Blockchain: Research and Applications 5 (2024) 100169

Contents lists available at ScienceDirect

Blockchain: Research and Applications

journal homepage: www.journals.elsevier.com/blockchain-research-and-applications

Review Article

Promoting rigor in blockchain energy and environmental footprint research: A systematic literature review

Ashish Rajendra Sai [a,b,*], Harald Vranken [a,c]

[a] Department of Computer Science, Open Universiteit, Heerlen, 6419 AT, the Netherlands
[b] Department of Advanced Computing Sciences, Maastricht University, Maastricht, 6211 LK, the Netherlands
[c] Institute for Computing & Information Sciences, Radboud University, Nijmegen, 6525 XZ, the Netherlands

ARTICLE INFO

Keywords:
Blockchain
Energy consumption
Environmental footprint
Scientific rigor
Reliability

ABSTRACT

There is a growing interest in understanding the energy and environmental footprint of digital currencies, specifically in cryptocurrencies such as Bitcoin and Ethereum. These cryptocurrencies are operated by a geographically distributed network of computing nodes, making it hard to estimate their energy consumption accurately. Existing studies, both in academia and industry, attempt to model cryptocurrency energy consumption often based on a number of assumptions, for instance, about the hardware in use or the geographic distribution of the computing nodes. A number of these studies have already been widely criticized for their design choices and subsequent over- or under-estimation of energy use.

In this study, we evaluate the reliability of prior models and estimates by leveraging existing scientific literature from fields cognizant of blockchain, such as social energy sciences and information systems. We first design a quality assessment framework based on existing research, and we then conduct a systematic literature review examining scientific and non-academic literature demonstrating common issues and potential avenues of addressing these issues.

Our goal with this article is to advance the field by promoting scientific rigor in studies focusing on blockchain energy footprint. To that end, we provide a novel set of codes of conduct for the five most widely used research methodologies: Quantitative energy modeling, literature reviews, data analysis and statistics, case studies, and experiments. We envision that this code of conduct would assist in standardizing the design and assessment of studies focusing on blockchain-based systems' energy and environmental footprint.

1. Introduction

> All models are wrong, but some are useful.
>
> George E.P. Box

This famous quote from the British statistician George E.P. Box highlights both the merits and limits of statistical modeling. All models designed to represent real-world systems are inherently limited due to their reductive nature; however, they may serve a useful purpose if designed and tested well and if their scope and assumptions are clearly indicated. This is particularly true in the case of energy consumption models designed for sociotechnical systems [1]. Designing these models is a non-trivial task that requires a number of social, economic, and technical assumptions. The intent behind many of these models is often to provide useful insight in the form of an estimate of the energy requirement or environmental footprint, rather than an absolute measurement of energy consumed or carbon emission by these systems.

Some of the early models from the 2000s predicted the electricity requirements of the internet and computers to varying degrees of accuracy. Some early reports suggested that all computers could consume up to 50% of US electricity in 2010 [2]. These claims have since been debunked through further research and empirical data [3]. This pattern of inaccurate or misleading predictions and measurements regarding the energy consumption of a fast-growing information technology is considered problematic, as it may influence policymakers [4] and may feed misinformation to the general public when picked up by popular media.

Decentralized digital assets are one such class of fast-growing information technology that has garnered significant interest from both academia and industry due to its unique energy profile [5]. Bitcoin

* Corresponding author. Department of Computer Science, Open Universiteit, Heerlen, 6419 AT, the Netherlands.
E-mail address: ashish.sai@ou.nl (A.R. Sai).

https://doi.org/10.1016/j.bcra.2023.100169
Received 14 October 2022; Received in revised form 18 October 2023; Accepted 24 October 2023
Available online 10 November 2023
2096-7209/© 2023 THE AUTHORS. Published by Elsevier B.V. on behalf of Zhejiang University Press. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

and other similar decentralized digital assets often employ an energy-intensive consensus mechanism[1] known as proof-of-work (PoW).

By its design, the participants in the PoW-based digital assets are incentivized to spend considerable effort, typically by executing compute-intensive or memory-intensive tasks, on a dynamically calibrated problem.[2] The first participant to find and broadcast the solution to this problem within a dedicated time frame is rewarded for their participation in the form of newly minted cryptocurrencies. For example, on June 1, 2022, the reward to find the solution or mine one Bitcoin block was approximately 200k USD [7]. This substantial incentive initiates a competitive race to mine the next block. Participants invest additional computational cycles in solving the problem to increase their chances of receiving this reward. Each attempt to find a solution to the problem incurs an energy cost in the form of electricity spent to power the device that solves the problem.

Similar to the early days of the internet and computers, we have seen numerous attempts at measuring the electricity consumption of decentralized digital assets such as Bitcoin [8]. It has been a frequent sight to see news headlines indicating the colossal energy and environmental footprint of Bitcoin. Many of the non-academic literature and (highly rated) academic sources used in these news headlines have been criticized for inaccuracy or misleading interpretations [9–11].

While we acknowledge that it is worthwhile to explore the energy and environmental footprint of cryptocurrencies such as Bitcoin, we stress that this should be done with utmost care to avoid inaccurate analysis and unjustified assumptions that may lead to sensational news headlines. For instance, the article published by Mora et al. [12] suggests that Bitcoin alone could push global warming above 2 °C as soon as 2033. This article has been widely criticized for provably inaccurate underlying assumptions, such as participants using unprofitable hardware [10,11,13,14]. The article also assumes that total electricity consumption is proportional to the number of transactions, a misconception that several publications explicitly contradict [5,14,15].

As it is inherent to energy modeling, each of these models relies on several assumptions to provide an estimate; thus, their accuracy is subject to the validity of their underlying assumptions. The scientific expectation is that these assumptions are not only mentioned explicitly but also backed by verifiable, preferably empirical evidence or justification [16].

Unfortunately, as seen in the case of Mora et al. [12], this is not always the case. Further research into the reliability of these studied by Lei et al. [8]. And Koomey [11] suggested that these issues are not isolated to one particular study. However, as they both only focused on a small set of models, it is difficult to generalize the results to the whole field.

Our study attempts to overcome this limitation by conducting a systematic literature review of both scientific and non-academic literature focusing on the energy and environmental footprint of cryptocurrencies, specifically focusing on PoW. We assess the quality of the shortlisted literature against the guidelines put forth by Lei et al. [8] and Sovacool et al. [16].

This article attempts to expand upon the findings of Lei et al. [8] more formally and rigorously. Unlike Lei et al. [8], we take a holistic approach to rigor assessment and base our quality measures on existing studies from blockchain energy science as well as information systems, social energy sciences, and computer science. This allows us to conduct a more rigorous and comprehensive analysis of the field. We also systematically conduct our literature review using the guidelines put forth by Kitchenham and Charters [17], allowing our study to be more transparent and verifiable.

We iteratively refine our quality assessment framework to account for domain-specific variations[3]. Thus, in this work, we present the first in-depth analysis of the scientific rigor of blockchain energy and environmental models to assess the following question:

Are the existing energy and environmental footprint models and resulting estimates for blockchain-based systems trustworthy?

Analyzing PoW's energy consumption presents unique challenges, mainly due to the opaque nature of Bitcoin mining. Comprehensive data are hard to come by, leading researchers to use indirect sources such as IP addresses or initial public offering (IPO) filings. These methods, although grounded in reality, offer only rough estimates. While our article critiques existing studies, it is essential to interpret these findings with these limitations in mind.

It is important to note that the purpose of our article is not to discuss whether or to what extent specific studies are flawed but to provide tools to transparently discuss the rigor of these studies while allowing for improvements in the design and prediction of these models. We support and encourage the work done in blockchain energy sciences over the last few years and intend to expand on it through this review.

To study the reliability of energy models, we first code and analyze relevant scientific and non-academic literature. We review the literature published from 2008, i.e., post the introduction of Bitcoin's white paper [18]. This is done by following the guidelines proposed by Kitchenham and Charters [17]. As suggested by Kitchenham and Charters [17], our review is broken down into five steps: search, selection, quality assessment, data extraction, and analysis. This review resulted in an article pool of 128 studies. These articles are then assessed for their scientific rigor by using the quality assessment framework based on the guidelines of Lei et al. [8] and Sovacool et al. [16].

Following the assessment of the scientific rigor, we consolidate our findings in the form of commonly occurring issues in different types of studies. We also document potential avenues for addressing these known issues. This subsequently led to the development of a novel code of practices to promote scientific rigor in blockchain energy studies.

We believe that this study assists the reader in understanding the reliability of the current energy and environmental studies in a blockchain context. This article also assists researchers and developers in designing or refining their existing models through adherence to the code of practices. The paper makes the following contributions:

- We systematically review the existing literature to document common issues with energy and environmental impact studies for blockchain-based systems (Section 3).
- We develop a novel quality assessment framework for blockchain-specific studies that can assist in understanding the scientific rigor of the energy or environmental impact model (Section 3).
- We identify research gaps specifically with regard to the lack of non-Bitcoin-specific investigations in the academic literature. We also report on the lack of empirical data for these models (Section 4.1).
- We manifest the findings of our review in a set of best practices that can assist in designing or improving existing models (Section 5).

2. Background

Cryptocurrencies, depending upon their consensus mechanism, may cause two prime concerns from an environmental perspective: the electricity consumption and the carbon emission associated with the energy

[1] In distributed computing systems such as peer-to-peer network-based cryptocurrencies, a consensus mechanism is employed to achieve an agreement on a single view of the data such as a ledger of transactions. We refer the reader to Zheng et al. [6], for further information on consensus mechanisms in blockchain-based systems.

[2] In Bitcoin like PoW-based cryptocurrencies, the participants are tasked with the problem of finding a block hash value below a set threshold. The difficulty of this problem is periodically changed to maintain the system property of a 10-minute average time difference between two blocks of transactions.

[3] This is particularly important for the guidelines provided by Sovacool et al. [16], as these guidelines are not specific to the blockchain domain.

consumption[4]. In this section, we provide an overview of how the energy and environmental footprints of these cryptocurrencies are usually measured.

2.1. Note on nomenclature

In the context of blockchain systems, it is important to note that the term "energy consumption" is predominantly used in the literature to refer to "electricity consumption". The energy consumed by blockchain-based systems is primarily in the form of electricity utilized by computer systems, which comprise hardware, software, and networking components. As such, it may be more accurate to refer to this as "electricity consumption" rather than "energy consumption." However, there are instances where the use of the term "energy consumption" may be more appropriate, such as when conducting a life cycle assessment (LCA) of mining hardware. Throughout this article, we endeavor to use the term "energy consumption", where the use of "electricity consumption" may be too restrictive.

2.2. Electricity consumption

As alluded to in the introduction section, measuring the electricity consumption of a geographically distributed network is a non-trivial task. This problem is compounded when considering decentralized systems, as it is difficult to find a centralized source of information about the network's physical composition [20,21]. There are two main ways of estimating the electricity consumption of a blockchain-based system depending on the availability of reliable data on the computing network: bottom-up and top-down.

2.2.1. Bottom-up

A distributed computing network is composed of computational devices that consume a certain amount of electricity per unit of work[5]. Each of these computational devices can have different performance and energy efficiency profiles. For example, a network could be made up of 100 Raspberry Pi (see www.raspberrypi.com) devices generating X units of work in a single unit of time, or it could be made up of 2 consumer-level personal computers generating the same amount of work in the same temporal resolution.

One of the early attempts at using a bottom-up approach for modeling the electricity consumption of Bitcoin was made by Bevand [22]. In his analysis, Bevand outlined prominent modern Bitcoin mining hardware that typically employs application-specific integrated circuits (ASICs) designed for Bitcoin mining. For example, an Antminer S9 system released in 2017 could perform 13 TH/s, whereas a consumer CPU such as Intel i7 (2021) can only perform 2.5 kH/s.

If we are aware of the exact hardware used in the network, including the hardware distribution (how many units of each type of device are on the network)[6], we can use this information to calculate the total energy consumed by all the constituting computing devices.

This calculation requires accurate values of each device's computing power and energy efficiency. This in itself can be problematic in a real-world scenario, as most of the information about power and energy efficiency is obtained through data sheets provided by manufacturers. These data sheets in most cases are not verified by an independent auditor. Furthermore, tuning operational parameters such as clock frequency and supply voltage may even lead to different numbers in practice. For an accurate measurement, it is also important to know the uptime for each device and the actual work done during this uptime.

This gives us a partial understanding of the network's electricity consumption as this calculation does not consider operational electricity consumption for devices other than the computing device such as the networking or cooling infrastructure. These operational expenses are often considered in the form of a fractional value known as power usage effectiveness (PUE) [23]. Once we know the electricity consumption of each device in the network and the associated PUE value, we can calculate the total electricity consumed as follows[7]:

$$T = \sum \varepsilon(i) \times \mathrm{PUE}(i) \tag{1}$$

where T is the total electricity consumption, ε is the electricity consumption of each constituting computing device and PUE is the additional operational electricity requirement. This calculation is also visually illustrated in Fig. 1.

