



## A PROBABILISTIC CELLULAR AUTOMATON TO FORECAST PORT CHOICE DECISIONS

Mabel A. LEVA<sup>1</sup>, Alejandro LEÓN<sup>2</sup>, Rodrigo A. GARRIDO<sup>3\*</sup>, Ángel IBEAS<sup>4</sup>

<sup>1</sup>Research Group in Transport Systems, University of Cantabria, Spain

<sup>2</sup>Institute of Basic Sciences, Diego Portales University, Chile

<sup>3</sup>Faculty of Engineering and Sciences, Diego Portales University, Chile

<sup>4</sup>Dept of Transport and Technology Projects and Processes, University of Cantabria, Spain

Received 2 June 2016; revised 24 November 2016; accepted 25 July 2017

**Abstract.** The port choice problem consists in predicting the selection of a port, made by an agent who has alternatives to choose from. Most of the literature has tackled this problem assuming a discrete choice model dependent on the ports' characteristics and agents' attributes. However, in practice the port choice decision depends also on the choices made by other agents as well as decisions made by these agents in the past. There are only a few examples that incorporate the complexity generated by spatio-temporal interactions between agents. However, those modelling structures are rather cumbersome, precluding their use in practical cases. This article presents a new modelling framework to predict port choice decisions, based on the theory of Cellular Automaton (CA), which is simple in structure and can be quickly calibrated and applied. This framework is a probabilistic CA intended to imitate the decision processes made from multiple shippers that interact with each other. These shippers face similar alternatives of seaports for exporting their products within a certain time span. The port choice here is a dynamic decision that depends on the ports' characteristics and attributes of each shipper at a given time, as well as the decisions made by their neighbours. The outcome of the interaction is a discrete decision that evolves in time according to the dynamics of the system as a whole. The specified CA was applied to the case of vehicle exports from Brazil and the calibration was performed through a genetic algorithm. The results show that the probabilistic CA is able to replicate the historic behaviour of the port choice decisions in the Brazilian vehicle industry, with a high degree of success. The spatial component of the CA turned out to be of major relevance in the dynamic decision process along with the attributes and geographical location of ports.

**Keywords:** port choice, cellular automaton, probabilistic, port competition, hinterland formation, discrete choice.

### Introduction

More than 85% of the cargo transported globally is moved through the maritime transportation system. This characteristic seems to be a long-run trend, especially when the size of the vessels and the ports infrastructure has risen considerable in the last decade and so has the transference capacity and specialized labour in this industry.

Given this scenario, the port planning processes will need to depend on studies and models to predict (to a certain extent) the load of the different port nodes in a transportation network, in order to balance the transference capacity, cost and time consumption and keeping the negative externalities under control.

The distribution of cargo between the different available competing ports (also called demand capture) is usually analysed from two perspectives: the shippers' view-

point and the shipping company's viewpoint. The shipper's perspective is that of choosing the port that maximises some utility criterion. For example, the shippers may rely on models whose main explanatory variable is the generalized cost of transporting the cargo between ports and (from) destinations (origins), within their hinterlands. From the perspective of the shipping companies, the choice between alternative ports rely on charges, rates and ports' level of service, as explanatory variables that define the port's efficiency (Tongzon, Sawant 2007). There are though, additional aspects that condition the port choice. In fact, according to Chou *et al.* (2008) there are at least six relevant aspects, which are: local cargo volume; terminal handling charge; berth availability; port location; transshipment volume and feeder network.

\*Corresponding author. E-mail: [rodrigo.garrido@udp.cl](mailto:rodrigo.garrido@udp.cl)

In both cases, only a subset of the explanatory variables is included into the models, mainly because of difficulties to measure them. In fact, in some cases these variables are economically sensitive for the stakeholders and, most importantly, it is difficult to predict them for the entire planning horizon. For example, key aspects such as bargaining power, specialization in a niche (e.g. refrigerated cargo), long-term contracts, commercial integration, among others, are left out of the models for the difficulties explained above.

The problem to be tackled in this article is that of the prediction of port choice from the shippers' perspective, considering variables related to port performance as well as spatial and temporal interaction, i.e. the influence of one shipper's decision over another that is geographically close and the effect of decisions along the time line.

The hypothesis about the influence of the shipper's decisions over one another is backed up by empirical evidence. In fact, large shippers affect the port choice decision of other shipping companies (Tiwari *et al.* 2003), due to the higher volumes (Chou *et al.* 2008). These large volumes generate the proper conditions for new services to flourish and also improve the shippers bargaining position *vis-à-vis* the shipping companies. With the presence of new shipping services, smaller shippers also improve their bargaining position and consequently smaller shippers will tend to follow the large shippers, most likely their neighbours, to take advantage of the long-term relationship between the large shippers and shipping companies. At the same time, the shippers' location is linked to similar economic activities and consequently there is also an indirect relationship between spatial location and port decisions for both actors.

This article attempts to develop a new modelling framework to describe and predict the behaviour of decision makers in the maritime transport chain, in which various shippers choose a port of origin to export their shipments during a period of time. To attain this main objective, we adapt the structure of a CA model, in order to capture the spatio-temporal interactions, that in other modelling contexts are rather difficult to calibrate and operate (Garrido, Leva 2004).

