BAYES THEORY AS A METHODOLOGICAL APPROACH TO ASSESS THE IMPACT OF LOCATION VARIABLES OF HYPERSCALE DATA CENTRES: TESTING A CONCEPT

David King, Dr Nadeeshani Wanigarathna, Prof. Keith Jones & Dr Joseph Ofori-Kuragu
Anglia Ruskin University, Bishops Hall Lane, Chelmsford CM1 1SQ, United Kingdom

ABSTRACT: *The theme of 'The Impact of Engineering Practices on a Sustainable Built Environment' emphasises the importance of considering various dimensions of resilient infrastructure. Selecting the location for a Hyperscale Data Centre is a crucial process that involves assessing the impact of various location variables. To determine the viability of a location, it is essential to identify the potential risks associated with each variable. This paper presents a proprietary methodological approach that includes a Delphi study to identify risks, a Likert scoring system to assess prior probabilities, and a Bayesian theory-based decision tree to assess the impact through risk prediction. The paper's contributions are significant, and the proposed methodology makes it possible to predict the risk level of each location variable by identifying the appropriate contingency percentage. The study's findings indicate that the paper's proposed approach is an effective way to mitigate the risks associated with selecting a location for a Hyperscale Data Centre. Embracing this knowledge allows us to align research and practise with the conference's call to studying the resilience of buildings and infrastructure to natural disasters and climate change, and developing strategies for adaptation and mitigation, ensuring that these practises become integral to shaping the future of Data Centres.*

KEYWORDS: *Bayes Theorem, Delphi, Data Centre, Location Variables*

1 BACKGROUND

Investments in Data Centres in the Nordic region have been on the rise, with significant contributions from cloud and hyperscale investors such as Facebook, Google, AWS and Apple, due to advanced technological progress and favourable cold climate conditions, significantly reduce the cooling energy demands of the facilities (Christensen et al.,2018; Avgerinou et al., 2017). However, the location of Data Centres outside of the UK presents a significant challenge for cost consultants during the capital cost estimation and modelling stages, which can impact investment decisions. At the feasibility stage, cost planning involves determining the possible cost of a building early in the design stage in relation to the employer's fundamental requirements before preparing a complete set of working drawings or quantities bills (RICS, 2011) Historical cost data is often used as base cases for cost consulting professionals, who adjust their costs to suit the circumstances of new projects. Although specific characteristics such as shape, inflation, and specifications are relatively easy to adjust based on case-based reasoning, predicting the impact of location is challenging for construction professionals, who rely on location cost indices for this purpose. Various location cost indices, such as Spon's Architects and Builders Price Book (AECOM, 2017) and the Building Cost Information Service (RICS, 2018) are available for cost consultants. However, such indices are less relevant for Data Centres as there are often no precedents set to use as a baseline for cost comparisons, and there are many variables ranging from macroeconomic, construction methodology, geographical, and geological categories. For example, 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), International location cost indices, such as those provided by Eurostat (EC, 2019), World Bank (2022) and the OECD (2022) are broad and mainly model variations at the country level, making them less effective during cost planning for individual projects specific to a particular region. Therefore, construction professionals must consider multiple factors and rely on a combination of indices and expert judgment to provide accurate cost estimates for Data Centres.

2 RESEARCH AIM

Whilst a wide range of variables impacts construction project costs and cost modelling, there is no evidence to suggest whether and how these variables would affect site location in cost planning for the capital expenditure of Hyperscale Data Centres. Although there is published data on traditional construction costs and location indices in the UK, they do not provide enough information to assess the impact of location variables, especially considering the specific design requirements of Data Centres (King et al., 2023). This highlights a significant knowledge gap in the existing body of research. This paper aims to validate a methodological concept using Delphi

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
David King, Nadeeshani Wanigarathna, Keith Jones, Joseph Ofori-Kuragu, *Bayes Theory as a Methodological Approach to Assess the Impact of Location Variables of Hyperscale Data Centres: Testing a Concept,* pp. 398-406, DOI 10.36253/979-12-215-0289-3.39

and Bayesian theory to assess the probability and impact of location variables. This approach aims to aid in selecting the appropriate methodological approaches for the research topic of ”the impact of location variables on the modelling and forecasting of Hyperscale Data Centres". By utilising this method, the study seeks to identify the potential risks and impacts of various location variables, contributing to a more comprehensive understanding of the relationship between site location and capital expenditure in the Hyperscale Data Centre industry.

