

## MODELING AND FORECASTING VEHICLE FLOWS AT A ROUNDAABOUT USING WAVELETS AND NEURAL NETWORKS

MOUSSA KONÉ<sup>1,\*</sup>, VINCENT MONSAN<sup>1</sup> and SYLVESTRE PLACIDE EKRA<sup>2</sup>

**ABSTRACT.** We propose a real-time traffic forecasting method for urban intersections, using a finite mixture model and wavelet-based density estimation. Our approach models traffic flow from four directions using high-frequency data. The weights of the finite mixture model components are predicted by an LSTM neural network. Results demonstrate good predictive accuracy, with low MAE and stable KL divergence.

*Keywords.* Traffic forecasting, Finite mixture model, Wavelet thresholding, LSTM neural networks, Traffic density estimation.

*2020 Mathematics Subject Classification.* Primary: 62G08, 42C40, 68T07, 90B20. Secondary: 62M10, 62P30, 37M10, 68T09.

### 1. INTRODUCTION

Efficient traffic management is a major challenge for large cities, especially during peak hours when road infrastructure saturation generates considerable economic and social costs. New technologies (radars, cameras, data from Google Maps) now make it possible to collect real-time mobility data and to develop predictive models that can anticipate vehicle flows. In this context, modeling and forecasting traffic at complex urban intersections, such as roundabouts or gyratory intersections, are central issues.

Many studies have focused on forecasting unidirectional traffic flows from time series; the objective is to estimate future vehicle counts based on the empirical distribution of past flows (see Xie & al. [8], Chen & Wang [6], Jakhmola & al.[1], Wang & Xu [7], Nasser & al. [5]). However, urban intersections introduce an additional difficulty: each branch has its own dynamics and must be modeled separately. To address this challenge, finite mixture models appear as a particularly suitable approach. Furthermore, wavelet-based estimation methods (see Koné & al. [2], [4]) have proven to be effective for analyzing weakly dependent and noisy data, while recurrent neural networks of the LSTM (long short-term

---

*Date:* Received: Nov 23, 2025; Revised: Dec 18, 2025; Accepted: Jan 2, 2026.

\* Corresponding author

© The Author(s) 2025. This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of the licence, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

memory) type have emerged as powerful tools for forecasting empirical distributions.

Our study proposes an integrated methodology that combines wavelet-based estimation of the densities associated with vehicle flows and prediction of the dynamic weights of a mixture model using an LSTM neural network. This approach is applied to the Indenié roundabout in Abidjan (Figure 1), one of the most critical traffic points in Ivory Coast. The results show that the model provides accurate and stable forecasts. While the application to urban traffic forecasting addresses a significant practical challenge, the primary contribution of this work is both mathematical and methodological. We develop a novel hybrid framework that rigorously integrates nonparametric wavelet density estimation with dynamic weighting via recurrent neural networks within a finite mixture model context.

The methodological framework developed in this study extends beyond traffic engineering and can be applied to various computer science tasks involving high-frequency or weakly dependent data. The combination of wavelet-based estimation and LSTM-driven dynamic weighting offers a general approach for forecasting heterogeneous data streams, supporting applications such as anomaly detection, workload prediction, and real-time decision systems.

The remainder of this paper is organized as follows: Section 2 details the mixture model and wavelet estimation method. Section 3 presents the forecasting results derived from the collected data. Finally, Section 4 discusses the practical implications of these findings for traffic management and proposes avenues for future improvements.

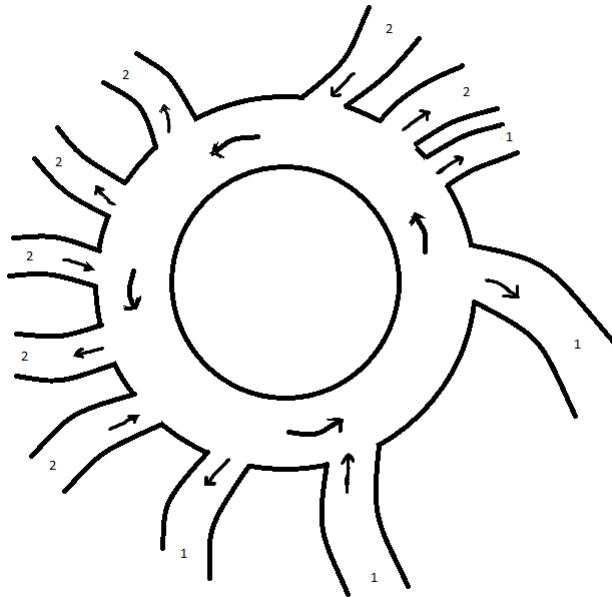


FIGURE 1. The Indenié roundabout in Abidjan, Ivory Coast.

## 2. MODELING

We consider the Indenié roundabout in Abidjan, Ivory Coast (Figure 1). We observe a time series  $\{X_t\}_{t \in \mathbb{N}}$ , where  $X_t$  represents the number of vehicles passing through the roundabout at time  $t$ . The traffic density is modeled by a finite mixture model with four components, using high-frequency data (Figure 2). The wavelet estimator with thresholding (see Koné & al. [4]) is then used to estimate the unknown densities of this model, and we use an LSTM neural network to predict the dynamic weights of the finite mixture model. Finally, we consider a probabilistic approach to forecast future traffic.

