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. 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. For recognizing the demonstrated importance of low harmonic frequencies in traffic pattern analysis, we quantify their relative contributions through the probabilities of each feature to be included in each model that we apply on the synthetic datasets (known as **Variable Inclusion Probabilities (VIP)** in the regression literature). The synthetic dataset have been generated using the **Maximum Entropy Bootstrap (MEB)** paradigm [3]. 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.
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 This approach builds upon the foundational work of F.Rodrigues [4]. In his study,we can see that extracting harmonic seasonal components (representing historical averages) from non-stationary time series yields stationary in mean residuals, which means that they have zero mean but the variance is not constant. Therefore, we can achieve forecasting accuracy comparable to computationally intensive deep learning methods by combining classical statistical models (e.g. ARMA models), with heteroskedasticity correction methods (e.g. GARCH).

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 volume and occupancy measurements taken every 90 seconds in April 2000 from seven loop detectors in different locations in the Alexandras Avenue. However, in our study, we are going to focus only on the traffic volume samples. These detectors are located on Alexandras Avenue, a key arterial avenue in Athens, Greece. In the figure below 1 we indicate the locations of the loop detectors as $L \in \{A, B, C, D, E, F, G\}$ (see appendix section for more info) and we can notice 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 9th of April in 2000 where the central-left party (named "PASOK") won, the president of that party had various election campaigns at party's office (which is located on Charilaou Trikoupi, a route that crosses Alexandras Avenue on F detector). An other interesting event is occurred on 20th of April, where a Greek basketball club (Panathinaikos) 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 volume 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 smooths the data and mitigates nonlinear and heteroskedastic effects ([7]). Traditionally, traffic volume 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]*

The figure 2 below depicts the traffic volumes of the aforementioned locations. In this illustration we can easily notice that we have noisy data where low values of detected vehicles are measured at the end of each workday and the traffic volume lies in the range 0 to 160. A notable anomaly is present on April 10th for every location, occurring after the aforementioned elections, which introduces a considerable challenge for effective model training.

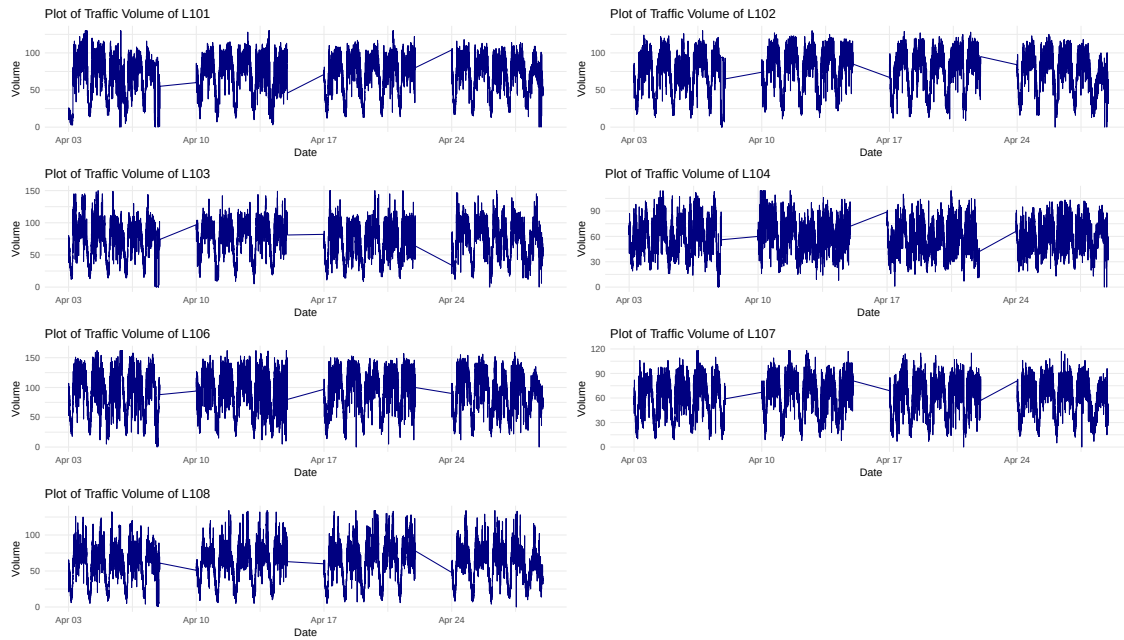

Figure 2: *Plot of traffic volume on 7 locations of Alexandras Avenue. The top left plot corresponds to location L101 (A), the top right to location L102 (B), the second row of plots corresponds to the locations L103 (C) and L104 (D), the third row to the locations L106 (E) and L107 (F) (since we do not have location L105) and the last plot depicts the traffic volume of the location L108 (G).*

3 Methods

3.1 Harmonic Seasonal Model

By plotting the data we can clearly see that the data of traffic 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 $\sin(2\pi f t)$ has a period $T = 1/f$, as it satisfies the condition:

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

Here, f denotes the frequency (cycles 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 $k = 1, 2, \ldots$:

$$x(t) = A_0 + \sum_{k=1}^{\infty} \{A_k \cos(2\pi f_k t) + B_k \sin(2\pi f_k t)\}, \text{ where } A_0, A_k, B_k \in \mathbb{R}, \forall k > 0. \tag{1}$$

In real-world applications (in our case too), this infinite sum is typically approximated by a finite number of terms. For this purpose, we can leverage the sinusoidal terms and take advantage of their non-linear structure to estimate the coefficients $A_0, \{A_k\}_{k>0}$ and $\{B_k\}_{k>0}$ and capture the periodicity of the data. For the coefficients estimation, we use two general families of methods: the minimization of the Manhattan (also known as ℓ_1 norm) and Euclidean distance (ℓ_2 norm) of the residuals, namely the difference between our prediction and the target variable. The second norm that we mentioned is the conventional way to estimate the coefficient, because it is smooth (there isn't any discontinuity) and we can find closed form solutions for the coefficients. The last few decades it is a common practice to use the ℓ_1 norm minimization, beside the lack of smoothness, because its more robust to extreme and rare values.
According to A.Metcalfe and P.Cowpertwait (see [9], section 5.6, pages 112-119) we can capture the seasonality pattern $\{x_t\}$ with T seasons, by defining the summation series of 1 with upper bound of the summation to be $\lfloor T/2 \rfloor$, and the harmonic seasonal model is defined as:

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

where m_t represents the trend component, and s_i and c_i are unknown parameters. The upper bound represent the possible cycles that we can use in order to consist a basis using the aforementioned trigonometric functions. This is the reason for constraining the summation in the set $\mathcal{C} = \{1, \ldots, \lfloor T/2 \rfloor\}$ harmonic frequencies. Furthermore, we can notice that the periodicity of sine and cosine functions allow us to leverage the harmonics in the set $\{i\}_{i=1}^{T}$, where T is the period, but $\forall x \in \left\{\lfloor \frac{T}{2} \rfloor + 1, \ldots, T\right\} \exists y \in \mathcal{C} : sin\left(2\pi x \frac{t}{T}\right) = -sin\left(2\pi y \frac{t}{T}\right)$, namely the $sin\left(2\pi x \frac{t}{T}\right)$ can be seen as linear combination of the $sin\left(2\pi y \frac{t}{T}\right)$. Therefore, taking into account the orthogonality of the trigonometric functions we can safely conclude that the set of $\left\{sin\left(2\pi x \frac{t}{T}\right), cos\left(2\pi x \frac{t}{T}\right)\right)\}_{x \in \mathcal{C}}$ consists of a basis for any time series dataset.
The trend can be modeled using a polynomial form. When T is an even number, the sine term at frequency $1/2$ (i.e., when $i = T/2$) 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 $1/T$ 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 T. The afore described model is also known as **Harmonic Seasonal Model** (HMS) since it uses the harmonics to capture the seasonal pattern.
It has been empirically shown through the literature that leveraging the low harmonic frequencies can help the accuracy of the model for short-term forecasting of traffic data. Therefore, we will focus our analysis on the first 500 harmonic frequencies (out of 2400 harmonic frequencies) reducing the complexity of the resulted model. This type of naive models can be used as a set of benchmark models against the more sophisticated models, since they are less computationally expensive and they are easier to implement. As a future step, we could implement more complex models that can better capture the seasonality pattern and we can check if these techniques can significantly outperform the benchmark models.

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 ([10]) proved that if $\lambda_n/n \to \lambda_0 \geq 0$, 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 ([11]) stated that the convergence rate of $\hat{\beta}^{(n)}$ is slower than $\sqrt{n}$, and the limiting quantity is non random. 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 an auxiliary LASSO regression (using 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 [11], where the two-step procedure is named **Adaptive LASSO** [11]. 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

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

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 $\sqrt{n}(\hat{\beta}_A^{(n)} - \beta_A^*) \to_d \mathcal{N}(0, \sigma^2 \times \hat{\Sigma}_{11}^{-1})$ (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][13]. Last but not least, we can conclude that the adaptive lasso solution is continuous by definition which assures that the procedure is optimal. [2]

In our investigation, the application of robust variable selection with the two-step 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 squares problems.One notable advantage of this methodology is that when a Taylor approximation of adaptive LASSO penalty and a BIC-type tuning parameter selector are employed, the resulting LSA estimator can be as efficient as the oracle.

Through the study of Wang and Leng [1], it is 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 $\tilde{\beta} = \underset{\beta \in \mathbb{R}^d}{\arg\min} \left(\mathcal{L}_n(\beta)\right)$ and the nullification of the first derivative of the aforementioned loss function. Taking into account that $\mathbb{E}\left[\frac{\ddot{\mathcal{L}}_n(\beta)}{n}\right] \approx \Sigma^{-1}$ where Σ^{-1} is the inverse of the

[1] The super-efficiency phenomenon ensures the boosting scheme of the efficiency on a two (or more) step procedure while the asymtotic 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 super-efficiency implies that asymptotically there exists no uniformly minimum-variance parameter estimator. [12]

[2] We can also find on the study where H.Zou proposed this model, that the Breiman's Nonnegative garrote [14] 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. Let $\{\hat{\beta}_k\}$ represent the original ordinary least squares (OLS) estimates. Define $\{c_k\}$ as the set of values that minimize the objective function $\sum_k \left(y_n - \sum_k c_k \hat{\beta}_k x_{kn}\right)^2$, 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.

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:

$$\begin{aligned}\arg\min_{\beta\in\mathbb{R}^d}\left(\frac{\mathcal{L}_n(\beta)}{n}\right) &= \arg\min_{\beta\in\mathbb{R}^d}\left(\frac{\mathcal{L}_n(\tilde{\beta})}{n}+(\beta-\tilde{\beta})^\top\frac{\ddot{\mathcal{L}}_n(\tilde{\beta})}{n}(\beta-\tilde{\beta})\right), \qquad \text{if } \mathcal{L}_n \text{ is the Loss function of an unpenalized estimator}\\ &\approx \arg\min_{\beta\in\mathbb{R}^d}\left((\beta-\tilde{\beta})^\top\hat{\Sigma}^{-1}(\beta-\tilde{\beta})\right)\end{aligned}$$

In the case of conventional unpenalized estimator of the ordinary least squares we use the approximation $\hat{\Sigma}^{-1} \approx cov(x^S)\sigma^{-2}$, where the x^S is the design matrix of data using the features that $S = \{j_1, \ldots, j_d\}$, indices set defines and σ^2 tjhe This is extremely intriguing in case of LAD minimization case, 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 Pollard's asymptotic covariance matrix [15]:

$$\hat{\Sigma}^{-1} \approx 4f_e^2(0)cov(x^S),$$

where

- → $S = \{j_1, \ldots, j_d\}$, denote an arbitrary candidate model,
- → $\{j_i\}_{i=1}^d$, (where d>1) is the set of predictors of the candidate model,
- → $f_e(0)$, is the probability density function of the noise at zero value.

Therefore, taking advantage of the LSA implementation[1] we can perform conventional LASSO, robust (ℓ_1) LASSO and their adaptive forms by applying the LARS [16] and LASSO approaches of the code. In our study, we are focusing on the latter part which competes the two-step adaptive lasso implementation.

Wang and Leng [1] proved in their study upon the various LSA scenarios, 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. Furthermore, 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 capture 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 $S_T = \mathbb{N}_{\leq d_0}$ (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 [17] and minimizes the BIC criterion, using the rule of thumb given by Zou [11] 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 $\mathbb{R}^d$. [3]

[3]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[18], J.G.MacKinnon[19], F.Cribaro- Neto [20], and W.B.D. Da Silva [21], keeping in mind that the last two estimators are useful in the small-sampled datasets, and Newey-West (NW-HAC, [22]) in case of autocorrelation persistence.

The model above has been implemented in R software by the authors of the corresponding publication. The corresponding R code scripts can be found in the github repository of the RobMixReg library.
The `lars.lsa` function efficiently solves LARS or LASSO linear regression problems using a path-following algorithm, particularly suited for precomputed covariance matrices. We are allowed to give a covariance matrix `Sigma_0`, coefficient vector `b_0`, intercept flag `intercept`, sample size `n`, algorithm type to be lasso or lars, precision tolerance `epsilon`, and maximum steps `max.steps`. The function iteratively builds the solution path by calculating correlations, adding variables to an active set, and updating the Gram matrix's Cholesky factor. For LASSO, it ensures sparsity by setting coefficients to zero when appropriate. Finally, it reconstructs coefficient pathways, adjusts for original scaling, and computes model selection criteria (for instance RSS, AIC, and BIC), providing a comprehensive output for analysis.

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 10 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 $\tilde{\beta} = \underset{\beta \in \mathbb{R}^d}{\arg\max} \left(\sum_{i=1}^{n} ln_t(f(y_i - x_i^T\beta)) - n \sum_{j=1}^{d} p_{\lambda_j}(|\beta_j|) \right)$ where $ln_t(f)$ is the p-th order Taylor expansion of, $ln(u) ln_t(u) = \mathbb{1}_{[u>t]} ln(u) + \mathbb{1}_{[0 \le u \le t]} ln(t) + \sum_{k=1}^{p} \frac{\partial^k ln(v)}{\partial v^k}\Big|_{v=t} \frac{(u-t)^k}{k!}$ with $t > 0$ is the tuning parameter [4]. 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 $w_i = \left(1 - \left(1 - \frac{f(y_i - x_i^T\beta)}{t}\right)^p\right)^{\delta_{[i \in \mathcal{A}^c]}}$ [5]. Then, the simple MTE problem can be induced to the following minimization problem

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

[4] Note that $\lim_{t \to 0^+} ln_t(u) = ln(u)$ and for $t = 0$ we get the MLE

[5] It's easily notable that $w_i \overset{t \to 0}{\to} 0$

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 ([23],[24]). 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 Likelihood[6] with respect to the true parameters β_0

$$\tilde{\beta} \underset{n\to\infty}{\overset{p}{\to}} \beta_t^\star, \text{ with } \beta_t^\star = \arg\max_{\beta\in\mathcal{B}} \mathbb{E}_{\beta_0}(ln_t(f(y - X^\top\beta_t^\star))$$

- ☞ 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 $t \to 0^+$ and then we get $\beta_t^\star \to \beta_0$. Namely,

$$\sqrt{n}\Omega^{-1/2}(\tilde{\beta} - \beta_t^\star) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0, I_{d\times d})$$

- ☞ 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 out 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, $a_n = \max_{\beta_{0j}\neq 0}(p'_{\lambda_{n_j}}(|\beta_{0j}|))$ and $b_n = \max_{\beta_{0j}\neq 0}(p''_{\lambda_{n_j}}(|\beta_{0j}|))$

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

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 $\liminf_{n\to\infty} \liminf_{\theta\to 0+} \left\{ \min_{s+1\le j\le d} p'_{\lambda_{nj}}(\theta)/\lambda_{nj} \right\} > 0$, and

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

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

$$\begin{cases} \Sigma_1 = diag\left(\left(p''_{\lambda_{n1}}(|\beta_{0i}|)\right)_{i\in\mathbb{N}_{\le s}}\right) \\ \Sigma_2 = Cov[\partial \ln_t(f(z;\beta_0))/\partial\beta_S] \\ J_S = \mathbb{E}[\partial^2 \ln_t(f(z;\beta_0))/\partial\beta_S\partial\beta_S^\top] \\ \mathbf{b} = \left(p'_{\lambda_{ni}}(|\beta_{0i}|)\mathrm{sgn}(\beta_{0i})\right)_{i=1}^{s}{}^{\top} \end{cases}$$

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

$$\mathbb{V}ar(\hat{\beta}_S) = \frac{1}{n}(J_S + \Sigma_1)^{-1}\Sigma_2(J_S + \Sigma_1)^{-1}$$

Let's define the distance of the candidate coefficients, $\hat{\beta}$ from the true values, β_0, as $\mathbf{\Delta} = \hat{\beta} - \beta_0$ and the set $\mathbf{\Delta}_S$ and $\mathbf{\Delta}_{S^c}$ are the projections of $\mathbf{\Delta}$ onto the coordinate sets S and S^c respectively. Therefore, the consistency of the penalized MTE scheme is achieved based on the restricted eigenvalue condition (REC)[7] over the constrain set $\mathbb{C}(S) = \{\mathbf{\Delta} \in \mathbb{R}^d : 3\|\mathbf{\Delta}_S\|_1 \le \|\mathbf{\Delta}_{S^c}\|_1\}$. The loss function of LASSO (under LS or LAD minimization) $\mathcal{L}_\lambda$ satisfies strongly restricted convexity (SRC) in the neighborhood of the true coefficients and its gradient is bounded on the true coefficients, β_0, taking into account that our regressors are bounded[8] and the design matrix, X, satisfies the REC[9] [24]. The SRC assumption ensures that the loss function, $\mathcal{L}_\lambda$, contains enough curvature in the direction of $\mathbf{\Delta} \in \mathbb{C}(S)$ such that small value of $\mathcal{L}_\lambda(\tilde{\beta}) - \cancelto{0}{\mathcal{L}_\lambda(\beta_0)} = \mathcal{L}_\lambda(\tilde{\beta})$ translates to a small ℓ_2 norm $\|\mathbf{\Delta}\|_2^2$. For sake of computational simplicity, we end up tuning the hyperparameter greedily in the domain defined which was carefully selected. After a series of experiments, we realized that the values of λ parameter that prevent the overshrinkage is constrained in the set partition $\mathcal{D}_\lambda \subset (0, 0.1] : 5\mathcal{D}_\lambda = \mathbb{N}_{\le 5}$. 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

[7]The 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 [25], 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, $(1-\delta_s)\|y\|_2^2 \le \|A_s y\|_2^2 \le (1+\delta_s)\|y\|_2^2$. Then, the matrix A is said to satisfy the s-restricted isometry property with restricted isometry constant δ_s.

[8]The regressors consist of the set$\{sin(2\pi\cdot xft), cos(2\pi\cdot xft)\}_{x=1}^{500}$ where f is the frequency and t is the time variable

[9]The design matrix X satisfies REC means that $\frac{\|X\mathbf{\Delta}\|_2^2}{n} \le \kappa_{REC}\|\mathbf{\Delta}\|_2^2\ (\kappa_{REC} > 0), \forall \mathbf{\Delta} \in \mathbb{C}(S)$

regressors can cause huge values of the coefficients, since the target values can reach upto in 0-162 vehicles. 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.
The previously described method has been implemented in the `MTE` package in `MTElasso` function. This method takes the design matrix X and a time series dataset y, handles intercepts by adjusting the design matrix and penalty, and allows for varied penalization across variables. The approach employs tuning parameters `t` and `p` to control the threshold and order of its Taylor approximation[10]. Its core is an iterative coordinate descent scheme, where coefficients are updated by solving univariate optimization problems, incorporating a tangent likelihood component and an ℓ_1 penalty term. The optimization continues until convergence or a maximum of 200 iterations, actively pruning small coefficients to ensure sparsity and accelerate the process, ultimately returning the estimated coefficients and fitted values.

3.5 Maximum Entropy Bootstrap

For sake of completeness, in order to make a proper comparison of the aforementioned methods, 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 because of the 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, MEBoot is very well suited to our objective.
This methodology has been proposed by Vinod ([26],[27], and [3]) based on the Wiener-Kolmogorov-Khinchine (WKK) process[11]. The WKK theory traditionally creates a time series population, known as a set Ω, with a strong dependence on the assumption of stationarity. Expanding this idea, the MEBoot algorithm provides a modern, computer-intensive method for constructing this ensemble Ω, particularly useful for time series that are short, non-stationary, or exhibit features, for instance regime changes, gaps, and sudden jumps. Beside that WKK is based on zero memory I(0) white noise type processes with circular events, the MEBoot method, offers considerable flexibility for time series analysis, but it has 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 MEboot 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

[10]The `t` parameter allows the use to trimm the data in both sides, droping out $t \cdot 100\%$ of the data, while `p` parameter defines the Taylor's approximation order

[11]The 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.

data structure, especially in the presence of temporal dependence, the resulting inference can be significantly flawed. Moreover, the MEboot 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 MEboot 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 ($z_t = \frac{x_{(t)}+x_{(t+1)}}{2}, \quad t = 1, \ldots, T-1$). 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: $z_0 = x_{(1)} - m_{\text{trm}}, \quad z_T = x_{(T)} + m_{\text{trm}}$. 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.

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 and keeps the initial location of each data point in the original time series. This ranking is used after finding the bootstraped samples and we can re-order the data points, respecting the temporal dependency of the time series data. The next couple of steps assists on the limitation of the maximum entropy density estimation and this is crucial step to ensure that this procedure enjoys the mass preservation property [28]. This property ensures that, on average, each defined interval of the underlying distribution retains an equal fraction ($1/T = 1/2400$) of the total probability mass. By applying the maximum entropy

boootstrap we ensure that 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 (f.i. 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. This procedure ensures that
Respecting the requirements of the method and paying attention of its limitations, we used the Maximum Entropy Bootstrap via `meboot` 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 adaptive LASSO implementation and in the MTE case we need to tune the hyperparameter λ. 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 [12] as it is proved by Burham and Anderson [29]. This occurs because the formula of AIC and BIC is $2k - log(L)$ and $klog(n) - log(L)$ 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 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's design inherently values structural accuracy over parsimony, allowing complexity that meaningfully improves model fit. This makes it suitable for scenarios where capturing nuanced data patterns is prioritized, though it requires careful validation to avoid overfitting[30]. This frequently prompts AIC to prefer more flexible models, which might result in predictions that have more fluctuations, especially in the presence of noise in the data. 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

[12] We assume that the set of candidate models includes the data generating mechanism

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 and follow the recommendations of Burham and Anderson [31].[13]

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.

The **Root Mean Square Error (RMSE)** is a commonly used statistical metric that measures the square root of the mean of squared deviations, where the deviations are the discrepancies between values predicted by a model or estimator and actual values observed. 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. 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 its dependence on the magnitude of the data, RMSE should not be used to evaluate performance across datasets measured in various units. If $\hat{y}_t$ is an unbiased estimator of y_t, then the RMSE simplifies to the standard deviation of the estimator.
One more common statistical metric that practitioners often use is the **Mean Absolute Error (MAE)** which calculates the mean average of absolute values of errors. The formal definition is $\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t| = \frac{1}{T}\sum_{t=1}^{n}|\varepsilon_t|$. 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 of different datasets. MAE is extensively used in time series forecasting and is occasionally confused with Mean Absolute Deviation (MAD); while similar in form, the two are fundamentally distinct. Although MAE is sometimes compared or confused with the RMSE, however the two differ in interpretation and sensitivity. While MAE is a linear average of absolute errors, 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.

