Quality & Quantity
https://doi.org/10.1007/s11135-025-02558-0

Forecasting traffic flow time series with vine–transform ARMA copula models

Sara Selvaggia Guerini[1] · **Rodolfo Metulini**[2]

Received: 30 June 2025 / Accepted: 23 December 2025

Abstract
Traffic–flow forecasting is gaining prominence in urban areas to facilitate urban planning for early warning systems and optimized logistics. Hence, there is a growing need for simple and high–performing statistical models. This study leverages the Vine–Transform AutoRegressive Moving–Average (VT–ARMA) copula model to forecast traffic data, and it emphasizes the evaluation of forecasting performance. Accordingly, we used real–life data on origin–destination signals extracted from mobile phone signals for specific areas in the province of Brescia (Italy). We conducted performance evaluation using the rank–graduation box approach along with a moving window cross–validation strategy, and incorporated rank–graduation accuracy for precision and rank–graduation explainability for component analysis. As a benchmark for comparison, the VARX–DHR and the Facebook Prophet models were used. Our results reveal that the VT–ARMA copula approach performed well in terms of accuracy, which was approximately 0.99 under the best specification. Furthermore, the copula component presented greater explainability than the autoregressive and moving–average components. In addition, residual diagnostics show significantly lower autocorrelation and partial autocorrelation with respect to the original data, and that the residuals are approximately normally distributed. Overall, the method developed in this study could provide valuable insights supporting urban planners and analysts in making informed decisions.

Keywords Vine–transform ARMA copulas · Rank–graduation approach · Predictive performance · Time series modeling · Urban traffic analysis

✉ Sara Selvaggia Guerini
s.guerini9@campus.unimib.it

✉ Rodolfo Metulini
rodolfo.metulini@unibg.it

[1] University of Milano-Bicocca, Milan, Italy

[2] University of Bergamo, Bergamo, Italy

Published online: 02 February 2026

1 Introduction

Studying traffic in urban areas has become increasingly important in recent years for analyzing various aspects of life in smart cities, playing a critical role not only in optimizing transportation systems (Tao et al. 2023) but also in enhancing urban planning (Shahriari et al. 2023), improving emergency response strategies (Yuan et al. 2012), and supporting environmental sustainability initiatives (Zhang et al. 2017). Studying human mobility also helps track the spread of diseases, identify high–risk areas, and design effective containment strategies, as seen during the COVID–19 outbreak (Ahdika et al. 2023; Guardabascio et al. 2024). Recently, research on this topic has proliferated, with many studies proposing methods – typically validated through case studies – to forecast traffic using approaches from various disciplines, such as informatics, machine learning, statistics, and operations research (Fernández-Ares et al. 2017; Alam et al. 2019; Nagy and Simon 2021; Yuan et al. 2025). For example, Long Short–Term Memory (LSTM) networks (Wei and Zhao 2025; Fujiang et al. 2025), a type of recurrent neural network designed to capture long–range temporal dependencies through input, forget, and output gates, as well as other neural architectures such as Graph Neural Networks (GNNs) (Jin and Xu 2025), which learn from both node features and edge relationships in graph–structured data, have already been applied in similar contexts.

Overall, accurately forecasting the movement of people between urban areas is essential for policy makers to make informed and effective decisions (World Health Organization 2022). Various types of data are commonly used to proxy urban traffic flows. One of the most traditional sources of traffic data comes from roadside sensors and inductive loop detectors embedded in roads. These devices provide real–time information on traffic volume, speed, and occupancy, which is essential to assess traffic flow (Po et al. 2019). Another increasingly valuable source of data is GPS information collected from vehicles, smartphones, and ride–sharing services. GPS data offer detailed information on vehicle speeds, travel times, and routes, allowing for the modeling of traffic patterns across various times of day and geographic locations (Kan et al. 2019). Giorgini and Sartori (2016) analyzed vehicle and pedestrian movements from GPS and video recordings related to urban traffic in Torino, Marseille, and Venice. Platforms such as Twitter, Waze, and Google Maps have emerged as valuable sources of crowdsourced traffic data. In addition, social media offers real–time, user–generated reports on traffic conditions, accidents, and road closures (Clemente et al. 2024). These types of data enable rapid identification of unforeseen traffic disruptions, providing authorities with the ability to make adjustments in real time. Data from public transportation systems, including buses, trams, and metros, provide valuable insights into traffic demand and congestion. These systems typically generate large amounts of real–time data that can be used to track vehicle locations, travel times, and capacity utilization (Burzacchi et al. 2024).

In this work, we used mobile phone anonymized location data, which have become a powerful tool for forecasting traffic flows. By analyzing the movement patterns of individuals, mobile phone data can provide insights into the flow of people across cities, thereby helping in forecasting traffic trends and commuter behavior.

Mobile phone data are being increasingly adopted in research for a wide range of objectives: to produce dynamic information related to population density (Metulini and Carpita 2021), to analyze mobility patterns (Tettamanti and Varga 2014), to monitor the impact of

social and cultural events (Carpita and Simonetto 2014), to identify variability in the distribution of population in the neighborhoods of a large city (Mariotti et al. 2022), to examine the seasonality of second homes in a tourist area (Curci et al. 2022), and to measure the growth of remote working in sparsely populated areas (Manfredini et al. 2022). The use of mobile phone data is not new to the natural risk literature as well. For example, Balistrocchi et al. (2020) developed spatio–temporal dynamic maps of flood exposure using mobile phone crowding data.

In this study, we used hourly–based origin–destination mobile phone data provided by TIM S.p.A., one of the largest companies in the telecommunications sector, primarily operating in Italy and Brazil. Previous studies have attempted to address the issue of forecasting traffic flow using mobile phone data by proposing a time series modeling strategy based on Vector AutoRegression model with eXplanatories and Dynamic Harmonic Regression (VARX–DHR, hereafter) (Perazzini et al. 2023; Metulini and Carpita 2024). Another class of advanced models, such as Facebook Prophet and Neural Prophet (Taylor and Letham 2018; Triebe et al. 2021), have recently gained popularity for their effectiveness in forecasting data with complex seasonality and they have already been applied to the same mobile phone data by Perazzini and Metulini (2024) with results that are not completely satisfying.

Traffic flow time series are characterized by complex temporal dynamics that include serial correlation, multiple seasonalities, and non–linearity, meaning that they require robust predictive models capable of capturing such patterns. VARX–DHR is a trivariate model with a double autoregressive structure accounting for, respectively, daily and weekly serial dependence, a dynamic harmonic regression component to account for temporal periodicity, a set of calendar–related dummy covariates, and cross–correlation among flows modeled via the variance–covariance matrix. The forecast of the model is quite good (despite improvable); however, authors have found leptokurtic distributions of estimated residuals with heavy tails, which have been used in Carpita et al. (2024) to cluster urban areas in terms of similarities in the tail using a copula approach. Moreover, the model is not parsimonious due to the large amount of covariates, and it does not consider non–linearity. This leaves ample room for the development of simple models that also effectively consider extreme events (which, in our case, are very high or very low traffic).

Consider a time series of traffic flows and its multivariate distribution. Copula models allow the separation of a multivariate distribution into its margins (i.e., univariate distributions of each time period) and a dependence structure linking them. These models have proven to be quite effective at capturing extreme events and non–linearities, and they have recently been applied to the analysis of traffic flow data. Fang et al. (2022) proposed some copula models, such as Gaussian, Farlie–Gumbel–Morgenstern, Gumbel, Clayton, Frank, and Ali–Mikhail–Haq, to construct bivariate joint distributions of traffic variables. Cervellera et al. (2022) introduced a data–driven method for scenario generation based on copula models, through which the marginals of single input parameters can be chosen freely without altering the joint multivariate dependence structure of the inputs. Lu et al. (2023) introduced the use of a regular–vine copula combined with Latin hypercube sampling. However, standard copulas often prove to be too rigid in high–dimensional datasets (Brechmann and Schepsmeier 2013), such as those arising in traffic–flow time series. Vine copulas overcome this issue by assembling complex dependencies from flexible bivariate pairs. In this study, we applied a recent approach based on vine copulas – Vine–Transform AutoRegressive Moving-Average (ARMA) (hereafter, VT–ARMA), proposed by McNeil

(2021). This approach is based on the combination of vine transforms, which allowed us to map the data to a structure in which marginals are independent, and the ARMA component, which captures temporal autocorrelation within the vine framework. To the best of our knowledge, this approach has never been employed for traffic flow data. Summing up, in contrast to existing approaches, our methodology addresses a key gap in traffic flow forecasting: the limited ability of statistical and machine–learning models to jointly capture nonlinear marginal behavior, temporal dynamics, and multivariate dependence structures. The VT–ARMA copula framework separates and models these components explicitly, offering a flexible and interpretable alternative to standard techniques. The proposed approach delivers practical value by improving short–term predictive accuracy and enhancing explainability, two aspects that are crucial for operational use in urban mobility systems.

The primary objective of this study is to assess the forecasting performance of the VT–ARMA model on traffic data. To ensure that the stationarity assumption is met, the data were first pre–processed using Multiple Seasonal–Trend Decomposition with LOESS (MSTL) before applying the VT–ARMA model. In doing so, we compared them against the VARX–DHR and the Prophet model, which serve as the benchmarks. We assessed the forecasting accuracy of the model by implementing a blocked k–folds cross–validation strategy for time series and Rank–Graduation Accuracy (RGA), in the spirit of Babaei et al. (2025). Since we also aimed to identify the key components of copula models contributing to the model's performance, we employed Rank–Graduation Explainability (RGE). Furthermore, we conducted a diagnostic analysis of the model residuals to assess the adequacy of the model fit. This involved examining residual plots and applying statistical tests to detect issues such as normality and residual autocorrelation.

We applied and validated our methodological approach using real–world data from a case study based on a one–year length hourly time series of traffic flows, derived from mobile phone data for Cellatica, a representative Italian National Institute of Statistics (ISTAT) "Area di CEnsimento" (ACE) located in the province of Brescia. To demonstrate the generalizability of our results, we applied the method to three additional nearby ACEs and assessed its robustness across different temporal resolutions.

This paper is structured as follows. Section 2 describes the data, Sect. 3 introduces the methods, Sect. 4 describes the application of the methods to the case study, Sect. 5 presents the results, and Sect. 6 concludes the paper.

2 Data

The dataset employed in this analysis is based on origin–destination (OD) flows derived from mobile phone data. It spans one full year of observations, from September 1, 2020 to August 31, 2021, and covers the ACEs within the province of Brescia.

The data capture traffic flows, denoted as $flow_{ij,t}$, representing the number of SIM cards moving from ACE i to ACE j during the t–th hour of a day. These OD flows are computed by identifying SIM cards first detected by an antenna in ACE i and then – after at least five minutes – detected by an antenna in ACE j, all within the same one–hour interval. SIM card locations are recorded at fixed five–minute intervals, namely: [00–05), [05–10), , [55–60). Only the first position observed within each five–minute interval is retained for analysis. For example, consider a one–hour time window t corresponding to 8:00–8:59 AM on

January 1, 2021. Suppose a SIM card is first detected in ACE *i* between 8:00 and 8:04 AM, moves to ACE *j* and is detected there between 8:05 and 8:09 AM, and finally reaches ACE *z* within the same five–minute window. In this case, the flow is counted only as $flow_{iz,t}$ and not as $flow_{ij,t}$ or $flow_{jz,t}$. For each time interval *t*, the dataset includes a nonsymmetric square matrix of size $N \times N$, where $N = 235$ corresponds to the number of ACEs in the province of Brescia. Each row of the matrix represents the ACE of origin, whereas each column indicates the ACE of the destination. The diagonal elements capture internal flows – i.e., those where origin and destination are the same ACE. Thus, for each ACE *i* and time *t*, three types of flows are identifiable: inflows to *i*, outflows from *i*, and internal flows from *i* to *i*. In total, the database comprises $24 \times 364 = 8736$ hourly matrices, each of dimension 235×235.[1] The data only include SIM cards connected to the TIM network, including foreign SIMs using roaming. Two types of SIM cards are present: human SIMs (approximately 85% of the total) and machine–to–machine (M2M) SIMs (approximately 15%), which are used in automated devices. To avoid double counting, when a user possesses both a human SIM and one or more M2M SIMs, the analysis is restricted to human SIM cards only.

