Advanced Engineering Informatics 40 (2019) 29–35

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Data-driven operational risk analysis in E-Commerce Logistics

Gangyan Xu[a,b], Xuan Qiu[c,*], Meng Fang[d], Xiaofei Kou[e], Ying Yu[f]

[a] School of Architecture, Harbin Institute of Technology, Shenzhen, China
[b] Intelligent Transportation Lab, Shenzhen Key Laboratory of Urban Planning and Decision Making, Harbin Institute of Technology, Shenzhen, China
[c] Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Hong Kong, China
[d] Tencent AI Lab, Shenzhen, China
[e] School of Management Engineering, Zhengzhou University of Aeronautics, Zhengzhou, China
[f] HKU-ZIRI Lab for Physical Internet, Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
E-Commerce Logistics
Operational risks
Data analytics
Risk analysis
Gaussian mixture model

ABSTRACT

The efficiency of E-Commerce Logistics (ECL) has become a major success factor for e-commerce companies in the competitive marketplace nowadays. However, the operation of ECL is complex and vulnerable to many risks, which would severely threaten its performance. A clear understanding of these risks would benefit a lot for conducting targeted measures to effectively mitigate their adverse effects. Therefore, this paper proposes a quantitatively analysis approach for operational risks in ECL based on extensive historical e-commerce transaction data. More specifically, the typical operation process of ECL is extracted through sequential analysis of key activities. After that, taking operation time as the key performance indicator, the performance patterns of different operation phases are analyzed. Then, considering the diverse distributions of operation time in different phases, especially the multimodal distribution of transportation time, a Gaussian Mixture Model (GMM) based risk analysis approach is proposed. Finally, an experimental case study is provided to measure the operational risks using real-life ECL data, and several managerial implications are also discussed based on the results.

1. Introduction

E-commerce is expanding rapidly worldwide and has been recognized as a major engine that drives the evolution of logistics [1]. On the one hand, the proliferation of e-commerce is creating enormous demands for logistic services. Take China for example, the market of retail e-commerce has reached US$1.05 trillion in 2017, with an annual increasing rate around 32%. Accordingly, the number of logistics orders exceeded 40 billion in 2017, increased 28% compared with the number of 2016 [2]. On the other hand, the increasingly competitive environment has forced e-commerce companies paying more attention on their logistic systems [3], whose performance is strongly correlated with their successes [4,5]. E-commerce also brings many new features to its logistics, including highly stochastic demands, huge number of orders, great diversities, and small and irregular items, which stimulate forming the new logistics paradigm, E-Commerce Logistics (ECL).

Along with the continuous rapid development of e-commerce, ECL has attracted many attentions from both industries and academics. For instance, JD.com, one of the largest B2C e-commerce platforms in China, has invested heavily in building its own logistics system [6]. Alibaba has built Cainiao Logistics to further improve the ECL efficiency through integrating various and distributed logistics resources [7]. Besides, Hu and Chang [8] developed an Automated Storage/Retrieval Systems (AS/RSs) using multilevel conveying device with three-dimensional movement to fit the small and irregular items or parcels in ECL, while Chen et al. [9] proposed efficient heuristic routing methods in e-commerce warehouses with ultra-narrow aisles and access restriction. Ruan and Shi [10] designed an Internet of Things (IoT) based framework to monitor and assess the fruit freshness in ECL. Shao et al. [11] adopted the concept of sliding time window and developed heuristics algorithm to synchronize the last mile delivery of mass e-commerce orders. And Xu et al. [1,12] proposed effective auction-based strategies to facilitate the matching of shippers (e-commerce platforms) and carriers (3rd party logistics) in ECL.

As a typical logistics system, ECL is vulnerable and open to many risks. For instance, the uncertainties on demand and supply, mistakes in warehousing activities, accidents during delivery would lead to great performance declining and heavy losses. The occurrence of disasters, like Typhoon Mangkhut in 2018, may damage ECL infrastructures, and brings significant delay on delivering goods and severe overstock at peripheral distribution centers. Thus risk management is essential in ECL, and becomes an important topic nowadays. Although a couple of

* Corresponding author.
E-mail address: xuanqiu@ust.hk (X. Qiu).

https://doi.org/10.1016/j.aei.2019.03.001
Received 30 September 2018; Received in revised form 31 January 2019; Accepted 2 March 2019

studies have been conducted [13,14], the research on ECL risk management is far from enough, leaving many topics remain unfold, including quantitative analysis of risks in ECL, optimization models on its operational risk, risk control strategies, etc.

This paper focuses on quantitative operational risk analysis in ECL. In literature, operational risks mainly refer to the risks owing to supply/demand uncertainties, human mistakes, and accidents that would decrease the service level or threat the normal operations [15]. Many efforts have been done on operational risk management in logistics and supply chain systems [16,17]. Specifically, Chen et al. [18] examined the effect of supplier, customer, and internal collaboration on mitigating operational risks in supply chain. Mo and Cook [19] developed a quantitative risk analysis approach of designing logistics network system to further improve the operational cost efficiency of automotive manufacturing industry. Considering the operation process of port logistics, Sutrisnowati et al. [20] proposed a bayesian network-based approach to analyze the lateness probability in container handling. Besides, Shang et al. [21] analyzed the pattern of transport risks in air cargo logistics, and developed a method using bayesian nonparametrics to forecast transport risks based on massive air cargo data.

However, to the best of our knowledge, there is still rare research on operational risks in ECL. In addition, ECL has distinct characteristics compared with traditional logistics systems, which may result in the differences of operation processes, risk factors, risk patterns, etc. Existing risk analysis approaches hence cannot be directly adopted.

