





Review

# Big Data and Climate Change

Hossein Hassani <sup>1,\*</sup> , Xu Huang <sup>2</sup>  and Emmanuel Silva <sup>3</sup>

<sup>1</sup> Research Institute of Energy Management and Planning, University of Tehran, Tehran 1417466191, Iran

<sup>2</sup> Faculty of Business and Law, De Montfort University, Leicester LE1 9BH, UK; xu.huang@dmu.ac.uk

<sup>3</sup> Fashion Business School, London College of Fashion, University of the Arts London, London WC1V 7EY, UK; e.silva@fashion.arts.ac.uk

\* Correspondence: hassani.stat@gmail.com

Received: 8 January 2019; Accepted: 30 January 2019; Published: 2 February 2019



**Abstract:** Climate science as a data-intensive subject has overwhelmingly affected by the era of big data and relevant technological revolutions. The big successes of big data analytics in diverse areas over the past decade have also prompted the expectation of big data and its efficacy on the big problem—climate change. As an emerging topic, climate change has been at the forefront of the big climate data analytics implementations and exhaustive research have been carried out covering a variety of topics. This paper aims to present an outlook of big data in climate change studies over the recent years by investigating and summarising the current status of big data applications in climate change related studies. It is also expected to serve as a one-stop reference directory for researchers and stakeholders with an overview of this trending subject at a glance, which can be useful in guiding future research and improvements in the exploitation of big climate data.

**Keywords:** big data; data mining techniques; climate change; review

## 1. Introduction

Big data analytics have been rapidly developing along with the emerging needs of big data technologies in numerous subjects (see, for example, [1,2]). The accessibility, availability and exponentially growing quantity of big data have further promoted the corresponding technological advancements and practical implementations. Earth is a complex dynamical system [3]; thereafter, big data analytics encountered more challenges in climate science than other subjects regardless of the extensive resources of big climate data. Climate change as an emerging topic and also a data-intensive subject has been the research focus of big data scholars over the past several decades [4,5]. Exhaustive big data analytics applications have been carried out on big climate data, while the Internet of Things, cloud computing, big data tools to investigate climate, as well as intelligent analytics platforms and new technological progressions, have further emphasized its significance and possible impacts on climate science and big data science development (see, for example, [6,7]).

Given the context of combating climate change, existing research has applied big data analytics in mainly the aspects of energy efficiency, intelligent agriculture, smart urban planning, weather forecast, natural disaster management, etc. Although overall this is not a new subject and there is a large amount of existing literature, there is no recent review to the best of our knowledge that particularly investigates the topic of big data in climate change, not to mention that the novel developments are progressing rapidly everyday along with the technological advancements. Therefore, this paper contributes to the existing literature by providing the most up-to-date overview of big data applications in climate change related studies at a glance

with the most recently published research that reflects the cutting edge of this topic. It is of note that over 80% of the listed applications are after 2016, which makes this review the latest comprehensive review of big data in climate change that is significantly different from the previously existing literature. This paper also contributes by serving as the one stop directory for researchers to gain the most up-to-date overview of this topic. Furthermore, we aim to summarize the popular practice court of research in this domain, and also seek to identify the non-mainstream applications that lack thorough exploration. It is expected that, by providing this comprehensive review, both researchers and practitioners can gain better knowledge of the current research trend and identify the research gaps with valuable potential.

As can be seen in Figure 1, it is identified that the applications of big data in climate change have two fundamental elements: the big climate data resources and the big data analytics techniques. We classify these studies by means of value creation as well as the specific topic of application. For convenient access of applications and clear guiding purposes, it is summarized that big data in climate change mainly function in four aspects of value creation: observing and monitoring, understanding, and predicting and optimizing, whilst the applications are grouped into five topics: energy efficiency and intelligence; smart farming and agriculture and forestry; sustainable urban planning and infrastructure; natural disaster and disease assessment; and other advanced supports.