In this analysis, we restricted the attention to a single ACE, Cellatica (Latitude: 45.584, Longitude: 10.180, depicted in orange in Figure 1), which belongs to the Mandolossa region, a critical zone with flood episodes located in the northwest outskirt of Brescia. We chose

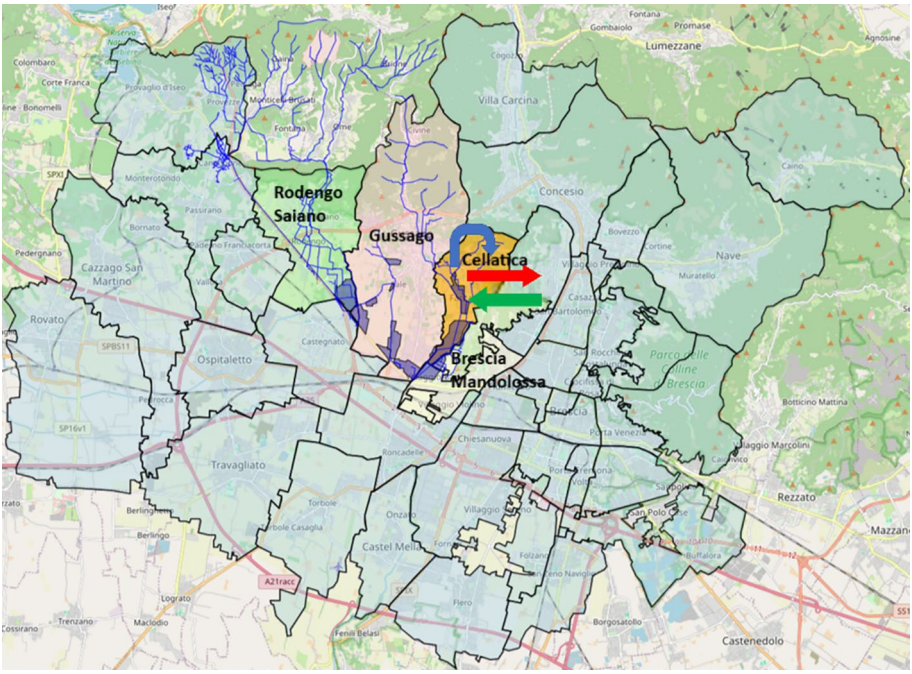


Fig. 1 Map of Mandolossa. Cellatica (orange colored), Gussago (light pink), Rodengo Saiano (light green), and Brescia Mandolossa (light yellow). The 10–year flood hazard map (blue area) along with the hydrographic network, obtained from Balistrocchi et al. (2020), and a stylized representation of outflows (red arrow), inflows (green arrow), and internal flows (blue arrow) from/to Cellatica to/from the 38 neighboring ACEs (in light blue)

[1] Data for September 2, 2020 were not collected.

Cellatica for three main reasons. The primary reason is that the area is subject to a high flood risk. As illustrated in Figure 1, a significant part of it falls within the 10–year flood hazard map depicted in blue. The second is that it constitutes a key node in the road network connecting to/from Brescia. The third reason is to facilitate methodological comparison, as previous studies based on the same mobile phone data have focused on the time series to/from Cellatica (Metulini and Carpita 2024).

We focused on three hourly time series: inf_t, out_t, and int_t, where:

$$inf_t = \sum_{j=1}^{n} flow_{ij,t},\ \text{i} = \text{Cellatica}, \tag{1}$$

$$out_t = \sum_{j=1}^{n} flow_{ji,t},\ \text{i} = \text{Cellatica}, \tag{2}$$

$$_t = flow_{ii,t},\ \text{i} = \text{Cellatica}, \tag{3}$$

and $n = 38$ is the number of neighboring ACEs identified in Metulini and Carpita (2024) as having strong traffic links to Cellatica, depicted in light blue in Figure 1.

As can be seen in Figure 2, which shows, as a sample, the time series of the first two weeks of February for Cellatica, the series presents a strong intra–day and intra–weekly pattern. The AutoCorrelation Functions (ACFs) presented in Figure 3 show the pattern of (linear) serial correlation for, from left to right, inflows, outflows, and internal flows. Together, these preliminary empirical patterns provide evidence of a complex temporal dependence, thereby supporting the suitability of our proposed approach.

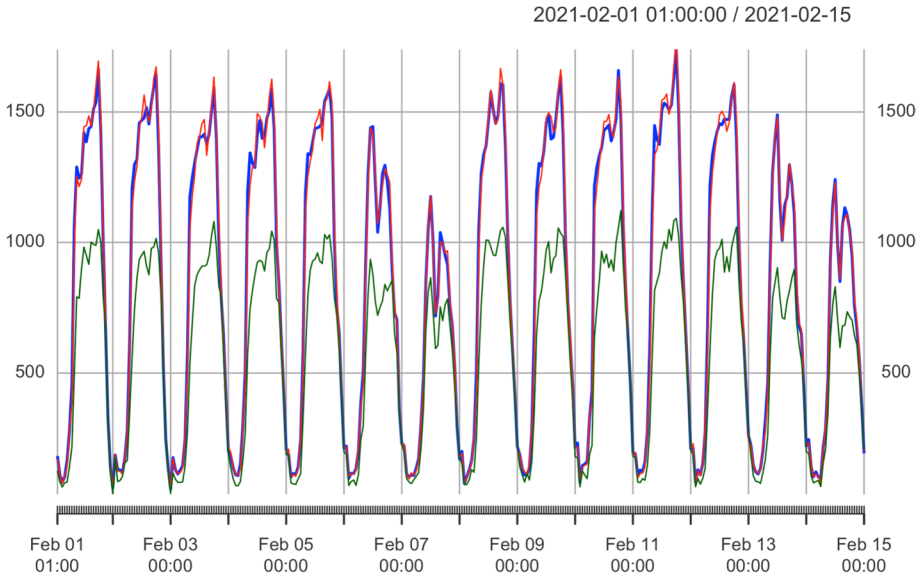
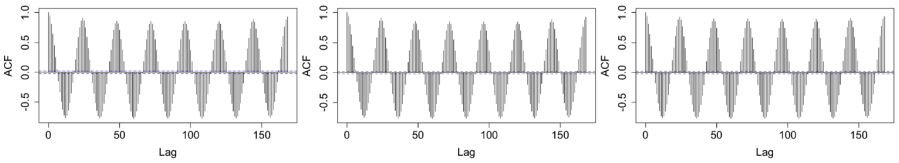


Fig. 2 Original data for the first two weeks of February in the ACE of Cellatica with outflows in blue, inflows in red, and internal flows in green



(a) Autocorrelation function of inflows (b) Autocorrelation function of outflows (c) Autocorrelation function of internal flows

Fig. 3 Autocorrelation functions (ACF) for the three time series of Cellatica

To ensure the robustness and generalizability of the findings, additional analyses were conducted on the other ACEs constituting the Mandolossa region, namely Rodengo Saiano (45.560, 10.107, depicted in light green in Figure 1), Gussago (45.584, 10.157, light pink), and Brescia Mandolossa (45.557, 10.178, light yellow).

3 Methods

Copula models provide a flexible statistical framework for modeling the dependence structure between random variables independently of their marginal distributions. By capturing nonlinear and tail dependencies, copulas allow the construction of multivariate distributions that better reflect complex relationships than traditional linear correlation measures. Hence, they are particularly useful for studying dependence among the components of a random vector.

Standard multivariate copulas, such as the Gaussian and exchangeable Archimedean copulas, often lack flexibility when modeling dependence among a large number of variables. As dimensionality increases, these models become either too rigid or require intricate parameterizations that are difficult to estimate and interpret (Brechmann and Schepsmeier 2013).

Vine copulas address this limitation by constructing multivariate dependence from bivariate building blocks. This approach provides greater flexibility and enables the modeling of complex, asymmetric, and tail–dependent relationships that traditional copulas cannot capture in high dimensions (Czado and Nagler 2022).

In this work, we adopted the VT–ARMA copula method proposed by McNeil (2021). This approach introduces v–transforms as a preprocessing step within the vine copula framework, thereby significantly improving the modeling of volatility and tail dependence in time series data, which are the key limitations of traditional copula–based methods in volatile, high–dimensional contexts.

VT–copula processes are stationary time series obtained by combining a v–transform with an ARMA process. As shown by McNeil (2021), these models capture nonlinear serial correlations that classical autoregressive models cannot identify. Thus, they can be viewed as copula–based dependence models that accommodate non–Gaussian marginals and nonlinear, non–Gaussian serial dependence, defining a tractable class of non–Gaussian ARMA processes.

Given $Y = \{Y_1, ..., Y_T\}$, a stationary time series of length T (where Y represents the variable *inf*, *out*, or *int*, with $T = 8712$), the vine–transform characterizes the relationship between the quantiles of Y and those of a predictable volatility proxy, such as $|Y|$, in which volatility appears through serial correlation. It is defined by:

$$\mathcal{V}_{\delta,\kappa,\zeta}(u) = \begin{cases} (1-u) - (1-\delta)\Psi\left(\frac{u}{\delta}\right), & u \leq \delta \\ u - \delta\Psi^{-1}\left(\frac{1-u}{1-\delta}\right), & u > \delta \end{cases} \tag{4}$$

where u is pseudo–copula data obtained from the standardized ranks of Y, δ is the vine–transform parameter of order 1, and Ψ is a continuous, strictly increasing distribution function on [0, 1], assumed to follow $\Psi(x) = \exp(-\kappa(-\ln x)^{\zeta})$ (McNeil 2021).

Here, $\mathcal{V}$ represents the relationship between the Probability Integral Transform (PIT) process (V_t) and the series PIT process ($\mathcal{U}_t$), given by $V_t = \mathcal{V}(\mathcal{U}_t) = |2\mathcal{U}_t - 1|$. The symmetric v–shaped function $\mathcal{V}(u) = |2u - 1|$ maps extreme values of $\mathcal{U}_t$ (close to 0 or 1) to high volatility levels ($V_t \approx 1$) and values near 0.5 to low volatility ($V_t \approx 0$). Setting $Z_t = \phi^{-1}(V_t)$ yields a normalized volatility proxy following a Gaussian ARMA process with mean 0 and variance 1. The variable Z_t affects only the dependence structure, leaving the marginal distribution of X_t unchanged.

The VT–copula framework can be intuitively viewed as a two–layer mechanism that separates what the data look like (their marginal distribution) from how they evolve (their temporal dependence). The v–transform acts as a lens that extracts volatility information from the series by mapping extremes to higher proxy values V_t. Transforming V_t into a Gaussian ARMA process Z_t captures serial dependence and volatility clustering without assuming linearity or normality. Finally, the vine–transform copula links these transformed variables pairwise, enabling flexible modeling of nonlinear, asymmetric, and tail–dependent relationships across multiple time series. Conceptually, this can be represented as

$$X_t \xrightarrow{F_X} \mathcal{U}_t \xrightarrow{\mathcal{V}} V_t \xrightarrow{\phi^{-1}} Z_t, \tag{5}$$

where each transformation isolates a different aspect of dependence: marginal behavior, volatility, and serial correlation.