## 1. Literature review

### 1.1. The port choice in maritime transportation

The usual explanation for the port choice behaviour in the specialized literature is the generalized transportation cost from/to the hinterland to/from the ports and the reliability of the total route within the context of a supply chain instead of the sole port choice. Other studies point out significant differences in the triggers of port choice between different commodities (Malchow, Kanafani 2001). According to Moya and Feo Valero (2017), the identification of the decision-maker is one of the main critical issues in port choice modelling. Furthermore, they point out that the decisions of different agents are highly interrelated and inherently iterative. Due to the latter, the determinants of

the decision process vary significantly with the characteristics of the decision-maker, its logistics requirements and objectives. Tongzon (2009) studies the port choice from the freight forwarders' viewpoint; finding out that the main factors in this decision are port efficiency, shipping frequency, adequate infrastructure and location. Magala and Sammons (2008), studied the port choice problem incorporating the concept of supply chain, pointing out that shippers select a set of logistics services, in which the port itself is only part of the whole chain.

There are authors who studied the port choice problem from the shipping lines perspective, where the port charges and level of service at the port become the fundamental decision variables (Tongzon, Sawant 2007). However, according to Chou *et al.* (2008), there are at least six relevant decision variables: local cargo volume, port's charges, berth availability, port's location, transferred volume and feeder connection.

Other authors have studied the problem of port choice considering various types of actors. For example, Song and Yeo (2004) identified 73 elements that contribute to the global competitiveness of the main ports in China, considering the opinions of ship-owners, shippers, shipping companies, terminal operators, researchers and scholars. After this broad analysis Song and Yeo (2004) concluded that location was the most relevant variable as far as competitiveness is concerned.

Lee *et al.* (2010), identify and evaluate the main determinants of the port choice from the carriers, shippers and terminal operator's perspectives. They identify 38 attributes for each of the seven analysed categories (port location, port characteristics, port's operation conditions, hinterland characteristics, ships characteristics, among others). The above-mentioned stakeholders evaluated these attributes, finding considerable variability in the assessment of each attribute by each stakeholder, and only one attribute that was commonly evaluated as relevant to all the stakeholders: port accessibility. However, further analysis of the shippers and containers' terminal operators, revealed that they had a larger number of common attributes and also giving a larger weight to "port's operation conditions" than that given by the shippers.

### 1.2. Port choice modelling

The common family of port choice studies is discrete choice models (Manski, McFadden 1981). These models assume the following: there are rational agents that face the selection of a single choice among a choice set formed by all the alternatives available to her. Each alternative can be conveniently described by its own attributes. The selecting agent has a utility function (a score) that depends on the attributes of each alternative, some individual characteristics plus a random component that explains deviations from the rational behaviour (given the available explanatory variables). Therefore, the agent chooses a certain alternative according to the probability that the random utility function is maximized when selecting that particular alternative. One of the most popular discrete choice

models based on random utility theory is the MultiNomial Logit (MNL) model (Hausman, McFadden 1984), which requires two conditions: all the available alternatives have to be uncorrelated between them, and the random component of the utility function has to be homoscedastic (i.e. the variance of all the alternatives' utility functions have to be equal). Examples of application of the MNL to the port choice problem are found in Veldman *et al.* (2016); Tiwari *et al.* (2003); Malchow and Kanafani (2003).

These MNL models show a good performance in terms of nominal prediction power. However, it is likely that the reported performance was highly biased due to the violation of MNL requirements (homoscedasticity and independence between alternatives) to produce unbiased and consistent estimators. Indeed, the choice set (i.e. their available alternatives) had significant correlation and heteroscedasticity. Those conditions are commonly found in the port choice problem. In fact, the spatial proximity of different port terminals and the type of long-term contracts established between shippers and carriers that define *ex-ante* the port terminals to be used within the duration of the contract. These effects create both spatial and temporal correlation that is not easy to deal with in the context of discrete choice models.

In addition, the maritime conditions, the ports infrastructure and the interface with the land transportation system (to name a few) impose large variability in the port services, which creates the ideal conditions for a heteroscedastic behaviour when considered as alternatives within a choice set. As a result, there are only a few examples of published studies in specialized journals that deal with more robust models that allow the violation of the MNL restrictions. One of these attempts is the modelling framework proposed by Veldman *et al.* (2013). This approach partially solves the problem of heteroscedasticity through the use of the Hierarchical Logit (HL) model (Papola 2004). However, the use of HL models avoids only a small fraction of the possible correlation/heteroscedasticity problem. The latter is due to the rigid structure of the variance-covariance matrix that the HL imposes.

To the best of the authors' knowledge, the only modelling approach that deals simultaneously with a general structure that allows alternatives with a general random component's specification is Garrido and Leva (2004). In that study the authors present a general MultiNomial Probit (MNP) model (Garrido, Mahmassani 2000). The MNP model put forward by those authors incorporates spatial and temporal correlations between alternatives and a general structure for the variance-covariance matrix. The model gives a good tool for forecasting port choice as well as representing the interactions between agents in a coherent way. However, there are complications with the calibration process (numerical stability, multiple maxima for the likelihood function, large computational effort, among others) that makes it difficult to be used in practice or by non-econometric experts.

There are other methodologies developed more recently to address this very same problem. One of them is

the approach proposed by Yeo *et al.* (2014). They propose a port choice model based on fuzzy logic. In this approach the authors combine various data sources and formats, both objective and subjective, to assess the performance of a given port, determined through a "fuzzy score" using degrees of belief to come up with a criterion to establish the port choice. This modelling approach was applied to the choice of container ports in northeast Asia. The outcome of the model is a preference score that allows the comparison of different ports either *vis-à-vis* other ports or the system as a whole. However, this methodology is strongly dependent on the opinion of the stakeholders considered in the sample, which can easily change in line with their preferences at that particular time.

