Benchmarking construction waste management using waste generation rates derived from big data

Weisheng Lu[1,*], Xi Chen[2], Yi Peng[3], and Liyin Shen[4]

[1]*Department of Real Estate and Construction, Faculty of Architecture 5/F, Knowles Building, The University of Hong Kong, Pokfulam, Hong Kong Hong Kong, P. R. China, Email: wilsonlu@hku.hk*
[2]*Department of Real Estate and Construction, Faculty of Architecture 5/F, Knowles Building, The University of Hong Kong, Pokfulam, Hong Kong, P. R. China, Email, chenx90@hku.hk*
[3]*School of Urban-rural Planning and Management, Zhejiang University of Finance and Economics, Hangzhou, P. R. China, Email: pengyihz@gmail.com*
[4]*Faculty of Construction Management and Real Estate, Chongqing University, Chongqing, P. R. China, Email: shenliyin@cqu.edu.cn*
* Corresponding author. Tel: +85228597981, Fax: +852 2559 9457, E-mail: wilsonlu@hku.hk

Abstract

The waste generation rate (WGR) is usually used as a key performance indicator (KPI) to benchmark construction waste management (CWM) performance, with a view to improving the performance continuously. However, existing researches, for different reasons, only investigated a relatively small amount of construction projects, whose WGRs cannot be confidently accepted as KPIs. This study develops a set of more reliable KPIs/WGRs using an available big dataset on CWM in Hong Kong. By mining the 2,212,026 waste disposal records generated from 5,764 projects in two consecutive years of 2011 and 2012, the WGRs/KPIs are revisited and refined. Demolition is found the most wasteful works. New building, and maintenance and renovation (M&R) works individually produce the least waste amount but by accumulating all M&R works, their contribution to the total amount of construction waste could be phenomenal. Based on the more reliable WGRs from the big data, CWM performance benchmarks for different categories of projects are set up. A contractor can benchmark its CWM performance against its counterparts or its past performance as 'Good', 'Average', and 'Not-so-good', and thus identify better CWM practices that induce superior performance. Based on the benchmarks, the government may consider setting up a WGR-step toll system to encourage those 'Not-so-good' contractors to perform well in the future, and initiate incentives to the companies conducting 'Good' projects to spur better CWM performance. Overall, the WGRs derived from the big data and more robust analyses provide a very powerful and handy tool for CWM.

Keywords: Construction waste management (CWM); Key performance indicator (KPI); Waste generation rate (WGR); Benchmarking; Big data; Data mining; Hong Kong

1. Introduction

Construction waste is defined as the waste that arises from construction, renovation, and demolition activities (Kofoworola and Gheewala, 2009). It may also include surplus and damaged products and materials arising in the course of construction work or used temporarily during the process of on-site activities (Roche and Hegarty, 2006). Sometimes, the terms 'construction waste' and 'C&D waste' are used interchangeably (Lu et al., 2015) and this is also the case in this paper. The Hong Kong Environmental Protection Department (EPD, 2014a) categorizes construction waste into two main types. They are inert construction waste (ICW), which are materials with stable chemical properties (e.g. soil, earth, silt, bricks, blocks, rocks and concrete), and non-inert construction waste (non-ICW) such as timber, bamboo, and paper board and other organic materials. ICW is suitable for public fill works, e.g. site formation and land reclamation, while non-ICW depletes land resources and contaminates surrounding environment after it is disposed of at landfills (Poon, 2007; Lu, 2013; Lu and Yuan, 2013; Yuan et al., 2013; Yuan, 2011). There are some hazardous wastes, such as asbestos and contaminated soil, that arise from construction works but in many countries they are not classified as construction waste (Mou, 2008) and therefore are not considered in this paper. With the increasing embracement of sustainable development, it is highly important to take measures to mitigate the waste generation from the construction industry.

Waste generation rate (WGR) has been broadly used as an indicator to measure CWM performance (Bossink and Brouwers, 1996; McDonald and Smithers, 1998; Formoso et al., 2002; Tam et al., 2007; Lu et al., 2011). It can be used as key performance indicators (KPIs), based on which contractors can benchmark their CWM performance and in turn identify the best practice that can seek for continuous improvement. Previous studies on WGRs, which adopted research methods, for instance, literature review, case studies, interviews, site inspections and questionnaire survey, provided subjective and limited understanding of the performances (Formoso et al., 2002; Lu et al., 2011; Lin, 2006; Tam et al., 2007; Gangolells et al., 2014). Most of the studies on CWM performance (measured by WGR) have a relatively small sample or sampled relatively small sites due to the difficulties involved in conducting a survey on large-scale projects over a relatively long period of time (Katz and Baum, 2011; Lu et al., 2011). As a consequence, these WGRs reportedly ranged from one study to another without any form of reliability. Results of such studies thus cannot be utilized with a high level of confidence as yardsticks for benchmarking.

The aim of this study is to develop a set of more reliable KPIs/WGRs by making use of a big dataset that has been collected in the past years. Complying to the Law of Large Numbers (LLN), the average of the results obtained from a large number of trials should tend to become convergent to a certain value as more trials are performed (Sen and Singer, 1994). The representative WGRs of non-ICW and ICW for different categories of construction works are identified to measure the CWM performance that epitomizes each category. Benchmarks are set to compare the performance of construction projects with various natures of waste generation. The introduction is followed by a detailed review of KPIs, WGRs, big data, and data mining. Based on the review, detailed research design was put forward in the section of research methodology. The process of analyzing the collected big dataset is presented in the data analysis and results section. Accordingly, the results and relevant implications are discussed in the section of discussion. Suggestions for enhancing the CWM are raised for policy-makers, contractors, researchers and other stakeholders in the final section.

2. Literature review

2.1 Benchmarking based on key performance indicators (KPIs)

In recent decades, the construction industry has become increasingly competitive. In order to gain competitive advantages, construction companies are pursuing an approach to assessing the management performance. Benchmarking was introduced as a continuous process of improving performance in a systematic and logical way by measuring products, services, and practices by learning from the best to make targeted improvements (Camp, 1989). Benchmarking systems are targeted for development in the construction industry in a few countries via typically analyzing the performance of a system based on a set of key performance indicators (KPIs) (Horta et al., 2009; Cheung, 2010). KPIs represent a set of metrics measuring how well a system performs an operational, tactical or strategic activity, which are the most critical for the current and future success of the system (Parmenter, 2007; Eckerson, 2006). An organization can benchmark its performance by taking the results of its KPIs and comparing these with the performance of their counterparts or with its own past performance as appropriate (Thomas and Thomas, 2008). Therefore, KPIs not only serve as early warning signs that give decision-makers information to reduce uncertainty, but are also expected to indicate what measures should be taken to make sustained improvement in efficiency and quality (Kerzner, 2011).

Researchers have attached their attentions to KPIs in benchmarking performance of CWM. For example, Lin et al. (2011) measured the success of construction projects through benchmarking the performance with the identified KPIs. Hegazy and Hegazy (2012) produced a benchmarking model based on financial KPIs for construction companies to benchmark and evaluate their business performance at the corporate level in the UK. Horta et al. (2009) tried to benchmark the performance assessment of the construction industry by integrating KPIs and data development analysis. More frequently, benchmarking with KPIs also exists in pursuing the success of CWM. Through studying the construction waste generated in a number of hotel projects, Ball and Taleb (2011) found that the benchmarks in existing CWM legislation need to be amended. In measuring waste management performance in the construction industry, waste generation rates (WGRs) are usually replaceable by the KPIs.

