Chapter 11
A Delphi Pilot Study to Assess the Impact of Location Factors for Hyperscale Data Centres

David King, Nadeeshani Wanigarathna, Keith Jones, and Joseph Ofori-Kuragu

Abstract The rapid expansion of digital technology used by individuals and organisations has resulted in the need for many data centres on a rapid scale. Data centres are high energy-consuming buildings during their operation. Therefore, data centres are increasingly being built in regions with good access to energy infrastructure and cold climates to reduce the cooling energy demands of the facilities. This paper presents location factors that must be considered when estimating and modelling the capital cost of data centres built outside the UK. The research presented here is based on a questionnaire survey conducted as a pilot study of an impending Delphi study. The participant's responses were obtained via an open-ended questionnaire with a cross-sectional timeframe. A thematic analysis of the responses revealed six overarching themes: ease of doing business, design, customer pricing, land, power and fibre, and supply chain. The three most dominant are land, power and fibre and supply chain. This pilot study confirmed the knowledge gap and supported the need that further investigatory work is required. A complete Delphi study will develop the themes identified in this pilot study to achieve consensus on their significance and support an assessment of the impact of location on the modelling and forecasting of capital expenditure for hyperscale data centres.

Keywords Data centres · Location · CAPEX · Delphi

D. King (✉) · N. Wanigarathna · K. Jones · J. Ofori-Kuragu
Anglia Ruskin University, Chelmsford, UK
e-mail: dk555@pgr.aru.ac.uk; nadeeshani.wanigarathna@aru.ac.uk; keith.jones@aru.ac.uk; joseph.ofori@aru.ac.uk

11.1   Introduction

Data centres are buildings that house information technology (IT) equipment to store and process data and services such as the internet (Whitehead et al., 2015). The rapid expansion of digital technologies resulted in building many data centres on a rapid scale (Parker, 2019). Hyperscale data centres are purposefully built large-scale data centre facilities above 20 MW, owned, and typically operated by the organisation they support, such as Apple or Facebook (Christensen et al., 2018). They are usually service platforms for social media, search engines, communication, entertainment, virtual reality, artificial intelligence, machine learning and e-commerce (ibid).

Data centres are high energy-consuming buildings during their operation. Therefore, hyperscale data centre site selection is often predicated on access to cheap power and utilities and in locations with substantial energy infrastructures already in place (Hogan, 2015: 4; Rosenwald, 2011; cited in Parker, 2019). Due to IT demands and cooling equipment primarily and lighting, power distribution and other requirements, they are very energy-intensive during their operation (Avgerinou et al., 2017). Therefore, energy efficiency became a primary concern of data centre development (Wahlroos et al., 2018). The Nordic region has attracted significant investment in new data centres with massive investments from the cloud and hyperscale investors like Facebook, Google, AWS and Apple (Christensen et al., 2018). These regions are preferred due to advanced technological progress in the sector and the cold climate conditions that reduce the cooling energy demands of the facilities (Avgerinou et al., 2017). This trend in locating data centres outside of the UK presents a significant challenge to cost consultants within the UK during the estimating and modelling capital cost of these buildings.

Cost planning (or cost modelling) at the feasibility stage is defined by the Royal Institution of Chartered Surveyors (RICS) as 'the determination of possible cost of a building(s) early in the design stage in relation to the employer's fundamental requirements. This takes place before the preparation of a full set of working drawings or bills of quantities and forms the initial build-up to the cost planning process' (RICS, 2012).

During the cost planning in early development phases, cost consulting professionals often use historical cost data as base cases and adjust their costs to suit the circumstances of new projects. Whilst the impact of specific characteristics such as shape, inflation and specifications is easy to adjust based on case-based reasoning, the impact of location is a difficult factor to predict for construction professionals. Therefore, they rely on location cost indices for this purpose. For example, several location cost indices such as Spon's Architects' and Builders' Price Book 2022 (Aecom, 2021) and the Building Cost Information Service (Royal Institution of Chartered Surveyors, 2021) are available for cost consultants to use. Similar indices are available in other European countries, yet they are less relevant for data centres. Likewise, there are often no precedents set to use as a baseline for cost comparisons. An example would be that specific standards and regulations for noise attenuation

for hyper-size generators for data centres did not exist in Sweden and had to be modelled on regulations from other countries (Vonderau, 2017). Furthermore, compounding the lack of available cost data is the secrecy surrounding data centres, their operations, cost and locations (Holt & Vonderau, 2015).