3 METHODOLOGY

3.1 Risk

Risk refers to situations that involve uncertainties that may occur, risk mitigation refers to actions taken to optimise the impact of risk. By selecting a comprehensive risk management strategy that considers all types of risk, one can ensure the implementation of a planned Data Centre investment within the specified time and budget. Various organisations have developed several approaches to risk. Notable among these are the Project Management Institute (PMI., 2001) and PRINCE2 (Bentley., 2012). This paper aims to introduce a concept that can quantify the impact of risk through a Delphi and Bayesian approach. A risk is defined as the probability of an event occurring and the subsequent consequence, as expressed in Equation (1). Here, R represents a risk, P is the probability of the event occurring, and C is the impact or consequence of the event.

$$R = (P, C). \tag{1}$$

Various methodologies exist for identifying risk including identification, assessment, response, and monitoring. Risk identification is identifying potential risks that may impact the project. Risk assessment involves analysing and evaluating the likelihood of occurrence, impact, and consequences of the identified risks. The risk response involves developing a plan to manage or mitigate the identified risks. Lastly, risk monitoring and control. Quantifying the impact of risk, especially with location variables, can provide invaluable information to decision makers and stakeholders and can be used to make informed decisions, develop contingency plans, and allocate resources appropriately. Therefore, developing a method to assess the impact of location variables on project risk can significantly improve the success of a project. Risk decisions involve assessing the factors that contribute to the emergence of risk and the likelihood and potential impact of the event.

3.2 Delphi Study

A pilot Delphi study (King et al., 2023) has been conducted to obtain expert opinions on the key themes that affect the location variables of Hyperscale Data Centres and their impact on the modelling and forecasting of capital expenditure. The analysis of the pilot study data has provided rigour and validity to the questionnaire for the main forthcoming Delphi study. This has allowed for identifying and assessing potential risks associated with the location variables of Hyperscale Data Centres. The pilot study results indicate the current understanding of the variables that impact the modelling and forecasting of capital expenditure for Hyperscale Data Centres. These variables have been identified as potential risks and are an essential consideration in the risk management strategy for the planning and implementing Hyperscale Data Centres. Previous research found that pilot Delphi studies are rarely reported in academic literature, making it difficult to establish best practices (Clibbens., 2012). For this pilot study, industry expert knowledge was obtained through several expert participants (n=5). The response rate was 100%. Through an open-ended questionnaire, experts could respond freely and without restriction. Having completed the thematic analysis of the data arising from the questionnaire, the pilot study identified categories and themes that are considered risk items; the following items were among those rated by the participants as having an impact on capital expenditure when locating a data centre:

- Requirement for cooling towers due to sub-zero climate
- Requirement to import generators due to in-country shortages.
- Acoustic screens to generators due to proximity of residential neighbours
- In-country technical labour shortages require backfilling with imported, experienced technical labour.

The themes arising from the Delphi study provide the data that will be used to provide the data that will be used for the assessment of the impact of location variables within a Bayesian framework.

3.3 Bayes Theory

Bayesian theory is based on the probability theory given by Thomas Bayes in 1763 (Bayes., 1763). Bayes's theory relates the conditional probabilities of random variables to each other. It provides a framework that allows for the integration of a prior belief about the distribution of a quantity of interest (the prior distribution) and the observed data (through the likelihood term). as shown in Equation (2).

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)} \tag{2}$$

To clarify, in this instance:

- P(B) denotes the prior belief (for example, the probability of occurrence of the variable, such as the probability of encountering ground conditions)
- P(B|A) denotes the level of impact should that variable occur.
- P(B) denotes the new site-specific evidence (for example, when new information arises, i.e., a higher probability of occurrence of encountering ground conditions)