Statisticians often use a metric that combines the properties of both MAE and RMSE and limits the influence of outliers. This metric is called **Huber loss (HL)** and it is commonly used in robust regression. HL behaves quadratically near the origin and linearly for large residuals making it less sensitive to outliers than the pure squared loss, while still retaining sensitivity to small errors. Originally introduced by Huber ([32]), the loss is defined piecewise as:

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

[13]This 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.

where $\varepsilon = y - f(x)$ denotes the residuals. The function smoothly transitions between the quadratic and linear regimes at the threshold $|\varepsilon| = \delta$, preserving continuity, differentiability and a matching derivative at the join. Geometrically, the Huber loss can be interpreted as a smoothed version of the absolute error loss. This smoothing effect which eliminates the sharp corner at the origin present in the absolute loss function. The combination of this effect with the strong convexity of the metric in a uniform neighborhood around the origin $\varepsilon = 0$ provides desirable stability properties for optimization.

The aforementioned metrics are scale dependent. This means that their values depend on the scaling of the data. For that reason it is a common practice to apply percentage error metrics. The most well known metric is the **Mean Absolute Percentage Error (MAPE)** which is 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: $MAPE = 100 \cdot \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right|$, 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. $g_{MAPE}(x) = \arg\min_{g \in \mathcal{G}} \mathbb{E}\left[\left| \frac{g(X)-Y}{Y} \right| \middle| X = x \right]$, Empirical risk minimization can be used to estimate the optimum model g_{MAPE}. The relevant empirical optimization issue is: $\widehat{g}_{\mathrm{MAPE}}(x) = \arg\min_{g \in \mathcal{G}} \sum_{i=1}^{n} \left| \frac{g(X_i)-Y_i}{Y_i} \right|$. 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:

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

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 [33]. It's worthy to note also that MAPE is criticized for its asymmetry ([34], [35], [36]), 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 [34] 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 [36] introduced **Symmetric Mean Absolute Percentage Error (sMAPE)**, $sMAPE = Mean\left(100 \cdot \frac{2|y_t - \hat{y}_t|}{|y_t + \hat{y}_t|}\right)$. [14]. 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 ([37]) and the final version proposed by Chen and Yang ([38]), 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 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.

Therefore, our analysis starts with the original dataset, which is then resampled using the *Maximum Entropy Bootstrap* technique. This technique creates 20 synthetic datasets MEB $SD_1, SD_2, \ldots, SD_K$ that retain the dependence structure and temporal properties of the original series, allowing for robust inference under uncertainty. Afterwards, we perform *Harmonic Seasonal Model* against each synthetic dataset, capturing periodic variations inherent in traffic data by estimating coefficients of harmonic basis functions using either ℓ_1- or ℓ_2-norm coefficients estimation. These models extract the historical seasonal trends found in each resampled dataset. Next, *Least Squares Approximation* is used to refine the estimated coefficients, the previously estimated coefficients. Model selection is guided by minimizing the *Akaike Information Criterion* and the *Bayesian Information Criterion*, ensuring a balance between model fit and complexity. In parallel, each bootstrapped dataset undergoes *Penalized Maximum Tangent Likelihood Estimation*, a robust penalized approach intended to handle outliers and heavy-tailed distributions. To improve model performance, a greedy Hyperparameter Tuning 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. The rationale behind this procedure is that the VIPs give us the significance of each feature. Therefore, these features that have higher VIPs should have less penalty than those of smaller VIP values. This information is given to the *Adaptive* ℓ_1*-Lasso* regression framework in the form of the penalty factor of the lambda parameters. Afterwards, we implement a 10-fold *Bi-daily Leave-One-Block-Out Cross-Validation* (LOBOCV) for each

[14] These 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

model performed separately. This validation structure preserves temporal coherence by separating two-day blocks, which improves the model's ability to generalize to previously unexplored time periods. Within this Lasso phase, a supplementary hyper-parameter optimization 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 pipeline is depicted in Figure 3 below.

The aforementioned analysis is written in R statistical software. In 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]).

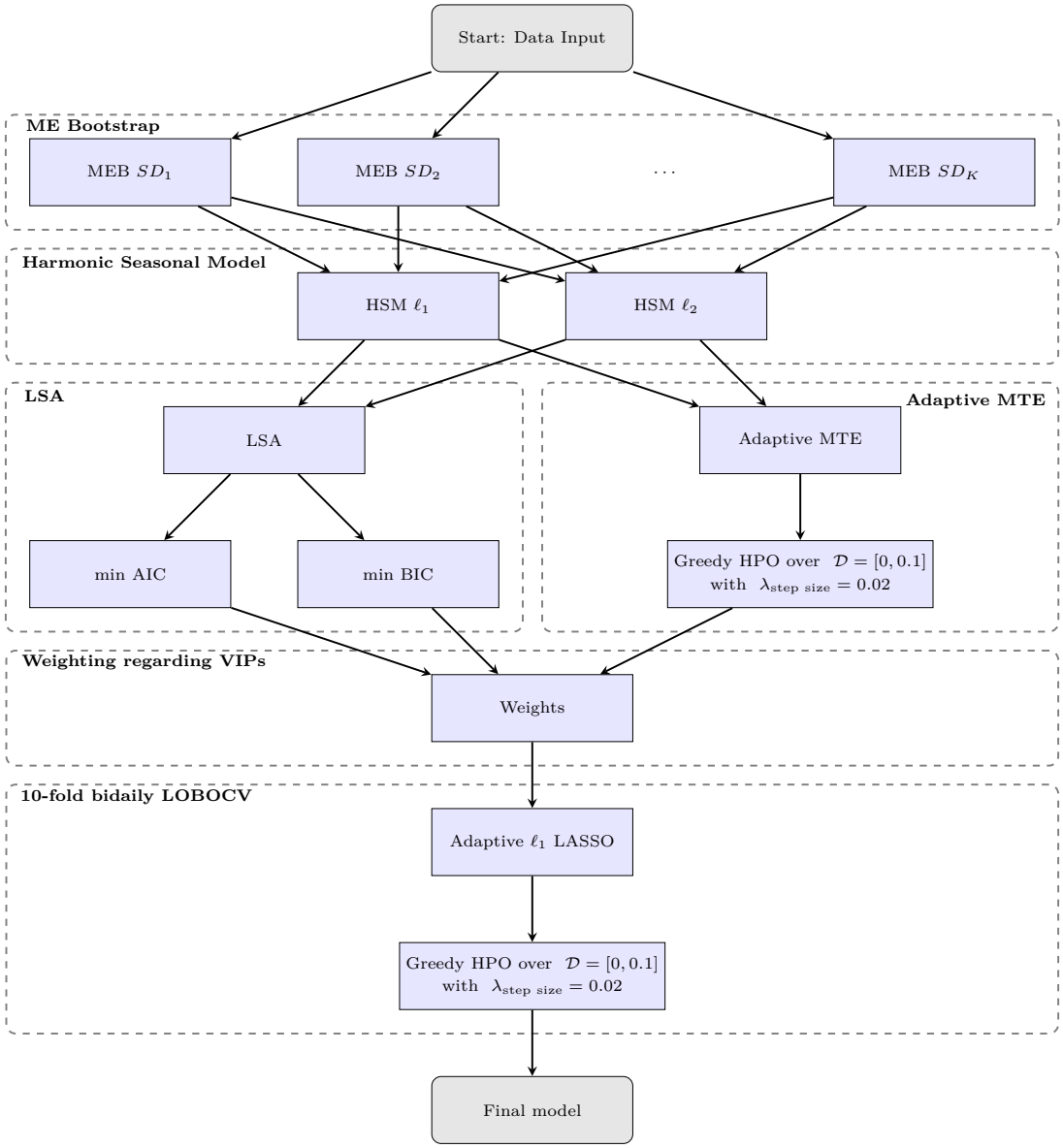

Figure 3: *The flowchart of our comparative analysis. We first generate the MEBoot datasets we apply Harmonic Seasonal Model using ℓ_1 and ℓ_2 norm. Afterwards, we apply the models we want to compare: LSA and MTE for ℓ_1 and ℓ_2 norm minimization. In the MTE scenario we apply a greedy fine tuning algorithm for λ in a set of $\mathcal{D} = [0, 0.1]$ with step values to be 0.02. On the other hand, we leverage the minimization of AIC and BIC on the LSA based methods. These procedures are applied on each MEBoot dataset and we collect the Variable Inclusion probabilities. The VIPs are fed into the Adaptive $\ell 1$ LASSO as weights. We apply again the greedy hyper parameter tuning algorithm over the set $\mathcal{D}$ with step values to be 0.02 for λ values.*

4 Results

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

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 chosen for its anticipated consistent traffic flow, in distinction to other locations where external factors, such as human behaviors, might exert considerable influence. 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 [15] 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. These results reveal a consistent pattern of identified significant harmonic frequencies (including the intercept) across most penalized variable selection methods, regardless the minimization criterion (AIC or BIC). The primary exception to this trend is the ℓ_1-norm Least Squares Approximation (ℓ_1LSA), which yields notably 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 ℓ_2LSA, will be presented more compactly. This will be followed by a dedicated discussion of the robust ℓ_1LSA 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.

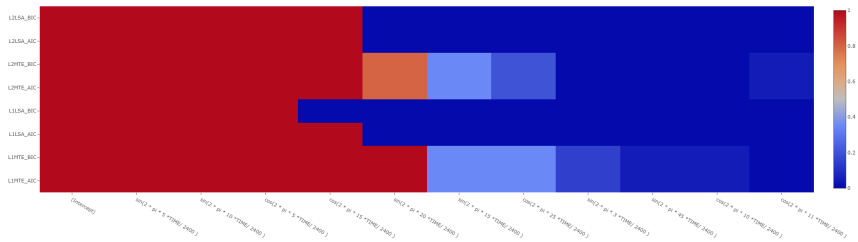

Figure 4: *Heatmap of Variable Inclusion Probabilities across all the models performed against 30 MEBoot samples regarding L102 location's data dynamics.*

[15] Charilaou Trikoupis's street is a unidirectional street, which intersects Alexandras Avenue from the opposite direction

Robust Maximum Tangent Likelihood Estimation (ℓ_1MTE)

The evaluation of the ℓ_1MTE approach starts with the AIC minimization ($\ell_1\text{MTE}_{\text{AIC}}$). Employing this model, features with a VIP of 100% (denoting consistent selection across all 20 MEBoot samples) include the Intercept and the harmonic terms $sin\left(2\pi \cdot x\frac{t}{T}\right)$ for $x \in \{5, 10, 20\}$, and $cos\left(2\pi \cdot x\frac{t}{T}\right)$ for $x \in \{5, 15\}$, where $T = 2400$ represents the period. Other regressors exhibit lower inclusion probabilities: $sin\left(2\pi \cdot 15\frac{t}{T}\right)$ and $cos\left(2\pi \cdot 25\frac{t}{T}\right)$ were selected with a probability of 35%; $sin\left(2\pi \cdot 3\frac{t}{T}\right)$ with 15%; and both $sin\left(2\pi \cdot 45\frac{t}{T}\right)$ and $cos\left(2\pi \cdot 10\frac{t}{T}\right)$ with 5%. These VIPs are visually depicted in Figure 5 and correspond numerically to the $\mathbf{L1MTE_{AIC}}$ column in Table 2.

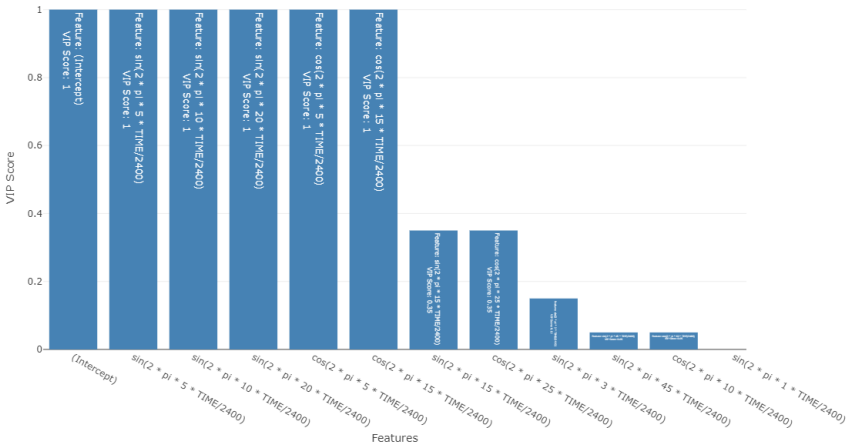

Figure 5: *Variable Importance Plot for the $\ell_1\text{MTE}_{\text{AIC}}$ model applied to location L102 data. The vertical axis refers to the probability of each feature to be included in the applied model and in the x axis the corresponding features in descending order (regarding its VIP value).*

This variable selection process leads to highly sparse coefficient vectors, as illustrated in Figure 6 and in table 10. In the figure below we can 6 see that the estimations of coefficients $\{cos(2\pi \cdot x\frac{time}{2400}), sin(2\pi \cdot y\frac{time}{2400})\}\big|_{x\in\{5,15,10\},y\in\{3,5,10,20,45\}}$ and the intercept have similar values across the MEBoot samples. This consistency verifies their significance. However, the features $cos(2\pi \cdot 25\frac{time}{2400})$ and $sin(2\pi \cdot 15\frac{time}{2400})$ seem to have more fluctuations across the datasets, taking into account their magnitude. These significant variability on their values amplitude indicates that their selection might have not a crucial impact on the inference of the original dataset dynamics.

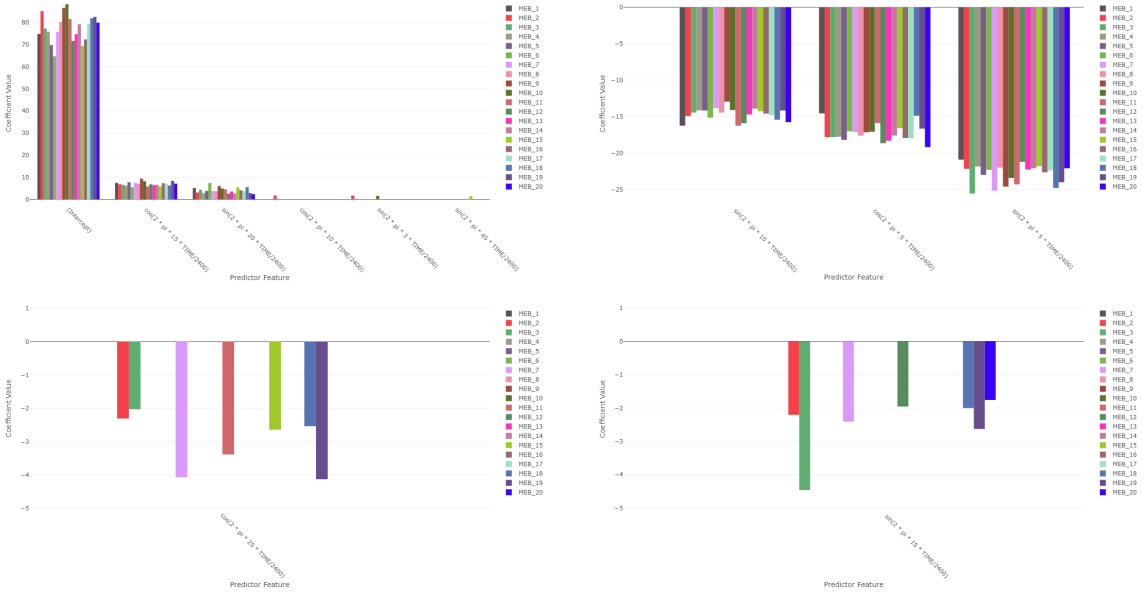

Figure 6: *Barplot of coefficient estimates for each MEBoot dataset using the* $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ *model at location L102. The top left barplot presents the values of the coefficients that consists of the set* $\{cos(2\pi \cdot x\frac{t}{T}), sin(2\pi \cdot y\frac{time}{T})\}\big|_{x\in\{15,10\},y\in\{3,20,45\}}$ *(where* $T = 2400$*, and* t *corresponds to time) and the intercept. The top right corresponds to* $\{cos(2\pi \cdot x\frac{t}{T}), sin(2\pi \cdot y\frac{t}{T})\}\big|_{x=5,y\in\{5,10\}}$*, the bottom left is for the* $cos(2\pi \cdot 25\frac{t}{T})$ *and the bottom right refers to the* $sin(2\pi \cdot 15\frac{t}{T})$*. The gap between the bars at the bottom line plots indicate that the model applied on the corresponding MEBoot samples avoid selecting those features.*

Figure 8 illustrates the performance of the $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ estimator at location L102, providing in-samples and out-of-sample evaluations. The time frame of the training set is the first two weeks of April restricted in the weekdays, between 3rd and 14th of April, while the testing set is the week between 17th and 21st of April in 2000. During the training phase, the fitted values (top-left panel) demonstrate similar patterns with data across the MEBoot samples, signifying accurate capture of primary temporal characteristics, particularly underlying periodicity. The corresponding training residuals (top-right panel) fluctuate around zero, but it still have heteroskedasticity evidence.
In the test phase, the model preserves its predictive structure (bottom-left panel), maintaining the naive seasonal trend and exhibiting robust generalization. The residuals of the test period (bottom-right panel), while indicating the violation of the homoskedastic property, remain centered around zero without a strong discernible evidence of remaining seasonality. The impact of robustness of the method is evident during specific periods with known external factors, for instance the late hours of April 5th and 7th (coinciding with political campaigns[16]) and April 20th (following a significant local sporting event[17]). The model yields stationary in mean residual time series, and the absence of residual periodicity confirms effective capture of seasonal patterns.

[16]Political campaigns occurred in central Athens during this week.
[17]Panathinaikos basketball club, with a fan club near Alexandras Street, won a European title on April 20th-21st.

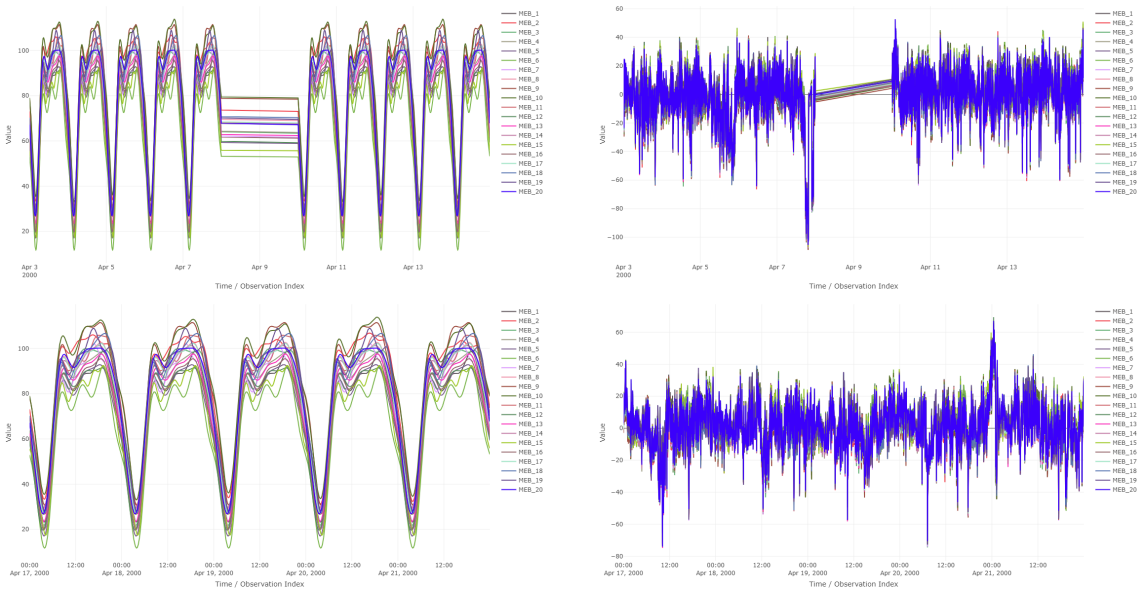

Figure 7: *Performance of the* $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ *model at location L102. The couple of the top line plots corresponds to the training period (time period between 3rd–14th April), while the bottom-line plots illustrates the results from the testing set (between 17th and 21st of April). The top left plot exhibits the fitted values for training period and the top right plot illustrates the residuals in the same time frame. The bottom left plot represents the forecasted values, while the bottom right is the testing residuals. The straight lines between 8th and 10th of April in the couple of plots of the top lane are caused from the absence of the weekend data and we connected the last data point of the first training week with the first data point of the last training week*

Figure 8: *Performance of the* $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ *model at location L102. The couple of the top line plots corresponds to the training period (time period between 3rd–14th April), while the bottom-line plots illustrates the results from the testing set (between 17th and 21st of April). The top left plot exhibits the fitted values for training period and the top right plot illustrates the residuals in the same time frame. The bottom left plot represents the forecasted values, while the bottom right is the testing residuals. The straight lines between 8th and 10th of April in the couple of plots of the top lane are caused from the absence of the weekend data and we connected the last data point of the first training week with the first data point of the last training week*

After collecting the VIPs from the application of the model against the MEBoot samples, we perform adaptive ℓ_1 LASSO as mentioned in methods section. A comparison of coefficient estimates from the model average of the $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ over MEBoot samples, from the model applied directly to the original sample and from a post-hoc adaptive ℓ_1-Lasso (`ADL1Lasso`) model is presented in Figure 9. The model average of the $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ yields sparser estimations compared to the `ADL1Lasso`. The comprehensive view (top panel) illustrates the corrective influence of the adaptive Lasso, particularly relative to the model average from 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. For sake of simplicity, we zoom in the parts where the non-zero coefficients of the model average and the actual model (model applied on the original dataset directly) located.