In our application, we employed a first–order vine–transform with $\kappa = 1$ and $\zeta = 1$ which we denote as VT(1), or simply VT, corresponding to a uniform Cumulative Density Function (CDF) for Ψ and encompassing the symmetric case $\mathcal{V}_{0.5} = |2u - 1|$ (Bladt and McNeil 2022a):

$$\mathcal{V}_{\delta}(u) = \begin{cases} (\delta - u)/\delta, & u \leq \delta \\ (u - \delta)/(1 - \delta), & u > \delta \end{cases} \tag{6}$$

According to Bladt and McNeil (2022b), the most suitable configuration for time series is the drawable vine (d–vine), which ensures strict stationarity and translation–invariant dependence structures.

We began the modeling process by transforming the original data X into pseudo–copula data via standardized ranks. Using maximum likelihood estimation, we jointly estimated the vine–transform parameter δ and the ARMA parameters $\{\phi_i\}_{i=1}^{p}$ and $\{\theta_i\}_{i=1}^{q}$, where p and q denote the autoregressive and moving average orders, respectively. With these parameters,

we constructed the copula model and then applied an appropriate marginal distribution. Among possible choices (Normal, Student–t, Laplace), we selected the Student–t margin because it effectively captures heavy tails and extreme values, features that are characteristic of our data, as discussed in Sect. 4.

4 Empirical procedure

In this section we describe how original data undergoes data processing and how the methods are applied to processed data.

Our main analysis focused on the time series of Cellatica. In addition, the time series of the ACEs of Rodengo Saiano, Gussago, and Brescia Mandolossa were examined as comparative benchmarks to validate and contextualize the results. Accordingly, to support the generalizability of our results, we adopted the same empirical strategy employed for Cellatica.

First, since the observations for September 2, 2020 were missing in the original data, we decided to remove the first two days of the data (i.e., September 1 and September 2).

Because of the stationarity assumption of the vine–transform copula model, the first step in our application aims at obtaining a stationary series. We do so through a MSTL decomposition (Bandara et al. 2025), a recently proposed method to decompose a time series into trend, cycle, and multiple seasonal components. This decomposition model was chosen because it enables the separation of the time series into more than one seasonal component, making it particularly appropriate for our context, where the data display both daily and weekly seasonal patterns. This approach uses LOcally Estimated Scatterplot Smoothing (LOESS), a non–parametric method that locally fits data to capture complex trends in order to estimate time series components through a multistage decomposition process. First, LOESS estimates the trend component. Then, seasonal decomposition techniques identify and model the seasonal patterns of the data. Finally, residuals (also called “irregulars") are obtained by subtracting the trend and seasonal estimates from the original series, representing unexplained variation. Since our data showed seasonality patterns that are both daily and weekly, as seen in Figure 2, our MSTL can be specified through an additive decomposition model of the following form:

$$Y_t = T_t + C_t + S_{d,t} + S_{w,t} + I_t \qquad t = 1, ..., n \ , \tag{7}$$

where Y_t denotes the original time series at time t representing either inflows, outflows, or internal flows, T_t denotes the long–term trend, C_t denotes the business cycle, $S_{d,t}$ denotes the daily seasonality, $S_{w,t}$ denotes the weekly seasonality, and I_t the irregulars (or residual term). Decomposition was performed using the `mstl` function in the `R` package `forecast` (Hyndman et al. 2008).

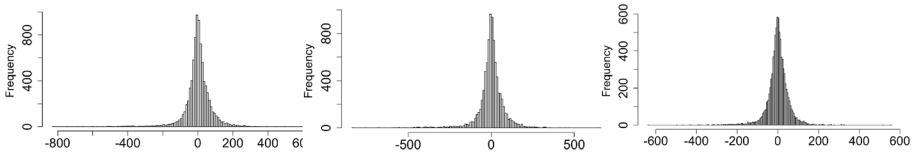
The irregular components were proposed as the dependent variable in the VT–ARMA model. To support this modeling decision, we assessed their stationarity properties by applying both the Augmented Dickey–Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests, using a lag order of $k = 24$ due to the strong daily seasonality of the time series.[2] Both tests supported the hypothesis that I_t is stationary, as presented for the

[2] We conducted additional tests with k = 168 (i.e., a full week). The findings are consistent with those reported in the main text and are available upon request.

Table 1 Results of the Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests applied to outflows, inflows, and internal flows, conducted with lag order of $k = 24$. The values refer to the ACE of Cellatica

		Out	Inf	Int
ADF				
	Statistic	-16.633	-16.732	-18.576
	p value	$< 0.01^*$	< 0.01	< 0.01
KPSS				
	Statistic	0.007	0.007	0.007
	p value	> 0.1	> 0.1	> 0.1

* The results of the p–values shown in the table are approximated as a consequence of the construction of the R functions employed



(a) Distribution of the irregulars for inflows (b) Distribution of the irregulars for outflows (c) Distribution of the irregulars for internal flows

Fig. 4 Distributions of the irregulars for the ACE of Cellatica

ACE of Cellatica in Table 1. Looking at the ADF statistic and the corresponding p–value, we can reject the null hypothesis that the series have a unit root, implying that they are not non–stationary. Moreover, based on the KPSS statistic and its p–value, we fail to reject the null hypothesis that the series are stationary.

Then, we considered the VT–copula models on the irregulars I_t with a vine–transform of order 1 and all possible combinations of AR and MA orders in the set $\{0, 1, 2\}$.

Using this approach, we first fit the vine–transform copula with AR and MA components to our pseudo–copula data u. Then, we fit the Student–t marginal distribution to the irregulars. Lastly, we fit the fitted copula and the margins to the irregulars.

To support the choice of the Student–t margins, we can refer to the distribution of the irregulars in Figure 4, which shows a normal distribution with heavy tails.

To implement the copula modeling, we applied the semiparametric estimation method developed by Genest et al. (1995) using the R package `tscopula` (McNeil and Bladt 2024).

To select the best model among all combinations of AR and MA orders, we conducted an analysis based on the Akaike Information Criterion (AIC). However, since the differences in AIC values were not substantial, all models were retained.

To assess the forecasting performance of the models, a blocked k–folds cross–validation for time series with a moving window, as depicted in Figure 5, was adopted. This cross–validation strategy is an adaptation of the proposed blocked k–folds cross–validation strategies for time series (Snijders 1988; Hyndman and Athanasopoulos 2021) and presents many similarities with the approach proposed in Metulini and Carpita (2024) and Perazzini et al. (2023).

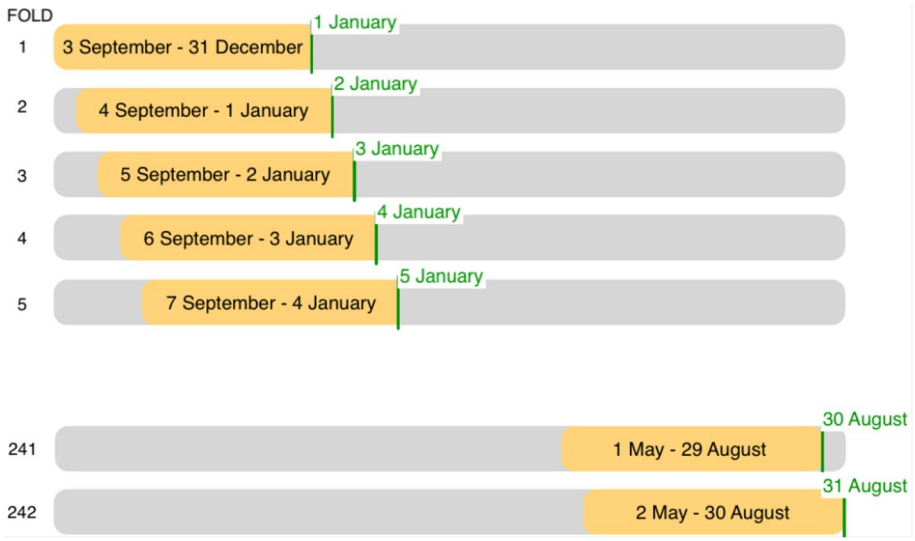


Fig. 5 Illustration of the blocked k–fold cross–validation approach adopted. The yellow time intervals represent the training periods, each with a sample size of 2902, whereas the green intervals correspond to the validation periods, each of length 24

According to our strategy, one day of observations was chosen as the validation set, and data from the previous four months were taken as the training set. Consequently, each validation set had sample size $n_v = 24$ intervals of one hour, and each training set had sample size $n_t = 2902$ (24 observations $\times$ 121 days -2 hours employed for the autoregressive lag term). To evaluate the performance consistently over all the days of the year, we replicated the analysis on different sets of training and validation samples, for a total of 242 different folds to be used for validation.[3]

To evaluate the model accuracy and the explainability of each model's component, the recently proposed rank–graduation box (RGB) measures (Babaei et al. 2025) were used. RGB measures are statistical tools that are used to assess and compare the fit of models involving both quantitative as well as ordered or ranked data. These measures evaluate the alignment between observed and predicted rankings. By letting Y be our original data, Y^* and Y^{**} be any two statistical distributions, $\hat{Y}$ be the predicted data, and $\hat{Y}^{(-X_k)}$ be the predicted data removing the covariate k, as well as letting

$$RG\cdot = \frac{1}{2} + \frac{1}{2}\frac{cov\left(Y^*, rank(Y^{**})\right)}{cov\left(Y^*, rank(Y^*)\right)} \tag{8}$$

be the general form for a rank–graduation box measure, the RGA was obtained by replacing Y^* with Y and Y^{**} with $\hat{Y}$, and the RGE was obtained by replacing Y^* with $\hat{Y}$ and Y^{**} with $\hat{Y}^{(-X_k)}$ in $1 - RG\cdot$. RGA values near 0 indicate poor predictions, values near 1 indicate accurate predictions, and a value of 0.5 suggests random predictions. An RGE close to

[3] Since the data were available for 363 days, 121 days (from September 3 to December 31) could not be validated because they are adopted for training.

1 indicates that the k_{th} predictor provides maximum explainability, and a value close to 0 means that it does not contribute.

It is important to note that the forecasted values are not directly comparable to the irregular components themselves, but rather to their standardized ranks. Therefore, we adopt a reconstruction procedure to transform the forecasted values back to a scale comparable with that of the original data. This step is essential to ensure proper application of performance evaluation metrics. The reconstruction procedure is detailed in Algorithm 1 in the Appendix. This algorithm reconstructs a time series observation by identifying where the copula–generated irregulars most closely match the standardized ranks; then, it retrieves the corresponding seasonal, trend, and remainder components, and combines them to produce the final reconstructed data point.

We compared the performance of the VT–ARMA model to that of the VARX–DHR model (Metulini and Carpita 2024; Perazzini et al. 2023), which for $t = 1, ..., T$, is defined as follows:

$$\mathrm{Y}_t = \boldsymbol{\nu} + \sum_{h_d=1}^{P_d} \mathrm{A}_{h_d} \mathrm{Y}_{t-24\times h_d} + \sum_{h_w=1}^{P_w} \mathrm{A}_{h_w} \mathrm{Y}_{t-168\times h_w} + \mathrm{B}\mathrm{x}_t + \boldsymbol{\epsilon}_t \tag{9}$$

where $\mathrm{Y}_t = [inf_t, out_t, int_t]'$, $\boldsymbol{\nu}$ is a constant vector of length 3, p_d and p_w are, respectively, the daily and the weekly autoregressive parameters, A_{h_d} and A_{h_w} are two 3×3 matrices of coefficients to be estimated, and $\boldsymbol{\epsilon}_t$ is the 3×1 vector of the error terms at time *t*. Moreover, x_t is the vector of the *l* exogenous variables at time *t*, and B is the $3 \times l$ matrix of coefficients of the exogenous variables such that Bx_t is a 3×1 vector.