### 1.3. CA modelling

The Cellular Automaton (CA) seems to be a direct tool to deal with a phenomenon that is dynamic, non-linear and complex (Goles, Martínez 2010) with significant computational efficiency and with reasonable data needs. The CA can be defined, for our purposes, as a mathematical model belonging to the field of computational physics. Proposed originally by John von Neumann during the 1950's to simulate and describe complex systems as a massive collection of simple objects that interact with each other locally. These simulation models proved to be especially adequate to study dynamic systems that evolve in discrete steps (Griffeath, Moore 2003; Chopard, Droz 2005).

The simulation structure is based on rather simple interactions among shippers and between shippers and ports to predict the port selection. The port choice emerges as a result of the dynamic interactions with feedback between these decision-makers. In spite of the simplicity of these interactions, the proposed methodology captures the complex behaviour of the system and predicts the port choice evolution with high accuracy. In addition, the modelling framework is simple enough to be quickly calibrated and applied to real cases.

This tool would be of great value for methodological and practical reasons alike. From the theoretical perspective, this approach could be a tool for searching complex patterns (chaotic behaviour, sensitivity to initial conditions, etc.) in the dynamics of market share formation, as well as identifying attributes that influence a port's market share. On the practical side, this approach is useful to establish not only the current market share of a port but also its hinterland and the variables that affect its boundaries within a competitive ports' system.

## 2. Design of a CA for port choice

### 2.1. A simulation structure for the port choice problem with a CA model

Consider a given spatial region that serves as a shared hinterland for a network formed by  $m$  ports. We divide this region into  $n$  mutually exclusive and collectively exhaustive Freight Generation Zones (FGZ). Each FGZ can have

$m + 1$  states of activation: one for each available port and one state of inactivity, meaning that if a FGZ “has chosen” port  $i$  as a port of call, then its current state is  $i$ , and if the FGZ has not chosen any port then it is not active, hence its state is Inactive. Thus, the  $i$ -th FGZ has a state function as follows:  $e_i = j$ , with  $j = 0, 1, 2, \dots, m$ , where 0 means inactivity and  $j$  means that  $i$ -th FGZ has chosen port  $j$ .

Figure 1 shows a scheme of the FGZ and ports on which a port choice CA model will operate. Each FGZ, represented as a circle, corresponds to a single cell of the CA. Thus, each FGZ acts like a CA that is either active or inactive. A given FGZ is active during discrete time  $t$  if it sends a shipment to an available port. The port choice decision of that FGZ at time  $t$  depends on the distance to the port, the operational characteristics of the port (e.g. tariffs, availability of dedicated space for incoming vehicles, additional services, expected waiting times, transfer capacity, specialization, long term contracts, bargaining position, among many others). Our novel approach assumes that the condition of activity at time  $t$  also depends on the state of its neighbouring FGZ.

In addition, this modelling framework assumes that the port choice not only depends on the attributes mentioned above, but also incorporates a stochastic element that would explain (similar to the assumption of random utility theory) at least two phenomena. First, the modeller ignores the actual attributes of each port at any given time. Secondly, a FGZ could become active/inactive triggered by exogenous variables not included in the model specification. Thus, our modelling approach will be called Probabilistic Port Choice Cellular Automaton (PPCCA).

### 2.1.1. The neighbourhood concept in the PPCCA model

There are several ways to interpret the concept of neighbourhood in CA (Goles, Martínez 2010; Griffeath, Moore 2003; Chopard, Droz 2005). For example, Von Neumann introduced the concept of neighbourhood in a rectangular grid as the cells sharing a common edge (four neighbours in a squared grid). Other classical concept is that of Moore neighbourhood, in which a neighbour is each one of the eight cells surrounding a given cell in a squared grid. In the port choice case, the FGZ not necessarily are ordered in a Manhattan metric array. Consequently, the concept of regular grids does not fit the actual spatial pattern. Instead, we incorporate the concept of radial neighbourhood, in which two cells are  $R$ -neighbours if they both are circumscribed in a circle of radius  $R$ , as shown in Figure 2.

### 2.1.2. Transition function

In the PPCCA model each cell corresponds to a FGZ that transits from its current state to the next one (in discrete time steps) according to a stochastic transition function. Probabilistic CAs update their state according to a rule that follows a certain probability distribution. Accordingly, they belong to the category of discrete-time random dy-

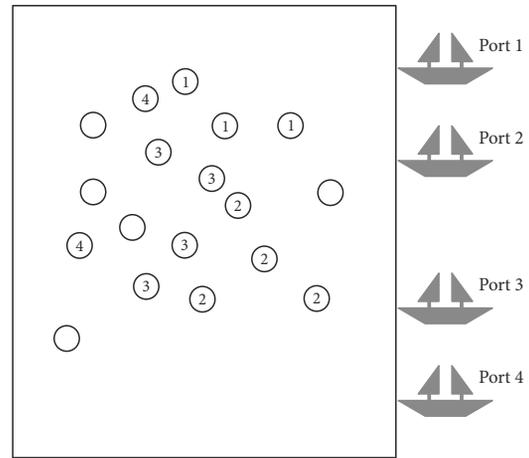


Figure 1. Conceptual scheme of FGZ and ports within a shared hinterland

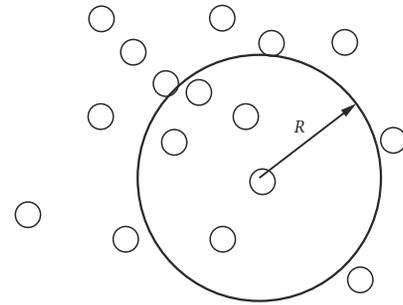


Figure 2.  $R$ -neighbours in the PPCCA are those FGZ within the circle

namical systems. Incorporating stochasticity in the spatial interaction between cells allows these models to identify complex behaviours that may emerge from the system.

