ℓ_2 MTE _{BIC}						
Method	Execution	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO					
		L101	18.05	17.77	15.98	18.71	18.30	15.96	18.85	18.66	16.17	19.18	18.26	16.10	18.77	15.41	20.56	20.10	15.52	18.85	18.66	16.17	19.00	18.26	15.96	18.85	18.66	16.17	19.00	18.26	15.96	18.85	18.66	16.17		
L103	16.38	16.17	13.57	16.41	16.19	13.57	16.36	16.10	14.07	16.27	16.13	14.60	16.38	16.17	13.57	18.16	17.94	13.62	19.38	18.10	14.07	16.27	16.13	14.60	19.38	18.10	14.07	16.27	16.13	14.60	19.38	18.10	14.07	16.27	16.13	14.60
L104	16.95	16.71	14.35	16.93	16.68	14.35	17.07	16.86	14.76	17.08	16.85	14.44	17.66	17.43	13.97	17.61	17.38	13.97	17.07	16.86	14.76	17.10	16.85	14.43	19.40	17.84	14.43	19.40	17.84	14.43	19.40	17.84	14.43	19.40	17.84	14.43
L105	17.27	17.20	14.70	17.29	17.14	14.70	18.02	17.81	14.76	18.02	17.81	14.76	18.26	17.98	14.76	20.34	20.54	14.66	19.55	19.25	14.66	19.30	19.09	14.66	19.55	19.25	14.66	19.30	19.09	14.66	19.55	19.25	14.66	19.30	19.09	14.66
L106	23.46	23.29	21.30	24.62	24.34	20.06	25.40	25.60	22.54	25.82	25.55	22.71	25.65	25.38	19.93	15.52	25.25	25.93	19.93	22.54	25.82	25.55	22.71	25.82	25.55	22.71	25.82	25.55	22.71	25.82	25.55	22.71	25.82	25.55	22.71	
L107	16.12	15.96	12.96	16.13	15.96	12.96	15.86	15.83	14.02	16.06	15.93	14.21	18.65	19.07	14.17	23.42	23.37	14.69	15.86	15.83	14.02	16.06	15.93	14.21	15.86	15.83	14.02	16.06	15.93	14.21	15.86	15.83	14.02	16.06	15.93	14.21
L108	15.26	15.02	12.51	15.94	15.70	12.45	15.36	15.08	12.72	15.37	15.08	12.51	17.24	16.98	12.53	24.23	24.00	13.03	15.36	15.08	12.72	15.37	15.08	12.51	15.36	15.08	12.51	15.36	15.08	12.51	15.36	15.08	12.51	15.36	15.08	12.51

Table 25: Comparison of all the methods (ℓ_1 MTE, ℓ_2 MTE, ℓ_1 LSA, ℓ_2 LSA under AIC and BIC minimization) under Root Mean Square Error (RMSE) metric , evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model.

		Metric: Mean Absolute Percentage Error (MAPE)																																		
		ℓ_1 LSA _{AIC}				ℓ_2 LSA _{AIC}				ℓ_1 MTE _{AIC}				ℓ_2 MTE _{AIC}				ℓ_1 LSA _{BIC}				ℓ_2 LSA _{BIC}				ℓ_1 MTE _{BIC}				ℓ_2 MTE _{BIC}						
Method	Execution	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO					
		L101	99.31	99.27	97.97	99.33	99.03	103.08	95.15	98.47	107.43	94.70	107.11	110.03	99.21	104.67	114.75	109.24	105.51	105.48	98.47	98.47	107.43	94.70	105.11	110.03	99.31	99.27	97.97	99.33	99.03	103.08	95.15	98.47	107.43	94.70
L102	99.41	99.46	101.65	99.39	100.12	104.70	99.88	98.99	102.56	109.28	101.61	102.25	99.41	100.20	104.70	109.24	105.12	98.89	98.99	102.56	109.28	101.61	102.25	99.41	100.20	104.70	109.24	105.12	98.89	98.99	102.56	109.28	101.61	102.25		
L103	99.22	99.28	101.84	99.13	99.19	101.84	102.78	103.79	102.05	101.93	102.78	103.05	99.33	99.40	102.78	103.05	103.07	102.78	103.07	103.07	102.78	103.07	103.07	102.78	103.07	103.07	102.78	103.07	103.07	102.78	103.07	103.07	102.78	103.07	103.07	102.78
L104	99.27	99.24	101.86	99.16	99.20	101.88	102.53</																													