Bayes theory can be applied to numerous components by using the product rule (Pearl., 2022) and, therefore, Bayes theory is applied for calculating the probability of occurrence of a phenomenon or hypothesis using multiple factors or variables. It is also considered a powerful method for hypothesis testing (Wetzels et al., 2012) making assumptions and having wide-ranging decision-making applications related to artificial intelligence, machine learning, and bio-statistics approaches. Prediction theory is a sub-field of statistics and machine learning that involves the development of mathematical models and algorithms for predicting future outcomes or events (Sarker., 2021). It uses data from past observations to create models that can be used to forecast future outcomes. Prediction theory employs various data analysis techniques like regression, clustering, and classification. It also involves identifying essential variables and patterns within the data, calculating the probability of specific outcomes, and selecting desirable outcomes based on the model generated. Although prediction theory and Bayes theory are related, they differ in terms of their fundamental principles. Bayes's theory concerns conditional probability and allows for the revision of probabilities based on new information or evidence (Ajzen et al., 1975). On the other hand, prediction theory is focused on building models and computing algorithms to predict outcomes from complex data sets. While prediction theory may incorporate probabilities, it does not involve the revision of probabilities like Bayes's theory. Using Bayesian theory and correlation analysis is a common practice for predicting future outcomes and events. In addition, integrating prediction theory with the Delphi method is a recognised technique used to forecast future outcomes based on expert opinions (Turoff et al., 2002). The Delphi method involves obtaining consensus opinions from subject matter experts through a series of planned interviews or surveys, which can then be used to forecast future outcomes. Furthermore, the Delphi method can be combined with Bayesian theory to revise established opinions based on the likelihood of different outcomes. This study highlights that expert opinions gathered through a structured sampling technique such as the Delphi method can be utilised to estimate probable outcomes, which can then be inputted into the Bayesian formula to provide current outcomes based on updated information gathered through qualitative risk assessments. The combination of the Delphi method and Bayesian theory enhances the accuracy and decisiveness of the mathematical model compared to using prediction theory alone. Previous research supported this approach, including Bijak (2011), who identified Bayesian theory as a natural methodology for combining expertise and data with expert judgments. Additionally, Bernardo (2003) suggests that Bayes's formula allows for expert opinions to be incorporated into projections in the form of prior distributions. However, a limitation of Bayesian forecasts is that they may contain subjective elements due to their dependence on expert opinions and history obtained from the data series (Abel et al., 2013). In conclusion, the combination of Bayesian theory and the Delphi method can provide a robust methodology to model and forecast the impact of location variables on Hyperscale data centres.

4 DATA COLLECTION

4.1 Likert

Psychologist Rensis Likert invented the Likert scale (Likert., 1932). It is a rating scale used to measure attitudes, opinions, or perceptions. The scale can have anywhere from 5 to 11 points, with the most common being a 5-point scale. It is widely used in social sciences, especially in survey research, as it allows researchers to gather information about people's attitudes, opinions, or perceptions systematically and standardised. The scale is also commonly used in market research, customer satisfaction, and employee engagement surveys. The Likert scale has several advantages, including ease of use, simplicity, and flexibility. It is easily understood by respondents, which can improve the accuracy and reliability of the data collected. However, it is important to remember that the Likert scale also has limitations, such as possible response bias, limited ability to capture complex attitudes, and the potential for data to be misinterpreted if it is not used appropriately. It is important to carefully consider the wording and format of the questions in the Likert scale to minimise these limitations and ensure accurate data collection. Additionally, it is essential to use appropriate statistical techniques when analysing the data obtained through the Likert scale to avoid misinterpretation of the results.

4.2 Probability

To establish the likelihood of events, a Likert ranking has been proposed with two extremes at either end of the scale. A score of 1 denotes an event highly unlikely to occur, whereas a score of 5 represents a highly likely scenario, as shown in Table 1. For instance, when assessing power availability, one might score it as one because the likelihood of that event occurring is low. On the other hand, if there is a substation on-site and the site is situated in the centre of a seismic zone, a score of 5 may be assigned since the probability of a seismic event causing damage is very high. These scoring descriptions outline the scoring criteria and help prevent ambiguity when experts score as part of the Delphi study.

Table 1: Likert ranking for probability.

Likert scale	Probability
1	Very unlikely
2	Unlikely
3	Neutral
4	Likely
5	Very likely

The variables identified and presented in Table 2 are derived from a previous Delphi study by King et al (2023).

Table 2: Likert scoring results for the probability of the event occurring.

Variable	Very unlikely	Unlikely	Neutral	Likely	Very Likely
Cooling towers	4	1	4	41	16
Imported generators	4	4	32	23	3
Acoustic screens	4	1	4	39	18
Technical labour shortage	4	28	27	5	2

These Likert scoring values are intended to illustrate the proof of concept. They are based on the authors' professional judgment regarding the probability of each item occurring in the real world. However, it is essential to note that these scores are hypothetical for illustration only to demonstrate the proof of concept. They will be subject to revision based on new available information, resulting in updated posterior probabilities that may differ significantly from the initial estimates.

