

Mutual information analysis of the factors influencing port throughput

Factors
influencing
port
throughput

129

Majid Eskafi

Faculty of Civil and Environmental Engineering, School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland

Milad Kowsari

Faculty of Civil and Environmental Engineering, School of Engineering and Natural Sciences, Earthquake Engineering Research Centre, University of Iceland, Reykjavik, Iceland

Ali Dastgheib

Department of Coastal and Urban Risk and Resilience, IHE Delft Institute for Water Education, Delft, The Netherlands

Gudmundur F. Ulfarsson

Faculty of Civil and Environmental Engineering, School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland

Poonam Taneja

Faculty of Civil Engineering and Geosciences, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands, and

Ragnheidur I. Thorarinsdottir

Faculty of Civil and Environmental Engineering, School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland

Received 13 May 2020
Revised 30 June 2020
14 September 2020
20 October 2020
25 November 2020
30 November 2020
Accepted 1 December 2020

Abstract

Purpose – Port throughput analysis is a challenging task, as it consists of intertwined interactions between a variety of cargos and numerous influencing factors. This study aims to propose a quantitative method to facilitate port throughput analysis by identification of important cargos and key macroeconomic variables.

Design/methodology/approach – Mutual information is applied to measure the linear and nonlinear correlation among variables. The method gives a unique measure of dependence between two variables by quantifying the amount of information held in one variable through another variable.

Findings – This study uses the mutual information to the Port of Isafjordur in Iceland to underpin the port throughput analysis. The results show that marine products are the main export cargo, whereas most imports

© Pacific Star Group Education Foundation. Licensed re-use rights only.

The authors are grateful to Professor Gunnar Stefansson in the Faculty of Industrial Engineering, Mechanical Engineering and Computer Science at the University of Iceland for his valuable remarks on this study. The authors thank the Icelandic Road and Coastal Administration, Statistics Iceland, Eimskip hf. and Samskip hf. companies for providing data. The authors thank the two anonymous referees for their review and fruitful comments. This research was supported in part by the University of Iceland Research Fund, the Municipality of Isafjordur and the Icelandic Road and Coastal Administration Research Fund.



are fuel oil, industrial materials and marine product. The aggregation of these cargos, handled in the port, meaningfully determines the non-containerized port throughput. The relation between non-containerized export and the national gross domestic product (GDP) is relatively high. However, non-containerized import is mostly related to the world GDP. The non-containerized throughput shows a strong relation to the national GDP. Furthermore, the results reveal that the volume of national export trade is the key influencing macroeconomic variable to the containerized throughput.

Originality/value – Application of the mutual information in port throughput analysis effectively reduces epistemic uncertainty in the identification of important cargos and key influencing macroeconomic variables. Thus, it increases the reliability of the port throughput forecast.

Keywords Iceland, Macroeconomics, Mutual information, Port throughput

Paper type Research paper

1. Introduction

Demand projection and selection of promising markets play an important role in the port planning process (Geweke and Whiteman, 2006). Identification of the key cargo for a port characterizes the strategy and direction of port planning projects and aids the preliminary design of basic infrastructure (Chen *et al.*, 2016). Financial viability and infrastructure-based investments should be supported by potential (cargo) demands (De Langen *et al.*, 2012). Appropriate investment in port capacity, based on the promising cargos, helps to win market share and strengthen the competitive position of the port (Taneja *et al.*, 2010). On the other hand, inaccurate statements about the likely course of demand lead to an improper development plan (Peng and Chu, 2009).

Port throughput analysis prior to resource allocation decisions in port capacity planning and development is critical. The information about the flow will have a substantial impact on port operations and planning and the use of assets (Milenković *et al.*, 2019). Port throughput analysis depends on the flow of different types of cargo and numerous macroeconomic variables. However, in the dynamic and complex nature of a port system, identification of the demand has remained with multiple uncertainties including epistemic uncertainty (Taneja, 2013).

Epistemic uncertainties in demand analysis depend upon the degree to which the information pertaining to the system is available. Ping and Fei (2013) expressed that port throughput is affected by numerous variables including macroeconomic variables. Owing to complex nonlinear relations between port throughput and macroeconomic variables, a single linear model (Chen *et al.*, 2016), or application of traditional regression methods, may result in inaccurate analysis performance (Gökkuş *et al.*, 2017). Hui *et al.* (2004) stated that the classical regression methods are valid if the data series are stationary and without a time trend. Chou *et al.* (2008) emphasized the nonstationary relation between the cargo flow and macroeconomic variables.

Therefore, a comprehensive study on the identification of main cargos and macroeconomic variables, in port throughput analysis is important. For this purpose, the correlation coefficient, as the most known measure of dependence between two random variables, can be used. However, its application has been criticized, as it is only able to detect linear dependencies (Sagar and Guevara, 2005). To overcome this problem, mutual information can be used to capture nonlinear dependency among variables.

In this study, mutual information is applied to evaluate the dependencies among different types of cargo and port throughput and identifies the prominent cargos that would heuristically describe the port throughput. In information theory, mutual information measures the amount of information that one variable contains about the other. Mutual information quantifies the statistical dependence between two random variables. Thus, it provides a better criterion than the autocorrelation function, which only measures linear

dependence (Fraser and Swinney, 1986). In contrast to the linear correlation coefficient, it is also sensitive to dependencies that do not manifest themselves in the covariance (Kraskov *et al.*, 2004).

Mutual information has been used in previous maritime research, albeit not in the context of port throughput. For instance, Wu *et al.* (2020) used mutual information to reduce needs for expert judgment in the identification of input variables for the Bayesian network for consequence estimation of navigational accidents. Yang *et al.* (2018) applied mutual information to facilitate the recognition of insignificant variables that should be excluded from a Bayesian network. Hänninen and Kujala (2014) used mutual information to reduce uncertainty by identification of influential variables in a Bayesian model for ship accident involvement. Furthermore, Hänninen (2014) discussed the advantage of applying mutual information for determining the uncertainty of variables' dependency in a maritime safety model.

Hence, this study applied mutual information in the analysis of port throughput influencing factors. The explanatory power of mutual information delivers a more systematic way of analysis for identification of the most and least important cargoes and influencing macroeconomic variables based on an analysis-oriented approach of data, rather than arbitrary judgment (that may have biases), which is an advancement over present practice. This is especially the case where historic data of the port are not sufficiently indicative. Therefore, it contributes to improving the port throughput analysis and consequent forecast, as it reduces the epistemic uncertainties associated with the identification of main cargoes and key macroeconomic variables. It also offers a more transparent, simpler and easier way to interpret the result of analysis. Mutual information is applied to the multipurpose Port of Isafjordur in Iceland. The presented method can be used for other cases.

2. Influencing factors on port throughput

The long technical lifetime of port infrastructure, huge capital investments and a long payback period make port planning a challenging task (Taneja *et al.*, 2010). Port capacity should satisfy the demand of multiple stakeholders with various objectives (Eskafi *et al.*, 2020). Overcapacity leads to the lack of cost-effectiveness in port planning, and capacity shortage results in loss of competitive position of ports (Jugovic *et al.*, 2011). Before port capacity planning, port throughput analysis needs to be conducted, and for that, identification of the main cargoes and influencing factors is essential.

