Computers & Industrial Engineering 217 (2026) 112066

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A non-homogeneous negative binomial autoregressive model for forecasting port truck arrivals

Vicente Valenzuela-Riquelme [a], Marcel Favereau [b,c,*], Carlos A. Monardes-Concha [d]

[a] Universidad Andrés Bello, Departament of Engineering Sciences, Antonio Varas 880, Providencia, 7500971, Santiago, Chile
[b] Universidad de los Andes, Chile, Facultad de Ingeniería y Ciencias Aplicadas, Monseñor Álvaro del Portillo 12455, Las Condes, 8150513, Santiago, Chile
[c] Center for Advanced Transportation, Logistics, and Economic Competitiveness (CATLEC), ANID/CIN 250061, Santiago, Chile
[d] School of Engineering, Universidad Católica del Norte, Larrondo 1281, Coquimbo, 1781421, Coquimbo Region, Chile

ARTICLE INFO

Keywords:
Port logistics
Truck arrivals
Non-homogeneous negative binomial process
Autoregressive model
Counting processes
Stochastic models

ABSTRACT

Accurate forecasting of daily truck arrivals is crucial for optimizing port logistics operations. Truck arrivals are typically modeled as a counting process exhibiting two complex features: serial correlation and nonstationarity. Autoregressive models are commonly employed to address serial correlation, but they often assume stationarity, which limits their ability to capture non-stationary dynamics. Conversely, the non-homogeneous Poisson process is commonly used to model non-stationarity in arrival processes. Yet it fails to account for serial correlation and is insufficient to capture the overdispersion frequently observed in real-world counting processes. To overcome these limitations, we propose a novel forecasting model that integrates an autoregressive framework with a non-homogeneous negative binomial distribution to model daily truck arrivals, capturing serial autocorrelation, nonstationarity, and overdispersion. The proposed approach generates predictions using an autoregressive model, which are then transformed into realizations of the non-homogeneous negative binomial distribution via the inverse transform method. Based on real-world daily truck arrival data at a major Chilean port, the proposed model is evaluated against established benchmark models. The results show that the proposed model achieves competitive point-forecast accuracy while providing substantially more informative and better-calibrated predictive distributions, particularly in terms of prediction-interval sharpness and coverage balance. The broader implications for modeling and practical management are also discussed.

1. Introduction

Global trade has expanded rapidly over recent decades, surpassing 11 billion tons of goods transported by sea in 2022. This sustained growth has strengthened the role of ports as strategic nodes in global logistics networks, supporting international connectivity and the efficient movement of goods (Monios et al., 2018). Yet the increasing scale and complexity of maritime flows have intensified operational pressures on port systems, producing recurrent challenges such as congestion, reduced operational efficiency, and heightened environmental impacts (Wilmsmeier & Sánchez, 2017). These pressures translate directly into increased logistics costs and negative externalities—including higher pollutant emissions and deterioration in the quality of life for communities near port areas (Sultana et al., 2018).

To mitigate these issues, intermodal logistics chains have been promoted as a means of integrating maritime and land transport through coordinated "door-to-door" systems that reduce transit times, operational costs, and emissions (Martínez-López & Chica, 2020). However, their implementation remains limited by coordination gaps among stakeholders and by technological and infrastructural constraints (Raza et al., 2020; Wilmsmeier & Monios, 2015). These persistent barriers underscore the need for predictive tools that anticipate logistical variability and support proactive resource allocation.

The challenge is particularly salient in Latin America, where more than 90% of international trade is conducted by sea (Wilmsmeier & Monios, 2015). Operational bottlenecks in land-based reception — especially those arising from high variability in daily truck arrivals — generate inefficiencies and increased logistics costs (Rodriguez et al., 2021). The absence of quantitative forecasting tools limits ports and terminal operators' ability to anticipate demand conditions and coordinate resources effectively.

In Chile, ports such as San Antonio and Valparaíso play a central role in national and regional trade. In 2022, San Antonio handled more than 22 million tons of cargo, while Valparaíso exceeded 10 million tons, consolidating their roles as critical gateways for domestic

* Correspondence to: Monseñor Álvaro del Portillo 12455, Las Condes, 8150513, Santiago, Chile.
E-mail address: mjfavereau@uandes.cl (M. Favereau).

https://doi.org/10.1016/j.cie.2026.112066
Received 5 August 2025; Received in revised form 18 March 2026; Accepted 26 April 2026
Available online 28 April 2026
0360-8352/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

and international markets. Their proximity to densely populated urban areas exacerbates operational pressures, as limited capacity for expansion and recurrent road-access congestion compound daily variability in truck arrivals—a central driver of operational uncertainty (Brooks & Wilmsmeier, 2017; Martínez-López & Chica, 2020; Wilmsmeier & Sánchez, 2017).

Recent research has explored machine-learning and data-driven approaches to forecasting truck arrivals in logistics environments. For example, Li et al. (2022) develop a big-data framework for modeling categorized truck arrivals in ports, demonstrating how high-frequency operational information can improve short-term predictions. Similarly, Van der Spoel et al. (2016) analyze predictive analytics for estimating arrival times in a European distribution center, emphasizing the value of contextual data for granular operational planning. Reviews on truck appointment scheduling (Abdelmagid et al., 2021; Gracia et al., 2025) further indicate that many existing models assume exogenous arrival patterns or rely on forecasting tools that lack interpretability and do not provide probabilistic representations of count uncertainty. Collectively, these works reveal a methodological gap in settings where overdispersion, temporal dependence, and uncertainty quantification must be modeled explicitly. The approach proposed in this paper addresses this gap by developing an interpretable, distribution-based framework that captures nonstationarity, serial correlation, and overdispersion, and that produces probabilistic forecasts suitable for tactical decision-making.

Modeling discrete events such as daily truck arrivals is central to logistics and operations management, yet such processes typically exhibit heteroscedasticity, temporal dependence, and overdispersion—features that challenge standard modeling techniques (Lütkepohl, 2005). Classical non-homogeneous Poisson processes (NHPP) have therefore been widely employed to represent time-varying counting processes, with applications ranging from unscheduled clinic arrivals (Alexopoulos et al., 2008) and automotive warranty claims (Majeske, 2007) to acoustic monitoring (Guarnaccia et al., 2016), environmental forecasting (Achcar et al., 2010), air traffic estimation (Lancia & Lulli, 2020), public transportation (Kieu, 2018; Kieu & Cai, 2018), and epidemiological analysis (Dincer et al., 2022).

When serial correlation and dynamic variability are present, Integer-Valued Generalized Autoregressive Conditional Heteroscedasticity (INGARCH) models (Ferland et al., 2006) offer an alternative by jointly capturing temporal dependence and time-varying dispersion. A broad set of extensions — such as generalized Poisson INGARCH (GP-INGARCH) (Zhu, 2012a), zero-inflated and structurally adaptive versions (Chen & Lee, 2016), and empirical comparisons favoring INGARCH over ARIMA in low-count and overdispersed contexts (Quddus, 2008) — illustrate their versatility. Nevertheless, many real-world arrival processes display complex features such as irregular dispersion, regime shifts, and nonlinear dependencies, motivating continued exploration of flexible yet interpretable approaches.

In settings with substantial overdispersion, negative binomial (NB) models provide an attractive alternative to Poisson-based formulations. Their effectiveness has been documented in epidemic time-series modeling (Davis & Wu, 2009) and in high-variability environments such as air traffic (Omer & Hussian, 2023). More advanced NB autoregressive structures — including TNBAR (Liu et al., 2018), SEHNBAR (Liu et al., 2019), stochastic-intensity NBAR (Gouriéroux & Lu, 2018), and Bayesian ACNB (Chen et al., 2016) — capture richer dynamic behaviors and have found applications in epidemiology (Gning et al., 2022) and early outbreak detection (Sparks et al., 2011).

From an economic perspective, improved predictability in port truck arrivals enhances supply-chain productivity and competitiveness by reducing delays, congestion, and logistics costs (Álvaro et al., 2019; Chen et al., 2011; Islam & Olsen, 2014). Social and environmental benefits also arise from improved truck-flow management, including reduced emissions, improved road safety, and enhanced coexistence between ports and neighboring communities (Brooks & Wilmsmeier, 2017; Buldeo Rai et al., 2017; D'Amico et al., 2021; Kalahasthi et al., 2022).

Although NHPP-based and autoregressive variants capture important dynamic features, several works document their limitations when data exhibit pronounced overdispersion or more complex temporal dependence structures (Cameron & Trivedi, 2013; Fokianos et al., 2009; Gouriéroux & Lu, 2018; Majeske, 2007). To bridge these gaps, this research develops a non-homogeneous autoregressive negative binomial framework for daily port truck arrivals forecasting. It integrates an NB distribution with an autoregressive structure to capture non-stationarity, temporal dependence, trend, seasonality, and overdispersion—phenomena intrinsic to port-logistics flows. The proposed framework aims to support tactical planning (day-to-week staffing and congestion risk-anticipation), rather than intra-day terminal control, and is replicable and adaptable to other logistics settings with similar characteristics.

Building on the above, the contributions of this research are:

1. A structured probabilistic forecasting framework — the Non-Homogeneous Autoregressive Negative Binomial Process (NH-ARNBP) — for port truck arrival time series that exhibit temporal dependence, non-stationarity, and overdispersion in port-logistics settings.
2. A transformation-based simulation and mapping procedure that generates coherent probabilistic forecasts on the count domain via inverse transform sampling, while separating short-run temporal dependence from monthly varying distributional heterogeneity.
3. Empirical evidence from a real-world case study showing that the proposed framework improves point forecasting and yields more informative prediction intervals than Poisson-based alternatives, also providing a competitive uncertainty–accuracy balance when compared with an observation-driven GP-INGARCH benchmark.

The remainder of this paper is organized as follows: Section 2 introduces the theoretical foundations and modeling approach; Section 3 presents the computational experiments and comparative evaluation; and Sections 4 and 5 discuss managerial implications and conclusions.

2. Proposed methodology

In this section, we present the proposed stochastic modeling framework for daily truck arrivals at a port terminal. We first recall the time-series and count-process tools that underpin our construction, and then introduce the non-homogeneous autoregressive Poisson process (NH-ARPP) and the non-homogeneous autoregressive negative binomial process (NH-ARNBP).

2.1. Preliminaries

Let $\{Y_t\}_{t\in\mathcal{T}}$ denote a time series of daily truck arrivals. Following classical time-series analysis (Box et al., 2015; Hamilton, 1994; Hyndman & Athanasopoulos, 2018), we represent Y_t through an additive decomposition

$$Y_t = T_t + S_t + r_t, \tag{1}$$

where T_t denotes the long-term trend, S_t captures recurrent seasonal patterns (e.g., annual or monthly effects), and r_t is a residual component that concentrates the short-term stochastic variability.

The residual component r_t in (1) is typically modeled using an autoregressive structure. An autoregressive model of order P, denoted AR(P), is defined as

$$r_t = \beta_0 + \sum_{i=1}^{P} \beta_i r_{t-i} + \varepsilon_t, \tag{2}$$

where r_t is the value of the series at time t, $\boldsymbol{\beta}$ are the model coefficients, and ε_t is a white-noise term, typically Gaussian, i.e., $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ for all $t \in \mathcal{T}$ (Hamilton, 1994; Law, 2014). The order P is commonly selected via information criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC).

To model non-stationarity in the arrival process, we consider a non-homogeneous Poisson process (NHPP) as a building block (Lewis & Shedler, 1976; Ross, 2014). Let $N(t)$ be the number of events observed in $[0, t]$ and $\lambda(t)$ the time-varying intensity function. The probability of observing k events in $[0, t]$ is

$$\mathbb{P}\big(N(t) = k\big) = \frac{\Lambda(t)^k e^{-\Lambda(t)}}{k!}, \qquad \Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u, \tag{3}$$

where $\Lambda(t)$ is the cumulative intensity function. This representation allows the event rate to vary over time while preserving independent increments.