5 RESULTS AND DISCUSSION

5.1 Establishing Nodes

The scoring rankings for probability are derived from the Likert scoring results in Table 2 and weighted to generate the probability distribution required for the Bayesian analysis. A weighing method has been used to assess these conditional probabilities, as shown in Equation (3).

$$\frac{Occurance}{Total\ respondants} \qquad (3)$$

The variables and the conditional probability of these events occurring are shown in Table 3. The results subsequently creating the nodes for the Bayesian network.

Table 3: Conditional probability of the event occurring.

Item	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Cooling towers	6%	2%	6%	62%	24%
Imported generators	6%	6%	48%	35%	5%
Acoustic screens	6%	2%	6%	59%	27%
Technical labour shortage	6%	42%	41%	8%	3%

Therefore, the node describing the event of Cooling Towers together with possible scenarios of this likelihood together with possible assessment factors is as Figure 1

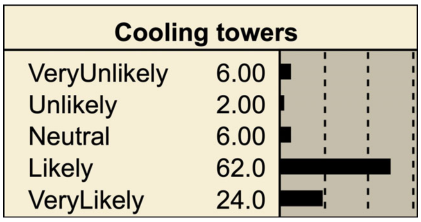

Figure 1: Conditional probability node of Cooling Towers being required.

5.2 Assigning event probabilities

A process of identifying possible events for each of the variables was established. The basis of the Bayesian network is related to determining the relationship of each individual node in the network. For this proof of concept, the relationship of individual node was based on the authors' own experience and assessed using a low, medium, and high ranking. For example, the impact of cooling towers is identified in Figure 2.

Cooling towers	Low	Medium	High
VeryUnlikely	100	0	0
Unlikely	50	50	0
Neutral	0	50	50
Likely	0	0	100
VeryLikely	33.3	33.4	33.3

Figure 2: Conditional probability node of Cooling Towers impact

These relationships have been used to identify scenarios that could occur because of events in the process of assessing the impact of location variables through four ranges for contingency between 0% and 20%, as shown in Figure 3. These contingency values have been presented based on the author's experience as proof of concept. Further research will be required to refine these contingency values.

Cooling towers	Technical labour sh...	0 to 5 percent	5 to 10 percent	10 to 15 percent	15 to 20 percent
Low	Low	100	0	0	0
Low	Medium	50	50	0	0
Low	High	0	0	100	0
Medium	Low	50	50	0	0
Medium	Medium	0	50	50	0
Medium	High	0	0	50	50
High	Low	0	50	50	0
High	Medium	0	0	50	50
High	High	0	0	0	100

Figure 3: Conditional probability node of Contingency for Mechanical

5.3 Performing calculations

The Bayesian network conditional probabilities were calculated using Netica software (Ni et al., 2011). This resulted in a functional and working network being developed to assess the impact of location variables. After calculations, the results of the conditional probabilities were established, as shown in Figure 4.