Luo and Grigalunas (2003) pointed out that (cargo) demand analysis is challenging, as it is influenced by many factors. Demand analysis may have an uncertain outcome, because of difficulties in identification of the main cargo, determination of exogenous and endogenous variables and their complex causal relations with cargoes (Taneja, 2013). Cargo flow is volatile over time and affected by the temporal demand of salient stakeholders in the projected lifetime of a port (Eskafi *et al.*, 2019). Van Dorsser *et al.* (2012) put forward that demand projection is sensitive to trend breaches and attaching much investment to one single cargo over a projected lifetime is not advocated.

As economic development is an important driver of maritime trade, there should be an interrelation between port throughput and macroeconomic variables (De Langen *et al.*, 2012). Jugovic *et al.* (2011) recognized gross domestic product (GDP) as a known macroeconomic variable that is the main reference of the elasticity factor for cargo flow in a port. Ping and Fei (2013) pointed out that the regional economy (i.e. regional GDP) considerably affects the port throughput. Van Dorsser *et al.* (2012) presented the correlation between GDP and port throughput by using a macroeconomic model. Moreover, Gökkuş *et al.* (2017) expressed that there is a strong correlation between GDP and the trade volume of a country.

Cargo flow is increasingly intertwined with the population, trade, global economic activity (Taneja *et al.*, 2012), fuel and energy prices (Van Dorsser *et al.*, 2012), competitive position, market share (Meersman *et al.*, 2003), country logistics system and supply chain, technology evolution and government policies (Günther and Kim, 2005). Frankel(1987) pointed out that, before port throughput analysis, deep knowledge of the hinterland is fundamental. Jugovic *et al.* (2011) echoed that cargo flow analysis should begin with the characterization of hinterland that gravitates to the port under study.

According to the cited literature, many factors should be considered when port throughput is analyzed. However, it is neither necessary nor possible (due to limited data) to take all variables into account. When the number of exogenous inputs is huge, numerous problems can occur related to high dimensionality and multicollinearity (Fuentes *et al.*, 2015), or the analysis is sensitive to false reduction (Bankes, 1993).

JICA(1994) used population and GDP variables to analyze and forecast a container port's throughput using a regression model. However, Gosasang *et al.* (2011) criticized them because of their use of inadequate variables in the analysis. Later, De Langen *et al.* (2012) and Van Dorsser *et al.* (2012) discussed using only GDP for port throughput analysis and forecast. Chou *et al.* (2008) used several macroeconomic variables to forecast port throughput but did not discuss the reason for the selection of macroeconomic variables. Gosasang *et al.* (2018) used multiple macroeconomic variables for developing a container throughput forecast. However, their work neither discussed the relation of the macroeconomic variables with port throughput nor showed which variables may have a higher influence on port throughput. Moreover, Dragan *et al.* (2020) used macroeconomic variables that presumably influence cargo throughput to forecast the throughput of different types of cargo. However, they did not discuss reasons for the selection of certain macroeconomic variables. Milenković *et al.* (2019) stressed the importance of determining an optimal set of input variables for developing a container throughput forecast model.

The complexity of port throughput analysis calls for a method to screen the main cargos and influencing macroeconomic variables. Mutual information analysis has been a useful method for data analysis. However, the application of mutual information has not been used in analysis of the factors influencing port throughput, which is the novelty of the present study.

3. Mutual information analysis

Mutual information is an important concept in information theory to handle uncertainties and abstraction of the notion of information. It measures the level of correlation among variables and then determines their dependency on each other by quantifying the amount of information held in a variable through another variable. In general, information refers to the ease of predictability of unknown outcomes provided by one probability distribution relative to another probability distribution (Soofi *et al.*, 2010). Mutual information gives a unique measure of dependence between the two variables, which is also connected to the concept of entropy and Kullback–Leibler divergence.

The advantages of mutual information in variable selection problems compared to the other criteria such as coefficient-based methods (Hall, 1999), RELIEF and RELIEF-F (Robnik-Šikonja and Kononenko, 2003), have been shown in previous studies (Zeng *et al.*, 2014). Kwak and Choi(2002) pointed out that in variable selection, mutual information can effectively eliminate redundant variables with relatively low computational effort. Mutual information has advantages to Pearson correlation in variable selection, as the former identifies the linear and nonlinear dependence and independence of variables, but the latter only recognizes the linear dependence of variables (Li, 1990; Steuer *et al.*, 2002). Peng *et al.* (2005) expressed mutual information as an appropriate variable selection method compared

to the max-dependency method, as the latter method requires much data and has relatively slow computational speed. Pethel and Hahs (2014) stated that chi-squared tests for variable selection are valid in the asymptotic limit of infinite data, whereas mutual information does not have the limitation of chi-squared tests. Zuur *et al.* (2003) pointed out the advantage of mutual information to a Markov Chain Monte Carlo method for parameter selection, as the latter method requires a large data set that imposes restrictions in its application. Li *et al.* (2009) stated that mutual information achieves better test performance and faster computation efficiency than Wrapper methods. Furthermore, Wrapper methods are faced with over-fitting problems. Also, they showed that mutual information outperforms Wrapper methods in variable selection when the number of variables is large. The merit of mutual information is to avoid unnecessary work and reduce the need for expert judgment, provide a straightforward, fast, with relatively low computational complexity, and cost-effective approach to recognize the influencing macroeconomic variables on port throughput (Wu *et al.*, 2020; Yang *et al.*, 2018). Mutual information also offers a useful visual tool for a better understanding of the dependencies among variables (Li *et al.*, 2009).

For a pair of random variables (X, Y) with marginal probability distributions of $\mu_x(x)$ and $\mu_y(y)$, mutual information uses the Kullback–Leibler measure to determine the distance between the joint probability distribution, $\mu(x, y)$ and the distribution associated with the case of complete independence (i.e. $\mu_x(x)\mu_y(y)$) and is expressed as (Kraskov *et al.*, 2004):

$$I(X, Y) = \iint \mu(x, y) \log \frac{\mu(x, y)}{\mu_x(x)\mu_y(y)} dx dy \quad (1)$$

Furthermore, mutual information is related to the concept of information entropy that was introduced by Shannon(1948) and quantifies how informative a random variable (X) with possible outcomes (x_i) , each with probability $p(x)$, could be:

$$H(X) = - \int_{x \in X} p(x) \log_2 p(x) dx \quad (2)$$

where the base 2 logarithm is corresponding to the unit of information measured in “bits” (Shannon, 1948). Thus, mutual information can be obtained by:

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X | Y) \\ &= H(Y) - H(Y | X) \end{aligned} \quad (3)$$

where $H(X)$ and $H(Y)$ are the entropy of random variables X and Y , respectively; $H(X, Y)$ is their joint entropy; and $H(X|Y)$ and $H(Y|X)$ are their conditional entropy and calculated as:

$$H(X | Y) = - \iint \mu(x, y) \log \mu(x | y) dx dy \quad (4)$$

where $\mu(x, y)$ is the joint probability distribution. The conditional entropy $H(X|Y)$ is the amount of uncertainty left in X when knowing Y . Thus, from these equations, the $I(X, Y)$ is interpreted as the reduction in the uncertainty of the random variable X by the knowledge of another random Y (Maes *et al.*, 1997). Mutual information illustrates the distributions of the information measures in terms of interdependency among variables. Mutual information takes the value of zero if and only if the two random variables are statistically independent and when the two variables are identical their mutual information reaches the maximum. To

calculate mutual information among variables, the equations can be coded in a computer programming language. In this study, MATLAB is used to code the equations and calculate mutual information for further analysis.