2.2 WGRs as KPIs

It has become the tide that construction industry measures performance of CWM with various data collection approaches by focusing on different KPIs, mainly found expressions in waste amount and WGRs. At early time, the method is to quantify construction waste amount, and digging out the causes of construction waste generation (Bossink and Brouwers, 1996). Poon et al. (2004a) also quantified waste amount and found the major causes of waste materials were improper preparation, handling, misuse, and incorrect processing. There are certain existing studies using WGRs as the KPIs for measuring the performance of CWM of individual construction projects. To this end, WGRs becomes the KPI of CWM in this study. Formoso et al. (2002) examined waste management in Brazil through estimating WGRs, which were waste percentage of purchased materials by weight. Poon et al. (2004b) measured the WGR with the volume of waste generated per gross floor area (GFA), which is probably the most frequently used KPI in the literature. WGR is also regarded as an important indicator for successful implementation of an integrated construction waste management plan (Bakshan et al., 2015).

In previous studies, diversified research methods were adopted to acquire the data to measure WGRs. Lin (2006) adopted the neural network method to measure the WGRs for the construction of factory and residential buildings in Taiwan. Interviewing waste manager is also a method for collecting data for calculating WGRs of some projects (Tam et al., 2007). Lu et al. (2011) examined the waste management effectiveness in a typical city, Shenzhen, China by focusing on WGRs of different materials from several construction sites. Visual inspection, tape measurement, and truckload records were used in the study of Poon et al. (2004b). However, these existing studies usually investigate WGRs with a small scale of data, which therefore cannot identify common rules and generalize their findings to other cases. With the help of convenient data collection and large record, big data and data mining are becoming possible to advance the research on WGRs.

2.3 Big data and data mining

Big data is defined as things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more (Mayer-Schönberger and Cukier, 2013). People tend to accept the definition that was asserted by IBM that big data has data volume, velocity and variety (three Vs) (Zikopoulos and Eaton, 2011). Volume is the quantities of terabytes, records, transactions, tables, or files; velocity finds expression in batch, near time, real time and streams; and variety can be structured, unstructured, semi-structured and a combination of them (Russom, 2011). Big data could be strategically used as a raw material and a vital input to create a new form of value in living, working, science and industry (Mayer-Schönberger and Cukier, 2013). Its value is found in finance and insurance industries, government, companies of computers and other electronic products, construction industries and others (Brown et al., 2011). Chen et al. (2012) studied how to better serve the needs of business decision-makers by emerging big data, managers and others. Howe et al. (2012) asserts big data analytics would become the mainstream of the future research in bio-curation.

Big data analytics can be useful in infering the likelihood of poor management performance in the construction industry. In managing construction projects, there is both physical and virtual data from procurement, controlling, sub-contracting, building information modelling, bidding, scheduling, tendering, site information, and many other aspects. Through detailed analysis of big data, an organization can gain business advantages by discovering new characteristics about their customers, markets, partners, costs, and operations (Labrinidis and Jagadish, 2012). Likewise, through analyzing the big data from various projects, it is able to find the reasons explaining the poor performance in this important sector. Recently, big data centers have been developed in construction markets for data capture, storage, security and analytics.

Data mining is a young, dynamic, and promising area, which is resulted from the urgent necessity of automatically discovering valuable information from a large collection of data and transforming it into organized knowledge (Han and Kamber, 2001). Rather than simply locating, identifying, understanding and citing data, data mining requires integrated, cleaned, trustworthy, and efficiently accessible data, declarative query and mining interfaces, scalable mining algorithms, and computing environments (Labrinidis and Jagadish, 2012). It is important to understand what should be the useful information. This resonates with Clifton (2010) that data mining serves as a computational process where patterns in large datasets can be discovered using diversified approaches, in particular the well-known machine learning and statistics. Characterization and discrimination, the mining of frequent patterns, associations, and correlations, classification and regression, clustering analysis and outlier analysis are patterns to mine in datasets (Han et al., 2012).

The overall objective of conducting data mining is to acquire information through analyzing a dataset and transform it into an understandable form for afterward uses (Clifton, 2010). A classical application of data mining is the finding that about 80% customers that buy beer also buy potato chips after analyzing supermarket transaction records to estimate customers' consumption behavior; the supermarket can then purposely place chips close to beer for promoting sales of both (Lee and Siau, 2001). Data mining has witnessed great success in numerous applications, such as business intelligence (Delmater and Hancock, 2001) and web search engine (Han and Chang, 2002), analysis of an energy efficient building design (Kim et al., 2011), education (Romero and Ventura, 2007) and finance (Zhang and Zhou, 2004). Therefore, this study aims to expand data mining to the waste management research.

3. Research methodology

After a detailed review of literature on KPI, WGR, big data, and data mining, the methodology for present research becomes clear, by following collecting big data, mining the big data in terms of WGR, setting benchmarks, and comparing the CWM performance. The analytical process is presented in Fig.1.

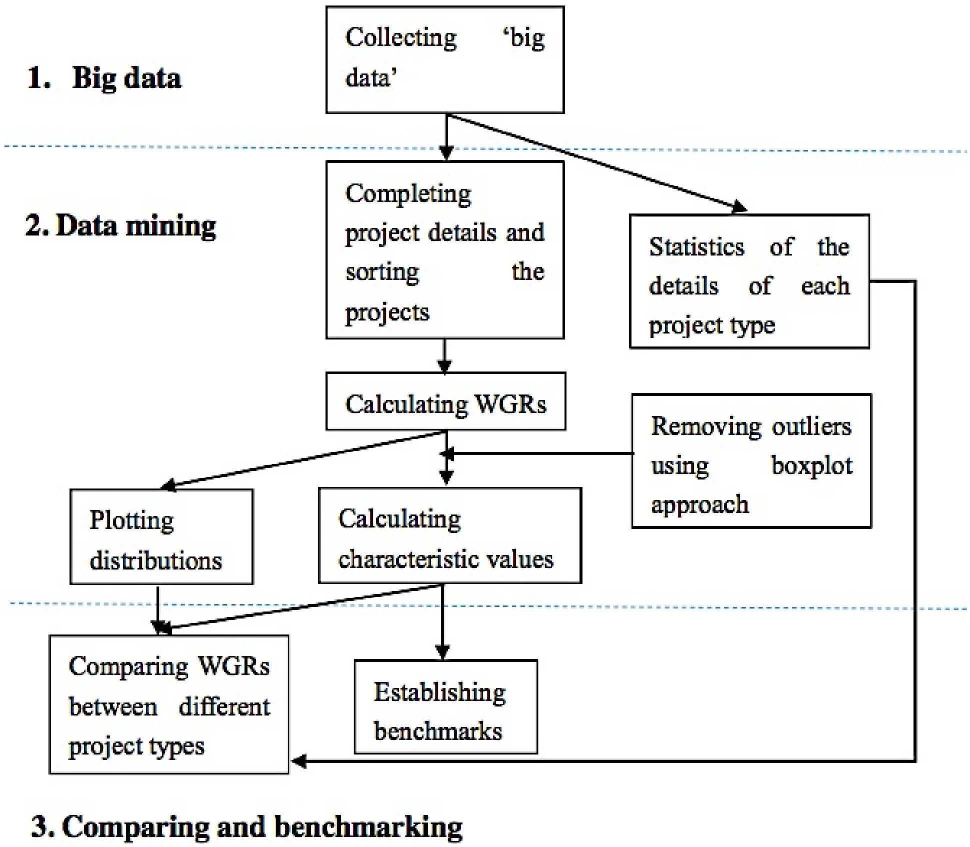

Fig. 1 The research methodology for deriving benchmarks for CWM performance

Step1 Collecting 'big data' of construction waste

Collecting big data is still a challenge, as it needs advanced sensors, transmission, and storage. With the aim of investigating WGR, this study mainly relies on existing data records rather than collecting data on field. Notably, the management practice of construction waste transaction and disposal in Hong Kong has led to a set of big data. To effectively manage construction waste, a Construction Waste Disposal Charging Scheme (CWDCS), on the basis of the 'polluter pays principle', has been enacted in Hong Kong since 2006 (Lu and Tam, 2013). In accordance with the CWDCS, a contractor should pay HK$125 per ton for the non-ICW generated from his construction site and accepted by landfills or outlying island transfer facilities (OITFs); HK$100 per ton of mixed ICW and non-ICW received by off-site sorting facilities (OSFs); and HK$27 per ton of waste mainly consisting ICW ended in public fill reception facilities (PFRFs) (HKEPD, 2014a). It is noticed that for every lorry of construction waste ended in any government-run facilities consisting the above four types, it leaves over a record at the HKEPD. This practice leads to more than 2 million transaction records of this kind in two consecutive years of 2011 and 2012 (see Fig. 2 for an excerpt of the big data).