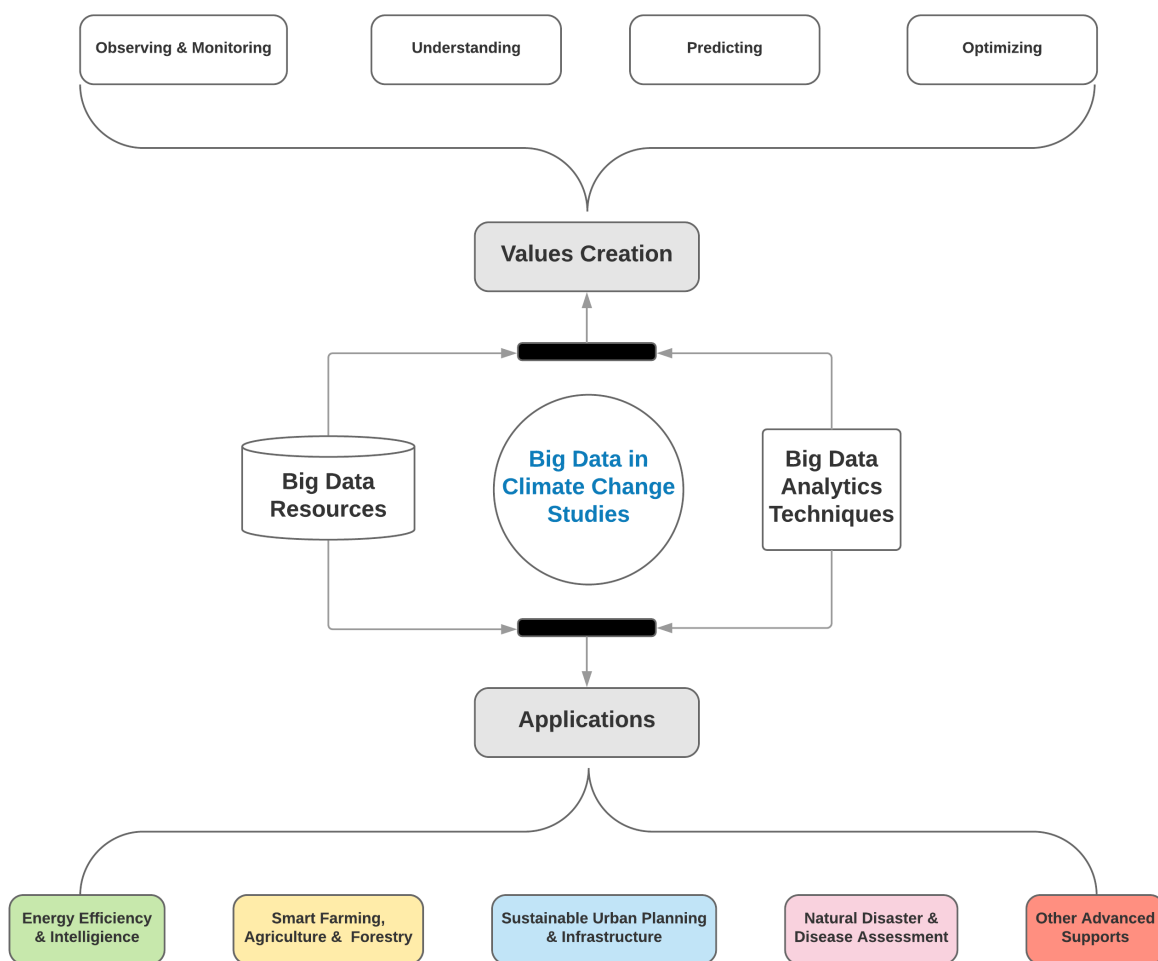


Figure 1. Framework of big data in climate change studies.

The remainder of this review paper is organized such that the values of big data to climate change study are summarized in Section 2 along with the trends of this focused topic. Section 3 lists a detailed review of the big data applications in climate change studies by topics. Finally, the paper concludes in Section 4 with current challenges and directions of future research.

## 2. Values of Big Data to Climate Change Study

### 2.1. Observing and Monitoring

One of the insights big data can bring is thoroughly revealing the realities from the large volume of data recorded. The exceptionally large sources of data contain significantly useful information and is also the fundamental asset of big data analyses. Monitoring the climate system is critical for better understanding the interactions within the system and its drives, respectively. Moreover, it is also beneficial for us to know the changes that may occur due to the global warming [3]. Thereafter, observing and monitoring can be considered as the fundamental value that big data brings along when it is incorporated with climate change study.