To evaluate the correlation between the port throughput and macroeconomic variables, at first, the dependency of the port throughput on different cargos is investigated. It helps to recognize the main cargo and eliminate cargos that are occasionally handled in (especially multipurpose) ports. The recognition of the main cargo reduces biases (e.g. a large amount of a specific cargo that is handled only a few times for a particular purpose) in the collection of data. On the other hand, including many cargos increases the complexity of the analysis and (epistemic) uncertainty in the results (Van Dorsser *et al.*, 2012).

4. Study area and data

The multipurpose Port of Isafjordur is the leading cargo port and hub in the Westfjords region of Iceland (Figure 1). This port plays a significant role in the logistic chain of the country. It is well connected to the hinterland in terms of coastal shipping and road transportation. It has a strategic location with short sailing time to the open sea and enough services for different types of vessels. Fisheries and industrial aquaculture are the core businesses of the region, where these activities are thriving (Icelandic Directorate of Fisheries, 2019), which increases cargo and container handling at the port. The port provides services to industries all over the country, and thus, its hinterland is the whole country.

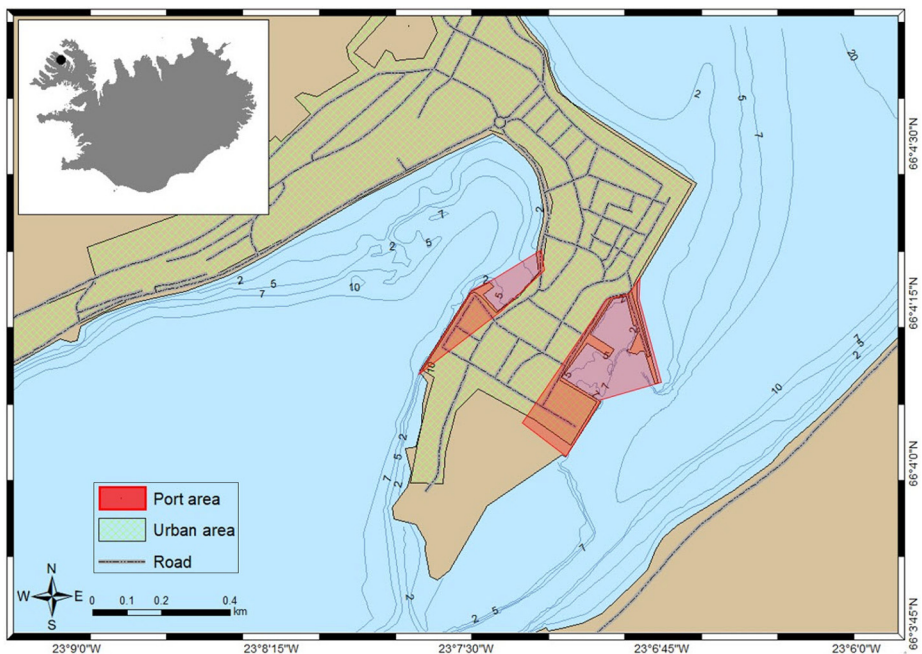


Figure 1.
The multi-purpose
Port of Isafjordur

Note: The location of the study area is marked with a block dot on the inset map of Iceland at the top left

The main functions of the port are:

- transfer and storage of containerized and non-containerized cargo;
- industrial value-added activities, including marine productions; and
- recreational activities, including servicing expedition and cruise ships, sailing boats and water sport activities.

Among the aforementioned factors (in Section 2) that influence port throughput, this study excludes those that are unavailable, cannot be accurately predicted (e.g. new technologies) and those that their influence cannot be quantified from observation of the past (e.g. growth in the hinterland) (Taneja *et al.*, 2010). Thus, six macroeconomic variables that are available and published by the relevant authorities are used for analysis. These are the national GDP, the average yearly Consumer Price Index (CPI), the world GDP, the volume of national export trade, the volume of national import trade and the national population. These variables were also used in previous studies (Gökkuş *et al.*, 2017). Table 1 describes the variables used in this study.

In case of the availability of more macroeconomic variables, these can be added. One of the scientific contributions of the present study is to identify which macroeconomic variable among the variables would have the highest influence on the port throughput. The identification of the most influential macroeconomic variables increases the accuracy of the port throughput forecast. Figure 2 shows the schematic application of mutual information in port throughput analysis.

Yet, there are many unknown variables that can affect the port throughput. These unknown unknowns (Walker *et al.*, 2013), and black swans (Smil, 2012), can considerably influence port throughput as their intensity and frequency cannot be meaningfully addressed. In testimony of this, the coronavirus disease (COVID-19) pandemic in 2020 is significantly affecting maritime sectors, as cruise ship calls have slumped, and there is a decline in port throughput. Considering these variables in port throughput analysis is out of the scope of this study.

In this study, non-containerized cargos are categorized as fuel oil, road construction and maintenance materials, fertilizer and fish feed, marine products and industrial materials. Small (in terms of quantity) cargos are considered as other general cargo. These cargos are measured in tonnes. The non-containerized port throughput data are garnered between 1990 and 2016. Containerized cargo (cargo that is transported in a [refrigerated] container) used in this study is based on a twenty-foot equivalent unit (TEU). The annual containerized port throughput and the macroeconomic variables are collected between 1990 and 2019.

In this study, the mutual information value between containerized port throughput and the macroeconomic variables is measured separately. The main reason is that containers have been attractive and promising to transport cargo. Determining the relation between containerized port throughput and the macroeconomic variables facilitates decision-making for strategic capacity development by the Port Authority.

The port throughput analysis does not include the recreational activities. Limited data, differing data sources and inconsistencies in terms of the accuracy and uniformity of recording variables and cargos could affect the results of this study. Thus, uniform collection and management of statistical data is recommended.

5. Results and discussion

To evaluate the correlation of (non-)containerized port throughput and macroeconomic variables, mutual information was conducted using equations (1)–(4).