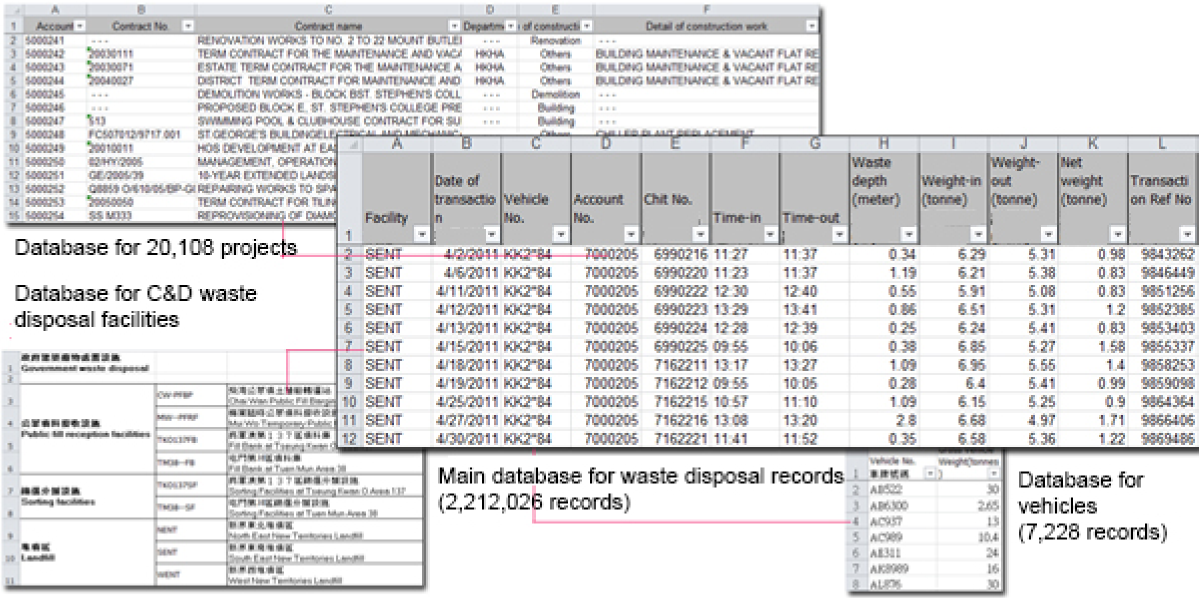

Fig. 2 Links between databases of construction waste disposal in Hong Kong

The waste disposal record includes information of the lorry of construction waste, including vehicle no., measured construction waste amount, waste disposal time, billing account for the construction project, and name of the facility that receives the waste. Account number, construction name, category, site and contract sum of 20,108 C&D projects are organized in another database. The unique account number acts as a bridge to link the information of a certain project and these waste disposal records. The third database is the information of the disposal facilities, including facility name, received waste type, and facility address. The links between the three databases are shown in Fig. 2. By mining the big data, it is possible to extract some meaningful CWM related patterns and insights for policy-makers and contractors.

Step 2 Data mining

Initial data process would be conducted to classify the generated construction waste as ICW and non-ICW while relevant projects would be classified as building, civil engineering, demolition, maintenance and renovation (M&R). The classification is useful to set feasible benchmarks for each category of projects and conduct effective comparisons among the different categories. After that, WGR of collected projects would be calculated by following Equation (1):

$$\text{WGR (t/mHK\$)} = \frac{\text{Waste net weight (ton)}}{\text{Project contract sum (million HK\$)}} \qquad (1)$$

This measurement indicates the level of waste generation in producing every million HK$'s (or any currency as concerned) worth of construction work. In existing research works waste volume and/or weight per GFA, i.e. m^3/m^2 and/or ton/m^2, are often used as the units of WGRs (Poon et al., 2004b; Lu, 2011). Contract sum is utilized to estimate WGRs owing to the fact that a large amount of construction works, such as maintenance, repair, civil and some minor works, are unavailable of GFA but with a contract sum. Therefore, this study adopts the WGR as shown in Equation (1), with a view to comparing CWM performance across different categories of projects. Nevertheless, it should be pointed out that contract sum differs from one country to another, and from one period to another, although in practice these can be adjusted by using construction cost indexes published in individual countries, and Consumer Price Indexes in different periods. The new KPI and others are not mutually exclusive. Particularly, the indicators with GFA as the denominator reflect building projects more objectively. In this sense, this KPI is introduced to supplement instead of replace existing CWM performance indicators.

In Hong Kong, construction waste is composed of non-ICW and ICW, readers are reminded of the hazardous construction waste was treated separately in another stream though. Waste materials disposed in landfill sites and OITF are regarded as non-inert waste (EPD, 2014a). The waste disposed in PFRFs consists entirely of inert construction waste. The OSFs receive mixed waste from construction sites, regarded consisting of at least 50% inert materials (EPD, 2014a). For a project, the non-inert and inert WGRs (i.e. $\mathrm{WGR}_{\text{non-inert}}$ and $\mathrm{WGR}_{\text{inert}}$ respectively) are consequently calculated as Equations (2) and (3).

$$\mathrm{WGR}_{\text{non-inert}}\ (\text{t/mHK\$}) = \frac{W_{\text{landfill}} + 50\%W_{\text{OSF}} + W_{\text{OITF}}\ (\text{ton})}{\text{Project contract sum (million HK\$)}} \qquad \text{Equation (2)}$$

$$\mathrm{WGR}_{\text{inert}}\ (\text{t/mHK\$}) = \frac{W_{\text{PFRF}} + 50\%W_{\text{OSF}}(\text{ton})}{\text{Project contract sum (million HK\$)}} \qquad \text{Equation (3)}$$

where W_{landfill}, W_{OITF}, W_{OSF} and W_{PFRF} are the construction waste disposed by landfills, OITFs, OSFs and PFRFs respectively.

However, it is common to find many outliers in calculating WGRs, which brings negative impact on followed statistics analysis and should be removed. R, which is an open source software for statistical computing and graphics, is applied in this study for removing the outliers through boxplot approach. After all outliers have been removed, WGR distribution figures can be drawn to visually compare existing cases while a characteristic values of the distributions, such as mean, standard deviation (SD), and median of a set of WGRs can be calculated for the followed use of benchmarking.

Step 3 Comparing and benchmarking

With the mined WGR distribution and characteristics values of relevant distributions, the benchmark of different categories of construction projects can be developed. By followed the common practice in benchmarking, the projects whose WGRs are in top 15% in the order of significance are benchmarked as the 'Non-so-good', those in bottom 15% are 'Good' ones, and the rest 70% projects between the 'Non-so-good' and 'Good' are 'Average' projects. In addition, with characteristics values of WGR distribution, C&D waste management performance of different categories of projects can also be compared. As well, A contractor can benchmark its C&D waste management performance against its counterparts or its past performance as 'Good', 'Average', and 'Not-so-good'.

4. Data analyses, and results

4.1 Project profiles

In the two consecutive years of 2011 and 2012, there were total 5,764 projects that disposed of construction waste in various government C&D waste management facilities, which maintained 2,212,026 waste disposal records in the EPD forming the 'big data' for the analyses of this study. Table 1 illustrates in detail the different categories of projects including their sample sizes and contract sums. The 'unclear' projects were those minor construction works, which only had a billing account without specific linkages to a client. Neither did they have any specific project information (e.g. construction category, GFA, or contract sum). These projects are excluded in the analyses in this study due to their information incompleteness, leaving 4,227 projects in the sample.