Fortunately, with the mature of IoT technologies, extensive and accurate information about logistics activities can be collected, recorded, and shared [22–25], which provides new opportunities for operational risk analysis. This is pretty true in ECL since e-commerce is born with the nature of digitalization and has been recognized as an engine that drives the application of information technologies in industry. Such information not only makes it possible to provide track and trace services for e-commerce customers, but also enables in-depth and comprehensive analysis of ECL, including its operational risks.

Therefore, take the advantages of IoT technologies, this paper aims at making quantitative analysis of operational risks in ECL based on real-life e-commerce transaction data. More specifically, the typical operation process of ECL is extracted through sequential analysis of key activities. Then taking operation time as the key performance indicator, the performance patterns of different operation phase are analyzed. After that, the abnormal detection method using Gaussian Mixture Model (GMM) is proposed, based on which the risks in different operation phases can be quantitatively analyzed.

The remainder of this paper is organized as follows. Section 2 introduces the data set adopted in this paper. Section 3 discusses the operational risks in ECL, including the operation process of ECL and the measurement of operational risks. Section 4 presents the method of identifying and evaluating the risks using GMM. An experimental case study is given in Section 5, and Section 6 concludes the whole paper and points out future research directions.

2. Data set

The data set adopted in this paper is provided by one of our collaborators, containing the e-commerce transaction records, warehouse information, and region information of an e-commerce platform in China.

In total, there are 905,422 historical e-commerce transaction records, each with around 120 properties denoting the detailed information of the transaction, including customer information, product information, payment information, delivery information, and time stamps of key activities (or status change). Meanwhile, the detailed information of 1045 self-owned and collaborative warehouses is provided, including their types, locations, corresponding distribution centers, and parent sub-companies. And the region information contains regional divisions of 520 cities (or districts) in China.

In order to improve the efficiency of data analytics afterwards, data preprocessing is conducted through three steps: (1) Data reduction. Besides the ECL operation data, the raw transaction data contains many other data which are irrelevant to the problem of ECL risks analysis. This step eliminates these useless data to reduce the overall data size, and only 46 key properties are kept for further processing. (2) Data cleaning. The raw transaction data inevitably contains inconsistent data, or some of the records may have incomplete data. In this research, these records with inconsistent data and incomplete data will be removed from the experimental data set. During this step, 279,222 records are filtered from the raw transaction data. (3) Data conversion. Based on the warehouse and region information, several new properties are added to denote whether the e-commence order is an inner region/province/city order (departure warehouse and the deliver address are in the same region/province/city). Finally, these processed data are integrated into one data table and stored in a CSV file for further analysis.

3. Operational risks in E-Commerce Logistics

The operation process of ECL is composed by many inter-dependent successive phases, which would face diverse risks, present different performance patterns, and accumulatively affect the overall performance of ECL. In this section, the operation process of ECL will be analyzed first, and then the quantitative measurement method of operational risks will be introduced.

3.1. Operation process of E-Commerce Logistics

ECL operation process is complex that consists of many activities that are conducted distributively by different stakeholders, including customers, e-commerce platforms, e-retailers, logistics service providers, etc. A clear understanding of it is a prerequisite for operational risks analysis. Therefore, take the advantages of extensive historical e-commerce transaction data, key activities involved in ECL are extracted first. Then their sequential relationships are examined through comparing their time stamps stored in these records, which help form the basic operation process of ECL. Besides, field studies with JD.com and Taobao were conducted to further confine the process and make some supplements. After that, the general operation process of ECL can be constructed, as depicted in Fig. 1.

Here, the blocks with hard borders and hard arrows are those analyzed directed from the e-commerce transaction data, while those blocks with dashed borders and dashed arrows are derived from field studies. Besides, the blocks with rounded corner are the activities conducted by e-commerce customers during their online shopping processes (e.g. search for items, add them to cart, place the order, etc.).

The operation process of ECL begins with customers placing their orders on e-commerce platform (Place Order). Then e-retailers could receive the orders and check the details through the platform, and begin to process the orders (Process Order). According to the content of ordered products, the orders will be sent to different warehouses, where they would be released to warehousing operators through printing them out or other channels (Print Order). The operators would then pick the products from shelves (Order Picking) and package products of the same order (Package) for delivery (Ship). Once these well packaged items reach the distribution centers, they will be sorted according to destinations (Sort) and then deliver to final customers (Ship). Before receiving the orders, the customers should pay either on-line or off-line (Pay), and the platform or e-retailer need also check the account (Check Account) during this process. After the orders are well received by final customers (Receive Order), the operation process of ECL ends.

According to these transaction data, an interesting phenomenon in real-life ECL operation process is detected, that around 8% orders are processed in advance before they are formally placed by customers. This may be because e-retailers could monitor the ordering process of