Table 1.
List of variables for
analysis of port
throughput

Acronym	Variable	Unit	Mean	SD	Minimum	Maximum
FOL	Fuel oil (gasoline, [marine] diesel oil)	Tonnes (T)	95.4	38.9	42.4	169.5
RCM	Road construction and maintenance materials (asphalt, salt, cement, etc.)	Tonnes (T)	45.2	103.9	0	540.5
FFF	Fertilizer and fish feed	Tonnes (T)	34,290.7	78,719.6	0	378,900.0
MAP	Marine products (fresh, frozen, (un)processed, wild and farmed fish, shrimp, etc.)	Tonnes (T)	167.0	93.7	19.7	346.0
INM	Industrial materials (fishing and maritime equipment, scrap, etc.)	Tonnes (T)	108.9	129.3	0	371.4
SGC	Small general cargo (construction material, etc.)	Tonnes (T)	211.2	409.8	0	1,357.8
-	Non-container port throughput	Tonnes (T)	117.8	63.8	30.2	279.5
-	Container and refrigerated container port throughput	Twenty-foot equivalent unit (TEU)	122.8	65.6	65.3	331.2
NGDP	National gross domestic product	Icelandic Krona (ISK)	95.4	25.8	60.8	141.7
ACPI	Average yearly consumer price index	No unit	115.2	45.3	59.6	191.9
WGDP	World gross domestic product	Billion USD	107.1	45.5	47.6	186.2
VNET	Volume of national export trade	Tonnes (T)	109.8	48.1	51.6	203.1
VNIT	Volume of national import trade	Tonnes (T)	74.5	26.9	36.4	124.9
NPOP	National population	Number	101.3	10.1	86.4	121.6

Notes: The values for the cargos (i.e. FOL, RCM, FFF, MAP, INM and SGC) are based on the summation of their export and import. The variables are indexed to their value in the year 2005

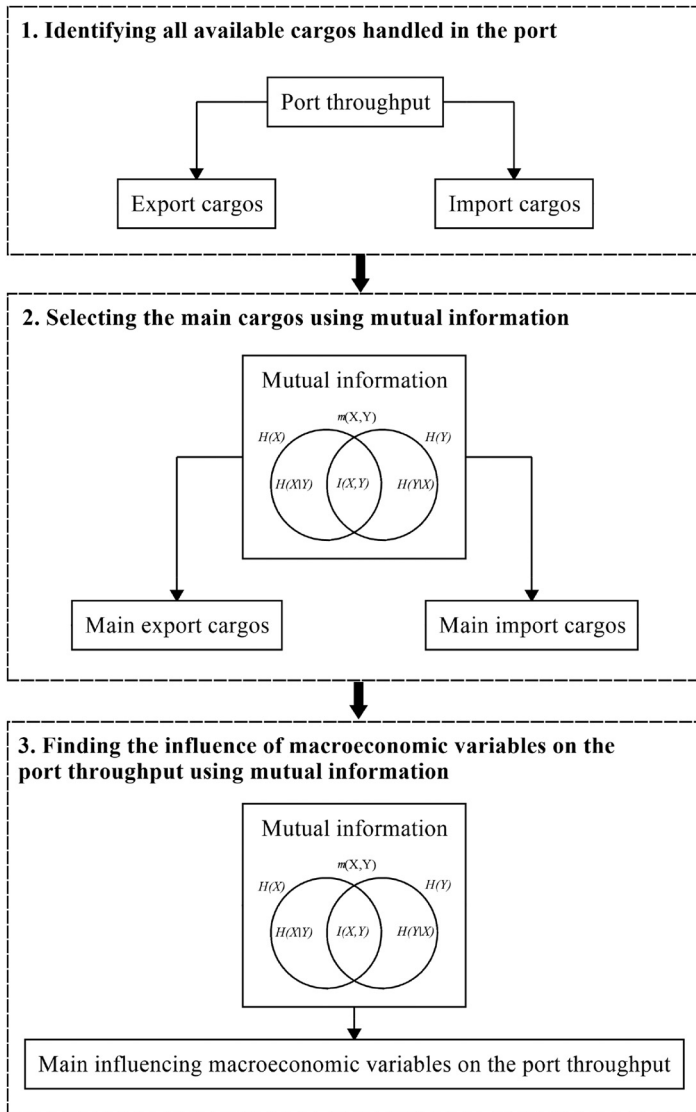


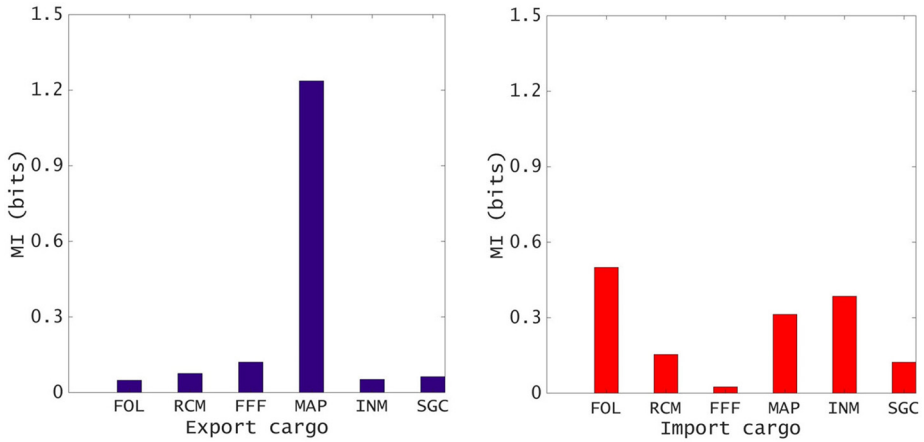
Figure 2. A schematic representation of the application of the mutual information in port throughput analysis

5.1 Identification of main cargoes

The results of the mutual information values of the handled non-containerized cargoes at the port with export and import are depicted in Figure 3. The values of mutual information of non-containerized cargoes with export and import are shown in Table 2.

From Figure 3, marine products cargo is the main pillar of export. This is because the core businesses in the region are fisheries and aquaculture. Therefore, these activities can significantly influence the export. As seen in Figure 3, the mutual information values of the other cargoes in export are considerably lower than the marine products. Fertilizer and fish

Figure 3. The mutual information (MI) values of the non-containerized cargos for export (left) and import at the port (right)



Notes: The cargos are fuel oil (FOL), road construction and maintenance materials (RCM), fertilizers and fish (FFF), marine products (MAP) industrial materials (INM), small general cargo (SGC)

Table 2. Mutual information values of non-containerized cargos with export and import

Acronym	Cargo flow	Mutual information	
		Export	Import
FOL	Fuel oil	0.044	0.499
RCM	Road construction and maintenance materials	0.085	0.156
FFF	Fertilizer and fish feed	0.141	0.018
MAP	Marine products	1.240	0.317
INM	Industrial materials	0.047	0.342
SGC	Small general cargo	0.051	0.134

feed and general cargo have relatively low mutual information values. These cargos with low mutual information values are mainly fisheries-related activities, for instance, processed products, industrial equipment, etc. The industrial materials cargo and fuel oil have the lowest mutual information values in the export.