Finally, we use the classical inverse transform method to map continuous random variables to a target discrete distribution (Law, 2014; Ripley, 1987). Given a cumulative distribution function (CDF) F_X and a uniform variable $U \sim \mathcal{U}(0, 1)$, a realization X following F_X can be generated as

$$X = F_X^{-1}(U). \tag{4}$$

In our context, this mechanism transforms autoregressive forecasts in a standardized space into coherent discrete counts by applying the inverse CDF of an appropriate count distribution (Poisson or negative binomial).

2.2. Non-Homogeneous Autoregressive Poisson process (NH-ARPP)

We first introduce the NH-ARPP, designed to model discrete-time arrival series while accounting for both nonstationarity and serial correlation. The methodology follows a sequence of transformations and statistical modeling steps formulated to preserve the structural information and intrinsic behavior of the original data.

We begin with the observed series $\{Y_t\}_{t\in\mathcal{T}}$, composed of integer values representing the daily count of events on day $t \in \mathcal{T}$. Due to heterogeneity, trend and seasonality patterns, and the discrete nature of the variable, direct fitting of traditional autoregressive models to $\{Y_t\}$ may be inadequate or statistically inefficient. Therefore, the first step involves standardizing the original series based on

$$U_t = \frac{Y_t}{\max\{\boldsymbol{Y}\} + 1} \quad \forall t \in \mathcal{T}, \tag{5}$$

where $\{U_t\}$ is the resulting standardized series, ensuring that each value lies within the interval $[0, 1]$ while preserving the relative behavior of $\{Y_t\}$. This transformation facilitates numerical handling and supports the subsequent recovery of the original count domain for prediction purposes.

Next, the standardized series $\{U_t\}$ is subjected to a logit transformation

$$L_t = \ln\left(\frac{U_t}{1 - U_t}\right) \quad \forall t \in \mathcal{T}, \tag{6}$$

where $L_t \in \mathbb{R}$ for each t. The purpose of this transformation is to project the data onto the real line, eliminating the restrictions of the unit interval and yielding a continuous random variable suitable for linear autoregressive modeling.

On the transformed series $\{L_t\}$, we fit a linear trend T_t using linear regression on t and a seasonal component S_t using harmonic regression with sine–cosine functions to model recurrent periodic patterns. The residual series $\{r_t\}$, obtained after removing these two components from $\{L_t\}$, corresponds to a stationary series on which an AR(P) model of the form (2) is fitted. The optimal order P and the statistical significance of the coefficients $\boldsymbol{\beta}$ are determined using AIC and BIC, ensuring parsimony and model accuracy.

Through this procedure, we capture the serial correlation present in the original series $\{Y_t\}$. To represent nonstationarity in the arrival intensity, we estimate a monthly parameter λ_m associated with each month $m \in M$ directly from the historical series $\{Y_t\}$, using the maximum likelihood estimator (MLE) for the Poisson distribution (Law, 2014)

$$\hat{\lambda}_m = \frac{1}{n_m} \sum_{t\in\mathcal{T}_m} Y_t, \tag{7}$$

where n_m is the number of observations in month m (with $\mathcal{T}_m$ denoting the set of days corresponding to month m). These monthly empirical parameters are subsequently used to simulate future counts.

Using the AR model fitted on the residual series $\{r_t\}$, multiple future daily trajectories are generated based on different realizations of the white-noise sequence $\boldsymbol{\varepsilon}$ in (2). The trend T_t and seasonal component S_t are then added to reconstruct the projected series on the logit scale $\{\hat{L}_t\}$. Once these future simulations are obtained on the logit scale, the inverse logit transformation

$$\hat{U}_t = \frac{e^{\hat{L}_t}}{1 + e^{\hat{L}_t}}, \tag{8}$$

is applied to return to the unit interval, where $\{\hat{U}_t\}$ represents the forecast values in the standardized scale.

Finally, using the inverse transform method, parameterized monthly with the rates $\hat{\lambda}_m$ previously estimated via MLE, we recover the discrete counts (in the original scale) by

$$\hat{Y}_t = F_{\mathrm{P}}^{-1}(\hat{U}_t | \hat{\lambda}_m), \tag{9}$$

where F_{P}^{-1} denotes the inverse CDF of the Poisson distribution with rate $\hat{\lambda}_m$. This procedure ensures that forecasts $\{\hat{Y}_t\}$ are consistent with the discrete nature of the data and reflect the variability observed in the original series, thereby enabling a coherent evaluation of the predictive performance of the NH-ARPP model.

To complement the above description, Fig. 1 summarizes the methodology, highlighting each step. Panel (a) shows the original series $\{Y_t\}$, composed of integer-valued observations at each period t. In (b), the data are standardized to the $[0, 1]$ interval to obtain $\{U_t\}$, preserving the relative behavior of $\{Y_t\}$. In (c), the logit transformation is applied to $\{U_t\}$, generating the series $\{L_t\}$ defined on the real line and enabling AR(P) modeling and forecast simulation. In (d), the transformed trajectories are additively decomposed into trend (T_t), seasonality (S_t), and residual (r_t), upon which the AR model is fitted. In (e), these forecasts are transformed back to the $[0, 1]$ domain via the inverse logit function to obtain $\{\hat{U}_t\}$. Finally, in (f), the values $\{\hat{U}_t\}$ are mapped to the original discrete scale through the inverse transform of the Poisson distribution, parameterized by the estimated monthly rate $\hat{\lambda}_m$ of the corresponding month m. Overall, this methodological sequence ensures traceability of transformations and guarantees that the forecasts preserve both the discrete nature of the data and their fundamental statistical properties (nonstationarity and serial correlation).

2.3. Non-Homogeneous Autoregressive Negative Binomial Process (NH-ARNBP)

Modeling count time series, such as the number of daily truck arrivals at a port, requires addressing both temporal autocorrelation and potential overdispersion, the latter understood as variance significantly exceeding the mean (Gouriéroux & Lu, 2018; Hilbe, 2011; Quddus, 2008). The Poisson distribution (such as in NHPP), assumes equidispersion (variance equal to the mean) (Goulding et al., 2016), which makes it inadequate when data exhibit overdispersion or structural heterogeneity attributable to external logistical or environmental

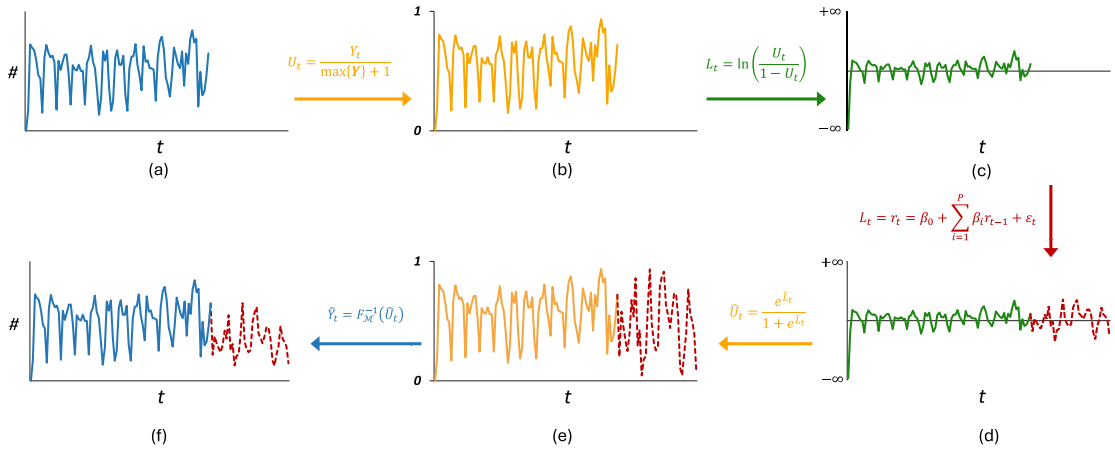


Fig. 1. Methodological scheme for the proposed model.

factors (Cameron & Trivedi, 2013; Omer & Hussian, 2023; Quddus, 2008).

In response, we introduce the NH-ARNBP, which, unlike NH-ARPP, captures the overdispersion present in the original series $\{Y_t\}$. The negative binomial distribution, parameterized by τ (shape parameter) and ρ (success probability), models the probability of observing k failures before registering τ successes, under the assumption of independent trials. Its probability mass function is

$$\mathbb{P}(X=k)=\mathrm{NB}(k|\tau,\rho)=\binom{k+\tau-1}{\tau-1}(1-\rho)^{\tau}\rho^{k} \qquad k\in\mathbb{N}_0. \tag{10}$$

The main appeal of the negative binomial distribution is the decoupling of its mean and variance, given by $\mathbb{E}[X]=\tau\frac{1-\rho}{\rho}$ and $\mathbb{V}(X)=\tau\frac{1-\rho}{\rho^2}$, respectively, allowing to model overdispersion when $\mathbb{V}(X)>\mathbb{E}[X]$ (Chen et al., 2016; Gouriéroux & Lu, 2018; Hilbe, 2011).

To adapt this distribution to the modeling of non-homogeneous count processes, the parameters τ_m and ρ_m are empirically estimated by period (e.g., by month m) based on the historical series $\{Y_t\}$. Let $\overline{Y}_m$ be the sample mean and s_m^2 the sample variance in month m. The distribution parameters are estimated by

$$\hat{\tau}_m=\frac{\overline{Y}_m^2}{s_m^2-\overline{Y}_m}, \qquad \hat{\rho}_m=\frac{\overline{Y}_m}{s_m^2}. \tag{11}$$

These estimators allow the negative binomial distribution to capture both the empirical mean and overdispersion in the data, overcoming the limitations of the classical Poisson model under excess variability (Chen et al., 2016; Omer & Hussian, 2023; Quddus, 2008). In our dataset (see later in Section 3.1), all monthly segments exhibit clear overdispersion ($s_m^2>\overline{Y}_m$), ensuring that the estimators of τ_m and ρ_m remain well-defined and positive. For robustness, in the rare case where $s_m^2\le\overline{Y}_m$, the negative binomial naturally reduces to its Poisson limit, and we could adopt $\lambda_m=\overline{Y}_m$ for that month, guaranteeing valid parameter values without altering the overall methodology.

The full procedure replicates the methodological logic of the NH-ARPP (summarized in Fig. 1) but is adapted to the NH-ARNBP structure. After the logit transformation of the original series and decomposition into trend and seasonality components, the AR(P) model is fitted to the logit-transformed series $\{L_t\}$. In this logit space, forecast trajectories are generated using the AR model. These predictions $\{\hat{L}_t\}$ are then transformed back to the standardized scale using the inverse logit transformation. In this standardized scale, the forecasts are mapped to probabilistic quantiles, and the inverse transform method is applied to map the values to a discrete count space. Specifically, for each month m, the inverse CDF (or Percent Point Function, PPF) of the negative binomial distribution, parameterized by the empirical estimators $\hat{\tau}_m$ and $\hat{\rho}_m$, is applied as

$$\hat{Y}_t=F_{\mathrm{NB}}^{-1}(\hat{U}_t|\hat{\tau}_m,\hat{\rho}_m), \tag{12}$$

where F_{NB}^{-1} denotes the negative binomial inverse CDF. This ensures that future forecasts $\{\hat{Y}_t\}$ simultaneously incorporate the discrete nature and overdispersion of the original series, along with its autoregressive dynamics and nonstationarity.

It is important to distinguish the proposed NH-ARNBP from existing classes of stochastic count models such as INGARCH. These models are observation-driven processes in which the conditional mean (and in some cases the conditional variance) evolves recursively as a direct function of past observations. In contrast, the NH-ARNBP adopts a hierarchically structured formulation: (i) a deterministic component that decomposes the logit-transformed series into trend and seasonal effects; (ii) a short-run dependence, handled by the AR dynamics on the transformed scale; and (iii) slow-moving non-stationary and overdispersed behavior, captured by month-specific negative binomial parameters. This is not presented as a replacement for observation-driven models, but as a complementary framework that can be preferable when the goal is to obtain calibrated and operationally informative predictive distributions under strong overdispersion and seasonal non-stationarity.

