Investigating the bulk density of construction waste: A big data-driven approach

Weisheng Lu, Liang Yuan*, and Fan Xue

Department of Real Estate and Construction, Faculty of Architecture, the University of Hong Kong, Hong Kong, China

This is the peer-reviewed post-print version of the paper:

Lu, W., Yuan, L., & Xue, F. (2021). Investigating the bulk density of construction waste: A big data-driven approach. *Resources, Conservation & Recycling*, 169, 105480. Doi: 10.1016/j.resconrec.2021.105480

The final version of this paper is available at https://doi.org/10.1016/j.resconrec.2021.105480.

Abstract

Construction waste contains inert (e.g., construction debris, rubble, earth, bitumen, and concrete) and non-inert materials (e.g., bamboo, plastics, wood, paper, and vegetation), while it is often a combination of the two when it is generated at source. The bulk density of construction waste is the yardstick information for many subsequent waste management efforts. One feasible way to derive the bulk density information is to segregate the mixture of inert and non-inert substances and examine their compositions, but clearly, this is an onerous task. This paper reports a data-driven approach to obtain the bulk densities of inert and non-inert construction waste by analyzing a big dataset of 4.9 million loads of construction waste in Hong Kong in the years 2017 to 2019. It is discovered that the means of bulk density are 336 kg/m^3 for non-inert waste, 528 kg/m^3 for mixed waste, and 991 kg/m^3 for inert waste, and their coefficients of variation are 69%, 43%, and 29%, respectively. The research not only proved our heuristic rules concerning the bulk densities of the three generic types of construction waste, but also articulated, for the first time, their converged means and ranges. The findings can be used in adjusting the admission criteria as adopted in the governmental waste management facilities. Future research is recommended to further narrow down the bulk density ranges to provide more accurate references for construction waste management.

Keywords: Construction waste management; inert waste; non-inert waste; bulk density; big data; data-driven approach

1. Introduction

Construction waste, sometimes also called construction and demolition (C&D) waste, is the solid waste arising from such construction activities as site clearance, excavation, new building, refurbishment, renovation, and demolition (HKEPD, 2019; Lu et al., 2019). In the U.S. or Europe, construction waste is usually classified into specific materials. For example, the U.S. Environmental Protection Agency (EPA, 2018) classifies construction waste into seven groups according to their composition: concrete, steel, wood products, gypsum wallboard and plaster, brick and clay tile, asphalt shingles, and asphalt concrete. The European Waste Catalogue classifies construction waste in line with its compositions into eight categories, including concrete bricks, tiles, ceramics, wood, glass, and plastic (SEPA, 2015). In other economies like the U.K., Australia, or Hong Kong, construction waste is often categorized into two types: inert waste, comprising primarily debris, rubble, soil, bitumen, concrete, and so on; and non-inert waste, comprising bamboo, plastics, wood, paper, vegetation, and so forth (HKEPD, 2019). In any case, construction waste generated at source is usually in mixed material dumps without knowing their detailed compositions or densities.

However, information on the detailed compositions or densities of a waste dump is of significant practical value. For example, the composition information is important for devising different technologies to sort them into different material groups for reusing or recycling (Clancy, 2019; Lu & Yuan, 2012). One also needs to understand the chemical and physical properties associated with specific materials for proper recycling strategies. For example, it needs to calculate their combustion value and emission (e.g., dioxin and furans) if using waste incineration (Eriksson & Finnveden, 2017; The World Bank, 1999), or examine their environmental degradation and nuisance production (e.g., carbon dioxide, methane, and leachate) if for landfilling (Salem et al., 2008; Xu et al., 2019). The overall density of a bulk of waste is also of significant practical value. For example, the UK WRAP (Waste and Resources Action Programme) published a dedicated report to investigate the bulk densities of commonly collected materials (e.g., food waste, mixed paper, cards, or plastic bottles). The information helps "inform the assessment of waste and recycling options and the planning and delivery of collection and recycling services" (WRAP, 2010). Li et al. (2020a) reported a bulk density-based method for recognizing kitchen and dry waste in Beijing. Li et al. (2020b) further investigated the seasonal variation impact on the bulk-densities by applying the 'intelligent supervision trashcan' in various climate areas across China. Bowan & Tierobaar (2014) characterized the composition and bulk density of solid waste in Ghanaian Markets for devising solid waste management strategies and policies. In the construction waste sorting facilities operated by the Environment Protection Department of Hong Kong (HKEPD), whether a load of waste is admittable is dependent on whether the inert substances exceed 50% of the bulk by weight (HKEPD, 2019). The bulk density is the yardstick information underpinning the waste sorting system.

A feasible way to obtain the bulk density information is to measure the weight and volume of a bulk of mixed waste materials and calculate its density. In fact, WRAP (2010) adopted similar approaches (e.g., self-reporting from contractors and researchers, and fieldwork) to measure the densities from containers, kerbsides, stillage vehicles, and so on. Ireland EPA (1996) also reported its approach to measure bulk densities of municipal solid waste, predominantly using weight and volume derived from fieldwork. Li et al. (2020a) collected and measured a sample of 270 bagged household solid waste and analyze their moisture content and bulk density.

Bowan & Tierobaar (2014) spent around two months collecting solid waste samples by placing many waste bins at determined sites, and then these samples were used for estimating the solid waste composition and bulk density. Apparently, these fieldworks are utterly onerous. Another concern is that the calculated results cannot be readily generalized to others with any confidence. Bulk density can be defined as the mass of many particles of the materials divided by the total volume. It is not an intrinsic property of a material. Rather, it depends on the compositional materials, voids, and porosities (Lyon & Buckman, 1922; Mattox, 2010). The combinations of inert and non-inert waste in waste dumps could be infinitive in terms of compositions and volumes. So could be their bulk densities, which are collectively determined by their compositions and volumes. Researchers around the world have endeavored to search for more feasible approaches to obtain the bulk densities of waste materials. Data-driven approaches come to the radar under this background.

Data-driven approaches are popularized in the era of big data. According to Mayer-Schönberger & Cukier (2013), big data has three defining characteristics, namely volume, variety, and velocity, or the three 'Vs'. *Volume* means the quantities of data incoming as terabytes or zettabyes; *velocity* means the data is increasing at a very high speed in batch, near time, real-time, and streams; and *variety* means the data can be structured, unstructured, semi-structured, and a combination thereof to indicate different aspects of a subject (Russom, 2011; Zaslavsky et al., 2013). Big data in the forms of records, transactions, tables, or files is relentlessly generated from such sources as weblogs, sensor networks, social networking, and streaming video and audio. Analytics have been developed to analyze big data to uncover hidden patterns, unknown correlations, and other useful information to guide better business predictions and decision-making that cannot be done in the small data contexts (Shen et al., 2014).

Data-driven approaches are exploratory to analyze big data to extract scientifically interesting insights (Kitchin, 2014). Unlike traditional theory-driven approaches to base the causal link between an intervention and its outcomes on an explicit theoretical model, data-driven approaches might not have an explicit theoretical model or causal link at the beginning. The stream of approaches relies on the big volume of data to inform a causation/pattern that might not be possible in the small data contexts. The major promises of data-driven approaches lie in patterns extracted from the analysis of large data sets, and insights derived from these patterns (Sivarajah et al., 2017). Data-driven approaches can find their theoretical root in probability theory, in particular, the law of large numbers (LLN) (Bernoulli, 1713), which is a theorem asserting that the average of the results obtained from a large number of trials should be close to the expected value and more converged as more trials are performed. Construction waste materials generated from a region are not entirely random in terms of compositions; rather, they are determined by prevailing construction materials, technologies, and recycling levels. If one can obtain the big data of the weights and volumes of C&D waste dumps, he/she might be able to derive a converged, reliable value or range of bulk densities regardless of the overwhelming combinations of the inert and non-inert substances.

The primary aim of this research is to determine the bulk density of construction waste by analyzing a precious big dataset in Hong Kong. Hong Kong's eminent construction activities have built an astonishing skyline and world-class infrastructure. However, they also generated a massive amount of C&D waste per annum, which requires careful management and efficient public policies. The remainder of the paper is organized as follows. Subsequent to this introductory section is Section 2 to describe the big data set on C&D waste obtained from Hong Kong's construction industry. Section 3 describes the methods by deploying both graphical and mathematical approaches. Section 4 reports the data analyses, results, and findings, followed by an in-depth discussion in Section 5. Conclusions are drawn in Section 6, which also proposes directions for future studies.

