ORIGINAL ARTICLE

Using big data to estimate the impact of cruise activity on congestion in port cities

Agustina Calatayud[1] · Santiago Sánchez González[1] · Jose Maria Marquez[1]

Abstract

By using big data that are generated by crowdsourced mobility platforms and GPS devices, policymakers can now better understand the relationship between cruise activities and traffic congestion, identify key areas in the road network that need traffic flow improvement, and design appropriate measures to take advantage of cruise tourism while mitigating its effects on urban mobility. This study uses big data and econometric techniques to analyze the relationship between traffic congestion and cruise activity for five of the busiest cruise destinations in South America and the Caribbean: Buenos Aires, Callao, Cartagena, Montevideo, and Rio the Janeiro. Over 80 million observations were collected from automatic identification system (AIS) and crowdsourced traffic data platforms for 2019. Results suggest that cruise activity causes a 12% increase in urban congestion. The analysis of big data presented in this paper facilitates the design of customized interventions to ease congestion according to the temporal dynamics of traffic flows, thus leading towards more efficient traffic management.

Keywords Cruise activity · Congestion · Maritime transport · Tourism · Big data · Latin America

1 Introduction

Prior to the outbreak of the coronavirus disease 2019 (COVID-19) pandemic, the cruise industry was one of the largest growing sectors of the tourism industry (Brida and Zapata 2010; Polat 2015; Sun et al. 2014). Between 2009 and 2019, the number of cruise passengers grew by 60%, eventually serving 30 million persons; creating 1.2 million jobs; and generating US$155 billion in revenue globally (CLIA 2021). Despite the unpredictable disruption of the growth trends caused by the pandemic, with less than 600,000 passengers completing their cruises in 2020 (Radic et al. 2021), current

✉ Agustina Calatayud
mcalatayud@iadb.org

[1]   Inter-American Development Bank, Washington, DC 20577, USA

trends suggest a quick recovery in the forthcoming years. After the industry essentially shut down in the early months of 2020, between July and December of that year the cruise sector was able to complete over 200 sailings (CLIA 2021). With the rapid deployment of COVID-19 vaccines in the USA in the first half of 2021, the industry is resuming operations in the Caribbean, the region with the highest cruise activity in the world (CLIA 2021).

The cruise industry is linked to a wide range of externalities: political, sociocultural, and environmental. These externalities can be either positive or negative (Brida et al. 2010). Cruise activity has proven to generate employment and increase income in host communities while decreasing crime owing to increasing government expenditures on vigilance (McNeill and Wozniak 2018). Nevertheless, the impacts of cruises are also associated with increasing pollution, noise contamination, loss of neighborhood lifestyle, and overcrowded public spaces (Ruiz-Guerra et al. 2019; Fabregat et al. 2021; Valverde and Capote 2021). To reduce these adverse effects, main cruise destinations such as Dubrovnik and Malaga have set a limit to the daily number of passengers received (Coffey 2018). This is particularly relevant since, according to residents' perspectives, among all forms of tourism, the type that is most detrimental to urban issues is cruising (Jordan et al. 2020). While the negative effects of cruise activities have gained attention in public opinion, academic literature is limited and mostly anecdotal, particularly regarding the effects on urban congestion. Traffic congestion in the port area has a twofold effect: On the one hand, it generates dissatisfaction for the residents of the area, increasing negative externalities. On the other hand, transport to and from the port is the first and last impression of the destination, which are among the most valued factors for cruise tourists (Baker 2015).

This research focuses on studying the impact of cruise activity on urban congestion adjacent to the port area in five of the most important destinations in South America and the Caribbean. This region is the biggest cruise region in the world, accounting for over a third of the global industry (CLIA 2021). Specifically, the ports that were chosen for the analysis are Buenos Aires (Argentina), Callao (Peru), Cartagena (Colombia), Montevideo (Uruguay), and Rio de Janeiro (Brazil). These cities were among the main tourist destinations in the region in 2018 (CLIA 2019). Given the high growth of cruise activities in South America and the Caribbean, the results from this research can provide useful insights for policymakers to mitigate the negative externalities in their cities, while leveraging the industry's benefits.

The paper is organized as follows: first, we review the existing literature on cruise activity and its negative externalities including congestion; second, we present the methodology used to analyze the impact of cruise activity on urban congestion in port cities; third, we present the results from the empirical analysis; fourth, we discuss how these results can help improve policymaking in port cities; and finally, we present our conclusions and further research needs on this topic.

2 Literature review

The cruise industry was, until the outbreak of the COVID-19 pandemic, the highest growing sector within maritime transport. Between 2009 and 2019, the number of passengers grew by 67%, reaching 29.7 million passengers globally (CLIA 2021). With an annual contribution to the world economy of over US$155 billion and the creation of more than 1.2 million jobs, the cruise industry has been recognized as a key economic pillar by governments in North America, Europe, Asia, and the Caribbean (CLIA 2021). Despite these broader economic benefits, there is substantial discussion of the short-term gains from cruise activity compared with the social costs generated in port cities (McNeill and Wozniak 2018; Rosa-Jiménez et al. 2018). Recent studies suggest that the impact of cruise tourism expenditure, although positive, is relatively low at the local level compared with the large economic impact of freight-related activities (Artal-Tur et al. 2018; Chen et al. 2019). In addition, evidence shows increased pollution and environmental costs for port cities (Perdiguero et al. 2020; Vicente-Cera et al. 2020). In some cases, environmental costs are up to seven times higher than the economic benefits from the cruise industry (Carić and Mackelworth 2014), creating a need for a more balanced consideration of the effects of tourism on the environment (Robinson et al. 2019).