2.4. Performance metrics

The quantitative evaluation of the predictive performance of a stochastic model requires metrics that assess not only the accuracy of point forecasts but mainly the probabilistic quality of the predictive distributions. These metrics, known as *proper scoring rules*, penalize the assignment of low probability to events that ultimately occur. One of the most widely used metrics in this context is the *Logarithmic Score* (LS), defined for discrete variables as

$$\mathrm{LS}_t=-\log f_t(Y_t),$$

where $f_t(\cdot)$ is the probability mass function (PMF) of the predictive distribution at time t. This metric exponentially penalizes the assignment of low probability to the realized observation Y_t, reaching its minimum when the PMF concentrates all mass on the observed value.

A key advantage of the LS is its *strict propriety* property introduced by Good (2018), which guarantees that its expected value is minimized only when the model reports its true predictive distribution (Gneiting & Raftery, 2007). This makes LS a suitable criterion for evaluating the statistical consistency of predictive distributions in discrete time series contexts (Jordan et al., 2019).

Despite these advantages, LS presents two critical limitations in practice (Gneiting & Raftery, 2007). First, if $f_t(Y_t)\to 0$, then $\mathrm{LS}_t\to\infty$, which results in extreme penalties for observations not well supported by the model. Second, as a *local rule*, it does not distinguish between distributions assigning equally low probabilities to the observed value.

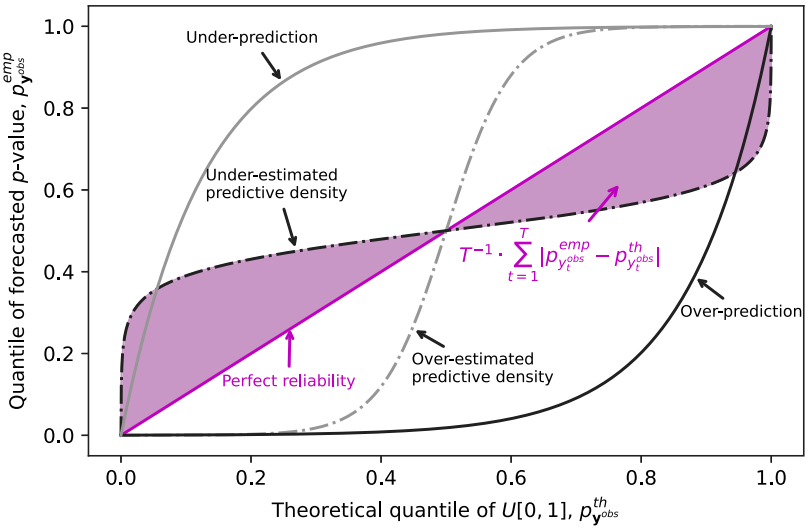


Fig. 2. Schematic representation of a predictive QQ plot illustrating under- and over-prediction zones, and predictive density.
Source: Adapted from Favereau et al. (2022).

For example, two distributions with $f_t(Y_t) = 0.01$ and $f_t(Y_t) = 0.02$ yield very different scores, even though both severely underestimate the observation (Jordan et al., 2019). In high-uncertainty scenarios or with wide support, it is convenient to complement LS with *non-local* rules such as the Continuous Ranked Probability Score (CRPS), which is sensitive to distance between distributions and penalizes underconfidence more gradually (Bracher et al., 2021).

CRPS is defined from a set of simulated trajectories or forecasts Ω, $\{y^{\text{fore}}_{t,\omega}\}_{\omega\in\Omega}$, and an observed realization y^{obs}_t for each time $t = 1, \dots, T$. Let

$$\boldsymbol{y}^{\text{obs}} = (y^{\text{obs}}_1, \dots, y^{\text{obs}}_T)^\top, \quad \boldsymbol{y}^{\text{fore}}_i = (y^{\text{fore}}_{1,\omega}, \dots, y^{\text{fore}}_{T,\omega})^\top. \quad \omega \in \Omega$$

Then, the CRPS at time t is defined as

$$\text{CRPS}_t = \underbrace{\frac{1}{|\Omega|}\sum_{\omega\in\Omega}\left|y^{\text{fore}}_{t,\omega} - y^{\text{obs}}_t\right|}_{\text{MAE}_t} - \underbrace{\frac{1}{2|\Omega|^2}\sum_{i\in\Omega}\sum_{j\in\Omega}\left|yt, i^{\text{fore}} - y^{\text{fore}}_{t,j}\right|}_{\text{Disp}_t}. \tag{13}$$

The average score across T time steps is defined as

$$\overline{\text{CRPS}} = \frac{1}{T}\sum_{t=1}^{T}\text{CRPS}_t,$$

which provides a single global measure of the predictive distribution quality (Gneiting & Raftery, 2007; Jordan et al., 2019). For (13), the first term quantifies the mean absolute error (MAE) between each simulated trajectory and the observation at time t, and the second term measures the internal dispersion (Disp) of the ensemble, penalizing excessive variability that does not contribute to improving forecast accuracy (Hersbach, 2000; Jordan et al., 2019). By subtracting both terms, CRPS simultaneously rewards accuracy and sharpness. Lower values of CRPS_t indicate better predictive distributions at time t (Matheson & Winkler, 1976).

Since CRPS summarizes overall predictive quality, it is useful to complement it with metrics that verify marginal calibration of prediction intervals. For this purpose, we use the concept of prediction interval at confidence level α, defined for each time t as

$$I^\alpha_t = \left[q_t^{\frac{1-\alpha}{2}}, \; q_t^{\frac{1+\alpha}{2}}\right],$$

where q_t^δ represents the δ-quantile of the predictive distribution f_t. This implies $\mathbb{P}(y^{\text{fore}}_t \in I^\alpha_t) = \alpha$. To assess whether the constructed intervals adequately cover observed variability, we compute the Prediction Interval Coverage (PIC) at level α, defined as the proportion of time steps in which the observed value lies within its corresponding interval

$$\text{PIC}_\alpha = \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}(y^{\text{obs}}_t \in I^\alpha_t),$$

where $\mathbb{I}(\cdot)$ is the indicator function, equal to 1 if the condition is satisfied and 0 otherwise. A value of $\text{PIC}_\alpha \approx \alpha$ indicates that the intervals are well calibrated and uncertainty is well quantified (Gneiting & Raftery, 2007; Hyndman & Athanasopoulos, 2018).

In practice, several confidence levels are typically evaluated simultaneously, e.g., $\alpha = 0.5, 0.8, 0.95$. The comparison of PIC_α with the nominal level α reveals whether the model consistently underestimates ($\text{PIC}_\alpha < \alpha$) or overestimates ($\text{PIC}_\alpha > \alpha$) uncertainty across the forecast horizon.

Additionally, we introduce two complementary metrics that capture fundamental properties of the predictive distribution: *reliability* and *resolution* (Favereau et al., 2022; Humphrey et al., 2016; Renard et al., 2010). Based on (Favereau et al., 2022), the reliability and resolution indices are defined as follows:

$$\text{Rel}(p^{\text{emp}}_{\boldsymbol{y}^{\text{obs}}}, p^{\text{th}}_{\boldsymbol{y}^{\text{obs}}}) = 1 - 2\frac{\sum_{t=1}^{N}|p^{\text{emp}}_{y^{\text{obs}}_t} - p^{\text{th}}_{y^{\text{obs}}_t}|}{N}, \tag{14}$$

$$\text{Res}(\bar{\boldsymbol{y}}^{\text{fore}}, \tilde{\boldsymbol{y}}^{\text{fore}}) = \frac{1}{N}\sum_{t=1}^{N}\frac{\bar{y}^{\text{fore}}_t}{\tilde{y}^{\text{fore}}_t}, \tag{15}$$

where $p^{\text{emp}}_{y^{\text{obs}}_t}$ and $p^{\text{th}}_{y^{\text{obs}}_t}$ denote the empirical and theoretical p-values associated with the observation y^{obs}_t. Likewise, $\bar{y}^{\text{fore}}_t$ and $\tilde{y}^{\text{fore}}_t$ correspond to the mean and standard deviation, respectively, of the simulated forecast values $\{y^{\text{fore}}_{t,\omega}\}_{\omega\in\Omega}$ at each time $t \in \mathcal{T}$.

The reliability index relies on the idea that, if y^{obs}_t is a realization of the forecast cumulative distribution function $F_t(\cdot)$, then the corresponding p-value $F_t(y^{\text{obs}}_t)$ should follow a uniform distribution on $[0, 1]$. The predictive quantile–quantile (QQ) plot is a diagnostic tool for comparing the empirical cumulative distribution of the sample $\{F_t(y^{\text{obs}}_t)\}_{t\in\mathcal{T}}$ with the theoretical $U(0, 1)$ distribution. Fig. 2 illustrates this concept, showing discrepancies from the 1:1 line, which are interpreted as deviations from perfect reliability. The reliability index takes values in the interval $[0, 1]$, where 1 indicates perfect alignment (ideal reliability), and 0 denotes a completely unreliable forecast.

In contrast, the resolution index quantifies how concentrated or sharp the predictive distributions are, regardless of whether they are well-calibrated. It is computed as the ratio between the forecast mean and standard deviation, averaged over all time points. Values of resolution close to zero indicate broad uncertainty ranges (i.e., high

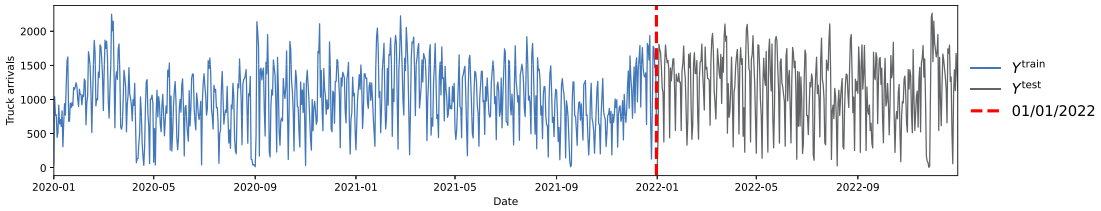


Fig. 3. Daily time series of truck arrivals at the port (Jan 1, 2020–Dec 31, 2023).

dispersion relative to the mean), while larger values suggest sharper, more concentrated forecasts. Although high resolution may be desirable, it must be balanced against reliability, as overly sharp forecasts that are not well calibrated can lead to overconfident and misleading predictions.

With the formalization of the metrics Rel and Res, we complete the set of tools used to evaluate the performance of probabilistic forecasting models in terms of calibration, statistical coherence, and structural accuracy. The metrics LS, CRPS, PIC_α, Rel, and Res together form a technically grounded and operationally replicable framework for the systematic comparison of models under inherent uncertainty in stochastic processes.

3. Analysis of results

This section presents the experimental analysis conducted to evaluate the performance of the proposed models. First, we describe the selected case study and the general characteristics of the analyzed time series. Then, we detail the calibration process of the proposed models, including the criteria used for parameter selection. The methodology for generating forecasts and simulations is also described, along with the validation procedures adopted. Finally, we present the results and quantitatively compare the models in terms of predictive accuracy and statistical fit.

3.1. Case study

This research is based on data from one of Chile's major ports, located in the Valparaíso region. Let $\{Y_t\}_{t=1}^{N}$ denote the daily time series of truck arrivals at the port, consisting of $N = 1096$ observations corresponding to the period from January 1, 2020, to December 31, 2023. The dataset was provided directly by the port authority, and no missing observations were present in the time series. Outlier inspection was performed using median absolute deviation, resulting in no extreme values requiring removal.

Fig. 3 shows the full evolution of $\{Y_t\}$. The blue curve represents the training subset $Y^{\text{train}} = \{Y_t\}_{t\in\mathcal{T}^{\text{train}}}$, with $\mathcal{T}^{\text{train}} = \{1,\dots,t^\star\}$, while the black curve represents the test subset $Y^{\text{test}} = \{Y_t\}_{t\in\mathcal{T}^{\text{test}}}$, with $\mathcal{T}^{\text{test}} = \{t^\star+1,\dots,N\}$. The vertical red dashed line marks the separation point $t^\star$ between the two subsets. In particular, $\mathcal{T}^{\text{train}}$ considers the first two years of data, whereas $\mathcal{T}^{\text{test}}$ the last year.

Section 3.2 details the transformations applied to Y^{train} to perform model fitting and calibration procedures. This sequence of transformations prepares the series for fitting autoregressive models and simulating future trajectories.