2. The big data set

The data was obtained from the HKEPD, which launched a Construction Waste Disposal Charging Scheme (CWDCS) in 2006, regulating that all solid waste generated from construction activities, unless being properly reused or recycled, must be disposed of at designated government waste disposal facilities such as landfills, public fills, or off-site sorting facilities. Prior to using the facilities, the responsible party (e.g., a main contractor if the contract worth is larger than HK$1 million, or an individual such as the owner or a small contractor of construction work under a contract with value less than HK$1 million) is mandated to open a billing account in the HKEPD. The billing account database thus retains basic information of all the projects, including the contract name, client, contract sum, site address, type of construction work, and so on. Responsible contractors or individuals who dispose of construction waste at the facilities will be charged a fee depending on the compositions of the waste (see Table 1).

Table 1. Government construction waste disposal facilities and respective charge levels

Government waste disposal facilities	Type of construction waste accepted	Charge per ton (HK$)	
		Before 7 April 2017	After 7 April 2017
Public fill reception facilities	Consisting entirely of inert construction waste	27	71
Sorting facilities	Containing more than 50% by weight of inert construction waste	100	175
Landfills	Containing not more than 50% by weight of inert construction waste	125	200

Source: Adapted from the HKEPD (2020). https://www.epd.gov.hk/epd/misc/cdm/introduction.htm. Access on 9 October 2020

Driven by the CWDCS, construction waste management in Hong Kong is shaped into some common practices, as shown in Figure 1. Even after proper reduction, reuse, or recycling, construction waste is unavoidably generated on various sites. On-site waste segregation (Point A) is highly recommended to sort the mixed waste materials into inert and non-inert portions. Inert waste will be sent to public fills (Point D) for reclamation, site formation, production of recycled aggregates, or other uses. Non-inert waste will be transported to landfills (Point B). When a site is too congested to allow on-site segregation, one can transport the mixed waste to the off-site sorting facilities (Point C) if it contains more than 50% inert substance by weight. Motivated by saving the waste disposal charging fees, one would put efforts to sort the waste into inert or non-inert types. They would also possibly "cheat" at the facilities, e.g., by transporting unqualified non-inert waste to public fills or off-site sorting facilities instead of landfills. Under some circumstances, e.g., to save time or labor cost, one would not bother to sort the waste but just transport it to landfills by paying a higher fee. From the government facility operators' perspective, it is important to make sure that qualified waste is accepted at proper facilities (see Table 1), e.g., by setting up technical gauges and conducting regular inspections.

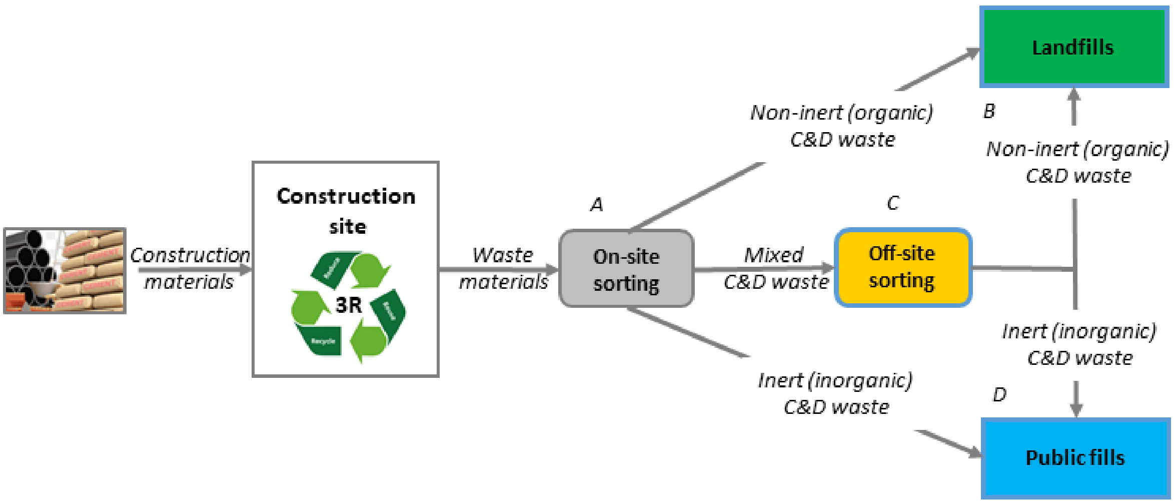


Figure 1. The common process of construction waste management in Hong Kong (Adapted from Lu & Tam [2013])

When construction waste is disposed of at the facilities, the HKEPD records information on every load of C&D waste, including the facility, date, vehicle number, net weight of the waste, the time when the vehicle enters and exits, and the billing account number the vehicle uses. The trucks delivering C&D waste must be registered at the HKEPD in a separate database, which contains the plate numbers and permitted gross vehicle weight (PGVW). An excerpt of the database can be perceived in Figure 2. Unintentionally, this practice generates a large secondary dataset, which makes it possible to probe into various aspects pertinent to construction waste management. The data covers the nine waste disposal facilities categorized into three types, namely landfills, public fills, and off-site sorting facilities, which receive qualified C&D waste, as shown in Table 1. We collected nine years' data ranging from 2011 to 2019 from the HKEPD's theme website, which publishes updated data every fortnight after made some necessary pseudonymization. The data contains more than 1 million highly structured records per annum, which means more than 1 million loads of waste are disposed of at the facilities yearly. The data covers various items, including facility names, vehicle plate numbers, PGVW, waste weight and depth, disposal date, times the lorries entering and leaving the facilities, and so on. The data is incoming more than 3,000 records per day. According to the 3Vs as elaborated above, clearly, the data is qualified as big data, although its volume is not as big as terra- or zetta-bytes. Unlike the fieldwork conducted by UK WRAP (2010) or Ireland EPA (1996) to manually measure the volumes and weights of municipal solid waste, the practice in Hong Kong generates a large set of secondary data for examining the bulk density of construction waste.

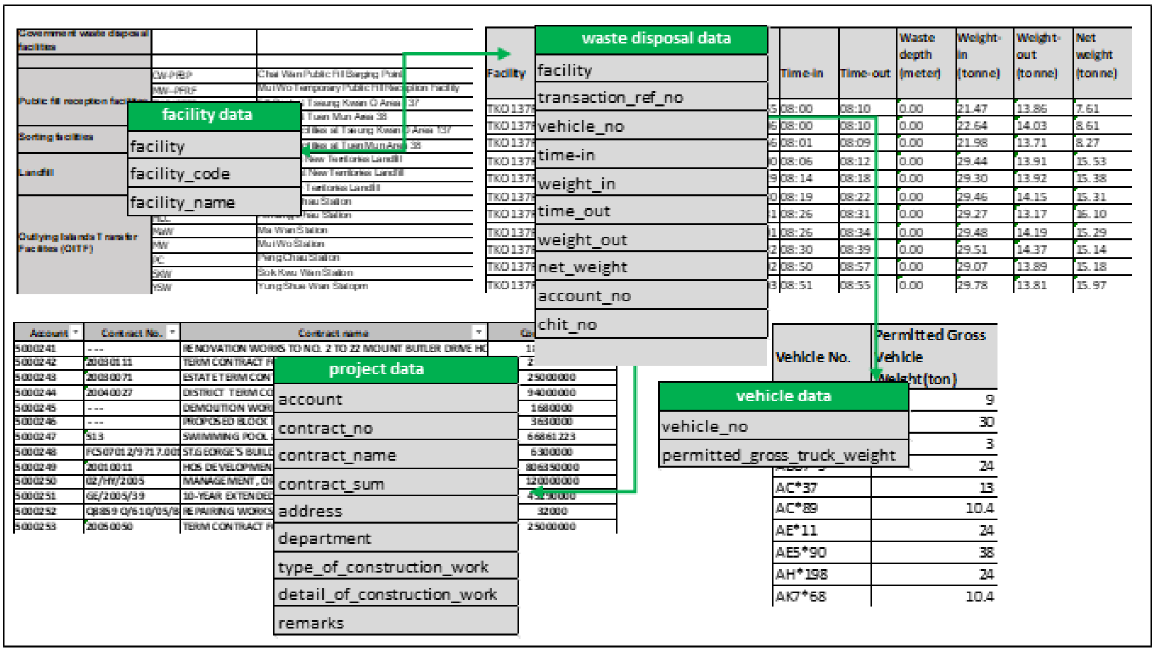


Figure 2. The big dataset

3. Methodology

Bulk density is defined as the mass of many particles of the material divided by the total volume they occupy, and the total volume includes particle volume, inter-particle void volume, and internal pore volume (Lyon & Buckman, 1922; Mattox, 2010). Unlike the true density of materials, which refers to the actual mass of a solid substance per unit volume (e.g., m^3) in an absolutely dense state, bulk density is not the intrinsic property of materials. It can change in line with the compositions of the materials and the voids. In either case, the density can thus be calculated by using Equation (1) below:

$$\rho = W/V \tag{1}$$

where ρ is the density, W is the weight of a particular dump of waste contents, and V is the volume.