		Metric: Median Absolute Percentage Error (MdAPE)																																			
TEST SET	Traffic	ℓ_1 LSA _{AIC}				ℓ_2 LSA _{AIC}				ℓ_1 MTE _{AIC}				ℓ_2 MTE _{AIC}				ℓ_1 LSA _{BIC}				ℓ_2 LSA _{BIC}				ℓ_1 MTE _{BIC}				ℓ_2 MTE _{BIC}							
		MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO						
L101	103.40	103.22	100.04	103.30	102.95	100.02	98.30	96.62	100.03	97.71	98.05	100.02	104.01	103.85	99.99	105.91	105.48	100.00	98.30	96.62	100.03	97.71	98.95	100.02	102.31	102.27	102.31	100.86	101.31	100.01							
L102	102.31	102.27	100.02	102.29	102.17	100.02	100.52	99.55	100.01	100.86	101.31	100.01	102.31	102.27	100.02	104.95	104.93	100.02	100.52	99.55	100.01	101.86	101.81	101.86	100.86	101.31	100.01	100.30	100.30	100.30	100.00	101.64	100.99	100.00			
L103	100.30	100.34	99.99	100.07	100.06	99.99	101.89	102.10	100.00	101.66	100.99	100.00	101.30	101.20	100.03	101.03	101.00	100.03	101.89	102.10	100.00	101.64	100.99	100.00	101.27	101.27	101.27	101.01	101.31	100.98	100.98						
L104	99.41	99.46	99.25	100.21	99.50	99.25	100.21	99.50	100.20	100.21	99.50	100.21	99.50	100.21	99.50	100.21	99.50	100.21	100.21	99.50	100.21	99.50	100.21	99.50	100.21	99.50	100.21	99.50	100.21	99.50	100.21	99.50	100.21				
L105	103.21	103.28	99.96	102.98	102.85	100.01	101.43	99.42	100.00	100.09	100.12	99.99	104.85	104.74	99.99	103.99	103.95	99.99	101.43	99.42	100.00	100.09	101.12	99.99	100.09	101.12	101.12	101.12	99.99	100.09	101.12	99.99	100.09	101.12	99.99		
L106	102.39	102.27	99.99	102.55	102.40	99.99	101.13	101.15	100.00	100.87	100.09	99.99	107.70	104.45	99.98	122.09	100.02	101.51	101.15	100.00	100.87	100.02	101.51	101.15	100.00	101.51	101.15	100.00	101.51	101.15	100.00	101.51	101.15	100.00	101.51	101.15	100.00
L107	101.66	101.16	100.00	101.93	101.43	100.02	100.61	99.37	99.99	99.86	98.74	100.77	107.14	104.11	99.98	121.10	120.32	99.98	100.66	99.37	99.99	99.86	98.74	99.98	100.66	99.37	99.99	99.86	98.74	99.98	100.66	99.37	99.99	99.86	98.74	99.98	
L108	101.66	101.11	102.28	103.34	102.72	102.24	104.78	104.40	101.72	105.18	104.49	102.28	106.37	105.76	102.58	126.68	129.68	102.16	104.78	104.40	101.72	105.18	104.49	102.28	106.37	105.76	102.58	126.68	129.68	102.16	104.78	104.40	101.72	105.18	104.49	102.28	

Table 27: Comparison of all the methods (ℓ_1 MTE, ℓ_2 MTE, ℓ_1 LSA, ℓ_2 LSA under AIC and BIC minimization) under Median Absolute Percentage Error (MdAPE) metric , evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model.