Table 1 Project categories and details of projects

Construction category	Sample size	Total contract sum(bHK$)	Average contract sum (mHK$)
Unclear	1537	N/A	N/A
Building	627	228.37	364.23
Civil	521	163.01	312.88
Demolition	282	3.80	13.49
Foundation	552	105.57	191.26
M&R	2119	84.89	40.06
Others	126	11.42	225.86

It can be noticed that maintenance and renovation (M&R) projects take more than half of all the 'information-clear' projects in Hong Kong owing to a decayed urban. According to the Housing, Planning and Lands Bureau (2005), there are about 39,000 private buildings in Hong Kong, about 13,000 of which are over 30 years' old, while in ten years' time, the number will increase to 22,000. Buildings and civil works are the two largest sectors in the construction market of Hong Kong. By the end of March 2012, there were 2,599,000 permanent residential flats in stock, of which 1,447,000 (56%) were private flats, and the rest is subsidized housing or public rental housing (PRH). The large building sector is further sustained by the ambitious public housing scheme in Hong Kong. According to the forecast of Hong Kong Housing Authority (2014), approximately 77,000 PRH and subsidized housing flats will be built 2014/15 to 2018/19. Accordingly, considerable amount of infrastructure projects were developed to support the economic and social activities in Hong Kong.

4.2 WGRs of all projects

By using Equation (1), the WGRs of each project in 2011 and 2012 are calculated and plotted in Fig. 3. There are altogether 4,227 projects available of WGRs, the values of which are from 0.005 to 7,115.12 ton/million HKD (t/mHK$). It can be seen that the majority of the WGRs distributed within a range between 0.1 to 100 t/mHK$. However, no apparent pattern of the WGRs can be easily detected.

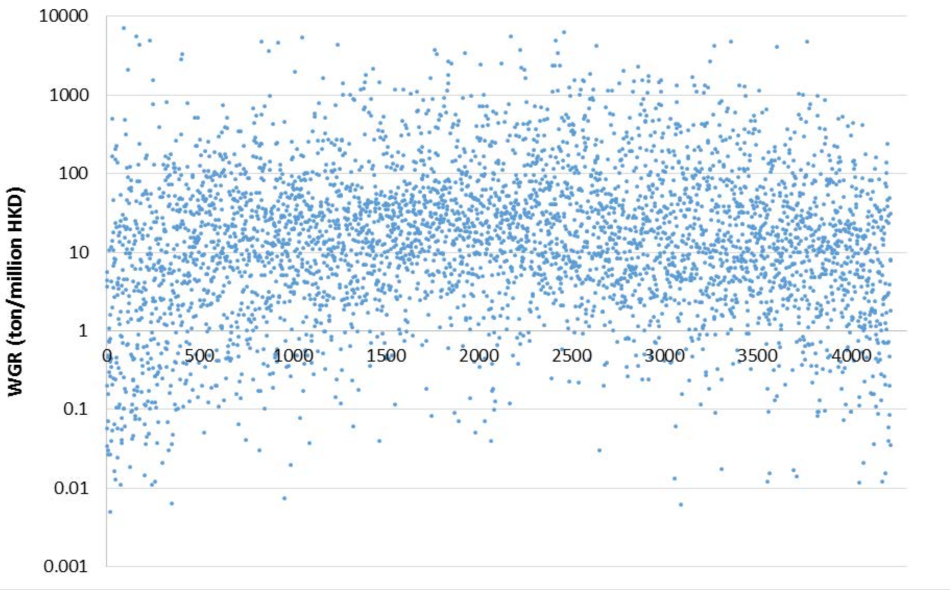

Fig. 3 WGRs of the individual projects (sample size=4227)

Noise reduction was performed by examining those obviously unreasonable WGRs (e.g. a WGR is larger than 10,000 t/HK$), and removing those outliers. Box plots can remove the possible outliers in a statistical population without making any assumptions of the underlying statistical distribution. This non-parametric approach is applied to remove the outliers out of *ln(WGR)*s for the 4,227 projects using *R*.

With outliers being excluded (now sample size=4062), *R* is used to produce the curve of density function of WGRs (see Fig. 4), which illustrates the distribution of the WGRs of all projects *per se*. The curve appears to be a positive-skewed distribution. Therefore, a log-normal distribution, which is one of the positive-skewed distributions, is applied to fit the distribution of WGRs of the projects. The natural logarithms of WGRs, i.e. *ln(WGR)s*, are calculated and the curve of density function of *ln(WGR)s* is also plotted as shown in Fig. 5. According to our curve fitting and statistical analyses using *R*, the distribution of *ln(WGR)s* is not an actual normal distribution, but appears similar to a normal distribution. Hence, it is legitimate to use the median of *ln(WGR)s* to reflect the average *ln(WGR)s* of the majority of the projects. The mean, SD, and median of WGRs of the projects are also calculated and tabulated in Table 2.The median of the new group of WGRs, 15 t/mHK$ is used to reflect the CWM performance of the major projects in the sample. It can be seen that simply using means without considering the distribution of the sample could be very misleading in understanding average C&D waste management performance due to the extremely skewed distribution and large range of the WGRs, which are presented in Table 2.

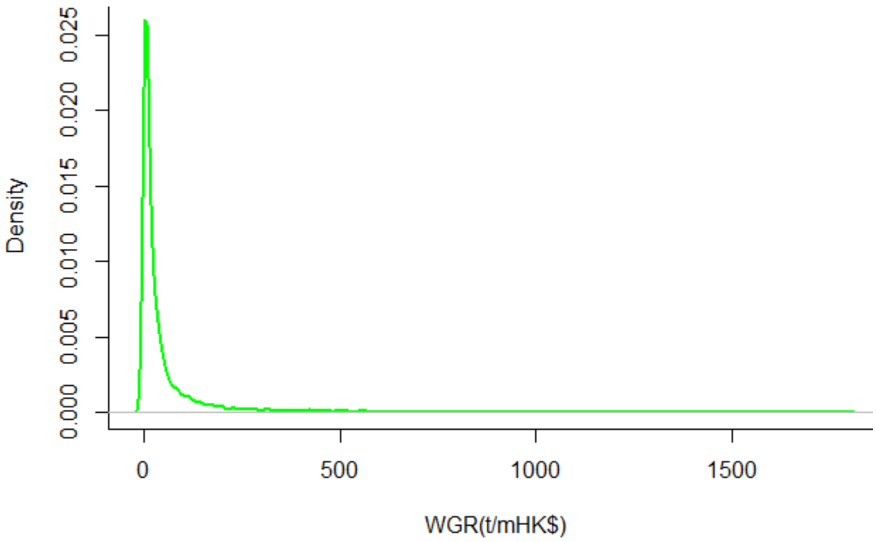

Fig. 4 Curve of density function of WGRs of all projects (sample size=4062)

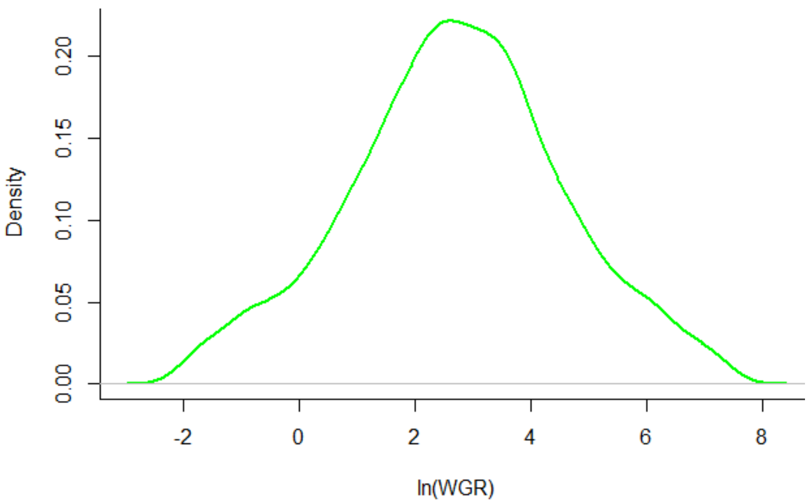

Fig. 5 Curve of density function of *ln(WGR)s* of all projects (sample size=4062)

