sustainability

MDPI

Article

Validation Method for a Multimodal Freight Transport Model Exploiting Floating Car Data

Dario Ballarano *, Marco Petrelli and Alessandra Renna

Department of Engineering, Roma Tre University, 00146 Rome, Italy; marco.petrelli@uniroma3.it (M.P.); alessandra.renna@uniroma3.it (A.R.)
* Correspondence: dario.ballarano@uniroma3.it

Abstract: The implementation of valid freight transport simulation models requires an extensive and detailed validation phase for understanding the feasibility of the outputs and the capacity of the structure of the proposed models in representing the real-world data. Traditional methods involve the use of surveys in order to describe the behaviour of stakeholders and to gather some aspects of the modal choices. Recent studies integrate this approach with Big Data as Floating Car Data to obtain better statistical information of the details at different levels. The current research involves the unexplored field of the validation of freight transport simulation models using a data-driven approach based on a large database of over 292 million Floating Car Data (FCD) signals generated by 29,298 commercial vehicles during the month of October 2019. The paper proposes an FCD processing methodology to identify freight vehicles using Ro-Ro/Ro-Pax services, and presents the results of an in-depth tracking analysis for combined transport and road transport. The validation phase permits the evaluation of the simulation tool results with real choices of heavy vehicles, referring also to the statistical information on travel times and the achievement of additional information through an in-depth analysis of tracking single vehicles.

Keywords: big data; floating car data; validation; freight transport model; multimodal; tracking; data driven

check for updates

Citation: Ballarano, D.; Petrelli, M.; Renna, A. Validation Method for a Multimodal Freight Transport Model Exploiting Floating Car Data. *Sustainability* **2022**, *14*, 5540. https://doi.org/10.3390/su14095540

Academic Editors: Marija Burinskienė and Rasa Ušpalytė-Vitkuniene

Received: 24 February 2022
Accepted: 27 April 2022
Published: 5 May 2022

1. Introduction

The modelling complexity of freight transport consists of several factors that influence the travel choices, the needs of the many companies involved, and the different points of view regarding the definition of the model (from the point of view of the goods or the vehicle, and therefore of the hauliers). Additionally, freight transport consists of several phases (e.g., loading and unloading phases, break and rest periods, and travel times), which are hard to clearly define, especially due to the presence of rest regulations and the different typologies of transport carried out (for instance, accompanied or not accompanied). To deal with such issues, the implementation of valid simulation models requires an extensive and detailed validation phase supported by the capabilities of Big Data, such as Floating Car Data (FCD), which have become exponentially important and largely utilised in recent years. FCD are signals generated by On Board Units (OBUs) usually built on vehicles with a mass greater than 3.5 tons, or on private or commercial vehicles in exchange for insurance price discounts. The OBU sends the recording to a data-processing centre, via a mobile network, at a fixed sampling rate, or variable based on speed, or on other variables that characterise the motion, such as acceleration, steering curvature, and the state of the vehicle's ignition as either on or off. The potentiality of FCD increases together with the use of OBUs, with implications not only for user information, security, and functionality, but also for planning. According to [1], the use of a geo-localised signal can be integrated into traditional approaches for the construction of transport models; in fact, FCD are commonly used to observe historical mobility patterns and, when integrated with traditional approaches, it allows to simplify, and make more effective, the construction and calibration of a transport model. Although the literature provides numerous studies mainly focusing on the analysis of geo-localised signals generated by vehicles in the field of private and public transport [2–4], there are only a few studies that address the possible applications of these signals for performance evaluations and the monitoring of freight transport. The structure of the FCD allows for investigations in relation to time and space, the single displacement of each vehicle, which, with the low cost of this type of information, are its strength [5]. Numerous studies have been carried out for light vehicles: the zoning phase and the selection of the infrastructures of the graph [1], and the monitoring of both runoffs along infrastructures presenting a greater sensitivity to traffic anomalies for accidental events and for the estimation of travel times, which can be defined deterministically or stochastically, based on the history of the speeds held by the probe vehicles [6]. Recently, due to sustainability and cost challenges, many studies are emerging to evaluate, model, and simulate freight interventions. For example, [7] aims to optimise delivery routes in urban areas by exploiting FCD to determine the travel times on the network at different times of the day and, in addition, to use them to identify the optimal route according to heuristic procedures. Other applications rely on the signal produced by on-board detectors to assess the performance of the tracking system for freight transport units moving along intermodal transport chains, specifically from the access port to the egress port. According to [8], it has been found necessary to monitor the cargo units during the navigation and to evaluate the logistics performance offered, recording the boarding time from the port operators and a wireless infrastructure for data exchange and transmission, in order to gain an insight into intermodal chains. In fact, the tracking system is particularly efficient on the road side, while, on the sea-side, the signals are of poor quality or totally absent because they are shielded by the thick walls of the vessels. However, the processing of FCD concerning heavy vehicles is a very complex operation. In fact, few studies refer to the definition of freight vehicle trips. As mentioned at the beginning of this paragraph, the difficulty of data management lies in the combination and distinction of the different phases of the trip and in the multiple points of view that might be adopted. Then, the poor use of FCD may be attributed to either the difficulty of arranging, manipulating, and managing such information due to the conceptual or operational issues, or the lack of interest of the various operators in sharing data and information. The definition of trips represents the main challenge, since the subdivision of the sequence of GPS points in a sequence of trips requires specific criteria [9]. No single methodology can be found in the literature, as this is strongly related to the amount and type of information in the database. [10] detects the areas in which trucks stop considering the concentration of records, and link these with the surveys provided by the drivers. Otherwise, it is possible to observe the use of purely temporal criteria [11] without foundation [12], or the combination of these with spatial criteria, matching the origin and destination of the trips to points of interest [13].

Usually, the traditional approach for validation is well-established in the literature and is based on the use of surveys [14]. Ref. [15] performs interviews with several trucking companies operating in Italy, in order to verify whether the actual behaviour of trucking companies agrees with their model and also to ascertain the main aspects with which trucking companies are concerned in relation to their modal choice. Unlike these forms of research, the current study proposes an innovative validation method for a freight transport simulation model in a multimodal context [16]. This model supports the decisions of transportation planners by assessing the impacts of policies in promoting more sustainable solutions for freight transport. This model simulates the mode choice between sea-road and road-only transport, focusing on the transport operators' points of view. The simulation tool provides two set of solutions, distinguishing between travel-time (T*) and travel-cost (C*) minimisations. Then, the model returns the potential demand in terms of Origin–Destination (OD) pairs identifying paths with regard to travel-time and travel-cost solutions for each mode of transportation. The model refers to all heavy vehicles moving different commodities in the Italian context. The evolution of the validation phase with the adoption of Big Data for statistical information consists also in the development of in-depth analyses; FCD might constitute a valid element to validate the results obtained, refine the definition of the model, or possibly entirely modify the approach to the problem studied. By comparing the model results with the information obtained from the FCD signals produced by the vehicles, it can be determined that a real validation problem is faced when returning the structured data analysis methodology and the main results. Indeed, this study intends to demonstrate the possibilities of both quantitatively extrapolating key information in the complex field of combined sea-road transport, such as the sailing times, service schedules, and port dwell times, and obtaining detailed information through an in-depth analysis by tracking each vehicle during travel by each mode of transportation, such as the origin and destination of the trips, and the travel and rest times. The validation phase highlights the importance to consider the chain trips in freight simulation model. For example, the hauliers can start the travel with different hours of driving influencing the modal choice.