For the import cargos, the mutual information values for several cargos are relatively large. These cargos, however, have small differences in their mutual information values. The results show that fuel oil has the largest mutual information value. The industrial materials cargo has a relatively lower value but higher than the marine products. These cargos could be considered as the raw materials and the main needs of the industries for their activities. The contribution of other cargos to the import is relatively weak.

As is shown in Figure 3, the main export cargo is marine products, whereas the aggregation of fuel oil, marine products and industrial materials can be indicated as credible import cargos. Therefore, it can be inferred that the non-containerized port throughput has a considerable dependency on marine products, fuel oil and industrial materials. This result is used to calculate the correlation between the non-containerized port throughput and macroeconomic variables.

5.2 Port throughput and macroeconomic variables data

Figure 4 gives the development of the port throughput and macroeconomic variables. To keep the confidentiality of (non-)containerized data and allow for comparison between the port throughput and the macroeconomic variables, the historical data are indexed to the year 2005.

As illustrated in Figure 4, the macroeconomic variables have generally been growing, though a trend breach can be observed at the world economic downturn in 2008. Afterward, containerized port throughput growth slowed until 2012, whereas non-containerized port throughput dropped in the same period. Then, the containerized port throughput increased again. Over the period, fluctuations are attributed to economic activity and trade. The containerized port throughput and macroeconomic variables indicate increasing trends. However, non-containerized port throughput shows a decreasing trend until 2012. The main reason is continuous growth in the use of containers for transporting goods. The significant jump in the port throughput after 2012 could be because of the rapid growth in aquaculture, especially the salmon industry in Iceland. The fast-growing aquaculture drives the growth of relevant activities including marine production and industrial equipment manufacturing. Furthermore, another shipping company (additional to the first one) started calling the port from 2013.

5.3 Identification of the relation between port throughput and macroeconomic variables

Based on the identified main cargos, Figure 5 shows the results of the mutual information values among non-containerized export, import, port throughput and macroeconomic variables. This means that export includes marine products cargo, import consists of fuel oil, marine products and industrial materials cargos, and port throughput represents the aggregation of export and import.

The results indicate that the non-containerized export is mostly related to the national GDP. As identified, marine products cargo is the major export, and this cargo has a high impact on the GDP of the country (Statistics Iceland Office, 2019). GDP has been determined

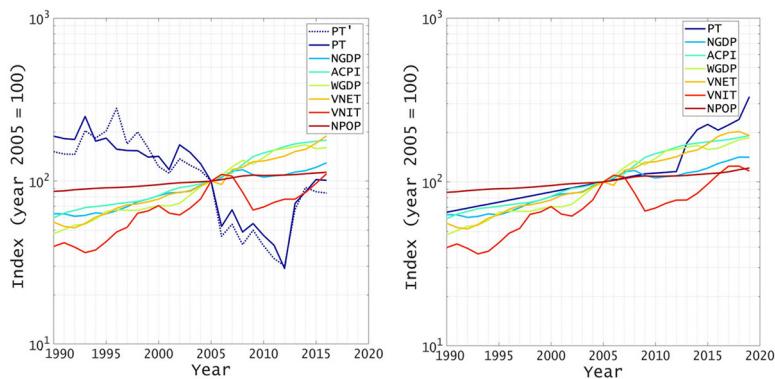


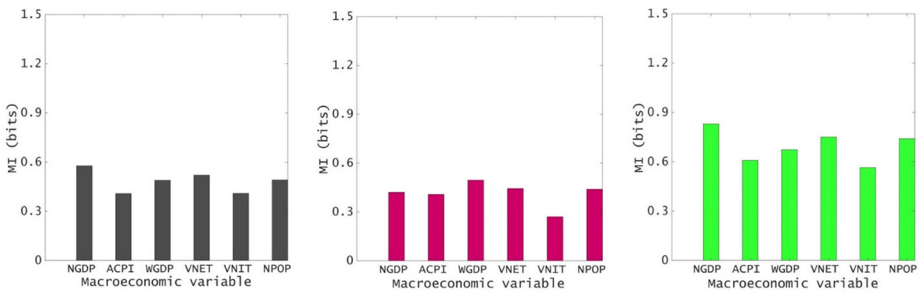
Figure 4. Historical development of the port throughput (left: non-containerized cargo, right: containerized cargo) and the macroeconomic variables

Notes: The variable are the port throughput (PT' before and PT after application of mutual information), and macroeconomic variables including the national GDP (NGDP), the average yearly CPI (ACPI), the world GDP (WGDP), the volume of national export trade (VNIT), the volume of national import trade (VNIT) and the national population (NPOP)

as one of the main influencing variables on port throughput (Van Dorsser *et al.*, 2012). On the other hand, the relation between the non-containerized import and world GDP is the strongest. This result indicates the influence of the economy (or GDP) of the country (the hinterland) and the world on the port activities (Jugovic *et al.*, 2011). There is a relatively high relation between the non-containerized port throughput and the national GDP. The non-containerized port throughput and the national GDP are intercorrelated, as they are affected by import (national consumption) and the export (productivity) of goods (Van Dorsser *et al.*, 2012). The result further shows the relation of the volume of national export trade with the non-containerized port throughput, as also mentioned by Gökkuş *et al.* (2017). The volume of national import trade has the lowest correlation with the non-containerized port throughput. The results of mutual information among containerized export, import, port throughput and macroeconomic variables are shown in Figure 6.

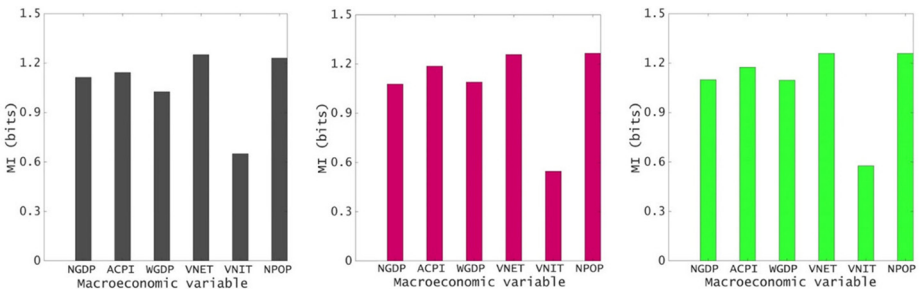
The results show similar patterns for the containerized export, import and throughput concerning the macroeconomic variables. The relation between the containerized port throughput and the volume of national export trade is the highest. This is because containerized cargo can be transported efficiently over long distances and easily transferred between modes of transport. The variables with slightly lower mutual information values

Figure 5. The mutual information (MI) values between non-containerized export, import, port throughput and macroeconomic variables (left to right, respectively)



Notes: The macroeconomic variables are the national GDP (NGDP), the average yearly CPI (ACPI), the world GDP (WGDP), the volume of national export trade (VNET), the volume of national import trade (VNIT) and the national population (NPOP)

Figure 6. The mutual information (MI) value between containerized export, import, port throughput and macroeconomic variables (left to right, respectively)



Notes: The macroeconomic variables are the national GDP (NGDP), the average yearly CPI (ACPI), the world GDP (WGDP), the volume of national export trade (VNET), the volume of national import trade (VNIT) and the national population (NPOP)

are the national population, the average yearly CPI and the national GDP. This is in line with literature where population, national GDP and inflation rate were used as the key influencing variables in container throughput forecast (Gökkuş *et al.*, 2017). The average yearly CPI partly determines the annual value of the national GDP (Gosasang *et al.*, 2011), and national GDP is a good indicator of container port throughput (Van Dorsser *et al.*, 2012). Population growth stimulates trade flow because of increased labor force and economic improvements (Hanushek and Kimko, 2000). Table 3 gives the values of mutual information of macroeconomic variables with (non-)containerized export, import and port throughput.