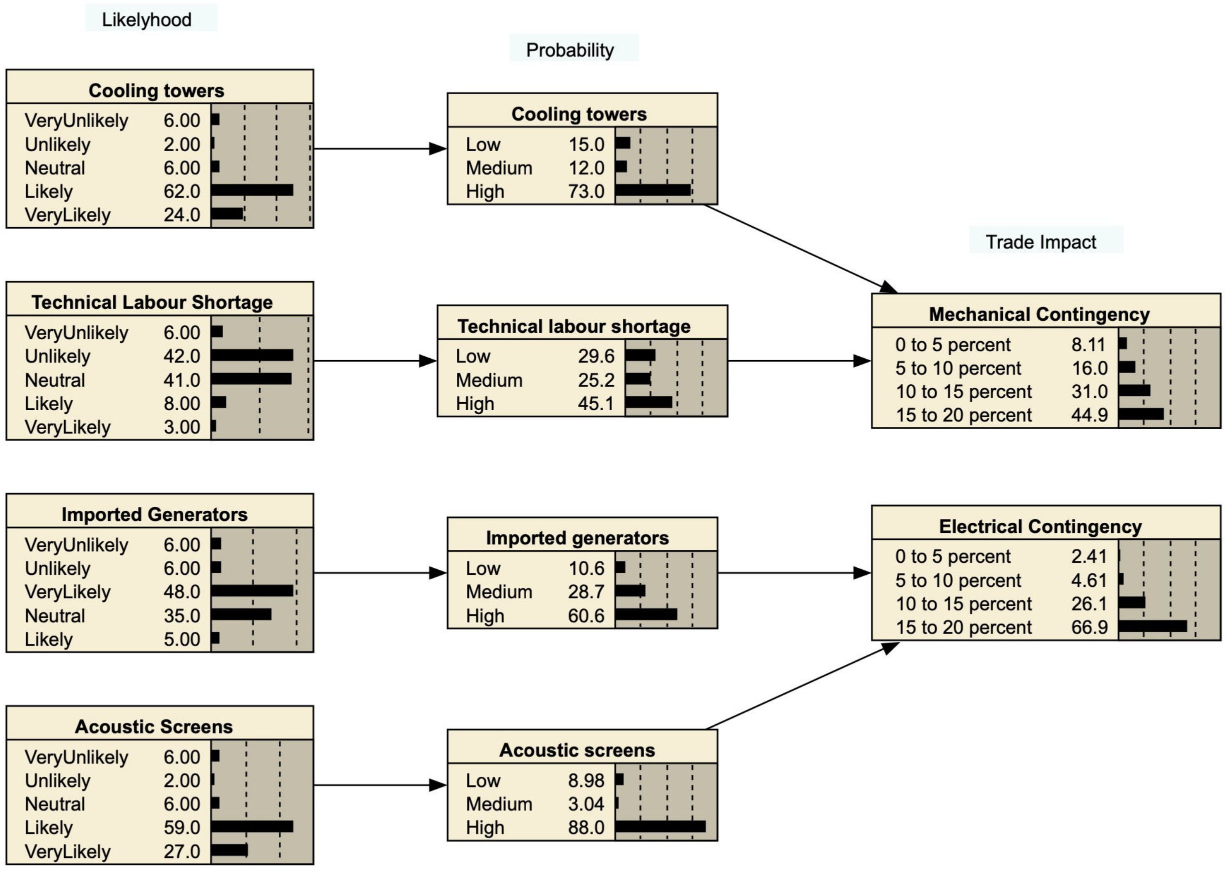

Figure 4: Bayesian network identifying trade Contingencies based on conditional probabilities.

5.4 Event scenario analysis

An example of the updated impact of Cooling towers is shown in Figure 5. This event has been modelled on the node 'Cooling towers'. A 100% likelihood of this event occurring has been assumed as 'Unlikely' in this hypothetical scenario.