2.2.2. Top-down

Bitcoin and other cryptocurrencies are often described as decentralized systems. Decentralization is a crucial component of the network on different levels of operations, such as applications (decentralized exchanges), protocols (consensus mechanisms), and networks (distributed peer-to-peer networks) [20,24]. This decentralized nature of the network makes it difficult for researchers to collect empirical data on the location and hardware of the consensus participants. In PoW-based systems, these consensus participants are also known as miners.

Due to the lack of reliable empirical information on network participants, a large number of energy and environmental models are based on a top-down approach [25]. In a top-down modeling approach, the model relies on high-level technical, economic, or social variables.

For instance, Vranken's top-down model [26] is based on the total computing power, also known as the hash rate. In De Vries's model [27], the author built an economic model based on the economic rationality of the miner.

In this subsection, we provide an abstract overview of how a top-down model conceptually works; however, we refer the reader to Refs. [8,25,26] for an in-depth discussion of top-down modeling.

For Bitcoin, we can calculate the total hash rate of the network by using the difficulty of mining [28]. First, an estimate of the hash rate on the network is required. This can be derived from a simple statistical model that considers the difficulty parameter and the time it takes on average to mine a block, which is 10 min for Bitcoin, as applied by O'Dwyer and Malone [28]. A more refined model may use empirical data on the exact amount of time it takes to mine blocks. For instance, for Bitcoin, the difficulty parameter is adjusted every 2016 blocks, and hence, some drift may occur in between.

The total hash rate of the network is composed of the combined hash rates of a number of different hardware in use, each with a specific energy and performance profile. A number of different combinations and permutations of available hardware can generate the required hash rate. Different models make different assumptions to obtain the total hash rate. For example, some models assume that the network is made up of only the most efficient commercially available hardware, while others consider a pool of hardware with different distributions. We can represent this calculation as follows[8]:

[4] There are other environmental impacts associated with cryptocurrency operations, such as E-Waste generation [19]. We briefly touch on this in a subsection; however, our focus in this article is primarily on energy consumption and carbon emissions.

[5] In PoW-based cryptocurrencies, the work is often performing hashing operations to find a nonce value such that the resulting hash value is lower than the target.

[6] When modeling electricity consumption, hardware distribution is a key factor. In practice, researchers and practitioners rely on energy efficiency data sheets for (more) precise information.

[7] It is worth noting that this equation is for illustrative purposes only as in the real model, the authors might account for additional factors such as the economics of operators.

[8] There are often many different combinations of devices possible here with different ρ. For instance, a small network can be made up of a large number of inefficient devices or a small number of highly efficient devices.

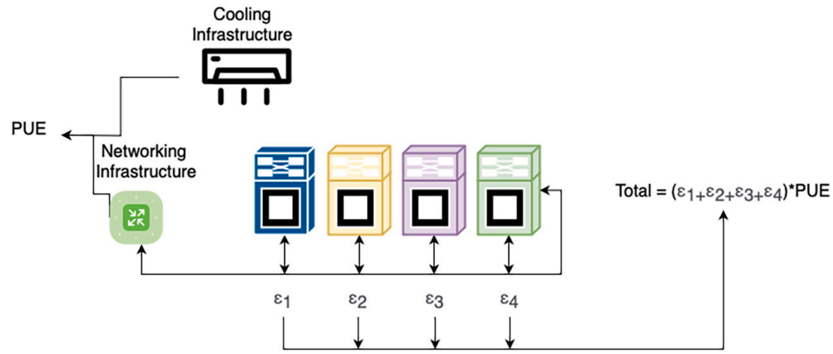

Fig. 1. Calculating electricity consumption using bottom-up approach.

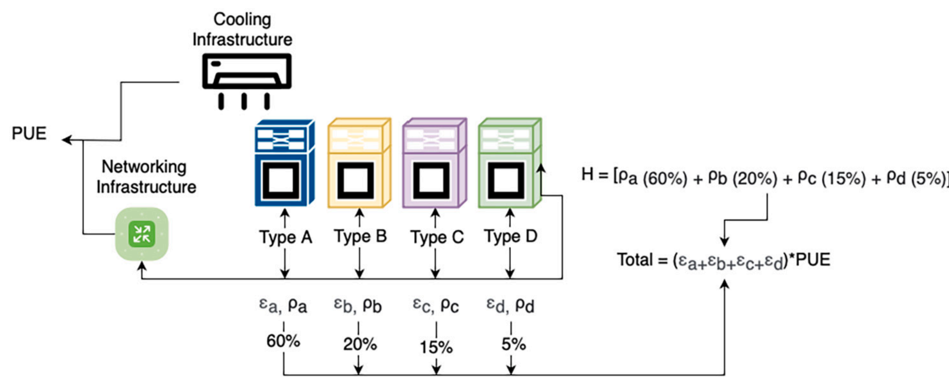

Fig. 2. Calculating electricity consumption using a top-down approach.

$$H = \sum \rho \tag{2}$$

Here, H is the hashing power of the network composed of all the individual hashing power (ρ) of the hardware used in the network. Once a pool of hardware is decided, we can similarly calculate the electricity consumption to that of bottom-up[9]:

$$T = \sum \varepsilon(i) \times \mathrm{PUE}(i) \tag{3}$$

We have also illustrated this process in Fig. 2.

2.3. Environmental impact measurement

The scope of the environmental impact of information technologies can be very broad ranging from the direct impact caused by E-waste to the consumption of electricity generated by non-renewable carbon-intensive operations such as coal-based power plants [29]. Through our literature review, we report that most of the studies in the blockchain context focus on the carbon emissions associated with the electricity consumption of the network. However, it is worth noting that there are a few studies that look at other aspects of the environmental impact of cryptocurrencies, such as E-waste generation [19], and scope 2 and 3 carbon emissions [30]. In this subsection, we focus on carbon emissions.

The carbon emission calculation consists of a five-step process as outlined below. It is important to note that these steps are only indicative of the process, and individual studies may differ in their approach:

1. Calculating the total electricity consumed by the network: This can be determined by either bottom-up or top-down approaches as discussed above.
2. Determining the geographic location of the devices in the network: In addition to understanding the pool of hardware and their respective share of the total network, we also need to know their geographic location.
3. Obtaining the energy mix for shortlisted geographic locations: For each geographic location, we need to understand the breakdown of electricity in the form of its sources of energy. An electricity grid in a country with heavy reliance on renewable energy is likely to produce considerably less carbon per unit of electricity than a country with a coal-based electricity grid.
4. Use grid emission factor to calculate the CO_2 emission: A geographic area with a predominantly green grid should account for less carbon emission per unit of electricity generation and consumption, while one with reliance on coal power should have high carbon emission. This is captured in the form of a grid emission factor. The grid emission factor measures the amount of CO_2 emissions per unit of electricity in a given geographic area.
5. Sum all the carbon emissions: In this step, we generate a final carbon emission value for the whole network by adding all the individual carbon intensity data points from each constituting geographic location.

2.4. Sensitivity of the blockchain energy consumption model

In the preceding subsections, we provide a brief overview of the prominent methodologies used to calculate both the energy and environmental impact of cryptocurrencies. It is important to note that the bottom-up method, despite seeming straightforward is quite difficult to execute for large cryptocurrencies due to the lack of reliable data. This may be due to the inherent pseudo-anonymous structure of these cryptocurrencies and the reluctance of participants to disclose their identities due to the fear of governmental retaliation [31].

Thus, most of the current studies are based on the top-down modeling approach, which requires a number of assumptions that may impact the final estimate of the model. For example, one of the most widely used models for Bitcoin is designed by the CCAF [25]. This model at-

[9] The ε is only for hardware that contributes to H above.

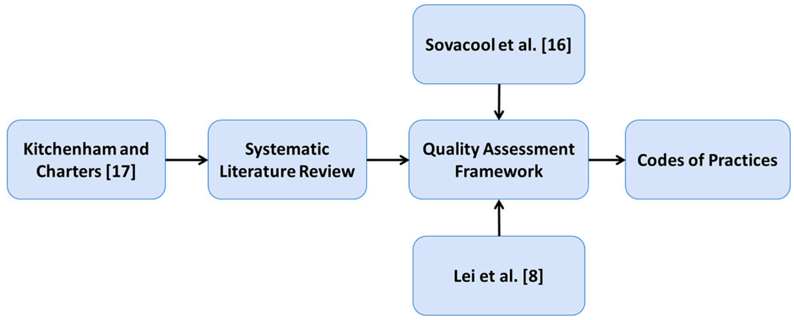

Fig. 3. Research flow.

tempts to use the profitability of mining to calculate the total power consumption of the network. In doing so, the authors made a number of assumptions including an assumption that on average the miners pay 0.05 USD/kWh. If we change the cost of electricity to 0.10 USD/kWh[10], the final model estimates drop by approximately 38%[11]. This is a significant difference based on the change in a single variable, which alludes to the sensitivity of the model to its parameters.

Unlike the CCAF model, where most of the parameters are based on empirical assumptions, other models do not fair well. For instance, the models developed by Mora et al. [12] and De Vries [27] have been widely criticized for their underlying assumptions [10,11,13].

In this article, we attempt to systematically document these variables and their impact on accuracy to allow researchers a means to improve their models or representation of the results from these models.

3. Methodology

In this section, we provide an overview of the methodology used to conduct the systematic literature review, quality assessment framework, and code of practices. Due to the diverse nature of blockchain research, it is difficult to come up with a single "ten steps to quality" measure. Thus, we begin our search by identifying and classifying prominent research methodologies employed in blockchain energy sciences. To this end, we define two primary research questions for our literature search:

1. **RQ1**: What are the different methods used to measure/model the energy or environmental footprint of blockchain-based systems and their associated modeling assumptions?
2. **RQ2**: What are the implicitly or explicitly acknowledged strengths or limitations of these models?

We attempt to answer these questions through a systematic literature review. For the systematic literature review, we follow the guidelines proposed by Kitchenham and Charters [17] to identify relevant literature in both academic and non-academic domains. Based on the literature review, we iteratively refine and generate a novel quality assessment framework based on the guidelines proposed by Lei et al. [8] and Sovacool et al. [16].

We then apply this quality assessment framework to the shortlisted studies and document common issues with the shortlisted studies. These issues form the basis for the code of practices proposed later in the study. This code of practices is based on well-established standards from information systems, statistics, social energy sciences, and blockchain-specific literature. The flow of this study is illustrated in Fig. 3.

[10] This change is in line with other studies focusing on Bitcoin's energy consumption. We will discuss how the cost of electricity varies significantly in different studies.

[11] This calculation is based on data collected on July 23, 2022.

3.1. *Systematic literature review*

As indicated earlier, we follow the guidelines put forth by Kitchenham and Charters [17]. We conduct our review in 3 phases: In the first phase, we construct the search query by explicitly documenting search strings. In the second phase, we conduct the search for relevant articles by first shortlisting appropriate repositories and sources of academic and non-academic literature. In the final phase, we extract the measurement/modeling technique used for energy/environmental footprint analysis from the shortlisted articles. These three phases are illustrated in Fig. 4.

3.1.1. *Phase 1: Search query formation*

RQ1 increases the coverage of our research by capturing different models and their associated assumptions. If a shortlisted article proposes a new model with assumptions, we include these new variables in our quality assessment framework. For each identified variable, we document all the assumptions made by the study.

With RQ2, we further solidify our understanding of the limitations associated with these models by documenting the acknowledged strengths and weaknesses of each shortlisted model. This documentation helps us prepare our code of practice.

These two primary research questions serve as the basis of our literature search strategy. We start by constructing a search string to identify relevant literature. We do this by first constructing an initial set of keywords based on both RQs and then further validating these keywords by performing backward and forward snowballing on Vranken [26][12]. We choose Ref. [26] for snowballing, as it is one of the first papers in the field while being one of the highest cited. The final search query is of the following structure:

```
Blockchain: "Blockchain" OR "DLT" OR "bitcoin" OR "
    blockchain" OR "cryptocurrencies" OR "cryptocurrency"
     OR "digital currency" OR "distributed ledger" OR "
    peer-to-peer computing" OR "smart contract platform"
Energy: "electricity" OR "power" OR "power supply"
Consumption: "expenditure" OR "use" OR "utilisation" OR "
    utilization"
Environment: "atmosphere" OR "carbon" OR "climate" OR "
    ecological" OR "emission" OR "environmental" OR "
    green" OR "footprint" OR "e-waste"
Sustainability: "green design" OR "green technology" OR "
    sustainable"
```

3.1.2. *Phase 2: Article search*

In the article search, we intend to ensure high coverage of models and their associated limitations and strengths by extracting relevant models from both academic and non-academic literature sources. For academic sources, we search prominent computer and energy science

[12] We plot a word cloud from the backward and forward snowballing in Appendix.

