Cross-sector sustainability benchmarking of major utilities in the United Kingdom

Walker, Nathan; Styles, David; Coughlan, Paul; Williams, Prysor

Utilities Policy

DOI:
10.1016/j.jup.2022.101417

E-pub ahead of print: 01/10/2022

Publisher's PDF, also known as Version of record

Utilities Policy 78 (2022) 101417

Contents lists available at ScienceDirect

Utilities Policy

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

Full-length article

Cross-sector sustainability benchmarking of major utilities in the United Kingdom

Nathan L. Walker[a,*], David Styles[b], Paul Coughlan[c], A. Prysor Williams[a]

[a] School of Natural Sciences, College of Environmental Sciences and Engineering, Bangor University, Gwynedd, UK
[b] School of Engineering, University of Limerick, Limerick, Ireland
[c] Business School, Trinity College Dublin, Ireland

ARTICLE INFO

Keywords:
Cross-industry
Efficiency
Performance analysis
Sustainability
Utility benchmarking

ABSTRACT

Benchmarking can be a useful tool for utility companies to improve their efficiency, offering many potential positives such as assessing performance objectively, exposing areas where improvement is needed, and identifying best performing companies, ultimately illuminating possible strategies for poorer performers to implement. Despite these positives, the challenge remains of how to compare the performance of different organisations from different sectors. This research aimed to develop a methodology to effectively compare companies across sectors using UK utilities across the water and sewage, energy, and communications sectors as a case study. A methodology was constructed based on service, environmental, and financial metrics, and cross-sector benchmarking was undertaken, which generated performance scores based on company metrics relative to sector peers. This circumnavigated issues of indicators often being mismatched across sectors and the lack of relevance and context when sectors do use similar indicators. Results showed that the sample of 18 utilities had two distinct clusters, one of eight sector leaders and the other of ten lower performers. *Sky* had the highest overall score of 13.5 (maximum 15), suggesting it significantly outperformed the rest of the communications sector. Similarly, *British Gas* and *SSE* lead the energy sector, whilst *Wessex*, *Severn Trent*, and *United Utilities* lead the way for water and sewage companies. The two distinct groups of sector leaders and lower performers can be employed to identify other companies that may offer learning opportunities. Top performers can assess top performers in other sectors to identify how they might continue improving, rather than be potentially limited within their sectors. Conversely, lower-performing companies can look within and across sectors to identify best practices to improve their performance. The methodological development and UK utility sustainability results collectively provide novel insight into the water, energy, and communication sectors and contribute to the international academic literature on benchmarking by illustrating an alternative and unique solution to comparing diverse sectors in any region.

1. Introduction

Benchmarking can be a key efficiency tool (Zhu, 2014), offering many positives such as assessing performance objectively, exposing areas where improvement is needed, and identifying best-performing companies, therefore highlighting potential adoption strategies for poorer performers and a roadmap to improved efficiency (Ecorys, 2012; Molinos-Senante et al., 2021). There are also many indirect benefits to benchmarking, promoting an understanding of company processes, questioning a company's objectives, verifying strategy, strengthening competitive position, and initiating the process of continuous improvement (Krishnamoorthy and D'Lima, 2014). The benefits are so widely understood that benchmarking is a common practice in many industries and sectors to optimise their resources and achieve ambitious goals (Castro and Frazzon, 2017). However, despite the positives of benchmarking, there remains the challenge of comparing the performance of different organisations from different sectors (Bititci et al., 2013).

The concept of cross-sector company benchmarking and performance comparison has many promising elements. Companies could gain value from being able to identify leaders in other sectors to begin the process of learning from them, ultimately instigating improvements in

* Corresponding author.
E-mail address: N.Walker@bangor.ac.uk (N.L. Walker).

https://doi.org/10.1016/j.jup.2022.101417
Received 16 February 2022; Received in revised form 15 August 2022; Accepted 15 August 2022
Available online 24 August 2022

their company, dependent on what they were measuring and analysing (Cankar and Petkovsek, 2013). These benefits, of course, can be gained from within-sector benchmarking (Walker et al., 2021a). However, by transcending sector boundaries and possibly common practice dogma, companies can learn something that could not have been garnered from peers, pushing the efficiency or productivity of their operations further. For example, a higher-performing company may have peaked or even become complacent at the top of their market, but new perspectives may develop when comparing cross-sector at other top companies. Furthermore, it is possible that when only benchmarking within the same sector, a company never beats their competitors, only follows them.

Many companies could benefit from cross-sector comparisons. However, this is particularly true of utilities. UK utilities such as water and sewage companies are private and monopolised, meaning an environment without competition can cause stagnation and relies heavily on regulators to drive efficiency and production progress (Walker et al., 2019). Other utilities may operate in a more competitive environment; however, significant market leaders often can stray toward quasi-monopolies if regulation is ineffective (Heath et al., 2015). This latter behaviour can limit innovation, meaning evaluating relative to leading companies from other sectors can be beneficial.

Camp (1989) developed various benchmarking categories that distinguish between internal, competitive, functional, and generic approaches. Internal compares performance within a company across teams and departments, competitive analyses company performance against the rest of their sector, functional compares cross-sector processes and practices, and generic benchmarking assesses foundational metrics to any industry (Edmondson and Harvey, 2017). Despite the useful divisions of benchmarking approaches, they all utilize similar functions, processes, or firms in their comparisons. An extension of these techniques from Watson (2007) attempted to overcome the limitations of this 'local' benchmarking by expanding the benchmarking geographical boundaries to encompass business and cultural process differences (Walker et al., 2021b). Despite opening the comparative sample wider, global benchmarking still does not necessarily satisfy cross-sector benchmarking, especially of whole companies, with Watson (2007) commenting that global performance benchmarking is chiefly carried out on financial metrics.