Table 2 Means, SDs, and medians of WGRs of the projects

Projects	Sample size	Mean (t/mHK$)	SD	Median (t/mHK$)	Range (t/mHK$)
Overall	4062	76	192	15	0.13~1793.33

Non-inert and inert WGRs

By using Equations (2) and (3), the WGRs of ICW and non-ICW in the two years can be calculated and presented in Fig. 6. Most projects usually generate both ICW and non-ICW, but a few projects generate either ICW or non-ICW only. Fig. 7 is the curves of density function of the natural logarithms of non-inert and inert WGRs, i.e. $ln(WGR_{non\text{-}inert})s$ and $ln(WGR_{inert})s$. Both curves are similar to a normal distribution but according to our curve fitting and statistical analyses using *R*, they are not statistically normal distributions. Nevertheless, as discussed above, it is legitimate to use median to reflect the average CWM performance of the majority of the projects. The medians are 3 and 12 t/mHK$ for non-inert and inert WGRs respectively (see Table 3).

Fig. 6 Inert and non-inert WGRs of individual projects

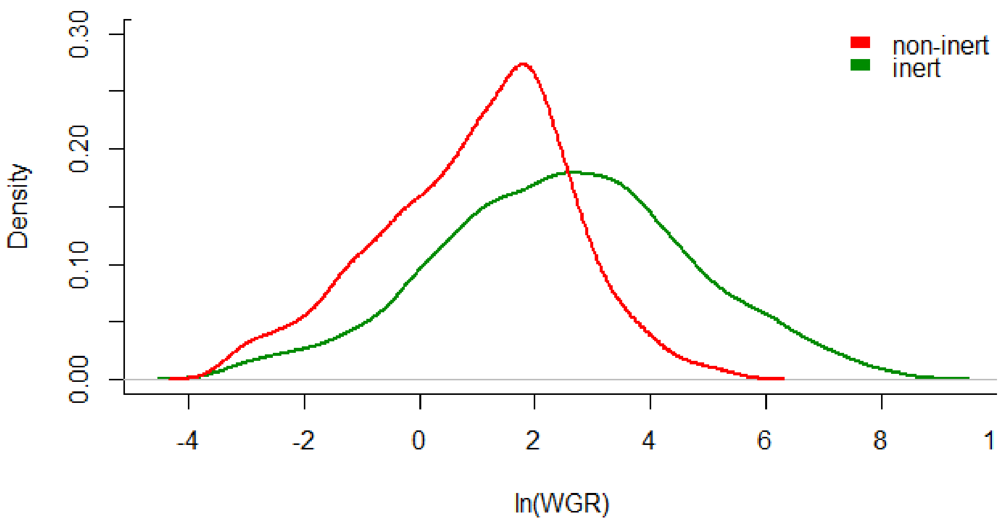

Fig. 7 Curves of density functions of $ln(WGR_{non\text{-}nert})s$ and $ln(WGR_{inert})s$

Table 3 Means, SDs, and medians of inert and non-inert WGRs of the projects

WGR type	Mean (t/mHK$)	SD	Median (t/mHK$)	Range (t/mHK$)
Non-inert	8	18	3	0.03~232.70
Inert	100	318	12	0.03~4188.89

4.3 WGRs of different construction categories

Different construction categories may make differences in construction waste generation. Building, civil, demolition, foundation and M&R are five main construction categories in the 4,227 projects, plotted in different styles in Fig. 8. A small amount of projects such as provision, cleaning, building service, material testing and equipment relocation are grouped into others. After noises and outliers in each construction category are removed by taking similar Box plots analyses using *R*, the curves of density functions of $ln(WGR_{non\text{-}inert})s$ and $ln(WGR_{inert})s$ are created and shown in Figs.9 and 10. The curves of $ln(WGR_{non\text{-}inert})s$ and $ln(WGR_{inert})s$ in Figs. 9 and 10 are all similar to a normal distribution, which means these distributions are similar with a log-normal distribution. Therefore, the median of the set of WGRs for each type is proper to reflect the general quality of CWM performed by that type of projects.

Fig. 8 Overall WGRs by construction categories

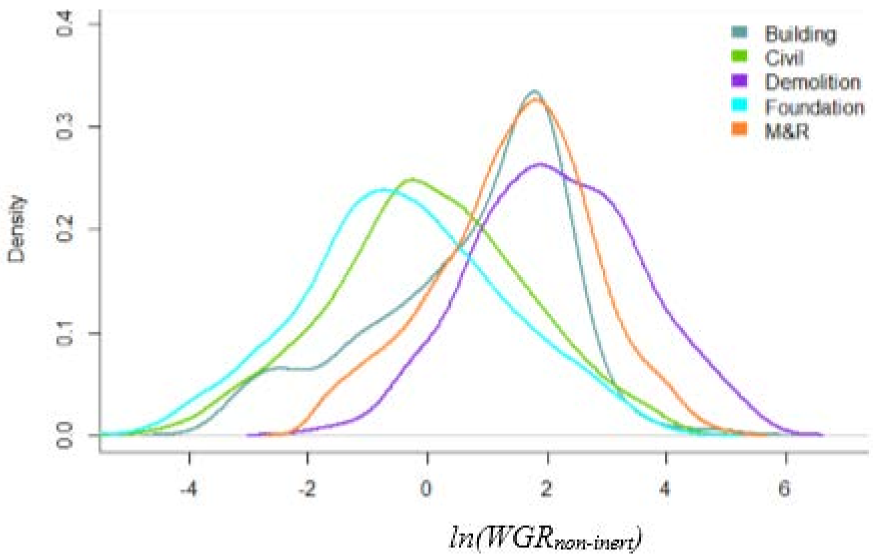

Fig. 9 Curves of density functions of *ln(WGR*$_{non\text{-}inert}$*)s* by construction categories

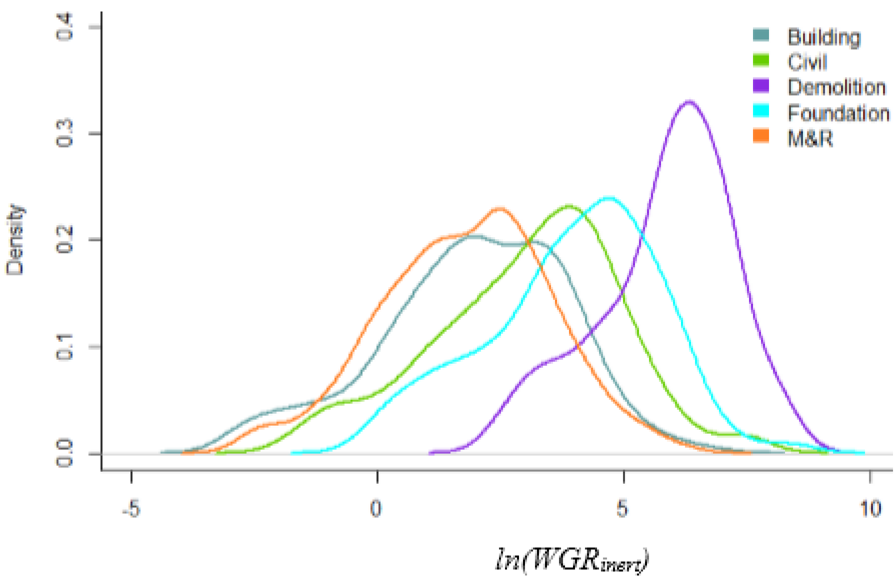

Fig. 10 Curves of density functions of *ln(WGR*$_{inert}$*)s* by construction categories

The medians of non-ICW and ICW WGRs for different construction categories are calculated and shown in Tables 4 and 5, respectively. From the tables, it can be seen that demolition projects are the most wasteful type among all the projects; the medians of both their non-inert and inert WGRs (8.15 and 423.23 t/mHK$) are much higher than other categories. Building and M&R, with higher non-inert WGRs (3 and 4.82 t/mHK$) has lower inert WGRs (8.05 and 6.58 t/mHK$), while foundation and civil works, with relatively higher inert WGRs (28.06 and 64.96 t/mHK$) however generate a small amount of non-inert waste per cost (0.96 and 0.65 t/mHK$).