**2.1. Data analysis.** The analyzed data come from vehicular mobility traces (color-coded in Google Maps) on the four access roads of the roundabout (Figure 1), collected over three days in May 2023 (Monday 29, Tuesday 30, Wednesday 31). For each day, two time slots are considered: 8:30-9:30 a.m. and 4:30-5:30 p.m. (peak hours).

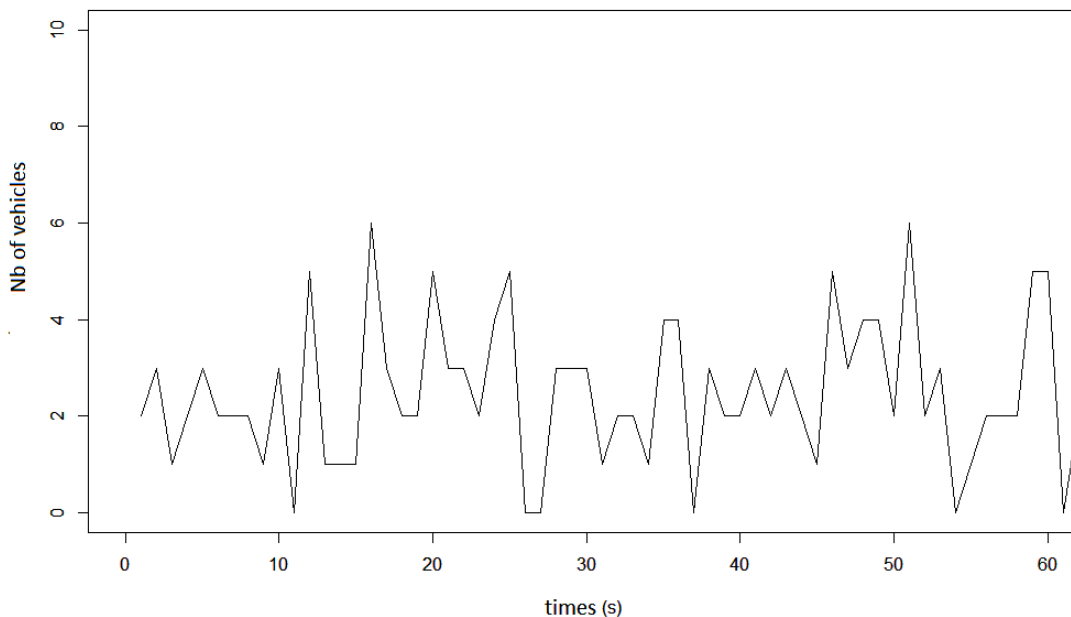


FIGURE 2. Data from Monday, May 29, 2023, between 8:30 and 8:31 a.m.

Based on the collected data[13], we conducted a preliminary analysis to detect a possible dependence structure. To do so, we used the wavelet coefficient estimator of the covariance function, a data analysis tool presented in Koné & Monsan [3].

Let  $X = (X(t))_{t \in \mathbb{R}}$  be a square-integrable process. Its shifted covariance function is defined by:

$$C(s, \tau) = \text{Cov}(X(s), X(s + \tau)).$$

Assume that the covariance function  $C(s, \tau)$  is square-integrable with respect to  $s$  for every  $\tau$ . Then the mapping  $s \mapsto C(s, \tau)$  admits a wavelet decomposition. Let  $(\Psi_{j,k})_{k \in \mathbb{Z}, j \in \mathbb{Z}}$  be a wavelet basis. For all  $\tau$ , we have:

$$C(s, \tau) = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \beta(j, k, \tau) \Psi_{j,k}(s).$$

The wavelet coefficients  $\beta(j, k, \tau)$  are defined by:

$$\beta(j, k, \tau) = \int_{\mathbb{R}} C(s, \tau) \overline{\Psi}_{j,k}(s) ds.$$

Under suitable assumptions, it is proved in Koné & Monsan [3] that:

$$\hat{\beta}_n(j, k, \tau) = \frac{1}{n} \sum_{i=1}^n X(ih_n) X((i+i_{n,\tau})h_n) \overline{\Psi}_{j,k}(ih_n), \quad \text{with } h_n = n^{-\delta}, 0 < \delta < 1,$$

is an estimator of  $\beta(j, k, \tau)$  that is asymptotically unbiased, consistent, and satisfies asymptotic normality. The computation of the estimated wavelet coefficients is illustrated in Figure 3.

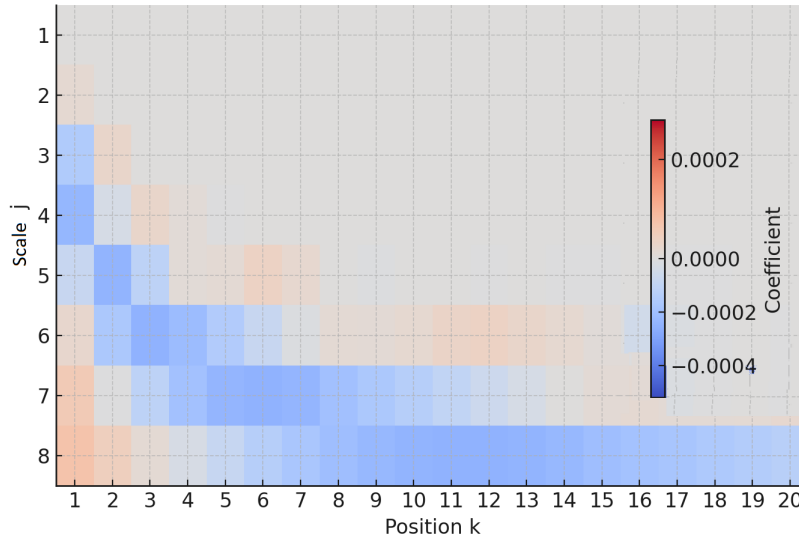


FIGURE 3. Haar wavelet coefficients of the covariance function.