In order to obtain thorough observation and comprehensive parameters of climate change, earth observation technology has played a significant part over the past decades [8]. A multi-dimensional big data system has been established and is still promptly developing, which has enabled us to observe and monitor changes on a global scale of diverse earth and climate parameters. According to [3], the climate data have generally four different sources: in situ, remote sensed, model output and paleoclimatic. A more detailed review of the climate data sources and corresponding features can be found in [9]. Later, Guo et al. [8] has thoroughly reviewed the Earth observation big data sources and relevant programs; for instance, the satellites that are working in climate change research and their functions, the remotely sensed oceanographic parameters and representative sensors, the essential climate variables by the United Nations Framework Convention on Climate Change, the atmospheric parameters by different international agencies, etc. Specifically, Sun et al. [10] provided a review of the global precipitation data sets regarding its sources and comparisons.

### 2.2. Understanding, Predicting and Optimizing

In the context of the nature of climate science, which investigates the tremendous global scale changes of various observations/parameters, its 3Hs feature (high dimension, high complexity and high uncertainty) has made it a great playground for big data researchers to explore and analyse even before proceeding to data mining. Besides the building up the multidimensional system of collecting and monitoring climate change, big data has also promoted the rapid progression of data-intensive analytics in climate change related studies. Here, we briefly categorizes them into the aspects of understanding, predicting and optimizing.

Understanding the big data (or data empathy) according to [3] is a challenging task considering its 5Vs feature (volume, variety, velocity, veracity and value), revealing the hidden valuable information from big data requires adequate knowledge of the purposes of corresponding data as well as the techniques/methods for collecting the data. As a trending and emerging topic, big data researchers who are also interested in climate sciences have been exposed to abundant established resources, for instance, the Global Climate Observing System (GCOS), Earth System Grid Federation ([esgf.llnl.gov](http://esgf.llnl.gov)), the National Center for Atmospheric Research ([ncar.ucar.edu](http://ncar.ucar.edu)), United Nations Global Pulse ([unglobalpulse.org](http://unglobalpulse.org)), the Climate Data Guide ([climatedataguide.ucar.edu](http://climatedataguide.ucar.edu)), NASA Global Climate Change ([climate.nasa.gov](http://climate.nasa.gov)), the NASA Center for Climate Simulation ([nccs.nasa.gov](http://nccs.nasa.gov)) and many other international and national climate

monitoring and analysing institutions over the world. A detailed report that introduces the core of global scale climate research and cyber-infrastructure can be found in [11].

The abundant resources above have enabled us to gain knowledge of what the big climate data are how the data are collected, and what the data can be used for. However, these barely scratch the surface of big data analytics. Big data have also been playing a vital role in predicting when it is incorporated with climate science, for instance weather forecasting, natural disasters monitoring and early warning, energy consumption forecasting, traffic forecasting, etc. By applying corresponding data mining techniques (the detailed introduction of data mining techniques can be found in [12,13]), it allows knowledge discovery of the potential relationships and causal inferences, which further contribute to the modelling and predicting [14].

Accurate forecasts can aid in adaptive policy making in relation to climate changes, whilst the value creation feature of big data puts emphasis on optimizing. Being able to understand and predict based on sufficient knowledge extracted from big data or drawing inferences across different cases/applications are relatively straightforward. However, optimization requires comprehensive theoretical understanding as well as adequate big data analytics skills to structure the optimal model/infrastructure so as to maximise performance, efficiency and utility, or, in some cases, for achieving sustainable development. In recent years, exhaustive relevant applications have been carried out on energy efficiency management, natural resource management, smart grid, smart farming, etc. More details of applications are provided in Section 3.