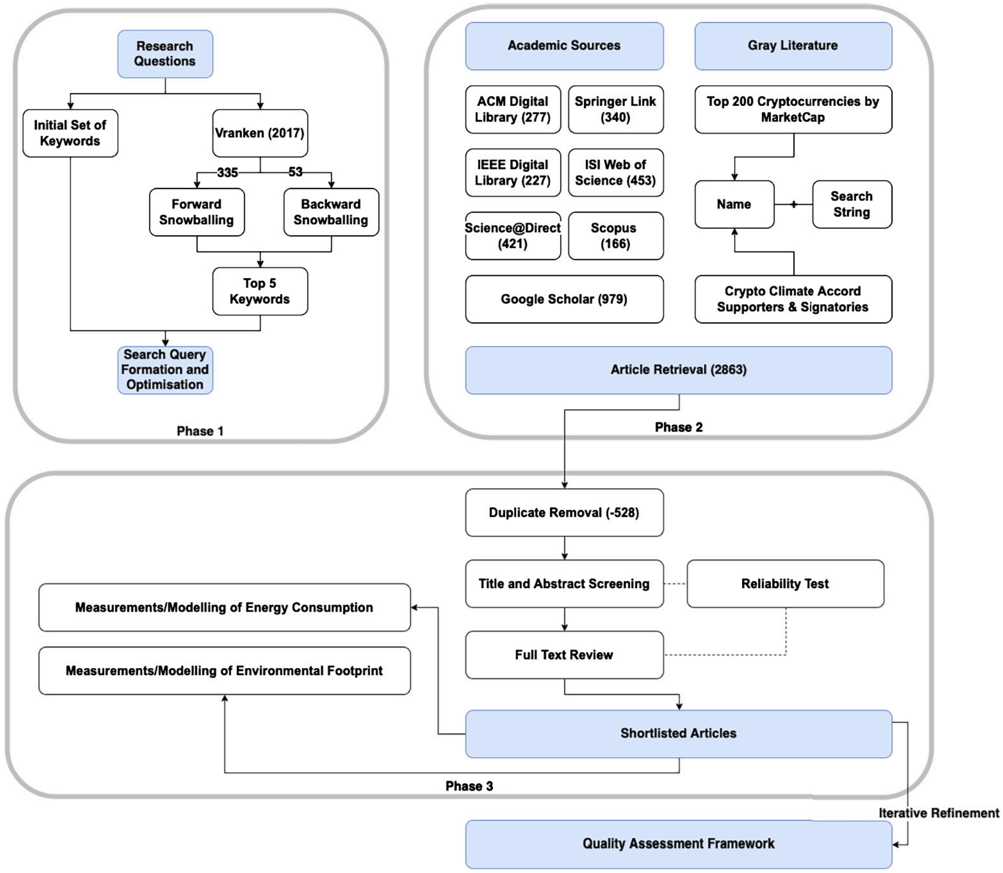

Fig. 4. Systematic literature review protocol.

research repositories: Google Scholar, ACM Digital Library, IEEE Digital Library, Web of Science, ScienceDirect, Scopus, and SpringerLink[13] Out of these search repositories, Google Scholar returned the highest number of articles due to its wide scope and coverage of potential non-academic literature as well. However, due to the limitations imposed by Google, it is only feasible to obtain the first 980 results. It is worth noting that the relevance of results depletes significantly in Google Scholar after the initial few articles [20].

For non-academic literature results, we identify two primary sources of information: the cryptocurrencies consuming the energy and third parties providing services around these cryptocurrencies. We conduct an exhaustive search of the top 200 cryptocurrencies by market cap using the Google search engine to locate any energy or environmental footprint models developed for these cryptocurrencies. A complete list of cryptocurrencies reviewed is presented in Appendix.

Due to the vast number of results, we limit our Google search to only the first 100 results. To capture the models developed by service providers, we consult the list of supporters and signatories of the Crypto Climate Accord[14]. As the nature of our query is left intentionally vague to improve coverage, we end up with an initial set of 2863 articles. Out of these 2863, a majority (2614) of the results are from academic sources. The non-academic literature consists of 259 results, mostly in the form of blog posts and white papers[15]. In the next phase, we document the process used to shortlist the relevant articles.

3.1.3. *Phase 3: Shortlisting relevant articles*

Due to a large number of articles returned for our initial search, we conducted a title and abstract screening to shortlist the articles using explicit inclusion criteria as outlined below:

1. The paper's title mentions energy or environment, or any of the synonyms mentioned in Section 3.1.1, or is potentially relevant to the study of blockchain energy or environmental impact research.
2. The abstract is relevant to the measurement of energy or environmental footprint.

This filtration was performed by the first author and resulted in 244 potentially relevant articles. To ensure that the title and abstract review process is reliable, we conduct a reliability test. To do this, we performed cross-validation using the mechanism proposed by Fleiss and Cohen [32]. We perform cross-validation on a sample of 67 with a confidence interval of **90%** and a margin of error of **10%** calculated in line with the suggestions of Sim and Wright [33]. These 67 shortlisted articles were independently screened and reviewed by the second author to perform cross-validation. The results from the cross-validation indicate a substantial agreement between the authors with Cohen's Kappa value of 0.71. This suggests that the shortlisting process is reliable.

The title and abstraction process significantly reduces the number of articles in our search pool to a set of 244 articles. We then reviewed the

[13] The exact search query used for each of these platforms is reported in Appendix.

[14] The Crypto Climate Accord is a recent industry-driven accord that limits the environmental impact of blockchain technologies. As of June 1, 2022, there are 200 supports and 150 signatories to the accord.

[15] The actual number of non-academic literature included in our study is likely significantly high due to the inclusion of results from Google Scholar, which tends to capture some non-academic literature sources such as academic blogs.

Table 1
Relevance matrix.

Factor	Yes	No
Measurement or modeling technique identified for energy/electricity consumption	0.5	0
Measurement or modeling technique identified for environmental footprint	0.5	0
Acknowledgment of strength and/or weakness of an energy/environmental footprint model	0.5	0

full text of these 244 articles to assess their relevance. To assess their relevance to our survey, we define a relevance matrix for shortlisting in Table 1. A study that complies with any one of these three factors is included in our final set of relevant articles. This results in the final set of 128 studies that are considered directly relevant to our research. A complete list of all the shortlisted studies is included in Appendix.

3.1.4. Completeness of quality assessment framework and reliability of coding

To ensure the completeness of the quality assessment framework, we iteratively refine it as we progress through the full review of the shortlisted articles. To ensure reliability in the refinement process, any suggested amendments to the framework are discussed with both the authors and an independent expert observer. For this study, Dr. Alan Ransil of Filecoin Green and Procotol Labs volunteered for the role of independent expert observer. There were no disagreements on the refinement of the quality assessment framework. The concluding framework is documented in the subsection.

The finalized quality assessment framework is then used for the quality assignment of all shortlisted articles. To ensure reliability in this process, we perform a reliability test by independently reviewing a subset of the shortlisted articles (8). Both authors agree on seven of the assignments with a disagreement on several aspects of one article. This disagreement was resolved when the article was reviewed by an independent expert observer.

3.2. Development of quality assessment framework

Objectively assessing and discussing the scientific rigor of a field as diverse as blockchain is a difficult exercise mostly due to the diverse nature of the research and application of this technology. In this study, we attempt to understand the common issues in the design and execution of these studies and provide guidelines for avoiding these mistakes. To do this, we first adopt the basic research elements proposed by Sovacool et al. [16] into four quality indicators: clear research question, building upon existing knowledge, explicit research design, and reliability assessment of the underlying data. We document these four basic research quality indicators, their associated measurement techniques, and the reference literature used to derive the categorization in Table 2.

These four indicators of basic research quality allow us to discuss research design independent of underlying research methodologies. Out of these four indicators, clear research questions and building upon existing knowledge are more applicable to the scientific literature, while the last two are equally applicable to studies in the non-academic literature as well.

It is expected that a scientific study should contain a research question, ideally explicitly stated within the text body [34]. The second indicator focuses on the use of existing knowledge to inform the construction of newer models, which can be in the form of building upon existing energy or environmental footprint models within the blockchain domain or the application of theories from fields such as economics or social sciences. The use of existing theory promotes the evolution of the research within a field while providing a more robust mechanism to discuss strengths and limitations.

Another important research aspect of studies that focus on the environmental impact of technology is the reproducibility of the analysis. We capture this through the third basic research quality indicator, explicit research design. Within this indicator, we expect that a study should contain sufficient details of the study design and execution to permit independent reproduction. It is also expected that a study reliant on a non-public dataset should share the dataset and appropriate source code where possible.

If the study is utilizing data sources from non-peer-reviewed studies or investigations, it should include a detailed description of potential reliability issues in the data. This is particularly important in the blockchain context, as the non-academic literature often contains unvetted datasets for which the reliability is not validated.

3.2.1. Different research methodologies

Through the full-text review of the final set of 128 studies, we identified prominent research methodologies employed in the shortlisted studies. We review all the shortlisted articles and classify them based on the scheme provided by Sovacool et al. [16]. In total, we identify five distinct research methodologies: quantitative energy modeling, literature review, data analysis and statistics, case studies, and experiments.

Out of these five, the quantitative energy modeling approach is highly dependent on the structure of cryptocurrency and the modeling approach adopted (top-down or bottom-up). Therefore, the quality indicators for quantitative energy modeling are closely tied to the variables discussed in Section 2. Unlike quantitative energy modeling, other research methodologies are not as dependent on the structure of the cryptocurrency and can benefit from more generic field-specific quality indicators. For instance, the way a literature review is conducted is independent of the subject under review. Thus, we reason that it is more appropriate to employ quality indicators that focus on the method irrespective of the subject. This focus on quality indicators centered on the method allows us to examine studies that might not be specific to popular crypto-assets such as Bitcoin and Ethereum.

In the following subsection, we briefly describe each of these research methodologies and provide an outline of the quality assessment mechanism.

- Quantitative energy modeling: Quantitative models rely on a robust dataset that is used in conjunction with a social, economic, or technical model of the blockchain. For instance, the CCAF model is an example of a computational economic model, while the Digiconomist model is an example of a socioeconomic model. Both of these models are based on the use of quantitative analysis for the estimation of energy consumption. The exact model composition within quantitative modeling can be based on any combination of the social [9,37–39], economic [8,11,16,26], and technical variables above; thus, the exact quality indicators vary significantly depending on the model formulation and the intended use. Technical variables considered in these studies may include hardware composition [8,11,16], geographic data [8,11,16,22], and temporal resolution of the model and its underlying data [16]. These categories are outlined in Table 3. It is worth noting that while most of the discussion here focuses on PoW, the identified quality indicators may also apply to other consensus mechanisms such as proof-of-stake. However, we leave the refinement of these quality indicators for other consensus mechanisms to future work.
- Literature review: A narrative review of the literature can provide useful insights into existing models and the potential strengths and weaknesses of these models. These reviews can be targeted toward different crowds, some of which focus exclusively on experts in the domain [8], while others are a more general introduction to the energy and environmental footprint of blockchain [26].
  Literature reviews are widely used in a number of cognizant fields, such as information systems [17] and computer science [16,41]; thus, there is a wealth of guidelines on the quality of these reviews. We shortlist a subset of these guidelines based on Refs. [16,17,41], and construct our quality indicators as outlined in Table 4.

Table 2
Basic research quality indicators.

Quality indicators	Measurement	Application area	Reference
BR1: Clear research question	Explicit or implicit research question	Academic literature	[16,34]
BR2: Building on existing knowledge	Use of a theory or framework from existing studies	Academic literature	[16,34]
BR3: Explicit research design	Clear research methodology, public data and source code (where appropriate)	Academic and non-academic literature	[8,16,34,35]
BR4: Reliability of external data	Reliability assessment	Academic and non-academic literature	[8,16,34,36]

Table 3
Quality indicators for quantitative energy modeling.

Quality indicators	Description	Measurement	Reference
QM1: Hardware composition	Hardware or pool of hardware used for calculation of energy consumption	Examination of the (accuracy of) exact hardware distribution, checking assumption (accuracy) related to hardware efficiency	[8,11,16]
QM2: Geographic data	Geographic data such as the location of miners in the network is crucial for calculating CO_2 emissions	Assessment of the methodology used for the extraction of geographic distribution	[8,9,11,16]
QM3: Economics	Participating in a cryptocurrency incurs both capital expenditure to acquire hardware and related operational expenses	Analyzing the cost of electricity, power usage effectiveness (PUE) value, hardware lifespan, impact of transaction fees and halving on rewards	[8,11,16,26,38,40]
QM4: Social	Incentive engineering behind the design of proof-of-work (PoW) requires the miners to be rational, this also needs to be accounted for when calculating energy consumption	Rationality of agents	[9,37,38]
QM5: Carbon intensity data	The carbon intensity data is crucial for calculating environmental impact of these crypto-assets	Source of the data and geographical resolution	[8,10,11,26]
QM6: Time resolution	The data used in quantitative energy modeling tends to evolve over time along with the model itself thus it is important to provide time resolution for the data along with appropriate archiving of legacy data or model parameters	Review of the data and model parameters	[16]

Table 4
Quality indicators for literature review.