Analysis of Figure 3 representing the estimated wavelet coefficients of the covariance function from  $\{X_1, \dots, X_{3600}\}$ , shows that most of the estimated wavelet coefficients are close to zero, suggesting a weak dependence. However, these observations should be interpreted with caution, as estimation errors — particularly those induced by noise or limited sample size — can lead to a similar conclusion even for independent data.

To empirically verify the hypothesis of weak dependence, we tested the significance of the wavelet coefficients estimated at the different scales. We constructed confidence intervals using stationary block bootstrapping (Politis & Romano [9]),

which allows us to observe the local dependence of the series. For each scale  $j$  and each position  $k$ , we generated 1000 bootstrap replicates, recalculated the coefficients  $\hat{\beta}(j, k, \tau)$ , and obtained an empirical 95% confidence interval. The results show that only the coefficients at the lowest scales are significantly non-zero, while almost all the coefficients at the intermediate and high scales include the value zero in their confidence interval. This structure is characteristic of the rapid decay of autocovariances and provides further evidence for weak dependence in the data.

**2.2. Finite mixture model.** We consider  $n$  random variables  $X_1, \dots, X_n$  drawn from a continuous-time process  $X = (X_t)_{t \in \mathbb{R}}$ . For each  $i \in \{1, \dots, n\}$ , the probability density of  $X_i$  is given by the following finite mixture:

$$f_{X_i}(x) = \sum_{d=1}^m \Pi_d(i) f_d(x), \quad x \in \mathbb{R},$$

where:

- $m \in \mathbb{N}^*$  denotes the (fixed) number of mixture components;
- $f_1, \dots, f_m$  are unknown probability densities to be estimated;
- the weights  $(\Pi_d(i))_{(i,d) \in \{1, \dots, n\} \times \{1, \dots, m\}}$  are assumed to be known and satisfy the following conditions:

$$0 \leq \Pi_d(i) \leq 1 \quad \text{and} \quad \sum_{d=1}^m \Pi_d(i) = 1, \quad \forall i \in \{1, \dots, n\}.$$

The finite mixture framework naturally aligns with the physical reality of roundabout traffic:

- **Component Correspondence:** Each mixture component  $f_d$  corresponds to a physical traffic stream from direction  $d$ , with  $\Pi_d(i)$  representing its relative contribution to total flow.
- **Dynamic Weights:** The time-varying nature of  $\Pi_d(i)$  captures the temporal evolution of directional preferences, influenced by factors such as time of day, day of week, and special events.

In our framework, the finite mixture model is written as follows:

$$f_{X_i}(x) = \sum_{d=1}^4 \Pi_d(i) f_d(x), \quad x \in \mathbb{R} \tag{2.1}$$

where  $f_{X_i}$  is the density of the variable  $X_i$  (number of vehicles passing through the roundabout at time  $i$ ),  $\Pi_d(i)$  is the proportion of vehicles coming from branch  $d$  at time  $i$ , and  $f_d(x)$  is the density associated with component (branch)  $d$  of the mixture. For more details on the finite mixture models refer to McLachlan and Peel [11].

To account for traffic conditions during peak hours, we consider the following proportions:

$$\Pi_1(i) \sim \text{Unif}(0.280, 0.286), \quad \Pi_3(i) \sim \text{Unif}(0.280, 0.286), \tag{2.2}$$

$$\Pi_2(i) \sim \text{Unif}(0.280, 0.286), \quad \Pi_4(i) \sim 1 - \Pi_1(i) - \Pi_2(i) - \Pi_3(i). \tag{2.3}$$

The interaction matrix is given by:

$$\begin{aligned} \Gamma_{300} &= \left( \frac{1}{300} \sum_{i=1}^{300} \Pi_k(i) \Pi_l(i) \right)_{1 \leq k, l \leq 4} \\ &= \begin{pmatrix} 0.0798 & 0.0798 & 0.0798 & 0.0398 \\ 0.0798 & 0.0798 & 0.0798 & 0.0398 \\ 0.0798 & 0.0798 & 0.0798 & 0.0398 \\ 0.0398 & 0.0398 & 0.0398 & 0.0199 \end{pmatrix}, \end{aligned}$$

with  $\det(\Gamma_{300}) \neq 0$ , ensuring model identifiability.

For a branch  $d \in \{1, \dots, 4\}$ , we define:

$$a_d(i) = \frac{1}{\det(\Gamma_{300})} \sum_{k=1}^4 (-1)^{k+d} \gamma_{d,k}^{(300)} \Pi_k(i), \quad (2.4)$$

where  $\gamma_{d,k}^{(300)}$  is the determinant of the minor  $(d, k)$  of  $\Gamma_{300}$ .