The academic literature mimics the varying benchmarking categories to an extent. The closest it gets to cross-sector and industry comparisons of whole companies is the various performance analysis studies that span multiple sectors; however, they do not capture the whole performance of companies. For example, specific processes and functions have been analysed across supply chains (Kim, 2007; Kojima et al., 2008; Akyuz and Erkan, 2009), manufacturing (Bukchin 1998; Laugen et al., 2005; Shou et al., 2017), product validation (Xu et al., 2018), and governments (Chatfield and Reddick, 2017), but there is little information

known beyond these narrow scopes. As affirmed by Richard et al. (2009), performance studies such as those documented above should be undertaken in the context of overall performance. Bititci et al. (2013) present the only study which has attempted to holistically compare companies across sectors with a sample of small and medium enterprises. They developed an effective framework to assess companies based on scoring shared indicators and chose a productive mixture of lagging and leading indicators. However, their indicator choices are still primarily based on financial metrics, much like global benchmarking, and they score indicators through interviews, as opposed to documented data, giving heed to potential inaccuracy and abstract outputs. There is no apparent method to compare the performance of whole companies across the sector in a data-driven and sustainability-focused manner.

Significant barriers must be overcome when attempting cross-sector benchmarking. Watson (2007) and Hawawini et al. (2003) comment that although benchmarking is sufficient when identifying the best performers in a specific industry, it does not function well across industries since a lack of context means comparisons become meaningless. These contextual differences can resonate in the form of the divergent services provided, market conditions, and a general lack of homogeneity. Meaningful comparisons can even be challenging to generate within the same sector when there are differences in the size of companies and even slight differences in services provided (Walker et al., 2020). Furthermore, there are practical limitations due to the variety in indicators, reporting, and auditing, even when overlapping operations and processes are present.

This study had two main objectives. Firstly, to develop a universal methodology to compare whole companies effectively across sectors. Secondly, to test the methodology with a UK case study and ultimately investigate the performance of UK utilities across the water and sewage, energy, and communications sectors. Collectively, these objectives provide novel insights for the water, energy and communication sectors and contribute to the international academic literature on benchmarking by illustrating a methodological framework that enables cross-sector performance comparisons in any region and contributing results and analysis of UK utility sustainability performance.

2. Methodology

2.1. Data description

Suitable organisations and the exploration of documentary evidence of each had to be collated to satisfy the study objectives. The sample consisted of 18 companies that spanned three sectors: water and sewage, energy, and communications. Despite covering multiple sectors, it was still essential to keep the peers within each sector homogenous in size to enable fair comparisons, which this study did by choosing the largest and most mature companies in each sector. By ensuring true peers were compared, scoring was fairer since all companies could compete and were not capped and limited in some metrics, especially when using normalised data.

The metrics collected for each company attempted to cover all aspects of sustainability and company performance, so service, environmental, and financial data were collected and analysed. These data came from numerous sources; however, they primarily consisted of self-reported statistics in audited company reports. Independently verified data were used where available; for example, this was the case for customer satisfaction and credit rating data across all industries; an extensive list of data sources is available in the Supplementary Information. Nine water and sewage companies were assessed between 2015 and 2020, using eight indicators (two service, three environmental, and three financial). The energy sector was represented by the largest five companies over seven years (2014–2020) using six indicators (two service, two environmental, and two financial). Lastly, four communications companies were included in the study over four years (2017–2020) and were analysed using seven indicators (four service, two environmental, and one financial). The dataset began as large as possible before narrowing due to data availability and quality, which was needed for all companies in a sector. For some sectors, gathering, analysing, and sorting data was more feasible than in others, which is why there is a mismatch of sample years; fortunately, this imbalance is not extreme enough to skew results. All indicators and their summary statistics for each sector are displayed in Table 1.

Data availability for all companies within a sector was the key reason for there not being more indicators and years covered. However, each company's significant operations and targeted themes (service, environmental, and financial) were encapsulated over enough time to overcome anomaly years. Moreover, chosen indicators for each group theme were analogous for all sectors where possible, ensuring any comparisons were credible. As Bititci et al. (2013) document, comparative measurement systems should be balanced, include financial and non-financial measures based on a time series, and be sensitive to contextual and environmental operation conditions, and the indicators chosen in this study satisfy these requirements. The positives and

Table 1
Summary statistics for all indicators in each sector.

	Water & sewage		Energy		Communication	
	Average	SD	Average	SD	Average	SD
Service indicators						
Customer Satisfaction	82.08	4.47	70.40	3.56		
Complaints received per 100 k			2284.06	794.75		
Customer Satisfaction - Broadband (%)					82.81	4.35
Customer Satisfaction - Broadband (Complaints/100 k)					66.63	34.74
Customer Satisfaction - Landline (%)					85.63	3.94
Customer Satisfaction - Landline (Complaints/100 k)					49.94	22.33
Water supply interruptions (mins/properties)	12.67	7.63				
Environment indicators						
Leakage (Ml/day)	207.22	190.25				
Renewable energy self gen (%)	18.41	8.44	23.06	15.49		
Carbon Intensity - $kgCO_2e$ per Ml	415.14	184.49				
Carbon Intensity - $CO_2(g)/kWh$			281.63	146.04		
Carbon Intensity - tCO_2e/100 k customers					1957.24	1206.98
Carbon Intensity - tCO_2e/£m revenue					13.04	12.80
Financial indicators						
Post-tax return on regulated equity (%)	7.00	4.46				
Gearing (%)	68.67	8.25				
Interest cover	2.05	0.75				
EBITDA (£m)			1650.34	571.13		
Credit Rating			7.97	0.81	10.94	1.96

*Water and sewage sector based on SIM scores; energy sector based on USwitch scores.
**EBITDA = Earnings Before Interest, Taxes, Depreciation, and Amortization.
*** Credit rating based on Fitch and Moody's rating scales and converted to numbers for ease of comparison.

limitations of indicator selection are more widely discussed in Section 3.2.