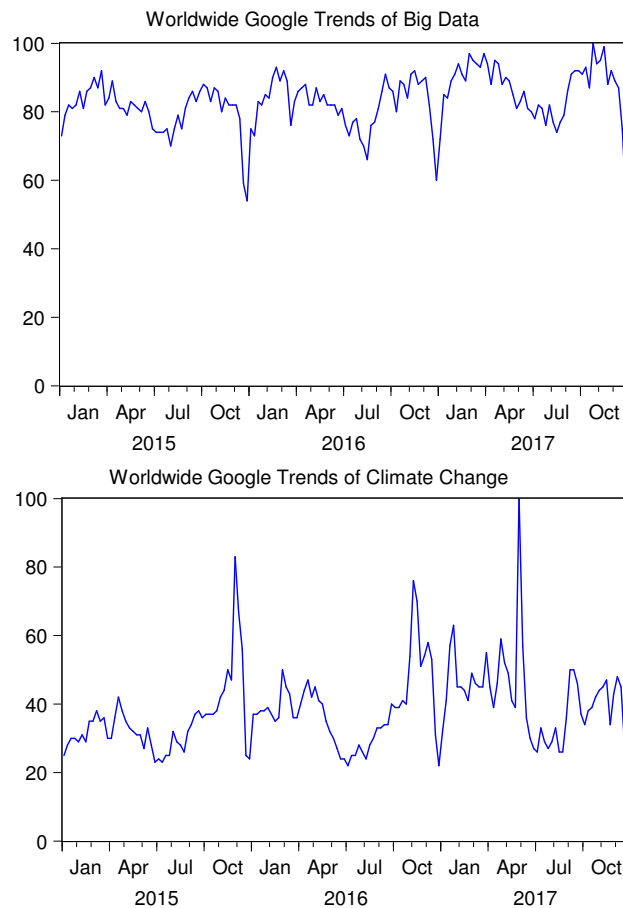
### 2.3. Trends

Google Trends provide the real-time index of interest for keywords that people search worldwide, and it has become a great tool for both academic research and practical implementations in marketing, journalism, entrepreneurship, etc. As can be seen in Figure 2, the worldwide Google Trends of big data have shown the interest index over time, which is generally between 80 to 100 with 100 the peak popularity. It reflects the emerging and trending topic of big data and its relevant terms over the recent years. Although climate change has the average interest index around 40 over time, the rapidly ascending interest from 20 with a wide span over 60 within three years also confirms the increasing popularity and growing attention to climate.

Yaqoob et al. (2016) [15] presented the emerging trend and sources of big data, i.e., Internet of Things [16], multimedia [17] and social media [18,19]. The authors also summarized the state-of-the-art big data processing technologies by categorizing them into the batch based and the stream based processing technologies (more detailed introduction of each technology and a list of big data vendors can be found in [20], a brief summary of the open source tools and initiatives can be seen in [21]). Another tool that is developed by the United States Environmental Protection Agency in 2015 named EnviroAtlas [22] is a web-based ecosystem service framework that brings together environmental, economic and demographic data.

In regard to big data analytics in climate science, some recent developments focus on the introduction of different platforms/services that adopt advanced technological progressions like cloud computing [23], in memory processing, real-time processing, etc. Schnase et al. [24] introduced Modern-Era Retrospective Analysis for Research and Applications (MERRA) analytic services that will enable cloud computing and analytics-as-a-service for climate science, it enables the capability of providing complete life cycle management of large scale scientific repositories, as well as incorporating MapReduce analytics for research and applications targeting a broad range of users. Another platform proposed by Hu et al. [25] named ClimateSpark is an in-memory distributed computing framework that aims to facilitate big climate data analytics and improve computational efficiency. Manogaran et al. [26] proposed the spatial cumulative sum

algorithm to be planted in a scalable big data processing framework for seasonal climate change detection, and the new algorithm has been validated, outperforming the existing change detection approaches. Later, in [27], the authors introduced the In-Mapper based combiner algorithm with higher performance than the widely used MapReduce algorithm in terms of processing big climate data. The proposed algorithm was implemented in seeking correlation between the climate parameters and incidence of dengue by integrating climate and health data.



**Figure 2.** Google Trends index of big data and climate change since 2015.

### 3. Big Data Research of Climate Change by Topics

Besides a few systematical reviews of big data and its techniques, applications and challenges in a broad sense (see [20,28–30]), there have been a number of subject-specific review papers that summarise big data applications in its selected subject (see, for example, crime analysis [13], causality analysis [31], petroleum industry [32], banking [1], blockchain and cryptocurrency, [2], etc.). This section further categorizes big data research in climate change mainly into five topics: energy efficiency and intelligence; smart farming, agriculture and forestry; sustainable urban planning and infrastructure; natural disaster and disease assessment; and other advanced supports. For each topic, a selective review of applications and research of recent years are presented in chronological order.