To capture the seasonality, the r^{th} element (r = 1, 2, 3) of the vector Bx_t is defined as the following combination of daily and weekly functions:

$$\beta_0^{(r)} + \sum_{k_d=1}^{K_d} \left[\alpha_{k_d}^{(r)} s_{k_d}(t) + \gamma_{k_d}^{(r)} c_{k_d}(t)\right] + \sum_{k_w=1}^{K_w} \left[\alpha_{k_w}^{(r)} s_{k_w}(t) + \gamma_{k_w}^{(r)} c_{k_w}(t)\right] \tag{10}$$

$$s_{k_a}(t) = \sin\left(\frac{2\pi k_a t}{m_a}\right), \quad c_{k_a}(t) = \cos\left(\frac{2\pi k_a t}{m_a}\right), \quad a = d, w \tag{11}$$

where β_0 is a constant term, K_d and K_w are the optimal numbers of Fourier bases for the daily and the weekly pattern, respectively, α_k and γ_k are regression coefficients to be estimated, $m_w = 24 \times 7 = 168$ is the weekly seasonal period, and $m_d = 24$ is the daily seasonal period.

Specifically, for comparison purposes, the chosen model was VARX($p_d = 3,\ p_w = 4$) with a DHR($K_d = 7,\ K_w = 6$) component, as the one adopted by Perazzini et al. (2023). The list of exogenous variables included monthly dummies, to control for the possible presence of changes in average levels among months, and weekdays dummies, to control for the possible presence of changes in average levels among weekdays.

Note that the model assumes that the three dependent variables (i.e., *inf*, *out*, and *int*) at time *t* are related to each other. It is also worth noting that the model is trivariate; therefore

it allows for the estimation of inflows, outflows and internal flows together, differently from the VT–ARMA, in which each flow is estimated in a separate model.

Furthermore, another comparison was carried out with the Prophet model (Taylor and Letham 2018), a time series model developed by Facebook's Core Data Science team.[4] This model is designed to handle non–linear trends, irregular observations, missing values and outliers, as well as to identify change points in the data. Prophet is defined, for $t = 1, \ldots, T$, as:

$$Y_t = g(t) + s(t) + r(t) + \varepsilon_t, \tag{12}$$

where Y_t is the observed univariate time series at time t, representing either inflows, outflows, or internal flows, $g(t)$ models non–periodic changes, $s(t)$ models periodic changes using a Fourier expansion, and $r(t)$ includes additional explanatory variables. It is important to highlight that the Prophet model also incorporates an ARMA component as part of its additive framework. In our case, for a fair comparison with our copula models, non–periodic changes were modeled through a linear trend function (i.e., using growth = "linear" in the `prophet()` function from the `prophet R` package), periodic changes included both weekly and daily seasonality through a Fourier expansion (i.e., yearly.seasonality=FALSE), and additional regressors were those related to the ARMA component.

In light of our empirical choices, it is now worth highlighting that copula models are generally simpler and more parsimonious than the VARX–DHR framework. Specifically, in the VT–ARMA specification, only a limited number of parameters need to be estimated. In contrast, the VARX–DHR model requires a substantially larger set of parameters due to the inclusion of two autoregressive components of orders 3 and 4, along with 6 and 7 Fourier basis terms, to capture weekly and daily periodicities, respectively. Furthermore, as previously mentioned, the copula model is univariate, which adds to its simplicity. The Prophet model occupies a middle ground between the copula approach and the VARX–DHR framework in terms of complexity and parsimony. Similar to the copula model, Prophet remains univariate and therefore avoids the high–dimensional parameterization required in multivariate settings. Its structure, comprising a linear trend, a small number of seasonal components represented through Fourier terms, and a limited set of additive regressors, keeps the parameter space relatively compact.

According to the recorded execution times, generating validation–period forecasting for the outflows of Cellatica–run on a Quad–Core Intel Core i5 1.4 GHz processor with 8 GB of RAM–required 12.176 min for the most comprehensive copula specification (the VT–ARMA(2,2) model), 3.063 s for the VARX–DHR model, and 2.228 h for the Prophet model with an ARMA(2,2) structure. Based on these figures, the copula model can be considered an intermediate option between the two alternatives in terms of computational cost.

5 Results

In this section, we present the main findings of the study, focusing on the forecasting performance of the VT–ARMA models in comparison to the VARX–DHR and Prophet benchmarks. We begin by reporting the results related to the ACE of Cellatica. In Sect. 5.1, we

[4] https://facebook.github.io/prophet/.

present the outcomes of the blocked k–fold cross–validation, discussing the results related to the accuracy metrics from the approach of Babaei et al. (2025). In this section, we also go more in–depth in terms of evaluating the models' performance by identifying and discussing periods in which the model's performance deteriorates. Subsequently, in Sect. 5.2, we analyze the contribution of individual components in the copula models through the rank–graduation explainability framework. A diagnostic analysis of model residuals is also provided in Sect. 5.3, including visual inspections and statistical tests to detect potential issues, such as normality and residual autocorrelation. In Sect. 5.4, we present results from additional case studies (specifically those identified within the ACEs of Rodengo Saiano, Gussago, and Brescia Mandolossa) to further demonstrate the generalizability of our approach.

5.1 Forecasting performance

Figure 6 shows the RGA values associated with the copula, VARX–DHR, and Facebook Prophet, computed over the entire forecast period (January 1, 2021 to August 31, 2021) for the ACE of Cellatica. For comparison with our copula models, after observing that the VT–ARMA(0,1) specification delivers the best forecasting performance and that VT–ARMA(2,2) represents the most comprehensive formulation, we decided to estimate two Prophet models: one incorporating an ARMA(0,1) component and the other an ARMA(2,2) component. The VARX–DHR model's features fall outside this modeling choice, as its autoregressive structure is defined by two distinct AR components, one operating at the daily frequency and the other at the weekly frequency. We can see that the forecasting performance of the copula models is comparable to that of the VARX–DHR model and definitely better than that of the Facebook Prophet model. Moreover, the best model among the three was the VT–ARMA(0,1).

For a more robust evaluation of whether the copula model outperformed the benchmark models, we also conducted the Diebold–Mariano (DM) test (Diebold and Mariano 2002).

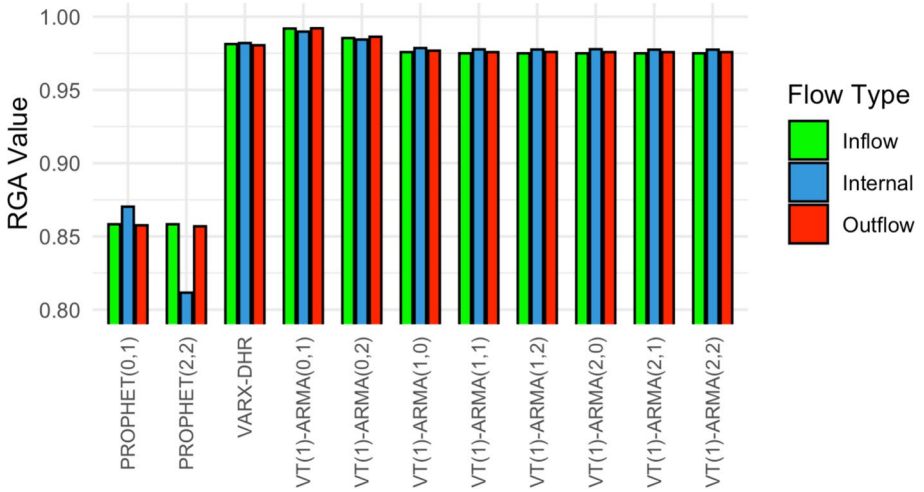


Fig. 6 Rank–graduation accuracy (RGA) values for the VARX–DHR, Facebook Prophet and all the copula models considered in the study, calculated on outflows, inflows, and internal flows. The results refer to the ACE of Cellatica

This test compares the errors for each pair of models. In our case, we compared the two copula models (with the ARMA(2,2) and ARMA(0,1) specifications) with the VARX–DHR model and the Prophet model (with the ARMA(2,2) and ARMA(0,1) specifications). We computed the DM test results using the `dm.test` function in the `forecast` package in `R`, setting the alternative hypothesis as "less", meaning that Model 1 has higher accuracy than Model 2. The results in Table 2 show that the copula model outperforms the benchmark models in 11 out of 12 cases, although only twice at the 5% significance level.

The Root Mean Square Error (RMSE) was also computed for each model. The results in Table 3 indicate that, based on this metric, both copula models outperformed the VARX–DHR and Prophet models, as they consistently achieved lower RMSE values.

Hereafter, we will go more in deep with the analysis by considering the VT–ARMA(2,2) and the VT–ARMA(0,1) models, which are, respectively, the augmented model (i.e., the one with the richer set of parameters) and the one with the highest RGA.

Figure 7a shows the original outflow time series in comparison with the forecasted one for the entire period of validation, which was from January 1, 2021 to August 31, 2021. The VT–ARMA(2,2) time series is depicted in purple, and the VT–ARMA(0,1) time series is depicted in orange. To provide a clearer view of short–term dynamics, Figure 7b focuses on the month of February. Analogous plots for inflow and internal flow are presented in Figures 12 and 13 in the appendix, with corresponding February zoom–ins.

It is worth noting that the models produced a limited number of negative values.[5] A summary of these values is presented in Table 4. It can be observed that the VT–ARMA(0,1) model performed better in terms of limiting the number of estimated negative values. Nevertheless, the number of negative values is very low in both cases. Fortunately, most of these negative values occur during the night and early morning hours, which are when the amount of traffic is generally the lowest, mitigating the impact of this issue on policy evaluation processes.

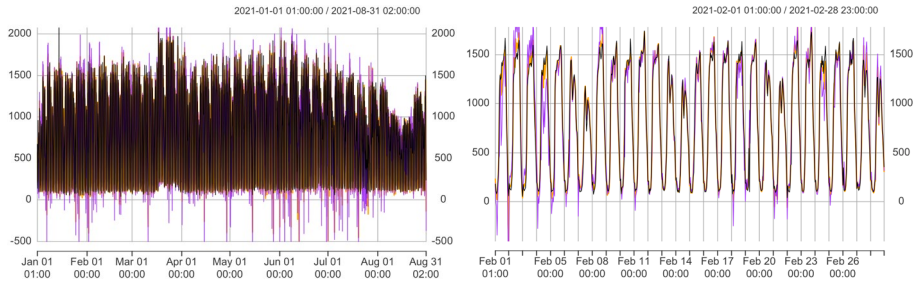
Table 2 Results of the Diebold–Mariano test comparing the performance of our VT–ARMA models with the VARX–DHR and Prophet benchmarks. The alternative hypothesis is set to "less", meaning that rejecting the null of equal accuracy implies that Model 1 achieves higher accuracy than Model 2. These results refer to the ACE of Cellatica. Outflows (Out), inflows (Inf), and internal flows (Int)

Flow	Model 1	Model 2	DM test	p value
Out	VT–ARMA(0,1)	VARX–DHR	−0.2741	0.3920
Out	VT–ARMA(2,2)	VARX–DHR	−0.3874	0.3492
Out	VT–ARMA(0,1)	Prophet–ARMA(0,1)	−1.2031	0.1145
Out	VT–ARMA(2,2)	Prophet–ARMA(2,2)	−1.3747	0.0846
Inf	VT–ARMA(0,1)	VARX–DHR	−0.3147	0.3765
Inf	VT–ARMA(2,2)	VARX–DHR	0.0668	0.5266
Inf	VT–ARMA(0,1)	Prophet–ARMA(0,1)	−0.9964	0.1596
Inf	VT–ARMA(2,2)	Prophet–ARMA(2,2)	−0.3430	0.3658
Int	VT–ARMA(0,1)	VARX–DHR	−1.9060	0.0284
Int	VT–ARMA(2,2)	VARX–DHR	0.6569	0.7444
Int	VT–ARMA(0,1)	Prophet–ARMA(0,1)	−2.5640	0.0052
Int	VT–ARMA(2,2)	Prophet–ARMA(2,2)	−0.4117	0.3403

[5] One possible cause of the negative values is that some components may have been reallocated incorrectly during the application of Algorithm 1. When the trend and seasonal components are summed with a mismatched irregular component due to a poor alignment in standardized ranks, it can lead to negative values. For instance, if the observation for May 22 is mistakenly matched with that of January 19 based on their ranks, the combined components may no longer correspond to a coherent time point, potentially resulting in a negative value.