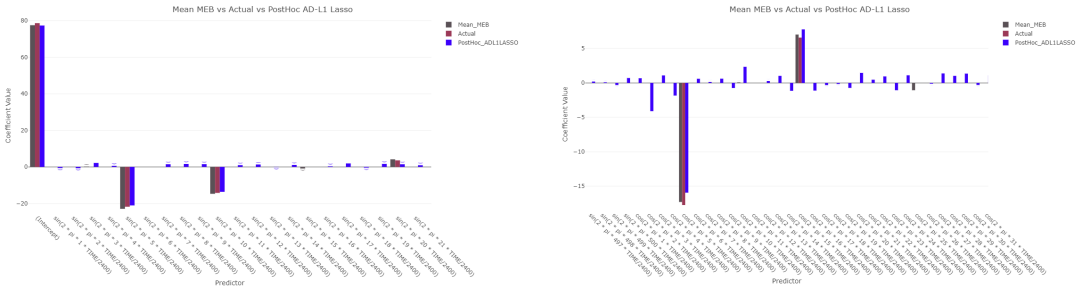

Figure 9: *Comparative barplot of average coefficients from the model averaging of ℓ_1MTE$_{\text{AIC}}$ applied on MEBoot samples, ℓ_1MTE$_{\text{BIC}}$ applied on the actual sample, and post-hoc adaptive ℓ_1-Lasso (weights from VIPs) for location L102. The features that the model average of model against the MEBoot samples and each model that applied on the initial dataset indicated as significant are estimated with similar values from the adaptive scheme of the ℓ_1 LASSO. The rest of the features have been estimated with substantially smaller values indicating that the adaptive LAD LASSO verifies the importance of the low frequencies.The model average, the actual model (model applied on initial dataset) and the adaptive LAD LASSO are distinguished using* ***dark grey-purple****,* ***muted wine red*** *and* ***electric indigo blue*** *colors correspondingly.*

Figure 10 displays violin plots for various error metrics, Huber Loss (HL), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Huber Loss, Symmetric Mean Absolute Percentage Error (SMAPE), and Symmetric Median Absolute Percentage Error (SMdAPE), for both training and testing datasets under the ℓ_1MTE$_{\text{AIC}}$ framework. MAE, RMSE, and Huber Loss exhibit proximate central tendencies and variability, with MAE and RMSE values typically between 11 and 13 units, according to table 17. It is worth highlighting that the model averaging of the ℓ_1MTE$_{\text{AIC}}$ across 20 Maximum Entropy Bootstrap samples consistently yields the second-lowest Huber loss across all three evaluation metrics (HL, MAE, and RMSE) in testing phases, deviating by merely 0.10-0.12 units from the absolute minimum. Moreover, the percentage-based error metrics further underscore its competitiveness as the method achieves the SMdAPE during testing dataset. For the remaining percentage error measures, the deviations from the minimum lie between 0.25-0.89 units in testing. In contrast, when the ℓ_1MTE$_{\text{AIC}}$ is applied directly to the original traffic dataset at location L102 (B), it attains the lowest RMSE, despite its foundation in minimizing absolute deviations. The values for the remaining performance metrics differ only slightly, ranging from 0.01 to 0.04 units relative to the optimal results in and testing contexts. 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. The symmetric Percentage-based metrics (SMAPE, SMdAPE) show wider distributions, centered around 100 (as they are percentages), reflecting sensitivity to low traffic volumes. Furthermore, they indicate 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_1MTE$_{\text{AIC}}$ 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. The low values of the SMAPE and SMdAPE reveals that this naive methodology has a nice balance between the forecasting performance and the model complexity. This is extremely useful for practitioners since they can retract missing values by leveraging the forecasting dynamics of this naive implementation, taking into account its robustness. Additionally, comparing the metrics results of the testing case against the training scenario, we can conclude the lack of overfitting.

Figure 10: *Violin plots of performance metrics (Huber Loss, MAE, RMSE, SMAPE, SMdAPE) for location L102 with* $\ell_1\text{MTE}_{\text{AIC}}$. *Navy blue diamond depicts the adaptive* ℓ_1*-Lasso results, the black diamond is for the* $\ell_1\text{MTE}_{\text{AIC}}$ *on initial data and the rest correspinds to the metrics distributions of the model average of* $\ell_1\text{MTE}_{\text{AIC}}$ *against MEBoot samples.*

Transitioning to the BIC type invariant, the $\ell_1\text{MTE}_{\text{BIC}}$ model yields remarkably similar results to its AIC counterpart regarding regressor significance and inclusion order across MEBoot samples (see table 2 [$\mathbf{L1MTE_{BIC}}$ column], and figure 4). Especially, coefficient estimates from $\ell_1\text{MTE}_{\text{BIC}}$ are indistinguishable from those obtained with $\ell_1\text{MTE}_{\text{AIC}}$. 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_1MTE framework without altering model structure or performance, thereby offering computational benefits. Therefore, we can now examine the ℓ_2 counterpart of the MTE methodology.

Conventional Maximum Tangent Likelihood Estimation (ℓ_2MTE)

This subsection details the performance of the conventional Maximum Tangent Likelihood Estimation, characterized by an ℓ_2-norm in its objective function (ℓ_2MTE), again employing the second-order Taylor approximation for the log-likelihood. An apparent observation is that both AIC and BIC minimization procedures for selecting the λ parameter yield identical sets of significant harmonic regressors, as in the ℓMTE case. Specifically, the intercept term and the regressors within the set $\left\{\sin\left(2\pi \cdot x\frac{t}{T}\right), \cos\left(2\pi \cdot y\frac{t}{T}\right)\right\}\Big|_{x\in\{5,10\},y\in\{5,15\}}$, where $T = 2400$, are unanimously identified as important predictors with VIP of 100%. Beyond these, $\sin\left(2\pi \cdot 20\frac{t}{T}\right)$ is included with VIP of 0.80 (16 out of 20 MEBoot samples). Less frequently selected features include $\sin\left(2\pi \cdot 15\frac{t}{T}\right)$ with VIP of 35% (7/20 samples), $\cos\left(2\pi \cdot 25\frac{t}{T}\right)$ with 20% (4/20 samples), and $\cos\left(2\pi \cdot 11\frac{t}{T}\right)$ with 5% (1/20 samples). These probabilities are documented in Table 2 (columns $\mathbf{L2MTE_{AIC}}$ and $\mathbf{L2MTE_{BIC}}$) and visualized in Figure 11.

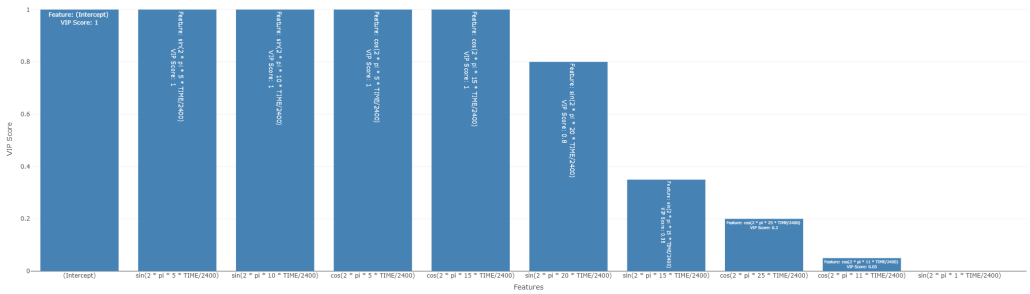

Figure 11: *Variable Importance Plots for both* $\ell_2\text{MTE}_{\text{AIC}}$ *and* $\ell_2\text{MTE}_{\text{BIC}}$ *applied to location L102 data.*

The coefficient estimates derived from the ℓ_2MTE approach (figure 12 and table 10) exhibit general similarity. Compared to the ℓ_1MTE models, a primary distinction lies in the VIP for $\sin\left(2\pi \cdot 20\frac{t}{T}\right)$, which is 80% under ℓ_2MTE rather than 100%. Furthermore, $\sin\left(2\pi \cdot 15\frac{t}{T}\right)$ maintains a VIP of 35%, while $\cos\left(2\pi \cdot 25\frac{t}{T}\right)$ has a reduced VIP of 20% (down from 35% in $\ell_1\text{MTE}_{\text{AIC/BIC}}$), and $\cos\left(2\pi \cdot 11\frac{t}{T}\right)$ 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 magnified views of the coefficient barplots if certain terms are frequently zero.
The coefficient values for $\{cos(2\pi \cdot x\frac{t}{T}), sin(2\pi \cdot y\frac{t}{T})\}_{x\in\{5,15\},y\in 5,10,20}$ and the intercept seem to have small variance across the MEBoot samples which can be explained from the slight variability of the data points od each dataset. This fact introduces the stability property of the methods to identify those features as significant. On the other side, the features $cos(2\pi \cdot 25\frac{t}{T})$, and $sin(2\pi \cdot 15\frac{t}{T})$ are identified as secondary features where as their significance is not presented in each MEBoot sample.

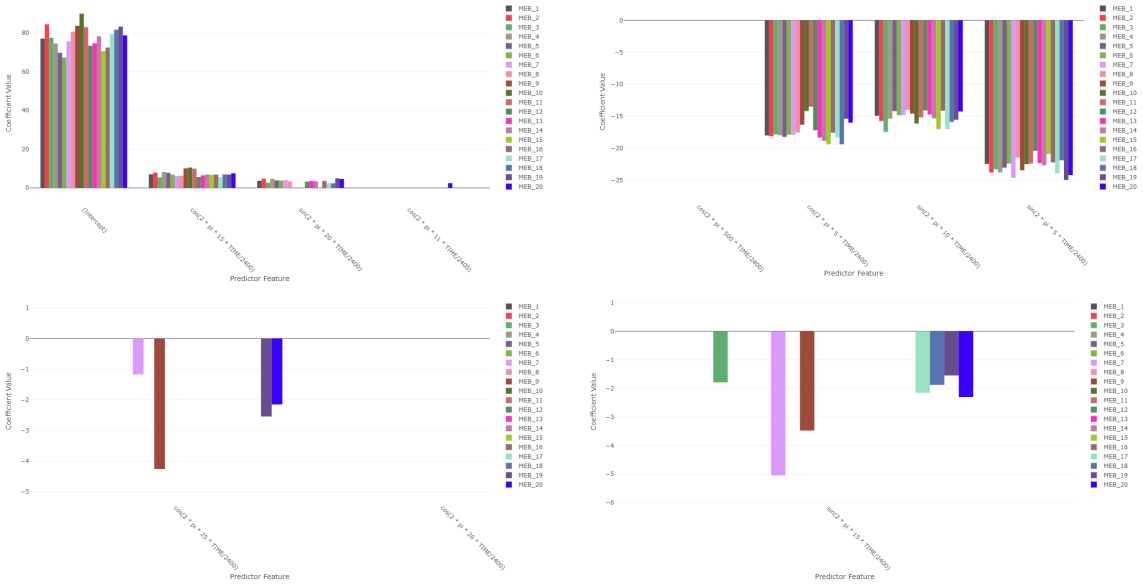

Figure 12: *Coefficients estimation barplot for each MEBoot dataset while performing* ℓ_2MTE *under any information criteria minimization.*

Given that the AIC and BIC minimization procedures for the penalization parameter λ lead to the same set of selected regressors within the ℓ_2MTE framework, the resulting fitted values, forecasted values, and associated residuals are have identical results. Figure 13 presents these outcomes for both $\ell_2\text{MTE}_{\text{AIC}}$ and $\ell_2\text{MTE}_{\text{BIC}}$. We adopt the same presentation style as before, where the panels display fitted values and residuals for the training period (3rd-14th April), alongside predicted values and forecast errors for the test period (17th-21nd April). Both estimators effectively capture the underlying traffic flow periodicities and demonstrate robust predictive capability.
The errors on the training part indicates the appearance of an anomaly on the data on the exact date of political campaigns and last weekday before the elections. The fitted values seem to have smooth oscillations which makes it better for inference and interpretability. The testing errors shows a significant pattern of heteroskedasticity and singularity in the data on the date of a sports great event that caused celebrations across that street on the first hours of the 21st of April.

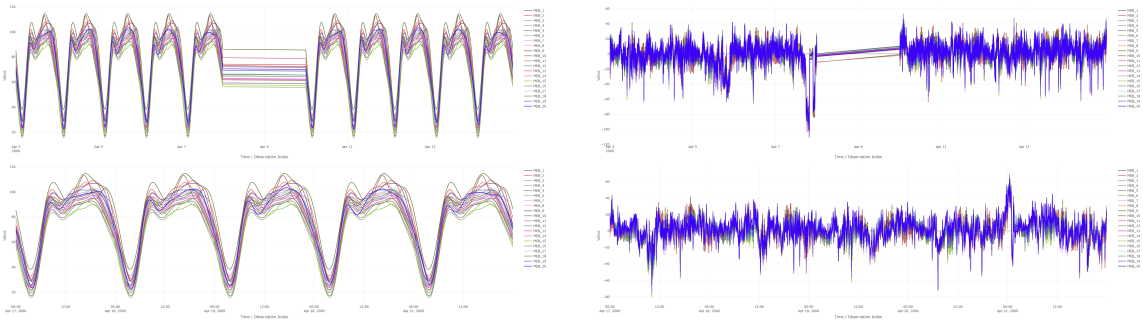

Figure 13: *Performance of* ℓ_2MTE *against training and testing scenaria. The top left plot corresponds to the training fitted values, and the top right to the errors on the training sample. The bottom left illustrates the forecasting on testing scenario and the bottom right plot depicts the test forecast errors. This quartet of plots represent both cases of AIC- and BIC-type model selection of the method.*

Once again, we conducted the adaptive LAD-LASSO incorporating the VIPs gathered by the application of the model against the MEBoot samples. Notably, the adaptive LAD LASSO identifies the significance of the same low harmonics that the model average of each model revealed. The most intriguing part is that the coefficient estimates that the adaptive ℓ_1 LASSO assigns to these features are similar to those of the estimations of the conventional MTE models and the model average. The rest of the harmonic features seem to have substantially lower values, which proves the concordance of those three models regarding the significance of the low frequencies.

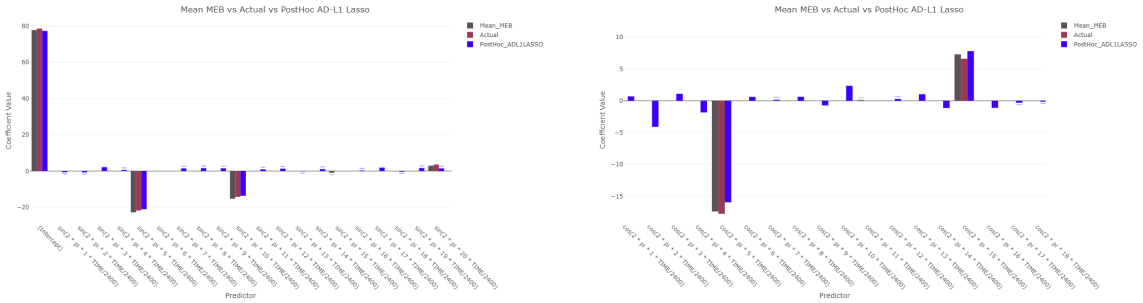

Figure 14: *Comparative barplots of coefficients for location L102 gathered from the model average of the* ℓ_2MTE *regardless the minimization of the AIC and BIC criteria. Comparisons are between coefficient estimates from the model average of the* ℓ_2 *MTE which applied against MEBoot samples, the model applied on the initial dataset, and adaptive* ℓ_1*-Lasso (incorporating weights from VIPs). The are distinguished using* ***dark grey-purple****,* ***muted wine red*** *and* ***electric indigo blue*** *colors correspondingly.*

The distributions of key performance metrics for conventional MTE models at location L102 are illustrated in Figure 15. Each plot shows the distribution across MEBoot samples, with the navy blue diamond indicating the adaptive ℓ_1-Lasso performance on the original data, and the black

diamond representing the ℓ_2MTE 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_2MTE framework and the adaptive estimation schemes in characterizing traffic flow dynamics.

Since the performance of the ℓ_2MTE 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 $\ell_2\text{MTE}_{\text{AIC/BIC}}$ achieves the lowest values for HL, MAE, and RMSE across the 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. Interestingly, the symmetric percentage error metrics (SMAPE and SMdAPE) are minimized by the robust $\ell_1\text{MTE}_{\text{AIC}}$ under model averaging, with the sole exception of SMdAPE in the test set, where the ℓ_2MTE result is merely 0.03 units above the minimum. When applied directly to the original traffic dataset, the ℓ_2MTE 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 testing timeframe. The post-hoc adaptive ℓ_1 LASSO model displays commendable generalization capabilities. In the testing scenario, this LASSO-based model remarkably attains minimum values for HL, MAE, and SMdAPE, with respective scores of 11.19, 11.68, 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_2MTE 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.

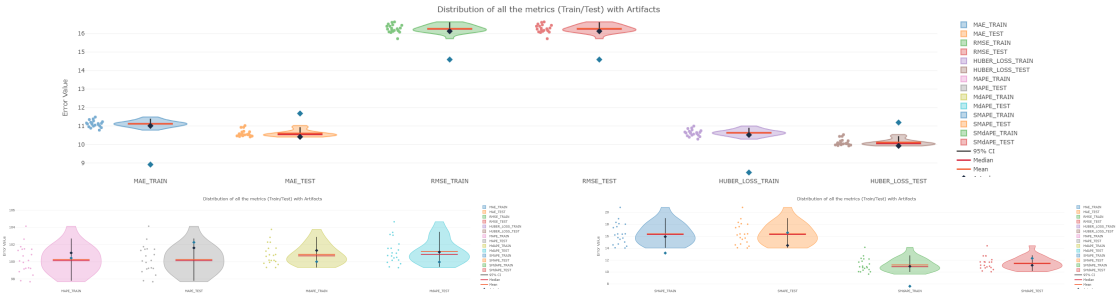

Figure 15: *Violin plots of performance metrics for location L102 of the previously presented models. The metrics presented here corresponds to* $\ell_2\text{MTE}_{\text{AIC}}$*. The* $\ell_2\text{MTE}_{\text{BIC}}$ *has identical results therefore these violin plots represent both models. The navy blue 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.*

Robust Least Square Approximation of Unified LASSO (ℓ_1LSA)

We now analyze the results obtained from the robust implementation of the least-squares approximation within the unified LASSO framework (ℓ_1LSA). 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 AIC ($\ell_1\text{LSA}_{\text{AIC}}$). When applied to 20 MEBoot samples, this technique consistently identified the following features as significant, demonstrating a 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 16 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.

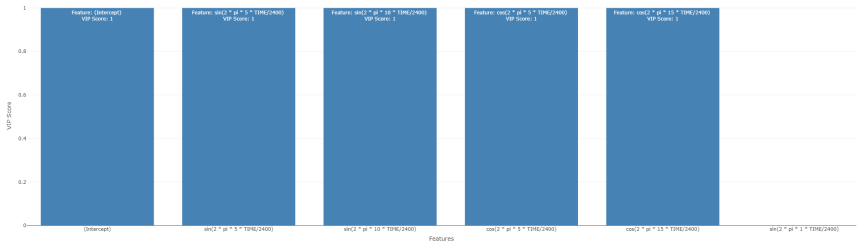

Figure 16: *Variable Importance Plot for* $\ell_1\text{LSA}_{\text{AIC}}$ *on L102*

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 17). 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 ℓ_1LSA framework, rendering them unnecessary for inference purposes. Furthermore, the coefficient estimates exhibit remarkable consistency across MEBoot samples, demonstrating minimal variance (Figure 17 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.

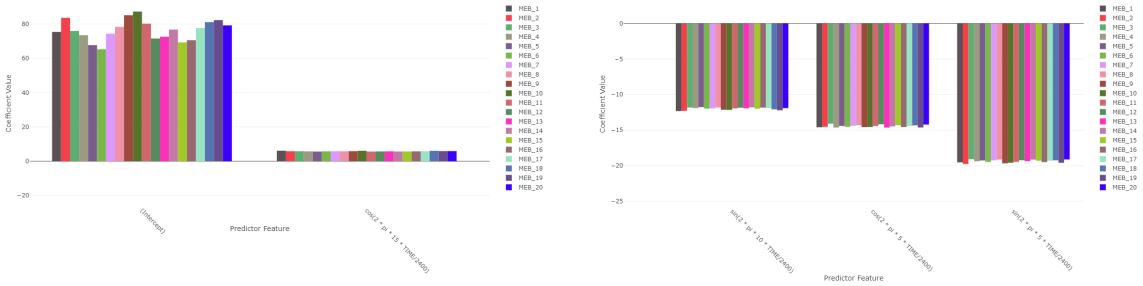

Figure 17: *Coefficients estimation barplot for each MEBoot dataset while performing* $\ell_1\text{LSA}_{\text{AIC}}$ *on L102. The right plot shows the coefficient estimation of the Intercept and the* $\cos\left(2\pi 15 \cdot \frac{t}{T}\right)$ *and the right plot shows the coefficient estimation of the* $\sin\left(2\pi \cdot 10\frac{t}{T}\right)$, $\sin\left(2\pi \cdot 5\frac{t}{T}\right)$ *and* $\cos\left(2\pi \cdot 5\frac{t}{T}\right)$. *The changes are subtle across the datasets.*

For computational feasibility, we employed a naive estimation of the variance-covariance matrix for the absolute deviation minimization within the ℓ_1LSA method [18]. This approach, based on Pollard's method [15], 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_1LSA, as illustrated in Figure 18.

The fitted data exhibit a strong and consistent periodicity, with distinct cyclical patterns indicating well-defined daily or subdaily seasonality (Figure 18, 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 18, fluctuate around zero, indicating minimal average bias and successful extraction of the primary deterministic components. It's noteworthy that the residuals in both training and testing scenario are stationary in mean, as it was in the MTE invariant models. 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, although the presence of the heteroskedasticity is discernible. 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. On the forecast errors, 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 ability to generalize and strong predictive performance.

It is really interesting that we can easily notice that the fitted and the forecasted values can be described as a family of curves with different location parameter. In other words, we can notice that for each pair of curved in each plot we can see that they have constant vertical distance. This fact can be described by an overshrinking effect which explains the excluding of the $sin\left(2\pi \cdot 20\frac{t}{T}\right)$. The overshrinkage will cause model with small number of features, which makes the corresponding curve to be simpler.

[18] We attempted to account for autocorrelation and heteroskedasticity of the samples using Barrett kernel filtering in Andrews's [39] implementations (Newey-West[22]). We attempted to use Cribari-Neto [20], Cribari-Neto-Da Silva [21], and Galvao-Yoon [40] 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 [15] naïve but robust estimation of the covariance matrix.

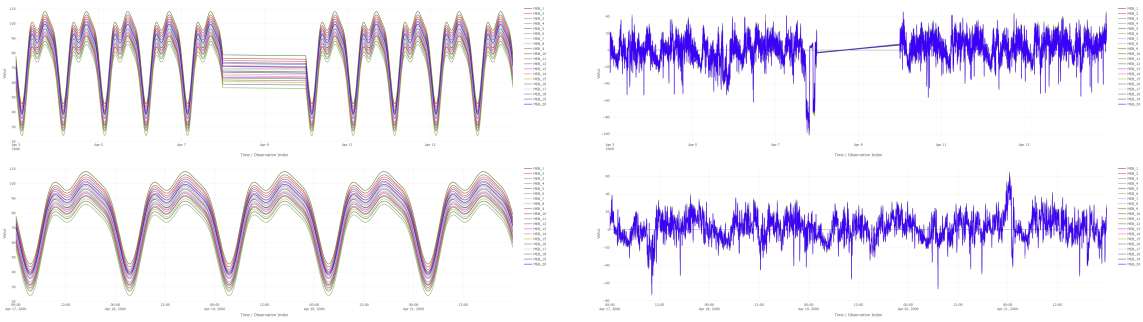

Figure 18: *The couple of panels in the top lane depict the fitted values and residuals, respectively, for the training period (3rd–14th April). The remaining pair illustrates the forecasted values and corresponding errors during the test period (17th–21st April), based on the application of* $\ell_1\text{LSA}_{\text{AIC}}$ *on location L102.*

As in previous analyses, we applied $\ell_1\text{LSA}_{\text{AIC}}$ 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 19 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 $\cos\left(2\pi \cdot 15\frac{t}{2400}\right)$, where the post-hoc LASSO estimate is approximately double that of the other two models.