### 3.1. Energy Efficiency and Intelligence

One of the primary roles of big data analytics in climate change is managing and utilizing the resources to fight global warming. There have been exhaustive applications focusing on improving energy efficiency and building up intelligent energy networks.

Zhou et al. [33] provided a review of smart energy management incorporating big data analytics, in which the authors presented the industrial infrastructure and resources for smart energy management. In the context of opportunities coming along with big data analytics in electric utilities, the relevant issues and opportunities are discussed and reviewed in [34] focusing on the policymaking perspective. The review of [35] summarised the data driven approaches for predicting building energy consumption at urban level, whilst the data driven technologies and applications are reviewed in regard to the development of intelligent energy network in [36].

The recent reviews of big data applications in smart grids/smart meters can be found in [37–39], in which the authors aim to provide a road map guide by comprehensively summarizing the most up to date applications in this specific subject, introducing the latest big data analytics technologies adopted, and listing the mainstream platforms. The papers that investigate big data applications and technologies of building energy management can be found in [40–42]. Similarly, the review of [43] focused on the prediction and classifications of building energy consumption on the future micro-scale change for a particular building; the review of [44] addressed the uncertainty analysis and big climate data in assessing building energy performance; from the perspective of data mining techniques and applications, a review of research on building operational performance improving is summarized in [45]; the review of [46] emphasises the cyber-physical systems of smart buildings and the diverse data driven approaches for energy efficiency control, particularly focusing on strategies for existing buildings.

Given that the available recent reviews have comprehensively covered the studies before 2017 and some of the research in 2017, here we focus on identifying the latest significant research and intercapillary research that have not been previously reviewed so as to present the most complete and up-to-date overview of this particular subject.

In terms of latest development on big data analytics platform/system, Linder et al. [47] introduce the Big Building Data (BBData) platform for smart buildings using big data technology. It enables observation of the state of room/building and automated implementation of certain rules to enhance comfort while economizing energy and achieving sustainable development. Another smart building energy management platform named the Internet of Things Energy Platform (IoTEP) is proposed in [48]; it is based on FIWARE and oriented to support and ease the analysis of heterogeneous big energy data with the assistance of complex event processing and clustering.

Zakovorotnyi and Seerig [49] presented a practical and semi-automated artificial neural networks clustering method to group the daily energy consumption profiles so to identify typical and untypical behaviour patterns and further improve building energy efficiency. Similarly, to enhance the building energy efficiency and operational performance, Fan and Xiao [50] applied gradual pattern mining to extract gradual rules from the real-world building energy big data.

Ashouri et al. [51] introduced a new methodology of building energy saving advisory systems by adopting data mining framework. Unlike the existing frameworks, this system considers the role of occupants. By extracting the consumption patterns and identifying the correlation between occupant activities and building energy consumption, the system can provide feasible energy saving recommendations for occupants.

To overcome the user confidentiality challenge when outsourcing big metering data to heterogeneous distributed systems, Jiang et al. [52] proposed the privacy-preserving query (P2Q) scheme over encrypted

multidimensional big metering data so as to achieve user privacy in a semi-trusted cloud with fast response and high search quality.

A decision support system for energy efficiency management is developed in [53] that enables the capability of effectively processing cross domain data and high-level energy/climate data management and analytics. The novel high level architecture addressed by the authors integrated the single-entry, processing and semantification point for multiple-sourced data, data storage cluster, data access control, advanced analytics services and dashboard interface service.

A specific focus of the energy efficiency assessment of China industry sectors can be found in [54], where the authors applied k-means clustering and the multi-dimensional association rules algorithm to evaluate the environmental performance and seek possible energy leakages and faults so to further guide regional energy conservation.