2. Validation Phase Using a Data-Driven Approach

The validation phase represents a key point to understand the capacity of the model of representing the real behaviour of all the represented stakeholders and elements. Therefore, the current research aims to demonstrate the possibility of quantitatively extrapolating the relevant information in the complex field of freight transport from an FCD dataset by determining the routes used by freight vehicles, the level of performance offered by access and egress ports, in terms of port dwell time and time of service. This information is used to compare the outputs of the model, in terms of travel times and choice of specific Ro-Ro-Ro-Pax service, but also to investigate the variables influencing the mode choice model. The present analysis proposes a comparison between the road-only alternative and the sea-road transport for the same ODs, from the point of view of the goods referring to the accompanied transport, tracking the vehicle's trip from origin to destination and carrying out an in-depth analysis. This study is based on the high number of detailed information, which can be obtained with in-depth analysis from a large sample of vehicles. This approach allows for the identification of the different behaviour of the hauliers, and therefore the choice variables, highlighting the possible criticalities of both transport modes. The paper contains a description of the dataset (FCD) and the problematic schematisation for the development of the methodology to correctly manage FCD, with regard to sea-road transport and specifically Ro-Ro services. Finally, the main results of the validation phase are reported, distinguishing a comparison between the model and the FCD trips and some additional conceptual information for the model setting.

2.1. Dataset Description

The survey was carried out on a database provided by VEM Solution S.P.A., which has more than 292 million records generated by 29,298 commercial vehicles registered in Italy and mainly circulating in Italy and Europe during the month of October 2019. Each record is composed of several fields, such as the unique numeric identification id associated with the OBUs; the position expressed in terms of geographical coordinates (latitude and longitude), according to the WGS84 coordinate system; and the instant time of recording the data expressed, according to the Italian time zone. Additionally, the records present information relating to the instantaneous speed (km/h), the direction of travel (degrees), the reliability of the GPS data (Horizontal Dilution of Precision), and the engine status distinguished as engine on, engine off, and switch off. The complexity of the management and use of these data, due to the large quantity of information associated with the signal, required a preliminary analysis of the spatial data, evaluating the variability of each field. Therefore, a preliminary analysisprovided information about the consistency ofthe spatial and temporal distributions of the dataset. Subsequent analysis was performed to identify, correct, or exclude critical data, such as duplicate records, localisation errors, values expressed according to different units of measurement, and fields with null or erroneous values, such as the temporal progress associated with a journey in which there was a reduction in or absence of variations in the kilometres travelled. In order to determine the set of heavy vehicles, after a filtering phase based on the vehicle's brand and model, a study of the sampling rate was conducted for the selected 14,222 freight vehicles. The way in which the records are produced greatly varies, depending on the technology fitted to the vehicle, gsm signal quality, satellite coverage, and the behaviour of the driver. Generally, devices are observed that emit signals according to a sampling time dependent on the speed, direction, and respective variations, and on the switching on and off of the engine and vehicle. Therefore, it was necessary to study the different sampling times of the vehicles, observing the average value, the standard deviation, and evaluating the need to adopt lower or higher threshold values to process homogeneous data. Specifically, the use of a higher threshold makes it possible to impose the level of accuracy for the operation, while the use of a lower threshold makes it possible to reduce the positioning errors at the boundaries of the port geometries. By selecting a sample of vehicles with average sampling times of more than 30 s and less than 3 min, it can be observed that the GPS tracks produced by these vehicles are sufficiently detailed to carry out analyses on the movements made on the land side. On the other hand, sea-side displacements are characterised by vehicles that are usually switched off on board vessels, which, for this reason, interrupt the transmission of information during the use of the Ro-Ro/Ro-Pax service and, at the end of the sea-side displacement, generate records for which the positioning may not be updated. The thick walls of the vessels prevent the positioning from the gsm network, and it does not prevent the generation of signals that contain incorrect information. Another problem with the data relates to the alteration of on-board instruments, operated by drivers to make up for delays and circumvent the rest of the regulations to which they are subject. In this case, the vehicle is able to disguise the journey as a rest by generating a fictitious record of the vehicle as being stationed at a service station. During the counterfeit period, it is impossible to trace the vehicle, so that only the previous and subsequent information is known.

2.2. Approach and FCD Processing Methodology

The methodology developed to analyse FCD investigating sea-road and road-only transport required a problem schematisation to manage the signals, and associated it with the different phases of transport (Figure 1). It was also useful to directly compare the model results with those identified via FCD. For road-only, the trip was characterised by the origin, destination, distance, and duration. For the combined sea-road transport, the following phases were distinguished:

- Access phase: The phase starts from the origin of the goods' journey to the departure of the Ro/Ro-Pax service. This phase is characterised by a journey on the road, whereby the full information of the trip is available, specifying the break and rest periods on the road trip or in the parking areas of the port, and the duration of the operational activities (loading/unloading operations and ticket control). Therefore, the relevant information is the origin of the voyage, the time and distance travelled by road, the rest periods, and the time spent in the port.
- Navigation phase: This phase occurs on board of the vessel between two ports. During the journey, trucks usually do not send signals because they are switched off or on the open sea. For this phase, the relevant information is the time of departure and the end of the Ro-Ro/Ro-Pax service.
- Egress phase: The phase starts from the time of the arrival of the vessel at the berth when the trucks lock on the signal, to the final destination. For this phase, the relevant information is the time spent in the egress port, the time and distance travelled by road, the rest periods, and the destination of the trip.

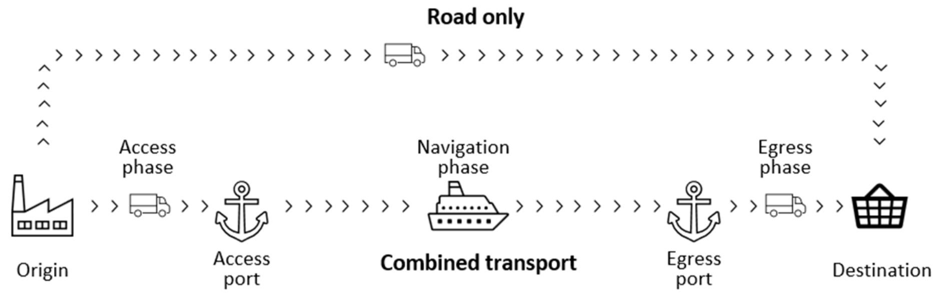

Figure 1. Problem schematisation.

The proposed methodology (Figure 2) is made up of three phases using the information from the previous analyses to refine the subsequent ones. The first phase allows for the identification of the users of Ro-Ro/Ro-Pax services as all those vehicles produce records in one port area and then consecutively in another by sea. The second phase consists of the analysis of vehicles using Ro-Ro/Ro-Pax services, which is necessary to correct the port geometries and to identify recurring problems of the data in the case of a loss of signal. In the third phase, the vehicles are tracked to define the origin and destination of the trip from the point of view of the goods in the combined transport mode. The variables contributing to the modal choice were investigated through a comparison of the trips with the same origin and destination, but which used different transport modes.

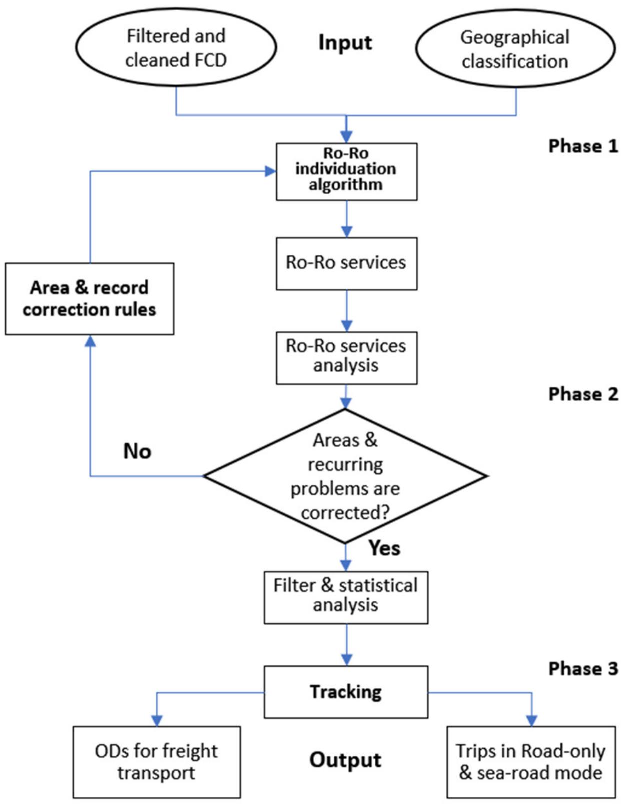

Figure 2. Methodology for data-driven validation.