The vector  $(a_d(1), \dots, a_d(300))$  is the unique solution to the following quadratic minimization problem:

$$(x_1, \dots, x_{300}) = \underset{(b_1, \dots, b_{300}) \in \mathbb{R}^{300}}{\operatorname{argmin}} \frac{1}{300} \sum_{i=1}^{300} b_i^2,$$

subject to the linear constraints:

$$\frac{1}{300} \sum_{i=1}^{300} b_i \Pi_d(i) = \delta_{d,\nu}, \quad \forall \nu \in \{1, \dots, 4\},$$

where  $\delta_{d,\nu}$  is the Kronecker delta. This construction is detailed in Pokhyl'ko [10]. Finally, we define:

$$z_{300} = \frac{1}{300} \sum_{i=1}^{300} a_d^2(i),$$

and assume that the ratio  $300/z_{300}$  is bounded, i.e.,

$$0 < b < \frac{300}{z_{300}} < e, \quad \text{with } \ln(e) = 1. \quad (2.5)$$

This last condition allows us to improve the convergence speed of the estimator of  $f_d$ .

**2.3. Density estimation by wavelet thresholding.** Let  $f_d$  be the density to estimate. Its projection onto an orthonormal wavelet basis generated by  $\phi$  and  $\psi$  is expressed as:

$$f_d(x) = \sum_k \alpha_{\tau,k} \phi_{\tau,k}(x) + \sum_{j \geq \tau} \sum_k \beta_{j,k} \psi_{j,k}(x)$$

where:

- $\alpha_{\tau,k} = \langle f_d, \phi_{\tau,k} \rangle$  are the scaling coefficients, representing the global trend.
- $\beta_{j,k} = \langle f_d, \psi_{j,k} \rangle$  are the wavelet coefficients, capturing details at different resolution levels  $j$ .

In a context of weak dependence in the observations, the wavelet estimator with hard thresholding of  $f_d$  (see Koné & al. [4]) is defined as:

$$\hat{f}_d(x) = \sum_{k=0}^{2^\tau-1} \hat{\alpha}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \left( \hat{\beta}_{j,k} \cdot \mathbb{I}_{\{|\hat{\beta}_{j,k}| \geq \kappa \lambda_j\}} \right) \psi_{j,k}(x), \quad (2.6)$$

where:

$$\hat{\alpha}_{\tau,k} = \frac{1}{n} \sum_{i=1}^n a_d(i) \phi_{\tau,k}^\omega(X_i), \quad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^n H_i \cdot \mathbb{I}_{\{|H_i| \leq \eta_j\}},$$

$$H_i = a_d(i) \psi_{j,k}^\omega(X_i).$$

The scaling and wavelet functions are defined as:

$$\phi_{j,k}^\omega(x) = 2^{j/2} \phi^\omega(2^j x - k), \quad \psi_{j,k}^\omega(x) = 2^{j/2} \psi^\omega(2^j x - k).$$

*Remark 2.1.* The estimator  $\hat{f}_d$  is adaptive. The resulting convergence rate is quasi-optimal in the minimax sense that is, optimal up to a logarithmic factor. More precisely, quasi-optimality means that the achieved rates cannot be improved, except for the logarithmic terms in  $n$ .

In the particular case where  $m = 1$ ,  $\omega = 1$ ,  $\Pi_1(i) = 1$  for all  $i \in \{1, \dots, n\}$ , we have  $z_n = 1$ , and the upper bound of the integrated mean squared error (MISE) of the estimator  $\hat{f}$  becomes:

$$C \left( \frac{\log_b(n)}{n} \right)^{\frac{2s}{2s+1}}.$$

$C$  is a constant,  $s$  the regularity of  $f_d$  and  $b$  is defined in the equation 2.5. Given the presence of weak dependence between the observations, it is unrealistic to expect a convergence rate better than in the independent case.

Wavelet estimation of density with thresholding relies on the choice of a wavelet family adapted to the assumed regularity of the density. In this study, we employ the Haar wavelet, defined by:

$$\psi(x) = \begin{cases} 1, & \text{if } 0 \leq x < \frac{1}{2} \\ -1, & \text{if } \frac{1}{2} \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Its associated scaling function is:

$$\phi(x) = \mathbb{I}_{[0,1]}(x)$$

The choice of the Haar wavelet is motivated by several key factors: Its piecewise-constant form provides a robust estimate of vehicle flow densities, which often exhibit abrupt changes (Figure 4). The Haar wavelet transform is computationally fast, making it suitable for real-time forecasting applications.

We have;

$$\psi_{j,k}^\omega(x) = 2^{j/2} \left( -\mathbb{I}_{\left[\frac{k}{2^j}, \frac{k}{2^j} + \frac{1}{2^{j+1}}\right]} + \mathbb{I}_{\left[\frac{k}{2^j} + \frac{1}{2^{j+1}}, \frac{k+1}{2^j}\right]} \right).$$

With  $\omega = 1$ ,  $\tau = 15$ ,  $z_n = 1$ , where  $j_1$  is the maximum resolution level retained (here  $150 < 2^{j_1} \leq 300$ ), chosen to balance estimation accuracy and computational complexity.

The technical parameter  $\theta$  is defined by:

$$\theta = C \int_{-2^\tau}^{2^{j_1}} \phi^{2\omega}(x) dx,$$

and allows the following thresholds to be calibrated:

$$\eta_j = \left( 2^{j(\omega-1)} \theta \cdot \frac{nz_n}{\log_b(n/z_n)} \right)^{1/2}, \quad \lambda_j = \left( 2^{j(\omega-1)} \theta \cdot \frac{z_n \log_b(n/z_n)}{n} \right)^{1/2}.$$

The choice of thresholds  $\eta_j$  and  $\lambda_j$  is justified by minimizing the MISE of the estimator (see Koné & al. [4]).