3.2. Transformations of the original time series

Following the methodology established in Section 2.2, three successive transformations were applied to the observed trajectory Y^{train}, defined as the training set with all observations before January 1, 2022.

First, the original series Y^{train} was standardized based on (5), resulting in the trajectory $U^{\text{train}} = \{U_t \in [0,1]\}_{t\in\mathcal{T}^{\text{train}}}$. This operation projects the data onto a relative scale, preserving their temporal structure. Then, the logit transformation defined in (6) was applied to U^{train}, yielding a new series $L^{\text{train}} = \{L_t \in \mathbb{R}\}_{t\in\mathcal{T}^{\text{train}}}$, suitable for autoregressive modeling on the real set. Fig. 4 depicts these three representations: the original series Y^{train} (panel a), the standardized series U^{train} on the $[0,1]$ scale (panel b), and the logit-transformed series L^{train} on the real set (panel c). The transformations faithfully preserve the relative shape of the original trajectory, a necessary property for future simulation steps using recursive prediction, where inverse transformations must be applied.

From the transformed trajectories L^{train}, an additive decomposition was performed into three structural components: trend T_t, seasonality S_t, and residual r_t. The trend component T_t was estimated as a linear regression over time t using ordinary least squares (OLS), i.e., $T_t = \alpha_0^{\text{trend}} + \alpha_1^{\text{trend}} t$ for all $t \in \mathcal{T}^{\text{train}}$, where α_0^{trend} is the intercept and α_1^{trend} is the slope. OLS also estimated seasonality S_t based on harmonic regression with periodicity ζ (Favereau et al., 2022), defined by

$$S_t = \sum_{h=1}^{H}\left[\beta_{1,h}^{\text{seas}}\sin\left(\frac{2\pi ht}{\zeta}\right) + \beta_{2,h}^{\text{seas}}\cos\left(\frac{2\pi ht}{\zeta}\right)\right], \quad \forall t \in \mathcal{T}^{\text{train}}$$

where H is the number of pairs of harmonics. To select the optimal number of pairs of harmonics, $\{S_t\}$ where estimated for $H \in \{1,\dots,12\}$, and it was evaluated using the BIC criterion, resulting in $H^\star = 3$ the value that minimized it. The residual series is then obtained after subtracting trend and seasonality from the original series, i.e., $r_t = Y_t - T_t - S_t$, and constitutes the primary input for fitting the autoregressive model presented in the next section.

3.3. Fitting and calibration of the autoregressive model

An AR(P) model was fitted to the residual series $\{r_t\}_{t\in\mathcal{T}^{\text{train}}}$ to capture its temporal dependence structure. The general formulation is

$$r_t = \beta_0^{\text{AR}} + \sum_{i=1}^{P}\beta_i^{\text{AR}} r_{t-i} + \varepsilon_t,$$

where β_0^{AR} is the intercept, β_i^{AR} are the AR coefficients for lags i, and $\varepsilon_t \sim \mathcal{N}(0,\hat{\sigma}_\varepsilon)$ for all t ($\hat{\sigma}_\varepsilon$ is estimated below). Before fitting the AR(P) model, the stationarity of $\{r_t\}_{t\in\mathcal{T}^{\text{train}}}$ was tested using the Augmented Dickey–Fuller (ADF) test (Dickey & Fuller, 1981). The test statistic was 2.872 with a p-value of 0.0487, allowing rejection of the null hypothesis of a unit root at the 5% significance level. Therefore, the series is deemed stationary and suitable for autoregressive modeling.

To select the optimal order P, AR(P) models were estimated for $P \in \{1,\dots,10\}$, and evaluated using the AIC and BIC. Fig. 5 shows the evolution of both criteria for P, where red circles denote the values at which AIC and BIC are minimized. Although AIC reaches its minimum at $P = 10$, BIC is minimized at $P = 8$, where the latter is considered due to BIC's stricter penalization for model complexity, which favors parsimony and reduces overfitting risk (Cavanaugh & Neath, 2019; Lütkepohl, 2005).

The AR(8) model was estimated using conditional maximum likelihood. Table 1 reports the estimated coefficients, standard errors, z-statistics, p-values, and 95% confidence intervals. A statistical significance filter was applied, retaining only terms with $p < 0.05$. The final model includes lags r_{t-1}, r_{t-4}, r_{t-6}, and r_{t-8}. The standard deviation of residuals $\hat{\sigma}_\varepsilon = 1.1021$ was also computed and will be used to generate white noise realizations in future simulations.

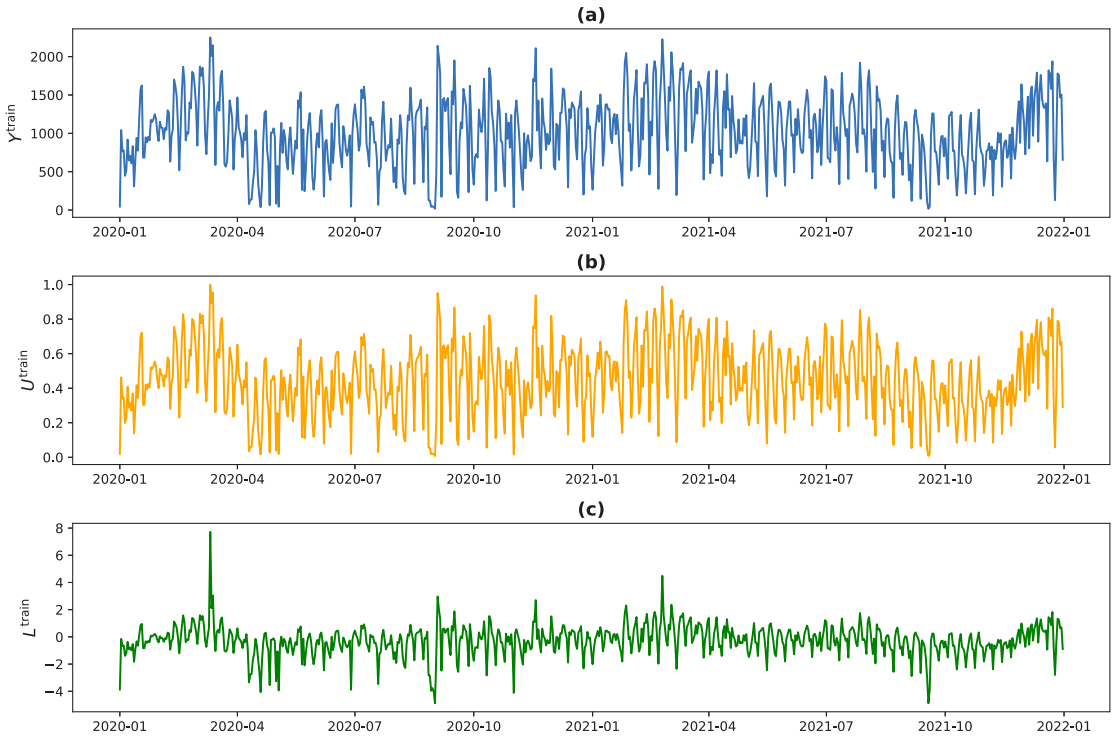


Fig. 4. Transformations of the arrival series: (a) original, (b) standardized, and (c) logit-transformed.

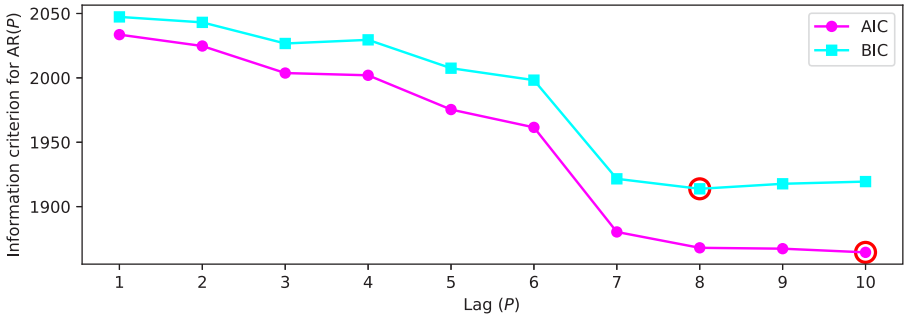


Fig. 5. Information criteria (AIC and BIC) for AR(P) models with different orders.

Table 1
Results of AR(8) model estimated via conditional maximum likelihood.

Lag	Coefficient (β_i)	Standard error	z-statistic	p-value	95% CI
r_{t-1}	0.4643	0.037	12.590	0.000	[0.392, 0.537]
r_{t-4}	−0.1274	0.038	−3.333	0.000	[−0.202, −0.052]
r_{t-6}	0.0843	0.038	2.196	0.028	[−0.009, 0.160]
r_{t-8}	−0.1307	0.037	−3.559	0.000	[−0.203, −0.059]

3.4. Simulated trajectories of NH-ARPP and NH-ARNBP models

Forecast trajectory simulation is carried out by recursively extending the autoregressive model fitted to the residual series $\{r_t\}_{t\in\mathcal{T}^{\text{train}}}$, incorporating the uncertainty component via a white noise matrix defined as

$$\varepsilon_t^{\omega} \sim \mathcal{N}(0, \hat{\sigma}_{\varepsilon}), \quad \forall t \in \mathcal{T}^{\text{test}}, \omega \in \Omega,$$

where $|\Omega| = 1000$ is the number of simulated trajectories (realizations). Each ε_t^{ω} corresponds to an independent realization of the white noise. The AR(8) model dynamics are applied recursively for each realization ω, using the corresponding sequence $\{\varepsilon_t^{\omega}\}_{t\in\mathcal{T}^{\text{test}}}$ to generate a matrix of residual predictions $\{\hat{r}_t^{\omega}\}_{t\in\mathcal{T}^{\text{test}}}$ according to the AR(8) specification (Table 1).

For each trajectory $\omega \in \Omega$ and period $t \in \mathcal{T}^{\text{test}}$, the full logit-transformed path is reconstructed by adding the deterministic components estimated earlier

$$\hat{L}_t^{\omega} = \hat{T}_t + \hat{S}_t + \hat{r}_t^{\omega},$$

where $\hat{T}_t$ and $\hat{S}_t$ correspond to the extrapolated (to $t \in \mathcal{T}^{\text{test}}$) trend and seasonality components from the training phase ($\mathcal{T}^{\text{train}}$). The inverse logit transformation is then applied to each trajectory to obtain the standardized predictions

$$\hat{U}_t^{\omega} = \frac{e^{\hat{L}_t^{\omega}}}{1 + e^{\hat{L}_t^{\omega}}} \in [0, 1],$$

i.e., $\{\hat{U}_t^{\omega}\}_{t\in\mathcal{T}^{\text{test}},\omega\in\Omega}$. This series $\{\hat{U}_t^{\omega}\}$ allow direct application of the inverse transform method to retrieve counts in the original scale using the appropriate count distribution per model (NH-ARPP and NH-ARNBP). For each trajectory $\omega \in \Omega$ and time $t \in \mathcal{T}^{\text{test}}$, the recovery of the discrete count scale is performed as

Table 2
Monthly parameters estimated for NH-ARPP and NH-ARNBP models.

Month	$\hat{\lambda}_m$ (Poisson)	$\hat{\tau}_m$ (NegBin)	$\hat{\rho}_m$ (NegBin)
January	991.00	6.443	0.00646
February	1285.28	9.078	0.00701
March	1338.19	8.331	0.00619
April	912.23	4.443	0.00485
May	882.92	4.547	0.00512
June	977.87	8.502	0.00862
July	1041.48	7.266	0.00693
August	899.45	3.692	0.00409
September	925.07	3.037	0.00327
October	870.15	4.946	0.00565
November	926.93	5.528	0.00593
December	1178.60	7.522	0.00634

$$\hat{y}^{\mathcal{M}}_{t,\omega} = F^{-1}_{\mathcal{M}_{m(t)}}(\hat{U}^{\omega}_t) \quad \forall \mathcal{M} \in \{\mathrm{P}, \mathrm{NB}\}$$

where $F^{-1}_{\mathrm{P}_{m(t)}}$ denotes the inverse CDF for the Poisson distribution (for NH-ARPP) and $F^{-1}_{\mathrm{NB}_{m(t)}}$ the inverse CDF for the negative binomial distribution, both for the corresponding month $m(t)$ where the day t belongs. Each distribution is monthly parameterized for each year-calendar month, from January ($m = 1$) to December ($m = 12$), with $\hat{\lambda}_m$ and $\{\hat{\tau}_m, \hat{\rho}_m\}$ for the Poisson and negative binomial cases, respectively. Both $\hat{\lambda}_m$ and $\{\hat{\tau}_m, \hat{\rho}_m\}$ are the monthly parameters estimated from the training set $\mathcal{T}^{\mathrm{train}}$, as described in Sections 2.2 and 2.3. Their estimations are shown in Table 2.