Quality indicators	Description	Measurement	Reference
LR1: Type of review method	Some review methods such as narrative review are considered less rigorous than a meta-analysis of the literature, we follow the guideline to classify rigor based on method type	Review of method	[16,17,41]
LR2: Explicit criteria	To avoid bias, it is important to document explicitly the RQs, search strings, inclusion and exclusion criteria	Review of method	[16,17,41]
LR3: Appropriate search database	In fields like blockchain, a significant portion of the research is conducted within non-academic literature thus it is important to have a high coverage	Review of method	[16,17,41]
LR4: Sampling process documentation	If only a specific sample is analyzed, it is important to document the process	Review of method	[16,17,41]

Table 5
Quality indicators for data analysis and statistical models.

Quality indicators	Description	Measurement	Reference
DA1: Type of analysis performed	Some analysis methods such as univariate are considerably less rigorous than a longitudinal multivariate analysis	Review of method	[16,42]
DA2: Clear hypothesis	It is important that in statistical analysis, the hypothesis is clearly stated and evaluated	Review of method	[16,34,42]
DA3: Practical vs. statistical significance	High statistical significance does not necessarily mean that the relationship analyzed or predicted is of practical significance thus the practical significance must be discussed when presenting a statistical significance value	Review of method	[16,42]

- Data analysis and statistics: Unlike quantitative energy modeling, a method that relies on data analysis and basic statistics is far less sophisticated. These models often rely on univariate or multivariate analysis of variables associated with the energy or environmental footprint of cryptocurrencies. Sovacool et al. [16] provided an extensive guideline on quality indicators for these studies, and we compound their guidelines with the 10 rules proposed by Kass et al. [42]. Based on these two studies, we identify three relevant aspects of studies within the blockchain energy space, and these three aspects are outlined in Table 5.
- Case studies: Case studies allow us to take an in-depth look at a specific instance of a broader phenomenon [43]. In energy studies of blockchain, these case studies are often specific to cryptocurrencies and their energy or environmental footprint. There is a wealth of literature on the selection and execution of case study analysis to ensure that the studies are of high relevance to the broader field [43,44]. We adopt and document these guidelines in Table 6.
- Experiments: In some instances where it might be difficult to extract information from observations, it may be useful to conduct experiments [16,45]. In the blockchain energy domain, these experiments are often performed to calculate the energy consumption

Table 6
Quality indicators for case studies.

Quality indicators	Description	Measurement	Reference
CS1: Case selection	Selecting an appropriate case study is important to ensure that the results are generalizable or of extreme cases that may explain a specific phenomena	Review of selection mechanism	[16,43,44]
CS2: Clear boundaries	Case studies can be very broad or hyper-specific thus it is important to document the bounds of the study	Review of method	[16,43,44]
CS3: Measurable dependent and independent variables	If the case study is analyzing a phenomenon, it is crucial to clearly define the dependent and independent variables	Review of method	[16,43,44]

Table 7
Quality indicators for experiments.

Quality indicators	Description	Measurement	Reference
EX1: Representative sample	Selection of the experiment object is crucial for the generalizability of results	Review of selection mechanism	[16,45]
EX2: Choice of setting	The settings of the experiment should closely resemble that of the real world object to ensure reliable results	Review of method	[16,45]

of a specific device and then use that information to generate measurements for a subset or the whole of the network. The quality indicators for the experiments are shown in Table 7.

In the next section, we employ these quality indicators and assess the quality of the academic and non-academic shortlisted literature through the systematic literature review.

4. Results

In this section, we provide an overview of our quality analysis. We begin by describing the trends in the literature followed by a description of the prominent results.

The field of blockchain energy science is quite new, with the first academic study published in 2014. The field has only seen traction since 2018. The non-academic literature has specifically seen large growth since 2020, a period also referred to as the DeFi summer [46].

The research methodologies in use have also evolved. In the early years, most of the models proposed were quite simple data analysis and statistical models, but since then, these have evolved into more mature quantitative models incorporating economic, social, and technical modeling. The field has also seen an evolution in each of these research methods as well. For example, we note a trend toward more sophisticated literature reviews such as meta-review.

The biggest development in terms of influential literature has been the introduction of the Digiconomist index [47] and then the Cambridge Bitcoin Electricity Consumption index [25]. Both of these indexes now underpin several of the assumptions made by newer studies. For instance, the value of electricity cost has seen a significant change since the introduction of the CCAF index, and a majority of newer studies assumed the cost of electricity to be 0.05 US Cents/kWh in line with the assumption made by the Cambridge Index.

These trends, however, do not indicate that the field is becoming more rigorous over time. In the following subsection, we discuss the basic research quality indicators followed by more specific quality checks for the five research methodologies outlined in Section 3.

4.1. Quality assessment

4.1.1. Basic research quality indicators

We have documented the results from our analysis in Table 8. It can be seen that a handful of the shortlisted studies do not describe their research goal clearly, making the document difficult to understand and contextualize.

A bigger and more systematic issue with research in this domain is the lack of building upon existing knowledge. As seen from our analysis, 74% of the studies do not build upon existing theories. For instance, in their analysis of the life cycle of Bitcoin mining, Köhler and Pizzol [29] utilized the well-established Life Cycle Assessment methodology allowing us to use existing theories to assess their estimates and independently verify their results.

Table 8
Basic research quality indicator results.

Quality indicators	Results
BR1: Clear research question	5% (7) studies do not contain a research question
BR2: Building on existing knowledge	74% (95) of the analyzed studies do not build upon existing theories in blockchain or any other related domain
BR3: Explicit research design	34% (44) of the studies do not have an explicit research design, while 43% (55) of studies do not share data whereas 67% (86) of studies do not share source code
BR4: Reliability of external data	79% (101) studies do not discuss the reliability of external data used in their analysis

In contrast, analyses conducted by studies such as Ref. [27] do not build on any existing theory and present their own methods without contextualizing it in the existing literature in blockchain or other scientific disciplines, making it difficult to compare the robustness of the underlying research method. We believe that there needs to be a more coordinated approach to build upon existing knowledge within the blockchain research domain as well as other cognizant research domains such as energy sciences, economics, and information systems.

Another important aspect of a reliable scientific study is the reproducibility of its results; however, this is not trivial in blockchain energy sciences, as 34% of the analyzed studies do not contain enough information to reproduce their analysis. This issue is even more prevalent when a supporting dataset is needed for reproduction. A large portion of studies (43%) do not share their underlying dataset, either requiring the reviewer to seek the dataset or, in some cases, deferring the publication of the dataset to an unspecified period in the future. The widely cited CoinShare report [48] is an example of this.

Some of the analyzed studies employ mathematical models that can assist in generating a system-wide figure for energy consumption or environmental footprint. These mathematical models are often deployed in the form of an Excel sheet or a computing script. In either of these situations, it is desirable to have access to the underlying calculation; however, a vast majority (79%) of the studies do not provide access to their source code.

As alluded to in the previous section, the field has seen significant interest from both academic and non-academic sources in the past four years, resulting in a plethora of models based on different assumptions and datasets. However, the reliability of these models and the dataset is not well established in most cases. For instance, a common assumption about the lifecycle of typical Bitcoin mining hardware is that the hardware will only be profitable for 1–2 years.

This assumption is based on a study by De Vries [27] published in *Joule* as a commentary. However, since its publication, this assumption has been widely criticized for its oversimplified view of mining operations and dependence on anecdotal examples to back the assumption [11].

Several studies have also established that the mining hardware used in Ref. [27], for example, was profitable for over four years since its production, undermining the assumptions in the initial study. However, as Ref. [27] is the first to give an estimate of the lifecycle of a typical Bitcoin miner, it is still being used in numerous studies despite the potential flaws. This is one of several instances where a flawed or outdated study has been widely used to develop newer models. Based on our analysis, we report that a majority of the studies (79%) analyzed rely on an external dataset but do not acknowledge any potential validity issues in these studies.

The basic research quality indicators suggest that the field is likely evolving toward more mature research methodologies; however, the reproducibility of this research is hampered by poor data and source code-sharing practices. It can also be seen from this analysis that most of this evaluation is performed in small silos rather than the whole field evolving together. This is most likely due to the lack of reuse of theory and high dependence on some non-academic or unvalidated datasets. We address these potential research limitations in Section 7, where we propose a set of novel codes of practices that may assist in improving the quality of research.

4.1.2. Quality of quantitative energy modeling studies

A majority of the studies analyzed are quantitative energy models; thus, we start our discussion by providing an overview of the quality of quantitative energy models. We highlight our results per quality indicators from Table 3. The following subsection highlights some of the common issues while providing a thematic overview of each of these quality indicators. We refer the reader to the supporting material available at: www.github.com/ashishrsai/energy for more detailed results.

1. **QM1:** Hardware assumptions
   (a) Improper hardware efficiency assumptions:
      - In Ref. [12], the authors kept the power efficiency of mining hardware constant while conducting their simulations for the next 100 years. They did not provide clear reasoning behind their choice. This has been criticized by many matters published in Nature Climate Change [10,13,14]. These studies have demonstrated that this choice alone impacted Mora et al.'s results [12] considerably, rendering their analysis provably inaccurate.
        It is also worth noting that in economic equilibrium where miners' revenues and costs are equal, hardware efficiency does not have an impact on a PoW blockchain's total electricity consumption anymore [5,49].

   (b) Assuming single hardware in use:
      - A more common error in many of the analyzed studies is in their decisions regarding the composition of the hardware pool, including the choice of hardware and the proportion of each machine. In their initial attempt O'Dwyer and Malone [28] did not incorporate a diverse range of hardware in their analysis. Instead, their estimates were derived from the lowest (commodity) and highest (specialist) hardware configurations. While this methodology could have influenced the accuracy of their results, it is essential to understand the context in which they were operating. At that time, detailed data about the global distribution of Bitcoin mining, especially in relation to electricity costs, were not as comprehensive as they are now. It should be acknowledged that O'Dwyer and Malone [28] did posit that Bitcoin might consume electricity equivalent to Ireland's consumption, although without detailed calculations to back this claim—a point later critiqued by Koomey [11] and Vranken [26]. However, their estimates, while representing the extreme boundaries, provided an essential early glimpse into the potential energy consumption of Bitcoin. It is important to recognize that both the lower and upper bounds, while offering a range, may not be representative of the entire network, as it is unlikely that all participants would be operating either the most or the least efficient hardware.
      - In another study [50], it is worth mentioning that the methodology employed focused on a single hardware device's efficiency rather than a pooled hardware assessment. It is essential to recognize that our critique in this context should be regarded within the context of our survey article's broader objective, which is to scrutinize and evaluate existing research works. Gallersdörfer's study did not specifically aim to deliver a solitary estimate; rather, it attempts to approximate the additional electricity consumption attributable to various PoW currencies.
        For instance, in their examination of Bitcoin, the authors selected the Bitmain Antminer S17 Pro 53TH as the representative hardware device. However, we note that they did not offer compelling evidence or a robust rationale for this particular choice as a suitable proxy for the entire network's characteristics. This lack of justification may raise questions regarding the generalizability of their findings to the broader Bitcoin network and, by extension, other PoW currencies.

   (c) Filling in the missing data:
      - In a more recent study by Stoll et al. [37], the authors attempted to improve the accuracy of the hardware pool by reviewing IPO filling for three major Bitcoin mining hardware providers. They assumed that the top three hardware providers control all of the supply despite each listed IPO acknowledging that there is approximately 5% to 15% of supply beyond these three providers. The exact distribution between these three suppliers also varies considerably between the three IPOs and the article itself. We believe this limits the reliability of this model considerably. The authors additionally made many assumptions that are not based on the data provided within the IPO reports, such as the assumption in their supplementary Sheet 3.4, these assumptions lack empirical basis while being consequential to the final prediction. Assumptions made by Stoll et al. [37]:
         i. The IPO filing does not specify the sales figures per model; thus, the authors assumed an equal distribution of sales on all available ASIC models.
         ii. Post the publication of the IPO filings, the authors assumed that the number of ASICs sold per month in 2018 stays constant.

        Both of these assumptions may skew the results in favor of old hardware that might not be as efficient as newer iterations. Similar assumptions of the equal spread of device sales over a time horizon are also foundational to the work presented in Ref. [27].
      - In the follow-up work, De Vries [51] attempted to justify how old devices (specifically Antminer S9) are still the dominant mining hardware, this is supposedly backed by the evidence in supplementary data Sheet 1, however, on inspecting the data sheet we notice that in 2019, for H1 (first half of the year), the author had assumed (*Assumption b*) that the sales of Bitmain devices are equally distributed among all models (including the old Antminer S9 and newer S11, etc.). This assumption is not justified by any empirical evidence

or rationalization. Similarly, for Q3 in 2019, the author applied *Assumptions e and f*. *Assumption e* establishes an arbitrary distribution of sales between the three producers; this distribution is not backed by any empirical evidence. *Assumption f* assumes that the ratio of sales between two ASIC devices (including S9) remains the same as H1 of 2019 and is equally spread between available models.