		Metric: Symmetric Mean Absolute Percentage Error (SMAPE)																																					
TEST SET	Traffic	ℓ_1 LSA _{AIC}				ℓ_2 LSA _{AIC}				ℓ_1 MTE _{AIC}				ℓ_2 MTE _{AIC}				ℓ_1 LSA _{BIC}				ℓ_2 LSA _{BIC}				ℓ_1 MTE _{BIC}				ℓ_2 MTE _{BIC}									
		MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO								
L101	23.44	22.28	17.68	24.47	23.18	18.04	23.05	21.86	17.89	23.20	22.16	18.75	25.53	24.25	18.04	28.43	27.04	18.04	23.05	21.86	17.89	23.20	22.16	18.75	18.16	18.07	17.68	23.30	22.16	18.75	18.16	18.07	17.68						
L102	18.16	17.61	12.47	18.22	17.62	12.47	16.65	16.03	12.79	16.40	15.91	12.47	17.45	17.14	12.47	21.74	21.11	12.48	16.65	16.16	12.79	16.40	15.91	13.18	17.16	17.07	12.47	17.16	17.07	12.47	17.16	17.07	12.47	17.16	17.07	12.47			
L103	17.89	17.28	12.94	17.79	17.16	12.94	17.15	16.58	13.21	19.73	19.07	12.94	19.44	19.00	12.94	19.40	18.97	12.94	17.15	16.58	13.21	17.13	16.52	13.03	17.15	17.06	12.94	17.15	17.06	12.94	17.15	17.06	12.94	17.15	17.06	12.94			
L104	20.20	19.86	9.75	20.32	19.51	9.75	19.85	19.68	9.75	20.45	20.25	9.75	20.21	19.73	9.75	20.45	20.25	9.75	20.21	19.73	9.75	20.45	20.25	9.75	19.85	19.68	9.75	20.21	19.73	9.75	19.85	19.68	9.75	20.21	19.73	9.75	19.85		
L105	20.88	20.34	16.34	22.19	21.53	15.93	21.68	21.27	17.39	21.65	21.14	15.93	21.42	21.30	15.93	21.68	21.27	15.93	21.42	21.30	15.93	21.68	21.27	15.93	21.42	21.30	15.93	21.42	21.30	15.93	21.42	21.30	15.93	21.42	21.30	15.93			
L106	22.69	22.09	14.74	22.76	22.11	14.74	22.65	22.04	14.74	22.76	22.11	14.74	22.65	22.04	14.74	22.76	22.11	14.74	22.65	22.04	14.74	22.76	22.11	14.74	22.65	22.04	14.74	22.76	22.11	14.74	22.65	22.04	14.74	22.76	22.11	14.74			
L107	22.69	21.82	20.24	22.76	21.29	20.24	20.43	19.87	21.15	20.07	18.50	21.90	20.74	18.50	22.48	22.26	18.50	20.64	20.35	18.50	22.48	22.26	18.50	20.64	20.35	18.50	22.48	22.26	18.50	20.64	20.35	18.50	22.48	22.26	18.50	20.64	20.35	18.50	
L108	19.91	19.64	22.33	22.18	22.33	20.72	19.42	18.86	19.96	19.48	18.87	19.96	20.37	18.87	19.48	25.37	25.77	18.02	18.35	15.37	18.02	23.34	22.30	8.37	12.60	12.27	8.35	12.70	12.27	8.35	12.70	12.27	8.35	12.70	12.27	8.35	12.70	12.27	8.35

Table 28: Comparison of all the methods (ℓ_1 MTE, ℓ_2 MTE, ℓ_1 LSA, ℓ_2 LSA under AIC and BIC minimization) under Symmetric Mean Absolute Percentage Error (SMAPE) metric , evaluated using model averaging from MEBoot samples (MEB), the corresponding models applied on the original traffic data (Actual), and post-hoc variable selection algorithm ADL1LASSO models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model.

5.2 More Details for data

For reason of completeness, we enlist here the exact locations of the loop detectors, given the image and that they have been installed on traffic lights:

- * **L101 (location A):** Intersection of Mavromataion Street and Alexandras Avenue (Central Bus Station of Attiki). Direction towards Patision Avenue.
- * **L102 (location B):** Intersection of Ioannis Varvakis Street, Alexandras Avenue, and Charilaou Trikoupi Street. Direction towards Patision Avenue.
- * **L103 (location C):** Intersection of Panagiotaras Street and Alexandras Avenue. Direction towards Patision Avenue.
- * **L104 (location D):** Intersection of Panormou and Alexandras Avenue (Ampelokipoi). Direction towards Patision Avenue.
- * **L106 (location E):** Intersection of Ioannis Varvakis Street Alexandras Avenue, and Charilaou Trikoupi Street (Republic of Argentina Square, Panathnea bus station). Direction towards Panormou Avenue.
- * **L107 (location F):** Intersection of Nikolaos Gyzi Street and Alexandras Avenue (Lampetis' Theater, Hypokratous bus station). Direction towards Panormou Avenue.
- * **L108 (location G):** Intersection of Panormou and Alexandras Avenue (Ampelokipoi bus station). Direction towards Panormou Avenue.

In the implemented R-shiny software, we display each location, regarding the user input.