The relationship between congestion and cruise activities has gained public attention. This debate has acquired particular relevance because of the public's view that cruise ships bring the least positive impacts to their destinations, compared with other types of tourism, especially affecting urban concerns like traffic and crowding of public spaces (Jordan 2020). The large number of visitors from cruise ships, often arriving all at once, often exerts tremendous pressure on cities, already suffering from mobility challenges (ITF 2020). Commute peak hours in the morning usually coincide with the arrival of cruise ships. This is particularly problematic since the disembarkation process is a critical point, as passengers usually exit at the same time (Chaos et al. 2017). Moreover, most of the cruise terminals are not well served by public transport services, thus generating a large amount of coach and taxi rides that add to urban traffic (Stefanidaki and Lekakou 2014). In this context, congestion has been identified as a significant source of stress for local inhabitants (Jordan and Vogt 2017), negatively impacting their well-being (Brida et al. 2011). Congestion is also a key concern for port terminals, as it affects the passenger's experience and thus reduces the attractiveness of the port city (Santos et al. 2019).

Compared with the increasing public debate on this topic, a review of academic literature shows that there is little research available on the relationship between congestion and cruise activity. A search on Scopus using the keywords "congestion" and "cruise" resulted in 252 articles, with only 22 specifically referring to cruise liner activity. Among these articles, many are case studies that, aside from relying on perception data, do not focus specifically on congestion but on a broader number of effects related to the cruise industry. Therefore, this paper contributes in filling a research gap by addressing the following research

question: To what extent does cruise activity cause road congestion in the areas near cruise terminals? Crowdsourced big data and econometric techniques are used to provide a novel, granular understanding of traffic behavior in the vicinity of cruise terminals. We hope, through the insights provided by our paper, to facilitate the design of customized interventions, to ease congestion according to the temporal dynamics of cruise-related traffic flows, thus leading towards a more efficient traffic management in port cities.

3 Methods

To explore the research question, data on both cruise activity and traffic levels in the areas near the port were collected for five cities: Buenos Aires (Argentina), Callao (Peru), Cartagena (Colombia), Montevideo (Uruguay), and Rio de Janeiro (Brazil). These cities were chosen because they are among the busiest tourist destinations in South America and the Caribbean, representing also significant urban areas.

Data were collected from two sources. One source was AIS, which provided the data for the cruise activity. This is a very representative source of data for cruise ships, as every passenger ship, regardless of its size, is required by International Maritime Organization (IMO) regulations to have an AIS transponder onboard. The information provided by this source is made of two sets of data: Port Calls, which register every ship arrival and departure; and Position, which describes the type of vessel, vessel size, status, speed, and direction once the ship has entered a port area. Together, these two datasets accounted for more than 10 million observations in our research. *Cruise activity*, the way the term is used here, refers to the moment when cruise ships are docked to port, following the methodology proposed by Feng et al. (2020).

The second source was traffic data collected from Waze, a navigation mobile app with a high rate of adoption in the respective cities (Waze 2019). To explore the direct impact of cruise activity on urban congestion, a radius of 2 km from the port terminal was selected for data gathering, equivalent to an area of $12.57km^2$ for each city (Fig. 1). This area represents a significant part of the urban environment around the port, and if congestion was relieved within this delimitation, it is unlikely that cruise-induced congestion would be present in the rest of the city. For the five port areas, around 68 million observations were retrieved from Waze's API, for the full year 1 January to 31 December 2019. This source provides two types of data:

(i) Alerts: each Waze user self-reports alerts noticed on the road. Once an alert is reported, other users validate it by reporting on the mobile app if the alert is still present. Based on the information received by users, Waze calculates a reliability factor between 1 and 10, 10 being the most reliable. Users can report three kinds of alerts: (i) Accident, which regards collisions of every type; (ii) Hazard, which is a type of alert that can be reported for stranded vehicles or objects on the road, adverse weather conditions, and floods, among others; and (iii) Road Closed, which stands for lane closures due to demonstrations, events, maintenance, and others.

A. *Buenos Aires*

B. *Callao*

C. *Cartagena*

D. *Montevideo*

E. *Rio de Janeiro*

Fig. 1 Definition of port areas in the five selected cities

(ii) Jams: this dataset is retrieved actively by Waze through smartphones' GPS signals. When the API identifies a significant group of vehicles moving at an irregular speed, in contrast to free-flow speed, it classifies it as a jam. For each jam, Waze collects information on average speed, expected delay in overcoming the jam compared with free-flow conditions, geographical coordinates, and jam length. Information on road status is updated every 2 min.

We applied the methodology developed by Calatayud et al. (2021) to estimate congestion, based on the data provided by Waze. Here, congestion was defined as "the impedance vehicles impose on each other, due to the speed-flow relationship, in conditions where the use of a transport system approaches its capacity" (Goodwin 2004). Consequently, congestion could be estimated as the extra time

that road users experience due to the excess of vehicles on a portion of road at a particular time, resulting in slower-than-normal or free-flow speeds.

As mentioned above, the data obtained from Waze included the coordinates where a jam was detected (henceforth referred to as a "segment"), the effective speed at which traffic was moving at the jam, and the time (delay) it would take to get through the segment at the effective speed of the jam. The information regarding road status was updated every 2 min. Given that Waze does not provide data on the number of vehicles at the jam, this difficulty was overcome following Calatayud et al. (2021). First, the road network was built using a neural network model, allowing the number of lanes on each road to be calculated based on the information provided by the app, and some socioeconomic figures, such as gross domestic product (GDP) and population. Subsequently, following Yperman et al. (2005), the fundamental triangular relation between traffic flow and vehicle density was assumed, to estimate the number of vehicles that were in the jam, the distance between these vehicles, given speed, and the number of vehicles that *would* circulate given the previous conditions (Xu and González 2017; Newell 1993). Furthermore, following Daganzo (1997), the average vehicle length is assumed to be 6 m, and according to Tirachini and Gomez-Lobo (2020), the average number of persons per vehicle is assumed to be 1.4. For each jam reported by Waze, the total delay is estimated as follows:

- *Scenario 1* In this scenario, all trapped vehicles progress through the jam during interval t (the 2-min interval between data captures), meaning that their actual time was less than the interval t:

$$D_{ts} = \text{OR} * \text{La}_s * d_{ts}^{\text{w}} * \Omega_{ts} \tag{1}$$

where D refers to the total delay or the extra time it took all car users to move through the segment s during the 2-min interval t due to congestion, compared with the time it would have taken under free-flow speed; the subscripts t and s stand for the interval of time and segment, respectively; OR stands for vehicle occupancy rate, which is assumed to be 1.4 persons per vehicle; La denotes the number of lanes the congested road has; d^{w} is the extra time borne by a vehicle to get through the congested segment; and Ω refers to the number of vehicles that get into the jam within the 2-min timeframe.