The **first phase** consists of three activities:

(1) Pre-processing the data to determine a dataset of FCD relating to heavy vehicles. Therefore, the procedure includes a preliminary analysis of the sampling times, the ranges of variability of the fields, and the geographical distribution of the signals. Based on the analysis results, the study area and the filters to be applied are deter-

mined in order to obtain a dataset of FCD produced by vehicles with homogeneous behaviour.

(2) The classification of the territories by geometries in which the study area is determined; the relevant ports and a coastal area cordon are defined, for which the services used should not be considered. Subsequently, the definition of the maritime and port geometries occurs (Figure 3), which is of fundamental importance due to the different behaviours and degrees of accuracy of the OBUs of the vehicles. The land-side port geometry follows the port boundaries, while the sea-side geometry has to consider the size of the ferries.

Figure 3. Port area geometry with berth offset.

(3) Preliminary identification of the trips for the combined transport through the identification algorithm, which is constrained by the port and maritime geometries. This algorithm provides two databases:

- The database of FCD relating to vehicles using combined transport (a subset of the analysed sample).
- The database of trips in combined transport for the Ro-Ro/Ro-Pax services. It is characterised by specific information: the vehicle identifier; the instant of entry into the port area of access and egress, equal to the instant of generation of the first record within the port geometry; the instant of exit from the port area of access and egress, equal to the instant of generation of the last record within the port geometry; the service used, equal to the name of the port of access followed by the name of the port of egress; and the identifiers of the FCD records associated with the instants of entry into and exit from the ports.

The geographical identification criterion on which the algorithm is based, consists of a chronological scanning of the records for which the vehicle moving from one port area to another is recorded considering the definition of the geometries and areas in activity 2.

The **second phase** consists of the analysis of vehicles using Ro-Ro/Ro-Pax services to correct the port geometries and to identify recurring problems. Before conducting the statistical analysis, it is necessary to filter this result from the outliers, leading to a clear picture on a national scale of the use and performance of maritime services for this sample of vehicles.

When a vehicle is within the port geometry, it may encounter shielding elements, such as a subpass or the thick walls of vessels. These elements might cause the OBU GPS to generate incorrect location records, resulting in the generation of inaccurate results. By adapting the port geometries to the data sample and implementing record correction rules based on the recurring errors in the algorithm, the algorithm is run to identify the combined transport movements for each vehicle, computing the navigation time for each maritime service and also the time spent at port. Therefore, it is possible to calculate the dwell times at the ports of access and egress, defined as the time between the first and the last record generated by the vehicle within the port geometry and the navigation time equal to the interval between the last signal issued in the port of access and the first issued in the port of egress. Subsequently, the statistical analysis and filtering of the results occurs in order to characterise each port on the basis of the performance offered.

The **third phase** consists of tracking vehicles in order to identify the origin and destination of the trips. The previous phase provides the subset of services used for each vehicle. The tracking of the single vehicles using the specific service provides all the origins and destinations of the combined trips during the month of survey. However, the tracking follows a temporal order, and therefore the activities of the vehicle are considered between one service and another, which may coincide or be different. This activity makes it possible to chain the movements of the same vehicle, highlighting the uses of different services and possibly different modes of transport between the same origin and destination, finding relationships between one trip and another.

3. Application of the Methodology for the Validation of a Model in the Italian Case Study

Due to the heterogeneity of the record production methods, a subset of vehicles with homogeneous sampling was considered. The sample was identified on the basis of an analysis of the average sampling times varying between 2 s and 818 s and the standard deviation between 0.5 s and 671 s, considering the total number of records produced. The filter reduces the sample to 7529 vehicles, meeting the constraint of having generated at least 4000 signals in a month with average sampling times between 30 s and 180 s. The application was carried out on a sub-sample of Italian ports, considered also in the maritime network of the simulation model, formed by the 26 ports where Ro-Ro traffic is expected: Ancona, Arbatax, Bari, Brindisi, Cagliari, Catania, Civitavecchia, Genoa, Gioia Tauro, La Spezia, Livorno, Messina, Monfalcone, Naples, Olbia, Palermo, Piombino, Porto Torres, Ravenna, Salerno, Savona, Taranto, Termini Imerese, Trieste, Venice, and Villa San Giovanni. Moreover, the following elements were defined (Figure 4): the port geometries of the berths, considering the maximum size of the ferries circulating in Italy; the geometries relating to the Italian counties; and the cordon geometry (which replaces the land-side geometry outside the Italian state) corresponding to a coastal strip covering the entire Mediterranean Sea.

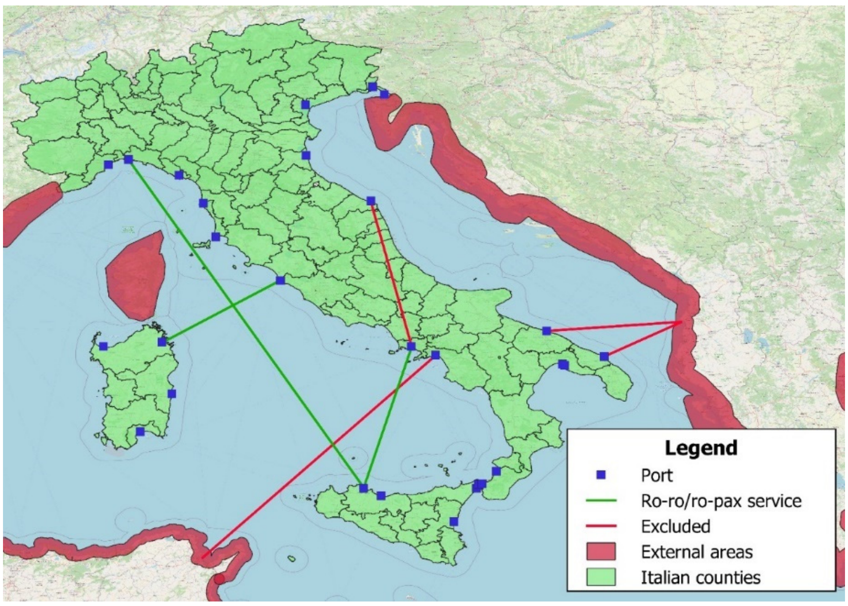

Figure 4. Identifying scheme of the use of Ro-Ro and Ro-Pax services.

Specifically, for October 2019, the execution of the algorithm identified that the Ro-Ro/Ro-Pax services were used 1201 times. The Villa San Giovanni—Messina connection was the most used, with a total of 944 trips, highlighting the propensity of Italian freight transport to use the road-only mode. The second phase represents an iterative phase of the correction of geometries, areas, and signals identified in the first phase. It consists of the analysis of records prior and subsequent to the use of maritime services, showing the following localisation problems:

- Two sets of records internal to a port geometry interspersed with one or more records located geographically in Italy, in the cordon or in the unclassified area.
- Two sets of records, respectively, internal to the geometries of the access and egress ports, interspersed with one or more records located geographically in Italy, in the cordon or in the unclassified area. Usually, the record in the egress port is incorrectly recorded in the access port, due to the failure to update the GPS position of the vehicle.
- Several series of pings with oscillating locations between several port geometries make the information unreliable, occurring mainly for obsolete vehicles.
- A series of records characterised by constant sampling, located within a port geometry where the last record of the vehicle is in another port geometry. This case is attributable to the falsification of the information recorded by the OBU, in order to deceive the regulations of rest periods for hauliers.