Whilst international location cost indices, such as those provided by Eurostat (European Union, 2021), World Bank (World Bank, 2021) and the OECD (Organisation for Economic Co-operation and Development, 2021) otherwise known as purchasing power parities, are broad and primarily model variations at country level, they are less effective during the cost planning for individual projects specific to a particular region as there are many variables ranging from macroeconomic, construction methodology and geographical and geological categories. However, there is a gap in knowledge about what variables and types are relevant to the modelling and forecasting of hyperscale data centres. Through the exploration and literature review, it has been established that a wide range of variables impact construction projects' cost and cost modelling. However, there is no evidence identifying if and how these variables would impact the site location on the cost planning for the capital expenditure of hyperscale data centres. This review identifies that further research is required to establish and define specific variables relevant to hyperscale data centres.

It is recognised above that whilst there is published data on traditional construction costs in the UK along with published location indices. Neither of these provides sufficient information on establishing construction costs overseas. This is further compounded due to the specific design requirements of data centres and identifies a knowledge gap in the existing body of research.

This paper aims to identify consensus and an assessment for probability to identify themes that will inform the first round of a questionnaire of a complete Delphi study to identify the impact of location on the modelling and forecasting of capital expenditure for hyperscale data centres. The research question is to establish the impact of site location on the capital expenditure of hyperscale data centres and the variables that impact the modelling and forecasting as identified in the research gap above. This will assist in selecting the correct location to make informed decisions and reduce the financial risk to capital expenditure. The indicative research question raised is 'the impact of location on the modelling and forecasting capital expenditure for hyperscale data centres'.

To support the research topic, four subtopics have been identified:

- Does the location of a data centre change the risk of overspend? If so, how?
- What variables of capital expenditure are directly affected by location?
- What elements of capital expenditure cost are the most likely to overspend?
- What are the essential items to consider when choosing the right location for a hyperscale data centre?

The research will improve knowledge, linking innovation and infrastructure and the development of resilient infrastructure, which are all items targeted by the United Nations Sustainable Development Goals.

11.2 Methodology

11.2.1 Research Instrument

A modified Delphi technique was used as the research instrument. Delphi is a methodological technique where consensus can be formed from responses arrived at from a panel of experts based on questions that are uncertain (Pill, 1971). This is expanded on further in the seminal book by Linstone and Turoff (1975), where they state 'Delphi may be characterised as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem'.

Therefore, the research uses a Delphi technique as the most appropriate method as Keeney et al. (2006) stated a Delphi approach 'is an important method for achieving consensus on issues where none previously existed'. In addition, Hasson et al. (2000) note that 'insufficient information has led to an increased use of consensus method' such as Delphi.

Whilst there are several modified uses of Delphi (Thompson, 2009), this pilot study has used a qualitative approach as it is suggested as the preferred method. Jairath and Weinstein (1994) note that pilot studies should be carried out for all Delphi studies in advance of the leading research questionnaire being developed. As a result, this pilot study creates a robust framework for developing the first-round questionnaire in a complete Delphi study.

11.2.2 Data Collection

To inform this pilot study, a sample (n = 5) of data centre industry experts was selected to identify the topics to inform the complete Delphi study. According to Connelly (2008), experts suggest that a pilot study sample should be 10% of the total Delphi study sample size. It is anticipated that due to the corporate sensitivity of obtaining the data, the number of participants for the full Delphi will be less than 50. Therefore, the number of participants selected for this pilot study is within range. It is known that a fundamental component of Delphi research is the identification of a 'panel of experts' (Baker et al., 2006) as they form an established method for determining consensus (Beech, 2001). Using a Delphi method with industry experts makes it an ideal solution where such information is lacking (Graham et al., 2003). The identification and selection of an expert is determined as a person regarded or consulted as an authority on account of special skill, training or knowledge, a specialist (Stevenson, 2010).