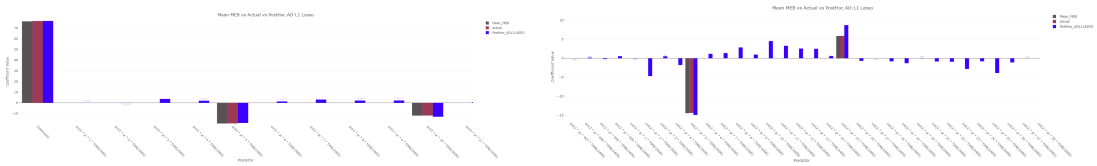

Figure 19: *Comparative barplots of coefficients for location L102 gathered from the model average of the* $\ell_1\text{LSA}_{AIC}$ *. Comparisons are between coefficient estimates from the model average of the* ℓ_1 *LSA which applied against MEBoot samples, the model applied on the initial dataset, and adaptive* ℓ_1*-Lasso (incorporating weights from VIPs). The are distinguished using* ***dark grey-purple****,* ***muted wine red*** *and* ***electric indigo blue*** *colors correspondingly.*

Figure 20 provides a comprehensive depiction of various error metrics through violin plots, illustrating their distributional properties and central tendencies for $\ell_1\text{LSA}_{\text{AIC}}$. 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 thin violin shapes, signifying

consistent accuracy and resilience. According to table 17, when applied through model averaging, the ℓ_1LSA method exhibits stable yet non-optimal performance. In testing scenario, 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 testing datasets under MEB sampling. In terms of percentage-based error metrics, this model performs relatively weakly. The 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_1LSA 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. The SMAPE improves marginally to 17.61, and SMdAPE to 12.29.

Overall, the ℓ_1LSA 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_2MTE and adaptive ℓ_1LASSO.

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 20 effectively highlights the efficiency of the $\ell_1\text{LSA}_{\text{AIC}}$ 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.

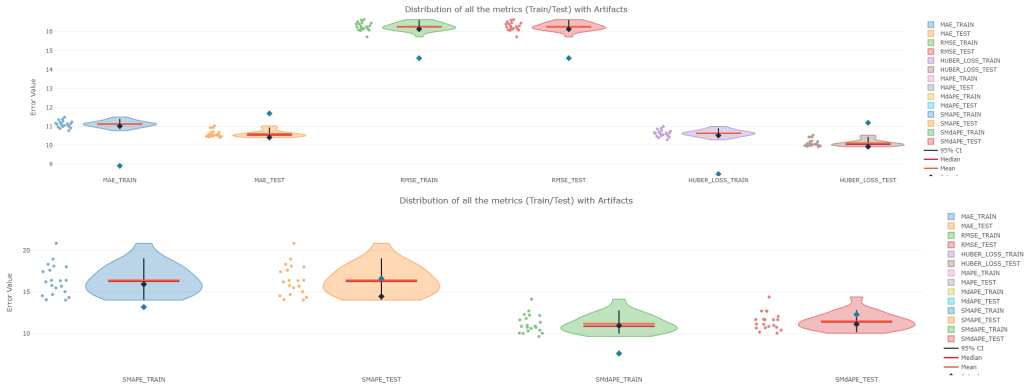

Figure 20: *Violin plots of performance metrics for location L102 while applying* $\ell_1\text{LSA}_{\text{AIC}}$*. The navy blue diamond presents the adaptive* ℓ_1 *Lasso and the black one shows the* $\ell_1\text{LSA}_{\text{AIC}}$ *on the initial data of L102 location. The rest came from the application of* $\ell_1\text{LSA}_{\text{AIC}}$ *on the MEBoot samples of L102 location data.*

Focusing on the BIC minimization for the ℓ_1LSA method ($\ell_1\text{LSA}_{\text{BIC}}$), we observe that the same features were selected as for $\ell_1\text{LSA}_{\text{AIC}}$, with the notable exception of the harmonic *cos* $\left(15\frac{t}{T}\right)$ (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 21). 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 21 and detailed in Table 2. We can also see these results in Figure 4.

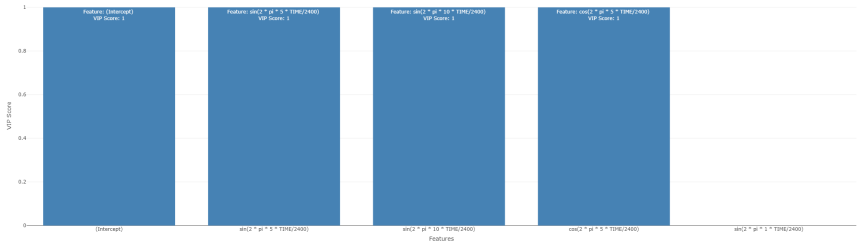

Figure 21: *Variable Importance Plot for* $\ell_1\text{LSA}_{\text{BIC}}$ *on L102*

The coefficient estimates for $\ell_1\text{LSA}_{\text{BIC}}$ 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 the estimation of coefficients using this invariant ℓ_1LSA. As shown in Figure 22 and

Table 10, the dispersion of value estimations is not statistically significant, indicating that, even with a limited set of 20 MEBoot samples, the ℓ_1LSA invariant models converge to specific values for each regressor.

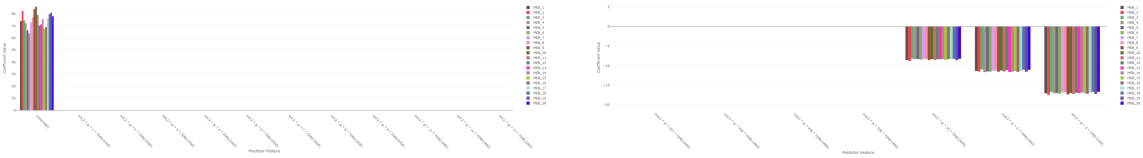

Figure 22: *Coefficients estimation barplot for each MEBoot dataset while performing* $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$ *on L102*

Excluding the $\cos\left(2\pi \cdot 15\frac{t}{2400}\right)$ regressor results in fitted values exhibiting less steep and wide fluctuations. However, the residual plots reveal more irregularities (Figure 23). Specifically, the $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$ 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 23, where the right-hand plots, representing residuals on training and testing data, appear to exhibit a seasonal pattern. This is caused by the smoothed forecasted values which might be caused by the naive implementation of the variance covariance matrix.

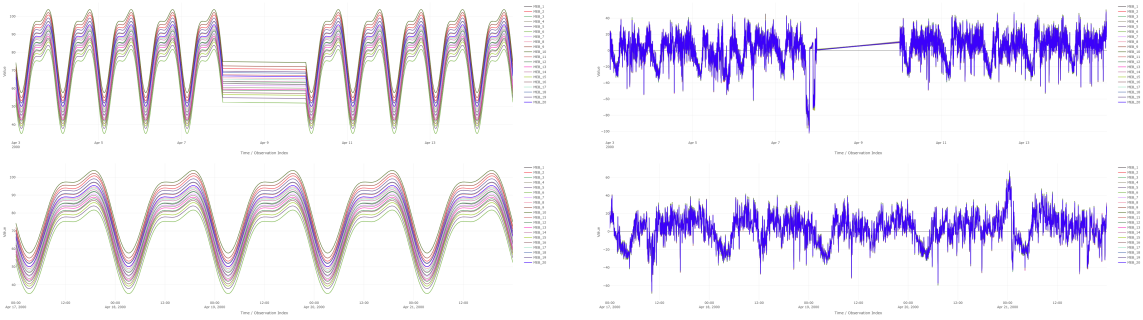

Figure 23: *The couple of panels in the top lane depict the fitted values and residuals, respectively, for the training period (3rd–14th April). The remaining pair illustrates the forecasted values and corresponding errors during the test period (17th–21st April), based on the application of* $\ell_1\mathrm{LSA}_{\mathrm{AIC}}$ *on location L102.*

As anticipated, Figure 24 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 $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$ 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.

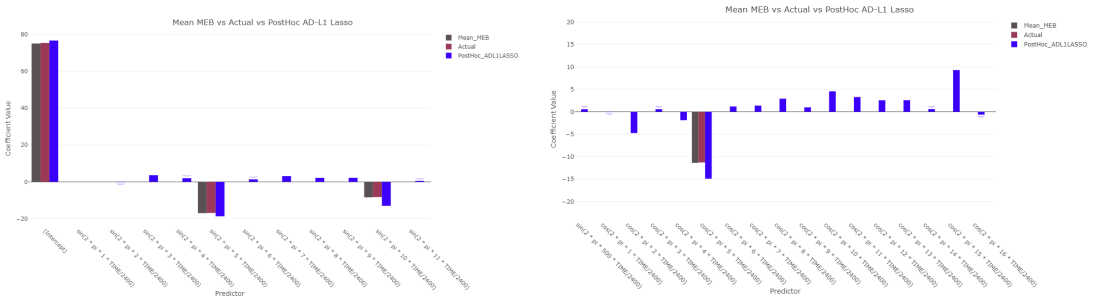

Figure 24: *Comparative barplot between average coefficients from models applied against MEBoot samples, from models against initial sample and post-hoc adaptive ℓ_1-LASSO which conducted using VIPs as weights for L102 while applying $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$.Comparisons are between coefficient estimates from the model average of the ℓ_1 LSA which applied against MEBoot samples, the model applied on the initial dataset, and adaptive ℓ_1-Lasso (incorporating weights from VIPs). The are distinguished using* ***dark grey-purple****,* ***muted wine red*** *and* ***electric indigo blue*** *colors correspondingly.*

Figure 25 demonstrates that the $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$ 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.

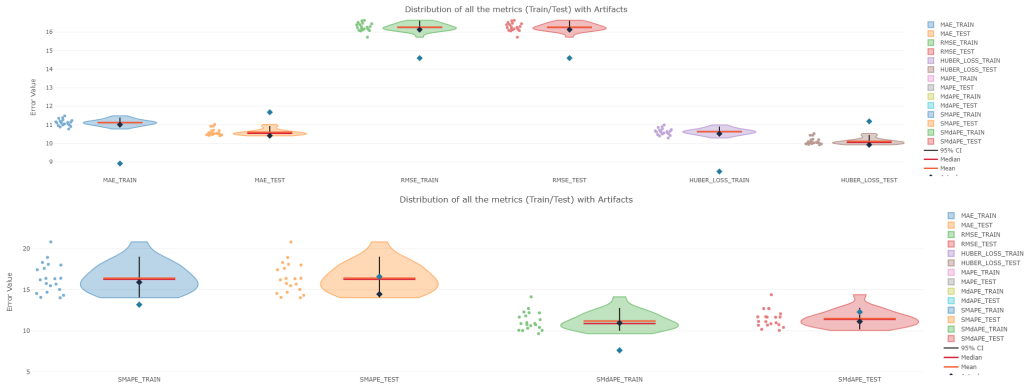

Figure 25: *Violin plots of performance metrics of Huber Loss, MAE, RMSE, SMAPE, and SMdAPE for location L102 while applying $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$. The navy blue diamond presents the adaptive ℓ_1 Lasso and the black one shows the $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$ on the initial data of L102 location. The rest came from the application of $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$ on the MEBoot samples of L102 location data.*

Conventional Least Square Approximation of Unified LASSO (ℓ_2LSA)

We now turn our attention to examining the performance of the ℓ_2LSA method under both AIC and BIC minimization frameworks. Notably, both the $\ell_2\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{BIC}}$ 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 $\ell_1\text{LSA}_{\text{AIC}}$ procedure. The behaviors of the $\ell_2\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{BIC}}$ are identical, therefore we are going to focus on $\ell_2\text{LSA}_{\text{AIC}}$. Same results and remarks occur for the $\ell_2\text{LSA}_{\text{BIC}}$.
It is crucial to emphasize that both ℓ_2LSA procedures, similar to $\ell_1\text{LSA}_{\text{AIC}}$, uniformly selected these characteristics in all MEBoot sampled scenarios, underscoring their consistent role as crucial predictors for inferring and forecasting the underlying periodic behavior (Figure 26).

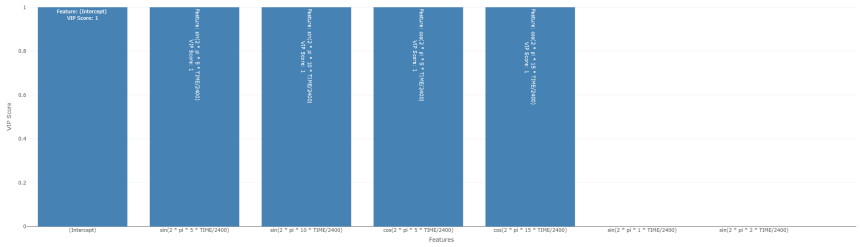

Figure 26: *Variable Importance Plot for ℓ_2LSA on L102 location.*

Given the consistent identification of the same predictor set by the ℓ_1LSA procedures, it is entirely expected that the resulting coefficient estimates from $\ell_2\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{BIC}}$ are nearly indistinguishable. This observation is clearly evidenced in Figure 27 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.

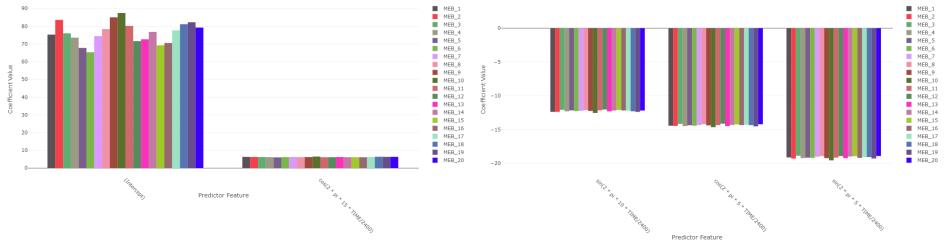

Figure 27: *Coefficient estimates for each MEBoot dataset while performing* ℓ_2LSA *on L102 location We can see that the variation of the coefficient values is not significant. This variability is cause by the variation among the MEBoot samples .*

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 28. 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 28 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.

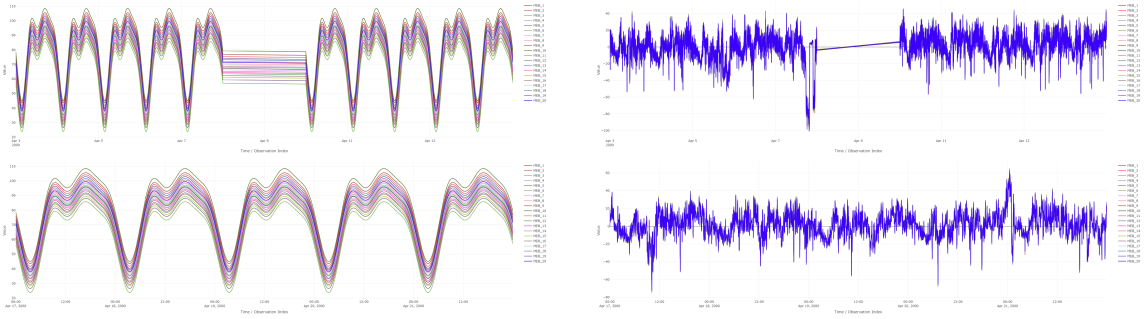

Figure 28: *The first two panels (top-left and top-right) depict the fitted values and residuals for the training period (April 3rd–14th), respectively. The subsequent two panels (bottom-left and bottom-right) illustrate the forecasted values and corresponding errors during the test period (April 17th–21st), 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 29 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.

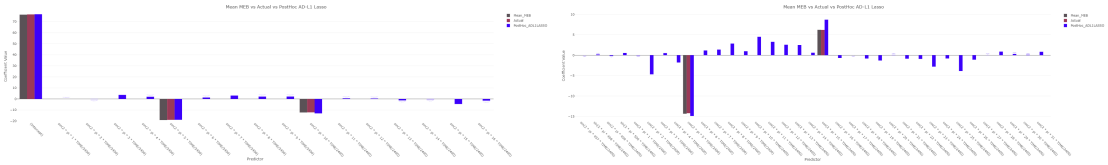

Figure 29: *These barplots provide a comparative analysis of average coefficients obtained from maximum entropy bootstrap samples, the actual sample, and the actual samples performing adaptive ℓ_1-Lasso with weights constructed by VIPs for L102, under both criterion minimization processes.Comparisons are between coefficient estimates from the model average of the ℓ_2 LSA which applied against MEBoot samples, the model applied on the initial dataset, and adaptive ℓ_1-Lasso (incorporating weights from VIPs). The are distinguished using* ***dark grey-purple****,* ***muted wine red*** *and* ***electric indigo blue*** *colors correspondingly.*

The performance of the ℓ_2LSA 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 ℓ_2LSA achieves solid, though not leading, performance. 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. The SMAPE and SMdAPE values are moderate where 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_2LSA 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. 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 30, mirroring the analysis conducted for ℓ_1LSA, offers a detailed and statistically informative visualization of the error characteristics associated with the $\ell_2\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{BIC}}$ 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. 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_2LSA method's capacity for stable, unbiased predictive performance and robust generalization across bootstrapped samples, affirming its methodological soundness and practical applicability.

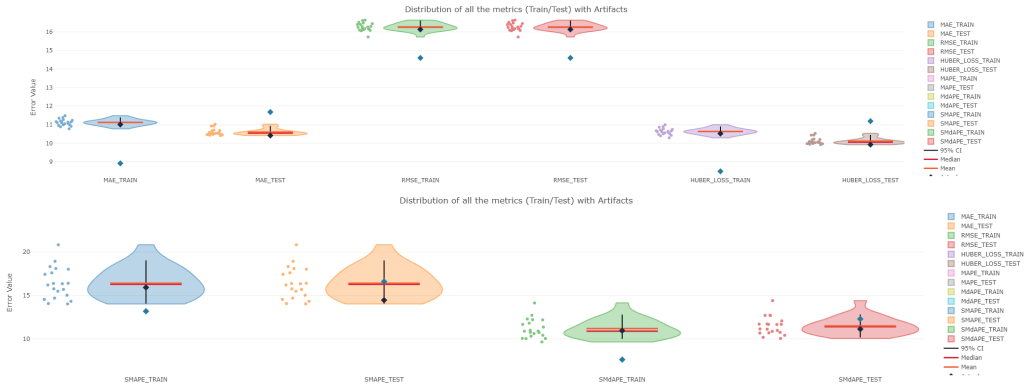

Figure 30: *These two sets of violin plots illustrate the performance metrics for location L102. The upper triad corresponds to* ℓ_2LSA$_{\text{AIC}}$ *and the lower triad to* ℓ_2LSA$_{\text{BIC}}$*. The navy blue diamond denotes the adaptive* ℓ_2 *Lasso, while the black diamond represents the corresponding 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 28, we can corroborate the persistence of periodic structures within the residuals of the model averaging across the MEBoot samples. The sinusoidal patterns identified in the ℓ_1LSA$_{\text{AIC}}$ 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_2LSA 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 **??**.

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 31) 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_2LSA models, as their behavior reveals more pronounced differences across the various information criterion minimization processes compared to the ℓ_1LSA models.
Specifically, the ℓ_2LSA method, when minimizing the AIC, consistently selects 100% of the predic-

tors 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 31, and Figure 32, the selected regressors under BIC minimization are:

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

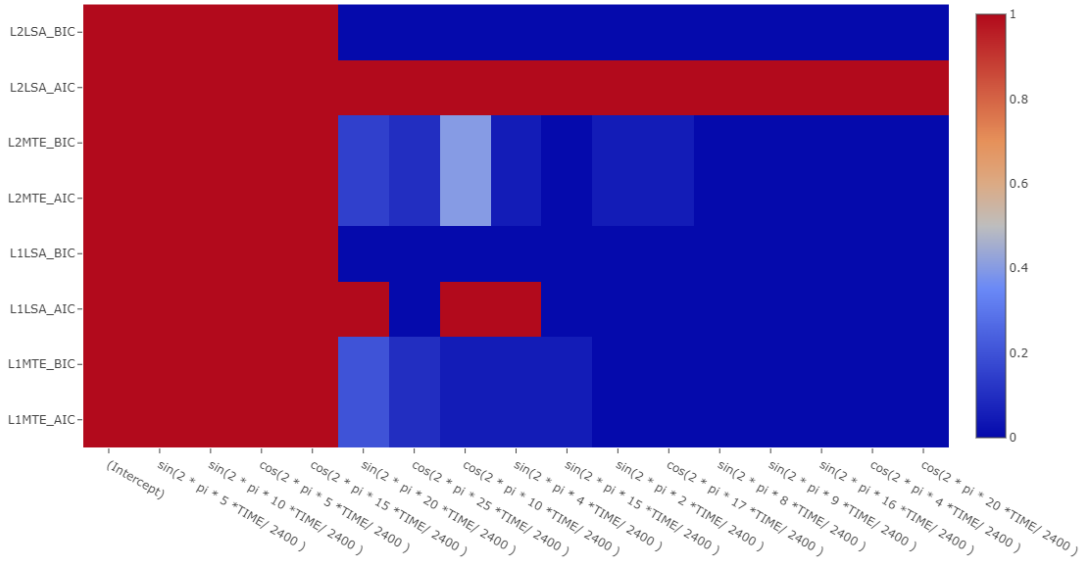

Figure 31: *Heatmap depicting the model-wise level of significance of regressors after applying the models against 20 MEBoot samples of L106 location data.*

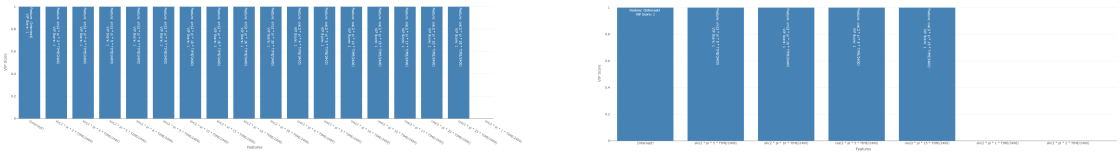

Figure 32: *Bar plots of Variable Inclusion Probabilities for* $\ell_2\text{LSA}_{\text{AIC}}$ *(left panel) and* $\ell_2\text{LSA}_{\text{BIC}}$ *(right panel) applied to 20 MEBoot samples based on L106 location.*

Naturally, this divergence in selected predictors leads to substantial differences in the coefficient estimates between the $\ell_2\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{BIC}}$ models, primarily through the elimination of numerous

coefficients under BIC minimization, as explicitly depicted in Figure 33. While the intercept and the coefficient for $\cos\left(2\pi \cdot 15\frac{t}{2400}\right)$ remain highly similar across both models, $\ell_2\text{LSA}_{\text{BIC}}$ estimates the remaining common coefficients with considerably reduced amplitudes.

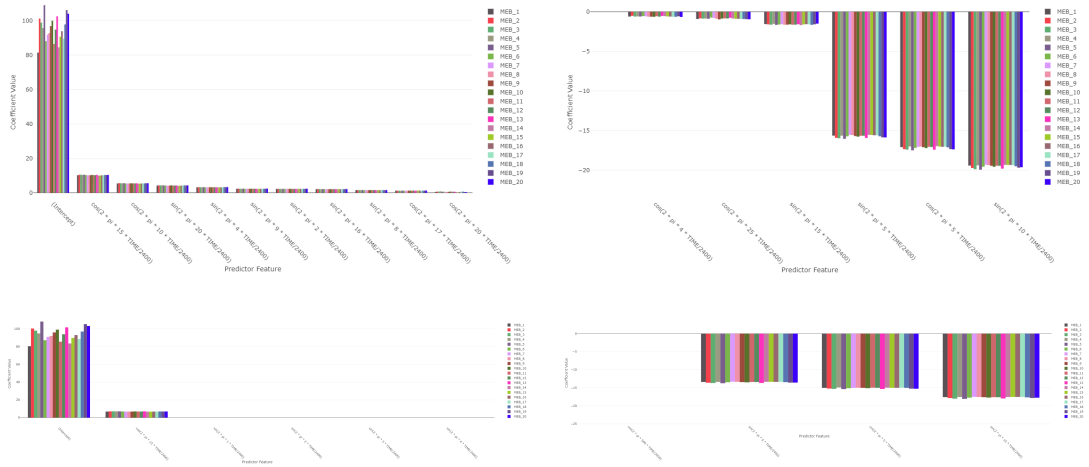

Figure 33: *Coefficient estimates for $\ell_2\text{LSA}_{\text{AIC}}$ (pair of plots in the top lane) and $\ell_2\text{LSA}_{\text{BIC}}$ (pair of plots in the bottom lane) applied to 20 MEBoot samples based on L106 location.*

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 34. 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.