These problems were solved by implementing detection rules to relocate the records to the respective ports where they were generated. The relocation was carried out from the identified services by evaluating the ratio between the distance of the signal locations and the time between them, estimating the average speed of movement. This was calculated starting from the port of egress backwards in time and, in the case it turns out to exceed the threshold value of 30/25 kn equal to or greater than the maximum speed declared by the manufacturers of the vessels, the algorithm reallocates the position of the record to the port of egress. In the case of incorrect positioning during the stay in the port, caused, for example, by the presence of shielding surfaces, the average speed of movement is evaluated considering a threshold of 100 Km/h, corresponding to the maximum speed allowed for heavy vehicles in Italy. The average speed is evaluated from the records defining the service, chronologically for the port of entry and anti-chronologically for the port of entry, and relocation occurs as follows:

- For the port of access, the position of the record is updated to the position of the next record, if the vehicle was previously at the port; otherwise, the record is deleted.
- For the port of egress, the position of the record is updated to that of the previous one, if the vehicle is subsequently in port; otherwise, the record is deleted.

The execution of the algorithm has shown a clear improvement in the quality of the outputs where, however, 24.4% is still unreliable. Therefore, additional filters were introduced based on a sensitivity analysis, which led to the exclusion of trips characterised by a single record in the port geometry of access or egress, imposing the strict positivity of the dwell time and excluding trips having an absurd service time less than 0.33 h and those with a duration of more than 16 h (except for the Genoa–Palermo service, which is about 20 h). A total of 907 movements by Ro-Ro/Ro-Pax services were identified, as shown in Table 1. The entity of the movements (about 20%) underlines a real use of the combined sea-road mode for accompanied transport, mainly involving the Genoa–Palermo, Naples–Palermo, Salerno–Messina, and Catania–Salerno routes. The Villa San Giovanni–Messina route, representing a natural road extension for the high frequency of services and the high performance offered, shows a predominant road-only mode of transport. The analysis was restricted to trips between Sicily and the Italian mainland, in order to make a comparison between road-only and combined sea-road transport considering the most used maritime services.

Table 1. Frequency of usage of maritime services derived by FCD.

Ro-Ro Services	N.
Villa San Giovanni-Messina	389
Messina-Villa San Giovsanni	351
Napoli-Palermo	37
Messina-Salerno	32
Palermo-Napoli	28
Genova-Palermo	18
Palermo-Genova	13
Salerno-Messina	11
Catania-Salerno	5
Civitavecchia-Olbia	4
Salerno-Catania	4
Civitavecchia-Palermo	3
Livorno-Olbia	3
Olbia-Civitavecchia	2
Olbia-Livorno	2
Salerno-Palermo	2
Cagliari-Palermo	1
Civitavecchia-Cagliari	1
Civitavecchia-Arbatax	1

3.1. Validation Results

Tracking provided the following information for each trip: the origin, the destination, the maritime service used, the departure and arrival times, the time spent in ports, the distance and duration of the trip, as well as the driving and resting time sequences. This was conducted on 122 trips by sea-road transport and 109 trips by road-only mode. The size of the sample depended on the complexity of analysing freight transport. In fact, reconstructing the trips of heavy vehicles is much more complex than it is for light ones. The in-depth analyses were used to evaluate the modal choices between specific ODs, and these provided two types of information:

(1) Quantitative and qualitative outputs to be compared to the model outputs for the validation phase.
(2) Conceptual information, both to support the modelling approach and to integrate some aspects into the freight transportation model.

3.2. A Comparison between the Model and FCD Trips

The comparison between the outputs of the model and FCD trips also permitted establishment of the representativeness of the simulation model in its setting. The validation might be conducted on the path choice and on the main components of the travel time, such as navigation, and access and egress times considering specific ODs.

3.2.1. The Road Network Used by Heavy Vehicles

Through a spatial analysis, the macro information about the most used road network by road freight transport was obtained, which also provided the origins, destinations, stops, and congestion areas. The outputs of the simulation model, with regard to the shortest path for road-only transport, show the same results of the distribution of the sample of FCD trips (Figure 5). In fact, in both cases, the most used infrastructures belong to the motorway network. The thickness of the bars indicates the degree of usage of the road network as the intensity of the colour. The dark blue links are characterised by the major density of the records produced by heavy vehicles. Therefore, both in terms of the paths and degree of utilisation, there is a correlation between the outputs of the model and the paths chosen in the real world.

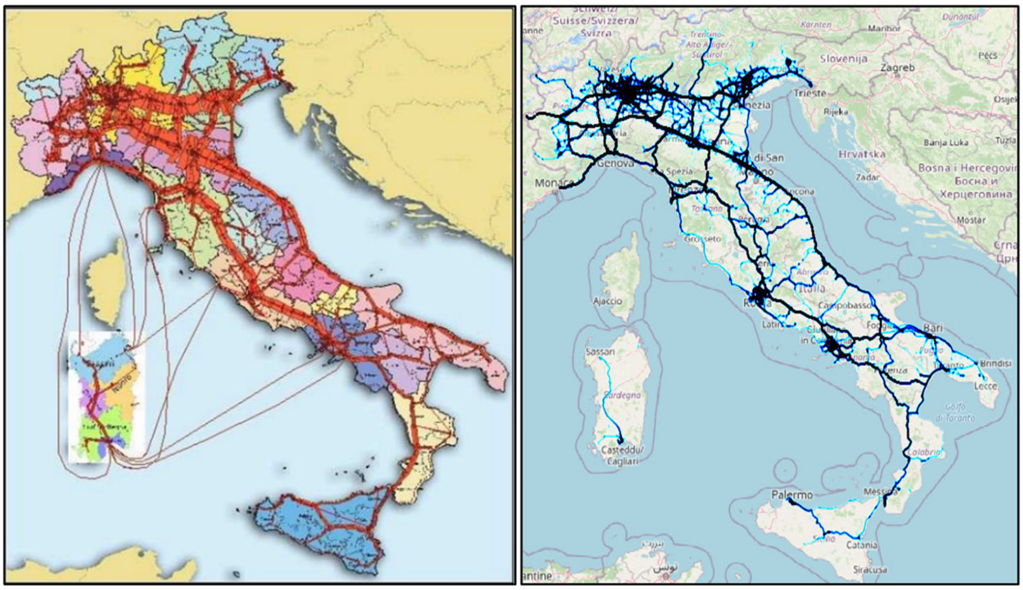

Figure 5. Comparison between the outputs of the simulation model and the distribution of FCD on the road network.

3.2.2. Path Choice for the Access and Egress Phases

Distinguishing the sample for each maritime service (Figure 6) shows both the prevailing use of motorways in the access and egress phases and the validity of the approach of the model. There is an effective representation of the total road distances. Many cases highlight the choice of longer road distances for the access and egress phases, in order to minimise the travel time. This consideration is evident for the origin or destination in proximity of more ports where more possibilities exist. Therefore, in evaluating the different combinations between the ports, as the model represents, the nearest port is discarded in favour of the shortest maritime route, which optimises the total travel time. These maritime routes are chosen in T* (travel time minimisation). Moreover, for the ODs that use the Palermo–Naples, the outputs of the model show an average total road travel distance of about 540 km, according to the analysis of FCD for trips using the same route, which indicates 450 km (230 km std deviation). A correspondence is also found for Palermo–Genova, where the average total road-travel distance is about 339 km for the model and 482 km for the analysis of FCD with a large variability (357 km std. deviation). For Palermo–Genoa, the model computing the total travel costs highlights that the choice of the sea-road mode is not convenient in terms of total travel costs, but for total travel times. In fact, the solution that minimises costs is the road-only mode. The results of the analysis of the real data (FCD) seem to indicate that the travel time is the most influencing variable.