In terms of the water energy efficiency management, a systematic review can be found in [55] that summarized the big data analytics applications in rainfall data management, flood risk management, water productivity optimizing, drinking water network management, etc. A recent paper [56] thoroughly investigated the big data challenges and coping strategies in water-energy metering intelligence. Moreover, it demonstrated four different case studies in details as showcase applications of big water-energy data analytics.

### 3.2. Smart Farming, Agriculture and Forestry

Another subject that big climate data has significantly benefited from over the past decade is agriculture. Agriculture is one of the most vulnerable domains when the negative impacts of climate change emerge from the past several decades. Its production heavily depends on the natural resources and agriculture itself is also one of the main sources of Green House Gas (GHG) emissions. The developments of big climate data and its analytics have prompted the widespread use of smart information management system, precision agriculture, as well as intelligent automatic agriculture. These implementations all contribute to the climate-smart aims of optimizing production with minimum costs and GHG emissions. Big data applications in agro-environmental science are summarized in [57] with analyses of theoretical framework and instructions of three different case studies so as to further address the limitations and challenges. A review of big data applications in smart farming can be found in [58]. Later, Kamilaris et al. [59] presented the review of big data analytics practices in agriculture, followed by the authors' recent review of deep learning applications in agriculture [60]. Pham and Stack [61] discussed the rapid development of precision agriculture and its impacts on both micro-level form capabilities and macro-level industry competitiveness, while an Internet of Things platform is designed for precision agriculture in [62]. These existing review papers aimed to present the state-of-the-art big data analytics technologies, frameworks and challenges in agriculture and farming related practices in or before 2017. Thereafter, here we focus on only the latest applications and developments in this particular subject for filling the gap and providing the most up-to-date overview.

Some recent development focus on the forest ecosystem. Franklin et al. [63] adopted four kinds of big data for the prediction of global climate change impacts on the plant populations and communities. Although the availability of data contributes to the ecosystem and climate change studies, the limitations of spatial representation, plot data, heterogeneity issues, and lacking of data standards still need to be improved by future research. It is also discovered that an emerging research trend of big data analytics in forestry resource focus on China [64]. The province level forestry resource efficiency in China is investigated in [65] by data envelopment analysis model and Malmquist total factor productivity index. The information extracted from the big data of numerous economics, social and ecological factors due to the widespread forest landscape of China has offered valuable recommendations for sustainable development.

Li et al. [66] presented a case study of big data analytics application in forest ecosystem service at Anhui province, China. Another case study in China was proposed in [67], which focuses on the ecosystem service in a karst region—Guangxi province of China—and identifies the positive influence of ecological big data analytics in ecosystem services. Xie et al. [68] focused on the influencing factors of ecological land change in China's Beijing–Tianjin–Hebei region. With the big climate/ecosystem data, the authors discovered substantial differences across factors regarding its impacts on ecological land changes in order to recommend on better strategies of natural resource management. A related research that also targeted China in [69], which introduced the big data analytics with provincial multisourced data on determining the optimal farmland conversion to reach the best balance between farmland protection and economic growth.

### 3.3. Sustainable Urban Planning and Infrastructure

Apart from the ecosystem and forestry resource management, the increasing greenhouse gas emission has brought attention to the urban planning and infrastructure in a sustainable manner. For instance, Hughes [70] analyzed the urban adaptation planning in the US in response to climate change. In terms of the role of big data analytics in achieving sustainable urban planning and infrastructure that also links to climate change, recent research mainly addresses the urban informatics, smart/green city and smart transportation (note that smart grid related research is summarized within the energy efficiency and intelligence topic).

In [71], Hashem et al. have systematically demonstrated the role of big data analytics in forming and supporting smart cities along with the review of important relevant literature. The authors have proposed the model of big smart city data management and evaluated the technological architecture and challenges. A few generic reviews or the discussion of big data in smart cities can be found in [72–75], while the role of sustainability in smart cities is exploited in [76,77].

A recent comprehensive review by Bibri and Krogstie [78] has summarized existing significant interdisciplinary literature regarding smart sustainable cities. Later, Bibri [79] identified the big data analytics applications enabled by Internet of Things for achieving environmental sustainability, where the author also summarized the state-of-the-art data processing platforms and associated big data analytics technologies. Another similar piece of research that introduced the cloud-based green Internet of Things architecture can be found in [80]. The most recent discussion regarding big data and smart cities by Lim et al. [81] seeks to identify the classification and challenges of big data use cases in smart cities.