As depicted in Figures 5 and 6, the volume of national import trade (VNIT) has the lowest mutual information with the port throughput. One of the main reasons for this low value is that the majority of import to the country is to the Port in Reykjavik, the capital of Iceland, not to the Port of Isafjordur.

Fuentes *et al.* (2015) stressed that including more inputs to the forecast models does not necessarily lead to better results with lower uncertainty. Hence, in the case of the throughput forecast for the port in this study, the VNIT can be excluded from the input to forecast models. Nevertheless, the volume of import trade (Ping and Fei, 2013) and the value of import trade (Hui *et al.*, 2004; Gökkuş *et al.*, 2017; Gosasang *et al.*, 2018) have been included in port throughput forecast in previous studies. The discussion on the selection of this macroeconomic variable (i.e. VNIT) in these studies was not rigorous, and the focuses on describing the relation are qualitative relations based on expert judgment and literature review. Therefore, the present study proposes that selection of influencing macroeconomic variables for forecast models only based on expert judgment and literature review may not be sufficient.

Qualitative (or expert-judgment-based) evaluation of influencing factors on port throughput may be time consuming, laborious, include biases (Wu *et al.*, 2020) and rely on incomplete and subjective knowledge, which is conditional on the background and experience of experts (Hänninen, 2014). Furthermore, experts are not always available to determine the influencing variables (Montewka *et al.*, 2014). Application of mutual information in port throughput analysis reduced the need for expert judgment and provided informative diagrams to recognize the influencing macroeconomic variables on port throughput (Li *et al.*, 2009; Wu *et al.*, 2020).

Selection of relevant variables from a large feasible set of input data improves forecast efficiency (Fuentes *et al.*, 2015). Thus, the application of mutual information for (macroeconomic) variable selection in port throughput forecast models increases the accuracy of models and reduces the uncertainty of the results (Hänninen, 2014; Hänninen and Kujala, 2014; Yang *et al.*, 2018). The premise of mutual information is to deliver a unique measure of dependence of factors influencing port throughput. The theoretical contribution

Macroeconomic variable	Mutual information					
	Non-containerized			Containerized		
	Export	Import	Port throughput	Export	Import	Port throughput
National GDP (NGDP)	0.596	0.405	0.804	1.123	1.088	1.099
Average yearly CPI (ACPI)	0.426	0.398	0.609	1.159	1.196	1.174
World GDP (WGDP)	0.496	0.486	0.684	1.070	1.089	1.096
Volume of national export trade (VNET)	0.537	0.440	0.759	1.265	1.257	1.258
Volume of national import trade (VNIT)	0.430	0.273	0.582	0.642	0.546	0.576
National population (NPOP)	0.491	0.437	0.746	1.238	1.258	1.258

Table 3.
Mutual information
values of
macroeconomic
variables with
export, import and
port throughput

of the application of mutual information in port throughput is to reduce the epistemic uncertainty in port throughput forecast. The managerial contribution of the application of mutual information in port throughput analysis is to provide a better insight into the major cargos handled in a port and a robust estimation of macroeconomic variables' influence on the port throughput.

6. Conclusion

Port throughput analysis provides valuable and fundamental inputs for port capacity planning and development. However, port throughput analysis is a complex task, as it is interwoven with a variety of cargos handled in the port, which are in turn influenced by numerous factors including macroeconomic variables. The analysis requires the selection of prominent cargos that meaningfully contribute to the port throughput. Furthermore, the analysis necessitates investigating the relation of port throughput with macroeconomic variables.

This study used mutual information analysis as a quantitative method to measure the linear and nonlinear correlation among variables. The presented method was able to indicate the important cargos handled in the port. Moreover, the method determined the relation between port throughput and macroeconomic variables.

The results showed that marine products cargo is the main non-containerized export, whereas the non-containerized import is mainly constituted by fuel oil, industrial materials and marine products. The aggregation of these cargos handled in the port would make up the non-containerized port throughput. Among the available macroeconomic variables in the present study, the national GDP has a relatively high relation with the non-containerized export. However, the non-containerized import is mainly related to the world GDP. The non-containerized port throughput showed relatively high correlations with the national GDP. The results unveiled that the relation between containerized port throughput and the volume of national export trade is more than other macroeconomic variables.

The new finding of this study is that the application of mutual information offers a solution for reducing epistemic uncertainty in port throughput analysis as it: first – determines the main cargos that significantly contribute to port throughput, and second – effectively identifies the relation between port throughput and macroeconomic variables. Thus, this approach can improve the reliability of port throughput forecast, which is recommended for further study.

References

- Bankes, S. (1993), "Exploratory modeling for policy analysis", *Operations Research*, Vol. 41 No. 3, pp. 435-449.
- Chen, Z., Chen, Y. and Li, T. (2016), "Port cargo throughput forecasting based on combination model", in *Joint International Information proceedings of the Technology, Mechanical and Electronic Engineering Conference*.
- Chou, C.C., Chu, C.W. and Liang, G.S. (2008), "A modified regression model for forecasting the volumes of Taiwan's import containers", *Mathematical and Computer Modelling*, Vol. 47 Nos 9/10, pp. 797-807.
- De Langen, P.W., Van Meijeren, J. and Tavasszy, L.A. (2012), "Combining models and commodity chain research for making long-term projections of port throughput: an application to the Hamburg-Le Havre range", *European Journal of Transport and Infrastructure Research*, Vol. 12 No. 3, available at: <http://resolver.tudelft.nl/uuid:e157a61d-3ac6-46df-a47f-308fc0b2ab3d>