Table 2 shows the true densities of construction materials that are often seen in the C&D waste dumps. An observation is that, generally, inert construction materials are of higher densities than the non-inert counterparts. This is particularly true in view of the fact that alloy and steel materials are quickly salvaged on-site without going to disposal. Common sense is that the bulk density of a dump of materials should be smaller than the true density of the dominant solid substance (e.g., concrete, bitumen, timber, or wood). In Hong Kong's practices, inert waste materials are disposed of at public fills (Point D in Figure 2); non-inert waste materials are disposed of at landfills (Point B); and the mixed materials at off-site sorting facilities (Point C), with some caveats of ignorance or disguising cases as mentioned in the above section. Drawing upon all these rationales, it would be legitimate to assume the following Inequation (2):

$$\bar{\rho}_{BD-inert} > \bar{\rho}_{BD-mixed} > \bar{\rho}_{BD-non-inert} \quad (2)$$

which means the average bulk density of inert materials ($\bar{\rho}_{BD-inert}$) is larger than that of mixed materials ($\bar{\rho}_{BD-mixed}$), and in turn, the average bulk density of mixed materials is larger than that of the non-inert materials ($\bar{\rho}_{BD-non-inert}$). It is also reasonable that the individual bulk densities could overlap with each other for two causes. One is the variable inter-particle voids between materials. Another is the true densities of different materials originally have overlap. As shown in Table 2, the true densities of plastic (non-inert) and bricks (inert) range from 913 to 2,159 kg/m^3 and from 1,500 to 1,800 kg/m^3, respectively. Other than these heuristic rules, we do not know the average bulk densities with any precision. The void volumes are infinite for the C&D waste consisting of either only one or more materials. Nevertheless, according to the Law of Large Numbers (LLN) (Bernoulli, 1713), when the data of bulk density is big enough, it is possible to indicate a converged bulk density, or some significant patterns which may provide clues for estimating the bulk density. A graphic illustration of the rationale behind the methodology is illustrated in Figure 3.

Table 2. The true density of common construction materials

Inert construction material	True density (kg/m^3)	Non-inert construction material	True density (kg/m^3)
Masonry	650~2,100	Wood	160~1,310
Asphalt	721	Paper	700~1,150
Cement	1,440	Leather	860
Bricks	1,500~1,800	Rubber	910~1,200
Rocks	1,600~3,500	Plastics	913~2,159
Sand	1,631	Bamboos	1,160

Lime mortar	1,760	Wool	1,314
Soil	1,800~2,000	Textile	1,560
Tiles	1,800~2,200	Aluminum alloy	2,640~2,810
Bentonite	2,200~2,800	Titanium alloy	4,429~4,512
Concrete	2,400~2,500	Steel	7,750~8,050
Glass	2,400~2,800	Stainless steel	7,850~8,060

Source: Adapted from the Engineering Toolbox. https://www.engineeringtoolbox.com/density-solids-d_1265.html Access on 9 October 2020

Figure 3. The rationale behind the methodology of this study

Data-driven research uses exploratory approaches to analyze big data to extract scientifically interesting knowledge (Kitchin, 2014), such as patterns underneath the large data sets, and insights derived from these patterns. Researchers (Jagadish, 2015; Shmueli & Koppius, 2011) describe the research as an iteration of the following steps: (1) identifying research questions; (2) creating/obtaining sources of data; (3) cleansing, extracting, annotating data streams to prepare for analyses; (4) integrating, aggregating, and representing data; (5) analyzing and modeling data; and (6) interpreting the patterns to arrive at solutions and insights. The big data-driven approach as adopted in this paper is developed by largely following these suggested steps.

4. The data-driven approach

Figure 4 presents the flow diagram of the data-driven approach. It includes five sections: Data sensing, cleansing, processing, analysing, and visualizing. They will be introduced in details.

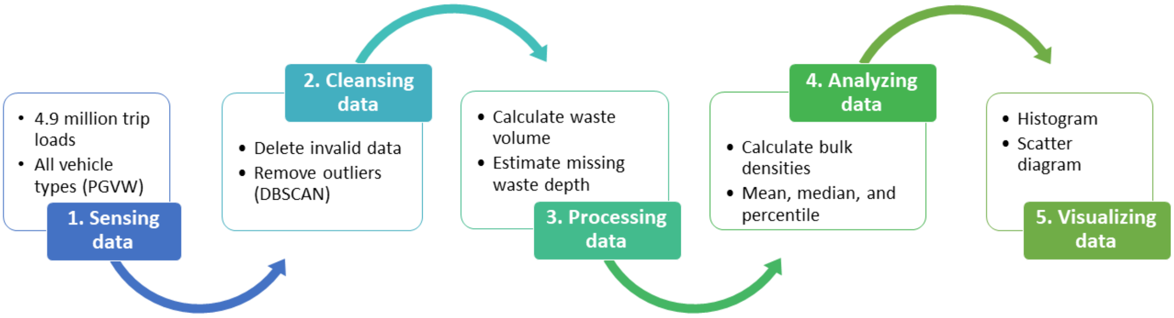


Figure 4. The flow diagram of the data-driven approach

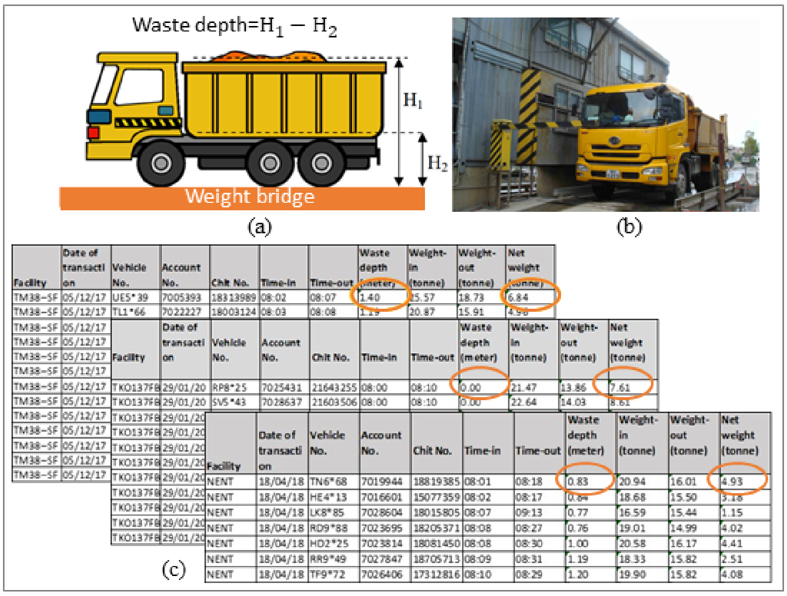
4.1 Sensing the data

We sourced the raw data of three years in 2017, 2018, and 2019. There are 4.9 million such records, including 1,178,427 from landfills, 468,961 from sorting facilities, and 3,280,550 from public fills, respectively, meaning that every year there around 1.64 million loads of C&D waste were disposed of at the various governmental waste management facilities. It is noticed that in the data set (see Figure 5), the landfills and off-site sorting facilities recorded the 'net weight' and the 'height' of each waste load. The net weight in the database means the net weight of a load of C&D waste that has been dumped in a waste disposal facility. It is calculated by weighing the vehicles at the in-weigh and out-weigh bridges and subtracting the two (see Figure 5b). The waste depth is determined by a method, as shown in Figure 5a. A set of sensors are installed above the in-weigh bridges to capture and calculate the waste depth.