The PPCCA will update the state of each FGZ, as a probabilistic CA, according to the following rules:

- Each zone has a real number associated with it:  $z \in \mathbb{R} \wedge 0 \leq z \leq 1$ . This number is calibrated by a genetic algorithm (that will search for the optimal combination of parameters that best fit a set of observed data). At each simulation step,  $z$  is replaced by another real number, drawn from a normally distributed sample whose mean value is  $z$  and its standard deviation is  $\sigma$  (also calibrated by the genetic algorithm).
- Each FGZ  $i$  has a function  $\Phi_1^i \in \mathfrak{R}^m$  whose domain is the  $R$ -neighbourhood of the  $i$ -th FGZ. This function's outcome gives a vector of  $m$  components. Each component is formed by the ratio between the number of FGZ within the  $R$ -neighbourhood that choose the port  $j$ , relative to the total number of active FGZ within the same  $R$ -neighbourhood. i.e.:

$$\Phi_1^i = \left( \frac{n_1}{N_i}, \frac{n_2}{N_i}, \dots, \frac{n_m}{N_i} \right), \quad \forall i = 1, 2, \dots, n, \quad (1)$$

where:  $n_1, n_2, \dots, n_m$  is the number of FGZs that chose the port 1 to  $m$  respectively in a given time

- step;  $N_i$  is the total number of active zones within the  $R$ -neighbourhood of  $i$ -th FGZ.
- Each FGZ  $i$  has a function  $\Phi_2^i \in \mathfrak{R}^m$  whose domain is the set of distances to the  $m$  ports.

$$\Phi_2^i = (\theta_d \cdot d_{i1} + \theta_c \cdot c_1, \theta_d \cdot d_{i2} + \theta_c \cdot c_2, \dots, \theta_d \cdot d_{ij} + \theta_c \cdot c_j), \quad \forall i=1, 2, \dots, n, \quad (2)$$

- where:  $\theta_d$  and  $\theta_c$  are coefficients to be calibrated by the genetic algorithm;  $d_{ij}$  represents the distance from the  $i$ -th FGZ to the port  $j$  (with  $j = 1, 2, \dots, m$ );  $c_j$  represents an aggregated proxy of the various costs involved in the choice of port  $j$  (note that these coefficients may implicitly include the costs associated to the carrier selection). As the  $c_j$  coefficients represent some generalised cost function, hence smaller values are preferred to larger ones (they can also be interpreted as a proxy for the *disutility* of choosing port  $j$ ).
- The function  $\Phi_2^i$  is mapped as follows: the smallest component of  $\Phi_2^i$  is assigned a probability  $p$ , the second smallest component is assigned a probability  $p \cdot (1-p)$  and so forth decreasing as a geometric law until reaching  $p \cdot (1-p)^{m-1}$ . The new mapped function, that contains the probability values (instead of distances) for each component, is renamed  $\Phi_2^{i'}$ .
  - Finally, a compound function is defined as follows:

$$\Phi^i = \alpha \cdot \Phi_1^i + (1-\alpha) \cdot \Phi_2^{i'}, \quad \forall i=1, 2, \dots, n \quad (3)$$

- with  $\Phi^i \in \mathfrak{R}^m$ . In this specification, the parameter  $\alpha$  represents the influence of the  $R$ -neighbours' behaviour on the port choice decision, i.e., the spatial component of the decision, associated to the FGZ  $R$ -neighbours' interaction, and  $(1-\alpha)$ , the influence of the distance and other systematic attributes related to the "attractiveness" of a port, on the port choice decision. The attractiveness here has opposite sign to the coefficients  $c_j$ .
- If  $T \leq z$ , (where  $T$  is an activation threshold calibrated by the genetic algorithm) then the state  $e_i$  of each FGZ is updated with a value between 0 and  $m$ , representing the choice (0 corresponds to inactivity) corresponding to the highest component of the function in Equation (3).

### 2.1.3. Model parameters

According to the specification described above, the following parameters need to be calibrated before the model can simulate the dynamics of the port choice decisions:

- $R$  – the radius of neighbourhood for all the FGZs;
- $\alpha$  – the weight of the neighbours' port choice decision;
- $T$  – activation threshold.

Additionally, the model requires the calibration of the following parameters:

- the value associated to the overall costs of each port  $c_1, c_2, \dots, c_m$ ;
- the standard deviation  $\sigma$  for the whole area and  $z_i$  for each FGZ  $i$ ;

- the probability  $p$  is used in a range manner to add variability to the system's dynamics – thus, the "best candidate" will not always be chosen but instead it will have the highest probability to be chosen.

### 2.2. Model calibration

The parameters described above were calibrated with the aid of a genetic algorithm (Goldberg 1989), implemented in Lazarus programming language (Van Canneyt *et al.* 2011).