The selection of participants for this pilot study is also experts as identified by their senior position in an organisation (Mead & Mosely, 2001), and as such, the participants selected are all at senior levels within an organisation. The experts were purposefully selected using a multistep iterative approach. Participants were

selected with design, development, engineering and commercial management expertise. The experts had similar expertise, experience and senior roles within the data centre industry.

11.2.3   Developing the Questionnaire

The survey was cross-sectional in its timeline, and the questions were open-ended with unlimited free text. The questions contained within the questionnaire are those identified as the research subtopics. One additional question was included within the questionnaire. This question was included to avoid any restrictions that the participants may experience whilst completing the four predetermined questions as above. In addition, it gave them the freedom to identify any other items they considered would impact capital expenditure outside the predetermined questions. This additional question was asked to the participants: what other variables may impact capital expenditure?

11.2.4   Details of the Data Collection

Participants were identified through known contacts based on their experience in cost consultation for data centre construction projects. The other criteria considered during the participant selection were their expertise and experience in hyperscale data centres both inside the UK and globally. The participants were not restricted to where and when the questionnaire was completed. The questionnaire was distributed to the participants utilising the online surveys distribution method via password-protected invites. The survey was open for 14 days to enable the participants' sufficient time to respond. All responses were received within this period.

11.2.5   Data Analysis

According to Kuhn (2021), having a paradigm as guidance for research is key to any research project as it is the 'set of beliefs and agreements shared between scientists about how problems should be understood and addressed'. Therefore, it is critical to establish ontological and epistemological approaches when forming the research methodology.

Williams (2016) states that ontology is 'the branch of philosophy concerned with the nature of things that exist'. The ontological approach taken within this pilot study and the subsequent interpretation and analysis of the participants observations identifies that the research has taken a realistic approach in that 'realism denies that there is any objective knowledge of the world' (Maxwell, 2012).

Epistemology is 'a way of understanding and explaining how we know what we know' (Crotty, 2020). It is also 'concerned with providing a philosophical grounding for deciding what kinds of knowledge are possible and how we can ensure that they are both adequate and legitimate' (Maynard, 1994). This pilot study has adopted a constructivist approach as the thematic analysis of the qualitative data is 'not discovered but constructed' (Crotty, 2020). This constructivist approach to the pilot study has also been supported by Gray (2021), who states that 'truth and meaning do not exist in some external world but are created by the subject's interactions with the world'.

Inductive thematic analysis was used to determine the themes and subsequent categories of the responses arising from the questionnaire as recognised by King (2004) that it assists in producing a concise analysis.

The thematic analysis was produced in NVivo software; the responses were coded to establish themes and analysed for themes and then grouped into categories. The coding and themes reflect the interpretation of the excerpts with the research question in mind.

The item frequency was calculated based on the aggregated items' responses. The two most dominant categories were explored in each question as most responses had a significant break between frequency counts.

11.2.6   Ethical Considerations

It is acknowledged that ethical considerations are fundamental to research for moral and institutional reasons (Farrimond, 2012). Therefore, before the questionnaire issue, written informed consent was obtained in advance from all participants and gatekeepers. Ethical approval for this study was received from Anglia Ruskin University Research School Research Ethics Panel (SREP) under the terms of Anglia Ruskin University's Research Ethics Policy.

11.3   Results and Discussion

11.3.1   Does the Location of a Data Centre Change the Risk of Overspend? If So, How?

The thematic analysis identified both land and the supply chain as the dominant categories in the participant's response when asked how the location of a data centre changes the risk of overspend. The analysis identified that land had a significant impact with overspending due to location choice. Results of the data analysis indicated that poor ground conditions, competing land use pressures and flood plain having significant impact.

In the thematic analysis, land equated to 50% of the themes. However, the perception of this impact varies, and whilst land appears to be the dominant category, there are many facets within this category. Following the data analysis, whilst geotechnical and environmental, land cost and infrastructure are identified, there is no consensus on impact ranking. The supply chain was also recognised as a dominant category in particular a lack of suitable labour, the cost of labour and materials being imported along with climatic conditions (hot/cold) impacting material selection.