The net weight is captured in all three types of facilities, as it is used for calculating the charges and preventing overloading. According to the CWDCS, if the total weight of a truck at the in-weigh bridges exceeds its PGVW by less than 5%, the truck can still be allowed in but will receive an overloading notice. For unknown reasons, the 'waste depth' is only captured in the off-site sorting facilities and landfills but not the public fills. If the missing data can be made up in a reasonable way, it is possible to covert the 'waste depth' into the 'volume' of a waste load by considering the bottom area of the truck's loading bucket, and calculate the bulk density of each waste load using Equation (3) below:

$$\rho_{BD} = \frac{W}{V} = \frac{W}{H \times A} \tag{3}$$

where ρ_{BD} is the bulk density of a waste dump, W is the net weight of the waste load, H is the height of the waste load, and A is the bottom area of the loading bucket.



Facility	Date of transaction	Vehicle No.	Account No.	Chit No.	Time-in	Time-out	Waste depth (meter)	Weight-in (tonne)	Weight-out (tonne)	Net weight (tonne)
NENT	18/04/18	TN6*68	7019944	18819385	08:01	08:18	0.83	20.94	16.01	4.93
NENT	18/04/18	HE4*13	7016601	15077359	08:02	08:17	0.84	18.68	15.50	3.18
NENT	18/04/18	LK8*85	7028604	18015805	08:07	09:13	0.77	16.59	15.44	1.15
NENT	18/04/18	RD9*88	7023695	18205371	08:08	08:27	0.76	19.01	14.99	4.02
NENT	18/04/18	HD2*25	7023814	18081450	08:08	08:30	1.00	20.58	16.17	4.41
NENT	18/04/18	RR9*49	7027847	18705713	08:09	08:31	1.19	18.33	15.82	2.51
NENT	18/04/18	TF9*72	7026406	17312816	08:10	08:29	1.20	19.90	15.82	4.08

Figure 5. An illustration of the methodology to capture the net weight and waste depth

The 4.9 million trip loads of waste were delivered by various types of waste hauling trucks, each with a PGVW ranging from merely 2.8 to 38 tons. Figure 6 illustrates the numbers of trip loads undertaken by each type of trucks with distinct PGVWs. It can be seen from Figure 6 that there are 21 types of trucks with different PGVWs in operation. The majority (96.4%) of the trip loads are delivered by five types of trucks with PGVWs of 9-ton, 16-ton, 24-ton, 30-ton, and 38-ton. We will, therefore, focus on these types of trucks and their transported C&D waste in the following analyses.

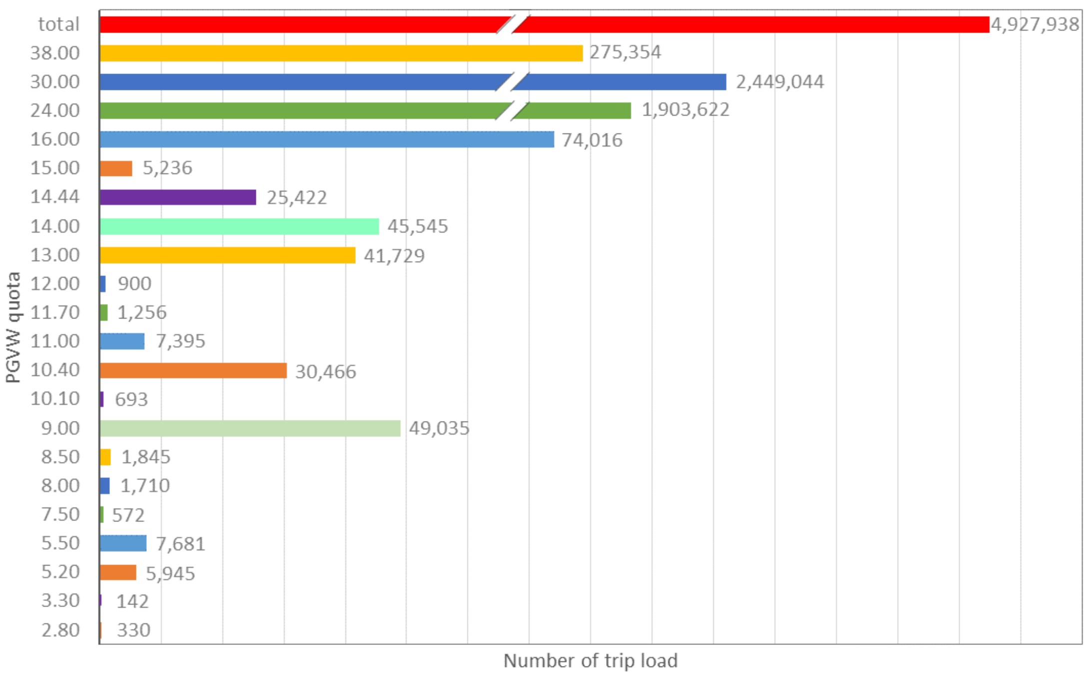


Figure 6. The C&D waste hauling trips conducted by various types of trucks (from 2017 to 2019)

4.2 Data cleansing

It is noticed that some of the data is apparently unreasonable. For example, some of the net weights are as high as 12 tons in the case of H being 0.1 m in a 16-ton truck. It means that the bulk density reaches approximately 10,900 kg/m^3, which even exceeds the maximum true density of stainless steel (8,060 kg/m^3). Proper data cleansing is thus conducted before processing it any further. Two primary steps are adopted. The first is to delete the invalid data by examining the following five criteria: (1) missing waste depth; (2) missing net weight; (3) missing PGVW; (4) net weight smaller than 0 ton; and (5) waste depth smaller than 0 m. The second step is to remove outliers in the waste depth and waste net weight data resulting from measurement errors, sloppy operations (e.g., the staff may simply input a 9,999 kg), or other unknown reasons. Whether a data point being outlier should be made based on the combination of waste depth and net weight, although some depth or net weight value separately are considered reasonable.

The input data of outlier removal is a two-dimension matrix consisting of waste depth and net weight. A total of 15 sub-datasets, including three waste types multiplying by five PGVW quotas, were treated as input data for outlier removal. This research adopted the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) model developed by Ester et al. (1996), which is a popular method for two-dimension dataset outlier removing, to detect and remove outliers. Epsilon Neighborhood (EN), which is specified as a numeric scalar that defines a neighborhood search radius around a core point, and Minimum Number of Neighbors Required for Core Point (MNNRCP) are the two critical parameters of DBSCAN model. The constraint relationship between the two parameters is that the EN of a core point in a cluster must contain at least MNNRCP neighbors. Figure 7 shows an example of removing outliers from a sub-dataset of non-inert C&D waste transported by 9-ton trucks. To select a suitable value for MNNRCP, it is required that the selected value should not be lower than the dimension number of the input dataset (n) plus one (i.e., $n + 1$). Using the two-dimension matrixes as input data, the least alternative MNNRCP value is three in this research. However, taking the computer calculation load into consideration, this research selected 50 as the MNNRCP value. One recommended strategy for estimating a value for EN is to generate a k-distance graph for the input dataset. For each point in the dataset, to find the distance to the k^{th} nearest point and plot sorted points against this distance, a k-distance graph that contains a knee interval can be obtained, as shown in Figure 7 (a). The knee interval [$P1$, $P2$] is an estimated region where data points start tailing off into outlier territory. In other words, the border of normal points and outlier points is an interval rather than a unique value. The longitudinal coordinate value D_i of Figure 7 (a) that corresponds to the knee interval is generally a good choice for the EN value. In the shown example, EN can be any value that belongs to the 50$^{\text{th}}$ nearest distances interval [$D1$, $D2$]. If $D1$=0.12 is selected as the EN value, it means 11,719 of 12,454 data points are normal points, and the rest 735 are outliers. The outlier border will be increasingly loosened when the EN moves from *D1* to *D2* as demonstrated in Figure 7 (b). To make the bulk density interval more convergent, this research uniformly selected *D1* as the EN value.