2.2. Scoring system

The performance indicators measured were split into each themed group (service, environmental, and financial), and each group was weighted the same. Themes were based on the traditional sustainability pillars of social, economy, and environment (Purvis et al., 2019), with social being reassigned as service due to the nature of the service indicators overlapping with social requirements. Scores were generated for each indicator and company based on their performance within their sector by assessing the total performance range for each indicator and splitting it into five equal segments. These five equal ranges made the boundaries on which a company's average performance over the sample years was scored. Alternative interval numbers were evaluated as a sensitivity analysis, such as with ten intervals; however, there was no significant difference when the companies were ranked, with an average rank change of 0.83 (available in the Supplementary Information). Tests using fewer than five intervals would have too large of a margin in assigning performance values, so companies performing considerably different could have the same score. The generated scores from indicator results were averaged within each themed indicator group (service, environmental, and financial) to evaluate a breakdown of the performance of the various companies. Since the distribution and scales used to score indicators were integral, the distribution of each of the 21 indicators was tested using the Shapiro-Wilk method. Indicators under non-normal distribution had company averages and subsequent scoring systems based on the geometric mean, whilst normal distributions used the arithmetic mean. This approach ensured accurate averages were displayed even with non-normal distributions and outliers, enabling indicators and companies to have non-skewed results.

Table 2 provides an example illustration of the scoring system; a full breakdown of each indicator is available in the Supplementary Information. Assessing company performance relative to their peers enabled a fair indicator comparison, and the generated scores allowed a cross-sector comparison. These generated scores and the comparisons are thus based purely on how companies perform relative to their own sector, enabling a cross-sector comparison of intra-sector ranking; they are not being directly compared to all companies in other sectors with raw metrics.

Table 2
Example of how an indicator was scored based on the range of companies in the same sector.

	Dummy indicator
Sample MIN	75.10
Sample MAX	86.68
Interval	2.32
1	75.10–77.41
2	77.42–79.73
3	79.74–82.05
4	82.06–84.36
5	84.37–86.68
Company x average	83.08
Company x score	4

2.3. Clustering and visualisation methods

The clustering techniques utilised were K-means and hierarchical. The K-means approach is an unsupervised machine-learning algorithm, where for a predefined number of centroids, data points (company scores, in this instance) are grouped depending on how close they are to one of the centroids, using the Euclidean distance metric (Mishra et al., 2012). The predefined number of centroids or 'clusters' was determined as two using the elbow and silhouette methods. The elbow method operates by varying the number of clusters from one to ten, then calculating the within-cluster sum of squares (distance between each point and the centroid). Then when plotted, the optimum number of clusters is revealed when the dataset becomes linear (Nanjundan et al., 2019). The silhouette method is a measure of how a clustering algorithm has performed. After computing the silhouette coefficient of each point in the dataset, a measure is taken to evaluate how similar a point is to its own cluster compared to other clusters. The optimum cluster is highlighted when most points have a high match to their own clusters (Kwon et al., 2018).

The hierarchical clustering used was agglomerative based, where each data point is initially assigned to a cluster, then iterations are made, and the closest pairs of clusters are merged until only one cluster is left (Wu et al., 2021). The 'Ward's Method' was applied, which minimises the increase in the total within-cluster sum of squared error, where the increase is proportional to the squared Euclidean distance between cluster centres (Murtagh and Legendre, 2014). The dendrogram produced from this showed the optimum number of clusters. All analysis and visualisation were performed using statistical computing software 'R'. The K-means computation was conducted with built-in functions within R; however, the elbow and silhouette methods used the 'factoextra' package (Kassambara and Mundt, 2020), and the 3D plots were written within 'rgl' (Murdoch et al., 2021). Lastly, the hierarchical clustering utilised the package 'cluster' (Maechler et al., 2021), and the dendrogram was made again in 'factoextra'.

3. Results and discussion

3.1. Utility performance

The water and sewage, communication and energy sectors control most major utilities, responsible for the operation of every country. They are suitably often large and operate as monopolies, and the UK companies within these sectors are no exception. These conditions make performance comparisons between these sectors desirable due to the validity of results and the effectiveness of outcomes. This section analyses companies from all sectors in terms of total and thematic performance and individual indicator performance in instances of exceptional results before clustering sector leaders and lower performers together.

The performance of companies was scored based on individual indicator performance relative to their sector peers, with 1 being the worst and 5 being the best scores. These scores ultimately were representations of sector-normalised relative performance across sectors. The indicators were grouped according to service, environmental, and financial; indicator scores were averaged within those groups. The results for these groups for each of the 18 companies are displayed in Fig. 1.

The classification of indicators into the three groups enables an assessment of broad areas in which companies perform well or could improve upon relative to their peers. For example, with a score of 5, *Talk Talk* performed the best in environmentally focussed metrics compared to other large communications companies; however, their service and financial performance was the worst, with scores of 1 for both groups. A similar mismatched performance in the communications sector is evident for *Virgin Media*, although to a lesser extent, where they performed relatively well in their service metrics but lagged and scored 1 for the environmental and financial indicators. As the grouping scores reflect the indicators, it is possible to trace which aspects can be improved. For example, for *Talk Talk* and *Virgin Media*, their financial group scores depend on their credit scores. Although credit scores are given to measure the likelihood of companies defaulting on their debt, which depends on many aspects such as financial history, liabilities, and assets (Duffie and Singleton, 2012), it gives a good starting point for evaluating areas to improve. Some indicators within groups can offer more specific targets to pinpoint improvements. For example, both companies could have improved their service scores by reducing broadband and landline complaints and increasing customer satisfaction. This assessment provides the initial phase of investigation where further internal auditing can reveal opportunities for progress, possibly requiring additional customer service staff to reduce solution times or improved asset management to reduce service outages (Chambers and