Scenario 2 If a vehicle cannot go through the congested segment during interval t, then the total delay would be:

$$D_{ts} = \text{OR} * \text{La}_s * d_{ts}^{\text{max}} * \Gamma_{ts} \tag{2}$$

$$d_{ts}^{\text{max}} = \frac{d_{ts}^{\text{w}}}{t_{ts}^{\text{ef}}} \cdot t \tag{3}$$

where d^{max} refers to the maximum extra time that a vehicle could lose due to congestion in the 2-min interval reported by Waze; d^{w} is the extra time borne by a vehicle to get through the congested segment; Γ refers to the number of vehicles inside the

jam; t^{ef} refers to the total time it would take a vehicle to get through the entire jam; and t stands for the 2-min interval.

The Alerts dataset, specifically information on road closures and hazards, was used to obtain independent variables for the model, proposed to estimate the impact of cruise activity on congestion. Only the alerts with reliability level of 5 or higher were used. Given that several alerts may correspond to the same event, alerts were filtered according to the following spatial criterion: alerts of the same type reported within a radius of less than 20 m and within the next 20 min of the first report were considered to be the same alert. After this processing, a database was obtained of approximately 260,000 alert records for the five port areas analyzed in this study.

Cruise disembarkation and embarkation processes are not evenly distributed over the day; usually, offboarding and boarding take place within certain time-frames (Chaos et al. 2017). Moreover, congestion will most likely be affected by cruise activity that takes place during peak hours. Therefore, for this study, we only considered the total delay that occurred in our selected areas around the port during the following two timeframes: 8:00 AM to 1:00 PM, mainly used for passenger offboarding (morning timeframe); and 3:00 PM to 8:00 PM which principally refers to passenger boarding (afternoon timeframe). Correspondingly, cruise activity was classified according to these two timeframes, as follows: if the cruise ship berthed before 12:00 PM, which was the case for most cruise ships, passengers would likely disembark within the morning timeframe. For the modeling, *that* cruise activity would therefore be assigned to the morning timeframe. Likewise, if the ship departed before 3:00 PM, *that* cruise activity would also regard the morning period. Conversely, if the cruise arrived after 12:00 PM, cruise activity for this operation was assumed to take place in the afternoon timeframe. In this sense, all cruises that left the port after 3:00 PM would carry out the embarkation process in the afternoon timeframe, and thus this activity will be assigned to the afternoon timeframe. A sensitivity analysis was conducted by changing the hours when arrivals/departures were assigned to morning or afternoon, and results were consistent and robust to these changes, mainly driven by the arrival–departure cruise schedule (Fig. 4).

Figure 2 presents how the final data were built from the three databases (Jams, Alerts, and Cruise Activity). After filtering the database according to the morning and afternoon periods selected, 3960 observations (or timeframes) were obtained.

Next, a two-phased approach was applied to answer the research question. First, an analysis of covariance (ANCOVA) was undertaken to explore the relationship between cruise activity and urban congestion in the areas near port terminals. For each city, the time periods were identified (morning and afternoon) with and without cruise activity. A morning or afternoon period *has* cruise activity if at least one cruise ship is berthed at the port during that period. The ANCOVA was carried out to test whether there was a significant difference in the mean congestion levels in the surrounding areas of the ports between the periods with and without cruise activity. The ANCOVA results were obtained considering the following independent variables: port of origin fixed effects—a dummy variable that takes the value of 1 if cruise activity was present in the timeframe—and the interaction term between port of origin and cruise activity; a dummy variable if the day was a working day; the number of roads closed; the number of hazards reported in the area; and a
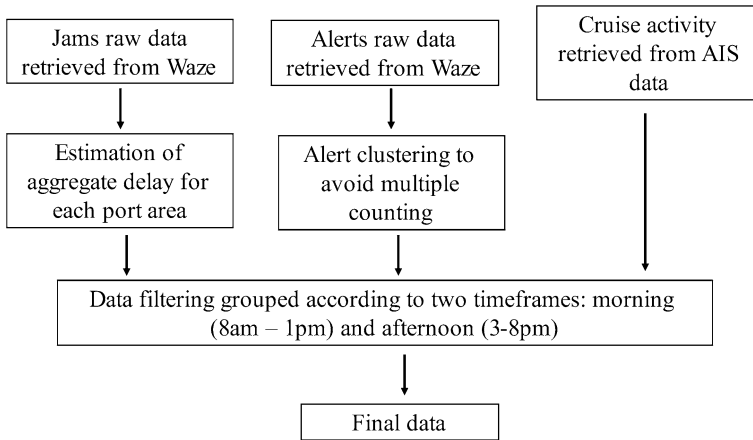
```
┌──────────────────────┐   ┌──────────────────────┐   ┌──────────────────────┐
│   Jams raw data      │   │   Alerts raw data    │   │   Cruise activity    │
│ retrieved from Waze  │   │ retrieved from Waze  │   │ retrieved from AIS   │
│                      │   │                      │   │        data          │
└──────────┬───────────┘   └──────────┬───────────┘   └──────────┬───────────┘
           │                          │                          │
           ▼                          ▼                          │
┌──────────────────────┐   ┌──────────────────────┐              │
│   Estimation of      │   │  Alert clustering to │              │
│ aggregate delay for  │   │   avoid multiple     │              │
│   each port area     │   │     counting         │              │
└──────────┬───────────┘   └──────────┬───────────┘              │
           │                          │                          │
           ▼                          ▼                          ▼
┌────────────────────────────────────────────────────────────────────────┐
│  Data filtering grouped according to two timeframes: morning            │
│          (8am – 1pm) and afternoon (3-8pm)                              │
└───────────────────────────────────┬────────────────────────────────────┘
                                     │
                                     ▼
                          ┌──────────────────────┐
                          │      Final data      │
                          └──────────────────────┘
```

Fig. 2 Database structure

variable indicating if the occurrence was during the morning or the afternoon. Using a dummy variable for the presence of cruise activity during the given timeframes enabled us to control for all time invariant factors (i.e., characteristics of each port such as infrastructure endowment, motorization rates, and location).