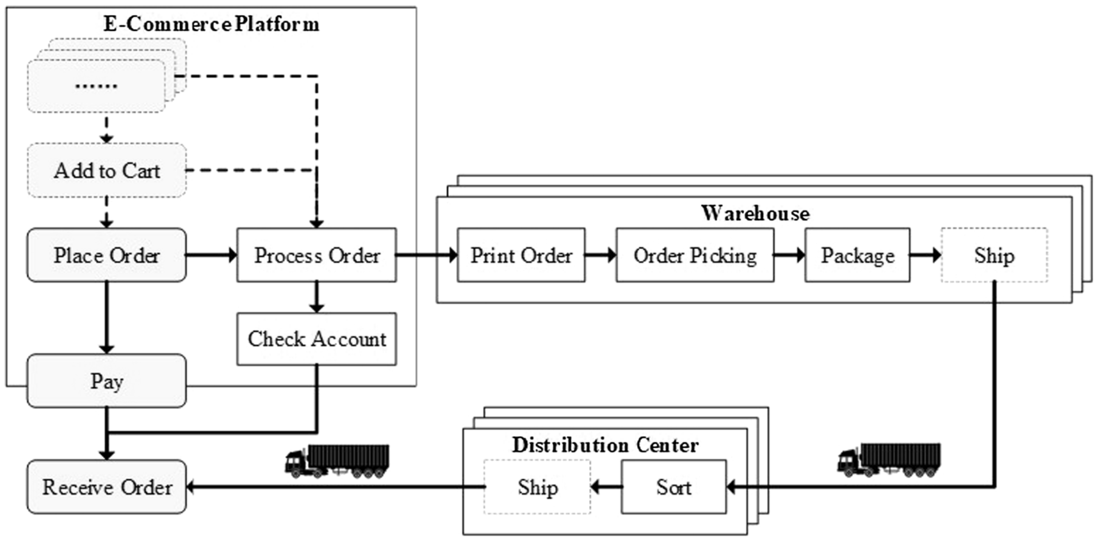

Fig. 1. General operation process of E-Commerce Logistics (ECL).

customers before they releasing their orders, such as adding products into their e-carts (Add to Cart), editing delivery addresses, or selecting payment methods. Despite it is not clear whether the timing advance could benefit the efficiency of entire ECL, some e-retailers would like to process the orders once they get all the necessary order information and consider the customers are very likely to place the order.

Besides, the operation process shown in Fig. 1 is a general process without considering the very detailed and specific operation activities. On the one hand, it may be much complex in reality. Take JD.com for example, it has several layers of distribution centers and the orders would pass from the top layer to the bottom layer before reaching final customers. On the other hand, there would be slightly differences between different operation modes of ECL. For example, JD.com has self-built ECL system hence it is responsible for conducting all the activities in warehouses and distribution centers, while in Taobao, e-retailers need to manage their own warehouses and outsource the delivery activities to 3rd party logistics providers. The different operation processes would thus influence the performance of ECL, including their reliability, service quality, operation time, etc. In the following of this paper, without loss of generality, the typical operation process shown in Fig. 1 is adopted for further analysis.

3.2. Measurement of operational risks

During the operation process of ECL, there are many threats that deteriorate or even disrupt its normal operation activities. In addition, these activities involved are usually highly inter-dependent with each other, and the delay of one activity could easily put its successive activities off schedule in real-time, leading to undesirable delays of the entire ECL process.

To minimize and avoid service level reduction during the logistics operation process, some researchers examined the causes of threats, their occurrence probabilities, and risk avoidance strategies. Nevertheless, the scenarios of logistics systems in reality vary from each other and the causes of threats may also have substantial differences, which make it challenging to comprehensively analyze the causes based on specific scenarios. Thus some researchers analyzed operational risks from the perspective of their adverse effects, such as the earliness or tardiness in logistics operations [21,26]. Considering the dynamic and diverse scenarios in ECL, this should also be appropriate for analyzing the operational risks in ECL. Meanwhile, since on-time delivery is the most important success factor of ECL [13,27], the tardiness of operation activities is selected as the indicator to represent the operational risks in ECL.

According to the above analysis, the operational risks of ECL can be defined as the expected delay of operation activities in ECL, which can be formulated as:

$$R = P \times D \tag{1}$$

where P is the probability of delay and D refers to the amount of delay.

Given n observations of ECL operation data, if m orders are delayed, its risks can be calculated as follows:

$$R = \frac{m}{n} \times \frac{\sum_{i=1}^{m}(T_i^a - T_i^r)}{m} = \frac{\sum_{i=1}^{m}(T_i^a - T_i^r)}{n} \tag{2}$$

Here, T_i^a refers to the actual operation time and T_i^r refers to the required operation time.

4. Operational risk analysis using Gaussian mixture model

The entire operation process of ECL consists of several inter-dependent successive phases. In order to measure the operational risks of every phase using Eq. (2), it is a prerequisite to have the required operation time (T_i^r) of each phase. However, in practice, most e-retailers and e-commerce platforms only provide the maximum delivery time guarantee (required operation time for entire ECL), while lacking clear regulations on the operation time of each phase. Meanwhile, since ECL scenarios vary from each other, it is difficult to determine the required operation time for each phase.

Taking these difficulties into consideration, this section analyzes the distribution of operation time in each phase based on historical transaction data, then proposes a Gaussian Mixture Model (GMM) based approach to detect the maximum time for normal operations (the required operation time of each phase). After that, the method for operational risks analysis of every phase in ECL is presented.

4.1. Operation time analysis

According to the ECL operation process depicted in Fig. 1, this research divides the whole process into six phases, that consists of order processing, order releasing, order picking, packaging, sorting, and transportation. The definitions of these phases are presented in Table 1. The operation time of each phase can then be calculated through comparing the time stamps of its end and start activities.

In the following, the patterns of ECL operation time are analyzed using the data set described in Section 2.

Fig. 2 shows the operation time distribution of the entire ECL process. It can be clearly observed that the entire ECL operation time

Table 1
Division of ECL operation phases.

Phase ID	Phase Description	Start Activity	End Activity
1	Order Processing	Place Order	Process Order
2	Order Releasing	Process Order	Print Order
3	Order Picking	Print Order	Order Picking
4	Packaging	Order Picking	Package
5	Sorting	Package	Sort
6	Transportation	Sort	Receive Order
–	ECL	Place Order	Receive Order

follows a multimodal distribution and most e-commerce orders can be delivered with 3 days.