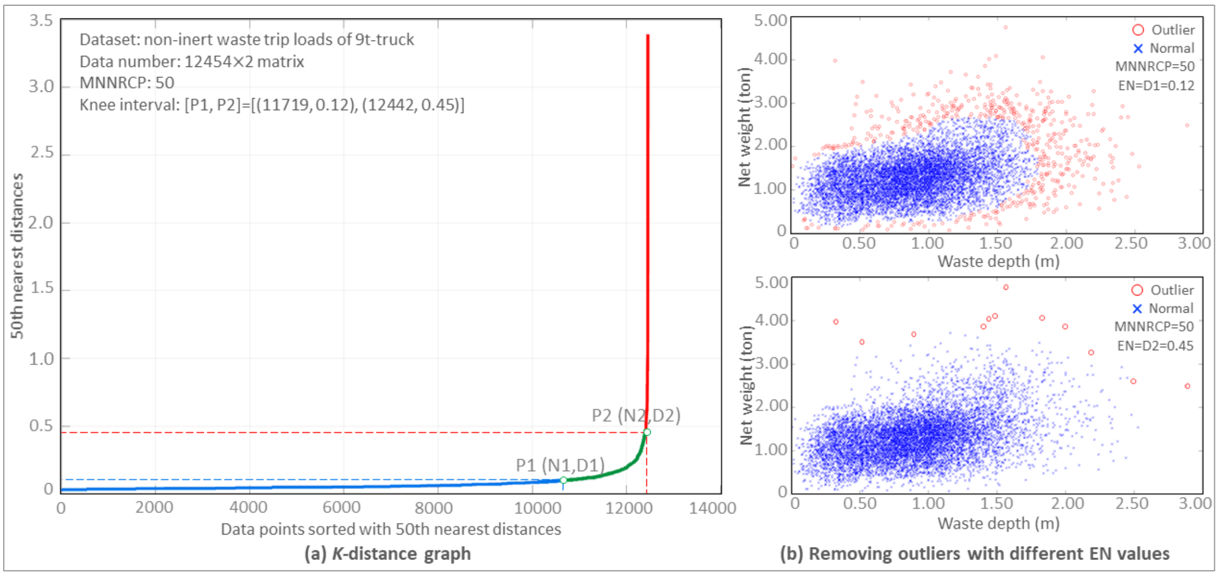


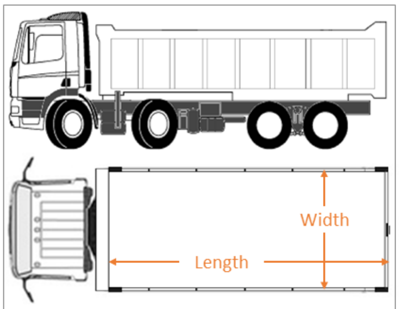
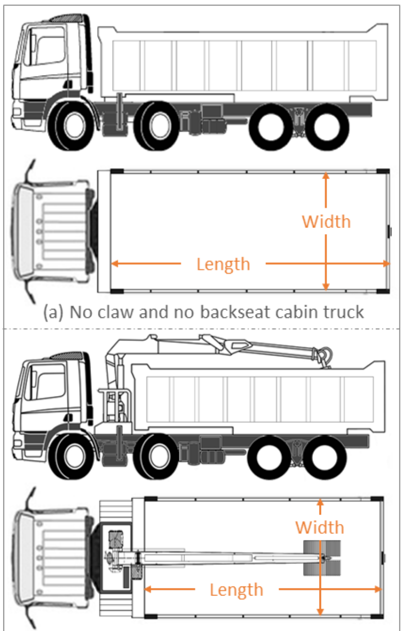
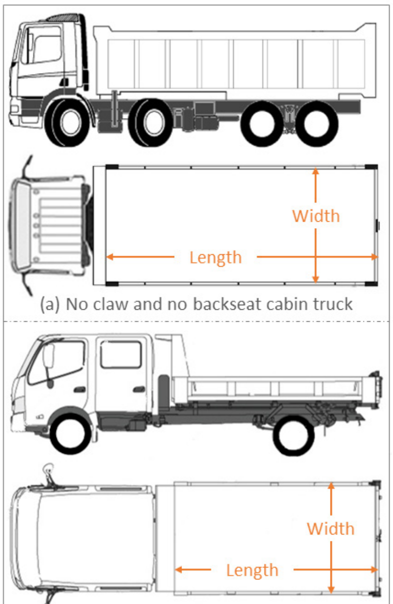
Figure 7. Removing outliers from the non-inert C&D waste data of 9-ton trucks

The approach has also been implemented to other 14 sub-datasets in removing outliers. Owing to the page limit, they are not elaborated here. This step removed 26,281, 6,987, and 66,781 outlier points from non-inert, mixed, and inert waste materials, respectively.

4.3 Calculating the volume of each waste load

According to Equation (3), the bulk density is calculated by using the recorded waste weight (W) and volume (V). The latter is further calculated by multiplying waste depth (H) and the bottom area (A). However, in current practice, only at landfills and off-site sorting facilities, they record waste depth; in public fills, no waste depth information is recorded altogether. In any case, there is no bottom area of the buckets recorded. We have assumed that the bottom areas are uniform across different types of trucks with different PGVW. However, we discovered that they are not uniform because of different styles, as presented in Table 3. Even within the same PGVW trucks, their bottom areas are different. Hence, we searched the official websites of several representative truck manufacturers (e.g., FUSO, ISUZU, and HINO) serving Hong Kong to obtain the vehicle dimensions. We also spent a significant amount of effort to collect the data from various truck owners, contractors, and service providers. Some of the details can be seen from Table 3. In the end, it is assumed that the bucket bottom area (A) will range in the intervals, as shown in the last column of Table 3.

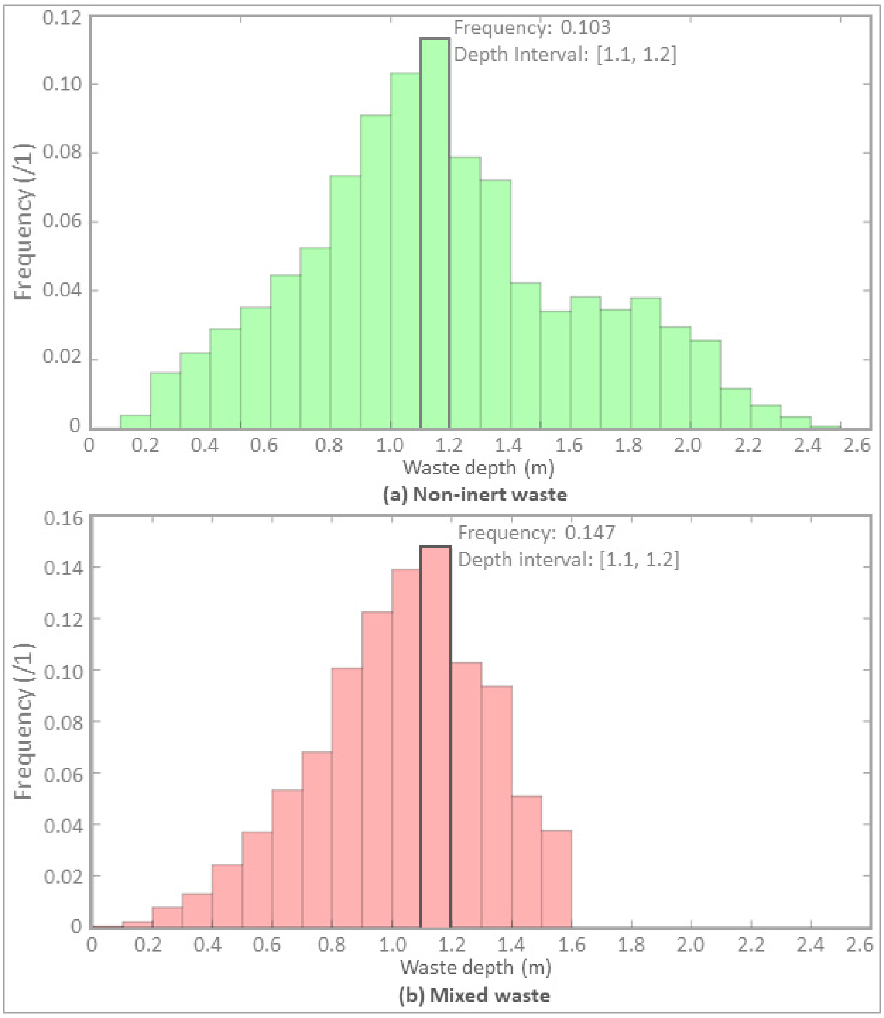
Table 3. The bucket bottom area of several typical trucks