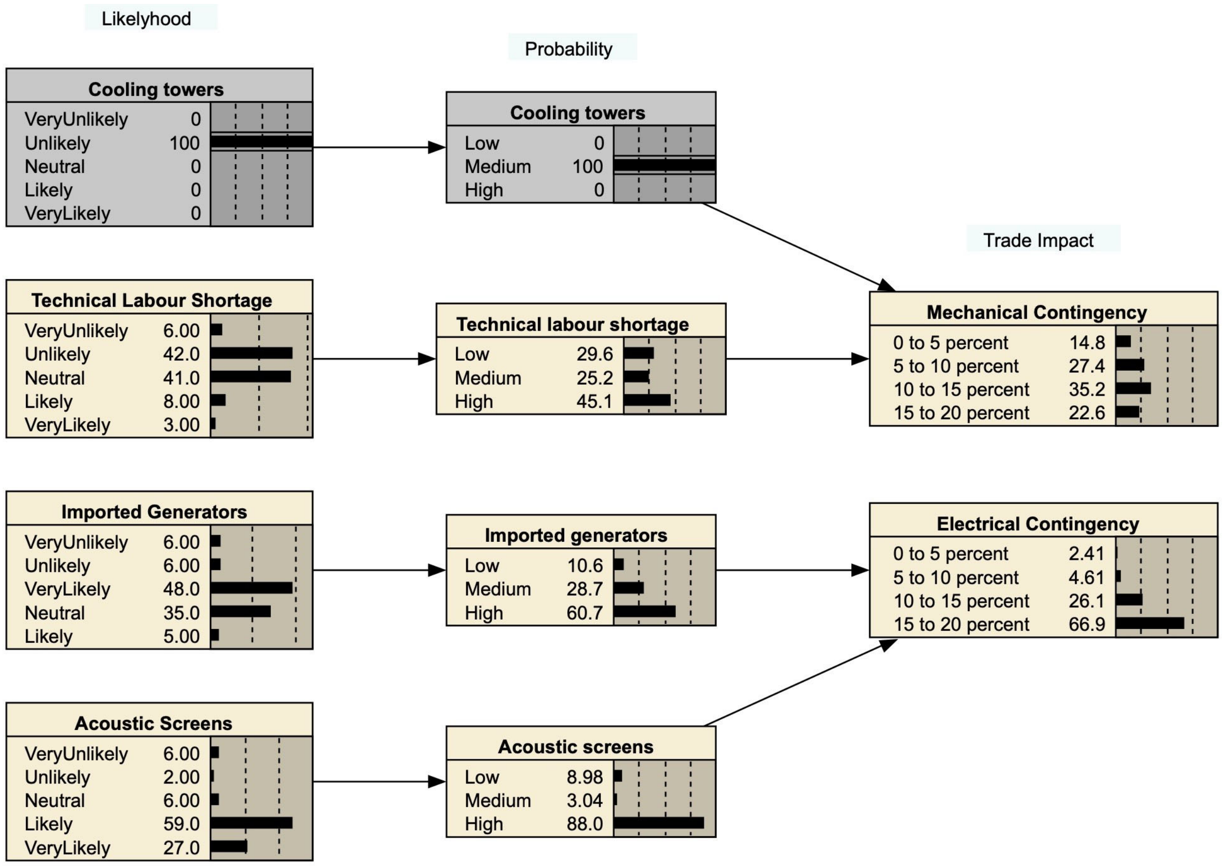

Figure 5: Bayesian network updated with new probabilities impacting Mechanical contingency.

In this scenario, we have also selected the node for a probability of an increase in cost to 'Medium.' Using this Bayesian network, this updated information has impacted the node for Mechanical Contingency, changing from 15%-20%, as identified in Figure 4, to 10%-15%, as shown in Figure 5. Therefore, in this example, the impact of location variables has, using the Bayesian theory, identified an improved risk and reduced contingency for the Mechanical Works.

6 CONCLUSION

Using a combination of the Delphi study, Likert scale, risk, and Bayesian theory to evaluate the impact of site location on capital expenditure for Hyperscale Data Centres has been demonstrated to be a feasible approach. The study findings indicate that it is possible to identify the likelihood of specific location variables impacting capital expenditure by conducting a Delphi study to obtain expert opinions and utilising a Likert scale to acquire subjective information about the probability and perceived risk of occurrence. These probabilities can be integrated into Bayesian analysis as prior knowledge, and as new information becomes available, they can be updated to calculate the posterior probability. The resulting percentage impact can then be applied to assess individual or multiple items and incorporated into the total capital expenditure, providing a method for determining the percentage impact, cost increase, or contingency. The findings of this study have significant implications for evaluating the impact of location variables for Hyperscale Data Centres, where variables can be identified and quantified as a percentage variance to capital expenditure. By utilising a Delphi study, the method can gather expert opinions, increasing the reliability and validity of the data obtained.

Furthermore, using a Likert scale allows for quantifying subjective information, which can be challenging to measure using other methods. Finally, by incorporating the probability and risk of occurrence, the Bayesian analysis provides a more accurate assessment of the impact of location variables on capital expenditure. The methodology described in this study can be applied to various industries, providing a comprehensive framework for determining the impact of various factors on capital expenditure and informing decision-making processes.