Short Proof of Wiener-Kinchine Theorem

Consider the definition of the autocorrelation function $C(t)$ of a function $E(t)$:

$$C(t) = \int_{-\infty}^{\infty} \overline{E(\tau)} E(t + \tau) d\tau$$

Also, remember that the Fourier transform of $E(t)$ is defined as

$$E(\tau) = \int_{-\infty}^{\infty} E_{\nu} e^{-2\pi i \nu \tau} d\nu,$$

Providing a complex conjugate of

$$\overline{E(\tau)} = \int_{-\infty}^{\infty} \overline{E}_{\nu} e^{2\pi i \nu \tau} d\nu.$$

Plugging $\bar{E}(\tau)$ and $E(t + \tau)$ into the autocorrelation function therefore yields

$$\begin{aligned}
C(t) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \bar{E}_{\nu} e^{2\pi i \nu \tau} d\nu \right] \left[\int_{-\infty}^{\infty} E_{\nu'} e^{-2\pi i \nu' (t+\tau)} d\nu' \right] d\tau \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{E}_{\nu} E_{\nu'} e^{-2\pi i \tau (\nu' - \nu)} e^{-2\pi i \nu' t} d\tau d\nu d\nu' \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{E}_{\nu} E_{\nu'} \delta(\nu' - \nu) e^{-2\pi i \nu' t} d\nu d\nu' \\
&= \int_{-\infty}^{\infty} \bar{E}_{\nu} E_{\nu} e^{-2\pi i \nu t} d\nu \\
&= \int_{-\infty}^{\infty} |E_{\nu}|^2 e^{-2\pi i \nu t} d\nu \\
&= \mathcal{F}_{\nu} [|E_{\nu}|^2](t),
\end{aligned}$$

Interestingly, the autocorrelation is simply supplied by the Fourier transformation of the absolute square value of the E_{ν} . The Wiener-Khinchine theorem is a special case of cross-correlation theorem, that is proved below, with $f = g$.

Proof: Let the conditional loss function be:

$$\begin{aligned}
L &= \mathbb{E}[|y - a| \mid X = x] \\
&= \int_{-\infty}^{\infty} |y - a| f_{Y|X}(y) dy \\
&= \int_{-\infty}^a (a - y) f_{Y|X}(y) dy + \int_a^{\infty} (y - a) f_{Y|X}(y) dy.
\end{aligned}$$

Differentiating with respect to a :

$$\frac{\partial L}{\partial a} = \int_{-\infty}^a f_{Y|X}(y) dy - \int_a^{\infty} f_{Y|X}(y) dy = 0.$$

This implies:

$$F_{Y|X}(a) = 0.5,$$

meaning a is the conditional median of y given x . □

Short Proof of Cross-Correlation Theorem

Let $f \star g$ represent the cross-correlation between the functions $f(t)$ and $g(t)$. Then

$$\begin{aligned}
f \star g &= \int_{-\infty}^{\infty} \bar{f}(\tau)g(t+\tau)d\tau \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \bar{F}(\nu)e^{2\pi i\nu\tau}d\nu \int_{-\infty}^{\infty} G(\nu')e^{-2\pi i\nu'(t+\tau)}d\nu' \right] d\tau \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{F}(\nu)G(\nu')e^{-2\pi i\tau(\nu'-\nu)}e^{-2\pi i\nu't}d\tau d\nu d\nu' \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{F}(\nu)G(\nu')e^{-2\pi i\nu't} \left[\int_{-\infty}^{\infty} e^{-2\pi i\tau(\nu'-\nu)}d\tau \right] d\nu d\nu' \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{F}(\nu)G(\nu')e^{-2\pi i\nu't}\delta(\nu'-\nu)d\nu' d\nu \\
&= \int_{-\infty}^{\infty} \bar{F}(\nu)G(\nu)e^{-2\pi i\nu t}d\nu \\
&= \mathcal{F}[\bar{F}(\nu)G(\nu)]
\end{aligned}$$

where \mathcal{F} denotes the Fourier transform, \bar{z} is the complex conjugate, and

$$f(t) = \mathcal{F}_{\nu}[F(\nu)](t) = \int_{-\infty}^{\infty} F(\nu)e^{-2\pi i\nu t}d\nu g(t) = \mathcal{F}_{\nu}[G(\nu)](t) = \int_{-\infty}^{\infty} G(\nu)e^{-2\pi i\nu t}d\nu$$

Applying a Fourier transform on either side yields the cross-correlation theorem, $f \star g = \mathcal{F}[\bar{F}(\nu)G(\nu)]$.