These perceptions arise from the supply chain theme whilst showing the effect does not identify the cause. Therefore, there is a consensus that the supply chain does have an effect; there is no consensus identifying the cause, particularly in identifying the elements that could impact the schedule.

11.3.2 What Variables of Capital Expenditure Are Directly Affected by Location?

The thematic analysis identified both land and power and fibre as the two dominant categories in the participant's response when asked to identify what they considered the most significant cost impact to capital expenditure through location choice.

All of the participants of this study held a view that land had a significant impact, with examples from the questionnaire to support the theme categories being land parcel size, neighbouring high-risk activities (above/below ground) and identification of contamination risk profile.

In analysing these references from the questionnaires, land was mentioned 18 times. However, the perception of impact varies significantly from purchase price to contamination. This identifies that land appears to be the dominant category. However, identifying what constitutes a theme within the land category is unclear, and whilst geotechnical, climate and flooding are identified, there is no consensus on the dominant theme. All participants cited power and fibre as a significant theme, referenced 14 times. Results of the data identified specifically that the availability of power, fibre, existing network backbone structure and 'dark' fibre infrastructure have significant impacts. Typical statements being 'how much power is available' along with 'where does the fibre come in and how is it connected to the regional network'.

In analysing this data, most respondents agree that the availability of power and fibre has the most significant impact on this category. However, there is no consensus on whether power or fibre is the dominant theme.

These results largely agree with the existing literature and confirm that infrastructure requirements specific to data centres are a primary determinant of capital cost. Accurately estimating the impact of these factors on capital cost could improve the accuracy of development appraisal exercise and early decisions on location selection.

11.3.3   What Elements of Capital Expenditure Are Most Likely to Overspend?

The thematic analysis identified both design and the supply chain as the dominant categories in the participant's responses when asked to identify what elements of capital expenditure are most likely to overspend.

The design of the data centre was the dominant category at 35%, with response examples of design-related themes being permitting and planning, the method of colling the building and physical security of the site. Elements such as seismic design we also identified with comments such as 'a 20 MW two-story design for a Data Centre in the UK cannot be transposed to a site in Gebze region of Istanbul as no seismic cost/design have been accounted for within the substructure or superstructure'.

Whist this question asks what elements are most likely to overspend, the themes do not form a consensus within the category.

The supply chain was also recognised as a dominant category particularly labour and material costs. Other items resulting from the supply chain themes identified construction delays along with associated preliminary costs as having a significant impact.

Therefore, there is a consensus that the supply chain does have an effect; there is no consensus identifying the cause, particularly in identifying the elements that could impact the schedule.

11.3.4   What Are the Essential Items to Consider When Choosing the Right Location?

The thematic analysis identified land, power and fibre as the dominant categories in the participant's response when asked what essential items to consider when choosing the right location for a hyperscale data centre. The data identified that land had a significant impact on overspending due to location choice, property ownership laws, natural hazard risks, the previous use of the land, proximity to tenant/customer requirements and the accessibility of the sites.

In the thematic analysis, land equated to 43% of the categories. However, the perception of this impact varies, and whilst land appears to be the dominant category, there are many facets to the potential consensus of a category. This has identified that whilst geotechnical and environmental are dominant. Land cost and infrastructure are equally referenced, and subsequently, a consensus has not been formed. Likewise, all participants also cited power and fibre as a significant theme, referenced 11 times. Particular themes identified proximity to the data's point of use (latency), the availability of large utility power supply and redundancy power sources such as hydro, gas, oil, diesel and solar.

The category analysis identifies that most respondents agree that power and fibre have a significant impact. However, there was no consensus on whether power or fibre was the dominant theme when the participants were asked what the essential items are to consider when choosing the right location for a hyperscale data centre.

11.3.5   What Other Variables May Have an Impact on Capital Expenditure?

When asked to identify what variables of capital expenditure are directly affected by location, the most prominent category in the participant's response is design, ease of doing business and power and fibre. The design was the dominant category at 27%, with the key themes being the power and cooling technologies used, the substructure and superstructure design and noise attenuation. Whist this question asks what variables of capital expenditure are directly affected by location, there is no consensus within the category.