Table 4 Medians and ranges of non-inert WGRs for building, civil, demolition, foundation and M&R projects (t/mHK$)

Construction category	Building	Civil	Demolition	Foundation	M&R
Median	3	0.96	8.15	0.65	4.82
Range	0.03~143.74	0.01~48.00	0.17~211.82	0.01~53.40	0.14~135.78
Non-so-good	(8.42,143.74)	(5.77,48.00)	(36.82,211.82)	(4.4,53.40)	(15.62,135.78)
Average	(0.31,8.42]	(0.19,5.77]	(2.12,36.82]	(0.12,4.4]	(1.01,15.62]
Good	[0.03,0.31]	[0.01,0.19]	[0.17,2.12]	[0.01,0.12]	[0.14,1.01]

Table 5 Medians and ranges of inert WGRs for building, civil, demolition, foundation and M&R projects (t/mHK$)

Construction category	Building	Civil	Demolition	Foundation	M&R
Median	8.05	28.06	423.23	64.96	6.58
Range	0.05~904.31	0.17~2105.90	9.51~5411.72	0.68~4883.86	0.05~681.03
Non-so-good	(46.93,904.31]	(125.43,2105.90]	(1237.28,5411.72]	(297.04,4883.86]	(34.74,681.03]
Average	(1.08,46.93]	(2.64,125.43]	(75.41,1237.28]	(6.99,297.04]	(0.97,34.74]
Good	[0.05,1.08]	[0.17,2.64]	[9.51,75.41]	[0.68,6.99]	[0.05,0.97]

4.4 Benchmarks of CWM performance amongst different construction categories

The ranges of non-inert WGRs and inert WGRs for different categories as listed in Tables 4 and 5 represent the performances of C&D waste management of building, civil, demolition, foundation and M&R works in construction industries. This step sets up the benchmarks of C&D waste management performance of these categories of projects. The projects whose WGRs are in top 15% in the order of significance are benchmarked as the ‘Non-so-good’, those in bottom 15% are ‘Good’ ones, and the rest 70% projects between the ‘Non-so-good’ and ‘Good’ are ‘Average’ projects. Based on the ranges derived as shown in Tables 4 and 5, Fig. 11and Fig. 12 illustrate the benchmarks as set up for C&D waste management performance of building, civil, demolition, foundation, and M&R projects.

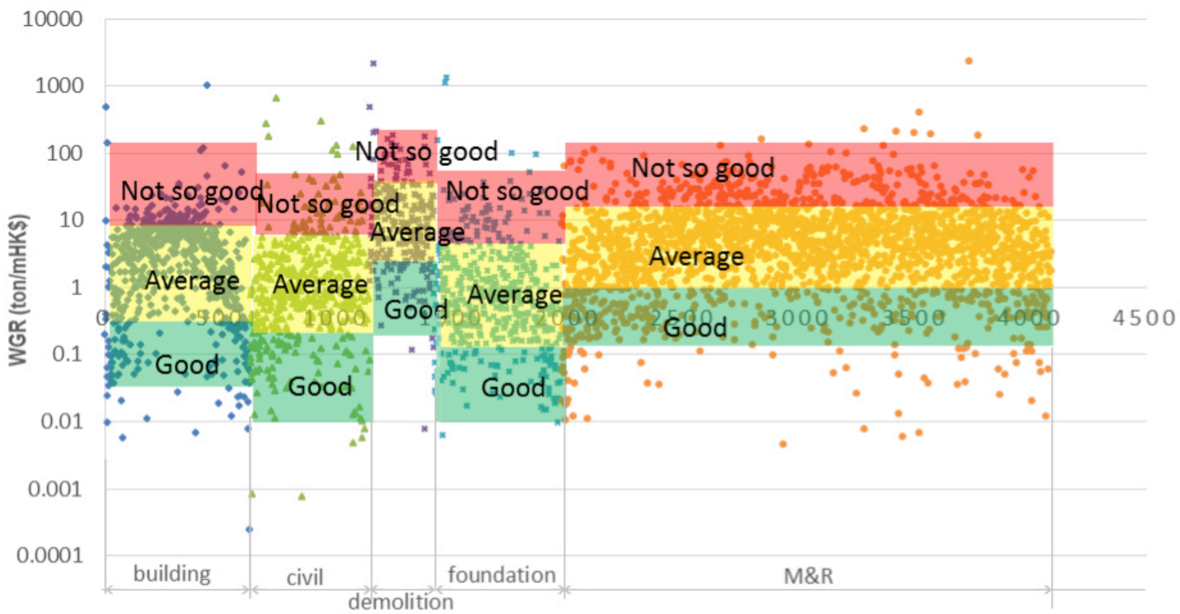

Fig. 11Benchmarks of non-inert C&D waste management performance

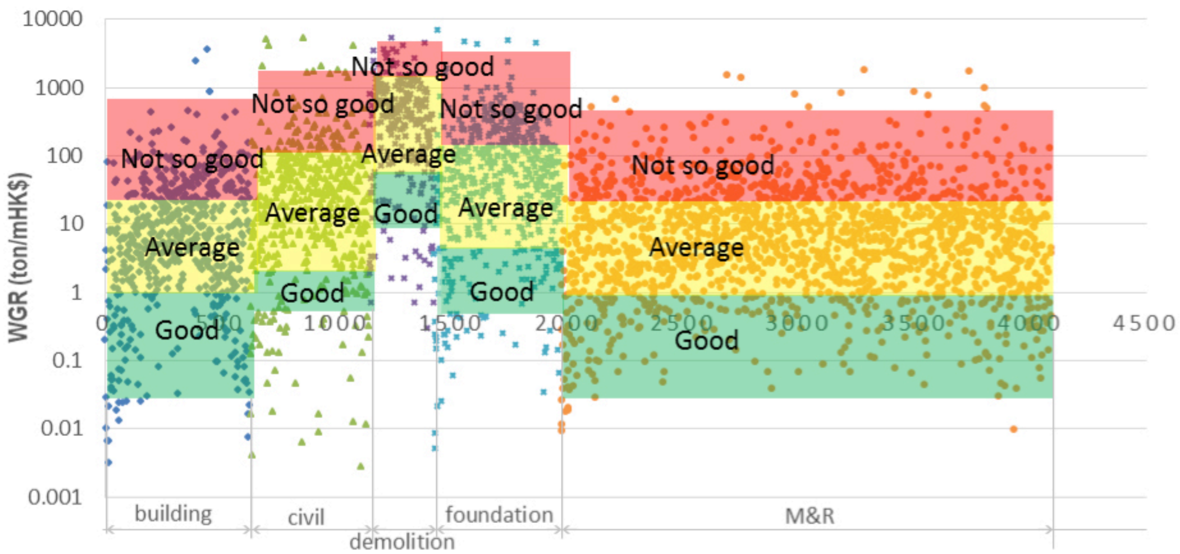

Fig. 12Benchmarks of inert C&D waste management performance

5. Discussions

5.1 WGRs acting as KPIs for benchmarking CWM performance amongst different categories of projects

By reducing the randomness of the sample using big data, a set of more reliable WGRs that can be accepted with high confidence is developed. By comparing both non-inert and inert WGRs between overall (i.e. medians in Table 3) and categorized situations (i.e. medians in Tables 4 and 5), it is notable that the waste generation of building and M&R are closest to the average level of overall construction projects. Since M&R projects took more than half of all the construction works, managing construction waste from them is crucial in determining the overall CWM performance in a region. M&R projects often generate non-inert waste, such as paperboard packages and wooden boxes owing to the large supply of materials, mechanical equipment and building service fittings. Too often, contractors of M&R projects place a roll-off container on a site and call in an *ad-hoc* waste hauler to dump it once it is full; normally, no systematic CWM is conducted on these projects but by accumulating all the M&R works together their contribution to total construction waste could be massive. That is probably why the Hong Kong Green Building Council (HKGBC) is initiating a green interior design guide that is particularly for minimizing renovation and decoration works. With no doubt, the most wasteful construction category is demolition works, which generate a large amount of ICW and non-ICW. Civil and foundation works generate little non-ICW but a large amount of ICW, because excavation usually arises earth and concrete. Managing the ICW from them is apparently an important direction to minimize the overall construction waste.