Table 3 Root Mean Square Errors (RMSE) computed on the validation set for the VT–ARMA, VARX–DHR, and Prophet models. Results refer to the ACE of Cellatica. Outflows (Out), inflows (Inf), and internal flows (Int)

Flow	Model	RMSE
Out	VT–ARMA(0,1)	98.8135
Out	VT–ARMA(2,2)	181.0320
Out	Prophet–ARMA(0,1)	226.4623
Out	Prophet–ARMA(2,2)	225.7890
Out	VARX–DHR	218.8769
Inf	VT–ARMA(0,1)	100.9052
Inf	VT–ARMA(2,2)	185.2211
Inf	Prophet–ARMA(0,1)	400.2837
Inf	Prophet–ARMA(2,2)	400.4174
Inf	VARX–DHR	217.7054
Int	VT–ARMA(0,1)	73.0764
Int	VT–ARMA(2,2)	116.2417
Int	Prophet–ARMA(0,1)	251.5107
Int	Prophet–ARMA(2,2)	251.9090
Int	VARX–DHR	119.4482



(a) Full validation period. (b) Focus on the month of February.

Fig. 7 Forecasting results for the VT–ARMA(2,2) model (in purple) and VT–ARMA(0,1) (orange) with Student–t, compared with the original data (black), during the entire validation period (left) and with a focus on the month of February (right). Outflows of the ACE of Cellatica

Table 4 Estimated negative values for the forecasting period (January 1, 2021 to August 31, 2021) of length 5808 for both the VT–ARMA(2,2) and VT–ARMA(0,1) models. The results refer to the ACE of Cellatica. Outflows (Out), inflows (Inf), and internal flows (Int). Presented are both the total negative values and the percentage over the total number of forecasted values (%)

Model	Measure	Out	Inf	Int
VT–ARMA(2,2)	Negative values	157	157	125
	(%)	(2.70)	(2.70)	(2.15)
VT–ARMA(0,1)	Negative values	52	50	53
	(%)	(0.89)	(0.86)	(0.91)

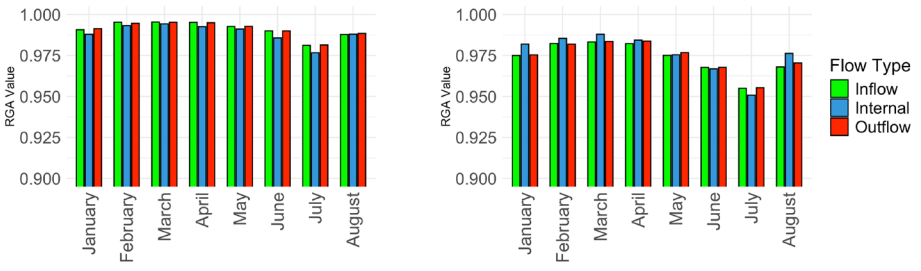
Although the model demonstrates strong overall accuracy across the entire forecast period, its performance may vary during specific times of the year. Understanding these variations is important, especially if the model is intended for predictive applications, as it allows users to identify periods that may be more challenging to forecast.

To further examine the temporal behavior of the model, we evaluated forecasting quality on a month–by–month basis. For each month, we aggregated the out–of–sample forecasts and computed the corresponding RGA by comparing them with the observed data. Figure 8 reports these monthly RGAs for both the best–performing specification, VT–ARMA(0,1), and the augmented model, VT–ARMA(2,2), allowing us to assess how forecast accuracy varies over time. The monthly RGA values in Figure 8a (VT–ARMA(0,1)) show consistently high forecasting accuracy across all months and flow types, with only a modest decline during July, likely reflecting atypical summer mobility patterns. In the VT–ARMA(2,2) model, the RGA values (reported in Figure 8b) exhibit slightly more variability compared with those of VT–ARMA(0,1). However, both models maintain RGA values well above 0.95 in nearly all months, confirming strong reliability. Overall, the results indicate that the proposed approach performs robustly throughout different periods, with only marginal seasonal sensitivity.

Moreover, we computed the RGA separately for each of the 242 forecast days using both the VT–ARMA(2,2) and VT–ARMA(0,1) models. The five days with the lowest RGA values, listed from lowest to highest, were January 6, which is a national holiday (Epiphany), May 14, March 10, July 4 (a Sunday), and July 19. In Figure 9, we display the original time series for January, 6 national holiday (Epiphany) versus the values estimated using VT–ARMA(2,2) and VT–ARMA(0,1). We can observe that for January 6th, especially in the afternoon, the predicted behavior differs from the actual data. The VT–ARMA(2,2) model (purple) shows a noticeable drop in the early hours and again during the night, which is not present in the actual data. In contrast, the VT–ARMA(0,1) model (orange) tends to overestimate both peaks and troughs, producing forecasts with more pronounced fluctuations than those observed in the real series.

5.2 Features explainability

Estimating the explainability of each feature or group of features in statistical models helps in identifying which components significantly contribute to the model's predictive power. This generally improves model interpretability, supports feature selection, and guides simplification without lowering performance. Therefore, we assessed the contribution of each component of the VT–ARMA(2,2) model using the rank–graduation explainability. We measured the explainability of the copula part, the AR component, and the MA component



(a) VT–ARMA(0,1) (b) VT–ARMA(2,2)

Fig. 8 Monthly calculations for the RGA for VT–ARMA models, obtained by aggregating the calculations of the RGA month by month

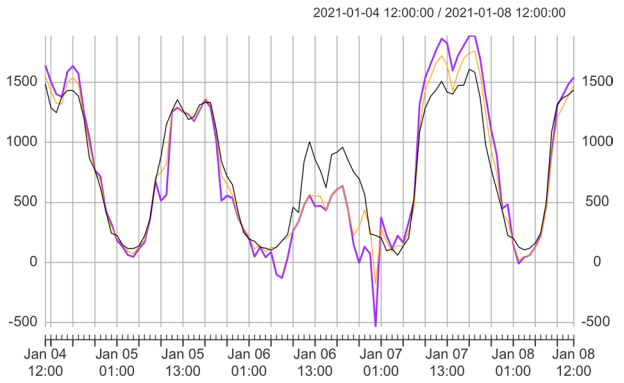


Fig. 9 Representation of the outflows corresponding to the window from midday on January 4 to midday on January 8, including the national holiday of Epiphany. Original data (black) against data forecasted using the VT–ARMA(2,2) model (in purple) and the VT–ARMA(0,1) model (in orange), with Student–t margins. The values refer to the ACE of Cellatica

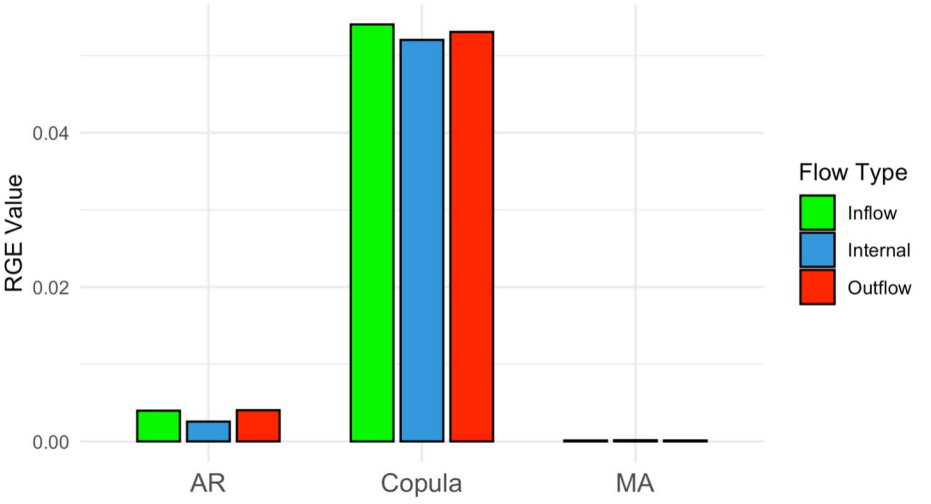


Fig. 10 Results of rank–graduation explainability (RGE) for each component of the VT–ARMA(2,2) model: copula (VT), the autoregressive component (AR), and the moving average component (MA). ACE of Cellatica

for the entire validation period. In doing so, we preferred the VT–ARMA(2,2) model, since it incorporates a larger number of terms in each component.

A summary of results is presented in Figure 10. It is worth noting that the explainability of each component of the model does not necessarily sum up to one. This means that a value such as that of the copula component, which is averaging approximately 0.05, should be regarded as high in this context.

We can note that the AR and MA components contribute very little to the model's explainability, especially the moving average part. This suggests that removing these components individually has minimal impact on the model's forecasting performance. In contrast, the copula component shows significantly higher explainability. This implies that excluding the copula part brings the model's performance closer to that of a random model, underscor-

ing the critical role of the copula component, which surpasses even the importance of the ARMA elements.

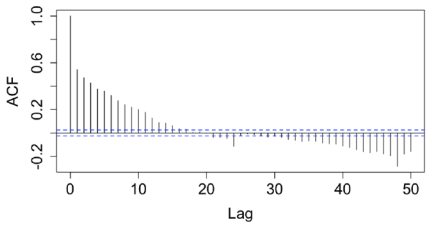
5.3 Residuals' diagnostics

To assess the adequacy of our VT–copula model, we performed residual diagnostics aimed at verifying two key assumptions: the normality of residuals with zero mean and the absence of significant serial autocorrelation. These checks are essential to ensure that residuals behave as random noise, indicating a well–specified model. Normality was evaluated through statistical tests and visual inspections, whereas autocorrelation was assessed using ACF and partial ACF (PACF). The residual diagnostics for outflows are shown in Figure 11, and similar results for internal flows and inflows are shown and discussed, respectively, in Figure 14 and Figure 15 in the Appendix. Looking at the ACF and PACF shown in Figure 11a and 11b, we can see that there is a lingering autoregressive dynamic, more accentuated for the VT–ARMA(2,2) model. However, this is significantly lower in comparison to the autocorrelation function of the original data, shown in Figure 3. Looking at the partial autocorrelation function in Figure 11c and 11d, there is still a limited partial autocorrelation; however, we can see that it is under the significance level for most of the observations. In the QQ–plots shown in Figure 11e and Figure 11f, the distribution of the residuals is basically normal, with a discrepancy in the tails, which is accentuated more in the VT–ARMA(2,2) model. The same thing is then shown in the histograms of the residuals, which show a normal behavior, with heavy tails.