The genetic algorithm runs as follows:

- An initial population of 100 individuals is generated. Each one of these individuals represent a set of values for the PPCCA's parameters. The genotype of each individual is formed by the values of the unknown parameters (randomly selected at the beginning of the optimization process).
- The phenotype associated to each individual corresponds to the observed data reflecting the port choice dynamics, i.e., active FGZ for each port at every time step during the whole sample. For each phenotype, various indicators can be computed to measure the goodness of fit between modelled choice and actual data.
- The selection of the optimal parameters' values must be done in terms of the best fit to the observed data. Therefore, we defined a score to be optimized as follows:

$$F_0 = \sum_{i,t} f_0(i,t); \quad (4)$$

$$f_0(i,t) = \begin{cases} \gamma, & \text{if } (e_i = er_i \neq 0), \gamma \in \mathfrak{R}^+ \\ 1, & \text{if } (e_i = er_i = 0) \\ 0, & \text{otherwise} \end{cases} \quad \forall t, \quad (5)$$

- where:  $i$  represents a FGZ and  $t$  is the observed time period. The best set of values is such that  $F_0$  is maximum. In fact, each time the set of values yields a successful assignment (i.e. the simulated port choice coincides with the observed data at a given time period), the value of  $f_0(i,t)$  increases. In this equation,  $e_i$  and  $er_i$  represent the state of a FGZ in the simulation and the observed dataset respectively at each time step.  $\gamma$  is an auxiliary parameter that allows a degree of freedom to induce the search towards more adapted generations. The greater the value of  $\gamma$  the greater the relevance given to an active success, i.e., a realization in which the simulation predicts an active FGZ, which is also active in the observed sample.
- Once the algorithm has evaluated the objective function  $F_0$ , the 100 individuals are arranged in a descending order. Then, crossover operators are defined between the top 30% of the list and the rest of the population.
  - To incorporate diversity in the next generations we defined mutation operators. The gene to be mutated is chosen randomly.

- The quality of each individual is evaluated along a diversity descriptor, to guarantee a heterogeneous population at each generation. Thus, at each future generation the “best” individual of the previous generation is preserved.
- A convergence criterion is defined, in accordance with the percentage of success between simulated and observed individuals. This percentage is an input from the modeller. The algorithm stops when a pre-specified threshold is reached.

### 3. Experimental design

#### 3.1. Application instance

Only a few countries have a public archive with detailed origin-destination data for international commerce. Brazil is one of those countries. In fact, they have publicly available datasets with freight transportation data disaggregated by type of commodity up to a 6-digit customs code (<http://www.aliceweb2.mdic.gov.br>).

For this reason, we have decided to apply the modelling framework to actual data obtained from this source. Among the non-bulk commodities available in the Brazilian database we have chosen the Roll-On Roll-Off (RORO) market, which is very active in Brazil, especially for vehicles exports. There are more than 30 brands with over 65 plants located in the South, Southeast, Center-West, Northeast and North of Brazil (ANFAVEA 2016). The RORO vessels operate mainly in five Brazilian seaports (both private and public terminals).

To apply and test the developed methodology, we selected a set of ports with public terminals and common hinterland, in order to control that all the FGZ had the same available choice set. The selected ports were Paranaquá, Santos, Sao Sebastiao and Rio de Janeiro. The Port of Sao Sebastiao moves a significantly lower volume of cargo compared to its competitors; however, it was included in the choice set to test the ability of the modelling framework to respond to changes in the ports' conditions and attributes when an improvement has been made (e.g. investment in capacity).

The FGZ correspond to *Municípios*, which are administrative geographic units in which the whole country is divided. The selected sample contains records of vehicle exports through any of the above mentioned ports since 1997 until 2015. The temporal resolution is a trimester, obtaining thus a total of 74 observations in a time series.

The proportion of active FGZ at each time interval ranges between 12 and 39%, and within the last 20 periods (five years of data) the average is 29%. Figure 3 shows the percentage of active FGZ at each period in the sample, as well as a smooth curve fitted to the data.

Regarding the historical market share of each port, Figure 4 shows Port 1 with a very low share, which is explained by its accessibility from the entire area (only 4% of the FGZs are closer to this port than to its competitors) as well as its lack of attractive market conditions (infrastructure, equipment, charges, etc.). On the other hand, Port 3 is the most attractive from the location perspective; 63% of the FGZs are closer to it, followed by Port 2 with a 23% and Port 4 with 12%.