Truck types by PGVW	Truck styles	Length×width (mm)	Bucket bottom area (m^2) [min, max]
38t	 (a) No claw and no backseat cabin truck	5200×2300 5200×2480 6380×2480	[11.960, 15.822]
30t	 (a) No claw and no backseat cabin truck (b) With claw but no backseat cabin truck	6095×2255 6095×2440 4800×2300	[11.040, 14.874]
24t		5480×2255 5480×2440 6095×2255 4800×2250 5180×2440	[10.800, 13.744]
16t		4880×2440 5480×2255 4570×1980 4880×2255	[9.049, 12.357]
9t	 (a) No claw and no backseat cabin truck (c) No claw but with backseat cabin truck	3960×1970 3960×2130 3660×2130	[7.796, 8.435]

4.4 Finding the missing depths of waste load in public fills

As mentioned above, the depth of the inert C&D waste received at the public fills has not been measured, which makes the volume information absent for calculating their bulk densities. However, their bulk densities should not be ignored, as the 3.1 million loads of waste received there occupied 63% of the total 4.9 million waste loads. Neither it is possible to re-measure the depth of the 3.1 million loads of C&D waste dumped. In the face of the difficulties, a hypothesis is made that the trucks will deliver similar depth of waste to what they do in the other two types of facilities, namely landfills and off-site sorting facilities. In real life, it is a matter of truck drivers' 'rule of thumb' to determine the depth of waste allowed to their trucks to avoid overloading or underloading.

We plot the frequencies of various waste depths in the two types of facilities. Figure 8 indicates surprisingly that both non-inert and mixed C&D waste present the same highest frequency in the depth interval ranging from 1.1 m to 1.2 m. This is the biggest serendipity of this data-driven approach. According to this result, it is confident to estimate the highest probability interval waste depth as H_p = [1.1, 1.2] m for inert waste as received at public fills. The estimated waste depth interval will be used to calculate the highest probability bulk density of inert C&D waste as received in the facilities.



* The cut-off value of 1.6m in (b) is owing to the fact that only waste loads with a waste depth <1.6m will be accepted

Figure 8. The waste depth frequency distribution histogram

4.5 Data analyses and visualization for calculating the bulk density

After the above efforts on data processing, a new dataset for the bulk density calculation of three types of C&D waste is established as presented in Figure 9.

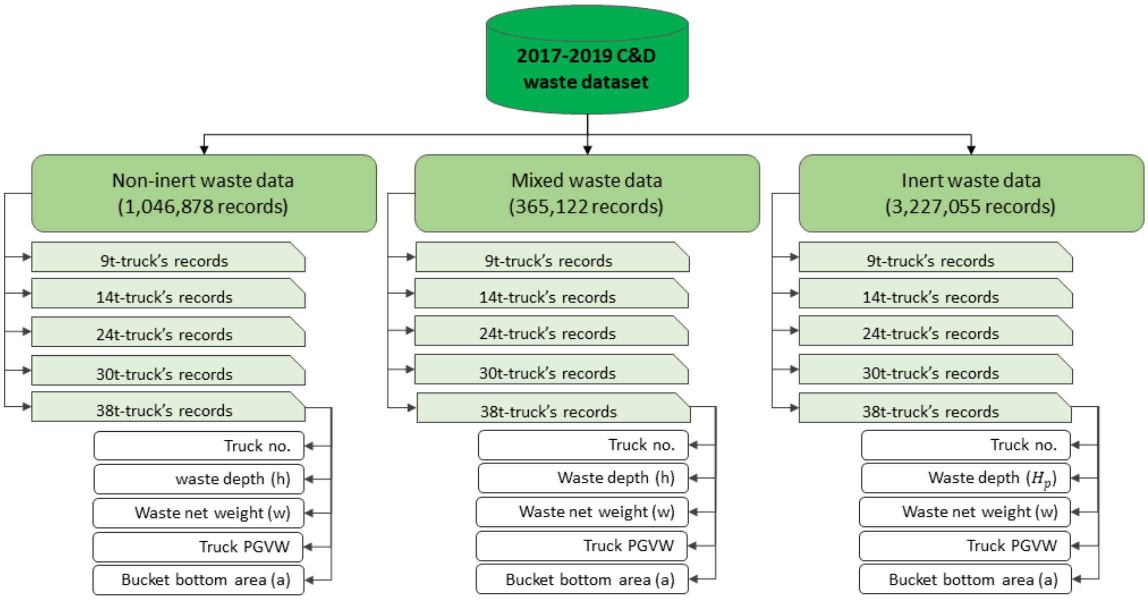


Figure 9. The established C&D waste dataset towards bulk density calculation

As for non-inert C&D waste, each data point has a unique waste depth value and waste net weight value, but the bucket area is an interval. Under this case, we first calculated the upper limit interval of bulk density (ρ_{upper}) when the trucks' bucket bottom areas are the minimum recorded in Table 3 (A_{min}) according to Equation (4):

$$\rho_{upper,\ i} = W_i/(H_i \times A_{min,\ i}) \tag{4}$$

where W_i is the net weight of each waste load; H_i is the waste depth; i is the number of non-inert C&D waste trip loads added by us for differentiation.

Then, the lower limit interval of bulk density (ρ_{lower}) when the trucks' bucket bottom areas being the maximum (A_{max}) was calculated according to Equation (5):

$$\rho_{lower,\ i} = W_i/(H_i \times A_{max,\ i}) \tag{5}$$

After obtaining the two bulk density intervals ρ_{upper} and ρ_{lower}, we used Merge Sort algorithm (Mehlhorn, 2013) to merge them. Based on the merged bulk density interval (ρ_{mg}), the mean value ($\bar{\rho}$), median value ($\rho_{0.5}$), 1% to 99% percentile interval ($\rho_{[1\%,99\%]}$), 5% to 95% percentile interval ($\rho_{[5\%,95\%]}$), and 10% to 90% percentile interval ($\rho_{[10\%,90\%]}$) of the bulk density were calculated. The bulk densities of mixed C&D waste can also be calculated by repeating the above procedures.

As for inert C&D waste, even though both the waste depth and the bucket area are intervals, the calculation method is the same as the other two types. We first calculated the upper limit interval of bulk density when both the bucket area and the waste depth are minimum values and then calculated the lower limit interval of bulk density when both the bucket area and the waste depth are maximum values. The other statistical values such as $\bar{\rho}$, $\rho_{0.5}$, $\rho_{[1\%,99\%]}$ can also be obtained after merging the two bulk density intervals.

Table 4 lists the different statistical values of the bulk density of different C&D waste obtained by using the data-driven approach. As shown in the table, the bulk densities of non-inert C&D waste, mixed C&D waste, and inert C&D waste respectively range from 39 kg/m^{3}to 2,434 kg/m^3, 146 kg/m^3 to 2,787 kg/m^3, and 207 kg/m^3 to 2,435 kg/m^3. This result shows surprisingly that the upper limit of bulk densities of three types of waste are rather close and comparable. It implies some loose waste disposal practices. Some contractors or waste haulers just dump waste in landfills, although the waste can be dumped at public fills or off-site waste sorting facilities to save levies. The result also illustrates that three types of C&D waste's bulk densities have a lot of overlap with each other. The inert and non-inert substances can be better separated for final disposal. The bulk density mean value presents a significant increasing trend from non-inert C&D waste (336 kg/m^3) to mixed C&D waste (528 kg/m^3), and in turn, to inert C&D waste (991 kg/m^3). This result verifies the proposition as shown in Inequation (2).

Table 4. The different statistics of bulk densities of three types of C&D waste

Bulk density	Non-inert C&D waste (kg/m^3)	Mixed C&D waste (kg/m^3)	Inert C&D waste (kg/m^3)
ρ_{lower}	[39, 1,656]	[146, 2,107]	[207, 1,807]
ρ_{upper}	[45, 2,434]	[158, 2,787]	[259, 2,435]
ρ_{mg}	[39, 2,434]	[146, 2,787]	[207, 2,435]
$\bar{\rho}$	336	528	991
$\rho_{0.5}$	287	476	949
$\rho_{[1\%,99\%]}$	[63, 1,280]	[220, 1,296]	[414, 1,536]
$\rho_{[5\%,95\%]}$	[98, 717]	[266, 971]	[564, 1,438]
$\rho_{[10\%,90\%]}$	[124, 571]	[297, 826]	[676, 1,414]
SD	231	227	286
CV	69%	43%	29%

A similar trend can also be observed in the median values of three bulk densities. To analyze the bulk densities further, three percentile intervals were introduced. Taking non-inert C&D waste as an example, its bulk density interval of 1% to 99% percentile is between 63 kg/m^3 and 1,280 kg/m^3. It represents that approximately 1% non-inert C&D waste's bulk densities do not exceed 57 kg/m^3, and approximately 99% non-inert C&D waste's bulk densities do not exceed 1341 kg/m^3. The 1% to 99% percentile intervals of mixed C&D waste and inert C&D waste are from 220 kg/m^3 to 1,296 kg/m^3 and from 414 kg/m^3 to 1,536 kg/m^3 respectively.