(d) Random distribution of hardware in the pool:
- Another type of error in constructing the hardware pool is a random distribution of hardware shares. For instance, Mora et al. [12] assumed a random assignment of mining hardware that leads to an equal probability that old hardware is used to mine a block as frequently as a newer, more efficient one, leading to an inflated electricity consumption figure.

2. **QM2:** Geographic data

(a) Using mining pool location:
- Mining pools are widely used in Bitcoin mining, as it is often profitable to mine in a group rather than individually [52]. Participation in a pool can lead to an overall improvement in the expected return from the mining operation. Studies such as Mora et al. [12] used the IP address of a mining pool to assign a geographic location to a miner. This approach has been questioned extensively in the literature, as anyone can join a mining pool irrespective of their geographic location [10]. This approach was also used by Onat et al. [53], where the study attributes the central server of a pool as the sole geographical location where all mining takes place. This is considered provably inaccurate [10].

(b) Extrapolation of data:
- Using the Cambridge dataset: The University of Cambridge, through their annual surveys, has been able to map 32% to 37% of Bitcoins mining power to a specific geographic location [25]. This dataset explicitly states that it only captures 32% to 37% of all the computing power in the network; despite this, De Vries et al. [54] used anecdotal examples to justify their reliance on the extrapolation of the Cambridge dataset to represent the whole network. The two non-academic literature examples used do not provide scientific evidence of actual distribution; thus, it is safe to assume that the validity of this study is considerably limited.
- Using public pools data: In Ref. [37], the authors used the location distribution of the mining pool BTC.com and assumed that it is representative of all Chinese pools, which is not backed by any reason or data. Similarly, the authors also used the location distribution of Slushpool and assumed it to be representative of all European pools. They additionally treated unknown/other pools as Chinese or European pools according to the ratio of Chinese to European pools. All of these assumptions have an impact on the accuracy of their model and their subsequent results.

3. **QM3:** Economics

(a) Cost of electricity:
- Cost assumption without empirical evidence: In their attempt, O'Dwyer and Malone [28] assumed the electricity cost may be considered high (0.10 USD/kWh). The authors reasoned for their choice by selecting the lowest electricity cost in the eurozone. As it is now established that a majority of mining occurs in countries with lower costs [25], this assumption may not hold true. In his work, Digiconomist assumes electricity cost to be 0.05 USD/kWh, however, their choice is not backed by any empirical data. In the University of Cambridge's Bitcoin Electricity Consumption Index, they also chose 0.05 USD/kWh as a reasonable estimate by citing "conversations with experts" as justification however there is no evidence, such as interview transcripts to make the claim traceable.

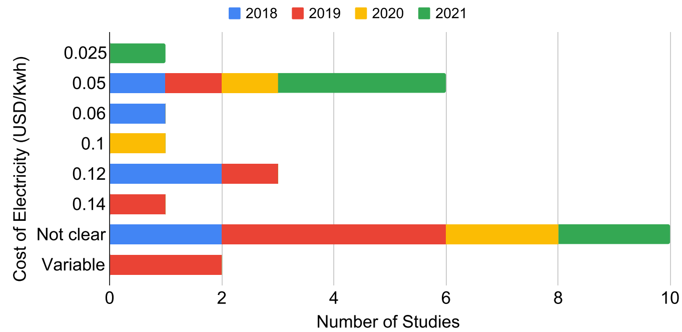

Fig. 5. Choice of the cost of electricity over time.

To demonstrate the variance in the choice of electricity cost, we plot the choice from all the reviewed studies in Fig. 5. Fig. 5 shows that there is a trend toward using 0.05 USD/kWh as the value since 2019. We attribute this trend to the introduction of the Cambridge index and their choice of electricity cost.

(b) Power usage effectiveness (PUE):
- Not considering PUE: A number of works such as Refs. [26,50] excluded PUE consideration from the analysis, likely limiting the validity of their results to just machines used for mining.
- No justification for the selected PUE value: This is surprisingly common in studies that we examined. We believe this to be a significant limitation of the current estimates. The Cambridge Index [25] provides no traceable justification for PUE value selection. They did state that their conversations with the industry suggest a low PUE value. However, they didn't provide any interview transcripts or supporting information. Similarly, De Vries [51] didn't provide any justification for their choice of PUE. In the supporting evidence, the PUE value is based on an assumption (*Assumption i*). This is also an issue with Stoll et al. [37] to a lesser extent, where the authors selected PUE values for small, medium, and large miners based on an interview with one miner. We believe this is very limited as their analysis demonstrates that a change in PUE from 1 to 1.3 results in approximately 26% more electricity consumption.
- Using hardware lifespan to estimate PUE: In De Vries's study [27], the author stated that the electricity consumption only accounts for 60% of all revenue costs. The author attempted to justify their choice of this arbitrary number through the aforementioned assumption of the limited life cycle of 1–2 years. This has been widely criticized both in academic [11] and non-academic literature [9].

(c) Impact of transaction fees and halving on rewards:
- In cryptocurrencies such as Bitcoin, network participants perform computationally expensive operations to receive rewards. However, the reward for mining decreases over time in a process known as halving [38,40]. This has a direct impact on the profitability of performing these computationally expensive operations. The economic model of Bitcoin assumes that the gradual decrement in block reward is compensated for by the increase in transaction fees paid by the users of the system. Most of the surveyed literature does not account for the economics of halving and transaction fees in their model. This is particularly worrisome for models that predict electricity consumption in time horizons involving the halving of the block reward.

4. **QM4:** Social

(a) Hardware lifespan:
- Non-empirical assumption: In De Vries's study [27], the author assumed that the life cycle of Bitcoin hardware is around

1–2 years. The author used Antminer S9 as an illustrative example; however, since the publication of this article in 2018, we have seen continuous use and sale of S9 up until the end of December 2021. This is also evident in IPO data analyzed by Stoll et al. [37]. This is a lifespan closer to 4–5 years[16] as opposed to the 1–2 years estimate given by the author. This work served as the foundation of follow-up work on the carbon emissions of Bitcoin [29], and we believe that the reliance on De Vries's estimate is questionable.

5. **QM5:** Carbon intensity data
   (a) Applying old energy mix and carbon intensity data to current or future predictions:
      - Mora et al. [12] used 2014 carbon intensity data in the 2017 analysis, which led to inaccurate estimates, as pointed out by Masanet et al. [10]. Similarly, De Vries et al. [54] relied on the data for the energy mix from 2019 and applied it to 2021. They did acknowledge that they are using old data due to the lack of availability of new datasets; however, they didn't explain how it impacts the reliability of their results, making it difficult for the reader to assess the reliability of their study. We also note that Onat et al. [53] used data from November 2018, and applied it to all estimates from 2015 up to 2020.
6. **QM6:** Time resolution
   (a) Not reporting time of measurement/prediction/analysis
      - A substantial chunk of the analyzed studies (32%) didn't document the time of their analysis or measurement. The lack of a temporal resolution makes it difficult to understand the reliability of the model. For instance, without explicitly mentioning the time of data collection for Bitcoin mining hardware, it is difficult to understand if the model captures appropriate hardware at the time. This is also true for the case of carbon intensity data, as illustrated above.
   (b) Not documenting the evolution of model parameters
      - Real-time indexes such as CCAF and Digiconomist evolve over time due to the changes in their underlying data or changes in assumptions. Documenting these changes can assist in understanding the evolution of these models; for instance, CCAF provides a change log. However, the Digiconomist index does not seem to have a change log documenting the changes over time.

4.1.3. Quality of remaining research methods

As is evident from our analysis presented above due to the premature nature of this research domain, there is a considerable number of assumptions, each of which may impact the accuracy of the results. Unlike the quantitative models that we have discussed so far, it is relatively easy to assess and improve the rigor of other methodologies due to the presence of vast literature on best practices associated with each of these methodologies. In this subsection, we provide an overview of our analysis. As before, we advise the reader to refer to the supporting evidence for an in-depth analysis of the results. (see Fig. 6)

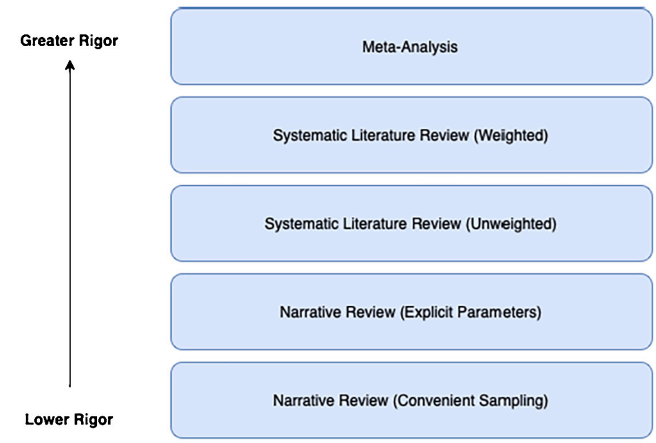

Fig. 6. LR1: Type of review method and associated rigor (Adopted from Refs. [16,55,56]).

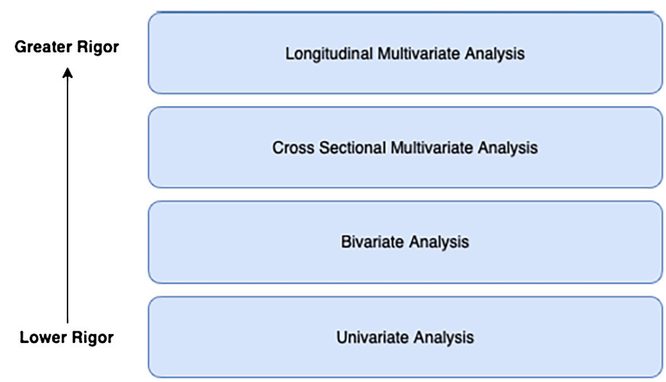

Fig. 7. DA1: Data analysis and statistics methods (Adopted from Refs. [16,57,58]).

1. Literature review: A majority of the reviews (13 out of 16) were narrative reviews that according to Ref. [16] are not as rigorous as other forms of reviews (see also Fig. 7) (LR1). Both systematic literature reviews and meta-reviews were not prominent. Only three of the analyzed studies had documented the search and inclusion criteria, making it possible to replicate the studies and independently estimate the coverage of the review (LR2). We report a similar trend in documenting the search databases and sampling process, as only 3 studies report these (LR3 and LR4). This is a major limitation in literature reviews, as it is difficult to assess the coverage and quality of the analysis without the ability to replicate it independently.
2. Data analysis and statistics: Unlike literature reviews which reported that a majority of the studies employed less rigorous research methods, in data analysis and statistics, a majority of the studies (7 out of 13) employed high-rigor multivariate longitudinal analysis. However, only four studies documented the hypothesis clearly, making it difficult to assess the results of the analysis. Despite the use of more rigorous research methods, these studies suffer a big limitation in acknowledging the practical significance of their results. As an example, Jana et al. [59] attempted to use a deep neural network to predict electronic waste generation from Bitcoin; however, it is not clear if their results are a good representation of the actual Bitcoin ecosystem or how the predictions could be used in a practical context.
3. Case studies: Only a small number of analyzed studies (6) used a case study approach. These studies employed a diverse range of selection mechanisms, illustrating being the most popular. Our analysis suggests that it is difficult to assess the appropriateness of a selected case due to the complicated nature of cryptocurrencies. For instance, Liu et al. [60] selected an influential case of mining in Sichuan and Xinjiang in China to understand the incentives associated with mining there; however, it could be argued that the case selection should have accounted for the seasonality of Bit-

[16] The IPO filling and other sources online tend not to differentiate between the different iterations of the same hardware. For instance, the Antminer S9 sold in 2019 might have a different performance and energy profile than the Antminer S9 sold in 2017. This lack of clarity might undermine the suggested 4–5 years of lifespan.

coin mining[17]. A bigger issue with these studies is the lack of clear boundaries; for example, Gundaboina et al. [61] examined the energy and resource consumption of several hashing algorithms in Dogecoin by using a selective piece of hardware without explicitly describing the bounds of the analysis. In a majority of these analyses (1 out of 6) it is unclear what the dependent and independent variables are.

4. Experiments: Experiments are increasingly popular in proof-of-stake cryptocurrencies, as it is not difficult to generalize the results from these experiments to the whole network, as they mostly rely on off-the-shelf hardware [62]. However, it is still important that the selected hardware for experiments is representative of the network's hardware composition. Out of the four analyzed experiments, only one used a large set of hardware devices to conduct the experiments. For instance, Roma and Hasan [63] performed an analysis of the Ripple network; however, they only used a single piece of hardware with a specific CPU without discussing how representative the hardware is of the whole network. All of these studies do not account for the settings of the experiment, such as geographic variation. This may play a significant role when calculating the real-world performance of these devices. For instance, Li et al. [64] conducted an experiment in an isolated room where they removed operational externalities such as heat, which may not be a good representation of the operational state of mining hardware.

4.2. Limitation of popular approaches

In the previous subsection, we provided a breakdown of common issues in energy consumption studies, highlighting specific instances from the shortlisted studies. In this subsection, we provide a summary of issues for two of the most popular models to contextualize our findings better.