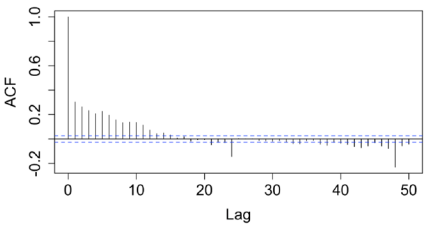
We finally examined whether the extremely high (low) estimated residuals correspond to extremely high (low) forecasted values. If this holds, it implies that the residuals are correlated with the forecasted values, suggesting that the model is misspecified. In general, this correspondence does not occur. According to the VT–ARMA(0,1) model (the one with the best performance), the five hourly periods with the highest estimated residuals are 7 a.m. on February 19, 7 p.m. on May 14, 7 p.m. on July 20, 5 p.m. on July 26, and 10 a.m. on July 28. Conversely, the five hourly periods with the lowest estimated residuals correspond to 3 p.m. on January 24, 4 p.m. on April 14, 1 a.m. on June 13, 8 p.m. on June 30, and 1 a.m. on July 1. These periods do not strictly coincide with the hourly intervals corresponding to the highest (lowest) estimated flow values.

5.4 Additional case studies

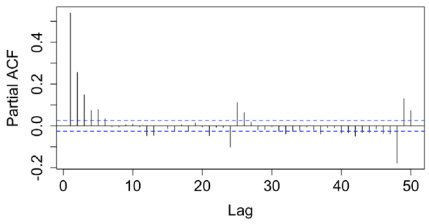
To examine whether the findings obtained for Cellatica are specific to this area or hold more generally, we conducted a validation exercise on three additional ACEs in close proximity (Rodengo Saiano, Gussago, and Brescia Mandolossa, as depicted in Figure 1). The full empirical strategy, comprising data preprocessing, MSTL decomposition, stationarity testing of the irregulars, VT–ARMA copula modeling and the choice of the order of the ARMA part, cross–validation, and reconstruction of forecasted values, was reproduced identically for each area. Similarly, the same benchmark specifications (VARX–DHR and Prophet) and evaluation metrics (Rank graduation box measures) were employed. By maintaining methodological coherence across all locations, this procedure provides a rigorous assessment



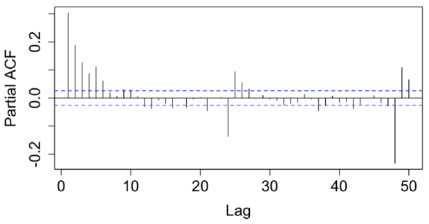
(a) ACF plot for VT–ARMA(2,2)



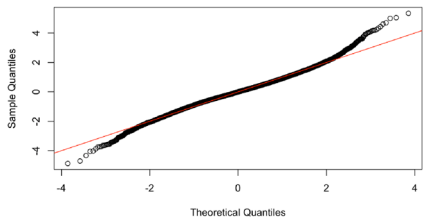
(b) ACF plot for VT–ARMA(0,1)



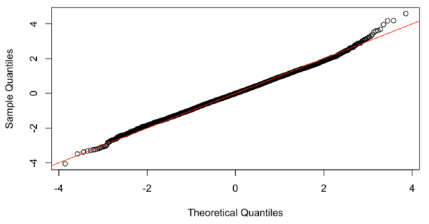
(c) PACF plot for VT–ARMA(2,2)



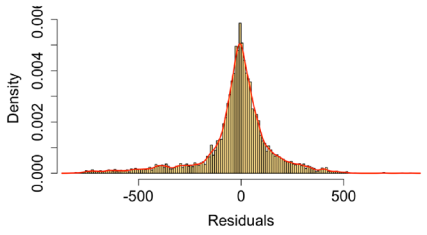
(d) PACF plot for VT–ARMA(0,1)



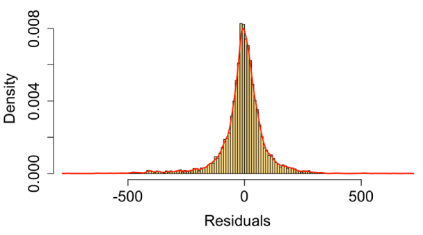
(e) QQ–plot for VT–ARMA(2,2)



(f) QQ–plot for VT–ARMA(0,1)



(g) Histogram of residuals for VT–ARMA(2,2)



(h) Histogram of residuals for VT–ARMA(0,1)

Fig. 11 Residual diagnostics for VT–ARMA(2,2) (left) and VT–ARMA(0,1) (right), calculated on the outflows for the ACE of Cellatica

of the robustness and generalizability of our modeling framework. The results reported in Table 5 show that the conclusions drawn for Cellatica extend consistently across the additional areas considered. In all three ACEs, namely Rodengo Saiano, Gussago, and Brescia Mandolossa, the VT–ARMA copula models maintain a clear performance advantage over both benchmark approaches. Although the magnitude of the improvement varies by loca-

Table 5 Rank–graduation accuracy (RGA) obtained for VARX–DHR, Facebook Prophet, and all the copula models considered for the ACEs of Rodengo Saiano, Gussago and Brescia Mandolossa. Outflows (Out), inflows (Inf), and internal flows (Int)

ACE	Model	Out	Inf	Int
Rodengo Saiano	VT–ARMA(2,2)	0.9750	0.9758	0.9883
	VT–ARMA(0,1)	0.9916	0.9914	0.9719
	Prophet–ARMA(2,2)	0.8427	0.8549	0.8663
	Prophet–ARMA(0,1)	0.8433	0.8545	0.8670
	VARX–DHR	0.9814	0.9842	0.9823
Gussago	VT–ARMA(2,2)	0.9746	0.9746	0.9717
	VT–ARMA(0,1)	0.9914	0.9921	0.9895
	Prophet–ARMA(2,2)	0.8458	0.8536	0.8599
	Prophet–ARMA(0,1)	0.8461	0.8538	0.8606
	VARX–DHR	0.9818	0.9827	0.9841
Brescia Mandolossa	VT–ARMA(2,2)	0.9695	0.9683	0.9678
	VT–ARMA(0,1)	0.9921	0.9920	0.9891
	Prophet–ARMA(2,2)	0.8552	0.8557	0.8537
	Prophet–ARMA(0,1)	0.8549	0.8550	0.8534
	VARX–DHR	0.9810	0.9813	0.9820

Table 6 Rank–graduation explainability (RGE) for each component of the VT–ARMA(2,2) model: copula (VT), the autoregressive component (AR), and the moving average component (MA). These results refer to the ACEs of Rodengo Saiano, Gussago, and Brescia Mandolossa. Outflows (Out), inflows (Inf), and internal flows (Int)

ACE	Component	Out	Inf	Int
Rodengo Saiano	VT	0.0547	0.0531	0.0574
	AR	0.0039	0.0037	0.0032
	MA	0.0002	0.0002	0.0002
Gussago	VT	0.0544	0.0542	0.0621
	AR	0.0045	0.0046	0.0039
	MA	0.0002	0.0002	0.0002
Brescia Mandolossa	VT	0.0575	0.0587	0.0635
	AR	0.0063	0.0064	0.0046
	MA	0.0001	0.0001	0.0003

tion, the copula–based specifications systematically achieved higher rank–graduation accuracy, indicating a more accurate reconstruction of the distributional features of traffic flows. Moreover, the results reported in Table 6 related to the rank–graduation explainability show that, across all ACEs, the vine–transform copula component (VT) consistently accounted for the largest share of explainability, confirming its central role in capturing dependence structures in traffic flows. The autoregressive component (AR) provided a smaller but stable contribution, reflecting short–term temporal dynamics, whereas the moving average component (MA) remains marginal across all flow types. This pattern was remarkably consistent across Rodengo Saiano, Gussago, and Brescia Mandolossa, indicating that the decomposition behaved robustly in areas with different traffic intensities. Overall, this multi–area validation confirmed that the proposed approach generalizes well across different ACEs and is suitable for deployment in diverse segments of the urban mobility network; however, testing areas with more heterogeneous characteristics would further strengthen this claim. The consistent pattern of the results shown in Tables 5 and 6 reinforces the empirical evidence supporting the vine–transform ARMA copula framework, indicating that its performance gains are systematic rather than specific to a single location.

6 Conclusions and future developments

This paper aimed to develop and evaluate a statistical framework that is flexible and robust to forecast urban traffic flows using mobile phone data. Specifically, we applied the Vine–Transform Autoregressive Moving-Average (VT–ARMA) copula model to some case studies of traffic flows recorded in the Mandolossa region, in the province of Brescia. Models that are accurate, interpretable, and parsimonious are needed more and more to capture the complex temporal and non–linear dependencies observed in mobility data. This need is tied to contexts in which mobility planning is crucial for emergency response policies linked to environmental issues.

The methodological strategy proposed in this paper involved incorporating ARMA components into a vine–transformed copula modeling by combining copula theory and traditional time series modeling. To ensure stationarity, the model was applied to the irregulars of the time series obtained by decomposing the original data through the MSTL process and isolating the trends and seasonalities of each observation. Out–of–sample forecasting was conducted using a blocked k–fold cross–validation with a moving window approach, and model performance was evaluated in terms of accuracy and explainability using the rank–graduation box framework. The model was compared with the VARX–DHR benchmark model, and it was revealed that the VT–ARMA copula model achieved comparable and, in some cases, superior predictive accuracy, even though it was simpler and required fewer parameters. Explainability analysis revealed that the AR and MA components had lower predictive value, whereas the copula component played a central role in capturing traffic flow dynamics, particularly by modeling nonlinearities and tail dependencies, which are often overlooked by traditional models.

The findings presented herein carry various practical implications for urban planners and traffic authorities. Short–term forecasts derived from the proposed model can play a crucial role in smarter mobility planning, such as adaptive traffic control, real–time traffic management, and investments in infrastructure. Logistics operators and public–transport agencies can exploit the model's forecasting capabilities to optimize routing, scheduling, and resource allocation in critical settings. Moreover, the ability to model events of high and low traffic helps in planning the responsiveness of emergency services, which is crucial in flood–prone and congested urban areas, such as Mandolossa. Finally, better traffic forecasting can aid in the creation of policies aimed at reducing emissions by redirecting traffic in cases of traffic congestion, or promoting alternative transportation during predicted traffic peaks.

This study presents some limitations that suggest directions for future research. Although the approach has been applied to multiple ACEs, all of them are located in a specific region, which may limit external validity and generalizability to particularly different urban contexts. Future work could extend the framework to other cities and incorporate heterogeneous mobility data sources.

Another limitation of this type of modeling is the impossibility of expanding this approach to the multivariate framework. It could be useful in jointly modeling more flows at the same time. Future work may extend this methodological framework by evaluating dynamic copulas in the multivariate modeling of interconnected areas. By considering mul-

tivariate copula–GARCH models, such as the Dynamic Conditional Correlation GARCH (DCC–GARCH) or the Baba–Engle–Kraft–Kroner GARCH (BEKK–GARCH) with copulas, it becomes possible to capture both time–varying volatility and nonlinear dependencies between locations. This would allow for a more nuanced representation of traffic volatility, which has thus far been overlooked in the current analysis. Another important aspect that warrants further investigation concerns the heavy tails observed in the residuals' distributions. Despite the efficient performance of the current model, some extreme residual values persist and may be more appropriately captured by incorporating extreme value distributions, such as the Generalized Extreme Value (GEV) or the Generalized Pareto Distribution (GPD).