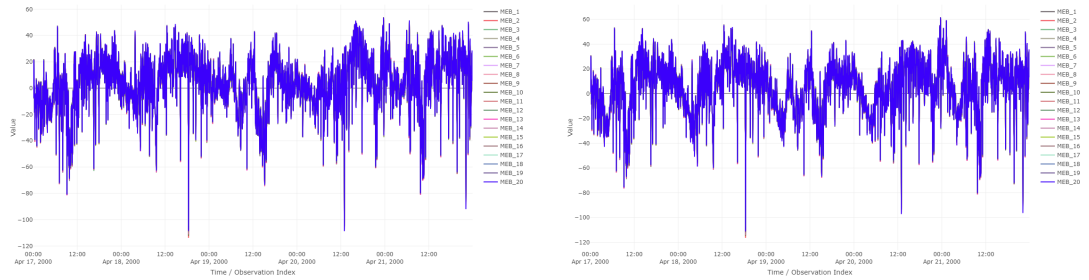

Figure 34: *Residual plots on the test data for $\ell_2\text{LSA}_{\text{AIC}}$ (left plot) and $\ell_2\text{LSA}_{\text{BIC}}$ (right plot) applied to 20 MEBoot samples based on L106 location.*

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 $\cos\left(2\pi \cdot 20\frac{t}{2400}\right)$. Conversely, BIC minimization leads to models that also share similar coefficients, albeit with the exception of the predictor $\cos\left(2\pi \cdot 15\frac{t}{2400}\right)$, as depicted in Figure 35.

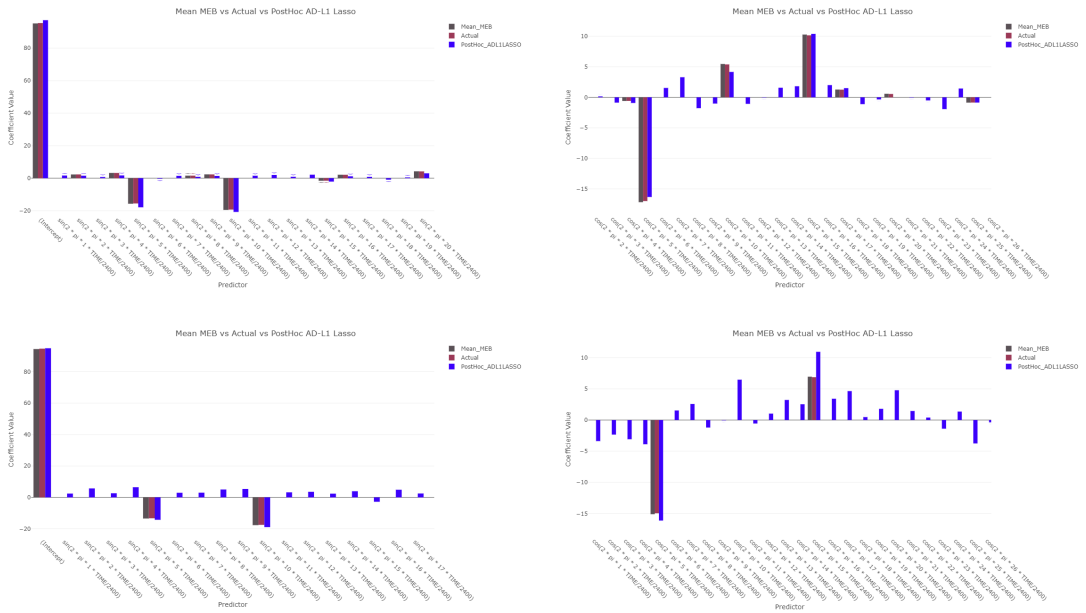

Figure 35: *Coefficient estimates for ℓ_2LSA$_{\text{AIC}}$ (upper pair of plots) and ℓ_2LSA$_{\text{BIC}}$ (bottom pair of plots) applied to 20 MEBoot samples, the original traffic volume data at L106, and the corresponding post-hoc variable selection technique.*

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_2LSA 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 36 and numerically in table 20. Using MEBoot sampling, ℓ_2LSA$_{\text{AIC}}$ achieves HL, MAE, and RMSE values of 15.66, 16.15, and 24.62 in the testing case. 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 13.80, 13.92, and 20.06 in testing scenario, respectively. This results in significant gains of 6-9 units across all metrics compared to the raw MEBoot output. In comparison, the ℓ_2LSA$_{\text{BIC}}$ model performs similarly at the MEB level, with HL, MAE, and RMSE of 16.64, 17.13, and 25.52 in testing week. Post-hoc adaptive LASSO improves performance, achieving the lowest RMSE (19.93) in the testing days. 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, 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, 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 36 compares the error distributions for $\ell_2\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{BIC}}$ 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_2LSA estimation, resulting in increased predictive performance and greater generalization as confirmed across both sampling-based and original data assessments.

Figure 36: *Violin Plots of MAE, RMSE, and Huber loss on both test and train 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.*

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, 2000). Unlike L106, which exhibits singularities in the testing data, the irregularities at L104 were predominantly observed in the testing data.

Comparison 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 (43, 44, 45, 46, 39, 40, 41, and 42) 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 $cos\left(2\pi \cdot 5\frac{t}{2400}\right)$, $sin\left(2\pi \cdot 5\frac{t}{2400}\right)$, $cos\left(2\pi \cdot 15\frac{t}{2400}\right)$, and $sin\left(2\pi \cdot 10\frac{t}{2400}\right)$, appear to be the most frequently selected and profoundly 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, except L105) within each distinct model and criterion combination. This

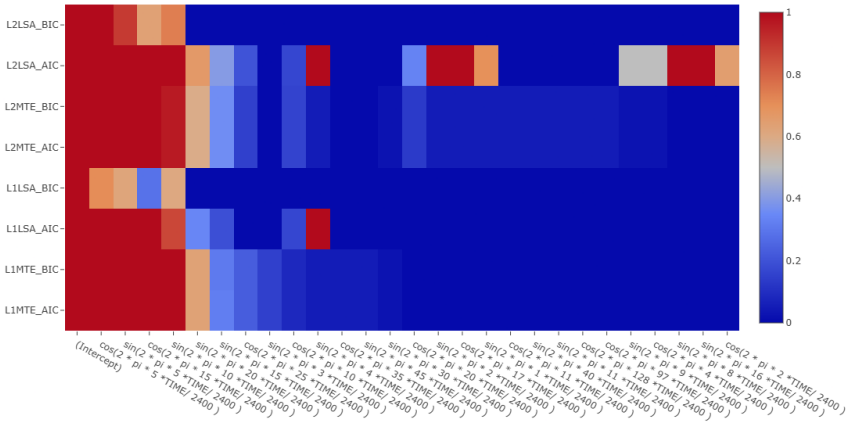

Figure 37: *Variable Inclusion Probabilities of each model over all the locations*

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 37 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 $\ell_1\text{LSA}_{\text{BIC}}$, $\ell_2\text{LSA}_{\text{BIC}}$, $\ell_1\text{MTE}_{\text{BIC}}$, and $\ell_2\text{MTE}_{\text{BIC}}$ 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_2LSA and ℓ_1/ℓ_2MTE model families. The 37 heatmap highlights that $\ell_1\text{LSA}_{\text{AIC}}$ and $\ell_2\text{LSA}_{\text{AIC}}$ 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 that the variable selection of the implementation of the LSA-invariant model we performed is significantly unstable, particularly concerning their sensitivity to the choice of information criterion. A compelling example is observed with terms such as $\sin\left(2\pi \cdot 20\frac{t}{2400}\right)$ and $\cos\left(2\pi \cdot 25\frac{t}{2400}\right)$, which register very high VIPs for $\ell_2\text{LSA}_{\text{AIC}}$ but approach near-zero VIPs for $\ell_2\text{LSA}_{\text{BIC}}$. In contrast, the ℓ_1MTE and ℓ_2MTE 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 37, 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 37, the intercept term, along with $\sin\left(2\pi \cdot 5\frac{t}{2400}\right)$, and $\cos\left(2\pi \cdot 5\frac{t}{2400}\right)$, 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 $\sin\left(2\pi \cdot 10\frac{t}{2400}\right)$ and $\cos\left(2\pi \cdot 15\frac{t}{2400}\right)$, 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.

Taking into consideration the metric values of each model across locations L101 to L108 (refer to Tables 23, 24, 25, 26, and 27), one observes substantial variability in performance, which is measured by HL, MAE, RMSE, SMAPE, and SMdAPE, among the different model-averaging strategies. Notably, the AIC-based variant of the ℓ_1LSA model exhibits increases of up to 55.70%, 53.17%, 55.3%, 63.14%, and 73.95% across the aforementioned metrics, while its BIC-based counterpart shows increase at most 60.31%, 57.70%, 51.16%, 70.30%, and 96.73%. Similarly, the ℓ_2LSA model under the AIC criterion shows increases of up to 50.19%, 48.0%, 49.78%, 62.36%, and 84.06%, whereas under the BIC criterion, the increases reach 69.24%, 66.72%, 62.62%, 97.82%, and 116.53%. The ℓ_1MTE model, irrespective of the information criterion employed, demonstrates rises of up to 66.02%, 62.53%, 61.91%, 76.93%, and 94.72%, while its ℓ_2 variant shows increases up to 9.51%, 56.67%, 59.44%, 77.58%, and 104.14% across the same metrics. Among all locations, L101 and L102 were found to be the most amenable to accurate profile estimation, whereas L106 and L104 presented the greatest challenges.

By plotting the forecast of seasonal profiles using all the performed models against the actual data of the testing week across the given locations 38 we can see that the ℓ_1MTE, regardless the information criteria, surpasses substantially the rest of the models. Furthermore, the AIC-type ℓ_1MTE model coincides the BIC counterpart and we can see that it fairly captures the periodicity. The LSA-based models, though, have really poor forecasting performance as we can see in the plot below. The BIC type ℓ_1LSA has the most poor forecasting capability since the forecasted values went off the actual data plot. The intriguing part of this plot is that we have smooth daily profiles, namely, beside the noisy data, the periodicity can be described by a smooth curve and a set of few low harmonic frequencies that have been identified as significant.

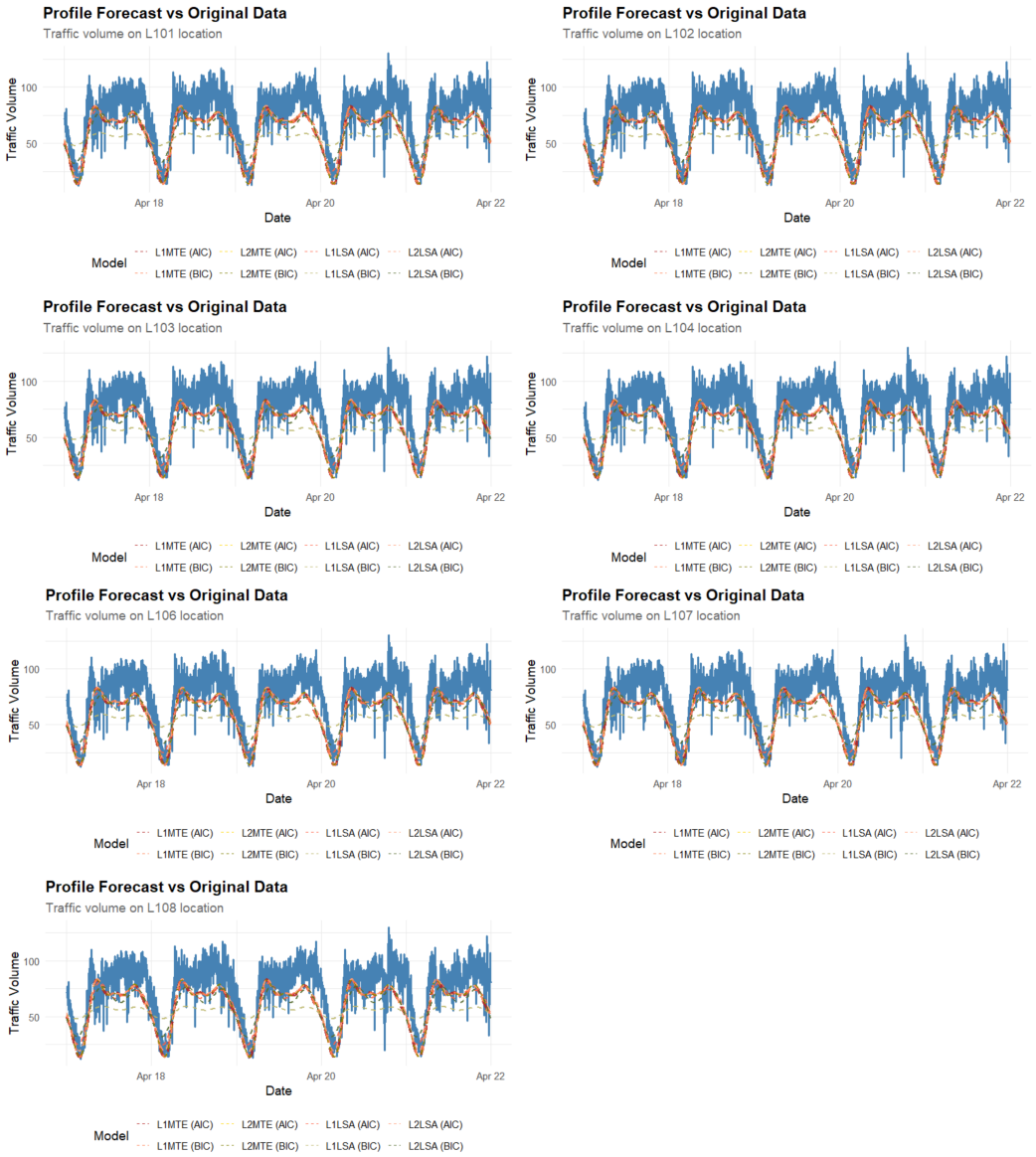

Figure 38: *Locationwise plots of forecasting results of each method plotted on the actual data. The curved line with firebrick is the AIC-type of the $\ell_1 MTE$, the coral is the BIC-type alternative, the gold and the olive is for $\ell_2 MTE$ with AIC and BIC minimization ,tomato and dark khaki corresponds to for $\ell_1 LSA$, and the last couple of for $\ell_2 LSA$ light salmon and dark olive green respectfully.*

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. The 37 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 (51, 52, 53, 54, 47, 48, 49, and 50) the intercept term, along with the low-frequency harmonic components $\sin\left(2\pi \cdot 5\frac{t}{2400}\right)$ and $\cos\left(2\pi \cdot 5\frac{t}{2400}\right)$, consistently exhibit uniformly exceptionally high VIPs, depicted by deep red in the heatmap. We present the magnified plots of the coefficients estimates of each model across the locations rather than the entire plots. This decision is made because the 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 $\sin\left(2\pi \cdot 10\frac{t}{2400}\right)$ and $\cos\left(2\pi \cdot 15\frac{t}{2400}\right)$, 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 criterion (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 stability.

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 which reveals distinct patterns concerning model efficacy and generalization capabilities. Upon examining the performance metrics for each model across locations L101 to L108 (see Tables 23, 24, 25, 26, and 27), notable variations emerge in the scores of HL, MAE, RMSE, SMAPE, and SMdAPE under different model averaging strategies. The AIC-based formulation of the ℓ_1LSA model exhibits increases of up to 63.66%, 61.03%, 70.26%, 54.02%, and 57.47%, respectively, across the aforementioned metrics, while its BIC-based counterpart demonstrates deviations reaching 67.66%, 64.94%, 59.06%, 46.04%, and 51.11%. Similarly, the ℓ_2LSA model records increases under the AIC criterion of up to 68.27%, 65.50%, 61.12%, 59.03%, and 55.39%, with the BIC criterion yielding respective rises of 67.38%, 64.67%, 52.95%, 42.69%, and 58.33%. The ℓ_1MTE model, consistent across information criteria, presents elevations of up to 67.42%, 64.55%, 77.20%, 56.87%, and 63.45% , whereas its ℓ_2 counterpart shows increases of 75.51%, 72.43%, 81.53%, 59.37%, and 57.61%. Among all assessed sites, L101 and L102 proved the most tractable for capturing accurate profiles, while L104 posed the greatest modeling challenges.

Therefore, the comprehensive analysis of these estimations reveals that, although a core subset of predictors (comprising the intercept and certain low-frequency harmonic features)is consistently identified as important, the choice between AIC and BIC minimization markedly affects the final selection of variables and, by extension, the overall model complexity, particularly within the LSA family. In contrast, the MTE models exhibit greater stability in variable selection, displaying less sensitivity to the choice of information criterion. The pronounced sparsity observed in the coefficient estimates collectively indicates that a parsimonious model is likely sufficient to capture the fundamental dynamics of traffic volume across the given locations. For a more nuanced understanding, a detailed inspection of the coefficient patterns (focusing on non-zero entries) or a direct numerical tabulation of these values would be highly informative.

While adaptive ℓ_1 Lasso models generally perform well across varied settings, the decision between

LSA and MTE frameworks appears to depend on the intrinsic characteristics of the data-generating process. In scenarios where datasets are explicitly constructed under post-hoc adaptive Lasso methodologies, both LSA and MTE approaches yield similarly high-fidelity outcomes. However, when applied to raw traffic volume data or in the context of model averaging, the MTE models consistently emerge as the more robust and dependable choice, offering superior generalization performance. This distinction underscores their practical effectiveness in real-world traffic forecasting, especially in situations where the underlying data structure may deviate from idealized Lasso assumptions.

4.1 Conclusions

Summary of Findings

This study presents an in-depth comparative analysis of penalized regression models, specifically, Maximum Tangent Likelihood Estimation (MTE) and Least Squares Approximation (LSA), under both ℓ_1 and ℓ_2 norms, optimized via the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The model family

$$\{\ell_x M_i \mid x \in \{1, 2\},\ M \in \{\text{MTE}, \text{LSA}\},\ i \in \{\text{AIC}, \text{BIC}\}\}$$

is systematically evaluated in the context of traffic volume forecasting. Variable selection is informed by Variable Inclusion Probabilities (VIPs), estimated through the Maximum Entropy Bootstrap (MEBoot) methodology.

The findings consistently highlight the superior performance and stability of MTE-based models over their LSA counterparts, particularly in terms of predictive accuracy, robustness, and error metric minimization. Notably, the ℓ_1MTE model exhibits a refined feature selection strategy, frequently identifying key low-frequency harmonic components as significant. In contrast, LSA models often omit these essential features due to over-shrinkage under stronger regularization, especially when BIC is employed, resulting in underfitting and compromised forecasting performance.

While the forecasting accuracy of ℓ_1 and ℓ_2 MTE variants is comparably strong, differences emerge in their treatment of less influential harmonics. Importantly, MTE models demonstrate stable feature inclusion regardless of whether AIC or BIC is used, whereas LSA models show notable variability in sparsity and selection, with results highly sensitive to the choice of criterion.

MTE models also generalize better to unseen data, particularly when evaluated through averaging over MEBoot samples. In these settings, MTE models yield consistently lower values for Huber Loss (HL), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) compared to LSA models. For instance, transitioning from MTE to LSA models results in average increases of 10.28% in HL, 9.97% in MAE, and 4.13% in RMSE, with peak degradations exceeding 22% in some cases.

The performance of the MTE-based models under the prism of the SMAPE indicates an interesting utility of the models. Based on the robustness of the SMAPE and SMdAPE key performance indicators, the low values of those metrics underscore the generalizability of this model. Its computational simplicity combined with its forecasting accuracy indicates that this model is useful for practitioners to use in the event of missing data.

The performance limitations of LSA models are partially attributable to their reliance on covariance matrix estimation techniques, such as that proposed by Pollard [15], which do not account for heteroskedasticity and serial correlation. To improve estimation robustness, future work could

adopt the methods of Galvao and Yoon [40] or Newey and West [22]. Furthermore, extending the Taylor expansion order for MTE and significantly increasing the number of MEBoot samples (e.g., to 999 as recommended by Vinod [26, 27]) may further enhance predictive reliability.

The variable selection results confirm that traffic flow dynamics are driven by a sparse set of low-frequency periodic components. Harmonics such as $\sin\left(2\pi \cdot 5\frac{t}{2400}\right)$ and $\cos\left(2\pi \cdot 5\frac{t}{2400}\right)$ consistently exhibit high VIPs, underscoring their foundational role in modeling diurnal traffic patterns. Other low-frequency components (e.g., $\sin\left(2\pi \cdot 10\frac{t}{2400}\right)$ and $\cos\left(2\pi \cdot 15\frac{t}{2400}\right)$) are also frequently selected, whereas higher-frequency features are often excluded, particularly under BIC minimization.

The empirical evaluation was conducted on traffic volume data collected from loop detectors at seven intersections along Alexandras Street, with a primary focus on location L102 due to its regular traffic flow characteristics. Results were benchmarked against the methodology of Rodrigues [4], which did not explore robust estimators. Motivated by this, we proposed an adaptive ℓ_1-LASSO framework informed by VIPs, though it showed limited forecasting improvements due to false positives in feature selection.

In conclusion, MTE-based models demonstrate consistently superior performance, robustness, and stability for traffic forecasting applications. They capture essential time-series dynamics with greater accuracy, particularly under complex conditions involving autocorrelation and heteroskedasticity. These findings establish MTE as a more reliable modeling approach in real-world traffic prediction contexts, where simplified linear assumptions often fall short. The R code scripts and the reproducible results of the code have been uploaded to the authors personal GitHub account in order to be accessible to everyone.