Fig. 3 shows the operation time distribution of each phase in ECL. Generally, transportation and sorting have been identified as the two most time-consuming phases in ECL, while order processing is the most efficient one, most of which can be finished within 5 min. Besides, it is also found that the distribution of transportation time is multimodal while the others are single modal.

Empirically, the transportation time is positively related with transportation distances. To learn the effect of transportation distance on the distribution of transportation time, this research divides these e-commerce orders into three groups according to the distances between warehouses and receiving addresses, they are inner-city order, inner-province order, and inner-region order. Fig. 4 shows the distributions of transportation time and total ECL time for different types of orders.

It can be found that the distributions of transportation time and total ECL time in different order groups are still multimodal. The phenomenon can be explained from the fixed schedules on the shuttle services in ECL. Similar with that in air cargo transportation [21], once the e-commerce orders failed to be loaded on its scheduled truck/train/flight, it has to be delivered through the next shuttle service on the same route later. The scheduled gap between shuttle services generates the gaps between adjacent peaks in the distributions of transport time and total ECL time.

4.2. Gaussian mixture model based anomaly detection

With the distribution patterns of different phases in ECL, the corresponding anomaly detection method can be designed to identify the threshold that separates the normal operation time and abnormal operation time, which is also treated as the required operation time in this research. Considering diverse distribution patterns, especially the multimodal distributions of transportation time and entire ECL time, Gaussian Mixture Model (GMM) is selected as the anomaly detection method, which could well fit various distribution patterns [28].

In general, GMM is the weighted sum of several Gaussian probability distributions $N(x|\mu_k, \sigma_k)$, $k = 1, \ldots, K$, and each of them is called a component. GMM can be denoted as:

$$p(x|\lambda) = \sum_{k=1}^{K} \omega_k N(x|\mu_k, \sigma_k) \tag{3}$$

where $\lambda = \{\omega_k, \mu_k, \sigma_k\}$ are the GMM parameters, K is the number of components, $\omega_k (0 < \omega_k < 1)$ is the weight of each component, which satisfies $\sum_{k=1}^{K} \omega_k = 1$. GMM is a flexible unsupervised clustering method, and has been widely adopted as an efficient tool for anomaly detection, such as in maritime navigation system [29], critical events detection [30], flight operation [31], building systems [32], health-care systems [33], etc.

In this research, the inputs of the GMM-based Anomaly Detection (GMM-AD) method are the operation time of each phase provided by the data set described in Section 2. Based on the assumption that the majority of operation time exhibit common patterns under routine ECL operations and a few outliers that deviate from those common patterns may be risky on the performance of entire ECL operations, GMM-AD automatically clusters the massive operation time records into groups and then the extremely long operation time with low occurrence probability is identified as the abnormal operation time.

In order to build the GMM for operation time in every phase of ECL, it is necessary to define the basic attributes of the model. In this research, the parameters among each component of GMM are not shared to maximize the goodness of fit. Besides, the number of mixture components (K) is determined through observations on the distributions of operation time.

After a GMM configuration is well defined, the parameters of the GMM can be obtained using Expectation-Maximization (EM) algorithm [34], which is the most popular and mature method. The working process of EM algorithm can be described as follows:

Step 1: Initialize GMM parameters $\lambda = \{\omega_k, \mu_k, \sigma_k\}$.
Step 2 (E Step): For every record of operation time, determine the posteriori probability for each component k using Eq. (3):

$$\gamma(k|x, \lambda) = \frac{\omega_k N(x|\mu_k, \sigma_k)}{\sum_{i=1}^{K} \omega_i N(x|\mu_i, \sigma_i)} \tag{4}$$

Step 3 (M Step): Update GMM parameters $\lambda = \{\omega_k, \mu_k, \sigma_k\}$ using Eqs. (4)–(6):

$$\omega_k^{new} = \frac{1}{N} \sum_{t=1}^{N} \gamma(k|x, \lambda) \tag{5}$$

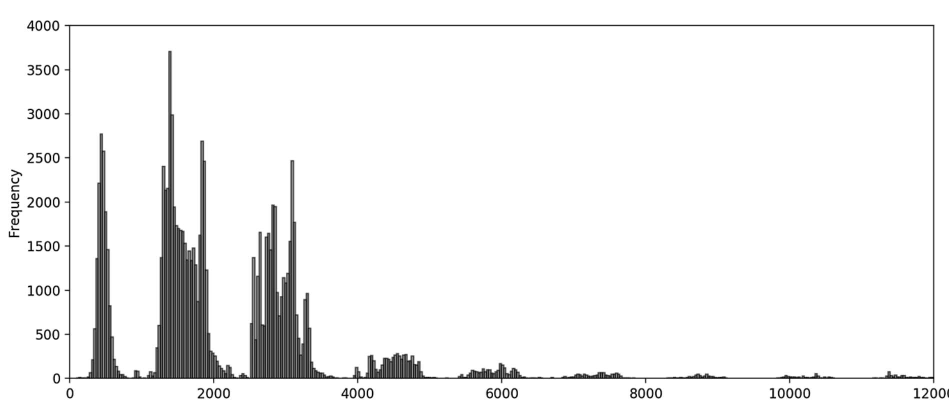

Fig. 2. Operation time distribution of entire ECL.