Proof of MAE's Optimality

Let the conditional loss function be:

$$\begin{aligned}
L &= \mathbb{E}[|y - a| \mid X = x] \\
&= \int_{-\infty}^{\infty} |y - a| f_{Y|X}(y) dy \\
&= \int_{-\infty}^a (a - y) f_{Y|X}(y) dy + \int_a^{\infty} (y - a) f_{Y|X}(y) dy.
\end{aligned}$$

Differentiating with respect to a :

$$\frac{\partial L}{\partial a} = \int_{-\infty}^a f_{Y|X}(y) dy - \int_a^{\infty} f_{Y|X}(y) dy = 0.$$

This implies:

$$F_{Y|X}(a) = 0.5,$$

meaning a is the conditional median of y given x . \square

References

- [1] H. Wang and C. Leng, “[Unified LASSO Estimation by Least Squares Approximation](#),” *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 1039–1048, 2007.
- [2] Y. Qin, S. Li, Y. Li, and Y. Yu, “[Penalized Maximum Tangent Likelihood Estimation and Robust Variable Selection](#),” 2017.
- [3] H. D. Vinod and J. L. de Lacalle, “[Maximum Entropy Bootstrap for Time Series: The meboot R Package](#),” *Journal of Statistical Software*, 2017.
- [4] F. Rodrigues, “[On the importance of stationarity, strong baselines and benchmarks in transport prediction problems](#),” 2023.
- [5] Y. Kamarianakis, “[Space-time modeling of traffic variables with adaptive LASSO](#),” *JSM Proceedings, American Statistical Association, Section on Statistical Learning and Data Mining*, pp. 2732–2738, 2014.
- [6] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, “[Short-term traffic forecasting: Where we are and where we are going.](#),” *Transportation Research Part C: Emerging Technologies*, 43, vol. 43, no. 1, pp. 3–19, 2014.
- [7] E. I. Vlahogianni and M. G. Karlaftis, “[Temporal aggregation in traffic data: Implications for statistical characteristics and model choice.](#),” *Transportation Letters*, 3, vol. 3, no. 1, pp. 37–49, 2011.
- [8] M. Giacomazzo and Y. Kamarianakis, “[Bayesian estimation of subset threshold autoregressions: short-term forecasting of traffic occupancy.](#),” *Journal of Applied Statistics: Advances in Computational Data Analysis*, vol. 47, no. 13-15, pp. 2658–2689, 2020.
- [9] H. E. Doran and J. Quilkey, “[Harmonic Analyss of Seasonal Data: Some Important Properties](#),” *American Journal of Agricultural Economics*, vol. 54, no. 4, pp. 646–651, 1972.
- [10] L. Sun, “[Spectral and time-frequency analyses of freeway traffic flow](#),” *Journal of Advanced Transportation*, vol. 48, pp. 821–857, 2014.
- [11] K. Knight and W. Fu, “[ASYMPTOTICS FOR LASSO-TYPE ESTIMATORS](#),” *The Annals of Statistics*, vol. 28, no. 5, pp. 1356–1378, 2000.
- [12] H. Zou, “[The Adaptive Lasso and Its Oracle Properties](#),” *Journal of the American Statistical Association*, vol. 101, pp. 1418–1429, 2006.
- [13] P. Stoice and B. Ottersten, “[The Evil of Superefficiency](#),” *Signal Processing*, vol. 55, p. 133–136, 1996.
- [14] J. F and R. Li, “[Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties](#),” *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [15] L. Breiman, “[Better Subset Regression Using the Nonnegative Garrote](#),” *Technometrics*, vol. 37, no. 4, 1995.