5 Appendix

5.1 Figures for all locations

5.1.1 Locationwise Coefficients Estimation plots of model averaging over 20 MEBoot regarding Methods

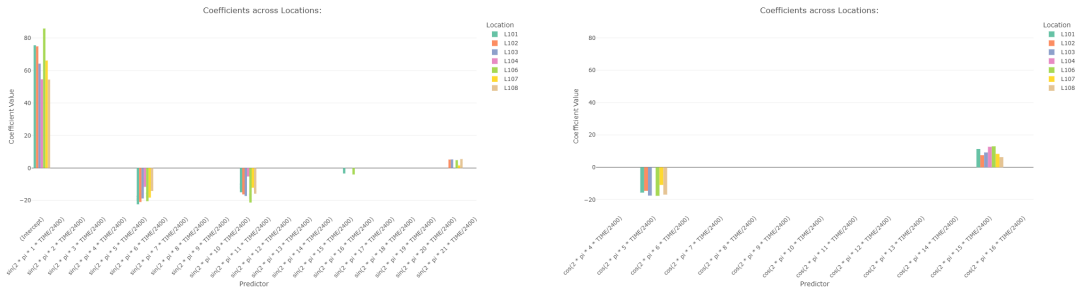

Figure 39: *All the coefficients estimations of* $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

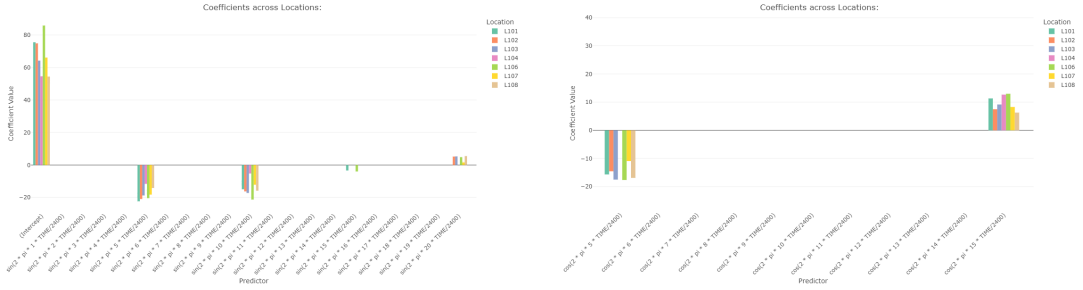

Figure 40: *All the coefficients estimations of* $\ell_1\mathrm{MTE}_{\mathrm{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

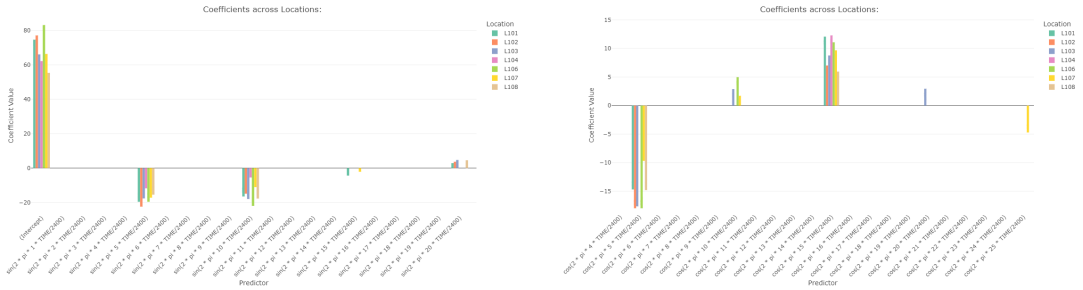

Figure 41: *All the oefficients estimations of* $\ell_2\mathrm{MTE}_{\mathrm{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

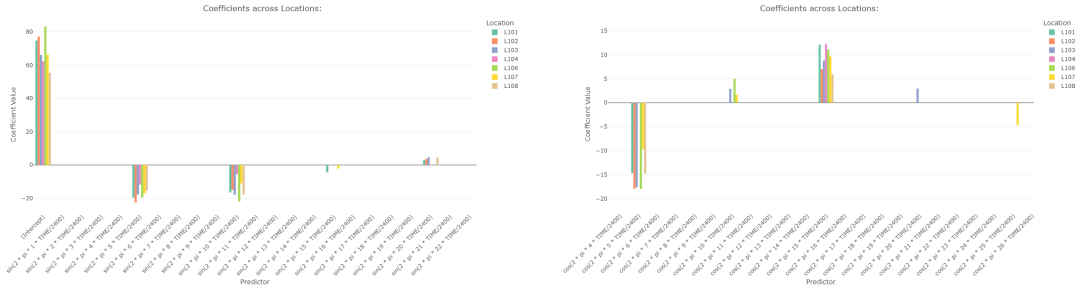

Figure 42: *All the coefficients estimations of* $\ell_2\mathrm{MTE}_{\mathrm{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

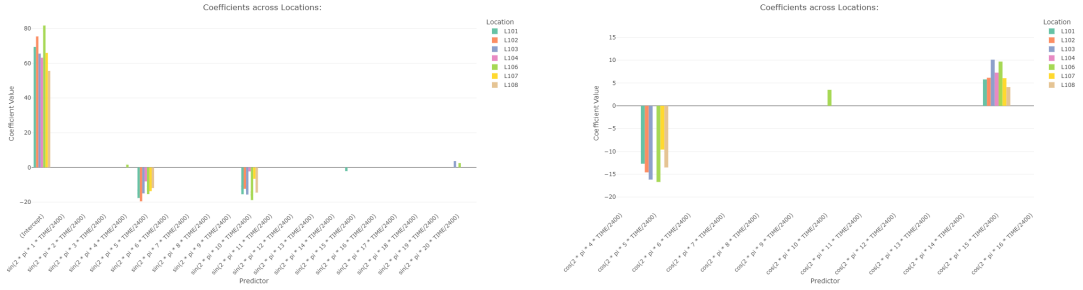

Figure 43: *All the coefficients estimations of* $\ell_1\mathrm{LSA}_{\mathrm{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

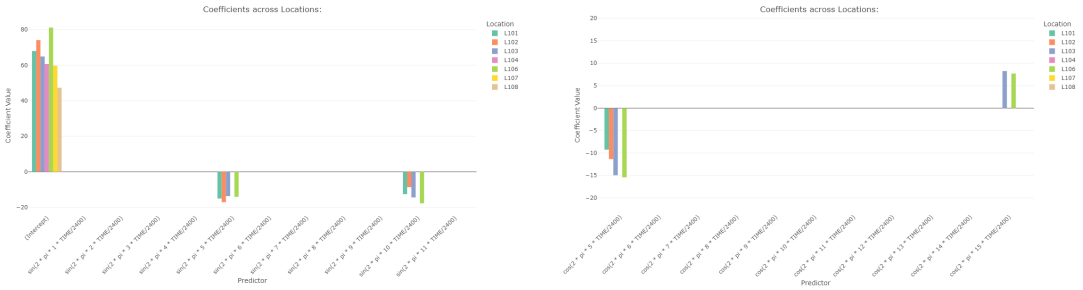

Figure 44: *All the coefficients estimations of* $\ell_1\mathrm{LSA}_{\mathrm{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

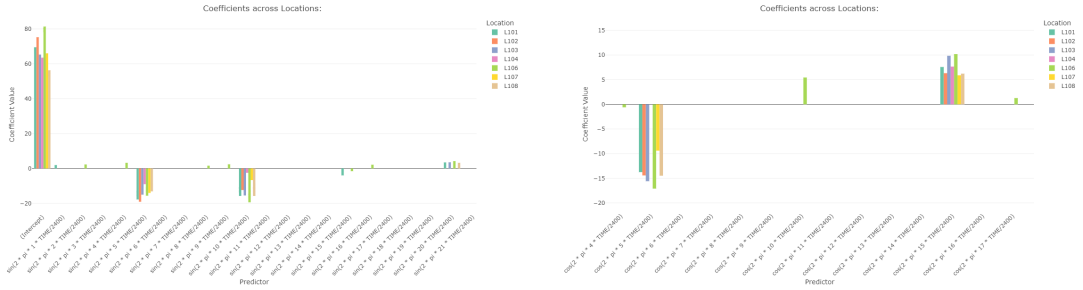

Figure 45: *All the coefficients estimations of* $\ell_2\mathrm{LSA}_{\mathrm{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

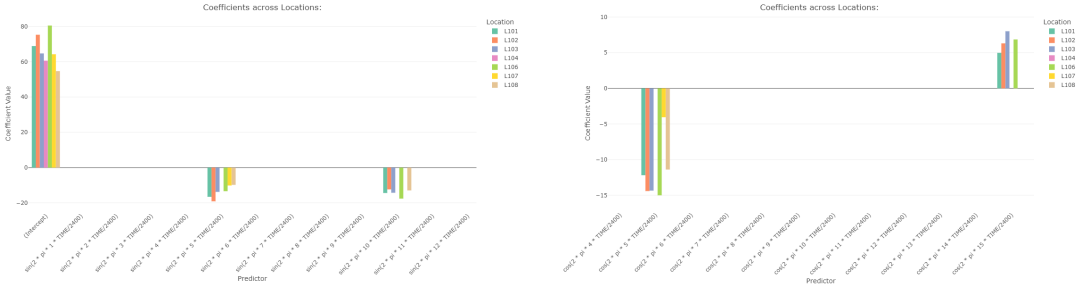

Figure 46: *All the oefficients estimations of* $\ell_2\mathrm{LSA}_{\mathrm{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

5.1.2 Locationwise Coefficients Estimation plots of Adaptive LAD LASSO regarding Methods

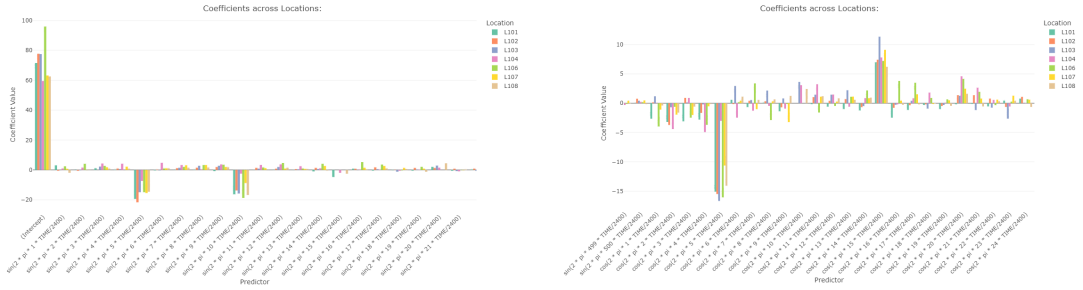

Figure 47: *Coefficients estimations of* $\ell_1\mathrm{MTE}_{\mathrm{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

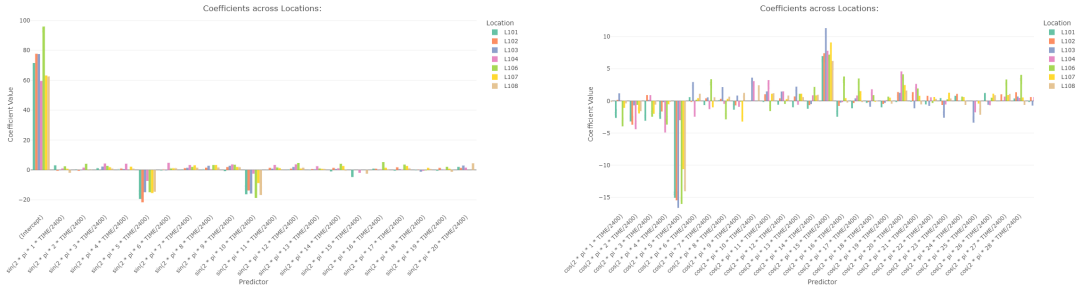

Figure 48: *Coefficients estimations of* $\ell_1\mathrm{MTE}_{\mathrm{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

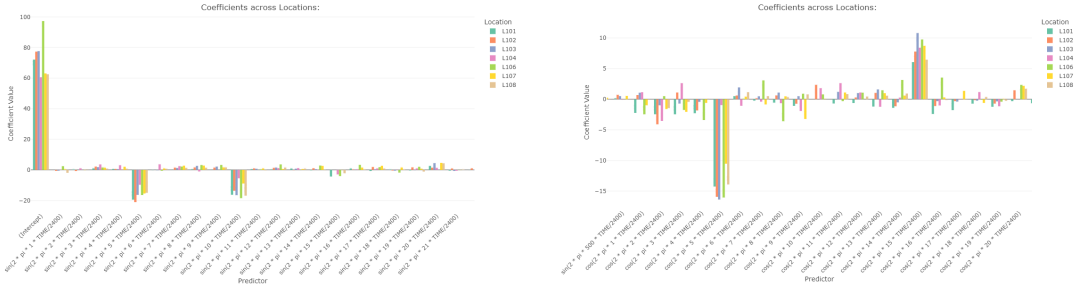

Figure 49: *Coefficients estimations of* $\ell_2\mathrm{MTE}_{\mathrm{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

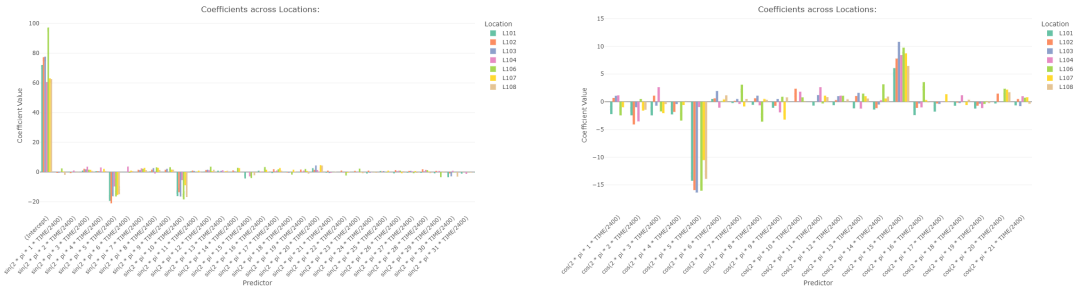

Figure 50: *Coefficients estimations of* $\ell_2\mathrm{MTE}_{\mathrm{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

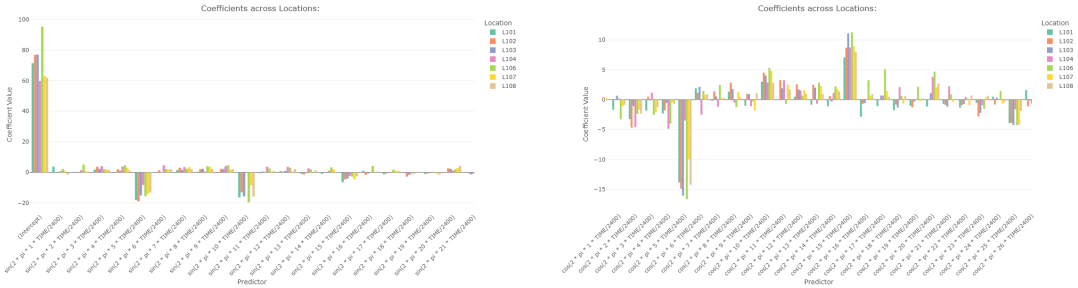

Figure 51: *Coefficients estimations of* $\ell_1\mathrm{LSA}_{\mathrm{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

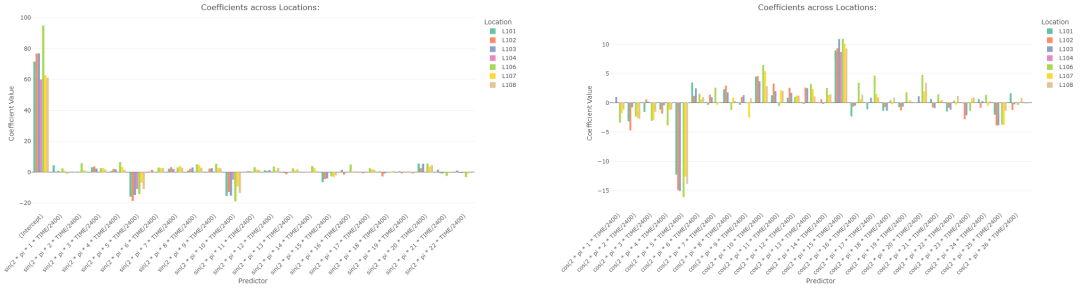

Figure 52: *Coefficients estimations of* ℓ_1LSA$_{\text{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

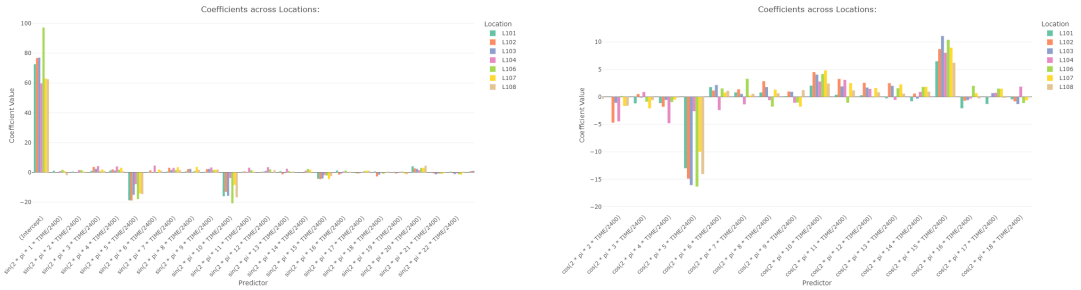

Figure 53: *Coefficients estimations of* ℓ_2LSA$_{\text{AIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

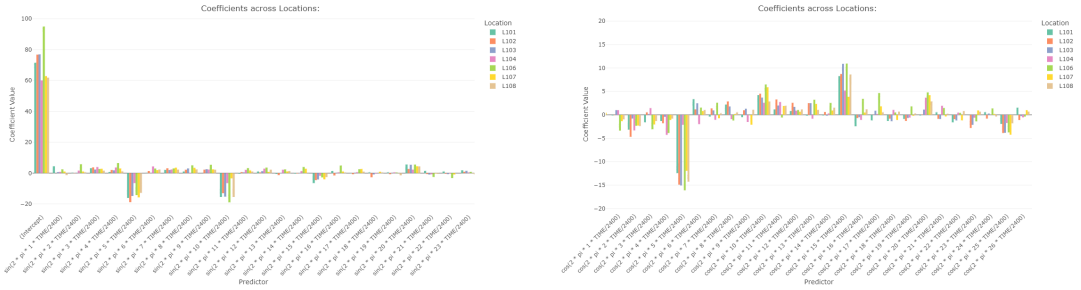

Figure 54: *Coefficients estimations of* ℓ_2LSA$_{\text{BIC}}$ *across the locations A - G (L101-L108) under model averaging over 20 MEBoot samples.*

5.2 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.

5.2.1 Variable Inclusion Probabilities (VIPs)

Features \ Methods	$\mathbf{L1MTE_{AIC}}$	$\mathbf{L1MTE_{BIC}}$	$\mathbf{L1LSA_{AIC}}$	$\mathbf{L1LSA_{BIC}}$	$\mathbf{L2MTE_{AIC}}$	$\mathbf{L2MTE_{BIC}}$	$\mathbf{L2LSA_{AIC}}$	$\mathbf{L2LSA_{BIC}}$
$(Intercept)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	0.95	0.00	0.90	0.90	1.00	0.00
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	0.40	0.40	0.05	0.00	0.60	0.60	1.00	0.00
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.25	0.25	0.00	0.00	0.15	0.15	0.00	0.00
$sin\left(2\pi \cdot 30\frac{t}{2400}\right)$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$cos\left(2\pi \cdot 35\frac{t}{2400}\right)$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$sin\left(2\pi \cdot 1\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.70	0.00
$cos\left(2\pi \cdot 2\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	0.65	0.00

Table 1: *Variable Inclusion Probabilities for location L101*

Features \ Methods	$\mathbf{L1MTE_{AIC}}$	$\mathbf{L1MTE_{BIC}}$	$\mathbf{L1LSA_{AIC}}$	$\mathbf{L1LSA_{BIC}}$	$\mathbf{L2MTE_{AIC}}$	$\mathbf{L2MTE_{BIC}}$	$\mathbf{L2LSA_{AIC}}$	$\mathbf{L2LSA_{BIC}}$
$(Intercept)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	1.00	1.00	0.00	0.00	0.80	0.80	0.00	0.00
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	0.35	0.35	0.00	0.00	0.35	0.35	0.00	0.00
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	0.35	0.35	0.00	0.00	0.20	0.20	0.00	0.00
$sin\left(2\pi \cdot 3\frac{t}{2400}\right)$	0.15	0.15	0.00	0.00	0.00	0.00	0.00	0.00
$sin\left(2\pi \cdot 45\frac{t}{2400}\right)$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$cos\left(2\pi \cdot 11\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 2: *Variable Inclusion Probabilities for location L102*

Methods / Features	$\mathbf{L1MTE_{AIC}}$	$\mathbf{L1MTE_{BIC}}$	$\mathbf{L1LSA_{AIC}}$	$\mathbf{L1LSA_{BIC}}$	$\mathbf{L2MTE_{AIC}}$	$\mathbf{L2MTE_{BIC}}$	$\mathbf{L2LSA_{AIC}}$	$\mathbf{L2LSA_{BIC}}$
$(Intercept)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	0.10	0.05	0.00	0.00	0.20	0.20	0.00	0.00
$sin\left(2\pi \cdot 40\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$cos\left(2\pi \cdot 20\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 3: *Variable Inclusion Probabilities for location L103*

Methods / Features	$\mathbf{L1MTE_{AIC}}$	$\mathbf{L1MTE_{BIC}}$	$\mathbf{L1LSA_{AIC}}$	$\mathbf{L1LSA_{BIC}}$	$\mathbf{L2MTE_{AIC}}$	$\mathbf{L2MTE_{BIC}}$	$\mathbf{L2LSA_{AIC}}$	$\mathbf{L2LSA_{BIC}}$
$(Intercept)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.25
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.25
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	1.00	1.00	0.05	0.00	0.75	0.75	1.00	0.00
$cos\left(2\pi \cdot 20\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.35	0.35	0.00	0.00
$sin\left(2\pi \cdot 9\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$sin\left(2\pi \cdot 11\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$cos\left(2\pi \cdot 4\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$cos\left(2\pi \cdot 128\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 4: *Variable Inclusion Probabilities for location L104*

Methods / Features	$\mathbf{L1MTE_{AIC}}$	$\mathbf{L1MTE_{BIC}}$	$\mathbf{L1LSA_{AIC}}$	$\mathbf{L1LSA_{BIC}}$	$\mathbf{L2MTE_{AIC}}$	$\mathbf{L2MTE_{BIC}}$	$\mathbf{L2LSA_{AIC}}$	$\mathbf{L2LSA_{BIC}}$
$(Intercept)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	0.20	0.20	1.00	0.00	0.15	0.15	1.00	0.00
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	0.10	0.10	0.00	0.00	0.10	0.10	1.00	0.00
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.05	0.05	1.00	0.00	0.40	0.40	1.00	0.00
$sin\left(2\pi \cdot 4\frac{t}{2400}\right)$	0.05	0.05	1.00	0.00	0.05	0.05	1.00	0.00
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	0.05	0.05	0.00	0.00	0.00	0.00	1.00	0.00
$sin\left(2\pi \cdot 2\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	1.00	0.00
$cos\left(2\pi \cdot 17\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	1.00	0.00
$sin\left(2\pi \cdot 8\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$sin\left(2\pi \cdot 9\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$sin\left(2\pi \cdot 16\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$cos\left(2\pi \cdot 4\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
$cos\left(2\pi \cdot 20\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00

Table 5: *Variable Inclusion Probabilities for location L106*

Features \ Methods	$\mathbf{L1MTE_{AIC}}$	$\mathbf{L1MTE_{BIC}}$	$\mathbf{L1LSA_{AIC}}$	$\mathbf{L1LSA_{BIC}}$	$\mathbf{L2MTE_{AIC}}$	$\mathbf{L2MTE_{BIC}}$	$\mathbf{L2LSA_{AIC}}$	$\mathbf{L2LSA_{BIC}}$
$(Intercept)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	0.10	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.20
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.20
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	0.65	0.65	0.00	0.00	0.40	0.40	0.00	0.00
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.15	0.15	0.00	0.00	0.30	0.30	0.00	0.00
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	0.10	0.10	0.00	0.00	0.35	0.35	0.00	0.00
$sin\left(2\pi \cdot 30\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 6: *Variable Inclusion Probabilities for location L107*