For more details, the bulk density intervals between 5% and 95% percentile interval and between 10% and 90% percentile interval were also given in Table 4.

Table 4 also lists the standard deviation (SD) of three types of waste's bulk densities. It is 231 kg/m^3 for non-inert C&D waste, 227 kg/m^3 for mixed C&D waste, and 286 kg/m^3 for inert C&D waste. The coefficient of variation (CV), also known as relative standard deviation, is a criterion for measuring and comparing the dispersion degree of a probability distribution or frequency distribution. With a CV of 69%, the bulk density dispersion degree of non-inert C&D waste is larger than mixed C&D waste counterparts (43%), and in turn, the bulk density of mixed C&D waste is more dispersed than the bulk density of inert C&D waste, which has a CV of 29%.

Referring back to Table 2, the true density of inert construction materials approximately ranges from 650 kg/ m^3 (masonry) to 3,500 kg/ m^3 (rocks), and the true density of non-inert construction materials approximately ranges from 160 kg/m^3 (wood) to 8,060 kg/m^3 (stainless steel). It can be theoretically derived that the true density of mixed construction materials is between 160 kg/m^3 and 8,060 kg/m^3. Figure 10 presents the bulk density intervals of three types of C&D waste and the true density range of common construction materials. It can be found that the bulk densities of three types of C&D waste comply with the heuristic rule. The results of bulk densities using a big-data approach are reasonable and acceptable.

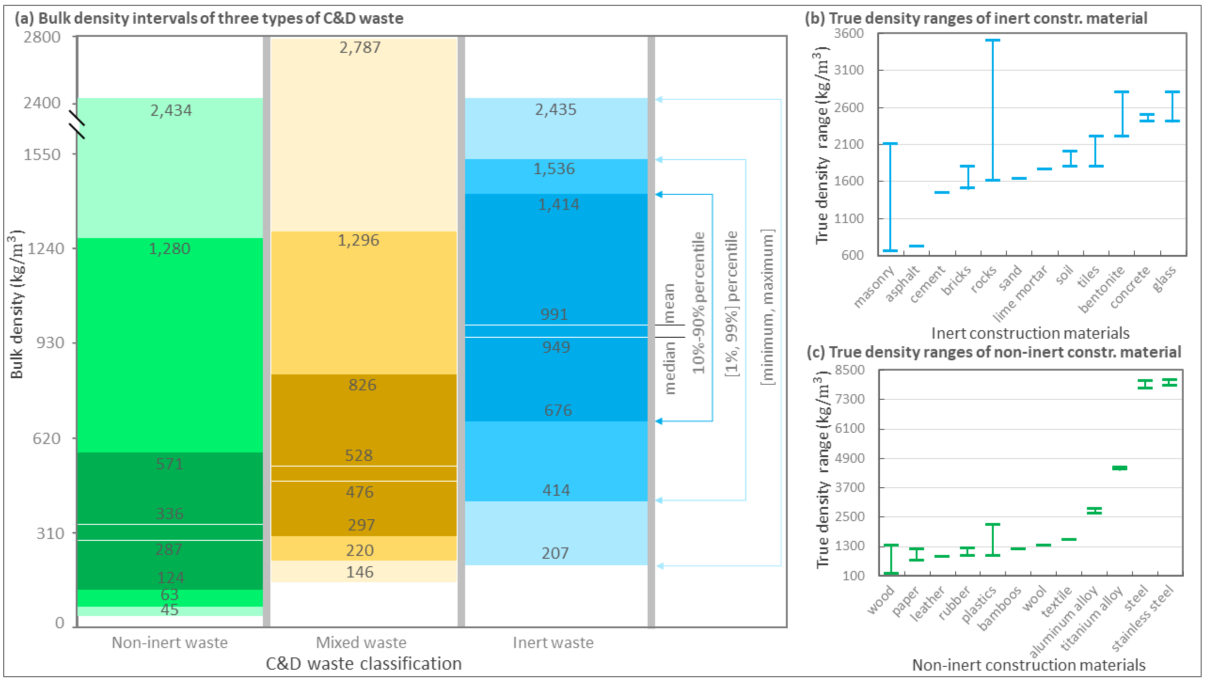


Figure 10. The bulk density of C&D waste and the true density of common construction materials

Discussions

Unlike previous studies concerning waste bulk density, this research presents a totally different approach that is motivated by the availability of a large set of secondary data. The big data-driven approach demonstrates a novel and powerful tool for scientific investigation. The research contributes to the following aspects concerning big data analytics and waste management.

Firstly, the power of big data lies in its volume, velocity, and variety, which can instigate value (e.g., patterns, insights, and knowledge) that may not be achieved in a small data context. For example, it plays an indispensable role in ruling out the outliers, finding the missing waste depth of construction waste loads, and finally, informing the reliable intervals of bulk densities of different types of C&D waste. The big data indicated the dominant types of waste hauling trucks to allow us to better use our research efforts in identifying the bucket bottom area. By discovering no statistically significant difference of waste depth as recorded in either landfills or off-site sorting facilities, the big data analytics proved our assumption that waste haulers based mainly on their experiences in determining the depth of a truckload. The range of [1.1, 1.2] m derived from the two types of facilities are readily transferred to make up the missing information in the third type of facilities. As Anderson (2008) put it, "with enough data, the numbers speak for themselves". This study shows that the big data allows useful patterns to come up even with the use of some simple analytics and visualization only.

Secondly, the case vividly illustrates the Law of Large Numbers (Bernoulli, 1713) in probability theories. The first glance of the waste bulk density problem seems to be an impossible mission as any waste, be it curbside solid waste (EPA, 1996; WRAP, 2010), kitchen waste (Li et al., 2020a), or construction waste (Lu & Yuan, 2011), is a heterogeneous mixture that is not formed by uniformed compositions in an absolutely dense state. Nevertheless, the heuristic rule is that waste is not generated randomly but conformed to certain conditions such as prevailing food structure, living habits, construction materials, or construction technologies. Therefore, the waste bulk density problem should follow the Law of Large Numbers and show some conformity. The big data is almost a full coverage of waste loads received at various facilities. It is able to paint a fuller picture of the subject matter to allow the insights of interest to surface.

Thirdly, although there are some generic steps of a big data-driven approach, such as data collection, extraction, cleansing, analysis, and interpretation; it should be pointed out that there is no one-size-fit-for-all approach for big data analytics. There is no advanced, fascinating analytics such as pattern finding algorithms, attended or unattended machining learning, or the like involved in this study. Lu et al., (2018) argued it would constitute a form of misunderstanding to assume that big data analytics only counts sophisticated data mining techniques without considering traditional functional applied statistics (Leek, 2014). That said, future studies are encouraged to mobilize powerful data analytics such as machine learning or the like to exploit the power of big data. In any case, having domain knowledge and asking the right questions is critical to harness the power of big data. Visualization, as shown in this study, is a powerful approach in parallel with data analytics.

One may argue, which is true, that the big data and its analytics are confined in Hong Kong only, and therefore, the research cannot be readily generalized to other settings with different economies or construction characteristics. Nevertheless, this research illustrates an example that some big datasets leftover unintentionally when businesses are done (Ekbia et al., 2015) are like buried treasure, which can be exploited to derive useful insights. This research provides an example to encourage researchers to explore big data in their respective domains consciously.