4.2.1. Cambridge bitcoin electricity consumption index (CBECI)

Based on our analysis, we consider CBECI to be one of the more carefully performed analyses of Bitcoin's electricity consumption; however, we would still caution policymakers and researchers from basing their decisions solely on this metric, as CBECI also has known limitations and issues.

The most prominent of these issues is the choice to use nonce analysis methodology by Coin Metrics [65] to improve their hardware pool distribution[18]. CCAF [25] utilized a basket of hardware devices, including two Antminer devices S7 and S9. To estimate the share of hashing power originating from the Antminer S7 and S9 line devices, CCAF [25] relied on the data provided by CoinMetrics [65]. On inspection of the nonce analysis methodology, we note that this methodology overestimates the share of Antminer S7 considerably[19]. Their analysis suggested up to four million active Antminer S7s, which is significantly higher than the reported values from the IPO of Antminer. We also note that the hardware pool used by CCAF [25] does not contain an up-to-date list of all ASIC miners.

Another critique of the CBECI is regarding their choice of PUE value, which is significantly better than Google Data Centers[20]. CBECI backs their choice by claiming that, based on interviews and discussions with experts, this is realistic. However, due to the lack of supporting evidence such as interview transcripts, we are unable to independently verify these claims.

As indicated in QM 6, CBECI does provide a log of changes made to their model; however, they do not document the exact parameters and their specific values used in each iteration of the model. This makes it difficult to clearly understand the evolution of the index over time, as the old predictions made by CBECI may also change retrospectively with newer model changes.

4.2.2. Digiconomist's bitcoin energy consumption index

The approach used by the Bitcoin Energy Consumption Index has been widely debated both in academia [11] and non-academic [9] literature. Our review suggests that the foundational approach used by Digiconomist as outlined in Ref. [27] is of questionable scientific rigor. We specifically question the choices made regarding the life span of the hardware and the subsequent calculation of the 60% ratio of electricity cost in miners' revenue. Unlike the CBECI model, the Digiconomist model does not provide a clear change log of the model evaluation.

We suggest that this work should be considered with caution. As indicated in the earlier section, the follow-up work [54] of Digiconomist suffers from some issues as well, most noticeably the assumptions made by the author regarding the equal distribution of hardware sales.

5. Code of practices

As alluded to in Section 1, one of the main contributions of this review is the preparation of novel research guidelines that could assist in improving the rigor of research in blockchain energy sciences. In the previous section, we highlight some of the common issues in terms of the scientific quality of the reviewed studies.

The primary recommendation is to avoid those pitfalls; however, we acknowledge that this might be difficult to accomplish due to the lack of reliable datasets and difficulty in the acquisition of newer data. In case it is difficult to avoid any of the abovementioned pitfalls, we give some specific recommendations on how the researcher could present their research to avoid giving inaccurate predictions, measurements, or estimates.

In this section, we propose a novel code of practices segmented into three broad categories: basic research design, quantitative energy modeling, and other research recommendations.

5.1. Validation and feasibility

The guidelines presented herein are based on recommendations from the literature and provide practical, actionable insights for both academic and industrial investigations into the energy consumption of blockchain systems. To ensure the validity of these guidelines, the first author conducted a review to assess their alignment with observations from the literature using the quality assessment framework. This alignment was subsequently independently validated by the second author. In instances where misalignment was identified, both authors discussed the results and refined the guidelines accordingly.

An additional important aspect of these guidelines is their practical applicability. To ensure that they could be feasibly implemented, their feasibility was evaluated by an independent external expert, Dr. Alan Ransil of Filecoin Green and Protocol Labs. These guidelines have since been incorporated into the energy model used by Filecoin.

5.2. Basic research design

We have three broad recommendations when designing and executing a blockchain energy study, and these guidelines are intended for both academic and non-academic studies.

[17] It is also worth noting that since the imposition of the Chinese ban on mining activities in China in 2021, it has become increasingly difficult to perform mining in the Sichuan and Xinjiang provinces.

[18] CBECI in their recent update removed the CoinMetrics API discussed here; however, this discussion still highlights that CBECI can also suffer from seemingly trivial issues such as reliance on an unvalidated methodology.

[19] This flaw in the methodology was also flagged by Digiconomist in his blog: https://digiconomist.net.

[20] See https://www.google.com/about/datacenters/efficiency/.

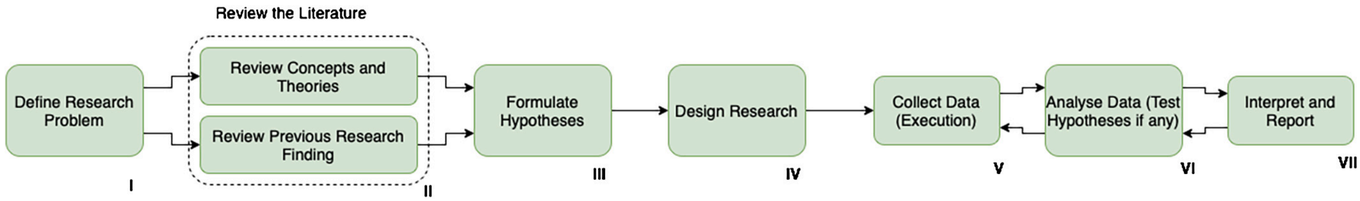

Fig. 8. Research design as proposed by Kothari [34].

5.2.1. *Explicit research methodology*

Any study measuring, estimating, or predicting the energy or environmental footprint of a crypto-asset should be designed to be easily reproduced. To this end, our first recommendation is to explicitly document the methodology employed in the study. We recommend that the research be documented using the 7-step model proposed by Kothari [34], as illustrated in Fig. 8. We provide a brief overview of these 7 steps in the following text; however, it is recommended that the reader refer to the detailed description as provided by Kothari [34].

1. Clear research problem: The research problem addressed by the project must be sufficiently documented in a clear manner. This assists the reader in understanding the project's intended goal without any ambiguity.
2. Review the literature: As discussed in Section 4, due to the lack of coherence in this newly developing space, a lot of research development is done in small silos, resulting in likely slow progress. Thus, we recommend that newer models review the existing literature thoroughly. To that end, our survey provides the reader with an extensive list of relevant academic and non-academic literature. We also document progress made by non-academic source[21].
3. Development of a hypothesis: Ideally, after performing an extensive literature review, the researcher should have a starting point within the literature, this could be an existing model that the researcher intends to build upon or a research methodology that the researcher intends to adopt. Before the execution of their research methodology, it is important to clearly state the working hypothesis or hypotheses. The hypothesis should be as concise as possible to allow testing through the research at the end.
4. Research method: The researcher is expected to succinctly document the steps taken by them to test a hypothesis to address the research problem. This should be documented in the form of a conceptual structure of the research conducted. In the preceding section, we have provided an overview of different research methodologies adopted to answer a specific type of research question. For instance, if the researcher wants to get an overview of the carbon emission of a popular cryptocurrency, they might benefit from first doing a literature review, ideally by following a more rigorous method within literature reviews. If the research question at the end requires data collection[22], it is important to perform the sampling of the data appropriately. For instance, if it is not feasible to collect data on all mining hardware employed in a specific PoW network, an appropriate selection mechanism should be employed rather than randomly sampling a selective set.
5. Data collection: There are many ways of collecting data for energy modeling, and documenting each of these methods is beyond the scope of this review. We refer the reader to Ref. [34] for more instructions. However, broadly speaking, it is important to test the reliability of the data collected either through a primary source or a secondary source. We provide more specific instructions on this for quantitative energy modeling in Section 5.4.
6. Analysis of the data: Analyzing the data is crucial for reporting at this stage, as the analysis can skew the interpretation considerably. For instance, an analysis conducted only in summer might overestimate the overall energy consumption of mining in China, as during rainy seasons, mining moves to hydroelectric power. It is worth noting that this phenomenon likely does not occur anymore due to the ban on mining in China; however, similar patterns might occur in other geographical areas. Similarly, solar-powered plants do not generate electricity during night and windmills only generate power if there is wind. The analysis should account for extremities as well as known limitations as outlined in our Section 4. At this stage, it is also crucial to test any hypothesis proposed in step 3. We provide more specific instructions on improving the analysis and highlighting the limitations based on research methods in the following Sections 5.3 and 5.4.
7. Interpret and report: It is important to contextualize the data generated through the analysis. For instance, the work of De Vries [54] uses a small (34–40%) geographic dataset from Cambridge and presents the results for the whole of the network justifying it through a selective set of qualitative non-academic literature sources. Our recommendation here is that the analysis of the data generated through these models or accumulated through a review should always be compounded with confidence intervals or sensitivity analysis to clarify the impact of assumptions or lack of data on the models' performance.

5.3. *Sharing the data and source code*

We identified accessing the data and source code as one of the main limitations to independently reproducing the results from a large set of studies we analyzed. To account for this limitation, we present two sets of recommendations in line with Stodden [66]:

1. Adhere to the following three long-term goals proposed by Stodden [66]:
   (a) Use a version-control system: The dataset should be version-controlled, this is specifically important for live indexes such as Digiconomist and Cambridge.
   (b) Provide standardized citations for data: In order to build upon existing datasets, the authors should provide standardized citations for the datasets. For instance, the authors in Ref. [67] provided a citable source of the dataset for mining hardware, allowing others to build on top of it.
   (c) Describe data using standardized terminology and ontologies: One of the biggest hurdles in building upon an existing dataset is the way data are reported by different studies. We argue that there is a need for standardization to incentivize improving existing datasets rather than replicating the work again. We sug-

[21] This document is public, and we hope researchers and practitioners will refine and update it as needed. It can be accessed at: https://bit.ly/3csZLg9.
[22] Based on our experience, this is required when conducting a quantitative energy study.

gest adhering to the template used by Cambridge for documenting hardware used for Bitcoin (see http://sha256.cbeci.org).
2. We strongly recommend that if a study is using data from an existing work, the authors supply data or a link to the data to promote transparency and verifiability of their analysis. Ideally, this should be done by using the above-listed guidelines.
3. Hardware distribution and location data should contain collection and validation steps.
4. If location data are used, we recommend that it be assessed for any seasonality patterns that might exist.

Another big limitation, as outlined in Section 4, is the lack of source code. Source code for energy analysis often takes two forms: mathematical models implemented using a programming language or Excel sheets. We provide two guidelines for sharing source code to account for both methods:

1. Excel sheets: The authors should provide details on information quality (IQ) and data quality (DQ) proposed by European Spreadsheet Risks Interest Group O'Beirne [68].
2. Source code: Similar to the long-term recommendations for sharing the data, Stodden [66] suggested that the source code should be version-controlled and should contain test routines that allow independent testing of the source code.

5.4. Quantitative energy modeling

For conducting quantitative energy modeling in a blockchain context, we provide six guidelines in addition to the basic research guidelines proposed above. It is also recommended that the common issues outlined in the last section be avoided.

1. Traceable and verifiable justification for hardware assumptions: One of the biggest issues with quantitative energy modeling studies is the lack of evidence for assumptions made by studies on the hardware. We specifically suggest that the authors:
   (a) State the assumptions within the text and not in supporting material.
   (b) Add sensitivity analysis or confidence intervals when filling in missing data.
2. Traceable and verifiable justification for economic assumption: Similar to assumptions about the hardware in use, in economic models, it is vital that the authors back their assumptions with evidence. We specifically suggest that the authors:
   (a) Should include both capital and operational costs for different agent types (small, medium, and large).
   (b) The cost of electricity should be as granular as possible; if data are missing, a location-based metric should be used.
   (c) The hardware lifespan assumption should be validated using real-world data.
3. Using or collecting geographic data: Geographic data play a crucial role in modeling the environmental footprint of these crypto-assets. We recommend that the authors:
   (a) Should avoid using mining pool IP address, if used, it should be accompanied by sensitivity analysis or appropriate confidence intervals.
   (b) Should avoid using non-academic literature and unvalidated sources for location.
   (c) Should include the date of data collection.
   (d) Should not extrapolate location data, if done, should be accompanied by sensitivity analysis or confidence intervals.
4. The PUE value should be based on empirical evidence. The modeling should also include different types of agents (small, medium, and large) if possible.
5. Avoid unreliable sources of data such as proven faulty studies [12]. Our study recommends referring to our quality assessment results, which include results per reviewed article on www.github.com/ashishrsai/energy. However, we caution readers that not all identified issues necessarily indicate equal flaws in the region of articles. We suggest that readers use our survey's results as a guideline and judge the severity of issues separately. If sources are not already reviewed by our survey, we recommend using our quality assessment framework to judge their reliability through assessment results.
6. The authors should also avoid using improper units of comparison. For instance, Bitcoin and Ethereum do not consume electricity per unit of transaction but per block [5,14,15]. Comparing the per-transaction electricity consumption of Bitcoin or Ethereum may be inaccurate or misleading according to Refs. [5,14,15].
7. The authors should account for the temporal resolution of the data and model parameters. We suggest using time stamps for all the data used in the model and the model itself if it evolves over time.