The choice of a monthly parameterization for the distributional component is primarily motivated by data availability and statistical stability. Estimating the Poisson or negative binomial parameters at a finer temporal resolution (e.g., daily or weekly) would result in very small sample sizes, producing highly unstable or non-identifiable estimates of the mean–variance structure. In contrast, monthly aggregation yields sufficient observations to estimate reliable distributional parameters while preserving the non-stationary behavior of the arrival process.

Shorter-term fluctuations are not ignored; instead, they are explicitly captured by the autoregressive component of the model, which operates on the logit-transformed residual series. This separation of roles — monthly non-homogeneous distribution for structural variability and AR dynamics for day-to-day fluctuations — allows the model to represent both slow-moving seasonal effects and high-frequency temporal dependence in a statistically robust manner. In this work, the forecasting target is daily arrivals and the intended use is uncertainty-aware planning at the daily-to-weekly level (e.g., anticipating congestion risk and staffing pressure), rather than intra-day control. As a result, the monthly distributional layer is designed to represent slow-moving structural heterogeneity, while intra-week dynamics are handled by the AR component.

Fig. 6 shows the monthly PMFs and CDFs for both models, based on the distributional parameters estimated in Table 2. The NH-ARPP model's PMF exhibits a narrow distribution concentrated around the expected values $\hat{\lambda}_m$. In contrast, the NH-ARNBP model's PMF reveals heavier tails and greater dispersion, controlled by the shape parameter $\hat{\tau}_m$ and success probability $\hat{\rho}_m$. For improved visual clarity and ease of comparison, discrete functions were smoothed while preserving structure and parameter-driven values.

The estimated mass function comparison highlights substantive differences in the predictive behavior of the two models. NH-ARPP shows limited representational capacity under overdispersed conditions, while NH-ARNBP produces more diverse trajectories concerning the empirical variability observed in the training data. The quantitative evaluation of these trajectories, evaluated with the performance metrics described in Section 2.4, is presented in the next section.

Table 3
Descriptive statistics of daily logarithmic score (LS_t) for each model during the test period.

Model	Mean	Std. Dev.	Min	Max
NH-ARNBP	8.0415	1.9128	6.6943	30.629
NH-ARPP	136.8879	145.1798	4.3249	690.376

3.5. Evaluation of performance metrics for NH-ARPP and NH-ARNBP forecasts

To further assess the predictive capacity of the NH-ARPP and NH-ARNBP models, we implemented an evaluation scheme based on the quantitative performance metrics defined in Section 2.4. These metrics were grouped into three complementary analytical blocks: (i) prediction interval coverage, (ii) logarithmic score and monthly distributions, and (iii) evaluation of predictive distributions.

3.5.1. Prediction interval coverage

The first dimension evaluated corresponds to the PIC of the simulated trajectories, defined as the proportion of observed values that fall within the prediction intervals generated by each model. Fig. 7 shows the prediction intervals at various confidence levels (from 50% to 99%) for both predictive distributions over the test horizon. The NH-ARPP model produces much narrower intervals centered around its mean, consistent with the probability mass functions observed in Fig. 6. In contrast, the NH-ARNBP model yields wider intervals, reflecting its greater ability to represent structural overdispersion in the data while maintaining temporal dynamics through the AR(8) model.

Analyzing the PIC, the NH-ARNBP model shows substantially better performance. For the 95%, 97.5%, and 99% intervals, the observed coverage reaches 72.6%, 79.5%, and 85.2%, respectively. In contrast, NH-ARPP yields significantly lower and insufficient coverage levels: 7.4%, 7.7%, and 7.9%, respectively. This clearly indicates a systematic underestimation of uncertainty. These results suggest that NH-ARNBP more realistically captures the variability observed in daily truck arrivals.

Beyond overall coverage metrics, it is useful to interpret the prediction intervals operationally. For example, in January 2022, the 50% central interval of the NH-ARNBP model spans approximately 500–1500 trucks per day, meaning that half of the expected daily arrivals fall within this range. Such central bands help operators anticipate "typical" demand levels for baseline resource allocation, while wider bands (e.g., 95%) quantify tail risk and support contingency planning for unusually high volumes.

The superior empirical coverage achieved by NH-ARNBP over NH-ARPP therefore reflects not only wider intervals, but more realistic and better-calibrated uncertainty quantification, which is crucial for avoiding systematic underestimation of required capacity in port operations.

3.5.2. Logarithmic score and monthly distributions

In the second analytical block, we consider the logarithmic score (LS), which evaluates the quality of the empirical distribution estimated by each model based on the probability assigned to the realized value. LS penalizes heavily when the observed value receives a low probability under the model's forecast distribution.

Fig. 8 shows the daily evolution of LS over the test period $\mathcal{T}^{\mathrm{test}}$. The NH-ARPP model exhibits high temporal instability, with multiple instances of severe penalties. Several peaks exceed 600 units, indicating moments when the observed value had nearly zero predicted probability. Summary statistics for LS_t are presented in Table 3.

Fig. 9 further reinforces this conclusion by displaying the histogram of LS for both models. NH-ARPP concentrates a significant proportion of days in the high-penalty interval (200–600), while NH-ARNBP concentrates more than 95% of its values below 30. This large difference

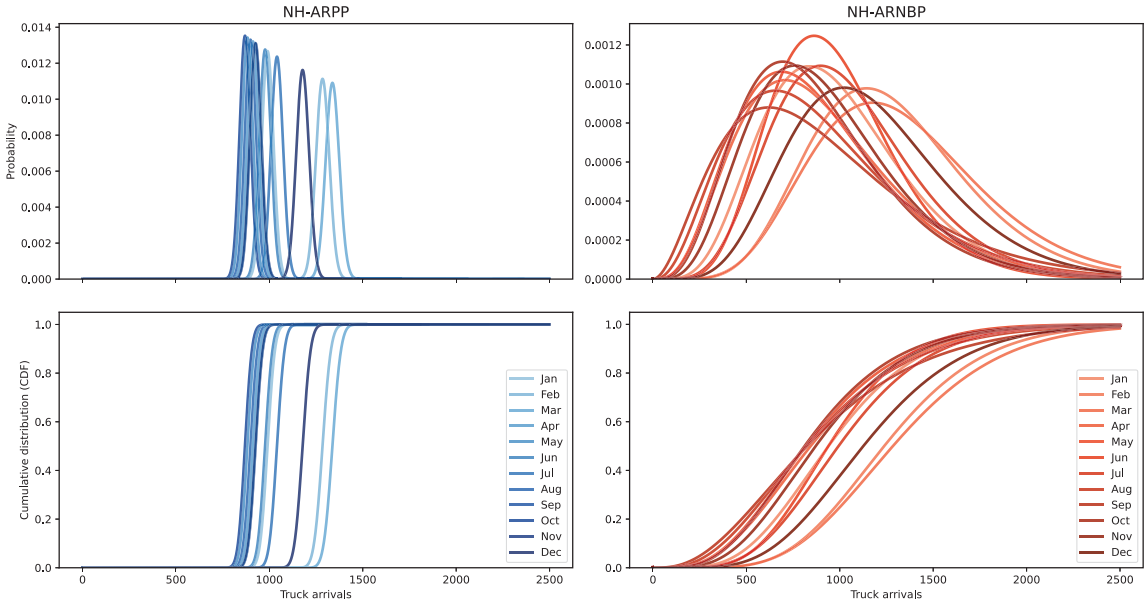


Fig. 6. Monthly probability mass functions (PMFs) and cumulative distribution functions (CDFs) for NH-ARPP and NH-ARNBP.

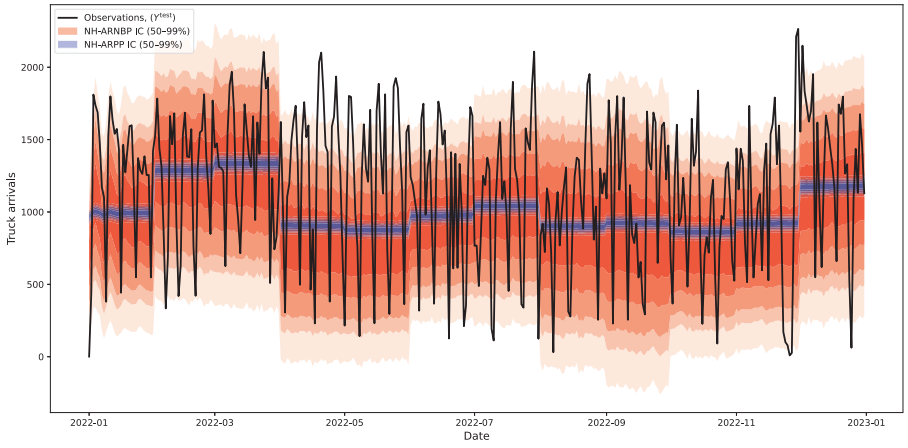


Fig. 7. Daily central prediction intervals generated by NH-ARPP and NH-ARNBP for various confidence levels.

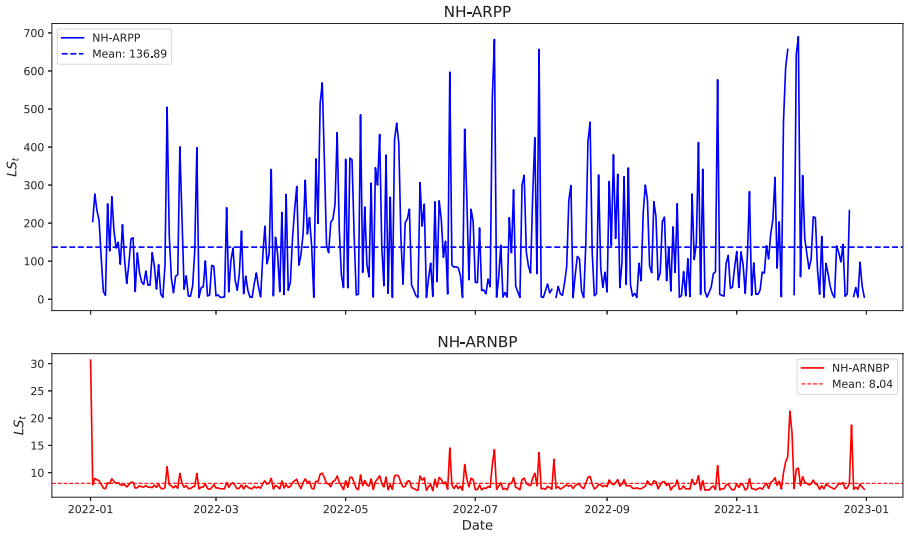


Fig. 8. Daily evolution of the logarithmic score (LS) for NH-ARPP and NH-ARNBP.

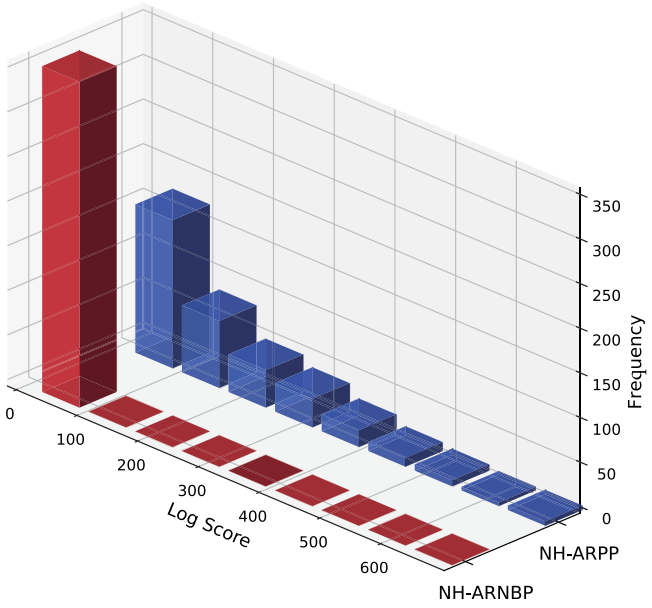


Fig. 9. Logarithmic score (LS) frequency distribution by model.

indicates that the predictive distributions generated by NH-ARNBP better match the observed data.