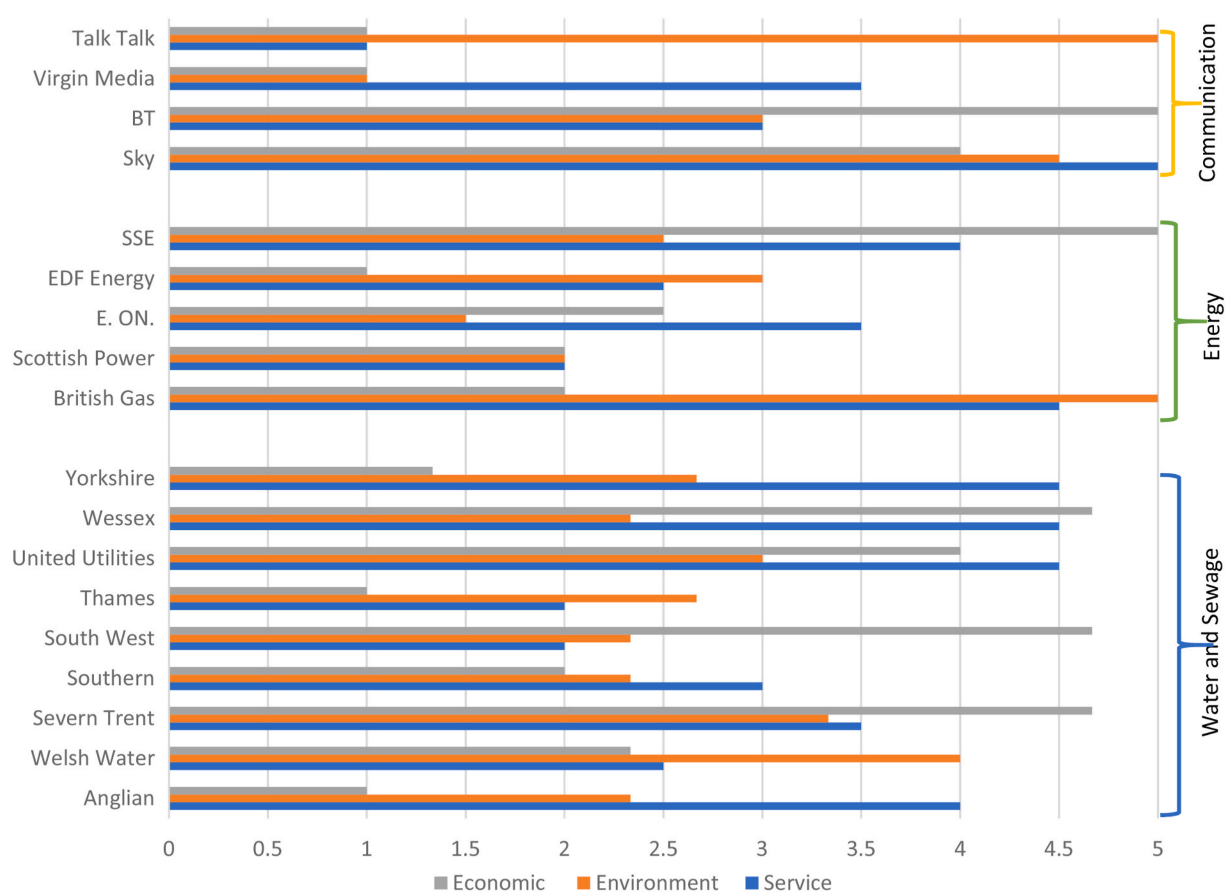

Fig. 1. Average service, environmental, and financial group scores for all companies, separated into their sectors.

Odar, 2015).

For some companies, there are no significant disparities between the grouping of indicators, and they are well balanced across service, environmental, and financial indicators. This situation applies to *Severn Trent,* which scored 3.5, 3.3, and 4.6 for service, environmental, and financial metrics, respectively. These findings suggest they can be content with their current investment and prioritisation strategies, although there are still growth opportunities. Balance across the indicator groups is not necessarily a positive attribute, as *Scottish Power* and *Thames* demonstrate with their relatively low scores. However, further exploration into specific indicators is again possible to begin to evaluate areas where progress needs to be made and make action plans to carry it out.

Fig. 2 presents the aggregated scores for each company to identify the sector leaders. Company scores can range from 3 to 15 since each group's minimum score was 1 and maximum was 5. This approach gives a clear view of who are the sector leaders. For example, *Sky* heads the communications sector with the highest aggregated score of 13.5. Similarly, *British Gas* and *SSE* lead the energy sector, whilst *Wessex*, *Severn Trent*, and *United Utilities* lead the way for water and sewage companies. With this approach of scoring and performance analysis, companies can see who is performing best overall in their sector with the metrics chosen, as well as the broad indicator grouping and varying individual indicator performances. The lower-performing companies have clear pathways to better efficiencies and performance since best practices in various areas of the company are distinguishable.

Despite the positives of the scoring system and intra-sector comparisons, the higher-performing companies in their sectors have limited opportunities to learn from best practices, which is where understanding the leaders in other industries and comparing cross-sector can be advantageous. Cluster analysis evaluated and highlighted the sector leaders and lower performers; K-means (linear) and hierarchical (quadratic) clustering validated the results since they utilize different algorithms (Garikapati et al., 2021). The K-means results depicted two distinct clusters (Fig. 3), which categorised higher performers into a group of seven and lower performers into a group of 11. The key difference between the two clusters appears to be the financial results since the leading companies perform consistently better in this indicator grouping, whereas for environmental and social groupings, there are mixed results for all companies.

The hierarchical clustering generated two clusters, supporting the 'elbow' and 'silhouette' methods that pre-assigned the number of clusters for the K-means analysis. The sizes of these clusters had a slight divergence from the K-means results, with the sector leaders comprising a group of eight and the lagging companies comprising a group of 10 (Fig. 4). The only difference in these clusters was the placement of *British Gas*, which was placed in the sector-leading group in the hierarchical clusters but in the low-performing group when using K-means. This finding appeared again to be due to financial performance, with *British Gas* having a score of just 2; however, when viewing the aggregated scores and cluster results displayed in Table 3, *British Gas* are clearly an energy sector leader, having an aggregated score of 11.5.