Power and fibre were identified as a significant theme, forming 20% of the categories, specifically distances from utilities (substation, fibre, water). Costs associated with incoming services and the need for an onsite substation were also frequent themes. Whilst the analysis identifies that power and fibre has a significant impact, there is not a consensus on whether it is either power or fibre that is the dominant theme.

11.3.6   The Relationship to the United Nations Sustainable Development Goals

By understanding the impact of power, fibre, land and design implications on the development of data centres provides knowledge and a greater understanding on the impact and importance of resilient infrastructure.

11.4   Conclusion

This pilot study was undertaken to seek expert opinion on the key themes impacting location on the modelling and forecasting of capital expenditure for hyperscale data centres, and the results of the data analysis are intended to provide rigour and to inform the questionnaire for the main Delphi study and increase the validity of the proposed questions.

This paper provides some indications of the current understanding of the variables that impact the modelling and forecasting for planning the capital expenditure of hyperscale data centres.

This review identified the gap in knowledge and the need for research in this area to move beyond prediction and forecasting of construction projects to that specifically for hyperscale data centres. Clibbens et al. (2012) state that pilot Delphi studies are rarely reported in the academic literature, making it difficult to establish best practice in this area. For this pilot study, industry expert knowledge was obtained through several expert participants ($n = 5$). The response rate was 100%.

Through having an open-ended questionnaire, experts were able to respond freely and without restriction.

Having completed the thematic analysis of the data arising out of the questionnaire, this pilot study has identified categories and themes that will address and inform the first round of a questionnaire of a complete Delphi study. Six overarching categories have been identified:

- Customer pricing
- Design
- Ease of doing business
- Land
- Power and fibre
- Supply chain

Additionally, the thematic analysis associated with each particular question has identified the themes within the categories as being:

- Access and infrastructure
- Geotechnical and environmental
- Land cost and availability
- Fibre
- Power
- MEPH (mechanical, electrical and public health)
- Substructure
- Superstructure
- General contractors

Therefore, from this pilot study, these categories and themes will inform round one of the main Delphi study. This pilot study has also confirmed the knowledge gap and supported the need that further investigatory work is required. Likewise, the results of this pilot study and the methodological stance taken have demonstrated that a Delphi approach would enable consensus in further research where none previously exist.

The results of this paper have identified that there are variables that impact the modelling and forecasting for planning the capital expenditure of hyperscale data centres that could not be identified through either existing published cost data or location variables thus connecting the results to the existing body of knowledge.

In addition, it is recommended that a further literature review is carried out to include the themes that have been identified through this pilot study, as this will subsequently inform and direct further research.

The results of this paper and the impending future research will provide knowledge to assist in achieving goal Nr 9 of the United Nations Sustainable Development Goals where it is identified that a key component of achievement should be that of a resilient infrastructure, particularly where there is a growing need for data led services.

11.4.1   Limitations

The respondent's location may have been biased as although the participants have experience in other geographic regions, they are all currently based in the UK. This may have led to biases in the answers, as some elements may differ at a country level. Such biases will be acknowledged as part of the main research study. However, it was expected that due to the extensive experience, the participants have the necessary expertise to respond to the questions in this pilot study, excluding their own biases.

Additionally, it is acknowledged that in undertaking this pilot study, the author has leveraged professional relationships, and this may have resulted in participant bias (Moore et al., 2010).

References

Aecom. (2021). *Spon's architects' and builders' price book 2022*. Taylor & Francis Group.

Avgerinou, M., Bertoldi, P., & Castellazzi, L. (2017). Trends in data centre energy consumption under the European code of conduct for data centre energy efficiency. *Energies, 10*(10), 1470.

Baker, J., Lovell, K., & Harris, N. (2006). How expert are the experts? An exploration of the concept of 'expert within Delphi panel techniques. *Nurse Researcher, 14*(1), 59.

Beech, B. (2001). The Delphi approach: Recent applications in health care. *Nurse Researcher (through 2013), 8*(4), 38.

Christensen, J. D., Therkelsen, J., Georgiev, I., & Sand, H. (2018). *Data centre opportunities in the Nordics: An analysis of the competitive advantages*. Nordic Council of Ministers.