To visualize how each model distributes probability mass in each month of the test set, Fig. 10 presents the predictive distributions for each model along with the observed values. This plot shows how the forecast support varies over time and its alignment with actual observations. NH-ARNBP provides higher density near the observed values in most months. In contrast, NH-ARPP tends to concentrate probability mass around its expected value $\hat{\lambda}_m$ (and far away from observations), leading to narrower and less flexible predictive distributions.

The results from this block support the conclusion that NH-ARNBP generates better-adapted predictive distributions under overdispersion, avoiding extreme penalties while preserving dynamic behavior.

3.5.3. Evaluation of predictive distributions

The final evaluation block compares the predictive distributions of NH-ARPP and NH-ARNBP using the Continuous Ranked Probability Score (CRPS) introduced in Section 2.4. Fig. 11 presents the daily difference in CRPS between NH-ARNBP and NH-ARPP, computed as

$$\Delta_t^{\text{CRPS}} = \text{CRPS}_t^{\text{NB}} - \text{CRPS}_t^{\text{PP}},$$

with $\text{CRPS}_t^{\text{NB}}$ and $\text{CRPS}_t^{\text{PP}}$ being the CRPS_t obtained from simulated trajectories of models NH-ARNBP (NB) and NH-ARPP (PP), respectively. The average CRPS for NH-ARPP was $\overline{\text{CRPS}}^{\text{PP}} = 431.67$, while NH-ARNBP achieved $\overline{\text{CRPS}}^{\text{NB}} = 346.48$. This difference offers an objective measure of the superior probabilistic performance of the NH-ARNBP model.

To explore the internal structure of this difference, Fig. 12 shows the empirical decomposition of CRPS_t for both models. Each point corresponds to a pair $(\text{MAE}_t, \text{Disp}_t)$ for each $t \in \mathcal{T}^{\text{test}}$. As seen here, NH-ARPP has a lower average absolute error (446.3) than NH-ARNBP (530.9), but its average internal dispersion is much smaller (14.7 vs. 184.4), making its forecasts overly concentrated. CRPS penalizes this underdispersion, leading to a higher overall score (the stars represent these values). This illustrates the typical subdispersive behavior of Poisson-based models, in contrast with the greater variability captured by NH-ARNBP.

Finally, we report reliability and resolution indicators defined in (5) and (6), respectively. NH-ARNBP achieves a higher reliability index (Rel = 0.2984) compared to NH-ARPP (Rel = 0.0282), indicating better marginal calibration. In terms of resolution, NH-ARPP reaches a higher value (Res = 39.21) than NH-ARNBP (Res = 2.942), consistent with its lower variability. However, as noted by Favereau et al. (2022), high resolution is not necessarily desirable when associated with underdispersed forecasts. In this case, the low reliability of NH-ARPP confirms that its high resolution stems from overconfident predictions. The higher dispersion of NH-ARNBP trajectories better matches the variability observed in the real series.

NH-ARPP exhibits a higher resolution index because its predictive distributions are much sharper (i.e., lower internal dispersion). However, this sharpness is caused by the fact that intervals are too narrow, which causes a very low PIC. In contrast, NH-ARNBP produces more dispersed predictive distributions that better match the empirical variability of truck arrivals, resulting in substantially higher reliability and more accurate prediction intervals.

3.6. Benchmarking against GP-INGARCH models

To further clarify the distinction between the proposed NH-ARNBP and established count time-series approaches, we include a benchmark analysis based on a Generalized Poisson INGARCH(p, q) model, known as GP-INGARCH (Ferland et al., 2006; Zhu, 2012b). The goal of this benchmarking exercise is not to provide an exhaustive comparison across all forecasting paradigms, but to contrast two fundamentally different modeling philosophies within count time series: (i) a hierarchically specified non-homogeneous framework that separates short-run dependence from slowly varying distributional heterogeneity (NH-ARNBP), and (ii) an observation-driven model where dynamics and dispersion are jointly embedded in a recursive conditional mean (GP-INGARCH). This controlled comparison helps isolate structural trade-offs in probabilistic behavior (calibration, sharpness, and coverage) that are central for operational use.

A GP-INGARCH(p, q) process is an observation-driven count model in which the conditional distribution of Y_t given the past information set $\mathcal{F}_{t-1}$ follows a Generalized Poisson law with conditional mean λ_t and dispersion parameter ϕ, that is,

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{GP}(\lambda_t, \phi).$$

The conditional mean evolves recursively according to

$$\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j},$$

where $\omega > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$ ensure positivity and stationarity under standard conditions. The Generalized Poisson distribution admits the probability mass function

$$\mathbb{P}(Y_t = k \mid \mathcal{F}_{t-1}) = \frac{\lambda_t (\lambda_t + \phi k)^{k-1}}{k!} e^{-(\lambda_t + \phi k)} \qquad k \in \mathbb{N}_0,$$

which accommodates both overdispersion and underdispersion depending on the sign and magnitude of ϕ (Zhu, 2012b). Thus, the GP-INGARCH framework jointly captures dynamic autoregressive dependence through the recursion for λ_t and dispersion heterogeneity through the parameter ϕ.

For $p, q \in \{1, 2, \ldots, 10\}$, we evaluated all GP-INGARCH(p, q) specifications and selected the model minimizing the BIC criterion. Fig. 13 summarizes the BIC surface over the full (p, q) grid and reveals a well-defined region of improved fit concentrated around larger autoregressive orders and shorter moving-average components. The global minimum occurs at $(p, q) = (10, 1)$, indicating that the best-performing INGARCH specification requires a relatively long memory of past counts but only limited dependence on past conditional means. This pattern is consistent with the dynamics of daily truck arrivals, where short-run shocks propagate for several days but the smoothing effect of λ_t is comparatively limited. Accordingly, GP-INGARCH(10,1) was selected as the benchmark model for subsequent comparisons.

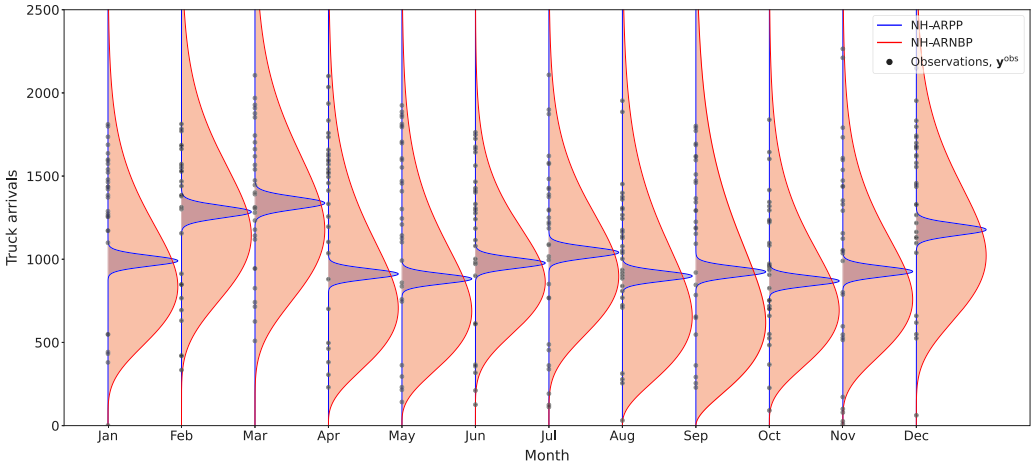


Fig. 10. Monthly predictive distributions by model with observed values.

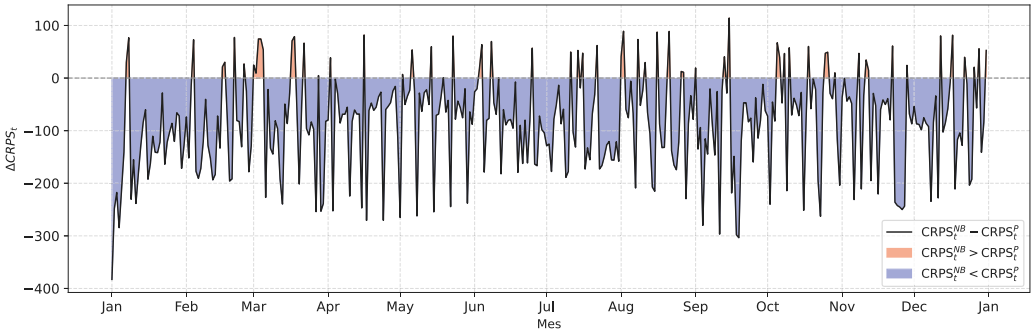


Fig. 11. Daily difference in CRPS between NH-ARNBP and NH-ARPP models.

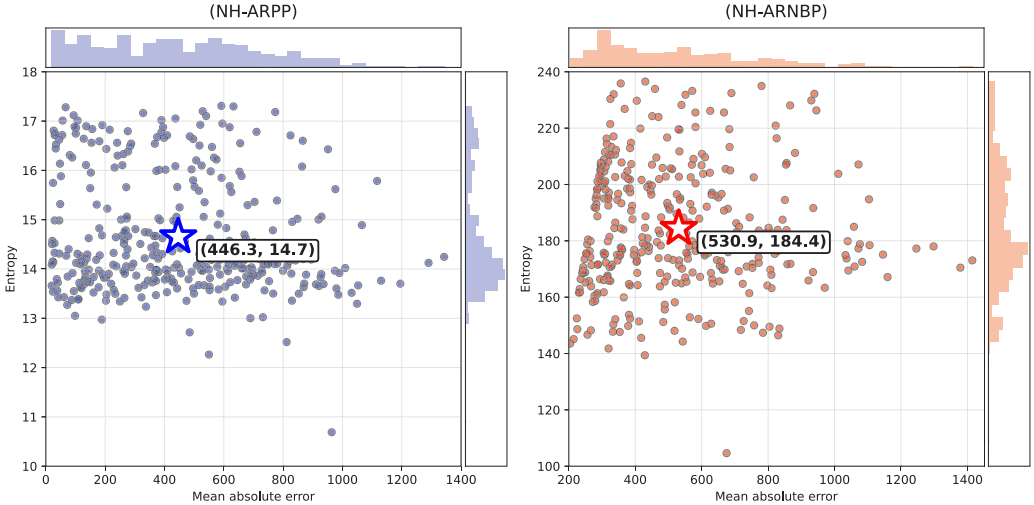


Fig. 12. Empirical decomposition of CRPS: mean absolute error vs. internal dispersion.

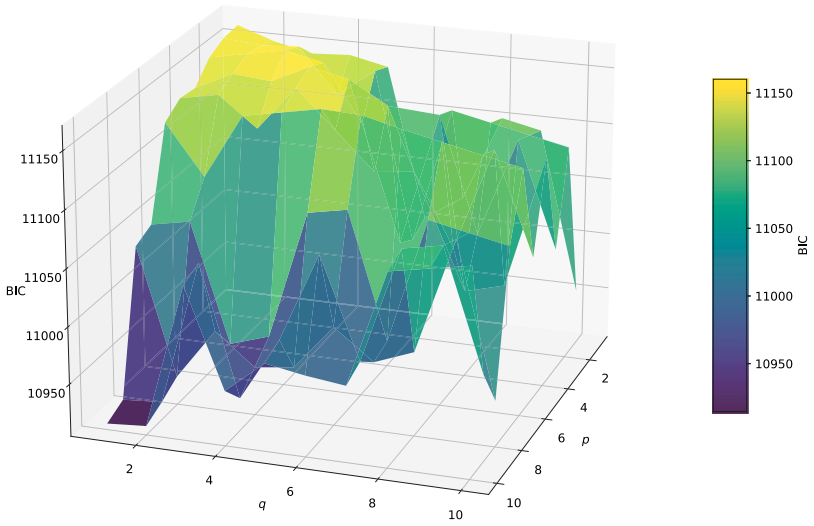


Fig. 13. BIC surface for all GP-INGARCH(p, q) specifications with $p, q \in \{1, 2, \ldots, 10\}$.