The advantage of clustering these grouping results is that companies can use the two distinct groups of higher and lower performers to identify other companies that may offer learning opportunities. This approach can be conducted with any inputted sample; however, with the UK companies, top performers such as *Sky*, *SSE,* and *Wessex* can assess leaders in other sectors, either as a whole or based on indicator groupings, to continue improving, rather than be potentially limited within their own sectors. Conversely, lower performers can look within and across sectors to identify best practices to improve their operations. The learning opportunities may be various since best practice is assessed cross-sector where different company processes are present. It is reasonable that the clustering has captured some factors that are not explicit in the traditionally reported indicators since companies generate their outputs and data based on many varying internal structures and procedures. By processing as much holistic information as possible and displaying the performance of companies in two distinct groups that possess diverging properties, it provides the initial steps and possibility to learn from others in overlapping processes and more abstract and lateral ways. These 'hidden' determinants are likely to be broadly based on business strategy, structure and culture, effective layered management for large companies, and creating unity between shareholder, regulator, and customer needs. These macro areas of potential improvement make a positive starting point before delving into the details of implementation, which can take the form of individual processes or strategies, from offsetting carbon to improving customer satisfaction.

3.2. Methodology review

The methodological approach in this study gathered 739 data points of 21 indicators for 18 different companies, spanning between 4 and 7

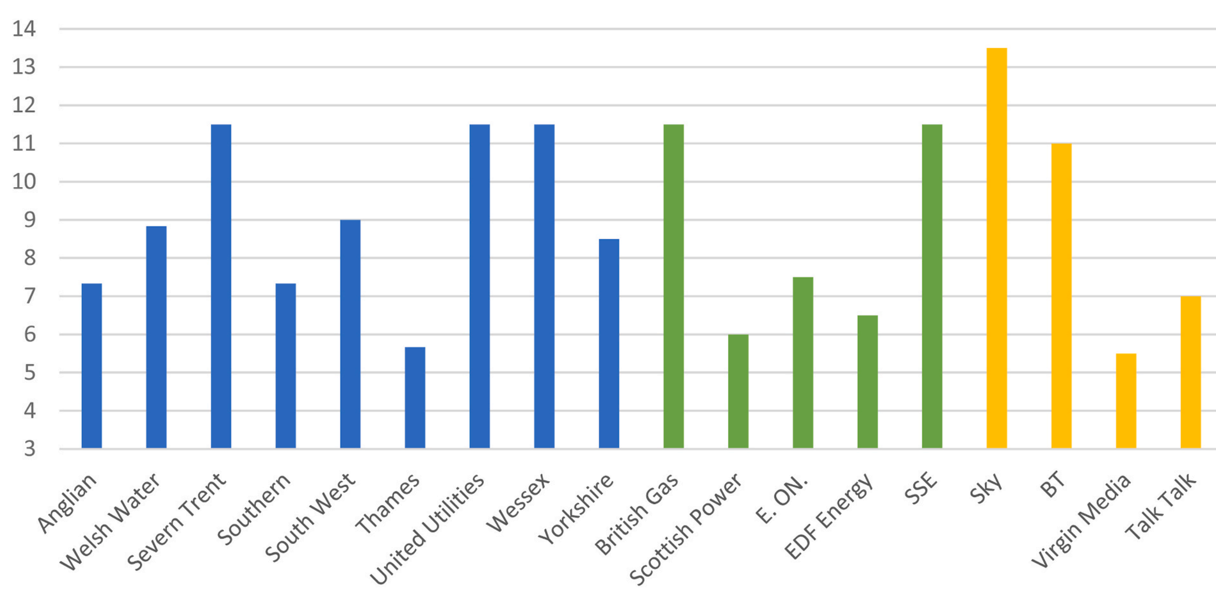

Fig. 2. Aggregated scores from the average service, environmental, and financial scores (blue = water companies, green = energy companies, yellow = communication companies).