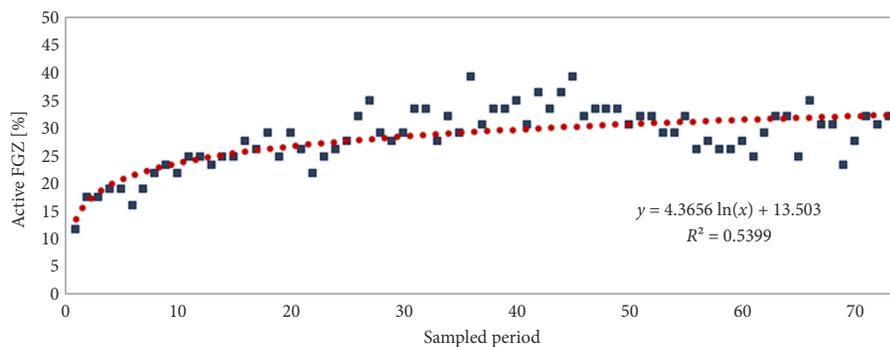


Figure 3. Percentage of active FGZ in each sampled period

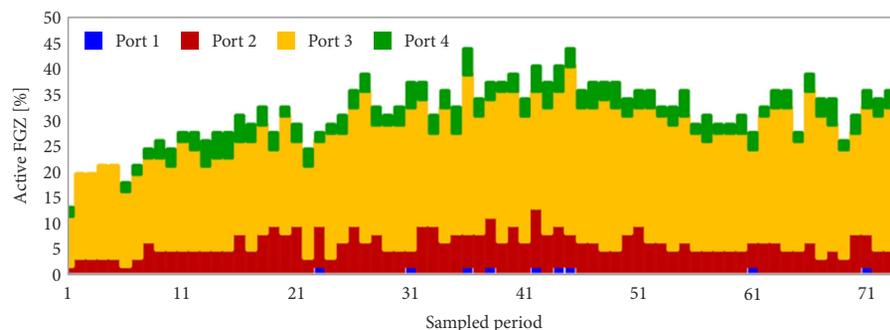


Figure 4. Historical market share of each port in the choice set

### 3.2. Calibration results

The PPCCA parameters were estimated through the genetic algorithm implementation explained in Section 2.2. The obtained parameters are:  $R = 69$  km;  $\alpha = 0.27$ ;  $T = 0.92$ . As stated in Section 2.2, the pre-specified threshold acts as stopping criterion for the algorithm. In this case that criterion was 70%, with was the highest value found for this sample. This prediction success is considered high when compared to other transportation demand models calibrated with actual data. In the case of discrete choice models, the proper way to establish their goodness of fit is by comparing the ratio of log-likelihood of an only-constants model and the log-likelihood of a given specification. That ratio is named  $\rho^2$ . The best published models reached values of  $\rho^2$  around 0.5 for the port choice problem. Table shows the goodness of fit reached by some discrete choice models applied to the port choice problem.

Table. Goodness of fit for discrete choice models applied to port choice

Author	Model	$\rho^2$
Tiwari <i>et al.</i> (2003)	MNL	0.57
Malchow and Kanafani (2004)	MNL	0.54
Veldman <i>et al.</i> (2016)	MNL	0.507

The percentage of success reached by the PPCCA was computed using a calibration parameter that assigns four times more weight to the success of an active FGZ than to an inactive one, i.e. the value of  $\gamma$  in Section 2.2 takes the value 4.

The obtained values show that the port choice in a FGZ is influenced by the choice of its  $R$ -neighbours with a weight of 27% (neighbours within a circle of 69 km of the FGZ). The remaining portion of the selection score is then explained by the distance to the port and the port's characteristics.

Note that the threshold  $T$  was calibrated obtaining a high value (0.92 and the maximum possible value was 1.0). This value is explained by the low proportion of active FGZs in the whole sample; about 30% of the FGZ remain active during the sample time span. This characteristic of the dynamic system is easily captured by a CA model

*vis-à-vis* a discrete choice model or a regression model, because those modelling approaches tend to present larger errors for agents with lower activity (or presence) within the calibration sample. The latter is due to the fact that the error term variance in discrete choice models is independent of the value of the choice probability. Thus, the error magnitude is larger when compared to alternatives with small market shares than it is to larger market shares. Consequently, those ports with significantly smaller market shares are subjected to error components that could be much larger than their systematic component.

### 3.3. Model's performance as a forecasting tool

To evaluate the model's performance as a forecasting tool we carried out several runs of the model keeping the ports' attributes fixed (i.e.  $(1-\alpha) \cdot \Phi_2^i$  remained constant in Equation (3)). Thus, the change in the FGZ's activity is explained mainly by the term  $\alpha \cdot \Phi_1^i$  in the  $\Phi^i$  function (the spatial effect of the  $R$ -neighbourhood) and the randomness added by the threshold condition for activation.

The model was executed 20 times during 80 periods (each period representing one trimester), i.e. an equivalent to 20 years simulation, in order to observe the long run behaviour of the system's evolution.

When comparing the average percentage of successful forecasts at each period (solid line in Figure 5) with the actual data for the 74 available periods (Figure 3), it can be seen that the trend in time is similar in both cases. The number of active FGZs varies between 12 and 40% in the simulated dynamics (i.e. the CA evolution) whereas the actual number of active FGZ lies between 12 and 39%, which is a good indicator of the accuracy of the calibrated CA in terms of capturing the system dynamics.

We studied the case of two ports close to each other (in terms of distance) when one of them increases its competitiveness by a 20% compared to its best competitor (i.e., the disutility of Port  $j$  was increased 20%). In the model, the latter would mean to decrease  $c_j$  by 20% in Equation (2). Figure 6 shows the market share in the baseline simulation. As can be seen, Port 1 does not capture any FGZs even though it is located close to the Port 3. The explanation for this is that Port 1 does not have sufficiently attractive conditions compared to its competitors (infrastructure, accessibility, equipment, charges, etc.).