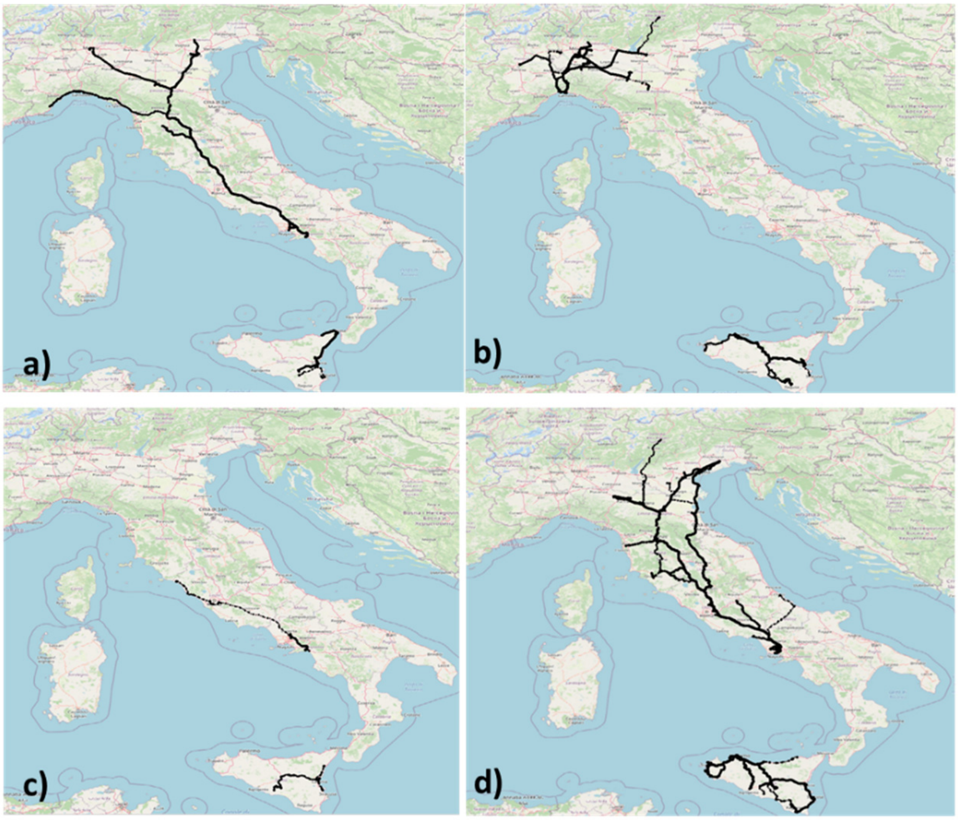

Figure 6. Path choice of FCD trips for each service: (**a**) Salerno–Messina; (**b**) Palermo–Genoa; (**c**) Catania–Salerno; and (**d**) Palermo–Naples.

3.2.3. Average Navigation Time

The comparison between the average values of the navigation times deriving from the processing of all the FCD and the ones computed in the simulation model shows a good fitting (Table 2). The model computed the navigation time as the ratio between the route length and the average speed of different maritime services for each route, calculated from the observation of services with marine traffic in the same analysis period, while the processing of FCD computed an average value from the average navigation time tracking for each vehicle moving on a specific maritime route. This means that differences can be detected among the different Ro-Ro/Ro-Pax services considered. The highest variability seems to be related to the shorter distances.

Table 2. Comparison between the average navigation time derived by FCD and average navigation time computed in the simulation model.

Ro-Ro Services	Avrerage Navigation Time (FCD) [h]	Avrerage Navigation Time (Model) [h]
Catania-Salerno	14.5	13
Civitavecchia-Olbia	10.6	7
Civitavecchia-Palermo	12.7	13.7
Genoa-Palermo	20.3	19.7
Messina-Salerno	9.9	8.4
Naples-Palermo	12.5	10.4
Palermo-Genoa	21.4	19.8
Palermo-Naples	12.9	10.4
Salerno-Catania	14.5	16.8
Salerno-Messina	9.5	7.9
Salerno-Palermo	9.7	10.75

3.2.4. Navigation, and Access and Egress Times

For the maritime services, the mean and standard deviation of navigation are also computed; the access and egress times are expressed in minutes to highlight the variability of the values (Table 3). The average navigation times are close to the times declared by the companies, and there is an increase in the standard deviation as the length of the route increases, which could be due to the different maritime conditions and the different performances offered by the vessels. A comparison of the variability of times in the port shows that Palermo has a greater variability, which may highlight problems of congestion on entry or exit. From the service times in the ports of access, the vocation of some ports (such as Messina, Villa San Giovanni, and Salerno) for this type of traffic emerges, managing to quickly satisfy the demand. In addition to the vocation of the port, the time in the access port and its variability seem to be related to the size of the yard and the extension of the service. The ports of Messina and Villa San Giovanni present particularly short average dwell times less than 35 min and characterised by a low variability; the ports of Salerno, Naples, and Palermo present average dwell times in the access port of around 60, 90, and 130 min and variabilities of around 50, 100, and 130 min, respectively. The time spent in the port of egress presents a measure of the delay induced before the vehicle enters the road system due to the congestion at the port or unloading operations. This information is less precise than the previous information, since the vehicle, depending on its position inside the vessel, may not produce records until the unloading operation is completed; therefore, this value can be cautiously considered as the time to leave the port system, excluding the unloading time. In line with the expectations, these values are significantly lower than those recorded for ports of entry with values less than or around 20 min, with the exception of the ports of Naples and Genoa, for which the average dwell times are around 42 and 56 min, respectively.

Table 3. Statistical time characteristics for Ro-Ro/Ro-Pax services between the Italian mainland and Sicily, derived by FCD.

Ro-Ro Services	Avg Navigation Time [min]	Stdev Navigation Time [min]	Avg Access Time [min]	Stdev Access Time [min]	Avg Egress Time [min]	Stdev Egress Time [min]
Villa San Giovanni-Messina	48	10	33	22	7	4
Messina-Villa San Giovanni	48	11	23	34	8	15
Naples-Palermo	760	66	91	97	19	25
Palermo-Naples	767	52	135	153	42	80
Messina-Salerno	594	45	29	19	23	11
Salerno-Messina	569	31	42	38	13	4
Salerno-Catania	867	35	60	54	20	9
Catania-Salerno	870	28	122	89	9	3
Palermo-Genoa	1284	47	122	51	56	35
Genoa-Palermo	1218	41	317	175	13	10

These values provide the following information: the different weights of loading and unloading phases as implemented in the model, with a consequent correct evaluation of the loading/unloading phases; the necessity to use different values of time for the loading and unloading phases; the possibility of introducing different values with regard to the specific port according to its characteristics also in terms of congestion; and the variability indicates a necessity to introduce a component of uncertainty in the model due to the occurrence of specific situations.

3.2.5. A Comparison of the Specific ODs

The in-depth analysis of the specific ODs that used both transport modes was carried out in order to compare some travel components resulting from the model and outputs of the analysis of FCD. In detail, Table 4 reports a comparison between the samples of road-only trips for some ODs. The comparison shows a good fitting in the total travel times that is comparable with those obtained with real data, demonstrating the representativeness of the model. Moreover, different rest-time values for the road-only transport concern two situations: the short discrepancy might be related to additional break times for refuelling, meals, the physiological needs of the hauliers, or the loading/unloading operations, and significant differences might be due to previous driving hours or the waiting times for the opening of the plant.