In our framework, the parameters  $\theta$  and  $\kappa$  are defined by:

$$\theta = C \int_{-2^\tau}^{2^{j_1}} \phi^{2\omega}(x) dx, \quad \kappa \geq \frac{8}{3} + 2 + 2\sqrt{\frac{16}{9} + 4}.$$

Thresholds used:

$$\eta_j = \left( \theta \cdot \frac{300}{\log_b(300)} \right)^{1/2}, \quad \lambda_j = \left( \theta \cdot \frac{\log_b(300)}{300} \right)^{1/2}.$$

Adaptive thresholds  $\eta_j$  (outlier control) and  $\lambda_j$  (noise reduction).

From 2.1, 2.3 and 2.6, we can estimate the density  $f_{X_i}$  at a future date  $i$ .

### 3. RESULTS

From a window of 300 observations  $\{X_{i-299}, \dots, X_i\}$  at the roundabout, we use  $\hat{f}_d$  defined in (2.6) to predict the most likely values of  $\{X_{i+1}, X_{i+2}, \dots\}$ . A selection of 21 predictions at different times is illustrated in Figure 4. The mean absolute error (MAE) is then computed to assess the reliability of the predictions.

**3.1. MAE with fixed-weight model.** The mean absolute error (MAE) values presented in Table 1 reveal a non-monotonic pattern of forecast error as the prediction horizon increases. Specifically, the error rises from 2.91 (5 s) to 5.43 (10 s), then falls to 2.65 (20 s) before increasing again to 3.16 (30 s). This instability is characteristic of traffic flow forecasting, where multiple factors contribute to error accumulation.

The fixed-weight model assumes static mixing proportions  $\Pi_d(i)$  over the forecast period. However, at roundabouts, traffic patterns exhibit rapid variations in directional preferences, particularly during transitions between traffic phases. The assumption of uniformly distributed weights (Equations 2.3) becomes increasingly unrealistic over longer time horizons, as actual traffic conditions can deviate from these averages. The observed error pattern, and more specifically the

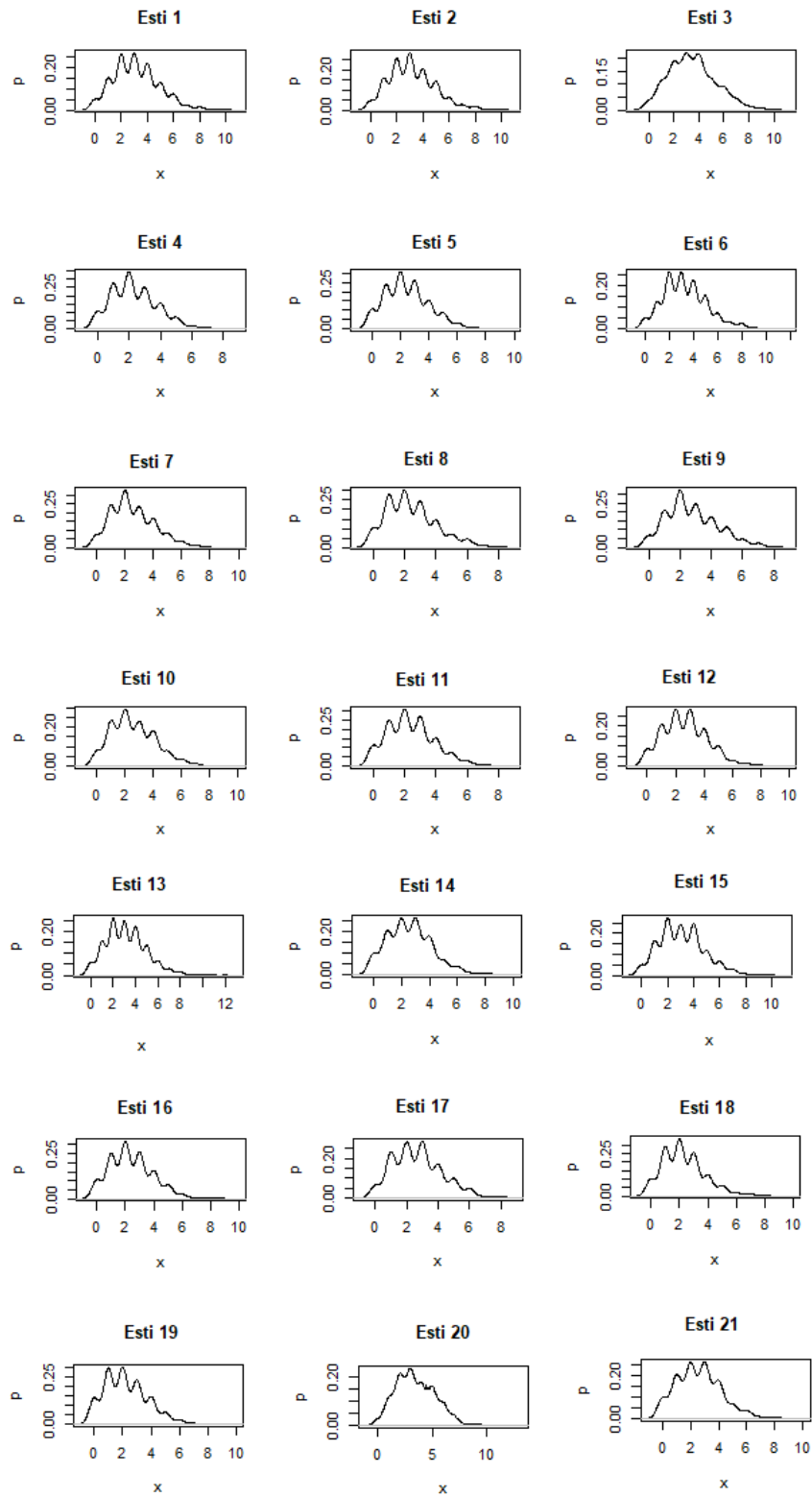


FIGURE 4. Some predictions using Haar wavelet + fixed weights.