Features \ Methods	$\mathbf{L1MTE_{AIC}}$	$\mathbf{L1MTE_{BIC}}$	$\mathbf{L1LSA_{AIC}}$	$\mathbf{L1LSA_{BIC}}$	$\mathbf{L2MTE_{AIC}}$	$\mathbf{L2MTE_{BIC}}$	$\mathbf{L2LSA_{AIC}}$	$\mathbf{L2LSA_{BIC}}$
$(Intercept)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	0.25	1.00	1.00	1.00	1.00
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	1.00	1.00	1.00	0.25	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	1.00	1.00	1.00	0.25	1.00	1.00	1.00	1.00
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	1.00	1.00	0.00	0.00	1.00	1.00	1.00	0.00
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00
$cos\left(2\pi \cdot 97\frac{t}{2400}\right)$	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00

Table 7: *Variable Inclusion Probabilities for location L108*

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

5.2.2 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, 75.10 , 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, 70.10 , 79.95)
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	10.20 (7.91, 10.04 , 12.71)	10.20 (7.91, 10.04 , 12.71)	5.93 (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)
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	1.22 (0.00, 0.00 , 4.52)	1.22 (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, 2.42 , 5.08)	2.23 (0.00, 2.42 , 5.08)	2.93 (1.72, 3.38 , 3.59)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.74 (0.00, 0.00 , 4.18)	0.74 (0.00, 0.00 , 4.18)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.21 (0.00, 0.00 , 1.55)	0.21 (0.00, 0.00 , 1.55)	0.00 (0.00, 1.90 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 35\frac{t}{2400}\right)$	0.08 (0.00, 0.00 , 0.81)	0.08 (0.00, 0.00 , 0.81)	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)
$sin\left(2\pi \cdot 1\frac{t}{2400}\right)$	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)	0.08 (0.00, 0.00 , 0.85)	0.08 (0.00, 0.00 , 0.85)	1.32 (0.00, 0.00 , 2.03)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 2\frac{t}{2400}\right)$	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)	−0.69 (-1.13, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	−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)
$sin\left(2\pi \cdot 30\frac{t}{2400}\right)$	−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)
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	−4.95 (-7.28, -4.93 , -2.64)	−4.95 (-7.28, -4.93 , -2.64)	−2.24 (-3.11, -2.27 , -0.92)	0.00 (0.00, 0.00 , 0.00)	−3.61 (-6.53, -3.88 , 0.00)	−3.61 (-6.53, -3.88 , 0.00)	−3.53 (-4.07, -3.93 , -2.53)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	−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)	−13.57 (-14.11, -13.71 , -12.95)	−12.18 (-12.54, -12.17 , -11.93)
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	−15.66 (-17.66, -15.81 , -14.08)	−15.66 (-17.66, -15.81 , -14.08)	−15.53 (-16.30, -15.45 , -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, -15.74 , -15.00)	−14.42 (-14.84, -14.37 , -14.10)
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	−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, -20.16 , -17.50)	−20.15 (-22.10, -20.16 , -17.50)	−17.74 (-18.34, -17.78 , -17.11)	−16.65 (-17.11, -16.57 , -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)
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	7.03 (5.65, 6.92 , 8.94)	7.03 (5.65, 6.92 , 8.94)	5.84 (5.65, 5.82 , 6.13)	0.00 (0.00, 0.00 , 0.00)	7.28 (5.49, 6.90 , 10.31)	7.28 (5.49, 6.90 , 10.31)	6.22 (6.07, 6.20 , 6.45)	6.22 (6.07, 6.20 , 6.45)
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	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 , 4.86)	2.95 (0.00, 3.52 , 4.86)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 3\frac{t}{2400}\right)$	0.17 (0.00, 0.00 , 1.68)	0.17 (0.00, 0.00 , 1.68)	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)
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.09 (0.00, 0.00 , 0.93)	0.09 (0.00, 0.00 , 0.93)	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)
$sin\left(2\pi \cdot 45\frac{t}{2400}\right)$	0.07 (0.00, 0.00 , 0.72)	0.07 (0.00, 0.00 , 0.72)	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)
$cos\left(2\pi \cdot 11\frac{t}{2400}\right)$	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)	0.12 (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)
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	−0.87 (-3.59, 0.00 , 0.00)	−0.87 (-3.59, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	−0.91 (-4.30, 0.00 , 0.00)	−0.91 (-4.30, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	−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)
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	−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 , -8.26)	−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)
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	−17.29 (-18.94, -17.62 , -14.71)	−17.29 (-18.94, -17.62 , -14.71)	−14.44 (-14.64, -14.44 , -14.13)	−11.36 (-11.62, -11.41 , -10.98)	−17.37 (-19.36, -17.88 , -13.80)	−17.37 (-19.36, -17.88 , -13.80)	−14.35 (-14.58, -14.33 , -14.14)	−14.35 (-14.58, -14.33 , -14.14)
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	−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, 77.09 , 87.79)	76.62 (66.08, 76.52 , 87.29)	75.88 (66.43, 75.44 , 86.71)	75.88 (66.43, 75.44 , 86.71)	77.00 (66.51, 76.95 , 87.55)	76.45 (65.90, 76.40 , 87.05)
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	9.68 (8.70, 9.67 , 11.47)	9.65 (8.62, 9.67 , 11.47)	10.19 (10.00, 10.17 , 10.37)	8.28 (8.08, 8.30 , 8.51)	9.48 (8.48, 9.42 , 11.00)	9.48 (8.48, 9.42 , 11.00)	9.93 (9.75, 9.91 , 10.17)	8.03 (7.88, 8.02 , 8.22)
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	5.05 (4.30, 5.12 , 5.73)	5.08 (4.30, 5.17 , 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)	3.63 (3.46, 3.61 , 3.78)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 20\frac{t}{2400}\right)$	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 , 1.55)	0.15 (0.00, 0.00 , 1.55)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	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)
$sin\left(2\pi \cdot 40\frac{t}{2400}\right)$	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)
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	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)
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	−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 (-2.39, 0.00 , 0.00)	−0.37 (-2.39, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	−17.85 (-19.27, -17.39 , -16.24)	−17.90 (-19.27, -17.39 , -16.68)	−15.74 (-16.81, -15.72 , -15.52)	−14.44 (-15.64, -14.42 , -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)
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	−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 , -13.52)	−17.23 (-19.36, -17.29 , -16.50)	−17.23 (-19.36, -17.29 , -16.50)	−15.80 (-15.50, -15.78 , -14.97)	−14.52 (-14.16, -14.49 , -13.63)
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	−18.15 (-19.65, -18.57 , -16.60)	−18.24 (-19.65, -18.57 , -16.75)	−15.09 (-16.13, -15.02 , -16.14)	−13.74 (-14.83, -13.68 , -14.81)	−18.33 (-19.49, -18.57 , -14.59)	−18.33 (-19.49, -18.57 , -14.59)	−15.16 (-15.97, -15.11 , -15.60)	−13.84 (-14.66, -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 (49.39, 57.13 , 64.58)	58.64 (51.26, 58.50 , 66.36)	58.64 (51.26, 58.50 , 66.36)	60.03 (52.61, 60.01 , 67.37)	57.63 (49.20, 57.06 , 66.98)
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	12.66 (12.66, 12.66 , 12.66)	12.66 (12.66, 12.66 , 12.66)	6.18 (5.95, 6.10 , 6.87)	0.00 (0.00, 0.00 , 0.00)	10.85 (8.15, 10.57 , 12.91)	10.85 (8.15, 10.57 , 12.91)	7.62 (7.50, 7.60 , 7.85)	1.60 (0.00, 0.00 , 6.44)
$cos\left(2\pi \cdot 20\frac{t}{2400}\right)$	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 , 4.60)	1.32 (0.00, 0.00 , 4.60)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 11\frac{t}{2400}\right)$	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 , 1.60)	0.15 (0.00, 0.00 , 1.60)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 9\frac{t}{2400}\right)$	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)
$cos\left(2\pi \cdot 128\frac{t}{2400}\right)$	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)
$cos\left(2\pi \cdot 4\frac{t}{2400}\right)$	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)
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	−5.23 (-5.23, -5.23 , -5.23)	−5.23 (-5.23, -5.23 , -5.23)	−0.11 (-1.20, 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)
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	−11.61 (-11.61, -11.61 , -11.61)	−11.61 (-11.61, -11.61 , -11.61)	−7.02 (-7.67, -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)

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 (81.95, 94.40 , 106.85)
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	12.36 (9.28, 13.06 , 14.71)	12.36 (9.28, 13.06 , 14.71)	9.67 (9.38, 9.69 , 9.90)	7.73 (7.43, 7.75 , 8.00)	13.04 (10.27, 13.34 , 14.57)	13.04 (10.27, 13.34 , 14.57)	10.28 (10.03, 10.28 , 10.49)	6.95 (6.75, 6.93 , 7.12)
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	1.00 (0.00, 0.00 , 5.11)	1.00 (0.00, 0.00 , 5.11)	2.48 (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)
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.29 (0.00, 0.00 , 3.04)	0.29 (0.00, 0.00 , 3.04)	3.44 (2.25, 2.47 , 3.68)	0.00 (0.00, 0.00 , 0.00)	1.91 (0.00, 0.00 , 5.96)	1.91 (0.00, 0.00 , 5.96)	5.46 (4.06, 4.23 , 5.58)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 4\frac{t}{2400}\right)$	0.21 (0.00, 0.00 , 2.16)	0.21 (0.00, 0.00 , 2.16)	1.49 (1.37, 1.50 , 1.64)	0.00 (0.00, 0.00 , 0.00)	0.19 (0.00, 0.00 , 1.98)	0.19 (0.00, 0.00 , 1.98)	3.23 (3.17, 3.23 , 3.31)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 2\frac{t}{2400}\right)$	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 (0.00, 0.00 , 1.77)	0.17 (0.00, 0.00 , 1.77)	2.29 (2.31, 2.34 , 2.36)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 17\frac{t}{2400}\right)$	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 , 1.58)	0.15 (0.00, 0.00 , 1.58)	1.28 (2.24, 2.30 , 1.33)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 9\frac{t}{2400}\right)$	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)	2.35 (2.09, 2.13 , 2.42)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 16\frac{t}{2400}\right)$	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)	2.13 (1.52, 1.57 , 2.18)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 8\frac{t}{2400}\right)$	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)	1.57 (1.20, 1.28 , 1.61)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 20\frac{t}{2400}\right)$	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)	0.58 (0.35, 0.59 , 0.80)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 4\frac{t}{2400}\right)$	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)	−0.60 (-0.67, -0.60 , -0.52)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	−0.20 (-2.09, 0.00 , 0.00)	−0.20 (-2.09, 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)	−1.60 (-1.67, -0.87 , -0.75)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	−0.48 (-4.80, 0.00 , 0.00)	−0.48 (-4.80, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	−0.50 (-5.05, 0.00 , 0.00)	−0.50 (-5.05, 0.00 , 0.00)	−0.87 (-0.96, -1.60 , -1.52)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	−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 , -13.16)	−16.79 (-20.82, -16.56 , -13.16)	−17.18 (-17.46, -17.10 , -16.99)	−15.09 (-15.35, -15.01 , -14.94)
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	−19.86 (-23.04, -19.50 , -16.55)	−19.86 (-23.04, -19.50 , -16.55)	−15.43 (-15.69, -15.38 , -15.22)	−14.09 (-14.40, -14.02 , -13.94)	−18.84 (-21.73, -19.25 , -16.53)	−18.84 (-21.73, -19.25 , -16.53)	−15.73 (-16.02, -15.68 , -15.54)	−13.46 (-13.71, -13.41 , -13.32)
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	−23.71 (-26.10, -23.77 , -20.87)	−23.71 (-26.10, -23.77 , -20.87)	−18.93 (-19.35, -18.86 , -18.72)	−17.82 (-18.28, -17.75 , -17.61)	−22.69 (-24.77, -22.58 , -19.36)	−22.69 (-24.77, -22.58 , -19.36)	−19.53 (-19.92, -19.45 , -19.33)	−17.69 (-18.04, -17.61 , -17.52)

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, 63.52 , 72.52)	63.84 (55.23, 63.52 , 72.52)	62.66 (54.24, 62.72 , 70.97)	56.40 (46.80, 56.29 , 67.78)	63.22 (54.85, 63.43 , 71.28)	63.22 (54.85, 63.43 , 71.28)	62.78 (54.39, 62.84 , 71.12)	61.24 (52.25, 60.98 , 71.12)
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	8.58 (6.39, 8.50 , 10.66)	8.58 (6.39, 8.50 , 10.66)	6.02 (5.88, 6.02 , 6.16)	0.00 (0.00, 0.00 , 0.00)	7.67 (6.04, 7.09 , 10.04)	7.67 (6.04, 7.09 , 10.04)	5.86 (5.77, 5.83 , 6.03)	1.19 (0.00, 0.00 , 6.03)
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	0.50 (0.00, 0.00 , 4.26)	0.50 (0.00, 0.00 , 4.26)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.70 (0.00, 0.00 , 3.66)	0.70 (0.00, 0.00 , 3.66)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	0.37 (0.00, 0.00 , 2.35)	0.37 (0.00, 0.00 , 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)
$sin\left(2\pi \cdot 30\frac{t}{2400}\right)$	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)
$sin\left(2\pi \cdot 15\frac{t}{2400}\right)$	−0.27 (-2.72, 0.00 , 0.00)	−0.27 (-2.72, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	−0.95 (-3.23, 0.00 , 0.00)	−0.95 (-3.23, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 25\frac{t}{2400}\right)$	−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 (-5.37, 0.00 , 0.00)	−1.73 (-5.37, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	−10.57 (-12.37, -10.34 , -9.62)	−10.57 (-12.37, -10.34 , -9.62)	−9.53 (-9.72, -9.53 , -9.39)	0.00 (0.00, 0.00 , 0.00)	−10.19 (-10.98, -10.43 , -9.20)	−10.19 (-10.98, -10.43 , -9.20)	−9.36 (-9.54, -9.35 , -9.26)	−5.10 (-9.54, -4.02 , -3.91)
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	−11.68 (-13.72, -11.67 , -9.99)	−11.68 (-13.72, -11.67 , -9.99)	−6.57 (-6.76, -6.56 , -6.43)	0.00 (0.00, 0.00 , 0.00)	−11.08 (-11.95, -11.18 , -9.23)	−11.08 (-11.95, -11.18 , -9.23)	−6.67 (-6.80, -6.66 , -6.58)	−1.35 (-6.80, 0.00 , 0.00)
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	−19.14 (-21.53, -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, -10.06 , -9.98)

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, 61.75 , 71.24)	60.23 (50.57, 60.21 , 69.77)
$cos\left(2\pi \cdot 15\frac{t}{2400}\right)$	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 , 6.66)	0.00 (0.00, 0.00 , 0.00)
$sin\left(2\pi \cdot 20\frac{t}{2400}\right)$	4.82 (3.98, 4.54 , 6.15)	4.82 (3.98, 4.54 , 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)
$cos\left(2\pi \cdot 10\frac{t}{2400}\right)$	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)	0.08 (0.00, 0.00 , 0.79)	0.08 (0.00, 0.00 , 0.79)	0.00 (0.00, 0.00 , 0.00)	0.00 (0.00, 0.00 , 0.00)
$cos\left(2\pi \cdot 97\frac{t}{2400}\right)$	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)
$sin\left(2\pi \cdot 5\frac{t}{2400}\right)$	−14.51 (-15.22, -14.50 , -14.02)	−14.51 (-15.22, -14.50 , -14.02)	−12.16 (-12.48, -12.19 , -11.85)	−2.46 (-9.94, 0.00 , 0.00)	−14.88 (-15.74, -14.84 , -14.23)	−14.88 (-15.74, -14.84 , -14.23)	−13.35 (-13.76, -13.25 , -13.14)	−9.91 (-10.13, -9.90 , -9.67)
$cos\left(2\pi \cdot 5\frac{t}{2400}\right)$	−16.08 (-17.44, -15.83 , -14.52)	−16.08 (-17.44, -15.83 , -14.52)	−13.60 (-15.07, -13.52 , -13.31)	−2.90 (-12.88, 0.00 , 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)
$sin\left(2\pi \cdot 10\frac{t}{2400}\right)$	−16.40 (-17.85, -16.54 , -15.29)	−16.40 (-17.85, -16.54 , -15.29)	−14.71 (-13.99, -14.70 , -14.40)	−3.20 (-11.68, 0.00 , 0.00)	−16.94 (-17.64, -17.04 , -15.36)	−16.94 (-17.64, -17.04 , -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.*

5.2.3 Error Metrics

Methods			Location: L101				
			HL	MAE	RMSE	SMAPE	SMdAPE
AIC	L1LSA	MEB	9.91	10.40	13.22	15.49	12.58
		Actual	9.95	10.44	**17.77**	15.63	12.68
		ADL1LASSO	**10.76**	**11.24**	15.98	**16.79**	**12.11**
	L2LSA	MEB	10.26	10.75	13.56	16.34	12.71
		Actual	10.24	10.73	18.30	16.40	12.57
		ADL1LASSO	11.98	12.47	15.96	19.34	13.27
	L1MTE	MEB	**8.77**	**9.26**	**12.34**	**13.87**	**10.23**
		Actual	**9.04**	**9.53**	18.66	**14.14**	**10.33**
		ADL1LASSO	11.02	11.51	16.17	17.68	12.12
	L2MTE	MEB	8.89	9.37	12.45	14.14	10.38
		Actual	9.22	9.71	18.26	14.98	10.75
		ADL1LASSO	11.88	12.37	16.10	19.57	13.31
BIC	L1LSA	MEB	10.89	11.38	14.25	17.29	14.02
		Actual	10.91	11.40	18.77	17.41	13.94
		ADL1LASSO	11.99	12.48	**15.41**	18.78	12.76
	L2LSA	MEB	12.81	13.31	16.16	20.81	16.48
		Actual	12.77	13.26	20.10	20.82	16.29
		ADL1LASSO	11.99	12.48	15.52	18.84	12.58
	L1MTE	MEB	**8.77**	**9.26**	**12.34**	**13.87**	**10.23**
		Actual	**9.04**	**9.53**	18.66	**14.14**	**10.33**
		ADL1LASSO	11.02	11.51	16.17	17.68	12.12
	L2MTE	MEB	8.89	9.37	12.45	14.14	10.38
		Actual	9.22	9.71	18.26	14.98	10.75
		ADL1LASSO	11.88	12.37	16.10	19.57	13.31

Table 16: *Comparison of five performance metrics (HL, MAE, RMSE, 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 selectoin 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				
			HL	**MAE**	**RMSE**	**SMAPE**	**SMdAPE**
AIC	**L1LSA**	**MEB**	10.48	10.97	14.21	15.96	11.94
		Actual	10.51	11.00	16.17	16.05	11.98
		ADL1LASSO	11.69	12.18	**13.57**	17.16	12.62
	L2LSA	**MEB**	10.50	10.99	14.22	16.02	11.86
		Actual	10.49	10.98	16.19	16.07	11.86
		ADL1LASSO	11.69	12.18	**13.57**	17.16	12.62
	L1MTE	**MEB**	**9.79**	**10.28**	**13.72**	14.36	**10.96**
		Actual	9.96	10.45	**16.10**	14.67	11.18
		ADL1LASSO	11.23	11.72	14.07	**16.46**	12.44
	L2MTE	**MEB**	9.81	10.30	13.75	**14.31**	11.04
		Actual	**9.93**	**10.41**	16.13	**14.46**	**11.12**
		ADL1LASSO	**11.19**	**11.68**	14.60	16.59	**12.29**
BIC	**L1LSA**	**MEB**	10.48	10.97	14.21	15.96	11.94
		Actual	10.51	11.00	16.17	16.05	11.98
		ADL1LASSO	11.69	12.18	**13.57**	17.16	12.62
	L2LSA	**MEB**	12.58	13.07	16.24	19.75	14.40
		Actual	12.58	13.07	17.94	19.80	14.34
		ADL1LASSO	11.71	12.20	13.62	17.24	12.55
	L1MTE	**MEB**	**9.79**	**10.28**	**13.72**	14.36	**10.96**
		Actual	9.96	10.45	**16.10**	14.67	11.18
		ADL1LASSO	11.23	11.72	14.07	**16.46**	12.44
	L2MTE	**MEB**	9.81	10.30	13.75	**14.31**	11.04
		Actual	**9.93**	**10.41**	16.13	**14.46**	**11.12**
		ADL1LASSO	**11.19**	**11.68**	14.60	16.59	**12.29**

Table 17: *Comparison of five performance metrics (HL, MAE, RMSE, 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 selectoin 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				
			HL	**MAE**	**RMSE**	**SMAPE**	**SMdAPE**
AIC	**L1LSA**	**MEB**	13.77	14.26	**19.24**	20.66	14.24
		Actual	13.82	14.31	16.71	20.82	14.21
		ADL1LASSO	15.99	16.49	14.35	22.99	16.66
	L2LSA	**MEB**	13.83	14.32	19.35	20.67	14.27
		Actual	13.87	14.36	**16.68**	20.79	14.32
		ADL1LASSO	15.99	16.49	14.35	22.99	16.66
	L1MTE	**MEB**	**13.49**	**13.98**	19.27	**19.71**	**13.96**
		Actual	**13.51**	**14.00**	16.86	**19.76**	13.96
		ADL1LASSO	**15.67**	**16.16**	14.76	**22.40**	**16.26**
	L2MTE	**MEB**	13.62	14.11	19.45	19.88	14.10
		Actual	13.86	14.36	16.85	20.13	**13.95**
		ADL1LASSO	16.36	16.86	14.44	23.36	17.17
BIC	**L1LSA**	**MEB**	14.34	14.83	19.49	22.20	15.00
		Actual	14.41	14.90	17.43	22.37	14.96
		ADL1LASSO	16.05	16.54	**13.97**	23.11	16.72
	L2LSA	**MEB**	14.38	14.87	19.57	22.19	15.09
		Actual	14.44	14.94	17.38	22.37	15.12
		ADL1LASSO	16.05	16.54	**13.97**	23.11	16.72
	L1MTE	**MEB**	**13.49**	**13.98**	19.27	**19.71**	**13.96**
		Actual	**13.51**	**14.00**	16.86	**19.76**	13.96
		ADL1LASSO	**15.67**	**16.16**	14.76	**22.40**	**16.26**
	L2MTE	**MEB**	13.63	14.12	19.47	19.89	14.02
		Actual	13.86	14.36	16.85	20.13	**13.95**
		ADL1LASSO	16.36	16.85	14.43	23.35	17.21