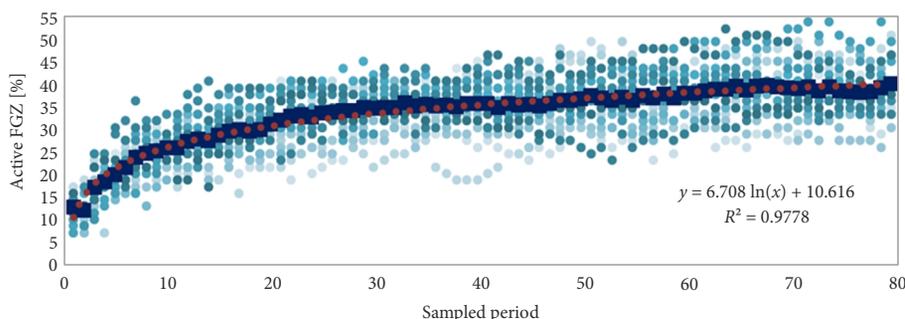


Figure 5. Forecasting of active FGZ for a 20 years simulation horizon

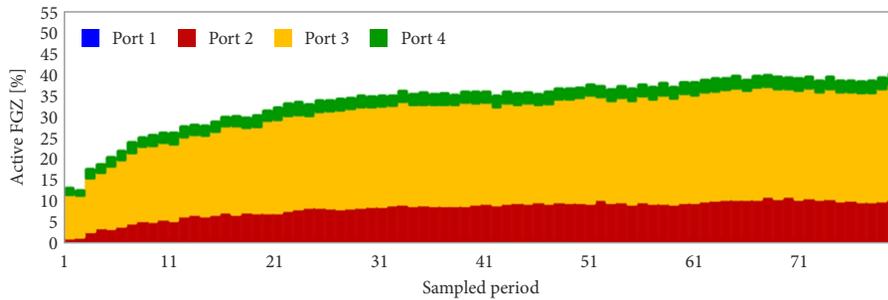


Figure 6. Market share forecast per port in a 20 years span

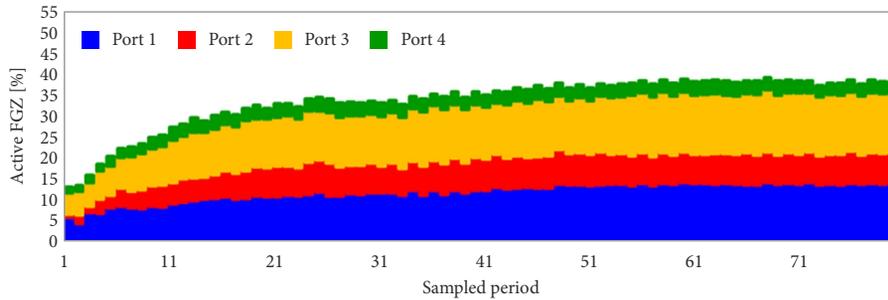


Figure 7. Market share forecast per port in a 20 years span with Port 1's attractiveness increased by 20% respect to Port 3

However, if Port 1 would manage to increase its attractiveness (or decrease its overall costs, represented by  $c_1$ ) by 20% (compared to its closest competitor Port 3), its market share would increase by 12% in the simulated horizon, as shown in Figure 7.

With this very same example is possible to estimate the average elasticity for the market share of Port 3 with respect to variation in the attractiveness of Port 1 (represented as lower overall costs). According to the obtained values, a 1 point variation in the attractiveness of Port 1 would trigger a 2.3 points of variation in the market share of Port 3. This high elasticity is due to two components. On one hand the attractiveness of the port itself, and on the other hand the spatial effect produced in the rest of the FGZs within the hinterland of both ports.

## Conclusions

In this article we developed a new modelling structure to solve the port choice problem, based on a probabilistic PPCCA that attempts to imitate the decision of multiple spatially located actors that interact with each other. These actors base their decision on various aspects such as their distance to each port, the attractiveness of each port as well as the spatial component of its own neighbourhood, in which the choice made by a FGZ's  $R$ -neighbours (those within a circle of radius  $R$ ) affects the decision of the FGZ itself.

This modelling framework applies only in cases where the decision makers interact between each other spatially through simple rules, but generating a complex system's behaviour. Thus, for cases where the decision makers are other than the exporter (e.g. importers), it would be nec-

essary to search for transition rules that capture the interaction between the various agents involved in the choice process.

The specification of the PPCCA was applied to a practical case of vehicle exports in Brazil. The parameters' calibration was done through a specially adapted genetic algorithm.

The results showed that the choice of a port by a given FGZ (i.e. the actors located within it) is strongly influenced (27%) by the choice made by its  $R$ -neighbours within a circle of 69 km from the given FGZ. The rest of the attributes influencing this decision rely on the distance from the FGZ to the ports as well as the particular characteristics of each port.

To test the forecasting ability of the proposed modelling framework, the system's behaviour was simulated by executing the PPCCA 20 runs during 80 periods (i.e. a 20 years horizon). The results showed that the simulated behaviour of the FGZs capture the dynamics of the actual data with high accuracy, both in the trends and the percentage of active FGZs at each period.

One of the practical uses of this framework is the estimation of elasticity of market shares with respect to any of the attributes involved in the decision process. In particular, the elasticity of market share with respect to the ports characteristics is a value of interest for many stakeholders. As an example, we estimated the value of such elasticity comparing two close ports: Santos and Sao Sebastiao in Brazil. In fact, if Sao Sebastiao would increase its attractiveness by 1 percentage point (over that of Santos), it would increase its market share by 2.3 points in the Brazilian vehicle export industry.

As a future line of research, the modelling framework presented in this article can be modified to represent not only the port choice but also the exported volume and the corresponding flows in the transport network. Another extension could be the inclusion of multiproduct structures. The latter would require the search of an alternative way to represent the concept of neighbourhood, capturing the interaction between exporters of various products, e.g. exporters of the same product could have more influence on their peers than closer neighbours that export a different product. In regard with the case presented in this article, the simulation could be improved with the inclusion of additional explanatory variates that may play a significant role in explaining the system's behaviour. For example, costs and times at the ports and for the deep-sea journey, could allow a more accurate description of the overall costs. However, the problem of measuring those values for each alternative and period represents a major challenge.

### Contribution

Mabel A. Leva was responsible for the conception of the problem, data selection/acquisition and interpretation as well as the results analysis and conclusions. Alejandro Leon built the PCCA simulator as well as the genetic algorithm implementation and data handling. Rodrigo A. Garrido designed the experiments, selected pertinent variables and wrote the paper. Angel Ibeas supervised the variables selection and reviewed the paper.