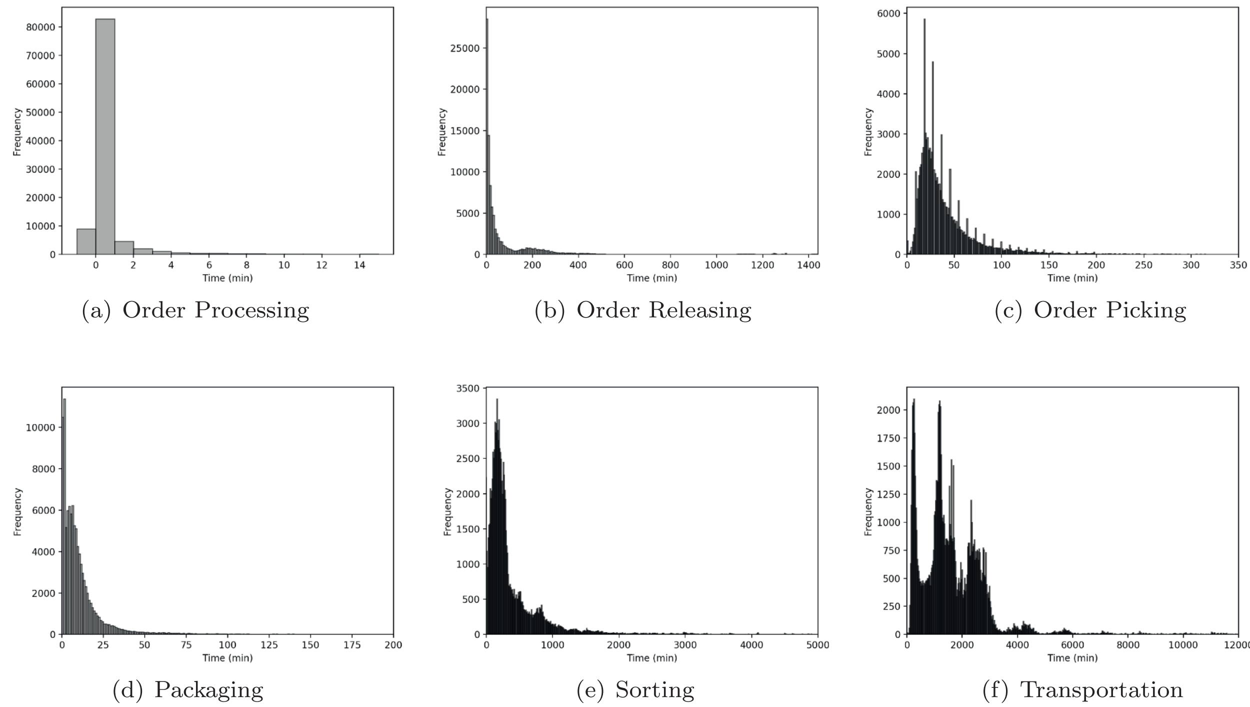

Fig. 3. Operation time distributions of each phase in ECL.

$$\mu_k^{new} = \frac{\sum_{t=1}^{N} \gamma(k|x, \lambda)x}{\sum_{t=1}^{N} \gamma(k|x, \lambda)} \tag{6}$$

$$\sigma_k^{new} = \frac{\sum_{t=1}^{N} \gamma(k|x, \lambda)(x - \mu_k^{new})^2}{\sum_{t=1}^{N} \gamma(k|x, \lambda)} \tag{7}$$

where N refers to the number of records.

Step 4: Evaluate log likelihood using Eq. (7):

$$ln(p(x|\lambda)) = \sum_{t=1}^{N} ln\left(\sum_{k=1}^{K} \omega_k N(x|\mu_k, \sigma_k)\right) \tag{8}$$

Step 5: If $ln(p(x|\lambda)) - ln(p(x|\lambda^{new})) < \varepsilon$, where ε is the threshold for termination, the EM algorithm stops and return λ as the parameter of GMM. Otherwise, set $\lambda = \lambda^{new}$ and go to Step 2 for a new iteration. After the GMM is generated, given the threshold p, the maximum