5.5. Other research methodologies

Similar to quantitative energy modeling, the prime recommendation here is to avoid the common pitfalls outlined in Section 4 and follow the basic research design guidelines proposed above. We have documented our guidelines in more detail on the research repository located at https://github.com/ashishrsai/energy/blob/main/COP.pdf. We strongly recommend that the reader refer to the repository to see the in-depth guidelines for these methodologies. Here, we provide an overview of the specific suggestions for other research methodologies:

1. Literature review: We strongly recommend that the authors adopt more rigorous forms of literature reviews such as meta-review or systematic literature reviews. It is advised that the authors follow the guidelines proposed by Kitchenham and Charters [17] to improve the coverage and reliability of the review.
2. Data analysis and statistics: For data analysis and statistics our prime recommendation is that the authors should avoid oversimplified analysis and account for the practical significance of their results.
3. Case studies: As indicated in Section 4, the selection of the case study should be appropriate. We recommend that the authors follow the guidelines put forth by Sovacool et al. [16] for conducting a case study analysis.
4. Experiments: Our main recommendation for conducting experiments is to choose an appropriately large sample size.

6. Discussion

There is a clear trend that both academic and non-academic researchers have paid increasing attention to the energy and environmental footprint of blockchain technologies in recent years. In academia, this trend is predominantly focused on accurately understanding and predicting the electricity consumption of popular cryptocurrencies such as Bitcoin and Ethereum. There is also a smaller fraction of academic literature that focuses on other smaller cryptocurrencies, such as IOTA and Ripple.

A majority of the academic research in this domain is either purely applied or uses inspired basic research with little to no focus on pure theory development[23] or refinement. This was also reflected through our quality indicator BR2, where we reported that a large chunk (74%) of academic literature does not build upon existing knowledge.

[23] We refer the reader to Sovacool et al. [16] for an in-depth discussion of different types of research in social energy sciences.

Unlike the academic literature, non-academic sources had a more broad focus on different types of cryptocurrencies. However, most of the research in the non-academic sphere is sponsored as we could anticipate. In terms of research design, we notice a trend of more applied research with little to no focus on theory building.

It is also clear that many previous studies have faced problems because they could not access important data from the industry. This lack of data sharing often causes studies to make mistakes and get the energy use estimates wrong. So, just criticizing the studies themselves might not be entirely fair because the industry's secrecy is a big reason for these problems. Our study highlights the pressing need for the industry to share data and cooperate more. This is crucial to making sure that research on blockchain's environmental impact is accurate and trustworthy. By concentrating on making data more transparent, we can get a better grasp of the issues related to digital currencies' energy and environmental impact.

6.1. Current problems and potential for improvement

We believe it is crucial to acknowledge the varying degrees of severity among conceptual errors highlighted in our article. Some errors, such as the discrepancies in interpolating the "energy per transaction" metric or the utilization of flawed time series analysis models, can have significantly larger implications compared to inaccuracies in PUE values or nuanced geographic distributions. While we recognize that certain aspects of works such as Ref. [51], there are indeed publications, such as Ref. [12], with more profound misconceptions that merit closer examination.

We support the efforts to model, estimate, and predict the energy and environmental footprint of these crypto-assets. Our research in this article clearly documents the evolution of the field. We believe this evolution will likely keep spurring more discussion and public debate on this worthwhile topic. However, we also note that the focus on unreliable results may blur or even misguide this discussion.

Our review and analysis make it clear that there are two broad problems:

1. The most prominent limitation is the lack of rigor predominantly caused by a lack of reliable data in many published studies. We have highlighted some specific instances of this throughout Section 4 while providing an abstract overview of the whole field where appropriate. This lack of rigor might suggest that the estimates and predictions on energy consumption and environmental footprint are likely questionable.
2. Our analysis also shows that the current system for publishing scientific/academic literature is lacking in appropriate scrutiny. The publishers, journal editors, program committees of conferences, and reviewers are responsible for ensuring the quality of publications. We show that quality is below what one may expect. However, we also acknowledge that it is always easy to comment when looking back. As a field evolves, prior errors become evidently visible.

Overall, based on our analysis using the quality indicators proposed in Section 3, we report that most of the models used for energy and environmental footprint estimation suffer many known flaws that limit their reliability. We argue that more empirical data has to be collected before these estimates can be considered accurate and scientifically rigorous for policy decisions.

Many of the reviewed studies suffer from trivial issues such as unsubstantiated claims or assumptions. This is uncharacteristic of a mature scientific domain; however, as this field is in the early stage, these measures are likely to see further refinement both in terms of their reliability and accuracy. We identify three potential avenues for improvement in the state of this field: standardization, quality assessment, and provision for more data collection.

1. Standardization: Before we can start meaningfully discussing how to improve the quality of these models, the field must agree on the scope and definitions of these models. For instance, a model designed to estimate the electricity consumption of a PoW-based cryptocurrency should describe all the fundamental building blocks, such as hardware efficiency, PUE value, and cost of electricity, in a transparent manner. The measurements generated by these models should also include a clear timestamp. The data used by these models should also adhere to the nomenclature. We suggest that the field begins with adopting the nomenclature used by the Cambridge dataset at http://sha256.cbeci.org/. This standardization not only helps in comparing the results of different models but also transparently understands the assumptions made by each of these models. It may also promote more collaboration by building upon existing open datasets rather than rebuilding the same datasets again.
2. Quality assessment: This study proposes an initial quality assessment framework in the form of numerous quality indicators. However, we believe this to be only a starting iteration of the framework. We have attempted to logically split the framework based on different research methodologies. This framework can easily be extended to account for different consensus mechanisms. We must be able to transparently discuss the robustness of a given model to make policy choices based on the outcome.
   One of the main recommendations of our model is the use of sensitivity analysis and confidence intervals to communicate results. This assists in ensuring that the results are better contextualized. Another important outcome of our survey is the development of a novel code of practice. We believe that a document structure that adheres to the guidelines proposed in Section 5.1 would allow for an easy evaluation of the scientific quality of the report.
3. Provision for more data collection: One of the prime issues associated with models for electricity and environmental footprint in this domain is the lack of real-world data regarding the machines in use or their geographic location. We believe that this problem needs to be addressed by both private and public stakeholders in this ecosystem. Private mining pool operators should attempt to document and validate their electricity consumption and use of renewable claims. Public bodies should attempt to design regulatory frameworks to either incentivize or impose reporting of electricity use. This can assist in validating the models that we already have in place while also promoting more rigorous reporting.

6.2. Severity of issues

While a large number of studies suffer identified issues, some of the studies, such as Mora et al. [12] made more fundamental and consequential errors in their model construction. Thus, it is important to note that not all the identified issues influence the reliability of the model equally. For instance, studies relying on energy per transaction measure are likely to extrapolate data a lot more than a study with a non-granular PUE value.

We strongly recommend that the reliability of these models should be looked at more granularly and that each identified issue should be independently assessed for its impact on reliability rather than assuming equal influence.

7. Conclusion

In this paper, we conduct a systematic literature review to provide a summary of research done on the energy and environmental footprint of blockchain-based technologies by both academic and non-academic sources. We contextualize our findings by analyzing the robustness of these studies, pointing out common issues while also highlighting potential avenues of either fixing the issues or presenting the results better to account for the limitations.

Given the significant growth of blockchain technologies in recent years and their potential impact on the environment, we believe this study to be pivotal in encouraging a constructive discussion on the reliability of the models used to measure and in some cases offset the carbon emissions generated through the use of blockchain. This refined understanding of the common issues faced by studies focusing on energy and environmental footprints allows us to generate a set of recommendations in the form of a code of practices that may improve the overall quality of these models.

7.1. Contribution

We systematically review the literature from both academic and non-academic sources. This allows us to build a large corpus of models used to measure, estimate or predict the energy, or environmental impact of blockchain technologies. We not only documented these models but also attempted to traceably discuss the potential limitations of these models. We do this by using the quality indicators from cognizant fields of social energy sciences, information systems, and computer science.

Our analysis suggests that a majority of these studies lack the scientific rigor expected from a mature scientific field. We make specific suggestions regarding the reuse of existing theories and datasets to promote more cohesive research and hopefully, iteratively improve these models. To measure or contextualize the quality of the research, we adopt and refine the code of practices from Ref. [16].

Through our review, we have documented the substantial progress made by blockchain energy researchers in obtaining novel datasets and constructing useful models under the constraints of a decentralized system. We strongly support and encourage the development of these models. We believe this article provides an enumerated list of common flaws to avoid while working on the blockchain energy model.

For quantitative energy modeling, we provide a list of common issues and identify instances from popular academic and non-academic sources that suffer these limitations. Our intention here is to highlight that even the most popular studies might be limited in terms of their validity. To this end, we present an overview of some of the issues present in both the Cambridge and Digiconimist models. In addition to the issues of quantitative energy modeling, we also highlight the common problems with other research methodologies and how they can be avoided through adherence to best practices from cognizant disciplines.

The main intention of our work is to promote rigor in blockchain energy sciences. To this end, we develop a set of guidelines in the form of a code of practices that can be used by both academic and non-academic researchers. We believe adherence to this code of practices will not only ensure that common issues and pitfalls are avoided but also help improve the quality of the model and the accompanying report. Our code of practices is also grounded in the needs of the field specifically in terms of building on existing knowledge and better contextualizing results for policy decisions.

To conclude, we strongly believe that this work paves the way for a constructive discussion on the topic of energy and environmental footprint by advocating for a common vocabulary while providing tools to compare different models and understand their reliability. We also believe that the code of practice will promote traceability, building upon existing work and better contextualizing the results. We note that the quality framework developed in this work is an initial framework that is intended to be built upon and iteratively refined as the knowledge around this topic improves.

7.2. Threats to validity

Our analysis despite adherence to the best practices from the literature might suffer from limitations. The primary limitation is the nature of the search for both non-academic and academic literature. The terms used for the search were intentionally kept vague to ensure high coverage; however, they returned a large number of both relevant and irrelevant articles that we then filtered through the title and abstract filtration process. This process might have omitted some relevant articles, as ensuring high reliability of filtration is difficult in a large dataset. To limit this potential issue, we perform cross-validation and obtain a reliable result, suggesting that the filtration process is reliable and may be replicated independently.

The selection of non-academic literature was particularly challenging due to the vastness of the cryptocurrency domain. We restrict our focus to a set of the top 200 cryptocurrencies and use keywords similar to the academic literature. However, this focus on a subset of cryptocurrencies might have led to the omission of some other methodologies. To account for this limitation, we also include all the signatories and supporters of the crypto climate accord. It is also worth noting that despite our inclusion of non-academic studies, we may have omitted some conflicts of interest. It is particularly difficult to both extract and assess potential conflicts of interest in non-academic articles. We acknowledge this as a potential limitation of our work and wish to account for this in potential future work.

Another potential source of limitation to our work is the selection of quality indicators from the work of Sovacool et al. [16]. Due to the fundamentally different nature of crypto climate research, not all the recommendations from Sovacool et al. [16] apply to all the models we reviewed. To account for this, we combine these guidelines with the work of Lei et al. [8]. We additionally propose our quality assessment framework as an initial step in the development of a more robust quality assessment framework. We leave it up to future work to iteratively refine and improve this model.

7.3. Future work

The work presented in this article provides an initial set of quality indicators that could benefit from refinement specifically for non-PoW-type cryptocurrencies. As a next step, we wish to develop a number of flavors of these quality indicators for proof-of-stake and other popular consensus mechanisms.

We also intend to make our code of practice more accessible to researchers and practitioners by developing a web application that provides a checklist and points out common flaws and potential avenues for fixing them. To this end, we have already designed a primitive version of these guidelines[24]; however, we intend to refine it further to improve user experience.

One of the intentions of this work is to allow researchers to revisit their existing models and refine them to avoid any of the issues pointed out in this work. To this end, we wish to refine the model developed and used by Vranken [26] while adhering to our code of practice.

CRediT authorship contribution statement

Ashish Rajendra Sai: Conceptualization, Data curation, Methodology, Visualization, Writing – original draft. **Harald Vranken:** Conceptualization, Investigation, 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.

[24] These guidelines can be viewed at https://github.com/ashishrsai/energy/blob/main/COP.pdf.

Appendix. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.bcra.2023.100169.

References

[1] A.C. Orgerie, M.D. de Assuncao, L. Lefevre, A survey on techniques for improving the energy efficiency of large-scale distributed systems, ACM Comput. Surv. 46 (2014) 1–31, https://doi.org/10.1145/2532637.

[2] Stephens, Emerging power technology: Industry report, Stephens Inc., 2000.

[3] J. Koomey, Turning Numbers into Knowledge: Mastering the Art of Problem Solving, Analytics Press, 2008.