- Dragan, D., Keshavarzsaleh, A., Intihar, M., Popović, V. and Kramberger, T. (2020), "Throughput forecasting of different types of cargo in the Adriatic seaport Koper", *Maritime Policy and Management*, pp. 1-27, doi: [10.1080/03088839.2020.1748242](https://doi.org/10.1080/03088839.2020.1748242).
- Eskafi, M., Fazeli, R., Dastgheib, A., Taneja, P., Ulfarsson, G.F., Thorarinsdottir, R.I. and Stefansson, G. (2019), "Stakeholder salience and prioritization for port master planning, a case study of the multi-purpose Port of Isafjordur in Iceland", *European Journal of Transport and Infrastructure Research*, Vol. 19 No. 3, pp. 214-260, doi: [10.18757/ejtir.2019.19.3.4386](https://doi.org/10.18757/ejtir.2019.19.3.4386).
- Eskafi, M., Fazeli, R., Dastgheib, A., Taneja, P., Ulfarsson, G.F., Thorarinsdottir, R.I. and Stefansson, G. (2020), "A value-based definition of success in adaptive port planning: a case study of the Port of Isafjordur in Iceland", *Maritime Economics and Logistics*, Vol. 22 No. 3, pp. 403-431, doi: [10.1057/s41278-019-00134-6](https://doi.org/10.1057/s41278-019-00134-6).
- Frankel, E.G. (1987), *Port Planning and Development*, Wiley-Interscience, New York, NY.
- Fraser, A.M. and Swinney, H.L. (1986), "Independent coordinates for strange attractors from mutual information", *Physical Review A*, Vol. 33 No. 2, p. 1134.
- Fuentes, J., Poncela, P. and Rodriguez, J. (2015), "Sparse partial least squares in time series for macroeconomic forecasting", *Journal of Applied Econometrics*, Vol. 30 No. 4, pp. 576-595.
- Geweke, J. and Whiteman, C. (2006), "Handbook of economic forecasting, chapter 1 Bayesian forecasting", in Elliott, G., Granger, C.W.J. and Timmermann, A. (Eds), *Handbook of Economic Forecasting*, Elsevier, pp. 3-80. [10.1016/S1574-0706\(05\)01001-3](https://doi.org/10.1016/S1574-0706(05)01001-3).
- Gökkuş, Ü., Yıldırım, M.S. and Aydin, M.M. (2017), "Estimation of container traffic at seaports by using several soft computing methods: a case of Turkish seaports", *Discrete Dynamics in Nature and Society*, Vol. 2017.
- Gosasang, V., Chandraprakaikul, W. and Kiattisin, S. (2011), "A comparison of traditional and neural networks forecasting techniques for container throughput at Bangkok port", *The Asian Journal of Shipping and Logistics*, Vol. 27 No. 3, pp. 463-482.
- Gosasang, V., Yip, T.L. and Chandraprakaikul, W. (2018), "Long-term container throughput forecast and equipment planning: the case of Bangkok port", *Maritime Business Review*, Vol. 3 No. 1, pp. 53-69.
- Günther, H.O. and Kim, K.H. (2005), *Container Terminals and Automated Transport Systems: logistics Control Issues and Quantitative Decision Support*, Springer-Verlag, available at: www.springer.com/gp/book/9783540223283
- Hall, M.A. (1999), "Correlation-based feature selection for machine learning", PhD diss., University of Waikato.
- Hänninen, M. (2014), "Bayesian networks for Maritime traffic accident prevention: benefits and challenges", *Accident Analysis and Prevention*, Vol. 73, pp. 305-312.
- Hänninen, M. and Kujala, P. (2014), "Bayesian network modeling of port state control inspection findings and ship accident involvement", *Expert Systems with Applications*, Vol. 41 No. 4, pp. 1632-1646.
- Hanushek, E.A. and Kimko, D.D. (2000), "Schooling, labor-force quality, and the growth of nations", *American Economic Review*, Vol. 90 No. 5, pp. 1184-1208.
- Hui, E.C.M., Seabrooke, W. and Wong, G.K.C. (2004), "Forecasting cargo throughput for the port of Hong Kong: error correction model approach", *Journal of Urban Planning and Development*, Vol. 130 No. 4, pp. 195-203.
- Icelandic Directorate of Fisheries (2019), "Find ship – individual vessels – web of directorate of fisheries", available at www.fiskistofa.is/english/quotas-and-catches/total-catches-by-harbours-months-and-vessel-type/ (accessed December 2019).
- JICA (1994), *Final Report for the Study on Modernization of Bangkok Port in the Kingdom of Thailand*, Japan International Cooperation Agency, No. 1, pp. 199-202.

- Jugovic, A., Hess, S. and Jugović, T.P. (2011), "Traffic demand forecasting for port services", *PROMET – Traffic&Transportation*, Vol. 23 No. 1.
- Kraskov, A., Stögbauer, H. and Grassberger, P. (2004), "Estimating mutual information", *Physical Review E*, Vol. 69 No. 6, p. 066138.
- Kwak, N. and Choi, C.H. (2002), "Input feature selection for classification problems", *IEEE Transactions on Neural Networks*, Vol. 13 No. 1, pp. 143-159.
- Li, W. (1990), "Mutual information functions versus correlation functions", *Journal of Statistical Physics*, Vol. 60 Nos 5/6, pp. 823-837.
- Li, Y.F., Xie, M. and Goh, T.N. (2009), "A study of mutual information based feature selection for case based reasoning in software cost estimation", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 5921-5931.
- Luo, M. and Grigalunas, T.A. (2003), "A spatial-economic multimodal transportation simulation model for US coastal container ports", *Maritime Economics and Logistics*, Vol. 5 No. 2, pp. 158-178.
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. and Suetens, P. (1997), "Multimodality image registration by maximization of mutual information", *IEEE Transactions on Medical Imaging*, Vol. 16 No. 2, pp. 187-198.
- Meersman, H., Van de Voorde, E. and Janssens, S. (2003), "Port throughput and international trade: have port authorities any degrees of freedom left?", in Loyen, R., Buyst, E. and Devos, G. (Eds), *Struggling for Leadership: Antwerp-Rotterdam Port Competition between 1870-2000*, Springer, pp. 91-113.
- Milenković, M., Milosavljevic, N., Bojović, N. and Val, S. (2019), "Container flow forecasting through neural networks based on metaheuristics", *Operational Research*, doi: [10.1007/s12351-019-00477-1](https://doi.org/10.1007/s12351-019-00477-1).
- Montewka, J., Ehlers, S., Goerlandt, F., Hinz, T., Tabri, K. and Kujala, P. (2014), "A framework for risk assessment for Maritime transportation systems—a case study for open sea collisions involving RoPax vessels", *Reliability Engineering and System Safety*, Vol. 124, pp. 142-157.
- Peng, W.Y. and Chu, C.W. (2009), "A comparison of univariate methods for forecasting container throughput volumes", *Mathematical and Computer Modelling*, Vol. 50 Nos 7/8, pp. 1045-1057.
- Peng, H., Long, F. and Ding, C. (2005), "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27 No. 8, pp. 1226-1238, doi: [10.1109/TPAMI.2005.159](https://doi.org/10.1109/TPAMI.2005.159).
- Pethel, S.D. and Hahs, D.W. (2014), "Exact test of independence using mutual information", *Entropy*, Vol. 16 No. 5, pp. 2839-2849.
- Ping, F.F. and Fei, F.X. (2013), "Multivariate forecasting mode of Guangdong province port throughput with genetic algorithms and back propagation neural network", *Procedia - Social and Behavioral Sciences*, Vol. 96, pp. 1165-1174.
- Robnik-Šikonja, M. and Kononenko, I. (2003), "Theoretical and empirical analysis of ReliefF and RReliefF", *Machine Learning*, Vol. 53 Nos 1/2, pp. 23-69.
- Sagar, R.P. and Guevara, N.L. (2005), "Mutual information and correlation measures in atomic systems", *The Journal of Chemical Physics*, Vol. 123 No. 4, p. 044108.
- Shannon, C.E. (1948), "A mathematical theory of communication", *Bell System Technical Journal*, Vol. 27 No. 3, pp. 379-423.
- Smil, V. (2012), *Global Catastrophes and Trends: The Next Fifty Years*, Cambridge, MA.
- Soofi, E.S., Zhao, H. and Nazareth, D.L. (2010), "Information measures", *WIREs Computational Statistics*, Vol. 2 No. 1, pp. 75-86.
- Statistics Iceland Office (2019), "Profitability in fishing and fish processing 2018", *Statistics Iceland*, Vol. 104 No. 14, available at: <https://static.is/publications/publication-detail/>
- Steuer, R., Kurths, J., Daub, C.O., Weise, J. and Selbig, J. (2002), "The mutual information: detecting and evaluating dependencies between variables", *Bioinformatics*, Vol. 18 No. Suppl 2, pp. S231-S240.