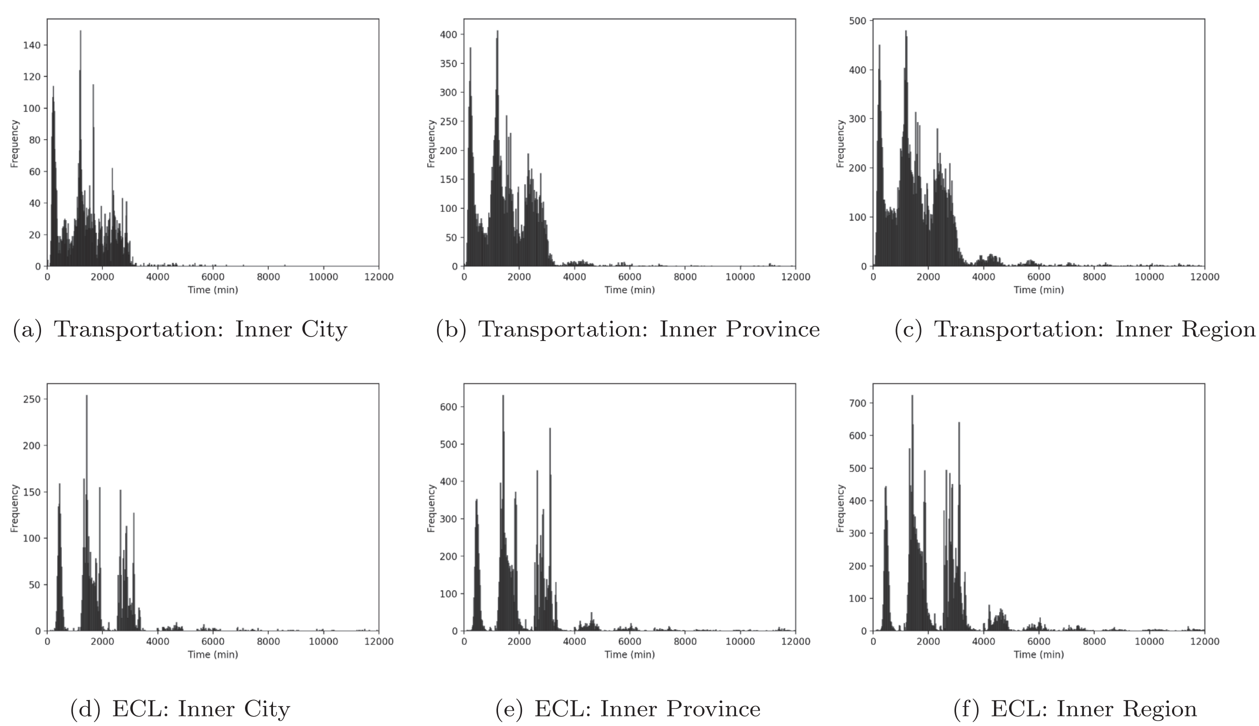

Fig. 4. Transportation and ECL time distributions for different types of orders.

operation time whose Probability Density Function (PDF) equals to p can be identified as the required operation time (T_i^r) for that phase of ECL. The operation time longer than T_i^r would be recognized as abnormal operation time.

4.3. Operational risk analysis

In the practice of ECL, it is common that only the required entire ECL time T_{ecl}^r is provided, which cannot be directly used to analyze the operational risks in each phase. Therefore, this part proposes an approach to identify the required operation time T_i^r for each operation phase using T_{ecl}^r and GMMs generated in previous stage.

The first step is getting the threshold value of entire ECL time based on GMM using:

$$p(T_{ecl}^r|\lambda) = \sum_{k=1}^{K} \omega_k N(T_{ecl}^r|\mu_k, \sigma_k) \tag{9}$$

It means $p(T_{ecl}^r|\lambda)$ orders cannot be delivered within T_{ecl}^r. Since the performance of ECL is an accumulated effect of different phases, it is reasonable to assume that in every operation phase, the worst $p(T_{ecl}^r|\lambda)$ orders are in risks that their operation time may exceed the required time and would lead to the delay of final delivery.

With this assumption, the second step is setting the threshold value for each operation phase as $p(T_{ecl}^r|\lambda)$, and the set of $H = h_1, \ldots, h_n$ that satisfies $p(h_n|\lambda_{op}) = p(p(T_{ecl}^r|\lambda))$ can be obtained, where λ_i is the parameters of GMM for ith operation phase. Then $T_i^r = Max\{H\}$ will be identified as the required operation time for ith operation phase.

According to the generated T_i^r for each operation phase, the operational risk can be analyzed using Eq. (2).

5. Experimental case study

In this section, an experimental case study is given based on the data set described in Section 2. The process of operational risk analysis using the proposed approach will be illustrated, and discussions on the risk analyzing results will also be provided.

5.1. Required operation time identification

According to the operation time distributions shown in Figs. 2 and 3, the number of GMM components K for different operation phases are decided by the number of modals they have. In this case, K_{ecl} is set as 7, K_2 is set as 3, K_6 is set as 4, and K_1, K_3, K_4, K_5 are set as 2. Besides, the maximum iteration times for the EM algorithm is set as 100.

The training algorithm is implemented using Python and running on a personal computer with 3.2 GHz CPU and 16G RAM. In this case, 7 GMMs were trained for ECL time and the operation time of each phase separately. Using the GMMs, the required operation time for each phase in ECL can be calculated.

In the practice of e-commerce in China, the maximum delivery time is frequently set as 3 days. Therefore, in this case, the required operation time for entire ECL time is set as 3 days ($T_{ecl}^r = 4320$ min). Using the GMM of ECL time, the PDF threshold of ECL time is calculated as $p(T_{ecl}^r|\lambda) = 0.097$.

Table 2
Required operation time for each phase of ECL.

Phase	Required Operation Time
Order Processing	1 (min)
Order Releasing	330 (min)
Order Picking	81 (min)
Packaging	23 (min)
Sorting	856 (min)
Transportation	2838 (min)

With the value of $p(T_{ecl}^r|\lambda)$, the required operation time for each phase can be calculated using its GMM, and the results are shown in Table 2.

From Table 2, order processing is the most efficient process ($T_1^r = 1$) while transportation is the most time-consuming process ($T_6^r = 2838$ min). This is consistent with the actual situation that order processing is usually done digitally and automatically through the e-commerce platform, which can be finished online in real-time. However, transportation has to move the goods physically for a long distance, its time is strongly related with the distance and speed (transport mode). Meanwhile, based on Fig. 1, it is found that although the operations in warehouses are more complex than in distribution centers, the required operation time in warehouses ($T_w^r = T_2^r + T_3^r + T_4^r = 434$ min) is shorter than that in distribution centers ($T_{dc}^r = T_5^r = 856$ min).

One thing should be noticed here is that the required operation time (T_i^r) presented in Table 2 is different with the regulated operation time given by individual company. T_i^r is based on the observations of extensive operation practices that refers to the maximum operation time that could still satisfy the overall service quality of entire ECL, while companies may set a more strict regulation on operation time to keep competitive in the ECL market.