Table 18: *Comparison of five performance metrics (HL, MAE, RMSE, 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 selectoin 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				
			HL	**MAE**	**RMSE**	**SMAPE**	**SMdAPE**
AIC	**L1LSA**	**MEB**	13.41	13.91	**17.16**	25.27	20.77
		Actual	13.47	13.96	**17.71**	25.39	20.88
		ADL1LASSO	13.81	14.30	**14.17**	25.86	**19.07**
	L2LSA	**MEB**	13.74	14.23	17.36	26.01	21.83
		Actual	13.81	14.30	18.21	26.14	22.01
		ADL1LASSO	14.39	14.88	14.68	27.29	19.61
	L1MTE	**MEB**	**13.29**	**13.78**	17.59	**24.54**	**19.92**
		Actual	13.48	13.97	17.91	24.79	20.02
		ADL1LASSO	14.22	14.71	14.76	25.82	19.81
	L2MTE	**MEB**	13.32	13.81	17.60	25.11	21.19
		Actual	**13.21**	**13.70**	17.84	**24.45**	**19.28**
		ADL1LASSO	13.96	14.45	14.48	26.44	19.37
BIC	**L1LSA**	**MEB**	14.33	14.82	17.97	27.18	23.49
		Actual	14.81	15.31	20.59	27.98	23.97
		ADL1LASSO	13.73	14.22	14.66	25.06	19.07
	L2LSA	**MEB**	14.80	15.29	18.55	27.95	23.71
		Actual	14.81	15.31	20.58	27.98	23.86
		ADL1LASSO	**13.18**	**13.67**	17.56	**24.60**	19.87
	L1MTE	**MEB**	**13.29**	**13.78**	17.59	**24.54**	**19.92**
		Actual	13.48	13.97	17.91	24.79	20.02
		ADL1LASSO	14.22	14.71	14.76	25.82	19.81
	L2MTE	**MEB**	13.32	13.81	17.60	25.11	21.19
		Actual	**13.21**	**13.70**	17.84	**24.45**	**19.28**
		ADL1LASSO	13.96	14.45	14.48	26.44	19.37

Table 19: *Comparison of five performance metrics (HL, MAE, RMSE, 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 selectoin 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				
			HL	**MAE**	**RMSE**	**SMAPE**	**SMdAPE**
AIC	**L1LSA**	**MEB**	15.43	15.93	20.53	17.51	13.87
		Actual	15.43	15.93	**23.20**	17.55	13.90
		ADL1LASSO	**17.61**	**18.10**	21.30	**19.97**	**14.19**
	L2LSA	**MEB**	15.41	15.91	20.31	17.78	13.99
		Actual	15.38	15.88	24.34	17.79	13.89
		ADL1LASSO	18.93	19.43	20.06	21.64	15.18
	L1MTE	**MEB**	14.56	15.05	19.98	16.69	12.21
		Actual	14.86	15.35	25.60	17.05	12.26
		ADL1LASSO	18.45	18.94	22.54	20.97	14.75
	L2MTE	**MEB**	**14.18**	**14.68**	**19.85**	**16.38**	**11.50**
		Actual	**14.10**	**14.59**	25.55	**16.37**	**11.54**
		ADL1LASSO	19.64	20.14	22.71	22.77	15.16
BIC	**L1LSA**	**MEB**	16.80	17.30	21.48	19.78	15.46
		Actual	16.80	17.30	25.38	19.83	15.39
		ADL1LASSO	19.60	20.09	**19.93**	22.56	15.73
	L2LSA	**MEB**	16.38	16.88	21.13	19.31	14.81
		Actual	16.39	16.88	25.25	19.37	14.73
		ADL1LASSO	19.60	20.09	**19.93**	22.56	15.73
	L1MTE	**MEB**	14.56	15.05	19.98	16.69	12.21
		Actual	14.86	15.35	25.60	17.05	12.26
		ADL1LASSO	18.45	18.94	22.54	20.97	14.75
	L2MTE	**MEB**	**14.18**	**14.68**	**19.85**	**16.38**	**11.50**
		Actual	**14.10**	**14.59**	25.55	**16.37**	**11.54**
		ADL1LASSO	19.64	20.14	22.71	22.77	15.16

Table 20: *Comparison of five performance metrics (HL, MAE, RMSE, 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 selectoin 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				
			HL	**MAE**	**RMSE**	**SMAPE**	**SMdAPE**
AIC	**L1LSA**	**MEB**	11.45	11.94	15.25	21.71	16.06
		Actual	11.48	11.97	15.96	21.82	16.01
		ADL1LASSO	**11.25**	**11.74**	**12.96**	**20.24**	**14.86**
	L2LSA	**MEB**	11.52	12.01	15.31	21.82	16.29
		Actual	11.55	12.04	15.96	21.92	16.25
		ADL1LASSO	**11.25**	**11.74**	**12.96**	**20.24**	**14.86**
	L1MTE	**MEB**	**10.26**	**10.74**	**14.45**	18.73	13.46
		Actual	10.67	11.16	**15.83**	19.87	13.96
		ADL1LASSO	11.77	12.26	14.02	21.15	15.41
	L2MTE	**MEB**	**10.26**	**10.74**	14.63	**18.41**	**13.12**
		Actual	**10.33**	**10.82**	15.93	**18.50**	**13.22**
		ADL1LASSO	12.10	12.59	14.21	21.90	15.74
BIC	**L1LSA**	**MEB**	13.78	14.27	17.45	26.34	19.42
		Actual	14.52	15.01	19.07	27.61	20.13
		ADL1LASSO	12.48	12.97	14.17	22.48	16.48
	L2LSA	**MEB**	19.06	19.55	22.61	34.84	30.52
		Actual	19.24	19.74	23.37	35.13	30.69
		ADL1LASSO	12.97	13.46	14.60	23.21	16.54
	L1MTE	**MEB**	**10.26**	**10.74**	**14.45**	18.73	13.46
		Actual	10.67	11.16	**15.83**	19.87	13.96
		ADL1LASSO	11.77	12.26	14.02	21.15	15.41
	L2MTE	**MEB**	**10.26**	**10.74**	14.63	**18.41**	**13.12**
		Actual	**10.33**	**10.82**	15.93	**18.50**	**13.22**
		ADL1LASSO	12.10	12.59	14.21	21.90	15.74

Table 21: *Comparison of five performance metrics (HL, MAE, RMSE, 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 selectoin 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				
			HL	MAE	RMSE	SMAPE	SMdAPE
AIC	L1LSA	MEB	11.34	11.82	**16.61**	19.41	14.56
		Actual	11.38	11.87	**15.02**	19.64	14.62
		ADL1LASSO	12.04	12.53	12.51	19.96	15.56
	L2LSA	MEB	12.38	12.87	17.56	22.05	16.49
		Actual	12.45	12.94	15.70	22.33	16.59
		ADL1LASSO	12.34	12.83	**12.45**	20.72	15.48
	L1MTE	MEB	**11.24**	**11.73**	16.73	**18.77**	**14.35**
		Actual	11.31	11.80	15.08	**18.86**	14.52
		ADL1LASSO	**11.97**	**12.46**	12.72	**19.86**	**15.43**
	L2MTE	MEB	11.36	11.85	16.88	18.94	14.67
		Actual	**11.31**	**11.80**	15.08	18.87	**14.38**
		ADL1LASSO	12.04	12.53	12.51	19.96	15.56
BIC	L1LSA	MEB	13.94	14.43	18.99	25.62	19.01
		Actual	13.97	14.46	16.98	25.77	18.96
		ADL1LASSO	12.47	12.96	12.53	21.02	15.48
	L2LSA	MEB	21.29	21.79	26.28	38.20	31.18
		Actual	24.07	24.57	26.00	42.87	36.40
		ADL1LASSO	13.04	13.53	13.03	22.34	15.89
	L1MTE	MEB	**11.24**	**11.73**	16.73	**18.77**	**14.35**
		Actual	11.31	11.80	15.08	**18.86**	14.52
		ADL1LASSO	**11.97**	**12.46**	12.72	**19.86**	**15.43**
	L2MTE	MEB	11.36	11.85	16.88	18.94	14.67
		Actual	**11.31**	**11.80**	15.08	18.87	**14.38**
		ADL1LASSO	12.04	12.53	12.51	19.96	15.56

Table 22: *Comparison of five performance metrics (HL, MAE, RMSE, 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 selectoin algorithm `ADL1LASSO` models under AIC and BIC minimization criteria, with bold values indicating the best-performing metrics within each model for location L108*

5.2.4 Location-wise Errors regarding Metrics Among Models

Methods / Locations	Metric: Huber Loss																							
	ℓ_1LSA$_{AIC}$			ℓ_2LSA$_{AIC}$			ℓ_1MTE$_{AIC}$			ℓ_2MTE$_{AIC}$			ℓ_1LSA$_{BIC}$			ℓ_2LSA$_{BIC}$			ℓ_1MTE$_{BIC}$			ℓ_2MTE$_{BIC}$		
	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO
L101	9.91	9.95	10.76	10.26	10.24	11.98	**8.77**	9.04	11.02	8.89	9.22	11.88	10.89	10.91	11.99	12.81	12.77	11.99	**8.77**	9.04	11.02	8.89	9.22	11.88
L102	10.48	10.51	11.69	10.50	10.49	11.69	**9.79**	9.96	11.23	9.81	9.93	11.19	10.48	10.51	11.69	12.58	12.58	11.71	**9.79**	9.96	11.23	9.81	9.93	11.19
L103	13.77	13.82	15.99	13.83	13.87	15.99	**13.49**	13.51	15.67	13.62	13.86	16.36	14.34	14.41	16.05	14.38	14.44	16.05	**13.49**	13.51	15.67	13.63	13.86	16.36
L104	13.41	13.47	13.81	13.74	13.81	14.39	13.29	13.48	14.22	13.32	13.21	13.96	14.33	14.81	13.73	14.80	14.81	**13.18**	13.29	13.48	14.22	13.32	13.21	13.96
L106	15.43	15.43	17.61	15.41	15.38	18.93	14.56	14.86	18.45	14.18	**14.10**	19.64	16.80	16.80	19.60	16.38	16.39	19.60	14.56	14.86	18.45	14.18	**14.10**	19.64
L107	11.45	11.48	11.25	11.52	11.55	11.25	**10.26**	10.67	11.77	**10.26**	10.33	12.10	13.78	14.52	12.48	19.06	19.24	12.97	**10.26**	10.67	11.77	**10.26**	10.33	12.10
L108	11.34	11.38	12.04	12.38	12.45	12.34	**11.24**	11.31	11.97	11.36	11.31	12.04	13.94	13.97	12.47	21.29	24.07	13.04	**11.24**	11.31	11.97	11.36	11.31	12.04

Table 23: *Comparison of all the methods (ℓ_1MTE,ℓ_2MTE, ℓ_1LSA, ℓ_2LSA 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.*

Methods / Locations	Metric: Mean Absolute Error (MAE)																							
	ℓ_1LSA$_{AIC}$			ℓ_2LSA$_{AIC}$			ℓ_1MTE$_{AIC}$			ℓ_2MTE$_{AIC}$			ℓ_1LSA$_{BIC}$			ℓ_2LSA$_{BIC}$			ℓ_1MTE$_{BIC}$			ℓ_2MTE$_{BIC}$		
	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO
L101	10.40	10.44	11.24	10.75	10.73	12.47	**9.26**	9.53	11.51	9.37	9.71	12.37	11.38	11.40	12.48	13.31	13.26	12.48	**9.26**	9.53	11.51	9.37	9.71	12.37
L102	10.97	11.00	12.18	10.99	10.98	12.18	**10.28**	10.45	11.72	10.30	10.41	11.68	10.97	11.00	12.18	13.07	13.07	12.20	**10.28**	10.45	11.72	10.30	10.41	11.68
L103	14.26	14.31	16.49	14.32	14.36	16.49	**13.98**	14.00	16.16	14.11	14.36	16.86	14.83	14.90	16.54	14.87	14.94	16.54	**13.98**	14.00	16.16	14.12	14.36	16.85
L104	13.91	13.96	14.30	14.23	14.30	14.88	13.78	13.97	14.71	13.81	13.70	14.45	14.82	15.31	14.22	15.29	15.31	**13.67**	13.78	13.97	14.71	13.81	13.70	14.45
L106	15.93	15.93	18.10	15.91	15.88	19.43	15.05	15.35	18.94	14.68	**14.59**	20.14	17.30	17.30	20.09	16.88	16.88	20.09	15.05	15.35	18.94	14.68	**14.59**	20.14
L107	11.94	11.97	11.74	12.01	12.04	11.74	**10.74**	11.16	12.26	**10.74**	10.82	12.59	14.27	15.01	12.97	19.55	19.74	13.46	**10.74**	11.16	12.26	**10.74**	10.82	12.59
L108	11.82	11.87	12.53	12.87	12.94	12.83	**11.73**	11.80	12.46	11.85	11.80	12.53	14.43	14.46	12.96	21.79	24.57	13.53	**11.73**	11.80	12.46	11.85	11.80	12.53

Table 24: *Comparison of all the methods (ℓ_1MTE,ℓ_2MTE, ℓ_1LSA, ℓ_2LSA 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.*

Methods / Locations	Metric: Root Mean Square Error (RMSE)																							
	ℓ_1LSA$_{AIC}$			ℓ_2LSA$_{AIC}$			ℓ_1MTE$_{AIC}$			ℓ_2MTE$_{AIC}$			ℓ_1LSA$_{BIC}$			ℓ_2LSA$_{BIC}$			ℓ_1MTE$_{BIC}$			ℓ_2MTE$_{BIC}$		
	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO
L101	13.22	17.77	15.98	13.56	18.30	15.96	**12.34**	18.66	16.17	12.45	18.26	16.10	14.25	18.77	15.41	16.16	20.10	15.52	**12.34**	18.66	16.17	12.45	18.26	16.10
L102	14.21	16.17	**13.57**	14.22	16.19	**13.57**	13.72	16.10	14.07	13.75	16.13	14.60	14.21	16.17	**13.57**	16.24	17.94	13.62	13.72	16.10	14.07	13.75	16.13	14.60
L103	19.24	16.71	14.35	19.35	16.68	14.35	19.27	16.86	14.76	19.45	16.85	14.44	19.49	17.43	**13.97**	19.57	17.38	**13.97**	19.27	16.86	14.76	19.47	16.85	14.43
L104	17.16	17.71	**14.17**	17.36	18.21	14.68	17.59	17.91	14.76	17.60	17.84	14.48	17.97	20.59	14.66	18.55	20.58	17.56	17.59	17.91	14.76	17.60	17.84	14.48
L106	20.53	23.20	21.30	20.31	24.34	20.06	19.98	25.60	22.54	**19.85**	25.55	22.71	21.48	25.38	19.93	21.13	25.25	19.93	19.98	25.60	22.54	**19.85**	25.55	22.71
L107	15.25	15.96	**12.96**	15.31	15.96	**12.96**	14.45	15.83	14.02	14.63	15.93	14.21	17.45	19.07	14.17	22.61	23.37	14.60	14.45	15.83	14.02	14.63	15.93	14.21
L108	16.61	15.02	12.51	17.56	15.70	**12.45**	16.73	15.08	12.72	16.88	15.08	12.51	18.99	16.98	12.53	26.28	26.00	13.03	16.73	15.08	12.72	16.88	15.08	12.51

Table 25: *Comparison of all the methods (ℓ_1MTE,ℓ_2MTE, ℓ_1LSA, ℓ_2LSA 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.*

Methods / Locations	Metric: Symmetric Mean Absolute Percentage Error (SMAPE)																							
	ℓ_1LSA$_{AIC}$			ℓ_2LSA$_{AIC}$			ℓ_1MTE$_{AIC}$			ℓ_2MTE$_{AIC}$			ℓ_1LSA$_{BIC}$			ℓ_2LSA$_{BIC}$			ℓ_1MTE$_{BIC}$			ℓ_2MTE$_{BIC}$		
	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO	MEB	Actual	ADL1LASSO
L101	15.49	15.63	16.79	16.34	16.40	19.34	**13.87**	14.14	17.68	14.14	14.98	19.57	17.29	17.41	18.78	20.81	20.82	18.84	**13.87**	14.14	17.68	14.14	14.98	19.57
L102	15.96	16.05	17.16	16.02	16.07	17.16	14.36	14.67	16.46	**14.31**	14.46	16.59	15.96	16.05	17.16	19.75	19.80	17.24	14.36	14.67	16.46	**14.31**	14.46	16.59
L103	20.66	20.82	22.99	20.67	20.79	22.99	**19.71**	19.76	22.40	19.88	20.13	23.36	22.20	22.37	23.11	22.19	22.37	23.11	**19.71**	19.76	22.40	19.89	20.13	23.35
L104	25.27	25.39	25.86	26.01	26.14	27.29	24.54	24.79	25.82	25.11	**24.45**	26.44	27.18	27.98	25.06	27.95	27.98	24.60	24.54	24.79	25.82	25.11	**24.45**	26.44
L106	17.51	17.55	19.97	17.78	17.79	21.64	16.69	17.05	20.97	16.38	**16.37**	22.77	19.78	19.83	22.56	19.31	19.37	22.56	16.69	17.05	20.97	16.38	**16.37**	22.77
L107	21.71	21.82	20.24	21.82	21.92	20.24	18.73	19.87	21.15	**18.41**	18.50	21.90	26.34	27.61	22.48	34.84	35.13	23.21	18.73	19.87	21.15	**18.41**	18.50	21.90
L108	19.41	19.64	19.96	22.05	22.33	20.72	**18.77**	18.86	19.86	18.94	18.87	19.96	25.62	25.77	21.02	38.20	42.87	22.34	**18.77**	18.86	19.86	18.94	18.87	19.96

Table 26: *Comparison of all the methods (ℓ_1MTE,ℓ_2MTE, ℓ_1LSA, ℓ_2LSA 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.*

Methods / Locations	Metric: Symmetric Median Absolute Percentage Error (SMdAPE)																							
	ℓ_1**LSA**$_{\mathbf{AIC}}$			ℓ_2**LSA**$_{\mathbf{AIC}}$			ℓ_1**MTE**$_{\mathbf{AIC}}$			ℓ_2**MTE**$_{\mathbf{AIC}}$			ℓ_1**LSA**$_{\mathbf{BIC}}$			ℓ_2**LSA**$_{\mathbf{BIC}}$			ℓ_1**MTE**$_{\mathbf{BIC}}$			ℓ_2**MTE**$_{\mathbf{BIC}}$		
	MEB	**Actual**	**ADL1LASSO**	**MEB**	**Actual**	**ADL1LASSO**	**MEB**	**Actual**	**ADL1LASSO**	**MEB**	**Actual**	**ADL1LASSO**	**MEB**	**Actual**	**ADL1LASSO**	**MEB**	**Actual**	**ADL1LASSO**	**MEB**	**Actual**	**ADL1LASSO**	**MEB**	**Actual**	**ADL1LASSO**
L101	12.58	12.68	12.11	12.71	12.57	13.27	**10.23**	10.33	12.12	10.38	10.75	13.31	14.02	13.94	12.76	16.48	16.29	12.58	**10.23**	10.33	12.12	10.38	10.75	13.31
L102	11.94	11.98	12.62	11.86	11.86	12.62	**10.96**	11.18	12.44	11.04	11.12	12.29	11.94	11.98	12.62	14.40	14.34	12.55	**10.96**	11.18	12.44	11.04	11.12	12.29
L103	14.24	14.21	16.66	14.27	14.32	16.66	13.96	13.96	16.26	14.10	**13.95**	17.17	15.00	14.96	16.72	15.09	15.12	16.72	13.96	13.96	16.26	14.02	**13.95**	17.21
L104	20.77	20.88	**19.07**	21.83	22.01	19.61	19.92	20.02	19.81	21.19	19.28	19.37	23.49	23.97	19.07	23.71	23.86	19.87	19.92	20.02	19.81	21.19	19.28	19.37
L106	13.87	13.90	14.19	13.99	13.89	15.18	12.21	12.26	14.75	**11.50**	11.54	15.16	15.46	15.39	15.73	14.81	14.73	15.73	12.21	12.26	14.75	**11.50**	11.54	15.16
L107	16.06	16.01	14.86	16.29	16.25	14.86	13.46	13.96	15.41	**13.12**	13.22	15.74	19.42	20.13	16.48	30.52	30.69	16.54	13.46	13.96	15.41	**13.12**	13.22	15.74
L108	14.56	14.62	15.56	16.49	16.59	15.48	**14.35**	14.52	15.43	14.67	14.38	15.56	19.01	18.96	15.48	31.18	36.40	15.89	**14.35**	14.52	15.43	14.67	14.38	15.56

Table 27: *Comparison of all the methods ($\ell_1 MTE$,$\ell_2 MTE$, $\ell_1 LSA$, $\ell_2 LSA$ under AIC and BIC minimization) under Symmetric Median Absolute Percentage Error (SMdAPE) 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.*

5.3 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-Khinchine 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

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

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

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

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

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

where $\mathcal{F}$ denotes the Fourier transform, $\overline{z}$ 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, $f \star g = \mathcal{F}[\overline{F}(\nu) G(\nu)]$.

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. □

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] A. Metcalfe and P. Cowpertwaite, *Introductory Time Series in R*. Springer, 2009.

[10] K. Knight and W. Fu, "ASYMPTOTICS FOR LASSO-TYPE ESTIMATORS," *The Annals of Statistics*, vol. 28, no. 5, pp. 1356–1378, 2000.

[11] H. Zou, "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, vol. 101, pp. 1418–1429, 2006.

[12] P. Stoice and B. Ottersten, "The Evil of Superefficiency," *Signal Processing*, vol. 55, p. 133–136, 1996.

[13] 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.

[14] L. Breiman, "Better Subset Regression Using the Nonnegative Garrote," *Technometrics*, vol. 37, no. 4, 1995.

[15] D. Pollard, "Asymptotics for Least Absolute Deviation Regression Es timators," *Econmetric Theory*, vol. 7, pp. 186–199, 1991.

[16] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least Angle Regression," *The Annals of Statistics.*, vol. 32, no. 2, p. 407–451, 2004.

[17] 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.

[18] H. White, "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity.," *Econometrica*, vol. 48, no. 4, pp. 817–830, 1994.

[19] 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.

[20] F. Cribari-Neto, "Asymptotic Inference Under Heteroskedasticity of Unknown Form.," *Computational Statistics & Data Analysis*, vol. 45, no. 2, pp. 215–233, 2004.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] H. D. Vinod, "Maximum Entropy Ensemble for Time Series Inference in Economics," *Journal of Asian Economics*, 2006.

[28] H. D. Vinod, "Maximum Entropy Bootstrap Algorithm Enhancements." 2013.

[29] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.

[30] M. Nguyen, *A Guide on Data Analysis*, ch. 8.4. 2020.

[31] D. Anderson and K. Burham, *AIC MYTHS AND MISUNDERSTANDINGS*, 2006.

[32] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, 1964.

[33] R. Flides, "The evaluation of extrapolative forecasting methods," *International Journal of Forecasting*, vol. 8, pp. 81–98, 1992.

[34] S. J. Armstrong, *Long-Range Forecasting: From Crystal Ball to Computer*. A WILEY-INTERSCIENCE PUBLICATION, 1985.

[35] 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.

[36] S. G. Makridakis, "Assessing Forecast Accuracy Measures," *International Journal of Forecasting 9*, vol. 4, pp. 527–529, 2004.

[37] P. Goodwin and R. Lawton, "On the Asymmetry of the Symmetric MAPE.," *International Journal of Forecasting*, 1999.

[38] Z. Chen and Y. Yang, "Assessing Forecast Accuracy Measures," 2004.

[39] D. W. Andrews, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, vol. 59, no. 3, pp. 817–858, 1991.

[40] 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.