REFERENCES

J. D. Christensen, J. Therkelsen, I. Georgiev, H. Sand, Data centre opportunities in the Nordics: An analysis of the competitive advantages, 1st Edition, Nordic Council of Ministers, Stockholm, 2018. https://doi.org/10.6027/TN2018-553

M. Avgerinou, P. Bertoldi, L. Castellazzi, Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency, Energies 10 (10) (2017). https://doi:10.3390/en10101470

RICS, Cost analysis and benchmarking, Tech. rep., Royal Institution of Chartered Surveyors (2011). https://www.isurv.com/downloads/file/3096/archived_cost_analysis_and_benchmarking_%E2%80%93_uk_1st_edition_august_2011%E2%80%93june_2022?restricted=true

AECOM, Spon's Architects' and Builders' Price Book 2018, one hundred Edition, CRC, Abingdon, 2017. https://doi.org/10.1201/b22288

RICS, BCIS Online (2018). https://bcis.co.uk/?gclid=EAIaIQobChMImqj1Lab_gIVB_tCh3e7g6KEAAYAiAAEgLWQPD_BwE

A. Vonderau, Technologies of imagination: Locating the Cloud in Sweden's North, Imaginations: Journal of Cross-Cultural Image Studies 8 (2) (2017) 8–21. https://doi.org/10.17742/image.ld.8.2.2

EC, Purchasing power parities (PPPs), price level indices and real expenditures - Eurostat, issue: November 2 Volume: 2019 (2019). https://ec.europa.eu/eurostat/web/purchasing-power-parities/publications

World Bank, The World Bank Consumer Price Index (2022). URL https://data.worldbank.org/indicator/FP.CPI.TOTL?locations=FR-DE-IT-GB-US

OECD, OECD Purchasing Power Parities (2022). URL https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm

D. King, N. Wanigarathna, K. Jones, J. Ofori-Kuragu, A Delphi Pilot Study to Assess the Impact of Location Factors for Hyperscale Data Centres, in G. Lindahl, S. C. Gottlieb (Eds.), SDGs in Construction Economics and Organization, Springer International Publishing, Cham, 2023, pp. 153–164. https://doi.org/10.1007/978-3-031-25498-7_11

PMI, Project management body of knowledge (pmbok® guide), Vol. 11, 2001, pp. 7–8. https://doi.org/10.1201/9781439882856-8

C. Bentley, Prince2: A practical handbook, Routledge, 2012. https://doi.org/10.4324/9780080497792

N. Clibbens, S. Walters, W. Baird, Delphi research: issues raised by a pilot study, Nurse researcher 19 (2) (2012). https://doi.org/10.7748/nr2012.01.19.2.37.c8907

T. Bayes, An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S, Philosophical transactions of the Royal Society of London (53) (1763) 370–418. https://doi.org/10.1098/rstl.1763.0053

J. Pearl, Reverend Bayes on inference engines: A distributed hierarchical approach, in Probabilistic and Causal Inference: The Works of Judea Pearl, 2022, pp. 129–138. https://doi.org/10.1145/3501714.3501727

R. Wetzels, E.-J. Wagenmakers, A default Bayesian hypothesis test for correlations and partial correlations, Psychonomic Bulletin & review 19 (6) (2012) 1057–1064, publisher: Springer. https://doi.org/10.3758/s13423-012-0295-x

I. H. Sarker, Machine learning: Algorithms, real-world applications and research directions, SN Computer Science 2 (3) (2021) 1–21, publisher: Springer. https://doi.org/10.1007/s42979-021-00592-x

I. Ajzen, M. Fishbein, A Bayesian analysis of attribution processes., Psychological Bulletin 82 (2) (1975) 261, publisher: American Psychological Association. https://doi.org/10.1037/h0076477

M. Turoff, H. A. Linstone, The Delphi method-techniques and applications (2002). https://doi.org/10.2307/1268751

J. Bijak, Forecasting migration: selected models and methods, in Forecasting International Migration in Europe: A Bayesian View, Springer, 2011, pp. 53–87. https://doi.org/10.1007/978-90-481-8897-0_4

J. M. Bernardo, Bayesian Statistics. Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics (2003). www.uv.es/~bernardo/BayesStat2.pdf

G. Abel, J. Bijak, A. Findlay, D. McCollum, A. Wisniewski, Forecasting environmental migration to the United Kingdom: an exploration using Bayesian models, Population and Environment 35 (2) (2013) 183–203. https://doi.org/10.1007/s11111-013-0186-8

R. Likert, A technique for the measurement of attitudes., Archives of psychology (1932). https://legacy.voteview.com/pdf/Likert_1932.pdf

Z. Ni, L. D. Phillips, G. B. Hanna, Exploring Bayesian belief networks using netica®, in Evidence Synthesis in Healthcare: A Practical Handbook for Clinicians, Springer, 2011, pp. 293–318. https://doi.org/10.1007/978-0-85729-206-3_12