5.2. Operation risk analysis

With the required operation time for each phase presented in Table 2, the operational risk of each phase in ECL can be quantitatively calculated using Eq. (2), and the results are shown in Table 3.

P_i refers to the proportion of orders that cannot be finished within required operation time. Since the generated GMM in previous step is difficult to perfectly fit the real distributions of operation time, there is slightly difference between the threshold value $p(T_{ecl}^r|\lambda)$ and P_i, which is calculated using real-life data. For example, only 6.3% orders in the experimental data set cannot be processed within the required operation time, rather than 9.7%.

Generally, from the view of entire ECL, the delivery of e-commerce order is expected to be delayed over seven hours, while for those delayed orders, they are expected to be delayed more than three days. Focusing on each operation phase, transportation is identified as the most risky phase in ECL ($R_6 = 363.368$). Although over 90% orders can be transported within 2838 min, every order is expected to be delayed over six hours. Moreover, for these delayed orders, they are expected to be delayed more than two and a half days ($D_6 = 3747.177$). Order releasing and sorting are the other two phases with high risks that are expected to be delayed around one to three hours ($R_5 = 73.013$, $R_2 = 173.834$). It is also identified that order processing, order picking, and packaging are with lower risks. All of their expected delay is within ten minutes ($R_1 = 1.010$, $R_3 = 8.306$, $R_4 = 2.823$).

5.3. Discussions

According to the risk analyzing results given above, several implications can be concluded as follows.

Firstly, transportation is identified as the most time-consuming and

Table 3
Operational risks in ECL.

Phase	P_i	D_i	R_i
Order Processing	0.063	16.005	1.010
Order Releasing	0.097	1790.844	173.834
Order Picking	0.098	84.880	8.306
Packaging	0.101	27.810	2.823
Sorting	0.097	754.526	73.013
Transportation	0.097	3747.177	363.368
ECL	0.097	4433.623	428.683

risky process in ECL. Its performance directly determines the performance of entire ECL. Therefore, to further decrease the risks of ECL, more reliable and resilient transportation systems should be designed. Meanwhile, more flexible and efficient transport services should be provided to further improve the efficiency of ECL.

Secondly, order releasing has the second highest risks. More efforts should be paid on analyzing its detailed working logics, the types of risks it faces, and the causes of risks. Then targeted measures should be implemented to mitigate the adverse effects of these risks.

Thirdly, sorting is the second most time-consuming process in ECL and its efficiency improvements may have a great impact on the efficiency of entire ECL. Therefore, more investment can be made on the sorting processes to further improve its efficiency, including adding more labors, adopting automation tools, optimizing its activities, etc.

6. Conclusions

Along with the rapid development of e-commerce, the importance of ECL has been widely recognized by both industries and academics. Nevertheless, ECL is vulnerable that its service quality is frequently threatened by various risks. In order to facilitate the understanding of operational risks in ECL, this paper proposed a data-driven approach to analyze the operation process of ECL, and then developed GMM-based approach to qualitatively analyze the operational risk in ECL.

The contributions of this paper can be concluded as follows. Firstly, this is a pioneer work on quantitatively analyzing the operational risks of ECL using real-life data. It enables a clear understanding of ECL operational risks and facilitates conducting targeted risk mitigation strategies. Secondly, the operational risk measurement method is introduced from the aspect of their adverse effects on operation time. It can adapt to dynamic scenarios with varied threats, and could be extended to many other fields. Thirdly, the GMM based operational risk analysis method is proposed, which could cope with diverse operation time distributions and automatically derive the required operation time of each phase for risk analysis.

In the future, this work can be extended from the following three aspects. Firstly, the causal factors of operational risks in different phases of ECL should be investigated to facilitate the design of effective risk mitigation strategies. Secondly, using the massive e-commerce transaction data, the operational risk prediction approaches could be studied, so that measures can be made in advance to eliminate adverse effects. Thirdly, efforts can be made on building resilient ECL systems to cope with the threats during its execution.

Acknowledgment

This work is supported by National Natural Science Foundation of China under Grant 71804034, New Faculty Start-up Fund of Harbin Institute of Technology, Shenzhen under Grant FB45001022.