Finally, a fixed effects panel data regression model was applied with an autoregressive process in the disturbance term to estimate the impact of cruise activity on congestion. The autoregressive process in the error term was used since delay, our dependent variable, and the independent variables are likely to be autocorrelated. Furthermore, notice that the residuals present an AR(1) structure when not considering Eq. 5, which supports the inclusion of this procedure. City fixed effects were also included in the model to capture the idiosyncratic effects of each city analyzed, such as quality of infrastructure, motorization rate, etc. (Eq. 4). Equation 5 presents the autoregressive process of the disturbance term:

$$D_{NTx1} = Z_{NTxL} + X_{NTxK}\beta_{Kx1} + U_{NTx1} \tag{4}$$

$$U_{NTx1} = \alpha U_{NT-1x1} + \in_{NTx1} \tag{5}$$

where D is a vector that contains the information for the total delay at the port area; N stands for the number of ports; T is the time period; Z is a matrix of unobserved fixed effects; U represents the error term assumed to be independent and identically distributed; and $\in$ is the error term of the autoregressive process. X is a matrix comprising the independent variables that explain congestion: "cruises," which refers to the number of cruise ships that either arrived or departed from the cruise terminal either in the morning or in the afternoon timeframe; "afternoon," which is a dummy variable taking the value of 1 if the timeframe is between 3:00 PM and 8:00 PM; "working day," which is also a dummy variable that takes the value of 1 if the observation is during a regular labor day; "hazard," referring to the logarithmic

value of the number of hazards in the area; and "road closure," which also refers to the logarithm of the number of roads closed in the area. These variables, as well as total delay, have been introduced in logarithms in the econometric model to obtain elasticities and semi-elasticities, to facilitate the interpretation of the results. Furthermore, using logarithms smooths the impact that outliers could have on results. Combining Eqs. 4 and 5 linearly, the model estimated was:

$$d_{it} = z_{it}^1 + \cdots + z_{it}^l + \beta_0 + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + u_{it}$$

$$d_{it} - \overline{d}_{it} = z_{it}^1 + \cdots + z_{it}^l + \beta_0 + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + u_{it} - \left(z_{it}^1 + \cdots + z_{it}^l + \beta_0 + \beta_1 \overline{x}_{1it} + \cdots + \beta_k \overline{x}_{kit} + \overline{u}_{it}\right) \quad (6)$$

$$\ddot{d}_{it} = \beta_0 + \beta_1 \ddot{x}_{1it} + \beta_2 \ddot{x}_{2it} + \cdots + \beta_k \ddot{x}_{kit} + \left(\alpha \ddot{u}_{it-1} + \ddot{\varepsilon}_{it}\right)$$

where two dots over the variables indicate the difference between the observed data $\left(y_{it}\right)$ and the mean of each port $\left(\overline{y}_t\right)$, namely: $\ddot{y}_{it} = y_{it} - \overline{y}_t$.

4 Results

Figure 3 shows cruise activity for each of the five ports under analysis. Overall, we identified 2120 cruise ships calling at the selected ports in 2019. Cartagena accounted for over 30% of the total and was also the port with the highest cruise activity: 56% of the timeframes in 2019 had at least one cruise activity, followed by Rio de Janeiro and Buenos Aires, with values of 43% and 37%, respectively. Callao was the port that registered the lowest cruise activity, with a value of 14%.

Cruise activity presented similar temporal patterns at each of the five port cities (Fig. 4). Most of the cruises arrived in the morning timeframe. In fact, 91% of the ships arrived before 12:00 PM, and two thirds of the total arrivals took place
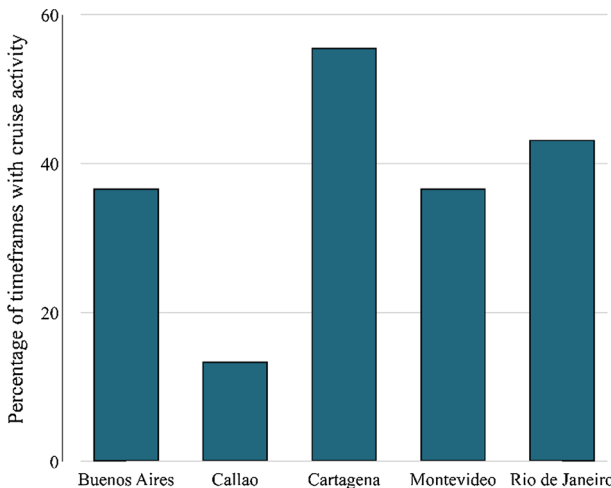

Fig. 3 Distribution of cruise activity (selected ports, 2019). *Source* own elaboration using Marine Traffic data
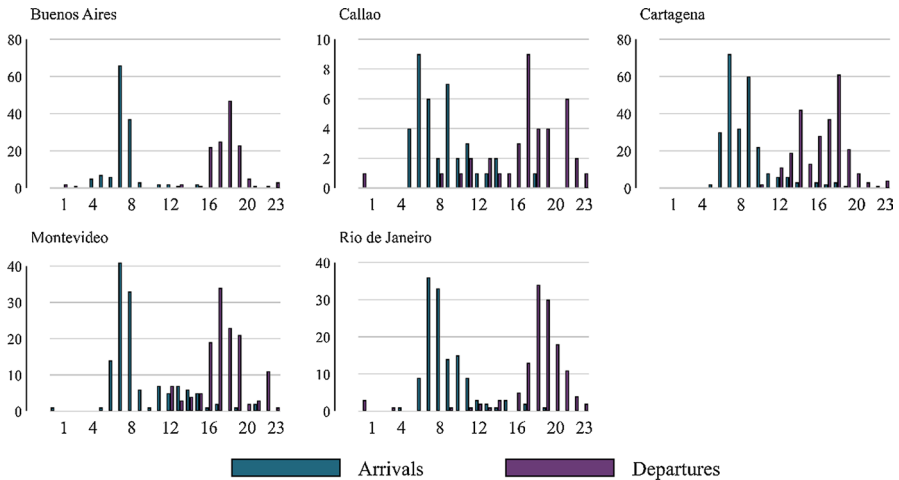
Fig. 4 Number of cruises by hour (selected ports, 2019). *Source* own elaboration using Marine Traffic data

between 7:00 AM and 9:00 AM. Arrivals at the port of Buenos Aires and Montevideo were more concentrated in this time period than in the rest of the ports. In the case of Buenos Aires, almost 80% of arrivals took place between 7:00 AM and 9:00 AM. Similarly, departures tended to concentrate in the afternoon. A share of 81% of departures took place after 3:00 PM.