From a computational viewpoint, the vine–transform ARMA copula framework can become demanding as dimensionality grows, making real–time forecasting challenging. In fact, estimating vine structures and dependence parameters may introduce latency in large networks. Future work on parallelization and/or simplified vine configurations can reduce runtime, making real–time implementation feasible.

Algorithm to reconstruct the data using the copula model and MSTL components

Algorithm 1 Reconstruction of the data using copula model and MSTL components

Input:
  $strank\_rem$: Standardized ranks of the irregulars
  $copula\_model\_rem$: Irregulars obtained from the copula model
  $seasonal\_comp$: Seasonal components (weekly, daily)
  $trend\_comp$: Trend component
  rem_comp : Original irregulars component
for each element i in $strank\_rem$ **do**
  $diff_i \leftarrow copula\_model\_rem[i] - strank\_rem[i]$ ▷ Compute the difference between copula model and standardized ranks
  $abs\_diff_i \leftarrow |diff_i|$ ▷ Compute the absolute difference
end for
$min\_diff\_idx \leftarrow$ index of $\min(abs\_diff)$ ▷ Find index with minimum absolute difference
Retrieve the components for the observation with minimum difference:
$seasonal\_comp \leftarrow$ extract seasonal components for observation
  at index min_diff_idx
$trend\_comp \leftarrow$ extract trend component for observation
  at index min_diff_idx
$rem\_comp \leftarrow$ extract original remainder component for
  observation at index min_diff_idx
Compute the final result:
$final\_result \leftarrow rem\_comp + seasonal\_comp + trend\_comp$ ▷ Combine components to reconstruct the data
Output: $final\_result$ ▷ The reconstructed data

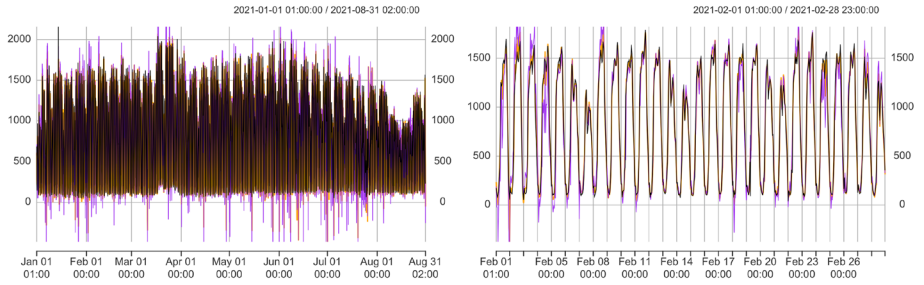


Fig. 12 Results of VT(1)–ARMA(2,2) (in purple) and VT(1)–ARMA(0,1) (orange) with Student–t, compared with the original inflow data (black), for the entire validation period (left) and with a focus on February (right). Inflows of the ACE of Cellatica

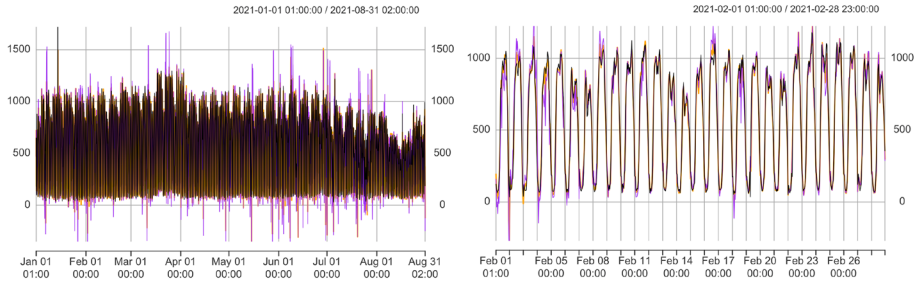


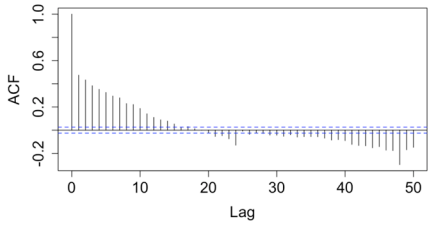
Fig. 13 Results of VT–ARMA(2,2) (in purple) and VT–ARMA(0,1) (orange) with Student–t, compared with the original internal flow data (black), as the whole validation period (left) and with focus on February (right). Internal flows of the ACE of Cellatica

Forecasting results for inflows and internal flows for the ACE of Cellatica

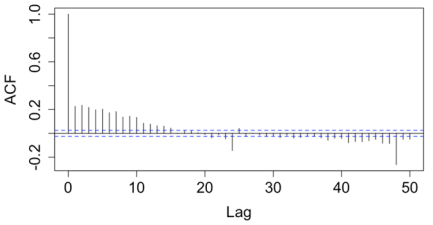
In this appendix, we report the results of the chosen models, VT–ARMA(2,2) and VT–ARMA(0,1), in terms of forecasting performance for inflows (Figure 12) and for internal flows (Figure 13) in the ACE of Cellatica. The results show that, similar to the results found by using data regarding outflows, as shown in Figures 7a and 7b, the VT–ARMA(0,1) outperformed the VT–ARMA(2,2) model in terms of how the forecasted data follows the trends of the original data.

Residual diagnostics for inflows and internal flows for the ACE of Cellatica

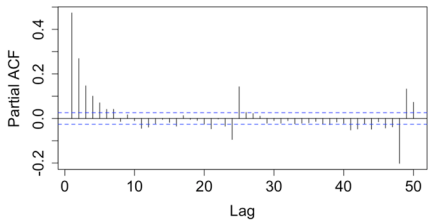
In this appendix, we report the diagnostics on the model's residuals for inflows (in Figure 15) and internal flows (in Figure 14) in the ACE of Cellatica, similar to what was presented in Figure 11 for the case of outflows. Moreover, in these cases, we can underline how the



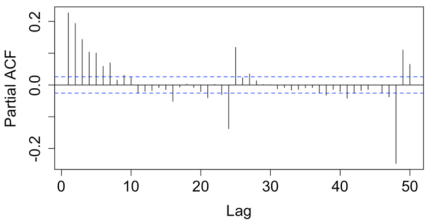
(a) ACF plot for VT–ARMA(2,2)



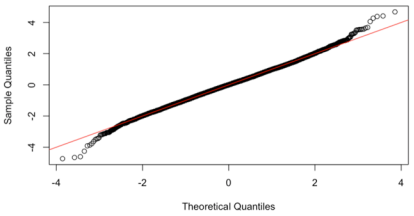
(b) ACF plot for VT–ARMA(0,1)



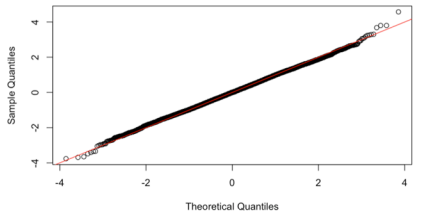
(c) PACF plot for VT–ARMA(2,2)



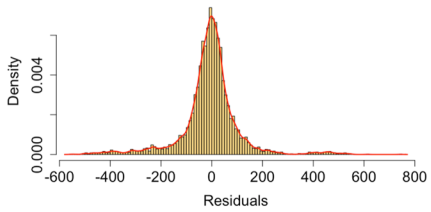
(d) PACF plot for VT–ARMA(0,1)



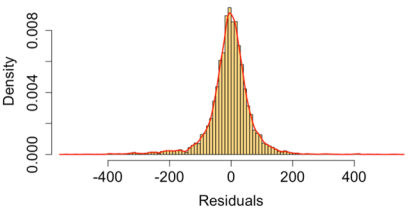
(e) QQ–plot for VT–ARMA(2,2)



(f) QQ–plot for VT–ARMA(0,1)



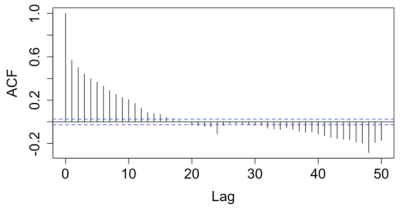
(g) Histogram of residuals for VT–ARMA(2,2)



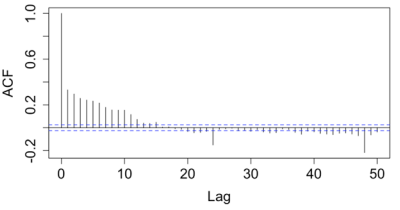
(h) Histogram of residuals for VT–ARMA(0,1)

Fig. 14 Residual diagnostics for VT–ARMA(2,2) model (left) and VT–ARMA(0,1) model (right) calculated on the internal flow data. The values refer to the ACE of Cellatica

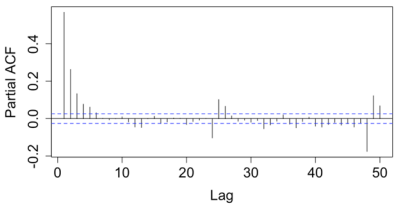
residuals related to the VT–ARMA(0,1) model follow a trend that behaves more normally than the residuals related to the VT–ARMA(2,2) model. Overall, in both inflow and internal flow cases, a moderate degree of residual serial autocorrelation can be observed, which is consistent with the pattern identified in outflows. The distribution of residuals appears approximately symmetric and normally distributed; however, deviations in the right and left tails are present and warrant further investigation in future research.



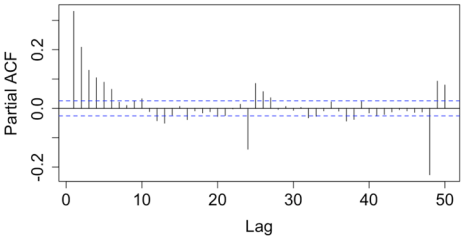
(a) ACF plot for VT–ARMA(2,2)



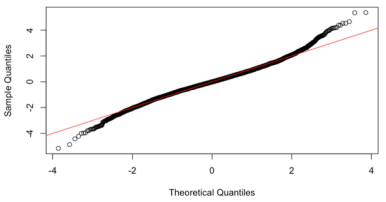
(b) ACF plot for VT–ARMA(0,1)



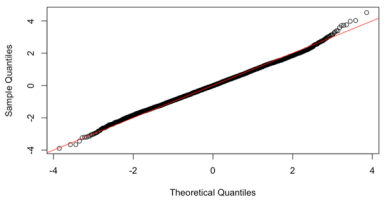
(c) PACF plot for VT–ARMA(2,2)



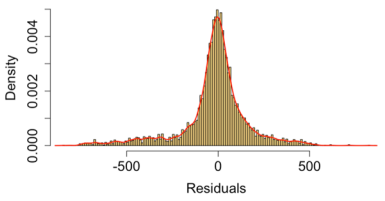
(d) PACF plot for VT–ARMA(0,1)



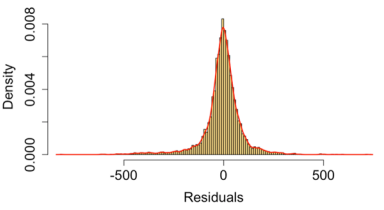
(e) QQ–plot for VT–ARMA(2,2)



(f) QQ–plot for VT–ARMA(0,1)



(g) Histogram of residuals for VT–ARMA(2,2)



(h) Histogram of residuals for VT–ARMA(0,1)

Fig. 15 Residual diagnostics for VT–ARMA(2,2) model (left) and VT–ARMA(0,1) model (right) calculated on inflow data. The values refer to the ACE of Cellatica

Author contributions CRediT author statement: SSG contributed to conceptualization, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing – original draft, writing – review and editing; RM contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, writing – review and editing.