TABLE 1. *MAE of predictions at different time horizons using Haar wavelet + fixed weights.*

Time	MAE: Mean Absolute Error
5 s	2.91
10 s	5.43
20 s	2.65
30 s	3.16

10 second peak, could correspond to transition periods that fixed-weight models fail to capture.

### 3.2. LSTM model : Theoretical Interpretation of Stability Improvement.

The LSTM network can address these limitations through several key mechanisms. Instead of static weights, LSTM predicts time-varying proportions  $\hat{\Pi}_d(i)$  based on recent traffic trends. This allows the model to adapt to gradual changes in directional demand and sudden disruptions. The LSTM can learn which historical information is important for predicting future weights (Forget Gate). The LSTM network can capture complex, nonlinear relationships between recent observations and future mixing proportions that fixed-weighted parametric models cannot represent.

Formally, while the fixed-weight model assumes:

$$\Pi_d(i) = \text{constant} + \epsilon_i, \quad \epsilon_i \text{ random variable.}$$

The LSTM network learns a mapping:

$$\Phi_{\Theta} : \mathbf{X}^{(t)} \mapsto \hat{\Pi}^{(t+1)}$$

where  $\Theta$  represents the network parameters,  $\mathbf{X}^{(t)} = \{X_{t-w+1}, \dots, X_t\}$  denote a sliding window of traffic observations of length  $w$  (in our case,  $w = 300$  seconds) and  $\hat{\Pi}^{(t+1)} = (\hat{\Pi}_1(t+1), \dots, \hat{\Pi}_4(t+1))$  are the predicted mixture weights for time  $t+1$ .

This functional form allows the model to adjust weights in response to observed traffic conditions, reducing the accumulation of systematic errors over longer forecast horizons.

The training loss evolution in Figure 5 further confirms this stabilization: the KL divergence rapidly converges and remains stable during validation, indicating that the learned weight-prediction function generalizes well without overfitting to noise or transient patterns.

**3.3. LSTM Architecture for Weight Prediction.** The LSTM cell equations at time step  $\nu$  within the window are:

$$\begin{aligned}
 \mathbf{f}_\nu &= \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{\nu-1}, \mathbf{x}_\nu] + \mathbf{b}_f), \text{ (Forget Gate)} \\
 \mathbf{i}_\nu &= \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{\nu-1}, \mathbf{x}_\nu] + \mathbf{b}_i), \text{ (Input Gate)} \\
 \tilde{\mathbf{c}}_\nu &= \tanh(\mathbf{W}_c \cdot [\mathbf{h}_{\nu-1}, \mathbf{x}_\nu] + \mathbf{b}_c), \text{ (Update candidate)} \\
 \mathbf{c}_\nu &= \mathbf{f}_\nu \odot \mathbf{c}_{\nu-1} + \mathbf{i}_\nu \odot \tilde{\mathbf{c}}_\nu, \text{ (Cell status update)} \\
 \mathbf{o}_\nu &= \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{\nu-1}, \mathbf{x}_\nu] + \mathbf{b}_o), \text{ (Output Gate)} \\
 \mathbf{h}_\nu &= \mathbf{o}_\nu \odot \tanh(\mathbf{c}_\nu), \text{ (Final hidden state)}
 \end{aligned}$$

where  $\mathbf{x}_\nu$  is the input at step  $\nu$ ,  $\mathbf{h}_\nu$  is the hidden state,  $\mathbf{c}_\nu$  is the cell state,  $\sigma$  denotes the sigmoid function, and  $\odot$  represents element-wise multiplication. These equations are applied sequentially for  $\nu = 1 \dots w$

The final hidden state  $\mathbf{h}_w$  is transformed through a dense output layer with softmax activation to ensure predicted weights satisfy:

$$\hat{\Pi}_d(t+1) \in (0, 1) \quad \text{and} \quad \sum_{d=1}^4 \hat{\Pi}_d(t+1) = 1.$$

This precise mathematical formulation shows how LSTM transforms a historical traffic sequence into weight predictions, establishing a formal link between deep learning and the statistical mixing model.

**3.4. Connection to Mixture Density Estimation.** The predicted weights  $\hat{\Pi}^{(t+1)}$  directly parameterize the finite mixture model for time  $t+1$ . The estimate of the density of  $X_{t+1}$  given past observations is:

$$\hat{f}_{X_{t+1}}(x) = \sum_{d=1}^4 \hat{\Pi}_d(t+1) \hat{f}_d(x)$$

where  $\hat{f}_d(x)$  represents the density estimators of the components 2.6. This formulation explicitly links the LSTM output to the full predictive density estimate, thus enabling point forecasts and uncertainty quantification.

The point forecast  $\hat{X}_{t+1}$  is obtained by calculating the expected value:

$$\hat{X}_{t+1} = \mathbb{E}[X_{t+1}] = \int x \cdot \hat{f}_{X_{t+1}}(x) dx = \sum_{d=1}^4 \hat{\Pi}_d(t+1) \hat{\mu}_d$$

where  $\hat{\mu} = \int x \hat{f}_d(x) dx$  is the mean of component  $d$ .