Based on the more reliable WGRs from the big data, CWM performance benchmarks for different categories of projects are set up for ICW and non-ICW, respectively. A contractor can calculate its own WGR and position itself as ‘Good’, ‘Average’, and ‘Not-so-good’. WGR, as an indicator of CWM performance, is considered the consequence of different casual factors, such as construction techniques, work procedures, and common practices (Bossink and Brouwers, 1996). Based on the relative positions, the contractor can benchmark its CWM performance and identify the better CWM practices that induce superior performance. A contractor can also benchmark its CWM performance by taking the results of its KPIs and comparing these with its own past performance periodically. By using the KPIs, the contractor can determine with greater certainty what measures should be taken to improve its CWM performance. From a regulator’s point of view, instead of adopting a uniform levy, the government may consider setting up a WGR-step toll system to encourage those ‘Not-so-good’ contractors to contribute more to CWM. On the other hand, incentives from government, such as awarding the companies conducting ‘Good’ projects can be initiated to spur better CWM performance, because encouragement, such as best practice measures was found to be effective in promoting CWM (Saez et al., 2013).

5.2 Projects with exceptionally high or exceptionally low WGRs

There were a handful of projects, which have been treated as ‘outliers’ and excluded in the data analysis owing to their exceptionally high WGRs. For example, there is a foundation project ‘XYZ’ with non-inert and inert WGRs of 1344.24 and 1940.38 t/mHK$. The contract sum is recorded as HK$ 1,000,000. It is understandable that 1940.38 tons of ICW waste is possibly generated from the excavation, but it is questionable that the 1344.24 tons of non-ICW is generated from this foundation project. The project contractors might have reported the wrong contract sum to the HKEPD, which is in charge of opening account numbers for every contract with HK$ 1,000,000 contract sum or more. By examining the projects with exceptionally high WGRs, it is able to inform the HKEPD of the potentially inaccurate project information registered. Lane et al. (2014) reported it often difficult or impossible to trace back to particular tortfeasors. However, in this big data set, it is possible to trace back the waste generation practice contributed by a particular contractor.

There are a few projects with exceptionally low WGRs, which also deserve further investigations. Some minor construction activities may in nature arise little construction waste in nature. However, if a contractor generates exceptionally low WGRs in construction works such as buildings, civil, demolition, foundation and M&R, the contractor should be treated as an exemplar in managing construction waste. It may introduce new CWM process, putting extra efforts, or new technologies in reducing, reusing, or recycling construction waste. There is an allegation that some contractors may be involved in illegal dumping, which may in turn lead to the exceptionally low WGRs. But unless the contractor was systematically involved in it and has not been caught, it is difficult to identify the contractor as the tortfeasor from mining the big data. To deal with this problem, knowledge for stimulating contractors' CWM, such as properly promoting CWM could bring net financial benefits for stakeholders (Yuan et al., 2011), should be disseminated among contractors.

6. Conclusions

The present study investigated the waste generation rates (WGRs) of inert and non-inert waste by various projects in the years 2011 and 2012 in Hong Kong. There are 5,764 projects, primarily including building, civil, demolition, foundation and M&R, that that generated construction waste and left over more than 2 million waste disposal records in the governmental department. By mining the waste disposal records, primarily using statistical analyses and nonparametric analyses, it found that the median WGR for all projects is about 15t/mHK$, with 3t/mHK$ for non-inert waste (non-ICW) and 12t/mHK$ for inert waste (ICW). The big data allows for a holistic investigation of all categories of projects over a relatively long period of time. It largely reduces randomness of sampling which is commonly seen in previous empirical studies of this kind. The results can thus be accepted with a high level of confidence for understanding average CWM performance.

After examining the WGRs of individual categories of projects, demolition is found the most wasteful works that generate both non-inert and inert construction waste. Civil and foundation generate much inert waste but little non-inert waste. Building and M&R works produce the least waste amount but with the large non-inert to inert WGR ratios; without systematic C&D waste management, and by accumulating all the M&R works together, their contribution to total amount of construction waste could be phenomenal. There are a few but not many projects, which have exceptionally high or exceptionally low WGRs. By examining these projects, it is able to trace back the tortfeasors that contributed the WGRs for two purposes: (a) informing the government department of the potential inaccurate project information registered, or (b) selecting them as exemplars for further investigation of their CWM. Overall, the findings provide more specific actionable information for CWM; specific CWM measures can be tailored to deal with different categories of projects, which have different waste generation profiles.

Based on the more robust WGRs from the big data, CWM performance benchmarks for different categories of projects are set up for ICW and non-ICW, respectively. A contractor can position itself in the benchmarks as ‘Good’, ‘Average’, and ‘Not-so-good’. The contractor can benchmark its CWM performance with its counterparts and identify the better construction techniques, work procedures, and common practices that induce superior performance. A contractor can also benchmark its CWM performance by taking the WGRs as KPIs and comparing them with its own past performance periodically. Based on the benchmarks, the government may consider setting up a WGR-step toll system to encourage those ‘Non-so-good’ contractors to perform better in the future. On the other hand, incentives from government, such as awarding the companies conducting ‘Good’ projects can be initiated to spur better CWM performance. Governmental departments are encouraged to improve the extant codes, standards, and practices relating to CWM. With the benchmarks developed in this study, it is believed that the works can be conducted in a more informed fashion. Overall, the WGRs derived from the big data and more robust analyses provide a very powerful and handy tool for CWM. Last but not the least, it should be pointed out that the WGRs are derived from Hong Kong which has its own construction profiles such as abounding with high-rise structures, unique construction technologies, high construction cost indexes, and different construction waste management systems. Researchers from other regions should be fully aware of these differences and try to avoid a “one-size-fit-all” stance when benchmarking CWM performance using the results reporting in this paper.

7. Acknowledgement

The research was supported by the National Nature Science Foundation of China (NSFC) (project no.: 71273219).

8. References

Bakshan, A., Srour, I., Chehab, G. and El-Fadel, M. (2015). A field based methodology for estimating waste generation rates at various stages of construction projects. *Resources, Conservation and Recycling*, 100, 70-80.

Ball, S. and Taleb, M.A. (2011). Benchmarking waste disposal in the Egyptian hotel industry. *Tourism and Hospitality Research*, 11(1), 1-18.

Bossink, B.A.G. and Brouwers, H.J.H. (1996). Construction waste: Quantification andsource evaluation. *Journal of Construction Engineering and Management*, 122(1), 55-60.

Brown, B., Chui, M. and Manyika, J. (2011). Are you ready for the era of “big data”? *McKinsey Quarterly*, 4, 24-35.

Camp, R. (1989). Benchmarking: The search for industry best practices that lead to superior performance, *ASQC*, Milwaukee, Wisconsin.

Chen, H., Chiang, R.H. and Storey, V.C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 36(4), 1165-1188.