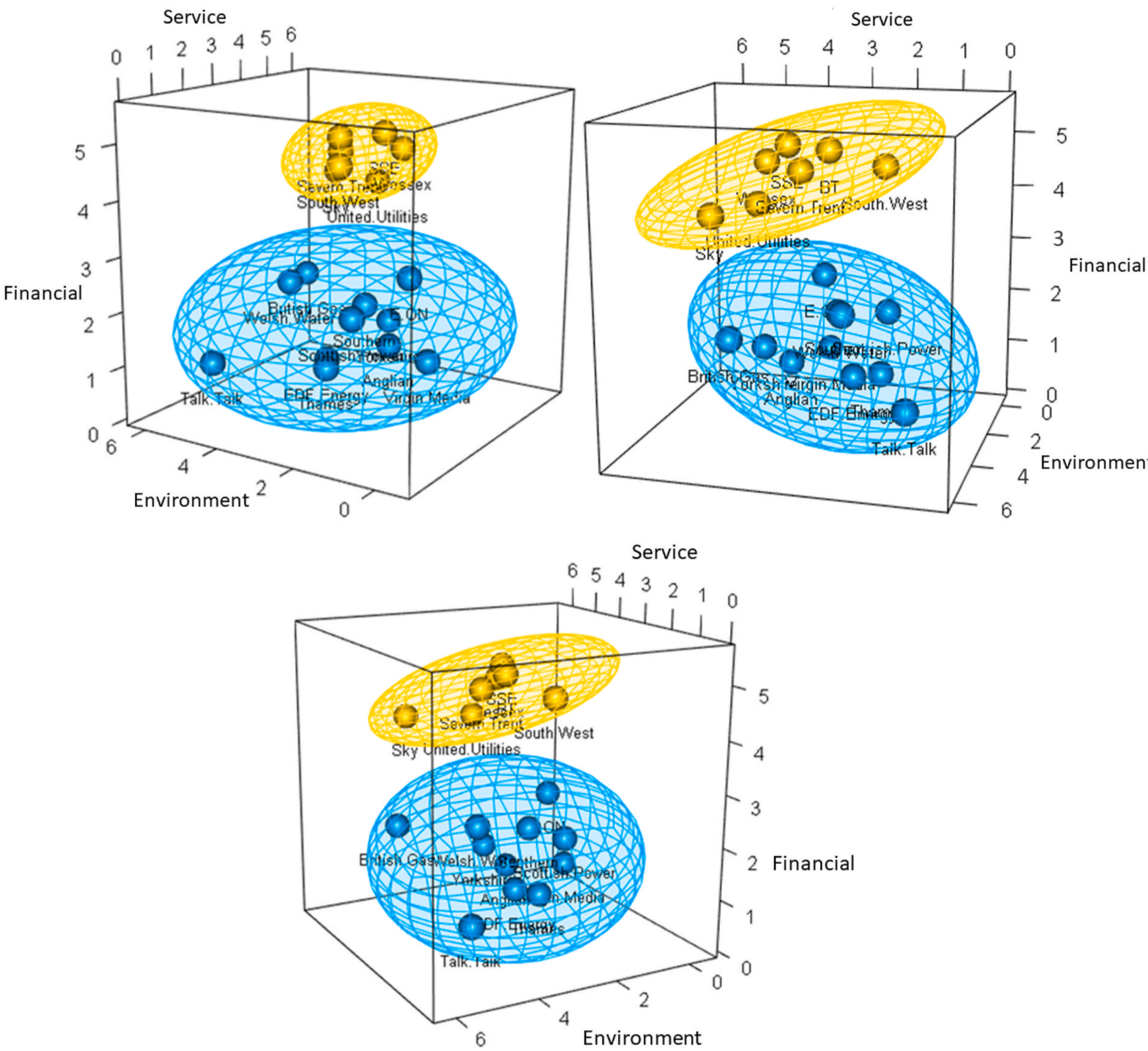

Fig. 3. Singular 3-D plot of company group scores with highlighted K-means clusters rotated on the z-axis.

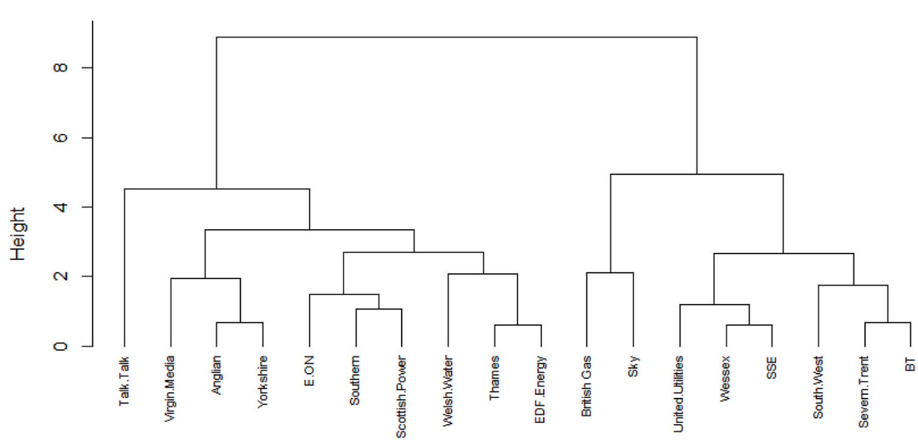

Fig. 4. A dendrogram of the hierarchical Ward method of clustering.

years and 3 sectors. These data and indicators were thematically grouped according to service, environmental, and financial, enabling cross-sector comparisons through re-generated performance scores based on company performance relative to their industry peers. Whilst the results are of interest themselves, they act as a practical illustration of how the methodology can be used in broader settings, other countries, and further studies. Comparing company performances via various metrics made it possible to see where companies ranked in each sector,

Table 3
Summarised company performance based on aggregated score and hierarchical and K-means clustering.

Company	Total (sum)	Rank	Hierarchical clusters	K-means clusters
Sky	13.5	1	1	1
Severn Trent	11.5	2	1	1
United Utilities	11.5	3	1	1
Wessex	11.5	4	1	1
British Gas	11.5	5	1	2
SSE	11.5	6	1	1
BT	11	7	1	1
South West	9	8	1	1
Welsh Water	8.83	9	2	2
Yorkshire	8.5	10	2	2
E. ON.	7.5	11	2	2
Anglian	7.33	12	2	2
Southern	7.33	13	2	2
Talk Talk	7	14	2	2
EDF Energy	6.5	15	2	2
Scottish Power	6	16	2	2
Thames	5.67	17	2	2
Virgin Media	5.5	18	2	2

revealing sector leaders and lower performers. This approach was deemed the only valid way to compare whole companies across different sectors since original indicators are often mismatched across sectors. Even when they use the same metrics, values lack relevance and context when used on companies that provide different services in different market conditions. By adding comparable validity, the scoring system significantly affected conclusions that would otherwise be drawn by directly comparing indicators if the metrics could even be compared. For example, if credit ratings were directly compared across the sectors, *Sky*, who ranked highest and had the second-best credit rating of the communications companies with an average of 10 (BBB- on the Fitch rating scale), would have ranked significantly lower since all energy companies scored between 7.29 and 8.43 (Fitch rating range of A-to BBB+). The natural differences between sectors are further emphasised across most indicators, such as customer satisfaction and carbon intensity, which cannot be compared due to a conflict of metrics. This result is not just a symptom of differing services; data availability and pre-conducted normalisation are also barriers.

Despite the positives that the scoring system can offer, there are limitations and the capacity to enhance and customise the methodology further. For example, it still cannot directly compare companies' raw metrics across the sector, it is based on comparing their intra-sector performances, so there are real trade-offs between detailed process-level benchmarking, especially within sectors, and macro-level cross-sector benchmarking. Furthermore, the practicality of the method relies on optimal indicator decisions. For example, company results highly depend on chosen indicators, which opens considerations of who chooses the indicators and why, and the role data availability and transparency can have. This study had the sample period mismatched for the sectors due to data availability, which could have influenced results since exogenous factors influencing all sectors may not have been captured, especially if the years were even more varied. The data utilised for the analysis incorporated self-reported and independently verified metrics to maximise the sample but possibly lowered the validity of some results. Other cross-sector benchmarking efforts could have significantly differed results with different data constraints.