References

[1] S.X. Xu, G.Q. Huang, Efficient multi-attribute multi-unit auctions for B2B e-commerce logistics, Prod. Oper. Manage. 26 (2) (2017) 292–304.
[2] F. Qian, S. Jia, e-commerce in China – 2017, Tech. rep., Department of Electronic Commerce and Informatization, Ministry of Commerce of the People's Republic of China, 2018.
[3] M. Zhang, G.Q. Huang, S.X. Xu, Z. Zhao, Optimization based transportation service trading in B2B e-commerce logistics, J. Intell. Manuf. (2016) 1–17.
[4] Q. Lu, N. Liu, Effects of e-commerce channel entry in a two-echelon supply chain: a comparative analysis of single-and dual-channel distribution systems, Int. J. Prod. Econ. 165 (2015) 100–111.
[5] P.-J. Wu, K.-C. Lin, Unstructured big data analytics for retrieving e-commerce logistics knowledge, Telematics Inf. 35 (1) (2018) 237–244.
[6] Q. Yang, X. Zhao, H.Y.J. Yeung, Y. Liu, Improving logistics outsourcing performance through transactional and relational mechanisms under transaction uncertainties: evidence from China, Int. J. Prod. Econ. 175 (2016) 12–23.
[7] J. Acimovic, M.K. Lim, H.-Y. Mak, Beyond the speed-price trade-off, MIT Sloan Manage. Rev. 59 (4) (2018) 12–15.
[8] K.-Y. Hu, T.-S. Chang, An innovative automated storage and retrieval system for B2C e-commerce logistics, Int. J. Adv. Manuf. Technol. 48 (1–4) (2010) 297–305.
[9] F. Chen, G. Xu, Y. Wei, Heuristic routing methods in multiple-block warehouses with ultra-narrow aisles and access restriction, Int. J. Prod. Res. (2018) 1–22.
[10] J. Ruan, Y. Shi, Monitoring and assessing fruit freshness in IOT-based e-commerce delivery using scenario analysis and interval number approaches, Inf. Sci. 373 (2016) 557–570.
[11] S. Shao, G. Xu, M. Li, G.Q. Huang, Synchronizing e-commerce city logistics with sliding time windows, Transp. Res. Part E: Logist. Transp. Rev. (2019) 1–12.
[12] S.X. Xu, M. Cheng, G.Q. Huang, Efficient intermodal transportation auctions for B2B e-commerce logistics with transaction costs, Transp. Res. Part B: Methodol. 80 (2015) 322–337.
[13] R. Ramanathan, The moderating roles of risk and efficiency on the relationship between logistics performance and customer loyalty in e-commerce, Transp. Res. Part E: Logist. Transp. Rev. 46 (6) (2010) 950–962.
[14] P. Xiao, D. Wang, Research on logistics outsourcing risk management in e-commerce enterprise based on agency theory, The Conference on Web Based Business Management, Scientific Research Publishing, 2010, pp. 345–348.
[15] T. Choi, C. Chiu, H. Chan, Risk management of logistics systems, Transp. Res. Part E: Logist. Transp. Rev. (2016) 1–6.
[16] I. Heckmann, T. Comes, S. Nickel, A critical review on supply chain risk–definition, measure and modeling, Omega 52 (2015) 119–132.
[17] C.S. Tang, Perspectives in supply chain risk management, Int. J. Prod. Econ. 103 (2) (2006) 451–488.
[18] J. Chen, A.S. Sohal, D.I. Prajogo, Supply chain operational risk mitigation: a collaborative approach, Int. J. Prod. Res. 51 (7) (2013) 2186–2199.
[19] J.P. Mo, M. Cook, Quantitative lifecycle risk analysis of the development of a just-in-time transportation network system, Adv. Eng. Inf. 36 (2018) 76–85.
[20] R.A. Sutrisnowati, H. Bae, M. Song, Bayesian network construction from event log for lateness analysis in port logistics, Comput. Industr. Eng. 89 (2015) 53–66.
[21] Y. Shang, D. Dunson, J.-S. Song, Exploiting big data in logistics risk assessment via bayesian nonparametrics, Oper. Res. 65 (6) (2017) 1574–1588.
[22] G. Xu, M. Li, L. Luo, C.-H. Chen, G.Q. Huang, Cloud-based fleet management for prefabrication transportation, Enterprise Inf. Syst. (2018) 1–20.
[23] Z. Li, G. Liu, L. Liu, X. Lai, G. Xu, IOT-based tracking and tracing platform for prepackaged food supply chain, Industr. Manage. Data Syst. 117 (9) (2017) 1906–1916.
[24] X. Qiu, H. Luo, G. Xu, R. Zhong, G.Q. Huang, Physical assets and service sharing for IOT-enabled supply hub in industrial park (ship), Int. J. Prod. Econ. 159 (2015) 4–15.
[25] G. Xu, J. Wang, G.Q. Huang, C.-H. Chen, Data-driven resilient fleet management for cloud asset-enabled urban flood control, IEEE Trans. Intell. Transp. Syst. 19 (6) (2018) 1827–1838.
[26] S. Wang, Z. Liu, X. Qu, Collaborative mechanisms for berth allocation, Adv. Eng. Inf. 29 (3) (2015) 332–338.
[27] A. Azadeh, V. Salehi, R. Salehi, S. Hassani, Performance optimization of an online retailer by a unique online resilience engineering algorithm, Enterprise Inf. Syst. 12 (3) (2018) 319–340.
[28] D. Reynolds, Gaussian mixture models, Encyclopedia Biometr. (2015) 827–832.
[29] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, G.-B. Huang, Exploiting AIS data for intelligent maritime navigation: a comprehensive survey from data to methodology, IEEE Trans. Intell. Transp. Syst. 19 (5) (2018) 1559–1582.
[30] F. Attal, A. Boubezoul, A. Samé, L. Oukhellou, S. Espié, Powered two-wheelers critical events detection and recognition using data-driven approaches, IEEE Transact. Intell. Transport. Syst. 19 (12) (2018) 4011–4022.
[31] L. Li, R.J. Hansman, R. Palacios, R. Welsch, Anomaly detection via a gaussian mixture model for flight operation and safety monitoring, Transp. Res. Part C: Emerg. Technol. 64 (2016) 45–57.
[32] A.J. Sonta, P.E. Simmons, R.K. Jain, Understanding building occupant activities at scale: an integrated knowledge-based and data-driven approach, Adv. Eng. Inf. 37 (2018) 1–13.
[33] B. Kang, G. Chhipi-Shrestha, Y. Deng, J. Mori, K. Hewage, R. Sadiq, Development of a predictive model for *Clostridium difficile* infection incidence in hospitals using gaussian mixture model and Dempster–Shafer theory, Stoch. Env. Res. Risk Assess. 32 (6) (2018) 1743–1758.
[34] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the em algorithm, J. Royal Stat. Soc. Ser. B (Methodol.) (1977) 1–38.