Cheung, S.O. (2010). Relational Contracting for Construction Excellence: Principles, Practices and Case Studies. *Construction Management and Economics*, 28(7), 805-806.

Clifton, C. (2010). Encyclopedia britannica: definition of data mining. Retrieved 2010-12-09.

Delmater, R. and Hancock, M. (2001). Data mining explained: a manager's guide to customer-centric business intelligence. *Digital Press*.

DEVB TC (2010). Trip ticket system for disposal of construction & demolition materials. Available on website: http://www.devb.gov.hk/filemanager/technicalcirculars/en/upload/308/1/C-2010-06-01.pdf

Eckerson, W.W. (2006). Creating Effective KPIs. *DM Review*, 16(6), 15.

Gangolells, M., Casals, M., Forcada, N. and Macarulla, M. (2014). Analysis of the implementation of effective waste management practices in construction projects and sites. *Resources, Conservation and Recycling*, 93, 99-111.

HKEPD (2014a). Construction Waste Disposal Charging Scheme. Retrieved from website: http://www.epd.gov.hk/epd/misc/cdm/scheme.htm in Sept. 2014

HKEPD (2014b). Introduction to Construction Waste. Retrieved from website: http://www.epd.gov.hk/epd/misc/cdm/introduction.htm in Sept. 2014

Formoso, T.C., Soibelman, M.L., Cesare, C.D. and Isatto, E.L. (2002). Material waste in building industry: main causes and prevention. *Journal of Construction Engineering and Management*, 128 (4), 316–325.

Han, J. and Chang, K.C. (2002). Data mining for web intelligence. *Computer*, 35(11), 64-70.

Han, J. and Kamber, M. (2001). Data mining: concepts and techniques. San Francisco, CA, itd: Morgan Kaufmann, 5.

Han, J., Kamber, M. and Pei, J. (2012). Data mining: concepts and techniques. *Waltham, Mass.: Morgan Kaufmann Publishers*.

Hegazy, M. and Hegazy, S. (2012). The development of key financial performance indicators for UK construction companies. *Accounting, Accountability & Performance*, 17(1/2), 49.

Hong Kong Housing Authority (2014). Forecast public housing production. http://www.housingauthority.gov.hk/en/about-us/publications-and-statistics/forecast-public-housing-production/index.html.

Horta, I.M., Camanho, A.S. and Da Costa, J.M. (2009). Performance assessment of construction companies integrating key performance indicators and data envelopment analysis. *Journal of Construction Engineering and Management*, 136(5), 581-594.

Housing, Planning and Lands Bureau (2005). Building Management and Maintenance - Public Consultation on Mandatory Building Inspection – Leaflet. Available on website: http://www.legco.gov.hk/yr05-06/english/panels/plw/papers/plw1021cb1-124-e.pdf.

Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., Pierre, S., Twigger, S., White, O. and Rhee, S.Y. (2008).Big data: The future of bio-curation. *Nature*, 455(7209), 47-50.

Kerzner H. (2011). Key performance indicators, 97-131 in Project management metrics, KPIs, and dashboards: a guide to measuring and monitoring project performance.

Kim, H., Stumpf, A. and Kim, W. (2011). Analysis of an energy efficient building design through data mining approach. *Automation in Construction*, 20(1), 37-43.

Kofoworola, O. F. and Gheewala, S. H. (2009). Estimation of construction waste generation and management in Thailand. *Waste management*, 29(2), 731-738.

Labrinidis, A. and Jagadish, H.V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, *5*(12), 2032-2033.

Lane, J., Stodden, V., Bender, S. and Nissenbaum, H. (2014). Privacy, Big Data, and the Public Good: Frameworks for Engagement. *Cambridge University Press*.

Lee, S.J. and Siau, K. (2001). A review of data mining techniques. *Industrial Management & Data Syste*ms, 1(1), 41-46.

Lin, G., Shen, G.Q., Sun, M. and Kelly, J. (2011). Identification of key performance indicators for measuring the performance of value management studies in construction. *Journal of Construction Engineering and Management*, 137(9), 698-706.

Lin, K.L. (2006). Feasibility study of using brick made from municipal solid waste incinerator fly ash slag. *Journal of Hazardous Materials*, 137(3), 1810-1816.

Lu, W. (2013). Beyond the inert and non-inert dichotomy: towards 'building a zero waste Hong Kong'. *Building Journal*, 13, 46-49.

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

Lu, W. and Yuan H. (2010). Exploring critical success factors for waste management in construction projects of China. *Resources, Conservation and Recycling*, 55, 201–208.

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

Lu, W. and Yuan, H. (2013). Investigating waste reduction potential in the upstream processes of offshore prefabrication construction. *Renewable and Sustainable Energy Reviews*, 28, 804-810.

Lu, W., Yuan, H., Li, J., Hao, J., Mi, X. and Ding, Z. (2011). An empirical investigation of construction and demolition waste generation rates in Shenzhen city, South China. *Waste management*, 31(4), 680-687.

Lu, W.S., Peng, Y. Webster, C. and Zuo, J. (2015). Stakeholders' willingness to pay for enhanced construction waste management: a Hong Kong study. *Renewable & Sustainable Energy Reviews*, doi:10.1016/j.rser.2015.03.008.

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

Mcdonald, B. and Smithers, M. (1998). Implementing a waste management plan during the construction phase of a project: a case study. *Construction Management and Economics*, 16, 71-78.

Mou, K.Y. (2008). The role of government and construction waste management: a case study of Hong Kong. *HKU Master Dissertation*.

Parmenter, D. (2007). Key performance indicators, Hoboken, NJ: *John Wiley & Sons Publishers*.

Poon, C.S., Yu, A.T.W., Wong, S.W. and Cheung, E. (2004a). Management of construction waste in public housing projects in Hong Kong. *Construction Management & Economics*, 2004, 22(7): 675-689.

Poon, C.S. (2007). Reducing construction waste. *Waste Management*, 27, 1715- 1716.

Poon, C.S., Yu, A.T.W. and Jaillon, L. (2004b). Reducing building waste at construction sites in Hong Kong. *Construction Management and Economics*, 22 (5), 461–470.

Roche, T.D. Hegarty, S., 2006. Best practice guidelines on the preparation of waste management plans for construction and demolition projects. Available at:http://www.envirocentre.ie/includes/documents/BPGConstructionand%20demolition.pdf.

Romero, C. and Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146.

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report*, Fourth Quarter.

Saez, P. V., del Río Merino, M., González, A. S. A. and Porras-Amores, C. (2013). Best practice measures assessment for construction and demolition waste management in building constructions. *Resources, Conservation and Recycling*, 75, 52-62.

Sen, P. K. and Singer, J. M. (1994). Large sample methods in statistics: An introduction with applications, 25. *CRC Press*.

Tam, V.W.Y., Tam, C.M., Zeng, S. X. and Ng, W. C. Y. (2007). Towards adoption of prefabrication in construction. *Building and Environment*. 42(10), 3642–54.

Thomas, G. and Thomas, M., 2008. Chapter 20. Benchmarking and key performance indicators, 109-114, in Construction Partnering & Integrated Team Working. *Published Online: 21 JAN 2008*.

Yuan, H. P., Shen, L. Y., Hao, J. J. and Lu, W. S. (2011). A model for cost–benefit analysis of construction and demolition waste management throughout the waste chain. *Resources, conservation and recycling*, 55(6), 604-612.

Yuan, H. (2011). A dynamic model for assessing the effectiveness of construction and demolition waste management, *Doctoral dissertation, The Hong Kong Polytechnic University*.

Yuan, H., Lu, W. and Hao, J. (2013). The evolution of construction waste sorting on-site. *Renewable & Sustainable Energy Reviews*, 20, 483-490.

Zhang, D. and Zhou, L. (2004). Discovering golden nuggets: data mining in financial application. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* 34(4), 513-522.

Zikopoulos, P. and Eaton, C. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data. *McGraw-Hill Osborne Media*.