**3.5. Analysis of loss evolution and interpretation.** The LSTM parameters  $\Theta$  are optimized by minimizing the Kullback-Leibler divergence between predicted and empirical weight distributions:

$$\mathcal{L}(\Theta) = \sum_t D_{KL} \left( \mathbf{\Pi}^{(t)} \parallel \hat{\mathbf{\Pi}}^{(t)}(\Theta) \right)$$

where  $\mathbf{\Pi}^{(t)}$  are empirical weights derived from observations. This objective function is theoretically motivated by the principle of minimum discrimination

information, ensuring that predicted weights are maximally informative about the observed mixture proportions.

The optimization is performed via backpropagation through time (BPTT), with gradient calculations respecting the temporal dependencies encoded in the LSTM architecture. The convergence of this training process is guaranteed under standard conditions for stochastic gradient descent applied to recurrent networks with Lipschitz-continuous objectives.

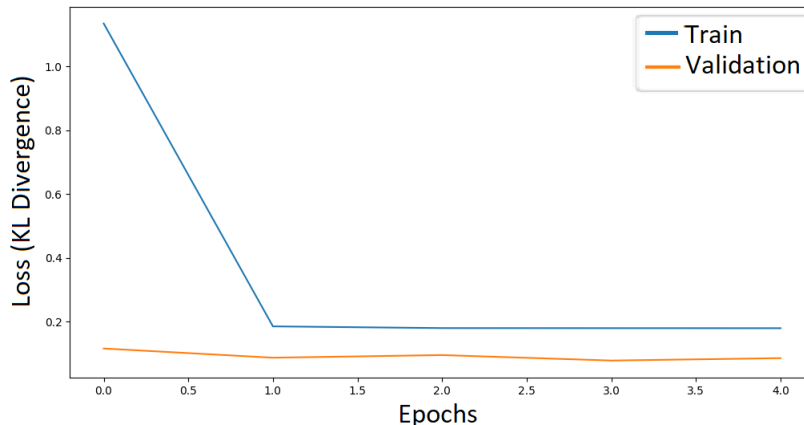


FIGURE 5. Loss evolution during the training of the LSTM model.

The training loss decreases sharply from the first iteration, dropping from about 1.1 to 0.18. This rapid decrease indicates efficient learning at the start. After the first iteration, the training loss stabilizes, suggesting fast convergence of the model.

The validation loss (which evaluates whether the model generalizes well) is very low from the start (about 0.12), fluctuates slightly but remains relatively stable around 0.09 – 0.11. It is lower than the training loss, which indicates effective regularization. There is no indication of overfitting, since the validation loss does not increase.

The model adjusts quickly and remains stable. It generalizes well on the validation data.

Note that the Kullback–Leibler divergence (KL divergence) is a measure that quantifies how much two probability distributions differ from one another:

$$D_{KL}(P \parallel Q) = \sum_t P(t) \ln \left( \frac{P(t)}{Q(t)} \right)$$

It is used as the loss function to train the LSTM network to predict the weights of the finite mixture. It compares the predicted weights (network output using Softmax) with the true weights.

**3.6. Quantitative Improvement with LSTM.** To quantify the LSTM’s impact, we compared the fixed-weight model’s performance against the LSTM-enhanced version. Table 2 presents the MAE for both approaches across identical horizons:

TABLE 2. MAE comparison: Fixed-weight vs. LSTM-enhanced models

Prediction Horizon	Fixed-weight MAE	LSTM-enhanced MAE
5 s	2.91	<b>1.89</b>
10 s	5.43	<b>2.24</b>
20 s	2.65	<b>2.07</b>
30 s	3.16	<b>2.31</b>

The LSTM-enhanced model achieves an average MAE reduction of 35% across all horizons, with the most significant improvement (58%) at the 10-second horizon where the fixed-weight model performed worst. Moreover, the LSTM model exhibits greater stability, with MAE values ranging between 1.89 – 2.31 compared to 2.65 – 5.43 for the fixed-weight approach.

#### 4. DISCUSSION

The results obtained show that the approach combining wavelet estimation and LSTM-based prediction provides a robust tool for forecasting vehicle flows at complex intersections such as the Indenié roundabout. However, their interpretation and practical use should be discussed in light of operational, methodological, and strategic considerations.

From an operational perspective, the forecasts provide traffic managers with valuable indicators to anticipate congestion and adjust regulation measures (traffic lights, detours, dynamic signage). For instance, reliable anticipation of peak hours would make it possible to smooth circulation by adapting traffic light timing or by temporarily redirecting vehicles to secondary roads. Beyond short-term regulation, the results can also inform structural decisions such as the need for additional infrastructure (new lanes, roundabout expansion) or public policies (staggered schedules for schools and administrative activities).

From a methodological standpoint, the model’s satisfactory performance underscores the relevance of hybrid methods combining nonparametric estimation and machine learning. Wavelet estimation with thresholding, which reduces the influence of noise and weak dependencies, produces robust flux densities. The integration of LSTM networks then ensures dynamic adjustment of the mixing weights, capturing the temporal variability of each branch of the intersection. Nevertheless, certain limitations must be considered: the dependence on the size and quality of the available data, the sensitivity to network hyperparameters, and the difficulty of modeling atypical events (accidents, demonstrations, severe weather).