Table 4. Comparison of road-only transport for specific ODs.

	Model			FCD Analysis		
	Road-Only			Road-Only		
O/D	**Total Travel Time [h]**	**Transit Time [h]**	**Rest Time [h]**	**Total Travel Time [h]**	**Transit Time [h]**	**Rest Time [h]**
Agrigento-Naples	21.66	11.91	9.75	11.83	10.49	1.35
Bologna-Catania	25.62	15.12	10.50	30.81	17.43	13.38
Caltanissetta-Salerno	9.64	8.89	0.75	11.33	9.44	1.89
Caserta-Catania	8.84	8.09	0.75	10.85	8.77	2.07
Caserta-Palermo	19.80	10.05	9.75	10.65	9.76	0.90
Caserta-Messina	7.62	6.87	0.75	9.98	7.97	2.01
Catania-Naples	8.66	7.91	0.75	9.60	8.71	0.90
Milan-Catania	28.10	17.60	10.50	39.00	20.2	18.8
Naples-Agrigento	21.66	11.91	9.75	10.77	9.95	0.82

Table 5 contains a comparison between the combined sea-road transport for the same ODs specifying the group of Solution (T* or C*) and the potentiality for the sea-road mode (1 or 0). The comparison was carried out on the tracked trips, and it shows a good fitting in the choice of maritime services and total travel times comparable with the average values obtained with real data, demonstrating the representativeness of the model. The possible differences depend on the maritime service or the waiting time, which represents the time spent in port due to the early arrival of the hauliers to the port. The greatest differences in terms of the total travel times were found for driving times close to the maximum value allowed by law, or for trips with an arrival time during the closure of the factory.

Table 5. Comparison of the sea-road transport for specific ODs.

				Model								FCD Anlysis			
				Combined Sea-Roade								Combined Sea-Roade			
O/D	**Solution/ Potentiality**	**Maritime Route**	**Total Travel Time [h]**	**Rest Time [h]**	**Access Time [h]**	**Navigation Time [h]**	**Egress Time [h]**	**Maritime Route**	**Total Travel Time [h]**	**Rest Time [h]**	**Access Time [h]**	**Time Spent in Port [h]**	**Navigation Time [h]**	**Time Spent in Port [h]**	**Egress Time [h]**
Agrigento-Naples	T*/1	Palermo-Naples	15.67	0.00	4.04	8.50	0.13	Palermo-Naples	18.79	0.47	2.78	1.28	12.87	0.24	1.16
Bologna-Catania	T*/1	Salerno-Messina	21.14	0.75	8.09	7.90	1.40	Naples-Palermo	19.49	0.65	7.91	0.66	9.49	0.10	0.68
Caltanissetta-Salerno	T*/0	Palermo-Naples	13.91	0.00	1.67	8.50	0.73	Palermo-Naples	16.13	0.00	2.91	0.58	11.93	0.19	0.52
Caserta-Catania	T*/0	Salerno-Messina	13.36	0.00	1.06	7.90	1.40	Salerno-Messina	13.72	0.25	1.20	0.80	9.17	0.20	2.09
Caserta-Palermo	T*/1	Naples-Palermo	14.05	0.00	0.46	10.40	0.19	Naples-Palermo	17.20	1.23	1.08	1.58	11.92	0.47	0.92
Caserta-Messina	T*/0	Salerno-Messina	12.14	0.00	1.06	7.90	0.18	Salerno-Messina	32.58	10.81	1.20	0.32	8.92	0.16	11.16
Catania-Naples	T*/0	Messina-Salerno	13.67	0.00	1.40	8.40	0.87	Catania-Salerno	18.60	0.00	1.17	1.98	14.43	0.20	0.82
Milan-Catania	T*/1	Catania-Genoa-Salerno-Catania	25.67	0.00	10.58	9.75	0.35	Salerno-Messina	33.19	10.50	10.88	0.32	9.49	0.33	1.67
Naples-Agrigento	C*/1	Catania-Genoa-Salerno-Catania	17.87	0.00	0.87	9.75	4.25	Naples-Palermo	16.72	0.00	0.42	1.49	11.80	0.17	2.85

3.3. Conceptual Information for the Model Setting

The in-depth analyses of FCD provide additional information that permits both to evaluate the model setting and identify other relevant aspects for the integration in further developments of this model, but also in the freight modelling approach. The analysis revealed some issues that both influence the modal choice and are strictly correlated: the type of commodities; the road distances differently weighting the access and egress phases regarding total travel times; recurring behaviours of the trucking companies; the complexity of evaluating the rest period and its incidence; and the existence and timetable of the maritime services.

3.3.1. Types of Goods

The geographical analysis using Open Street Maps (OSM) and Google Street View permittedto both distinguish the rest periods and the loading/unloading phases and to identify the origin and destination of the trips. The types of goods transported in each trip were identified by comparing the pings emitted by the vehicles during the loading/unloading phases with the satellite images of the yards and the information of the plants. In detail, a distinction was made between the transportation of perishable goods, characterised by companies at the origin and destination of the journey that only handle perishable goods, and non-perishable goods (Table 6). It can be observed that the majority of the identified goods (38%) are perishable goods. A total of 31% of the trips concern non-perishable goods and the remaining 31% is not distinguishable because the companies at the origin and destination deal with different types of goods. According to the transport mode, it can be observed that, in the case of combined transport, 42% of the trips concern perishable goods and 35% non-perishable goods, while 23% cannot be characterised. In the case of the road-only transport, 33% of the trips concern perishable goods, 27% of journeys concern non-perishable goods, and 40% cannot be characterised. Generally, for the analysed sample, it can be observed that most of the goods transported from Sicily are perishable, while most of the goods destined for Sicily are non-perishable.

Table 6. Distribution of trips distinguishing two macro categories of goods, transport mode, and origin and destination derived by FCD.

Type of Goods	All Trips	Sea-Road	Road-Only	From Sicily	To Sicily
Not-identified	31%	23%	40%	32%	31%
Non-perishable goods	31%	35%	27%	24%	38%
Perishable goods	38%	42%	33%	45%	31%

Perishable goods, specifically those that are frozen, mainly use combined sea-road transport. This might be due to the shorter total travel times for longer distances, where the rest time has a high incidence, and also to the timetable of maritime services, especially in relation to the arrival time at the destination. Moreover, the uncertainty of travel times is also a relevant component in this conceptual analysis. Therefore, the consideration of the type of commodities is a relevant aspect in freight modelling.

3.3.2. Impact of Trip Length in the Mode Choice

The trip length seems to be an important modal choice variable; with the increase in the distance, the data show an increase in the choice of the sea-road transport and, consequently, a decrease in the road-only transport (Figure 7).

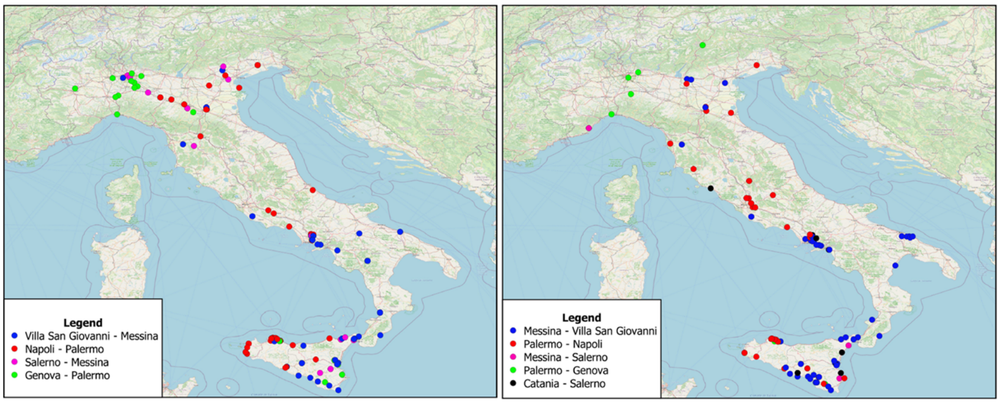

Figure 7. Distribution of ODs trips from/to Sicily derived by FCD distinguishing maritime service.