### Disclosure statement

The authors declare no competing financial interests.

### References

- ANFAVEA. 2016. *Brazilian Automotive Industry Yearbook*. Associação Nacional dos Fabricantes de Veículos Automotores (ANFAVEA). 152 p.
- Chopard, B.; Droz, M. 2005. *Cellular Automata Modeling of Physical Systems*. Cambridge University Press. 356 p.
- Chou, C.-C.; Chu, C.-W.; Liang, G.-S. 2008. A modified regression model for forecasting the volumes of Taiwan's import containers, *Mathematical and Computer Modelling* 47(9–10): 797–807. <https://doi.org/10.1016/j.mcm.2007.05.005>
- Garrido, R. A.; Leva, M. 2004. Port of destination and carrier selection for fruit exports: a multi-dimensional space–time multi-nomial probit model, *Transportation Research Part B: Methodological* 38(7): 657–667. <https://doi.org/10.1016/j.trb.2003.10.001>
- Garrido, R. A.; Mahmassani, H. S. 2000. Forecasting freight transportation demand with the space–time multinomial probit model, *Transportation Research Part B: Methodological* 34(5): 403–418. [https://doi.org/10.1016/S0191-2615\(99\)00032-6](https://doi.org/10.1016/S0191-2615(99)00032-6)
- Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional. 432 p.
- Goles, E.; Martínez, S. 2010. *Cellular Automata and Complex Systems*. Springer. 182 p.
- Griffeath, D.; Moore, C. 2003. *New Constructions in Cellular Automata*. Oxford University Press. 360 p.
- Hausman, J.; McFadden, D. 1984. Specification tests for the multinomial logit model, *Econometrica* 52(5): 1219–1240. <https://doi.org/10.2307/1910997>
- Lee, S.-Y.; Chang, Y.-T.; Lee, P. T.-W. 2010. Container port selection factors: heterogeneity among major market players, *Journal of International Logistics and Trade* 8(2): 73–90. <https://doi.org/10.24006/jilt.2010.8.2.004>
- Magala, M.; Sammons, A. 2008. A new approach to port choice modelling, *Maritime Economics & Logistics* 10(1–2): 9–34. <https://doi.org/10.1057/palgrave.mel.9100189>
- Malchow, M.; Kanafani, A. 2004. A disaggregate analysis of port selection, *Transportation Research Part E: Logistics and Transportation Review* 40(4): 317–337. <https://doi.org/10.1016/j.tre.2003.05.001>
- Malchow, M.; Kanafani, A. 2001. A disaggregate analysis of factors influencing port selection, *Maritime Policy & Management: The Flagship Journal of International Shipping and Port Research* 28(3): 265–277. <https://doi.org/10.1080/03088830110060840>
- Manski, C. F.; McFadden, D. 1981. *Structural Analysis of Discrete Data with Econometric Applications*. The MIT Press. 504 p.
- Moya, J. M.; Feo Valero, M. 2017. Port choice in container market: a literature review, *Transport Reviews* 37(3): 300–321. <https://doi.org/10.1080/01441647.2016.1231233>
- Papola, A. 2004. Some developments on the cross-nested logit model, *Transportation Research Part B: Methodological* 38(9): 833–851. <https://doi.org/10.1016/j.trb.2003.11.001>
- Song, D.-W.; Yeo, K.-T. 2004. A competitive analysis of Chinese container ports using the analytic hierarchy process, *Maritime Economics & Logistics* 6(1): 34–52. <https://doi.org/10.1057/palgrave.mel.9100096>
- Tiwari, P.; Itoh, H.; Doi, M. 2003. Shippers' port and carrier selection behaviour in China: a discrete choice analysis, *Maritime Economics & Logistics* 5(1): 23–39. <https://doi.org/10.1057/palgrave.mel.9100062>
- Tongzon, J. L. 2009. Port choice and freight forwarders, *Transportation Research Part E: Logistics and Transportation Review* 45(1): 186–195. <https://doi.org/10.1016/j.tre.2008.02.004>
- Tongzon, J. L.; Sawant, L. 2007. Port choice in a competitive environment: from the shipping lines' perspective, *Applied Economics* 39(4): 477–492. <https://doi.org/10.1080/00036840500438871>
- Van Canneyt, M.; Gärtner, M.; Heinig, S.; De Carvalho, F. M.; Ouedraogo, I. 2011. *Lazarus: The Complete Guide*. Blaise Pascal Magazine. 736 p.
- Veldman, S.; Garcia-Alonso, L.; Liu, M. 2016. Testing port choice models using physical and monetary data: a comparative case study for the Spanish container trades, *Maritime Policy & Management: The Flagship Journal of International Shipping and Port Research* 43(4): 495–508. <https://doi.org/10.1080/03088839.2015.1099754>
- Veldman, S.; Garcia-Alonso, L.; Vallejo-Pinto, J. Á. 2013. A port choice model with logit models: a case study for the Spanish container trade, *International Journal of Shipping and Transport Logistics* 5(4/5): 373–389. <https://doi.org/10.1504/IJSTL.2013.055277>
- Yeo, G.-T.; Ng, A. K. Y.; Lee, P. T.-W.; Yang, Z. 2014. Modelling port choice in an uncertain environment, *Maritime Policy & Management: The Flagship Journal of International Shipping and Port Research* 41(3): 251–267. <https://doi.org/10.1080/03088839.2013.839515>