Table 4
Comparison of predictive performance across NH-ARPP, NH-ARNBP, and GP-INGARCH models.

Metric	NH-ARPP	NH-ARNBP	GP-INGARCH
Error metrics			
RMSE (↓)	403.10	**351.29**	673.51
Pbias (→0)	**−2.77**	−12.43	−38.81
NSe (↑)	0.07	**0.30**	−0.79
Probabilistic metrics			
CRPS (↓)	431.67	**346.48**	596.05
Log Score (↓)	136.89	**8.04**	8.14
Reliability (↑)	0.028	0.298	**0.40**
Resolution (↓)	39.21	2.94	1.19
95%-PIC (↑)	7.4%	72.6%	**98.9%**

Note: Arrows indicate whether higher (↑), lower (↓), or values closer to zero (→0) represent better performance. No directional arrow is assigned to Resolution because, although larger values indicate sharper predictive distributions, excessively high values may also reflect overconfident forecasts if calibration is poor.

Table 4 summarizes the predictive performance of NH-ARPP, NH-ARNBP, and GP-INGARCH across error-based and probabilistic metrics. GP-INGARCH achieves very high 95%-PIC and reliability, but primarily by producing prediction intervals that are extremely wide, which reduces operational informativeness. In contrast, NH-ARNBP achieves a more balanced trade-off: it produces sharper intervals while improving calibration relative to NH-ARPP and maintaining a reasonable level of coverage. In strongly non-stationary settings, an observation-driven recursion may accommodate variability by spreading probability mass broadly, whereas the NH-ARNBP framework explicitly represents slowly varying distributional heterogeneity through its non-homogeneous negative binomial layer while reserving the AR component for short-run dependence. This leads to poor point-forecast performance (RMSE = 673.51, NSE = −0.79) and the worst overall CRPS among all models.

It is important to note that there is a significant difference in computational flexibility. While NH-ARNBP took approximately three minutes to train and simulate, GP-INGARCH lasted almost 2.5 h in this process, running on a personal computer with an Intel® Core® i5 1.6 GHz CPU (4 cores, 8 GB RAM).

4. Managerial insights

Access to probabilistic forecasts of truck arrivals at a port yields tangible operational, economic, and environmental benefits. In contrast, using a poorly calibrated forecasting model may lead to operational inefficiencies. From an operational standpoint, a probabilistic forecasting model that provides a detailed representation of uncertainty enables the anticipation of peak demand or short-term variability and the adjustment of resources accordingly. Prior studies have demonstrated the potential of such improvements. For example, Chen et al. (2011) reports that optimizing truck arrival patterns — such as through dynamic appointment systems — can reduce average waiting times by up to 50% under high-congestion conditions, compared to a baseline without such optimization. This reduction translates into hundreds of hours of waiting time saved daily at a high-traffic port.

To quantify the impact, consider a Chilean port with 2000 truck arrivals per day. If each truck waits on average 60 min due to poor coordination, approximately 6000 liters of diesel are consumed daily while idling (assuming 3 liters per hour per truck). This corresponds to around CLP (Chilean pesos) 5.7 million (US$6000) lost in fuel costs and more than 16 tons of CO_2 emitted per day in the port area. A forecasting model that reduces idle time by half would yield savings of 3000 liters per day (CLP 2.85 million) and avoid 8 tons of CO_2 emissions daily. Over a winter season (90 days a year), this would result in more than CLP 250 million in avoided costs and more than 720 tons of CO_2 not emitted, highlighting the substantial economic and environmental impact of improved arrival management.

These facts reinforce the value of tools such as the NH-ARNBP model, which enable the prediction of critical periods and support policies such as variable toll pricing to redistribute traffic flows. Following (Chen et al., 2011), the model proposed in this research could guide such policies by identifying periods of expected saturation in advance. Accurate prediction transforms port management from a reactive stance to an anticipatory planning approach, enabling efficient allocation of shifts, cranes, and yard space. In contrast, low-quality models — such as homogeneous Poisson models — fail to capture stochastic variability, leading to unexpected queues and infrastructure underutilization. Islam and Olsen (2014) showed that collaborative strategies such as truck-sharing, when based on reliable demand forecasts, avoided 17 million kilometers annually and saved €25.5 million in costs. This efficiency stems from synchronized flows and reduced empty trips.

From an economic and urban-planning perspective, access to reliable predictive tools is equally significant. Port logistics do not occur in isolation—they both influence and depend on the city and terrestrial transport networks. Wilmsmeier and Monios (2015) noted that port reforms in Latin America achieved "short-term technical efficiency gains at individual ports, but also led to a loss of public sector planning capacity and lack of integrated transport policy". Chile is a clear example: while private concessions have improved terminal-level operations

— doubling or even quadrupling container handling capacity over the past two decades — no coordinated national strategy was developed to manage the port system as a whole.

In this context, having accurate forecasts can help compensate for the lack of macro-level coordination by enabling each port to anticipate its operational needs. For example, the NH-ARNBP model can detect a sustained increase in truck flows into central Chile months in advance. Reliable predictions would allow for early identification of critical weeks and evaluation of modal shifts. Qi and Shen (2018) emphasized that modern urban management requires data-driven anticipatory decisions. The NH-ARNBP model fits into this framework by eventually transforming historical data into early warnings, enabling more resilient cities and better synchronization between urban planning and port operations.

Another key managerial aspect is the adoption of digital technologies and information platforms, which have proven effective in improving port logistics efficiency. In Qingdao, China, automation and real-time monitoring increased efficiency by 30% and reduced labor costs by 85% (D'Amico et al., 2021). In Europe, ports like Hamburg and Rotterdam implemented coordination systems that reduced waiting times and emissions. Hamburg, for instance, reduced its annual CO_2 emissions by 12,000 tons through energy management. Li et al. (2016) emphasized that smart cities transform operations through scalability and connectivity, as exemplified by collaborative platforms such as Uber Freight, which dynamically adjust services to meet demand. Buldeo Rai et al. (2017) proposed crowd logistics, in which stakeholders share logistical capacity and information through digital platforms, thereby achieving faster deliveries and lower emissions.

However, Islam and Olsen (2014) warned that a lack of trust and coordination between actors can hinder collaboration. Therefore, reliable data would enable port managers to negotiate effective agreements and build consensus. The NH-ARNBP model can be integrated into such systems as a predictive component to generate short-, medium-, and long-term forecasts. Additionally, it produces daily trajectories with explicit uncertainty estimates, facilitating the identification of periods of high demand. This information supports proactive planning, reduces waiting times, and mitigates operational risks. Deterministic or less informative models do not provide this level of support. The absence of probabilistic forecasts limits preventive action and exposes the system to higher costs, congestion, emissions, and overall risks.

The NH-ARNBP framework is intended to support daily probabilistic planning and congestion-risk anticipation, providing calibrated uncertainty estimates that can inform staffing decisions, yard preparedness, and appointment-policy adjustments at the day-to-week horizon. While intra-day forecasting is crucial for gate-level control, daily probabilistic forecasts constitute the natural planning layer for staffing, congestion anticipation, and contingency design, which motivates the modeling resolution adopted in this work. Additionally, developing an hourly-resolution model requires dedicated within-day structure (e.g., within-day seasonality, peaked arrival patterns, and queue spillover) and therefore lies outside the scope of the present work. Similarly, exogenous drivers such as holidays and weekday effects may influence arrivals; in this framework, part of their impact is absorbed through the trend–seasonal–residual decomposition, and incorporating explicit covariates constitutes a natural extension.

5. Conclusion

This research presents a methodological proposal for modeling and forecasting daily truck arrivals at port terminals that accounts for the intrinsic statistical characteristics of this type of logistics flow: discreteness, nonstationarity, serial correlation, and overdispersion. From this standpoint, we introduce the Non-Homogeneous Autoregressive Negative Binomial Process (NH-ARNBP). This stochastic model integrates an autoregressive time series structure with a non-homogeneous negative binomial distribution.

The proposed model decomposes the original count series into three interpretable components: trend, seasonal pattern, and stochastic residual. The latter is then modeled using an autoregressive structure. Future trajectories are generated by simulating innovations of the autoregressive model, to which the trend and seasonal components are added, and then projected back to the original count space using the inverse transform method for the non-homogeneous binomial distribution. This sequence ensures that predictions preserve the structure, discreteness, and stochasticity observed in historical observations.

To validate the model's performance, we use a real-world case study from a major Chilean port terminal and compare the NH-ARNBP model against a classical-based Non-Homogeneous Autoregressive Poisson Process (NH-ARPP), which shares the same autoregressive logic but uses a Poisson distribution to model non-stationarity. In addition, a Poisson-based Integer-Valued Generalized Autoregressive Conditional Heteroscedasticity (GP-INGARCH) is considered as a state-of-the-art benchmark for the proposed model. The results indicate that NH-ARNBP yields clear probabilistic advantages over NH-ARPP. In particular, it not only improves point forecasting but also improves prediction-interval coverage and reduces the logarithmic score, yielding an informative and sharp predictive distribution. When compared with GP-INGARCH, the NH-ARNBP exhibits a more operationally balanced trade-off between calibration, sharpness, and point accuracy. These improvements are particularly relevant in the presence of overdispersion, where Poisson-based models systematically underestimate variability and yield overly confident forecasts.

In addition to its methodological contribution, the model has practical implications for port logistics management. Accurate probabilistic forecasts of truck arrivals enable the anticipation of high-demand periods, the optimization of resource allocation, the reduction of waiting times, and the minimization of pollutant emissions. The model provides a quantitative basis for designing dynamic appointment policies, congestion mitigation strategies, and coordination mechanisms among stakeholders. Furthermore, it can be integrated into collaborative platforms, innovative port systems, or optimization models, thereby supporting day-to-week tactical planning and enhancing resilience to operational uncertainty.

A limitation of this work is that the empirical validation was performed using data from a single port in the Valparaíso region. While the results are encouraging, the statistical properties of truck arrivals may differ across ports with different scales, infrastructure conditions, and operating contexts. Future work should therefore evaluate the proposed approach using data from multiple ports to further assess its generalizability and robustness.

This research opens multiple avenues for future work. First, the NH-ARNBP framework could be extended to incorporate additional covariates — such as holiday effects, vessel schedules, or weather disruptions — within the autoregressive component or the distribution-mapping stage. Second, future efforts may explore alternative temporal resolutions, such as hourly truck arrivals, which are highly relevant for operational planning at port terminals. Third, model parameters could be updated using finer time granularities (e.g., weekly or rolling estimation) or via Bayesian methods to stabilize estimation in periods with limited data. Finally, expanding the evaluation to include optimization-based applications of the predictive distributions (e.g., truck appointment scheduling or yard resource planning) is a promising direction for further quantifying the model's managerial impact.

CRediT authorship contribution statement

Vicente Valenzuela-Riquelme: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Marcel Favereau:** Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Carlos A. Monardes-Concha:** Writing – review & editing, Validation, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment and funding sources

This work was supported by FIC-R-2023 Gobierno Regional de Valparaíso BIP 40058874.

Data availability

The authors do not have permission to share data.

References

Abdelmagid, A., Gheith, M., & Eltawil, A. (2021). A comprehensive review of the truck appointment scheduling models and directions for future research. *Transport Reviews*, *41*, 1–24. http://dx.doi.org/10.1080/01441647.2021.1955034.

Achcar, J. A., Rodrigues, E. R., Paulino, C. D., & Soares, P. (2010). Non-homogeneous Poisson models with a change-point: an application to ozone peaks in Mexico city. *Environmental and Ecological Statistics*, *17*(4), 521–541. http://dx.doi.org/10.1007/s10651-009-0114-3.