Table 1 presents the main congestion statistics by port city. Average total delay refers to the mean congestion over the timeframes (morning-afternoon) for 2019 in the selected area. Each port area had different levels of average delay; the highest average delays were observed in Rio de Janeiro and Buenos Aires. In these cities, port terminals are located very closely to the epicenter of economic and social activities, therefore resulting in cruise traffic mixing with regular heavy traffic in already highly congested areas. In contrast, the port terminals in Callao and Cartagena are located outside of central areas, with the Callao cruise terminal situated as far as 15 km away from downtown Lima. Other factors such as population size,

Table 1 Average delay in selected port areas (2019)

City	Average total delay (h)	Standard deviation	Average total delay morning (h)	Average total delay afternoon (h)
Buenos Aires	5492	3747	6380	4608
Callao	278	133	256	301
Cartagena	313	248	387	239
Montevideo	2420	2032	2515	2325
Rio de Janeiro	9371	5819	9449	9293

Source Own elaboration using Waze data

infrastructure endowment, and motorization rates also affected the various levels of average delay reported in each city. The second column of Table 1 presents the standard deviations, with all five port areas presenting a high deviation. The largest coefficient of variation corresponds to Montevideo, with a value of 0.84. Despite the differences among port cities, Table 1 shows similar time patterns in terms of total delay. Except for Callao, delays in the morning are more severe, especially in the cases of Cartagena and Buenos Aires.

Table 2 presents the average total delay for each city considering cruise activity. As explained in the *Methods* section, cruise activity was present in a port area when at least one cruise ship arrives at port or departs from it within the given timeframe. In Buenos Aires, Cartagena, Montevideo, and Rio de Janeiro, total delay levels were higher on days when there was cruise activity. Results are statistically significant at a 1% confidence level except for Rio de Janeiro, where they are significant at 10%. The largest difference in the level of total delay with and without cruise activity was found in Buenos Aires, where congestion increases by 30% around the port area with cruise activity. Only the port of Callao presented a larger average total delay when there was no cruise activity. However, this difference was not statistically significant. Considering all the ports in our sample, results showed that the average total delay is higher when there was cruise activity, and the difference in delay with and without cruise activity was statistically significant.

The results were further explored by applying the ANCOVA model to control for road closures in the area, hazards, and working days. This model takes into consideration an interaction term between the port and cruise activity, to control for time invariant factors such as motorization rate, idiosyncrasy, and commercial activity in the surrounding area. Figure 5 shows the marginal effects of the logarithm of the delay by port and cruise activity considering average hazards and road closures and a typical working day. The logarithm of the total delay was used as the dependent variable with a twofold purpose: to exclude outliers and to facilitate interpretation. Three of the five cities—Buenos Aires, Cartagena, and Montevideo—present significantly higher congestion when there is cruise activity at the port. On

Table 2 Average total delay with and without cruise activity (2019)

City	With cruises (h)	Without cruises (h)	Difference (with − without)	Mean difference test
Buenos Aires	6394	4963	1433	0.00***
Callao	271	280	− 8	0.72
Cartagena	358	257	101	0.00***
Montevideo	2880	2163	816	0.00**
Rio de Janeiro	9688	9102	585	0.08*
Average	4207	3189	1018	0.00***

The reported *p*-value stands for the alternative hypothesis of positive difference ($Ha : diff = mean(Delay|CruiseActivity) - mean(Delay|WithoutCruiseActivity) > 0$)

Source own elaboration using Marine Traffic and Waze data

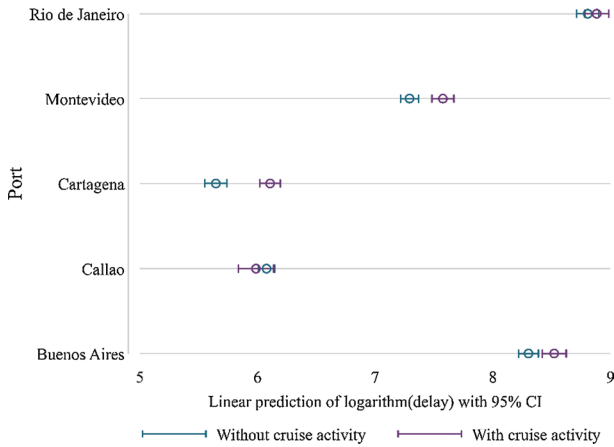*Significant at 10%, **significant at 5%, ***significant at 1%

Fig. 5 Marginal effects of ANCOVA

average, there is a 19% increase in congestion whenever there is at least one cruise ship berthed at port. The largest difference is found in Cartagena, where the delay increases by almost 50%. In Montevideo and Buenos Aires, the effect is 28% and 22%, respectively.

To complement these findings, the results from the econometric model to estimate the impact of cruise activity on congestion are presented below. As in the case of the marginal effect of the ANCOVA model, the logarithm of the total delay was used as the dependent variable. However, in contrast to the variable used in the ANCOVA model, *cruises* here take the value of the total number of ships berthed, and not just a dummy for the case when there is activity at the port. Table 3 presents the results of the model, estimated with the four different groups of independent variables explained in the *methods* section. Each column represents the results for each model, considering different exogenous variables. All the models reject the Hausman Test for the null hypothesis, suggesting appropriate estimation by assuming fixed effects instead of random effects. All the models contain fixed effects for month to control for seasonality.