Similarly, this study chose to incorporate a holistic view of sustainability, meaning that environmental indicators were utilised due to their increasing importance. However, if only the service and financial indicators were applied, the results would have been substantially different, with *Wessex* being the top-rated company and *Talk Talk* being the lowest. Furthermore, getting variation and representation within each indicator group ensures that one indicator does not skew the overall group score. To an extent, this can be controlled by carefully considering the weighting of individual indicators and the indicator groups, which also heavily influences conclusions. In this study, all weightings were equal; however, it is something to consider when extrapolating and attempting to apply results.

Future studies may want to contemplate stakeholder engagement to narrow down the priority of the research and ultimately the aim of the study, which will then help choose and weight the metrics. Future research could further develop the approach presented here by using the generated comparable values as inputs and outputs in production frontier analysis. This approach would enable an understanding of how companies are performing by comparing inputs and outputs relative to one another within the sample, e.g., companies performing the best in their outputs whilst minimising inputs would perform the production frontier for the other companies to be benchmarking against (Walker et al., 2021a). The limitations of this approach vary depending on the specific methodology. However, queries of sample size would have to be satisfied with regard to the number of indicators relative to the number of companies, which would have been the obstruction faced by this study.

4. Conclusions

The goals of this research were to investigate the performance of UK utilities across the water and sewage, energy, and communications sectors and to develop a methodology to compare whole companies across sectors effectively. Results support two main conclusions. Firstly, the methodology developed used a scoring framework enabling successful cross-sector comparisons with generated performance scores based on company performance relative to sector peers. This method was deemed a constructive way to compare whole companies across different sectors since original indicators are often mismatched across sectors. Even when they use the same metrics, values lack relevance and context when used on companies that provide different services in different market conditions. The second conclusion is that the sample used to illustrate the scoring framework of 18 utilities is in two distinct clusters, one of eight sector leaders and the other of ten lower performers. *Sky* had the highest overall score of 13.5, suggesting that it significantly outperformed the rest of the communications sector.

Similarly, *British Gas* and *SSE* lead the energy sector, whilst *Wessex*, *Severn Trent*, and *United Utilities* lead the way for water and sewage companies. The two distinct groups of sector leaders and lower performers can be employed to identify other companies that may offer learning opportunities. Top performers such as *Sky*, *SSE* and *Wessex* can assess top performers in other sectors to continue improving rather than being limited within their own sectors. Conversely, lower companies can look within and across sectors to identify best practices to improve their performance.

A further investigation into better or worse performance areas is possible since total scores were based on metrics spanning three main groups: service, environmental, and financial. To illustrate, *Talk Talk* performed the best in environmentally focussed metrics compared to other communications companies with a score of 5; however, its service and financial performance were the worst, with scores of 1 for both groups. Further research on indicator selection and weighting is recommended, possibly utilising stakeholder engagement, particularly in response to data availability and representation of core processes. The methodology developed fills a knowledge gap in comparing companies across sectors and with the UK utility sustainability performance results, provides novel insight into the water, energy and communication sectors and contributes to the international academic literature on benchmarking.

Author contributions

Nathan L. Walker: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualisation; Writing – original draft.David Styles: Conceptualization; Validation; Supervision;

Writing – review & editing. Paul Coughlan: Validation; Writing – review & editing. A. Prysor Williams: Conceptualization; Validation; Supervision; Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data used are present in Supplementary Information and sources are linked too.

Acknowledgments

This research was undertaken within the Dŵr Uisce project, which is partly funded by the European Regional Development Fund (ERDF) through the Ireland Wales Co-operation programme 2014–2020 (grant number 14122).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jup.2022.101417.

References

Akyuz, G.A., Erkan, T.E., 2009. Supply chain performance measurement: a literature review. Int. J. Prod. Res. 35 (1), 1–19.

Bititci, U.S., Firat, S.U.O., Garengo, P., 2013. How to compare performances of firms operating in different sectors? Prod. Plann. Control: Manag. Oper. 24 (12), 1032–1049. https://doi.org/10.1080/09537287.2011.643829.

Bukchin, J., 1998. A comparative study of performance measures for throughput of a mixed model assembly line in a JIT environment. Int. J. Prod. Res. 36 (10), 2669–2685.

Camp, R.C., 1989. Benchmarking: the Search for Industry Best Practices that Lead to Superior Performance. ASQC/Quality Press, Wisconsin.

Cankar, S.S., Petkovsek, V., 2013. Private and public sector innovation and the importance of cross-sector collaboration. J. Appl. Bus. Res. 29 (6), 1597–1606. https://doi.org/10.19030/jabr.v29i6.8197.

Castro, V.F.d., Frazzon, E.M., 2017. Benchmarking of best practices: an overview of the academic literature. Benchmark Int. J. 24 (3) https://doi.org/10.1108/BIJ-03-2016-0031, 750-744.

Chambers, A.D., Odar, M., 2015. A new vision for internal audit. Manag. Audit J. 30 (1), 34–55. https://doi.org/10.1108/MAJ-08-2014-1073.

Chatfield, A.T., Reddick, C.G., 2017. A longitudinal cross-sector analysis of open data portal service capability: the case of Australian local governments. Govern. Inf. Q. 34 (2), 231–243. https://doi.org/10.1016/j.giq.2017.02.004.

Duffie, D., Singleton, K.J., 2012. Credit Risk: Pricing, Measurement, and Management. Princeton university press, Princeton.

Ecorys, 2012. Study on Incentives Driving Improvement of Environmental Performance of Companies. Ecorys, Rotterdam.

Edmondson, A.C., Harvey, J.-F., 2017. Extreme Teaming: Lessons in Complex, Cross-Sector Leadership. Emerald Publishing, Bingley.