Funding This contribution has been developed for the project "Study of mobile phone siGNals for the evalUation of the interconnections between Mobility and the environment in Lombardia (SIGNUM)" CUP: F53D23010910001– PRIN 2022 PNRR M4C2 – financed by the European Union – Next Generation EU (DD MUR n. 1409 del 14/09/2022).

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest all authors declare that they have no conflict of interest.

References

Alam, I., Farid, D.M., Rossetti, R.J.: The prediction of traffic flow with regression analysis. In: Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, vol. 2, pp. 661–671. Springer (2019). https://doi.org/10.1007/978-981-13-1498-8_58

Ahdika, A., Primandari, A.H., Adlin, F.N.: Considering the temporal interdependence of human mobility and covid-19 concerning Indonesia's large-scale social distancing policies. Qual. Quant. **57**(3), 2791–2810 (2023). https://doi.org/10.1007/s11135-022-01497-4

Babaei, G., Giudici, P., Raffinetti, E.: A rank graduation box for safe AI. Expert Syst. Appl. **259**(125239), 10 (2025). https://doi.org/10.1016/j.eswa.2024.125239

Bandara, K., Hyndman, R.J., Bergmeir, C.: MSTL: a seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. Int. J. Oper. Res. **52**(1), 79–98 (2025). https://doi.org/10.48550/arXiv.2107.13462

Bladt, M., McNeil, A.J.: Time series copula models using d-vines and v-transforms. Econom. Stat. **24**(27–48), 10 (2022). https://doi.org/10.1016/j.ecosta.2021.07.004

Bladt, M., McNeil, A.J.: Time series models with infinite-order partial copula dependence. Depend. Model. **10**(1), 87–107 (2022). https://doi.org/10.1515/demo-2022-0105

Balistrocchi, M., Metulini, R., Carpita, M., Ranzi, R.: Dynamic maps of human exposure to floods based on mobile phone data. Nat. Hazard. **20**(12), 3485–3500 (2020). https://doi.org/10.5194/nhess-20-3485-2020

Brechmann, E.C., Schepsmeier, U.: Modeling dependence with c-and d-vine copulas: the R package CDVine. J. Stat. Softw. **52**, 1–27 (2013)

Burzacchi, A., Urbano, V.M., Arena, M., Azzone, G., Secchi, P., Vantini, S.: Spatio-temporal analysis of public transportation undercrowding: leveraging APC data for a comprehensive evaluation of usage rates. (2024). https://doi.org/10.48550/arXiv.2410.12618

Clemente, G.P., Della Corte, F., Zappa, D.: Hierarchical spatial network models for road accident risk assessment. Ann. Oper. Res. **208**, 1–36 (2024). https://doi.org/10.1007/s10479-024-06049-7

Carpita, M., De Luca, G., Metulini, R., Zuccolotto, P.: Traffic flows time series in a flood-prone area: modeling and clustering on extreme values with a spatial constraint. Stoch. Env. Res. Risk Assess. **38**(8), 3109–3125 (2024). https://doi.org/10.1007/s00477-024-02735-x

Curci, F., Kërçuku, A., Zanfi, F., Novak, C., et al.: Permanent and seasonal human presence in the coastal settlements of Lecce. An analysis using mobile phone tracking data. TEMA **2**(57–71), 10 (2022). https://doi.org/10.6093/1970-9870/8914

Cervellera, C., Maccio, D., Rebora, F.: Copula-based scenario generation for urban traffic models. Expert Syst. Appl. **210**(118389), 10 (2022). https://doi.org/10.1016/j.eswa.2022.118389

Czado, C., Nagler, T.: Vine copula based modeling. Ann. Rev. Stat. Appl. **9**(1), 453–477 (2022)

Carpita, M., Simonetto, A.: Big data to monitor big social events: analysing the mobile phone signals in the Brescia smart city. Electron. J. Appl. Stat. Anal.: Decis. Support Syst. Serv. Eval. **5**(1), 31–41 (2014). https://doi.org/10.1285/i2037-3627v5n1p31

Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. J. Bus. Econ. Stat. **20**(1), 134–144 (2002)

Fernández-Ares, A., Mora, A., Arenas, M.G., García-Sánchez, P., Romero, G., Rivas, V., Castillo, P.A., Merelo, J.: Studying real traffic and mobility scenarios for a smart city using a new monitoring and tracking system. Futur. Gener. Comput. Syst. **76**(163–179), 10 (2017). https://doi.org/10.1016/j.future.2016.11.021

Fujiang, Y., Yangrui, F., Xiaohuan, B., Zhen, T., Chunhong, Y., Yankang, L.: A decomposition-driven hybrid framework based on STL for accurate traffic flow forecasting. (2025). arXiv:2510.23668v1

Fang, Z., Zhu, S., Fu, X., Liu, F., Huang, H., Tang, J.: Multivariate analysis of traffic flow using copula-based model at an isolated road intersection. Phys. A **599**(127431), 10 (2022). https://doi.org/10.1016/j.physa.2022.127431

Guardabascio, B., Brogi, F., Benassi, F.: Measuring human mobility in times of trouble: an investigation of the mobility of European populations during covid-19 using big data. Qual. Quant. **58**(6), 5181–5199 (2024). https://doi.org/10.1007/s11135-023-01678-9

Genest, C., Ghoudi, K., Rivest, L.P.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika **82**(3), 543–552 (1995). https://doi.org/10.2307/2337532

Giorgini, B., Sartori, M.: Human mobility world lines on urban topologies. Qual. Quant. **50**(4), 1817–1831 (2016). https://doi.org/10.1007/s11135-015-0237-6

Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice, 3rd OTexts (OTexts.com/fpp3), , Melbourne, Australia (2021). Accessed 20 Mar 2025

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F.: Forecast: Forecasting Functions for Time Series and Linear Models. (2008). R package version 8.23.0 https://pkg.robjhyndman.com/forecast/

Jin, B., Xu, X.: China commodity price index (CCPI) forecasting via the neural network. Int. J. Financ. Eng. **12**(03), 2550003 (2025). https://doi.org/10.1142/S2424786325500033

Kan, Z., Tang, L., Kwan, M.-P., Ren, C., Liu, D., Li, Q.: Traffic congestion analysis at the turn level using taxis' GPS trajectory data. Comput. Environ. Urban Syst. **74**(229–243), 10 (2019). https://doi.org/10.1016/j.compenvurbsys.2018.11.007

Lu, H., Sun, D., Hao, J.: Random traffic flow simulation of heavy vehicles based on r-vine copula model and improved Latin hypercube sampling method. Sensors **23**(5), 2795 (2023). https://doi.org/10.3390/s23052795

McNeil, A.J., Bladt, M.: Tscopula: Time Series Copula Models. (2024). R package version 0.3.9 https://cran.r-project.org/web/packages/tscopula/index.html

Metulini, R., Carpita, M.: A spatio-temporal indicator for city users based on mobile phone signals and administrative data. Soc. Indic. Res. **156**(2), 761–781 (2021). https://doi.org/10.1007/s11205-020-02355-2

Metulini, R., Carpita, M.: Modeling and forecasting traffic flows with mobile phone big data in flooding risk areas to support a data-driven decision making. Ann. Oper. Res. **342**(1629–1654), 10 (2024). https://doi.org/10.1007/s11205-020-02355-2

McNeil, A.J.: Modelling volatile time series with v-transforms and copulas. Risks **9**(1), 14 (2021). https://doi.org/10.3390/risks9010014

Mariotti, I., Giavarini, V., Rossi, F., Akhavan, M.: Exploring the 15-minute city and near working in Milan using mobile phone data. TeMA-J. Land Use Mobil. Environ. **2**(39–56), 10 (2022). https://doi.org/10.6093/1970-9870/9309

Manfredini, F., Lanza, G., Curci, F., et al.: Mobile phone traffic data for territorial research. Opportunities and challenges for urban sensing and territorial fragilities analysis. TEMA **2**(9–23), 10 (2022). https://doi.org/10.6093/1970-9870/8892

Nagy, A.M., Simon, V.: Improving traffic prediction using congestion propagation patterns in smart cities. Adv. Eng. Inform. **50**(101343), 10 (2021). https://doi.org/10.1016/j.aei.2021.101343

Perazzini, S., Metulini, R.: Exploring urban mobility patterns in lombardia through advanced analysis of mobile phone data. In: Scientific Meeting of the Italian Statistical Society, pp. 422–427. Springer (2024). https://doi.org/10.1007/978-3-031-64350-7_71

Perazzini, S., Metulini, R., Carpita, M.: Integration of flows and signals data from mobile phone network for statistical analyses of traffic in a flooding risk area. Socioecon. Plann. Sci. **90**(101747), 10 (2023). https://doi.org/10.1016/j.seps.2023.101747

Po, L., Rollo, F., Bachechi, C., Corni, A.: From sensors data to urban traffic flow analysis. In: 2019 IEEE International Smart Cities Conference (ISC2), pp. 478–485. IEEE (2019). https://doi.org/10.1109/ISC246665.2019.9071639

Snijders, T.A.: On cross-validation for predictor evaluation in time series. In: On Model Uncertainty and Its Statistical Implications, vol. 307, pp. 56–69. Springer, Berlin Heidelberg (1988). https://doi.org/10.1007/978-3-642-61564-1_4

Shahriari, S., Sisson, S.A., Rashidi, T.: Copula ARMA-GARCH modelling of spatially and temporally correlated time series data for transportation planning use. Transp. Res. Part C: Emerg. Technol. **146**(103969), 10 (2023). https://doi.org/10.1016/j.trc.2022.103969

Tao, X., Cheng, L., Zhang, R., Chan, W., Chao, H., Qin, J.: Towards green innovation in smart cities: Leveraging traffic flow prediction with machine learning algorithms for sustainable transportation systems. Sustainability **16**(1), 251 (2023). https://doi.org/10.3390/su16010251

Triebe, O., Hewamalage, H., Pilyugina, P., Laptev, N., Bergmeir, C., Rajagopal, R.: Neuralprophet: explainable forecasting at scale. arXiv preprint arXiv:2111.15397 (2021) https://doi.org/10.48550/arXiv.2111.15397

Taylor, S.J., Letham, B.: Forecasting at scale. Am. Stat. **72**(1), 37–45 (2018). https://doi.org/10.1080/00031305.2017.1380080

Tettamanti, T., Varga, I.: Mobile phone location area based traffic flow estimation in urban road traffic. Adv. Civ. Environ. Eng. **1**(1), 1–15 (2014)

Urban planning, design and management approaches to building resilience – an evidence review: first report on protecting environments and health by building urban resilience. Technical report, Copenhagen: WHO Regional Office for Europe (2022)

Wei, L., Zhao, J.: Hybrid CNN-multivariate LSTM model for accurate short-term electricity price forecasting and energy system optimization. J. Circuits Syst. Comput. **34**(09), 2550198 (2025). https://doi.org/10.1142/S0218126625501981

Yuan, F., Fan, Y., Bing, X., Tian, Z., Yuan, C., Li, Y.: Traffic flow forecasting, STL decomposition, hybrid model, LSTM, ARIMA, XGBoost, intelligent transportation systems. arXiv preprint arXiv:2510.23668 (2025)

Yuan, J., Zheng, Y., Xie, X.: Discovering regions of different functions in a city using human mobility and pois. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 186–194 (2012). https://doi.org/10.1145/2339530.2339561

Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (1) (2017). https://doi.org/10.48550/arXiv.1610.00081