The first model only includes the *afternoon* and *cruises* variables. As expected, according to the results presented in Table 3, congestion tends to be more prominent in the morning period, reflected in the negative coefficient of *afternoon*. In this model, each cruise increases congestion in the port area by 11%. However, results suggest that, while the *afternoon* and *cruises* variables significantly impact congestion, they alone do not explain much of the variability of congestion, as reflected in the small R^2 value. The second model considers *working day* as an explanatory variable. As shown in model (ii), *working day* increases congestion by 78%. Next, model (iii) considers *hazards* and *road closures*—both in logarithms—as regressors. According to this model, both *hazards* and *road closure* are significant in explaining congestion. In fact, an increase of 1% in the number of hazards increases congestion by 0.26%. In turn, a similar increase in the number of road closures raises congestion by 0.17%.

Table 3 Model results

| Variable | Models | | | |
	(i)	(ii)	(iii)	(iv)
Cruises	0.11	0.13	0.10	0.12
	(0.00)	(0.00)	(0.00)	(0.00)
Afternoon	−0.18	−0.18	−0.12	−0.15
	(0.00)	(0.00)	(0.00)	(0.00)
Working day		0.78		0.73
		(0.00)		(0.00)
Hazard			0.26	0.09
			(0.00)	(0.00)
Road closure			0.17	0.19
			(0.00)	(0.00)
Constant	7.26	6.70	6.75	6.48
	(0.00)	(0.00)	(0.00)	(0.00)
α	0.58	0.55	0.56	0.55
Observations	3926	3926	3926	3926
R^2	0.02	0.07	0.60	0.31

Values in parenthesis refer to p-value of the t-statistic

With an R^2 of 0.6, this is the model that better explains the variability of the total delay in the port area. Model (iv) introduces all the regressors. All the variables are statistically significant explaining the dynamics of congestion. Among them, *working day* has the largest impact on congestion, raising congestion by 73%. *Cruises*, the variable of interest in this article, is consistently positive and statistically significant in every estimated model, showing the robustness of the estimated coefficient. Overall, there is a 12% increase in congestion for every cruise vessel that either arrives or departs from the port. This number is smaller than the one obtained with the ANCOVA model. The difference is explained by the fact that results illustrated by Fig. 5 refer to an increase in congestion due to the existence of at least one cruise ship, with many timeframes including more than one ship. Instead, results presented in Table 3 refer to congestion growth caused by a single cruise ship. Finally, while there is no information available on the number of passengers traveling on each ship, we used gross tonnage and the dead weight tonnage of the cruise vessel in additional analyses as regressors. The estimated coefficients of both variables are consistently not significant and do not change the overall findings from model (iv). This result is likely related to the presence of multicollinearity between the number of cruises and the total gross tonnage of all cruises.

5 Discussion

While the negative externalities of the cruise industry have been much debated in public opinion, little academic research is available on the impact of cruise activity on congestion. The evidence presented in this paper for port cities in South America and the Caribbean indicates a statistically significant effect of cruise activity on urban congestion in the surrounding area of the port. As these two regions represent more than one-third of global cruise traffic (CLIA 2020), these findings can help design activities to balance the negative and positive effects of an industry that is a key source of economic revenue for host cities. According to industry data, a single call from an average ship in the South American and Caribbean regions generates approximately US$415,000 in passenger spending ashore (FCCA 2020).

Initiatives to accommodate both higher passenger volumes and larger cruise ships have focused on upgrading port infrastructure. New cruise terminals in Panama (Perico Island) and Ecuador (Manta) required investments of US$165 million and US$2.4 billion, respectively (Mendoza 2018; Varela 2019). Interestingly, most infrastructure improvements have targeted the easing of maritime operations at cruise terminals, whereas less improvement has been seen on smoothing passenger mobility in highly congested areas. This insight is particularly relevant for cruise activity in South America, where key cruise destinations (Rio de Janeiro, Buenos Aires, and Lima) are also among the 50 most congested cities in the world (TomTom 2020). Based on the results of this research, cruise terminal investments such as new docks or larger terminal buildings should be complemented with initiatives aimed to reduce the negative impacts that cruise tourism has on highly congested urban areas. Our analysis suggests that each vessel increases congestion by nearly 12% in the area close to port terminals. In certain cases, road infrastructure investments may be needed to accommodate larger volumes of traffic around the port area. In the case of Buenos Aires, an exclusive highway for heavy traffic and long-distance buses was inaugurated in 2019, with the aim of easing congestion in a neuralgic area of the city that attracts business offices, the port—both cargo and passenger terminals—and the long-distance bus passenger terminal of Retiro. By separating light and heavy traffic, speed was significantly improved in the area (Calatayud et al. 2021). The government is currently evaluating allowing tourist buses to use the highway during cruise peak season to further reduce urban congestion. Indeed, our results suggest that congestion increases by 30% around the port area whenever there is cruise activity. Other measures may include rerouting traffic in streets adjacent to the terminal gate during periods of passenger boarding and offboarding, designating specific lanes for cruise-related traffic at certain times of the day, and coordinating with port authorities to avoid that the peak of the disembarking process takes place during traffic peak hours.