Garikapati, P., Balamurugan, K., Latchoumi, T.P., Malkapuram, R., 2021. A cluster-profile comparative study on machining AlSi7/63% of SiC hybrid composite using agglomerative hierarchical clustering and K-means. Silicon 13, 961–972. https://doi.org/10.1007/s12633-020-00447-9.

Hawawini, G., Subramanian, V., Verdin, P., 2003. Is performance driven industry- or firm- specific factors? A new look at the evidence. Strat. Manag. J. 24 (1), 1–16.

Heath, J., Moriarty, J., Norman, W., 2015. Business ethics and (or as) political philosophy. Bus. Ethics Q. 20 (3), 427–452. https://doi.org/10.5840/beq201020329.

Kassambara, A., Mundt, F., 2020. Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7. https://cran.r-project.org/web/packages/factoextra/index.html.

Kim, W.S., 2007. Organizational structures and the performance of supply chain management. Int. J. Prod. Econ. 106 (2), 323–345.

Kojima, M., Nakashima, K., Ohno, K., 2008. Performance evaluation of SCM in JIT environment. Int. J. Prod. Econ. 115 (2), 439–443.

Krishnamoorthy, B., D'Lima, C., 2014. Benchmarking as a measure of competitiveness. Int. J. Process Manag. Benchmark. 4 (3), 342–359. https://doi.org/10.1504/IJPMB.2014.063240.

Kwon, B.C., Eysenbach, B., Verma, J., Ng, K., De Fillipi, C., Stewart, W.R., Perer, A., 2018. Clustervision: visual supervision of unsupervised clustering. IEEE Trans. Visual. Comput. Graph. 24 (1), 142–151. https://doi.org/10.1109/TVCG.2017.2745085.

Laugen, B.T., Acur, N., Boer, H., Frick, J., 2005. Best manufacturing practices: what do the best-performing companies do? Int. J. Oper. Prod. Manag. 25 (2), 131–150.

Maechler, M., et al., 2021. 'cluster: finding Groups in Data'. R package version 2.1.2. https://cran.r-project.org/web/packages/cluster/cluster.pdf.

Mishra, B.K., Rath, A., Nayak, N.R., Swain, S., 2012. Far efficient K-means clustering algorithm. In: Proceedings of the International Conference on Advances in Computing, Communications and Informatics., pp. 106–110. https://doi.org/10.1145/2345396.2345414. August 2012.

Molinos-Senante, M., Maziotis, A., Sala-Garrido, R., 2021. Benchmarking the economic and environmental performance of water utilities: a comparison of frontier techniques. Benchmark Int. J. https://doi.org/10.1108/BIJ-08-2021-0481. Vol. ahead-of-print No. ahead-of-print.

Murdoch, D., et al., 2021. Rgl: 3D visualization using OpenGL. R package version 0.108.3. https://cran.r-project.org/web/packages/rgl/index.html.

Murtagh, F., Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J. Classif. 31, 274–295. https://doi.org/10.1007/s00357-014-9161-z.

Nanjundan, S., Sankaran, S., Arjun, C.R., Anand, G.P., 2019. Identifying the number of clusters for K-Means: a hypersphere density based approach. Int. Conf. Comp, Comm. Sig. Proc. https://doi.org/10.48550/arXiv.1912.00643.

Purvis, B., Mao, Y., Robinson, D., 2019. Three pillars of sustainability: in search of conceptual origins. Sustain. Sci. 14, 681–695. https://doi.org/10.1007/s11625-018-0627-5.

Richard, P.J., Devinney, T.M., Yip, G.S., Johnson, G., 2009. Measuring organizational performance: towards methodological best practice. J. Manag. 35 (3), 718–804. https://doi.org/10.1177/0149206308330560.

Shou, W., Wang, J., Wu, P., Wang, X., Chong, H., 2017. A cross-sector review on the use of value stream mapping. Int. J. Prod. Res. 55, 3906–3928. https://doi.org/10.1080/00207543.2017.1311031.

Walker, N.L., Norton, A., Harris, I., Williams, A.P., Styles, D., 2019. Economic and environmental efficiency of UK and Ireland water companies: influence of exogenous factors and rurality. J. Environ. Manag. 241 (December 2018), 363–373. https://doi.org/10.1016/j.jenvman.2019.03.093.

Walker, N.L., Williams, A.P., Styles, D., 2020. Key performance indicators to explain energy & economic efficiency across water utilities, and identifying suitable proxies. J. Environ. Manag. 269, 110810 https://doi.org/10.1016/j.jenvman.2020.110810.

Walker, N.L., Styles, D., Gallagher, J., Williams, A.P., 2021a. Aligning efficiency benchmarking with sustainable outcomes in the United Kingdom water sector. J. Environ. Manag. 287, 112317 https://doi.org/10.1016/j.jenvman.2021.112317.

Walker, N.L., Williams, A.P., Styles, D., 2021b. Pitfalls in international benchmarking of energy intensity across wastewater treatment utilities. J. Environ. Manag. 300, 113613 https://doi.org/10.1016/j.jenvman.2021.113613.

Watson, G.H., 2007. Strategic Benchmarking Reloaded with Six Sigma: Improve Your Company's Performance Using Global Best Practice. John Wiley & Sons, New Jersey.

Wu, C., Peng, Q., Lee, J., Leibnitz, K., Xia, Y., 2021. Effective hierarchical clustering based on structural similarities in nearest neighbor graphs. Knowl. Base Syst. 228, 107295 https://doi.org/10.1016/j.knosys.2021.107295.

Xu, Y., Tiwari, A., Chen, H.C., Turner, C.J., 2018. Development of a validation and qualification process for the manufacturing of medical devices: a case study based on cross-sector benchmarking. Int. J. Process Manag. Benchmark. 8 (1), 79–102. https://doi.org/10.1504/IJPMB.2018.088658.

Zhu, J., 2014. Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets. Springer, New York.