- Taneja, P. (2013), "The flexible port", PhD diss., Delft university of technology, available at: <https://doi.org/10.4233/uuid:a9f0c128-d4c3-41a2-8790-13aec89dca63>
- Taneja, P., Ligteringen, H. and Walker, W.E. (2012), "Flexibility in port planning and design", *European Journal of Transport and Infrastructure Research*, Vol. 12 No. 1, available at: <http://resolver.tudelft.nl/uuid:c72ec34f-90c1-4b24-ae86-dfea21b8efab>
- Taneja, P., Ligteringen, H. and Van Schuylenburg, M. (2010), "Dealing with uncertainty in design of port infrastructure systems", *J. of Design Research*, Vol. 8 No. 2, pp. 101-118.
- Van Dorsser, J.C.M., Wolters, M. and Van Wee, B. (2012), "A very long term forecast of the port throughput in the Le Havre-Hamburg range up to 2100", *European Journal of Transport and Infrastructure Research*, Vol. 12 No. 1, doi: [10.18757/ejtir.2012.12.1.2951](https://doi.org/10.18757/ejtir.2012.12.1.2951).
- Walker, W.E., Lempert, R.J., Kwakkkel, J.H., Gass, S.I. and Fu, M.C. (2013), "Deep uncertainty", in Gass, S. I. and Fu, M. C. (Eds), *Encyclopedia of Operations Research and Management Science*, Springer, Boston, MA, pp. 395-402.
- Wu, B., Yip, T.L., Yan, X. and Mao, Z. (2020), "A mutual Information-based Bayesian network model for consequence estimation of navigational accidents in the Yangtze River", *Journal of Navigation*, Vol. 73 No. 3, pp. 559-580.
- Yang, Z., Yang, Z. and Yin, J. (2018), "Realising advanced risk-based port state control inspection using data-driven Bayesian networks", *Transportation Research Part A: Policy and Practice*, Vol. 110, pp. 38-56.
- Zeng, Z., Zhang, H., Zhang, R. and Zhang, Y. (2014), "A hybrid feature selection method based on rough conditional mutual information and naive Bayesian classifier", *ISRN Applied Mathematics*, Vol. 2014
- Zuur, A.F., Fryer, R.J., Jolliffe, I.T., Dekker, R. and Beukema, J.J. (2003), "Estimating common trends in multivariate time series using dynamic factor analysis", *Environmetrics*, Vol. 14 No. 7, pp. 665-685.

About the authors

Majid Eskafi (ORCID: 0000-0002-4697-4754) is Doctoral Candidate in the Faculty of Civil and Environmental Engineering, University of Iceland. He received MSc degree in Physical Oceanography from Azad University in Iran, focused on physical modeling of coastal structures. He also has MS in Coastal and Marine Management from the University of Akureyri in Iceland, focused on numerical simulation of oceanographic parameters. Majid Eskafi is the corresponding author and can be contacted at: mae47@hi.is

Milad Kowsari (ORCID: 0000-0002-4331-4021) is Post-Doctoral Researcher in the Faculty of Civil and Environmental Engineering at the University of Iceland. Dr Kowsari received his PhD in Civil Engineering from the University of Iceland. The principal focus of his research is probabilistic seismic hazard assessment, earthquake ground motion modeling, engineering seismology and data analysis.

Ali Dastgheib (ORCID: 0000-0002-5914-7213) is Senior Researcher and Lecturer in Coastal Engineering and Port Development at IHE-Delft. He has combined his background in port planning and design of coastal structures with the knowledge of coastal engineering, coastal hazards and climate change, which leads to a better understanding of new coastal development in coastal zones and climate change adaptation. At IHE-Delft, he is carrying out research, advisory and capacity building projects in the field of tidal basins and estuaries, coastal morphology, climate change adaptation, coastal risk, coastal zone management, port planning and port traffic simulation. Dr Dastgheib received his PhD degree from the Delft University of Technology.

Gudmundur F. Ulfarsson (ORCID: 0000-0002-0094-7531) is Professor and Head of the Faculty of Civil and Environmental Engineering, University of Iceland. Prof Ulfarsson received PhD and MSc degrees in Civil Engineering, majoring in Transportation Engineering and Econometrics, from the University of Washington, WA, USA and BSc degrees in Physics and Computer Science from the University of Iceland. Prof Ulfarsson teaches courses in transportation engineering and geographical

information systems. His major research interests include traffic safety, econometric analysis and travel behavior analysis.

Poonam Taneja (ORCID: 0000-0002-0221-9829) is Senior Lecturer in the Faculty of Civil Engineering and Geosciences, Delft University of Technology. A civil engineer with a master's degree in port planning, she has extensive experience in port projects in various countries. After working at the Municipality of Rotterdam for 15 years, she defended her dissertation titled "The Flexible Port" in 2013. Her research areas include long-term planning under uncertainty, integrated and sustainable port design, building with nature, climate change, energy transition and supply chain optimization. Currently, she is a part of research teams carrying out projects for SmartPort and NWO-WOTRO.

Ragnheidur I. Thorarinsdottir (ORCID: 0000-0002-2490-1781) is Rector of the Agricultural University of Iceland and Visiting Professor at the University of Iceland. She received PhD and MSc degrees in Chemical Engineering from the Technical University of Denmark and MBA from the University of Iceland. Thorarinsdottir is former director and deputy director of the National Energy Authority in Iceland. She has a broad experience from various fields of engineering; energy use, material science and biochemistry both as a scientist and from administration. She has been responsible for several R&D and innovation projects internationally and has been an active entrepreneur with foundation of companies. She has been member of scientific committees and boards and is experienced in financial and strategic projects. She has supervised PhD, MSc, MPM and BSc students and has a long teaching experience in engineering and business courses.