The access distance of 81% of the trips originating in Sicily is between 50 and 150 km, highlighting the importance of the proximity to the port in the access phase. The simulation model considers the access and egress phases separately, and the formulation entails the possibility to introduce different weights to the two phases, according to the results of processing FCD. Moreover, it seems important to establish a constraint on the length of the access and egress phases, as formalised in the model. Hauliers seem to choose the alternative, which minimises the arrival time at the destination to complete the unloading operation before the plant closure.

3.3.3. Rest Periods

The rest time is the most difficult component to both study and evaluate. It is subject to the drivers' rest-time regulations, imposing numerous stops on the driver, which do not always take place in service areas or stations, but directly in factory yards. Additional rest-time components have to be considered in the freight model to correctly compute the total travel times. The setting of the model distinguishing a single trip might be correct, with regard to the systematic trips (recurring behaviour) between the same and few ODs. For a sort of trip chaining, the additional component of the times depending on previous driving hours have to be computed. In fact, the values are not deterministicas imposed by the regulations, but are characterised by a variability in the lower and higher values. The rest issue supports the view of the forwarder that he also has to take into account the link of a previous and a subsequent journey for the calculation of the rest time. Moreover, the rest time might be a decisive variable in the modal choice. If the navigation time is long enough to absorb the rest time, depending on the access phase or any previous trips, then the combined transport is a convenient choice. Finally, the analysis shows that hauliers for trips with a duration close to the legal limit, try to reach the destination by postponing the rest period.

3.3.4. Ro-Ro Services

The modal choice is influenced by the existence of the maritime services and also by the departure and arrival times. The timetables of the services do not coincide with the desired departure and arrival times, or otherwise entail longer journeys. For example, Ro-Pax services are often characterised by higher speeds entailing lower navigation times, but the timetables are poorly suitable for freight transport, which require the arrival at a destination before plant closure. The analysis of road-only trips originating near ports shows that another decisive variable in the choice of combined transport is the time of day. In fact, vehicles using combined transport leave and travel to the port at recurring times, highlighting the importance of the presence of maritime services. Otherwise, road-only trips are characterised by different departure times from the origin and/or arrival times at the destination, reflecting the greater flexibility of this mode in responding effectively to needs of the demand The trucking companies vary their departure times according to the service, which allows them to arrive at their destination during the opening hours of the companies. It means that the choice between the two transport modes depends not only on the distance, but also on the time of service and, consequently, the arrival at the destination. Two levels of choice are distinguished that are strictly linked to the length of access and egress phase: (1) the estimated time of arrival at the port from the origin, in order to use the service with the shortest waiting time, and (2) the estimated time from egress to destination in time for the unloading operation as a desired time of arrival. According to the in-depth analysis, the choice between the two modes is made by evaluating the existence of a service with respect to the desired arrival time and the possible departure time. If the vehicle can reach the port at the desired time, the service is attractive. Naples–Palermo and Palermo–Naples allow for the arrival at the egress port between 8 and 7.30 a.m., reaching their destination in the morning or during the day before plant closure. This route is mainly used by services transporting perishable goods, respectively, 56% and 51%. The possibility of resting on board and being able to make the trips without resting, increases the attractiveness of the service. It means considering a sort of basin of influence with a radius of less than the maximum driving hours per day, without rest times for the access or the egress phases depending on the navigation time. A long navigation time is a loss of time for the small access distance (for example, from Sicily), but not in the case where the vehicle previously accumulated numerous driving hours. The timetable of the services might influence the modal choice if the rest period is at night after the unboarding operation. In fact, Salerno–Messina seems attractive for egress trips with a duration close to the maximum allowed by the rest regulations. A long navigation distance is useful if the access time is about equal to the maximum daily driving hours and the navigation adsorbs the rest period. The arrival time in the egress port requires a waiting time for the opening of the destination plants. In this case, 33% of the trips concern logistics centres operating 24 h a day. Naples–Palermo, entailing a large basin of influence, allows vehicles to leave during the opening hours of the production plants, rest on board, and reach the destination without taking daily rest on the road. Vehicles arriving in the egress port at about 7.30 a.m. are able to reach their destination in the morning with measured distances of less than 200 km. Moreover, a preference of hauliers to rest on board the vessel has been observed when the travel time is close to the legal limit. Generally, it seems that it is the demand that has adapted to the services.

3.3.5. Recurring Behaviour and Systematic Trips

The analysis shows a recurring behaviour of trucking companies. In detail, the tracked vehicles were 64, and, during the period of analysis, 25% only used the road-only transport (about 8 times each). A total of 14% of the vehicles used both modes of transport, with an average use of Ro-Ro services of 5 times per month; 11% used different combined transport services 7 times per month; and 50% maintained the same combined transport service with an average number of monthly uses of 1.5. It follows that 75% of drivers did not change their mode of transport during the month under analysis, underlining the possibility of a propensity for a specific mode of transport. The investigation identifies the establishment of relationships between the plants and trucking companies that regularly move between the same ODs (two or more). For these vehicles, there is a strong sense of habit in the mode, route, destinations, and travel times. As an example, Figure 8 shows the route taken by a driver in the month of analysis, which illustrates the factors above.

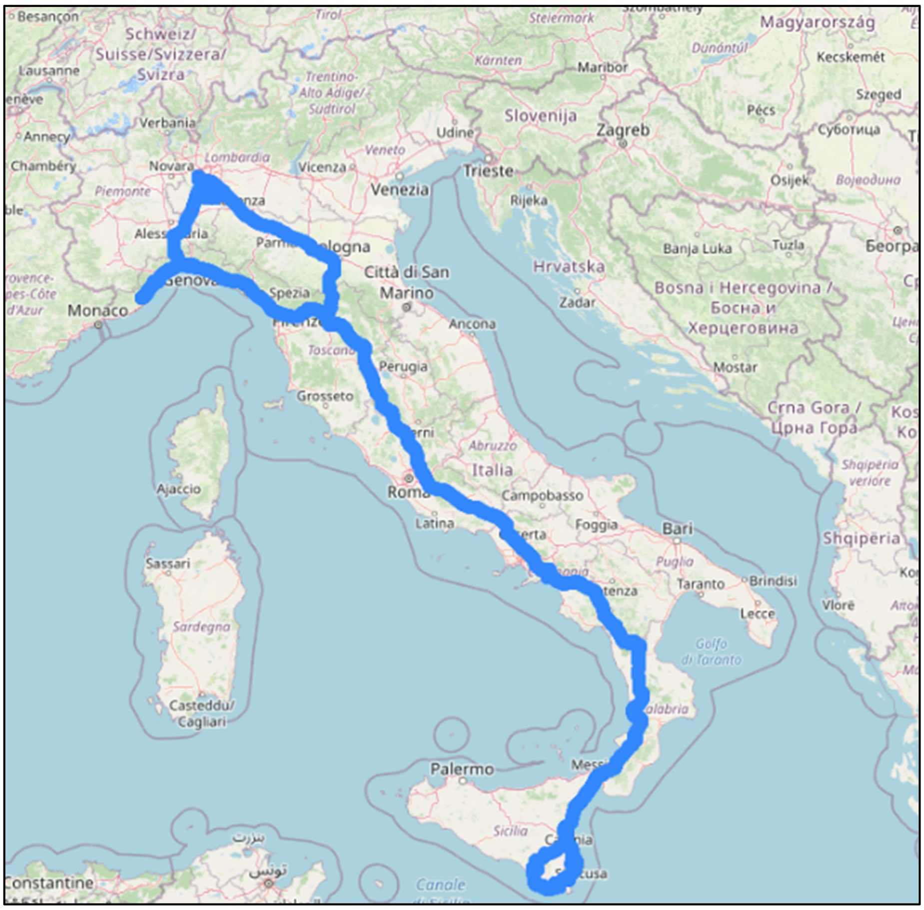

Figure 8. Example of recurring behaviour for a specific vehicle.