[4] J.G. Koomey, C. Calwell, S. Laitner, J. Thornton, R.E. Brown, J.H. Eto, C. Webber, C. Cullicott, Sorry, wrong number: the use and misuse of numerical facts in analysis and media reporting of energy issues, Annu. Rev. Energy Environ. 27 (1) (2002) 119–158, https://doi.org/10.1146/annurev.energy.27.122001.083458.

[5] J. Sedlmeir, H.U. Buhl, G. Fridgen, R. Keller, The energy consumption of blockchain technology: beyond myth, Bus. Inf. Syst. Eng. 62 (2020) 599–608, https://doi.org/10.1007/s12599-020-00656-x.

[6] Z. Zheng, S. Xie, H.N. Dai, X. Chen, H. Wang, Blockchain challenges and opportunities: a survey, Int. J. Web Grid Serv. 14 (2018) 352–375, https://doi.org/10.1504/IJWGS.2018.095647.

[7] bitcoinvisuals.com, Block reward per block chart, https://bitcoinvisuals.com/chain-block-reward, 2022.

[8] N. Lei, E. Masanet, J. Koomey, Best practices for analyzing the direct energy use of blockchain technology systems: review and policy recommendations, Energy Policy 156 (2021) 112422, https://doi.org/10.1016/j.enpol.2021.112422.

[9] M. Bevand, Serious faults in digiconomist's bitcoin energy consumption index, http://blog.zorinaq.com/serious-faults-in-beci/, 2017.

[10] E. Masanet, A. Shehabi, N. Lei, H. Vranken, J. Koomey, J. Malmodin, Implausible projections overestimate near-term bitcoin co2 emissions, Nat. Clim. Change 9 (9) (2019) 653–654, https://doi.org/10.1038/s41558-018-0321-8.

[11] J. Koomey, Estimating bitcoin electricity use: a beginner's guide, May, Coin Center report, 2019, https://www.coincenter.org/app/uploads/2020/05/estimating-bitcoinelectricity-use.pdf.

[12] C. Mora, R.L. Rollins, K. Taladay, M.B. Kantar, M.K. Chock, M. Shimada, E.C. Franklin, Bitcoin emissions alone could push global warming above 2 °C, Nat. Clim. Change 8 (11) (2018) 931–933, https://doi.org/10.1038/s41558-018-0321-8.

[13] N. Houy, Rational mining limits bitcoin emissions, Nat. Clim. Change 9 (9) (2019) 655.

[14] L. Dittmar, A. Praktiknjo, Could bitcoin emissions push global warming above 2 °C?, Nat. Clim. Change 9 (9) (2019) 656–657, https://doi.org/10.1038/s41558-019-0534-5.

[15] N. Carter, How much energy does bitcoin actually consume?, https://hbr.org/2021/05/how-much-energy-does-bitcoin-actually-consume, 2021. (Accessed 7 April 2023).

[16] B.K. Sovacool, J. Axsen, S. Sorrell, Promoting novelty, rigor, and style in energy social science: towards codes of practice for appropriate methods and research design, Energy Res. Soc. Sci. 45 (2018) 12–42, https://doi.org/10.1016/j.erss.2018.07.007.

[17] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, 2007.

[18] S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system, Decent. Bus. Rev. (2008) 21260, https://bitcoin.org/bitcoin.pdf, 2008.

[19] A. De Vries, C. Stoll, Bitcoin's growing e-waste problem, Resour. Conserv. Recycl. 175 (2021) 105901, https://doi.org/10.1016/j.resconrec.2021.105901.

[20] A.R. Sai, J. Buckley, B. Fitzgerald, A. Le Gear, Taxonomy of centralization in public blockchain systems: a systematic literature review, Inf. Process. Manag. 58 (4) (2021) 102584, https://doi.org/10.1016/j.ipm.2021.102584.

[21] A.R. Sai, J. Buckley, A. Le Gear, Characterizing wealth inequality in cryptocurrencies, Front. Blockchain 4 (2021) 730122, https://doi.org/10.3389/fbloc.2021.730122.

[22] M. Bevand, Electricity consumption of Bitcoin: a market-based and technical analysis, https://blog.zorinaq.com/bitcoin-electricity-consumption/, 2017.

[23] G.A. Brady, N. Kapur, J.L. Summers, H.M. Thompson, A case study and critical assessment in calculating power usage effectiveness for a data centre, Energy Convers. Manag. 76 (2013) 155–161, https://doi.org/10.1016/j.enconman.2013.07.035.

[24] A.R. Sai, A. Le Gear, J. Buckley, Centralization Threat Metric, 2019.

[25] CCAF, Cambridge bitcoin electricity consumption index (cbeci), https://ccaf.io/cbeci/index, 2022.

[26] H. Vranken, Sustainability of bitcoin and blockchains, Curr. Opin. Environ. Sustain. 28 (2017) 1–9, https://doi.org/10.1016/j.cosust.2017.04.011.

[27] A. De Vries, Bitcoin's growing energy problem, Joule 2 (5) (2018) 801–805, https://doi.org/10.1016/j.joule.2018.04.016.

[28] K.J. O'Dwyer, D. Malone, Bitcoin mining and its energy footprint, in: 25th IET Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014), 2014, pp. 280–285, https://doi.org/10.1049/cp.2014.0699.

[29] S. Köhler, M. Pizzol, Life cycle assessment of bitcoin mining, Environ. Sci. Technol. 53 (23) (2019), https://doi.org/10.1021/acs.est.9b05687.

[30] M. Johnson, S. Pingali, Guidance for accounting and reporting electricity use and carbon emission from cryptocurrencies, https://cryptoclimate.org/wp-content/uploads/2021/12/RMI-CIP-CCA-Guidance-Documentation-Dec15.pdf, 2021.

[31] S. Huckle, M. White, Socialism and the blockchain, Future Internet 8 (4) (2016) 49.

[32] J.L. Fleiss, J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educ. Psychol. Meas. 33 (3) (1973) 613–619, https://doi.org/10.1177/001316447303300309.

[33] J. Sim, C.C. Wright, The kappa statistic in reliability studies: use, interpretation, and sample size requirements, Phys. Ther. 85 (3) (2005) 257–268, https://doi.org/10.1093/ptj/85.3.257.

[34] C.R. Kothari, Research Methodology: Methods and Techniques, New Age International, 2004.

[35] B.J. Oates, M. Griffiths, R. McLean, Researching Information Systems and Computing. Sage, 2022.

[36] W.Q. Meeker, L.A. Escobar, F.G. Pascual, Statistical Methods for Reliability Data, John Wiley & Sons, 2022.

[37] C. Stoll, L. Klaaßen, U. Gallersdörfer, The carbon footprint of bitcoin, Joule 3 (2019) 1647–1661.

[38] A.R. Sai, J. Buckley, A. Le Gear, Assessing the security implication of bitcoin exchange rates, Comput. Secur. 86 (2019) 206–222, https://doi.org/10.1016/j.cose.2019.06.007.

[39] A.R. Sai, Towards a Holistic Assessment of Centralization in Distributed Ledgers, Ph.D. thesis, University of Limerick, 2022.

[40] D. Easley, M. O'Hara, S. Basu, From mining to markets: the evolution of bitcoin transaction fees, J. Financ. Econ. 134 (2019) 91–109, https://doi.org/10.1016/j.jfineco.2019.03.004.

[41] Y. Xiao, M. Watson, Guidance on conducting a systematic literature review, J. Plan. Educ. Res. 39 (1) (2019) 93–112, https://doi.org/10.1177/0739456X17723971.

[42] R.E. Kass, B.S. Caffo, M. Davidian, X.L. Meng, B. Yu, N. Reid, Ten simple rules for effective statistical practice, PLoS Comput. Biol. 12 (6) (2016), https://doi.org/10.1371/journal.pcbi.1004961.

[43] E.W. Tsang, Case studies and generalization in information systems research: a critical realist perspective, J. Strateg. Inf. Syst. 23 (2) (2014) 174–186, https://doi.org/10.1016/j.jsis.2013.09.002.

[44] I. Benbasat, D.K. Goldstein, M. Mead, The case research strategy in studies of information systems, MIS Q. (1987) 369–386, https://doi.org/10.2307/248684.

[45] A.R. Dennis, J.S. Valacich, Conducting experimental research in information systems, Commun. Assoc. Inf. Syst. 7 (1) (2001) 5, https://doi.org/10.17705/1CAIS.00705.

[46] T.A. Xu, J. Xu, A short survey on business models of decentralized finance (defi) protocols, arXiv preprint (2022), arXiv:2202.07742, 2022.

[47] Digiconomist, Bitcoin energy consumption index, https://digiconomist.net/bitcoin-energy-consumption/, 2022.

[48] C.B.a.M. Kimmell, The bitcoin mining network, https://coinshares.com/research/bitcoin-mining-network-2022, 2022.

[49] E. Budish, The economic limits of bitcoin and the blockchain, Technical Report, National Bureau of Economic Research, 2018.

[50] U. Gallersdörfer, L. Klaaßen, C. Stoll, Energy consumption of cryptocurrencies beyond bitcoin, Joule 4 (2020) 1843–1846, https://doi.org/10.1016/j.joule.2020.08.005.

[51] A. De Vries, Bitcoin's energy consumption is underestimated: a market dynamics approach, Energy Res. Soc. Sci. 70 (2020) 101721.

[52] Y. Lewenberg, Y. Bachrach, Y. Sompolinsky, A. Zohar, J.S. Rosenschein, Bitcoin mining pools: a cooperative game theoretic analysis, in: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, Citeseer, 2015, pp. 919–927.

[53] N. Onat, R. Jabbar, M. Kucukvar, N. Fetais, Bitcoin and Global Climate Change: Emissions Beyond Borders, 2021, https://doi.org/10.21203/rs.3.rs-564281/v1.

[54] A. De Vries, U. Gallersdörfer, L. Klaaßen, C. Stoll, Revisiting bitcoin's carbon footprint, Joule 6 (2022) 498–502.

[55] K.S. Khan, R. Kunz, J. Kleijnen, G. Antes, Five steps to conducting a systematic review, J. R. Soc. Med. 96 (3) (2003) 118–121, https://doi.org/10.1177/014107680309600304.

[56] G.M. Huebner, M.L. Nicolson, M.J. Fell, H. Kennard, S. Elam, C. Hanmer, C. Johnson, D. Shipworth, Are We Heading Towards a Replicability Crisis in Energy Efficiency Research? A Toolkit for Improving the Quality, Transparency and Replicability of Energy Efficiency Impact Evaluations, UKERC, 2017.

[57] B.G. Tabachnick, L.S. Fidell, J.B. Ullman, Using Multivariate Statistics, vol. 5, Pearson, Boston, MA, 2007.

[58] D.J. Denis, Applied Univariate, Bivariate, and Multivariate Statistics, John Wiley & Sons, 2015.

[59] R.K. Jana, I. Ghosh, M.W. Wallin, Taming energy and electronic waste generation in bitcoin mining: insights from Facebook prophet and deep neural network, Technol. Forecast. Soc. Change 178 (2022) 121584, https://doi.org/10.1016/j.techfore.2022.121584.

[60] Z. Liu, B. Chen, Z. Li, H. Li, D. Wang, Y. Lyu, L. Gao, B. Hou, Bitcoin mining recognition based on community detection with electricity consumption data, in: 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2), IEEE, 2021, pp. 3091–3096, https://doi.org/10.1109/EI252483.2021.9713449.

[61] L. Gundaboina, S. Badotra, S. Tanwar, Energy and resource consumption in cryptocurrency mining: a detailed comparison, in: 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, 2021, pp. 1–5, https://doi.org/10.1109/ICRITO51393.2021.9596387.

[62] M. Platt, J. Scdlmeir, D. Platt, J. Xu, P. Tasca, N. Vadgama, J.I. Ibañez, The energy footprint of blockchain consensus mechanisms beyond proof-of-work, in: 2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C), IEEE, 2021, pp. 1135–1144, https://doi.org/10.1109/QRS-C55045.2021.00168.

[63] C.A. Roma, M.A. Hasan, Energy consumption analysis of xrp validator, in: 2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), IEEE, 2020, pp. 1–3, https://doi.org/10.1109/ICBC48266.2020.9169427.

[64] J. Li, N. Li, J. Peng, H. Cui, Z. Wu, Energy consumption of cryptocurrency mining: a study of electricity consumption in mining cryptocurrencies, Energy 168 (2019) 160–168, https://doi.org/10.1016/j.energy.2018.11.046.

[65] CoinMetrics, Coin metrics' state of the network, https://coinmetrics.substack.com/p/coin-metrics-state-of-the-network-375, 2020.

[66] V.C. Stodden, Reproducible Research: Addressing the Need for Data and Code Sharing in Computational Science, 2010.

[67] M. Zade, J. Myklebost, P. Tzscheutschler, U. Wagner, Is bitcoin the only problem? A scenario model for the power demand of blockchains, Front. Energy Res. 7 (2019) 21, https://doi.org/10.3389/fenrg.2019.00021.

[68] P. O'Beirne, Information and data quality in spreadsheets, arXiv preprint, arXiv:0809.3609, 2008, https://doi.org/10.48550/arXiv.0809.3609.