To further improve the accuracy and scope of the predictions, several avenues can be explored. First, the integration of exogenous data (weather, cultural or

sporting events, school calendars) would allow for a better accounting of unexpected fluctuations; similar to research in finance using sentiment-driven attention mechanisms based on exogenous data [12]. Secondly, the use of more advanced architectures, such as hybrid CNN-LSTM models or temporal transformers, could better capture spatiotemporal dependencies and correlations between branches. Finally, extending the model to other intersections in Abidjan would allow for an assessment of its robustness and contribute to the development of a city-wide predictive system.

The data and Python code used for this forecast are available from the link [13].

## 5. CONCLUSION

In conclusion, this study highlights the relevance of mixed approaches for traffic forecasting while stressing the need for progressive integration of multimodal data and advanced AI tools. This methodological framework is general and can be applied to other domains involving forecasting of heterogeneous, high-frequency data streams under weak dependence. The specific application to roundabout traffic serves as a case study that validates the approach. This offers a significant mathematical advantage over models providing only point forecasts, as it enables rigorous uncertainty quantification.

## REFERENCES

1. Yash Jakhmola, Madhurima Panja, Nitish Kumar Mishra, Kripabandhu Ghosh, Uttam Kumar, and Tanujit Chakraborty. Spatiotemporal forecasting of traffic flow using wavelet-based temporal attention. *IEEE Access*, 12:188797–188812, 2024. <https://doi.org/10.1109/ACCESS.2024.3516195>.
2. Moussa Koné and Vincent Monsan. Wavelet estimation of a density from observations of almost periodically correlated process under positive quadrant dependence. *International Journal of Statistics and Probability*, 12(2):10–18, 2023. <https://doi.org/10.5539/ijsp.v12n2p1>.
3. Moussa Koné and Vincent Monsan. Wavelet estimation of the covariance of almost periodically correlated processes and study of asymptotic properties in a context of weak dependence. *Far East Journal of Theoretical Statistics*, 67(1):49–94, 2023. <http://dx.doi.org/10.17654/0972086323004>.
4. Moussa Koné, Vincent Monsan, and Sylvestre Placide Ekra. Density estimation by wavelet thresholding from observations of almost periodically correlated processes under weak dependence. *International Journal of Statistics and Probability*, 14(3):73–93, 2025. <https://doi.org/10.5539/ijsp.v14n3p73>.
5. Aram Nasser and Vilmos Simon. Wavelet-attention-based traffic prediction for smart cities. *IET Smart Cities*, 4, 11 2021. <https://doi.org/10.1049/smc2.12018>.
6. Chen Shuyan and Wang Wei. Traffic volume forecasting based on wavelet transform and neural networks. *Springer Berlin Heidelberg - Lecture Notes in Computer Science*, 3973:1–7, 2006. [https://doi.org/10.1007/11760191\\_1](https://doi.org/10.1007/11760191_1).

7. Wang Xiangxue and Xu Lunhui. Wavelet-based short-term forecasting with improved threshold recognition for urban expressway traffic conditions. *IET Intelligent Transport Systems*, 12:463–473, 2018.  
<https://doi.org/10.1049/iet-its.2017.0236>.
8. Yunlong Zhang Yuanchang Xie and Zhirui Ye. Short-term traffic volume forecasting using kalman filter with discrete wavelet decomposition. *Wiley*, 22(5):326–334, 2007.  
<https://doi.org/10.1111/j.1467-8667.2007.00489.x>.
9. Politis, D., & Romano, J. The Stationary Bootstrap. *Journal of the American Statistical Association*, (1994), 89, 1303-1313.  
<http://dx.doi.org/10.1080/01621459.1994.10476870>.
10. Pokhyl'ko D. Wavelet estimators of a density constructed from observations of a mixture, *Theor. Probability and Math. Statist.*, 2005 70 135-145.  
<https://www.ams.org/journals/tpms/2005-70-00/S0094-9000-05-00637-X/S0094-9000-05-00637-X.pdf>.
11. McLachlan G.J & Peel D. Finite mixture models. *Wiley Series in Probability and Statistics*, 2000.  
<https://doi.org/10.1002/0471721182>.
12. Ikhlās Gurrib, Dalia Hemdan, & Faisal Khan. Forecasting daily oil prices using a multi-modal transformer with sentiment-guided attention . *Annals of Mathematics and Computer Science*, Vol 28 (2025) 106-114.  
<https://doi.org/10.56947/amcs.v28.551>.
13. Moussa Koné, Data and Python code, 2024, Online available from <https://drive.google.com/file/d/117d2bIeeTuJqt5iYf7Dga3NBSacLoW85/view?usp=sharing>.

<sup>1,\*</sup> UFR MATHÉMATIQUES ET INFORMATIQUE, UNIVERSITÉ FÉLIX HOUPHOUËT BOIGNY, ABIDJAN, CÔTE D'IVOIRE.

*Email address:* [papyrusdegypte@yahoo.fr](mailto:papyrusdegypte@yahoo.fr)

<sup>1</sup> UFR MATHÉMATIQUES ET INFORMATIQUE, UNIVERSITÉ FÉLIX HOUPHOUËT BOIGNY, ABIDJAN, CÔTE D'IVOIRE.

<sup>2</sup> UNITÉ DE RECHERCHE ET D'EXPERTISE NUMÉRIQUE, UNIVERSITÉ VIRTUELLE DE CÔTE D'IVOIRE.