In Northern Italy, the vehicle repeatedly travels to a logistics operator, using the Messina–Salerno service and, in Sicily, to a production plant, using the Villa San Giovanni–Messina service. The choice of mode seems to be linked to the type of goods, the needs of the companies, the service timetables, and the travel times. The vehicle on a journey to Northern Italy is loaded at around 20:00, then takes the Messina–Salerno service at 00:30, arriving at around 11:00, and arrives at its destination at around 23:00. Otherwise, the vehicle is loaded around 17:30 and heads towards the port of Villa San Giovanni Messina where it arrives at around 20:30, with an average of 12 h of rest before using the service, and a further 8 h after using the service with an arrival time of around 8:00. The average journey time to the north is 27 h, while the average journey time to the south is 39 h.

4. Conclusions

Several studies indicate the new tendency of using Big Data and, specifically, FCD both for the validation phase and the development of behavioural models in the transport sector. FCD provide a wide range of geographically information which, if processed correctly, can support the transport planner's tools by providing a rich and robust database that can be used for model validation and model refinement. The present paper concerns an innovative use of FCD for the validation phase of a freight transport simulation model in a multimodal context. The validation process proposed is not limited to the comparison of some indicators, but also involves the development of in-depth analyses by tracking the single vehicles' movements, in order to obtain additional aspects for validating the freight transport model. The phase of processing FCD has been characterised by a high complexity, because reconstructing the trips of heavy vehicles is much more complex than private ones, since these are subject to different phases, such as loading and unloading, break times, and rest periods. Despite these complex activities required for identifying the freight trips, the analysis of FCD permits us to obtain some relevant information about the model validation with a comparison of the real world data, providing quantitative/qualitative information. Qualitative and quantitative analyses allow for the validation of the model outputs and a better setting for some elements. Specifically, the validation process confirms many choices of the simulation model as the prevailing use of the motorway network, according to the path choice in the simulation model or the validity of the approach of the model concerning the choice of longer road distances for the access and egress phases, in order to minimise the travel times. Additionally, the numerical outputs of the model are similar to the data provided by FCD analysis as the average total road distances for the combined sea-road mode, the average navigation time, the necessity to use different values of time for the loading and unloading phases, and the importance of considering the variability of the temporal components with an uncertainty variable. The analysis of FCD indicates that travel costs and travel times are the most important factors considered by trucking companies in the modal and path choice, thus also confirming the results provided by the interviews in [15]. The comparison of trips travelled both in the combined mode and in the road-only one allows us to understand the behaviour and the choices of the hauliers. Many choices concerning travel depend on the specific need of each trip, highlighting the importance of keeping the two variables separated (travel times and costs).

According to the obtained results, it is important to highlight the positive impact of such validation processes for the capacity of better specifying the simulation model. In this way, it is possible to obtain more realistic models that are able to support the transportation planner activities for defining the actions and policies for a more sustainable freight transport system.

Author Contributions: Conceptualisation, D.B., M.P. and A.R.; methodology, D.B., M.P. and A.R.; validation, D.B.; formal analysis, D.B.; investigation, D.B.; resources, M.P. and A.R.; data curation, D.B.; writing—original draft preparation, D.B. and A.R.; writing—review and editing, D.B. and M.P.; visualisation. D.B.; supervision, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Croce, A.I.; Musolino, G.; Rindone, C.; Vitetta, A. Transport System Models and Big Data: Zoning and Graph Building with Traditional Surveys, FCD and GIS. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 187. [CrossRef]
2. Chadil, N.; Russameesawang, A.; Keeratiwintakorn, P. Real-time tracking management system using GPS, GPRS and Google earth. In Proceedings of the 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, Thailand, 14–17 May 2008; pp. 393–396. [CrossRef]
3. Bojan, T.; Kumar, U.; Bojan, V. Designing vehicle tracking system—An open source approach. In Proceedings of the 2014 IEEE International Conference on Vehicular Electronics and Safety (ICVES), Hyderabad, India, 16–17 December 2014. [CrossRef]
4. Rahman, M.M.; Mou, J.R.; Tara, K.; Sarkar, M.I. Real time Google map and Arduino based vehicle tracking system. In Proceedings of the 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh, 8–10 December 2016; pp. 1–4. [CrossRef]
5. Ajmar, A.; Arco, E.; Boccardo, P.; Perez, F. Floating Car Data (Fcd) for Mobility Applications. *Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 1517–1523. [CrossRef]
6. Pascale, A.; Deflorio, F.; Nicoli, M.; Dalla Chiara, B.; Pedroli, M. Motorway speed pattern identification from floating vehicle data for freight applications. *Transp. Res. Part C Emerg. Technol.* **2015**, *51*, 104–119. [CrossRef]
7. Ehmke, J.F.; Steinert, A.; Mattfeld, D.C. Advanced routing for city logistics service providers based on time-dependent travel times. *J. Comput. Sci.* **2012**, *3*, 193–205. [CrossRef]
8. Fancello, G.; Serra, P.; Schintu, A.; Zoratti, A. Performance Evaluation of a Tracking System for Intermodal Traffic: An Experimentation in the Tyrrhenian Area. *Eur. Transp. Trasp. Eur.* **2020**, *11*, 142.

9. Shen, L.; Stopher, P.R. Review of GPS travel survey and GPS data-processing methods. *Transp. Rev.* **2014**, *34*, 316–334. [CrossRef]
10. Cich, G.; Knapen, L.; Bellemans, T.; Janssens, D.; Wets, G. Threshold settings for trip/stop detection in GPS traces. *J. Ambient. Intell. Humaniz. Comput.* **2016**, *7*, 395–413. [CrossRef]
11. Xiao, G.; Juan, Z.; Zhang, C. Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. *Transp. Res. Part C* **2016**, *71*, 447–463. [CrossRef]
12. Li, Q.; Zheng, Y.; Xie, X.; Chen, Y.; Liu, W.; Ma, W. Mining user similarity based on location history. In Proceedings of the 16th ACM Sigspatial International Conference on Advances in Geographic Information Systems, Irvine, CA, USA, 5–7 November 2008; p. 34. [CrossRef]
13. Alvares, L.O.; Bogorny, V.; Kuijpers, B.; De Macedo, J.A.F.; Moelans, B.; Vaisman, A. A model for enriching trajectories with semantic geographical information. In Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, Washington, DC, USA, 9–11 November 2007; pp. 1–8. [CrossRef]
14. Comi, A.; Russo, F.; Vitetta, A. Calibration of generalized cost models for road freight vehicles within an intermodal chain: An experimental application in Italy. In Proceedings of the European Transport Conference, Strasbourg, France, 3–5 October 2005.
15. Lupi, M.; Farina, A.; Orsi, D.; Pratelli, A. The capability of motorways of the sea of being competitive against road transport. The case of the Italian mainland and Sicily. *J. Transp. Geogr.* **2017**, *58*, 9–21. [CrossRef]
16. Renna, A.; Petrelli, M.; Carrese, S.; Bertocci, R. An Innovative Simulation Agent-Based Model for the Combined Sea-Road Transport as a DSS. *Sustainability* **2021**, *13*, 10773. [CrossRef]