Clibbens, N., Walters, S., & Baird, W. (2012). Delphi research: Issues raised by a pilot study. *Nurse Researcher, 19*(2), 37.

Connelly, L. M. (2008). Pilot studies. *Medsurg Nursing, 17*(6), 411.

Crotty, M. (2020). *The foundations of social research: Meaning and perspective in the research process*. Routledge.

European Union. (2021). Statistics | Eurostat. [Online]. Available at https://ec.europa.eu/eurostat/databrowser/view/prc_ppp_ind/default/table?lang=en. Accessed 19 Dec 2021.

Farrimond, H. (2012). *Doing ethical research*. Macmillan International Higher Education.

Graham, B., Regehr, G., & Wright, J. G. (2003). Delphi as a method to establish consensus for diagnostic criteria. *Journal of Clinical Epidemiology, 56*(12), 1150–1156.

Gray, D. E. (2021). *Doing research in the real world*. Sage.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*(4), 1008–1015.

Holt, J., & Vonderau, P. (2015). Where the internet lives. *Signal Traffic*, 71–93.

Jairath, N., & Weinstein, J. (1994). The Delphi methodology (Part one): A useful administrative approach. *Canadian Journal of Nursing Administration, 7*(3), 29–42.

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: Ten lessons from using the Delphi technique in nursing research. *Journal of Advanced Nursing, 53*(2), 205–212.

King, N. (2004). 21: Using templates in the thematic analysis of text. In *Essential guide to qualitative methods in organisational research* (pp. 256). Sage.

Kuhn, T. (2021). *The structure of scientific revolutions*. Princeton University Press.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method*. Addison-Wesley Reading.

Maxwell, J. A. (2012). *A realist approach for qualitative research*. Sage.

Maynard, M., 1994. Methods, practice and epistemology: The debate about feminism and research. Researching women's lives from a feminist perspective Routledge, 10(26), pp.10–26.

Mead, D., & Mosely, L. (2001). The use of the Delphi as a research approach. *Nurse Researcher, 8*(4), 4–23.

Moore, D. A., Tanlu, L., & Bazerman, M. H. (2010). Conflict of interest and the intrusion of bias. *Judgment and Decision Making, 5*(1), 37.

Organisation for Economic Co-operation and Development. (2021). *Prices and purchasing power parities (PPP)*. OECD. [Online]. Available at: https://www.oecd.org/sdd/prices-ppp/. Accessed 19 Dec 2021.

Parker, M. (2019). *An apple a day: Listening to data centre site selection through a ionospheric investigation*.

Pill, J. (1971). The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Economic Planning Sciences, 5*(1), 57–71.

RICS. (2012). *New rules of measurement-NRM 1: Order of cost estimating and cost planning for capital building works*. Royal Institution of Chartered Surveyors.

Royal Institution of Chartered Surveyors. (2021). *BCIS Indices – basic parameters*. [Online]. Available at: https://service-bcis-co-uk.proxy-lib.anglia.ac.uk/BCISOnline/Indices. Accessed 19 Dec 2021.

Stevenson, A. (2010). *Oxford dictionary of English*. Oxford University Press.

Thompson, M. (2009). Considering the implication of variations within Delphi research. *Family Practice, 26*(5), 420–424.

Vonderau, A. (2017). Technologies of imagination: Locating the cloud in Sweden's North. *Imaginations: Journal of Cross-Cultural Image Studies, 8*(2), 8–21.

Wahlroos, M., Pärssinen, M., Rinne, S., Syri, S., & Manner, J. (2018). Future views on waste heat utilisation–Case of data centers in Northern Europe. *Renewable and Sustainable Energy Reviews, 82*, 1749–1764.

Whitehead, B., Andrews, D., & Shah, A. (2015). The life cycle assessment of a UK data centre. *The International Journal of Life Cycle Assessment, 20*(3), 332–349.

Williams, M. (2016). *Key concepts in the philosophy of social research*. Sage.

World Bank. (2021). *PPP conversion factor, GDP* (LCU per international $) | Data. [Online]. Available at: https://data.worldbank.org/indicator/PA.NUS.PPP?view=map. Accessed 19 Dec 2021.