In particular, the detailed level of analysis provided by this research, enabled by the innovative use of big data, shall hopefully allow policymakers to identify key areas in the road network that need traffic flow improvement and design appropriate and dynamic measures to take advantage of cruise tourism while
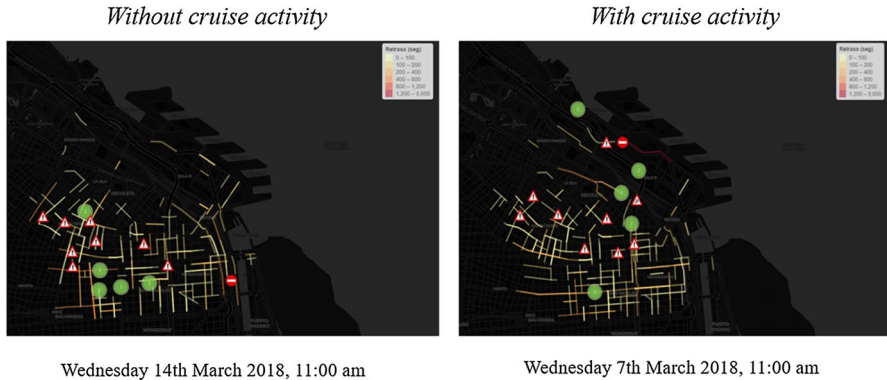
Without cruise activity   *With cruise activity*

Wednesday 14th March 2018, 11:00 am   Wednesday 7th March 2018, 11:00 am

Fig. 6 Congestion level and cruise activity in Buenos Aires. *Source* Own elaboration using Marine Traffic and Waze data

mitigating its effects on urban mobility. For example, Fig. 6 shows congestion levels in Buenos Aires for the same time of day on a weekday (Wednesday at 11:00 AM) with and without cruise activity. Congestion levels are higher not only on the streets next to the cruise terminal, but also along the main roads that connect to the main touristic attractions of Buenos Aires. Therefore, besides enhancing cruise terminal accessibility with infrastructure and traffic management measures, transport planners should also pay attention to easing the impact of cruise activities on roads that receive high levels of cruise-related traffic even if the roads are not located next to the terminal gate. Improving public transportation services that connect the cruise terminal with tourist attractions in the city is key to reducing the number of vehicles transporting cruise passengers, which add to the already high congestion levels in the area.

The adoption of appropriate measures to reduce traffic congestion around cruise terminals can have a twofold benefit for port cities. First, it can help reduce a source of stress for its inhabitants (Jordan et al. 2017). By addressing one of the main concerns that local communities have regarding the negative impact of cruise activity, these measures can in turn stimulate local support. Secondly, easing cruise-related congestion can increase the attractiveness of a port city as a cruise destination by enabling smoother, faster connectivity between the cruise terminal and tourist attractions.

6 Conclusions

In this paper, big data along with statistical and econometric models are used to uncover the impact of cruise activity on urban congestion. Specifically, the model analysis found that each vessel increases traffic congestion by nearly 12% in the areas close to port terminals. Our research helps advance knowledge in two specific areas. The first area demonstrates how to leverage big data and econometric techniques to understand the behavior of traffic flows related to cruise activity. This

level of accuracy was previously not available, with only the use of surveys or origin–destination matrices, since these can only collect perceptions or data not spatially partitionable. The second area developed here is filling the void of research in exploring the effect of cruise activity on urban congestion. Overall findings suggest a positive impact. The nonsignificant effect in the case of Callao might be explained by the location of the port, far from downtown Lima and urban activities. The evidence provided in this paper can support policymakers in the design of customized interventions to ease congestion related to cruise activities, thus mitigating the negative externalities of such activities for their cities, while leveraging the industry's benefits. Areas for further research include the application of our model to port cities in other regions, to compare the impact of cruise activity. The model proposed here can be further applied to understand the relationship between congestion and other maritime-related activities, such as container movements and congestion at port gates.

Annex

Port area coordinates

Buenos Aires		Callao		Cartagena		Montevideo		Rio de Janeiro	
Latitude	Longitude	Latitude	Longitude	Latitude	Longitude	Latitude	Longitude	Latitude	Longitude
−34.5913	−58.3498	−12.0446	−77.1416	10.41331	−75.5452	−34.9061	−56.2294	−43.1676	−22.9237
−34.5991	−58.3741	−12.0448	−77.1246	10.41863	−75.5327	−34.9162	−56.2176	−43.1955	−22.919
−34.5972	−58.4011	−12.0707	−77.139	10.40884	−75.5207	−34.9007	−56.1796	−43.2103	−22.8979
−34.5763	−58.4001	−12.0812	−77.1676	10.40175	−75.5201	−34.8879	−56.1866	−43.1922	−22.8904
−34.5697	−58.3933	−12.0583	−77.183	10.40099	−75.525	−34.8911	−56.2084	−43.168	−22.895
−34.5661	−58.3663			10.40327	−75.545	−34.8985	−56.2096	−43.1639	−22.9008
								−43.1623	−22.9074
								−43.1615	−22.9166

References

Artal-Tur, Andrés, Antonio Juan Briones-Peñalver, and Marina Villena-Navarro. 2018. Tourism, cultural activities and sustainability in the Spanish Mediterranean regions: A probit approach. *Tourism & Management Studies* 14: 7–18.

Baker, David. 2015. Exploring cruise passengers' demographics, experience, and satisfaction with cruising the Western Caribbean. *International Journal of Tourism & Hospitality Reviews* 1: 23–31.

Brida, Juan Gabriel, Giacomo Del Chiappa, Marta Meleddu, and Manuela Pulina. 2012. The perceptions of an island community towards cruise tourism: A factor analysis. *Tourism an International Interdisciplinary Journal* 60: 29–42.

Brida, Juan Gabriel, and Sandra Zapata. 2010. Cruise tourism: Economic, socio-cultural and environmental impacts. *International Journal of Leisure and Tourism Marketing* 1: 205–226.

Carić, Hrvoje, and Peter Mackelworth. 2014. Cruise tourism environmental impacts—The perspective from the Adriatic Sea. *Ocean & Coastal Management* 102: 350–363.

Coffey, Helen. 2018. Dubrovnik to cap the number of cruise ships allowed to dock each day. *Independent*. October 2.

Cruise Lines International Association (CLIA). 2021. State of The Cruise Industry Outlook 2021. https://cruising.org/en/news-and-research/press-room/2020/december/clia-releases-2021-state-of-the-cruise-industry-outlook-report. Accessed 13 May 2021.

Cruise Lines International Association. 2018. *The contribution of the international cruise industry to the global economy in 2018.* https://cruising.org/en/news-and-research/research/2019/december/the-contribution-of-the-international-cruise-industry-to-the-global-economy-in-2018. Accessed 29 July 2020.

Chen, Jaime M., James F. Petrick, Alexis Papathanassis, and Xinjian Li. 2019. A meta-analysis of the direct economic impacts of cruise tourism on port communities. *Tourism Management Perspectives* 31: 209–218.