Alexopoulos, C., Goldsman, D., Fontanesi, J., Kopald, D., & Wilson, J. R. (2008). Modeling patient arrivals in community clinics. *Omega*, *36*(1), 33–43. http://dx.doi.org/10.1016/j.omega.2005.07.013.

Álvaro, R.-S., Comendador, F. G., Valdés, R. A., Pérez-Castán, J., Montes, R. B., & Serrano, S. C. (2019). Assessment of airport arrival congestion and delay: Prediction and reliability. *Transportation Research Part C: Emerging Technologies*, *98*, 255–283. http://dx.doi.org/10.1016/j.trc.2018.11.015.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Hoboken, NJ: Wiley.

Bracher, J., Ray, E. L., Gneiting, T., & Reich, N. G. (2021). Evaluating epidemic forecasts in an interval format. *PLoS Computational Biology*, *17*(2), Article e1008618. http://dx.doi.org/10.1371/journal.pcbi.1008618.

Brooks, M. R., & Wilmsmeier, G. (2017). Coastal shipping in Chile: Opportunities and challenges. *Transportation Research Record*, *2611*, 32–40. http://dx.doi.org/10.3141/2611-04.

Buldeo Rai, H., Verlinde, S., Merckx, J., & Macharis, C. (2017). Crowd logistics: An opportunity for more sustainable urban freight transport? *European Transport Research Review*, *9*, 39. http://dx.doi.org/10.1007/s12544-017-0256-6.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press, http://dx.doi.org/10.1017/CBO9781139013567.

Cavanaugh, J. E., & Neath, A. A. (2019). The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, *11*(3), Article e1460. http://dx.doi.org/10.1002/wics.1460.

Chen, C. W. S., & Lee, S. (2016). Generalized Poisson autoregressive models for time series of counts. *Computational Statistics & Data Analysis*, *99*, 51–67. http://dx.doi.org/10.1016/j.csda.2016.01.009.

Chen, C. W. S., So, M. K. P., Li, J. C., & Sriboonchitta, S. (2016). Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology*, *31*, 73–90. http://dx.doi.org/10.1016/j.stamet.2016.02.001.

Chen, X., Zhou, X., & List, G. F. (2011). Using time-varying tolls to optimize truck arrivals at ports. *Transportation Research Part E: Logistics and Transportation Review*, *47*, 965–982. http://dx.doi.org/10.1016/j.tre.2011.04.001.

D'Amico, G., Szopik-Depczyńska, K., Dembińska, I., & Ioppolo, G. (2021). Smart and sustainable logistics of port cities: A framework for comprehending enabling factors, domains and goals. *Sustainable Cities and Society*, *69*, Article 102801. http://dx.doi.org/10.1016/j.scs.2021.102801.

Davis, R. A., & Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, *96*(3), 735–749. http://dx.doi.org/10.1093/biomet/asp029.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*(4), 1057–1072. http://dx.doi.org/10.2307/1912517.

Dincer, N. G., Demir, S., & Yalçin, M. O. (2022). Forecasting COVID19 reliability of the countries by using non-homogeneous Poisson process models. *New Generation Computing*, *40*(4), 1143–1164. http://dx.doi.org/10.1007/s00354-022-00183-1.

Favereau, M., Lorca, Á., Negrete-Pincetic, M., & Vicuña, S. (2022). Robust streamflow forecasting: a student's t-mixture vector autoregressive model. *Stochastic Environmental Research and Risk Assessment*, *36*, 3979–3995. http://dx.doi.org/10.1007/s00477-022-02241-y.

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, *27*(6), 923–942. http://dx.doi.org/10.1111/j.1467-9892.2006.00496.x.

Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, *104*(488), 1430–1439. http://dx.doi.org/10.1198/jasa.2009.tm08270.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. http://dx.doi.org/10.1198/016214506000001437.

Gning, L. D., Diop, A., Diagne, M. L., & Tchuenche, J. (2022). Modelling COVID-19 in Senegal and China with count autoregressive models. *Modeling Earth Systems and Environment*, *8*(4), 5713–5721. http://dx.doi.org/10.1007/s40808-022-01483-7.

Good, I. J. (2018). Rational decisions. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *14*(1), 107–114. http://dx.doi.org/10.1111/j.2517-6161.1952.tb00104.x.

Goulding, J., Preston, S., & Smith, G. (2016). Event series prediction via non-homogeneous Poisson process modelling. *Proceedings of the IEEE International Conference on Data Mining (ICDM 2016)*, 821–826. http://dx.doi.org/10.1109/ICDM.2016.0027.

Gouriéroux, C., & Lu, Y. (2018). Negative binomial autoregressive process with stochastic intensity. *Journal of Time Series Analysis*, *40*(2), 225–247. http://dx.doi.org/10.1111/jtsa.12441.

Gracia, M., Mar-Ortiz, J., & Vargas, E. (2025). Truck appointment scheduling: A review of models and algorithms. *Mathematics*, *13*, 503. http://dx.doi.org/10.3390/math13030503.

Guarnaccia, C., Quartieri, J., Tepedino, C., & Rodrigues, E. R. (2016). A time series analysis and a non-homogeneous Poisson model with multiple change-points applied to acoustic data. *Applied Acoustics*, *114*, 203–212. http://dx.doi.org/10.1016/j.apacoust.2016.07.022.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559–570. http://dx.doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511973420.

Humphrey, G. B., Gibbs, M. S., Dandy, G. C., & Maier, H. R. (2016). A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. *Journal of Hydrology*, *540*, 623–640. http://dx.doi.org/10.1016/j.jhydrol.2016.06.026.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts, URL https://otexts.com/fpp2/.

Islam, S., & Olsen, T. (2014). Truck-sharing challenges for hinterland trucking companies: A case of the empty container truck trips problem. *Business Process Management Journal*, *20*(2), 290–334. http://dx.doi.org/10.1108/BPMJ-03-2013-0042.

Jordan, A., Krüger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringrules. *Journal of Statistical Software*, *90*(12), http://dx.doi.org/10.18637/jss.v090.i12.

Kalahasthi, L. K., Sánchez-Díaz, I., Castrellon, J. P., Gil, J., Browne, M., Hayes, S., & Sentís Ros, C. (2022). Joint modeling of arrivals and parking durations for freight loading zones: Potential applications to improving urban logistics. *Transportation Research Part A: Policy and Practice*, *166*, 307–329. http://dx.doi.org/10.1016/j.tra.2022.11.003.

Kieu, L. M. (2018). Analytical modelling of point process and application to transportation. In J. Zhou, & F. Chen (Eds.), *Human and machine learning: human–computer interaction series* (pp. 385–408). Springer, http://dx.doi.org/10.1007/978-3-319-90403-0_19.

Kieu, L. M., & Cai, C. (2018). Stochastic collective model of public transport passenger arrival process. *IET Intelligent Transport Systems*, *12*(9), 1027–1035. http://dx.doi.org/10.1049/iet-its.2018.0085.

Lancia, C., & Lulli, G. (2020). Predictive modeling of inbound demand at major European airports with Poisson and pre-scheduled random arrivals. *European Journal of Operational Research*, *280*(1), 179–190. http://dx.doi.org/10.1016/j.ejor.2019.06.056.

Law, A. M. (2014). *Simulation modeling and analysis* (5th ed.). McGraw-Hill Education.

Lewis, P. A. W., & Shedler, G. S. (1976). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, *26*(3), 403–413. http://dx.doi.org/10.1147/rd.205.0465.

Li, F., Nucciarelli, A., Roden, S., & Graham, G. (2016). How smart cities transform operations models: a new research agenda for operations management in the digital economy. *Production Planning and Control*, *27*(6), 514–528. http://dx.doi.org/10.1080/09537287.2016.1147096.

Li, N., Sheng, H., Wang, P., Jia, Y., Yang, Z., & Jin, Z. (2022). Modeling categorized truck arrivals at ports: Big data for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, *PP*, 1–17. http://dx.doi.org/10.1109/TITS.2022.3219882.

Liu, M., Li, Q., & Zhu, F. (2018). Threshold negative binomial autoregressive model. *Statistics*, *52*(5), 1277–1300. http://dx.doi.org/10.1080/02331888.2018.1546307.

Liu, M., Li, Q., & Zhu, F. (2019). Self-excited hysteretic negative binomial autoregression. *Advances in Statistical Analysis*, *103*(1), 115–140. http://dx.doi.org/10.1007/s10182-019-00360-6.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer, http://dx.doi.org/10.1007/978-3-540-27752-1.

Majeske, K. D. (2007). A non-homogeneous Poisson process predictive model for automobile warranty claims. *Reliability Engineering & System Safety*, *92*(2), 243–251. http://dx.doi.org/10.1016/j.ress.2005.12.004.

Martínez-López, A., & Chica, M. (2020). Joint optimization of routes and container fleets to design sustainable intermodal chains in Chile. *Sustainability*, *12*(6), 2221. http://dx.doi.org/10.3390/su12062221.

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*(10), 1087–1096.

Monios, J., Bergqvist, R., & Woxenius, J. (2018). Port-centric cities: The role of freight distribution in defining the port-city relationship. *Journal of Transport Geography*, *66*, 53–64. http://dx.doi.org/10.1016/j.jtrangeo.2017.11.012.

Omer, P. A., & Hussian, T. M. (2023). Fitting of generalized Poisson regression and negative binomial regression models for analyzing of count time series event. *Mitanni Journal of Humanitarian Sciences*, *4*(2), 116–126. http://dx.doi.org/10.25156/ptjhss.v4n2y2023.

Qi, W., & Shen, Z.-J. M. (2018). A smart-city scope of operations management. *Production and Operations Management*, http://dx.doi.org/10.1111/poms.12928.

Quddus, M. A. (2008). Time series count data models: An empirical application to traffic accidents. *Accident Analysis and Prevention*, *40*(5), 1732–1741. http://dx.doi.org/10.1016/j.aap.2008.06.011.

Raza, Z., Svanberg, M., & Wiegmans, B. (2020). Modal shift from road haulage to short sea shipping: a systematic literature review and research directions. *Transport Reviews*, http://dx.doi.org/10.1080/01441647.2020.1714789.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*, W05521. http://dx.doi.org/10.1029/2009WR008328.

Ripley, B. D. (1987). *Stochastic simulation*. Wiley.

Rodriguez, S. A., Rodrigo, A., & Aguayo, M. M. (2021). A simulation-optimization approach for the facility location and vehicle assignment problem for firefighters using a loosely coupled spatio-temporal arrival process. *Computers & Industrial Engineering*, *157*, Article 107242. http://dx.doi.org/10.1016/j.cie.2021.107242.

Ross, S. M. (2014). *Introduction to probability models* (11th ed.). Academic Press.

Sparks, R. S., Keighley, T., & Muscatello, D. (2011). Optimal exponentially weighted moving average (EWMA) plans for detecting seasonal epidemics when faced with non-homogeneous negative binomial counts. *Journal of Applied Statistics*, *38*(10), 2165–2181. http://dx.doi.org/10.1080/02664763.2010.545184.

Van der Spoel, S., Amrit, C., & Hillegersberg, J. (2016). Predictive analytics for truck arrival time estimation: A field study at a European distribution center. *International Journal of Production Research*, *In Press*, 1–21. http://dx.doi.org/10.1080/00207543.2015.1064183.

Sultana, Z., Mishra, S., Cherry, C. R., Golias, M. M., & Tabrizizadeh Jeffers, S. (2018). Modeling frequency of rural demand response transit trips. *Transportation Research Part A: Policy and Practice*, *118*, 494–505. http://dx.doi.org/10.1016/j.tra.2018.10.006.

Wilmsmeier, G., & Monios, J. (2015). Institutional structure and agency in the governance of spatial diversification of port system evolution in latin america. *Journal of Transport Geography*, http://dx.doi.org/10.1016/j.jtrangeo.2015.02.004.

Wilmsmeier, G., & Sánchez, R. J. (2017). Evolution of national port governance and interport competition in Chile. *Research in Transportation Business & Management*, http://dx.doi.org/10.1016/j.rtbm.2017.01.001.

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, *389*(1), 58–71. http://dx.doi.org/10.1016/j.jmaa.2011.11.042.

Zhu, F. (2012b). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, *389*(1), 58–71.