- [16] D. Pollard, “[Asymptotics for Least Absolute Deviation Regression Es timators](#),” *Econometric Theory*, vol. 7, pp. 186–199, 1991.
- [17] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “[Least Angle Regression](#),” *The Annals of Statistics.*, vol. 32, no. 2, p. 407–451, 2004.
- [18] H. Zou, T. Hastie, and R. Tinshirani, “[ON THE “DEGREES OF FREEDOM” OF THE LASSO](#),” *The Annals of Statistics*, vol. 35, no. 5, pp. 2173–2192, 2007.
- [19] H. White, “[A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity.](#),” *Econometrica*, vol. 48, no. 4, pp. 817–830, 1994.
- [20] J. G. MacKinnon and H. White, “[Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties](#),” *Journal of Econometrics*, vol. 29, no. 3, pp. 305–325, 1985.
- [21] F. Cribari-Neto, “[Asymptotic Inference Under Heteroskedasticity of Unknown Form.](#),” *Computational Statistics & Data Analysis*, vol. 45, no. 2, pp. 215–233, 2004.
- [22] F. Cribari-Neto and W. B. D. Silva, “[A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model.](#),” *AStA Advances in Statistical Analysis*, vol. 95, p. 129–146, 2011.
- [23] W. Newey and K. D. West, “[Automatic lag selection in covariance matrix estimation](#),” *Review of Economic Studies.*, vol. 61, no. 4, pp. 631–653, 1994.
- [24] A. S. Hadi and A. Luceño, “[Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms.](#),” *Computational Statistics & Data Analysis*, vol. 25, no. 3, pp. 251–272, 1997.
- [25] A. C. Lozano and N. Meinshausen, “[Minimum Distance Lasso for robust high-dimensional regression](#),” *Electronic Journal of Statistics*, vol. 10, no. 1, p. 1296–1340, 2016.
- [26] G. Raskutti, M. J. Wainwright, and B. Yu, “[Restricted Eigenvalue Properties for Correlated Gaussian Designs](#),” *Journal of Machine Learning Research*, vol. 11, pp. 2241–2259, 2010.
- [27] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, “[A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers](#),” *Statistical Science*, vol. 27, no. 4, p. 538–557, 2012.
- [28] H. Wang, G. Li, and G. Jiang, “[Robust Regression Shrinkage and Consistent Variable Selection through the LAD-Lasso](#),” *Journal of Business & Economic Statistics*, vol. 25, p. 347–355, 2007.
- [29] L. Wang, “[The L1 penalized LAD estimator for high dimensional linear regression](#),” *Journal of Multivariate Analysis*, vol. 120, pp. 135–151, 2013.
- [30] X. Wang, Y. Jiang, M. Huang, and H. Zhang, “[Robust variable selection with exponential squared loss.](#),” *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 632–643, 2013.
- [31] D. Anderson and K. Burham, *[AIC MYTHS AND MISUNDERSTANDINGS](#)*, 2006.

- [32] H. D. Vinod, “Ranking mutual funds using unconventional utility theory and stochastic dominance,” *Journal of Empirical Finance*, vol. 11, no. 3, pp. 353–377, 2004.
- [33] H. D. Vinod, “Maximum Entropy Ensemble for Time Series Inference in Economics,” *Journal of Asian Economics*, 2006.
- [34] N. Wiener, “Generalized Harmonic Analysis,” *Acta Math*, vol. 55, pp. 117–258, 1930.
- [35] A. Kolmogorov, “Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung,” *Mathematische Annalen*, vol. 104, pp. 415–458, 1931.
- [36] A. Khinchine, “Korrelationstheorie der stationären stochastischen Prozesse.,” *Mathematische Annalen*, vol. 109, p. 604–615, 1934.
- [37] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.
- [38] P. J. Huber, “Robust Estimation of a Location Parameter,” *The Annals of Mathematical Statistics*, 1964.
- [39] R. Flides, “The evaluation of extrapolative forecasting methods,” *International Journal of Forecasting*, vol. 8, pp. 81–98, 1992.
- [40] S. J. Armstrong, *Long-Range Forecasting: From Crystal Ball to Computer*. A WILEY-INTERSCIENCE PUBLICATION, 1985.
- [41] S. J. Armstrong and F. Collopy, “Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons.,” *International Journal of Forecasting*, vol. 8, pp. 69–80, 1992.
- [42] S. G. Makridakis, “Assessing Forecast Accuracy Measures,” *International Journal of Forecasting* 9, vol. 4, pp. 527–529, 2004.
- [43] P. Goodwin and R. Lawton, “On the Asymmetry of the Symmetric MAPE.,” *International Journal of Forecasting*, 1999.
- [44] Z. Chen and Y. Yang, “Assessing Forecast Accuracy Measures,” 2004.
- [45] R. Hyndman and A. B. Koehler, “Another Look at Measures of Forecast Accuracy.,” *International Journal of Forecasting*, vol. 22, p. 679–88, 2004.
- [46] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “M5 accuracy competition: Results, findings, and conclusions,” *International Journal of Forecasting*, vol. 38, pp. 1346–1364, 2022.
- [47] N. Kouretzes and F. Petropoulos, “Forecast combinations for intermittent demand,” *Journal of the Operational Research Society*, vol. 66, 2015.
- [48] D. W. Andrews, “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, vol. 59, no. 3, pp. 817–858, 1991.
- [49] A. G. Galvao and J. Yoon, “HAC Covariance Matrix Estimation in Quantile Regression,” *Journal of the American Statistical Association*, vol. 119, pp. 2305–2316, 2021.