Calatayud, Agustina, Santiago Sánchez González, Felipe Bedoya-Maya, Francisca Giraldez, and José María Márquez. 2021. *Congestion urbana en América Latina y el Caribe: características, costos y mitigación.* IDB-MG-902

Daganzo, Carlos. 1997. *Fundamentals of transport and traffic operations* (30th ed.).

Fabregat, Alexandre, Lluís Vázquez, and Anton Vernet. 2021. Using Machine Learning to estimate the impact of ports and cruise ship traffic on urban air quality: The case of Barcelona. *Environmental Modelling & Software* 139: 104995.

Feng, Mingxiang, Shih-Lung. Shaw, Guojun Peng, and Zhixiang Fang. 2020. Time efficiency assessment of ship movements in maritime ports: A case study of two ports based on AIS data. *Journal of Transport Geography* 86: 102741.

Florida-Caribbean Cruise Association. 2018. Economic contribution of cruise tourism to the destination economies. https://www.f-cca.com/downloads/Caribbean-Cruise-Analysis-2018-Vol-I.pdf. Accessed 29 July 2020.

Goodwin, Phil. 2004. The economic costs of road traffic congestion. *University College London*.

International Transport Forum. 2016. Cruise shipping and urban development: The case of Venice. https://www.itf-oecd.org/sites/default/files/docs/cruise-shipping-urban-development-venice.pdf. Accessed 29 July 2020.

Jordan, Evan J., and Christine A. Vogt. 2017. Residents' perceptions of stress related to cruise tourism development. *Tourism Planning & Development* 14: 527–547.

Jordan, Evan J., José Cabral Vieira, Carlos M. Santos, and Ting-Yen Huang. 2020. Do residents differentiate between the impacts of tourism, cruise tourism, and Airbnb tourism? *Journal of Sustainable Tourism* 1–19.

MacNeill, Timothy, and David Wozniak. 2018. The economic, social, and environmental impacts of cruise tourism. *Tourism Management* 66: 387–404.

Mendoza, Mónica. 2018. Manta inauguró la primera terminal de cruceros con inversión chilena. *El Comercio*. October 19.

Navarro, Valverde, Francisco Antonio, and Alberto Capote Lama. 2021. ¿Overtourism en la ciudad de Granada?: Una aproximación a la percepción de turistas, residentes y partidos políticos locales. *Cuadernos Geográficos De La Universidad De Granada* 60: 35–53.

Newell, Gordon F. 1993. A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological* 27: 289–303.

Perdiguero, Jordi, and Alex Sanz. 2020. Cruise activity and pollution: The case of Barcelona. *Transportation Research Part D: Transport and Environment* 78: 102181.

Polat, Naci. 2015. Technical innovations in cruise tourism and results of sustainability. *Procedia-Social and Behavioral Sciences* 195: 438–445.

Radic, Aleksandar, Michael Lück, Amr Al-Ansi, Bee-Lia. Chua, Sabrina Seeler, António Raposo, Jinkyung Jenny Kim, and Heesup Han. 2021. To dine, or not to dine on a cruise ship in the time of the COVID-19 pandemic: The tripartite approach towards an understanding of behavioral intentions among female passengers. *Sustainability* 13: 2516.

Robinson, Danielle, Steven P. Newman, and Selina M. Stead. 2019. Community perceptions link environmental decline to reduced support for tourism development in small island states: A case study in the Turks and Caicos Islands. *Marine Policy* 108: 103671.

Ros, Chaos, David Pino Sergi, Sergi Saurí Roca, and Marchán, and Agustín Sánchez-Arcilla Conejo. 2018. Cruise passenger impacts on mobility within a port area: Case of the Port of Barcelona. *International Journal of Tourism Research* 20: 147–157.

Rosa-Jiménez, Carlos, Beatriz Perea-Medina, María J. Andrade, and Nuria Nebot. 2018. An examination of the territorial imbalance of the cruising activity in the main Mediterranean port destinations: Effects on sustainable transport. *Journal of Transport Geography* 68: 94–101.

Ruiz-Guerra, Ignacio, Valentín Molina-Moreno, Francisco J. Cortés-García, and Pedro Núñez-Cacho. 2019. Prediction of the impact on air quality of the cities receiving cruise tourism: The case of the Port of Barcelona. *Heliyon* 5: 01280.

Santos, Maria, Elena Radicchi, and Patrizia Zagnoli. 2019. Port's role as a determinant of cruise destination socio-economic sustainability. *Sustainability* 11: 4542.

Stefanidaki, Evangelia, and Maria Lekakou. 2014. Cruise carrying capacity: A conceptual approach. *Research in Transportation Business & Management* 13: 43–52.

Sun, Xiadong, Xuegang Feng, and Dinesh K. Gauri. 2014. The cruise industry in China: Efforts, progress and challenges. *International Journal of Hospitality Management* 42: 71–84.

Tirachini, Alejandro, and Andrés Gomez-Lobo. 2020. Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. *International Journal of Sustainable Transportation* 14: 187–204.

TomTom Traffic Index. 2019. https://www.tomtom.com/blog/road-traffic/urban-traffic-congestion/. Accessed 19 July 2020.

Varela, Gestión. 2019. Nueva terminal de cruceros lleva un 60% de avance en la construcción. *La Prensa*. May 14.

Vicente-Cera, Isaías, Asunción Acevedo-Merino, Juan Antonio López-Ramírez, and Enrique Nebot. 2020. Use of AIS data for the environmental characterization of world cruise ship traffic. *International Journal of Sustainable Transportation* 14: 465–474.

Waze for Cities. 2019. https://www.waze.com/ccp. Accessed 29 July 2020.

Xu, Yanyan, and Marta C. González. 2017. Collective benefits in traffic during mega events via the use of information technologies. *Journal of the Royal Society Interface* 14: 20161041.

Yperman, Isaak, Steven Logghe, and Ben Immers. 2005. The link transmission model: An efficient implementation of the kinematic wave theory in traffic networks. In *Proceedings of the 10th EWGT meeting*, 122–127.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.