The converged ranges of bulk densities of C&D waste derived from the big data analytics, albeit confined to Hong Kong's construction context, are of important referential uses. For example, the average bulk density of organic construction waste (i.e., 336 kg/m^3) is comparable with the average bulk density of food and garden waste (i.e., 338 kg/m^3) as reported by WRAP (2010). By examining some sample waste dumps and referring to the prevailing construction materials, it is possible to associate the bulk density range with main waste materials so as to estimate the compositions of such C&D waste. The estimated result can serve for the following waste sorting works, such as sorting the plastic, paper, and timber out from the mixed construction waste bulk. Currently, the segregation of inert and non-inert waste when it is generated on construction sites is highly recommendable. The large overlaps of the three ranges of bulk densities mean better segregation, e.g., more separated inert and non-insert waste, can be done, although in reality, one will also consider the labor cost, time constraints, and other factors. Lastly, the bulk densities of inert and non-inert construction waste present a significant difference, which can be used to develop more effective admission criteria as adopted in the licensed waste management facilities.

Conclusions

Construction waste, when it is generated at source, usually contains inert materials, non-inert materials, or a mixture of the two. Owing to the infinite combinations of the materials and their voids, the bulk density of construction waste, albeit important and meaningful, has never been calculated with any precision. Using a series of data-driven approaches, this research, for the first time, articulated that the average bulk density is 991 kg/m^3 for inert construction waste, 336 kg/m^3 for non-inert construction waste, and 528 kg/m^3 for mixed construction waste, all in Hong Kong's context. This research also reported a range of minimum and maximum bulk density of [564, 1,438] kg/m^3 for inert, [98, 717] kg/m^3 for non-inert, and [266, 971] kg/m^3 for mixed construction waste, all with a 5% to 95% percentile interval. The findings proved the heuristic rules that inert construction waste materials, in general, are denser than their non-inert counterparts owing to the main substances they contained, and mixed materials situated in the middle of the bulk density spectrum. The bulk densities can be used in gauging whether a truck load of C&D waste is qualified and admittable in Hong Kong's off-site construction waste sorting facilities. Segregation at source is a highly recommendable strategy for waste recycling. The large overlaps between the different groups of waste imply that there is room for clearer sorting of inert and non-inert materials from the dumps.

The big data-driven approach showed its power for scientific research. The approach is found indispensable in informing almost every key subject matter in this research, e.g., outliers, dominant types of trucks in operation, missing waste depth, and, ultimately, bulk density of C&D waste. The big data is able to portray a fuller picture of the subject matter to allow a stronger claim to the objective truth. In addition, the big data speaks for itself. By following the Law of Large Numbers in probability theory, the big data, with proper analytics and visualization, allows interesting patterns or insights to surface. There are some generic steps for big data-driven approaches. However, there is no one-size-fit-for-all approach to exploit big data in different domains. No fascinating big data analytics have been adopted in this study. However, future studies by using advanced algorithms such as machine learning, and supervised or unsupervised learning, are highly recommended to make use of the C&D waste big data. It is also important to ask the right questions to harness the power of big data.

Acknowledgment

This research is jointly supported by the Strategic Public Policy Research (SPPR) (Project No.: S2018.A8.010) Funding Schemes and the Environmental Conservation Fund (ECF) (Project No.: ECS Project 111/2019) of the Hong Kong SAR Government.

References

Anderson, C. (2008). *The end of theory: the data deluge makes the scientific method obsolete*. WIRED. https://www.wired.com/2008/06/pb-theory/

Bernoulli, J. (1713). Ars conjectandi: Usum & applicationem praecedentis doctrinae in civilibus. *Moralibus & Oeconomicis*, *4*, 1713.

Bowan, P. A., & Tierobaar, M. T. (2014). Characteristics and management of solid waste in Ghanaian markets-A study of WA municipality. *Civil and Environmental Research*, *6*(1), 114–119.

Clancy, H. (2019). *How artificial intelligence helps recycling become more circular*. GreenBiz. https://bit.ly/3268XyP

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S., & Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, *66*(8), 1523–1545. https://doi.org/https://doi.org/10.1002/asi.23294

EPA. (1996). *Municipal waste characterisation*. https://bit.ly/3ekrn3J

EPA. (2018). *Construction and demolition debris generation in the united states*. https://bit.ly/2JB2Z2B

Eriksson, O., & Finnveden, G. (2017). Energy recovery from waste incineration—the importance of technology data and system boundaries on CO2 emissions. *Energies*, *10*(4), 539. https://doi.org/https://doi.org/10.3390/en10040539

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, *96*(34), 226–231.

HKEPD. (2019). *Management of abandoned construction and management of abandoned construction and demolition materials*. https://www.aud.gov.hk/pdf_e/e67ch04sum.pdf

Jagadish, H. V. (2015). Big data and science: Myths and reality. *Big Data Research*, *2*(2), 49–52. https://doi.org/https://doi.org/10.1016/j.bdr.2015.01.005

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 2053951714528481. https://doi.org/https://doi.org/10.1177/2053951714528481

Leek, J. (2014). *10 things statistics taught us about big data analysis*. KG Nuggets.

Li, Z., Wang, Q., Zhang, T., Wang, H., & Chen, T. (2020a). A novel bulk density-based recognition method for kitchen and dry waste: A case study in Beijing, China. *Waste Management*, *114*, 89–95. https://doi.org/https://doi.org/10.1016/j.wasman.2020.07.005

Li, Z., Zhou, H., Zheng, L., Wang, H., Chen, T., & Liu, Y. (2020b). Seasonal changes in bulk density-based waste identification and its dominant controlling subcomponents in food waste. *Resources, Conservation and Recycling*, 105244.

Lu, W., Chi, B., Bao, Z., & Zetkulic, A. (2019). Evaluating the effects of green building on construction waste management: A comparative study of three green building rating systems. *Building and Environment*, *155*, 247–256. https://doi.org/https://doi.org/10.1016/j.buildenv.2019.03.050

Lu, W., & Tam, V. W. Y. (2013). Construction waste management policies and their effectiveness in Hong Kong: A longitudinal review. *Renewable and Sustainable Energy Reviews*, *23*, 214–223.

Lu, W., Webster, C. J., Peng, Y., Chen, X., & Chen, K. (2018). Big data in construction waste management: prospects and challenges. *Detritus*. https://doi.org/https://doi.org/10.31025/2611- 588 4135/2018.13737

Lu, W., & Yuan, H. (2011). A framework for understanding waste management studies in construction. *Waste Management*, *31*(6), 1252–1260. https://doi.org/https://doi.org/10.1016/j.wasman.2011.01.018

Lu, W., & Yuan, H. (2012). Off-site sorting of construction waste: what can we learn from Hong Kong? *Resources, Conservation and Recycling*, *69*, 100–108.

Lyon, T. L., & Buckman, H. O. (1922). *The nature and properties of soils: A college text of edaphology*. Macmillan.

Mattox, D. M. (2010). Film characterization and some basic film properties. *Handbook of Physical Vapor Deposition (PVD) Processing: Film Formation, Adhesion, Surface Preparation and Contamination Control*, 595.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Mehlhorn, K. (2013). *Data structures and algorithms 1: Sorting and searching* (Vol. 1). Springer Science & Business Media.

Russom, P. (2011). *Big data analytics. TDWI best practices report, Fourth Quarter*.

Salem, Z., Hamouri, K., Djemaa, R., & Allia, K. (2008). Evaluation of landfill leachate pollution and treatment. *Desalination*, *220*(1–3), 108–114. https://doi.org/https://doi.org/10.1016/j.desal.2007.01.026

SEPA. (2015). *Guidance on using the European Waste Catalogue (EWC) to code waste*. https://bit.ly/34Qgtzo

Shen, Y., Li, Y., Wu, L., Liu, S., & Wen, Q. (2014). Big data overview. In *Enabling the new era of cloud computing: Data security, transfer, and management* (pp. 156–184). IGI Global.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 553–572. https://doi.org/https://doi.org/10.2139/ssrn.1606674

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286.

The World Bank. (1999). *Municipal solid waste incineration: world bank technical guidance report*. https://bit.ly/2Gqfrko

WRAP. (2010). *Material bulk densities*. https://bit.ly/34UUpUx

Xu, Q., Qin, J., & Ko, J. H. (2019). Municipal solid waste landfill performance with different biogas collection practices: Biogas and leachate generations. *Journal of Cleaner Production*, *222*, 446–454. https://doi.org/https://doi.org/10.1016/j.jclepro.2019.03.083

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2013). Sensing as a service and big data. *ArXiv Preprint ArXiv:1301.0159*. https://doi.org/https://arxiv.org/ftp/arxiv/papers/1301/1301.0159.pdf