In terms of technological developments, Zhu et al. [82] demonstrated the latest progression regarding big data and sensor-cloud, and then introduced three different types of sensor-clouds for the big data analytics architecture of green cities. To achieve reduced dimensionality while maintaining a high level of long-term average transmission reward, Huang et al. [83] introduced the threshold-based scheduling policy for the energy harvesting sensor networks used in green cities.

Although the intelligent transportation system associated with big data analytics has been well exploited for the past several decades, the sustainable aspect of application is still in its early development stage. In a recent paper, the bike sharing scheme in Shanghai is evaluated in [84] by big data analytics and the environmental benefits from a spatiotemporal perspective are identified.

### 3.4. Natural Disaster and Disease Assessment

Over the past decade, there have been debates regarding the relationships between climate change and natural disasters and disease [85–87]. The climate change certainly increased the vulnerability and levels of risk with weather extremes and disease spread, and the natural disaster/disease assessment and management play an important role in climate science. Given the big data and climate change context, it



is also vital to gain the most up-to-date knowledge of big data analytics applications in natural disaster and/or disease management.

There are existing reviews on data mining applications in terms of combating natural disasters [88,89]—for instance, floods, storms, landslides, volcanic eruptions, tsunamis, earthquakes, etc., in which the authors summarized the relevant big data analytics applications in predicting, detecting and management strategy improvement for a variety of natural disasters, along with identifying existing challenges and future research directions. A review that specially targeted climate prediction applications in China can be found in [90].

With regard to computational development, Miyoshi [91] reported the encouraging performance of a Japanese flagship K supercomputer on big climate data processing comparing to the system operated at the world's weather prediction centers. The performance reached 120 times more rapid speed and the big data assimilation system is also validated with encouraging performance.

One of the highlights is disaster resilience, and Ofli et al. [92] addressed the importance of aerial imagery data by unmanned aerial vehicles for timely disaster response, and a hybrid crowdsourcing and real-time machine learning framework was proposed to aid in the existing artificial intelligence disaster response platform. In a recent piece of research, Papadopoulos et al. [93] employed big data analytics in multisourced online press and social media big data, in order to validate the disaster resilience in supply chain for sustainability.

Some recent research exploited the big data analytics in disaster assessment. Nguyen et al. [94] applied the deep convolutional neural networks (deep-CNN) on processing post disaster social media images, and the damage assessment showed effective and accurate performance, which also outperformed the existing technique bag-of-visual-words (BoVW). A case study in South Korea [95] predicted the regional level building damage risk under weather extremes or natural disasters using decision tree analysis. In regards to flood damage assessment mapping, the case study of the Myanmar flood in [96] emphasised the significance of global open geo-information, while later Cian et al. [97] adopted the earth observation big data and proposed the normalized difference flood index for flood mapping. Given the significant amount of emissions caused by fire disturbance on a global scale, a recent research work by Ramo et al. [98] investigated the global burned area mapping using classification data mining techniques.

It has been systematically reviewed in [99] of the big data analytics in global infectious disease surveillance, whilst its corresponding big data applications related to climate science still stay in their early stage of development. Traore et al. [100] applied data mining techniques on satellite imagery data for identifying risk areas that are exposed to epidemic crisis. Similarly, Manogaran and Lopez [101] focused on dengue and proposed a big climate data based surveillance system for continuous monitoring and timely warning.

### 3.5. Other Advanced Supports

Apart from the above domain topics of big data analytics in climate change studies, here we also seek to identify significant relevant research that emphasize other non-mainstream subjects, yet bring valuable insights on the role of big data in climate science.

Zhang and Chen [102] studied the prefecture-level big data of the Poyang Lake Eco-Economic Zone in China and offered local experience oriented recommendations to the national strategy given the aim of sustainable development.

Due to the increasing desire of enhancing sustainable competitive advantages, product lifecycle management is also playing a significant role in combating climate change. A big data driven product lifecycle management framework was introduced in [103] to assist with the decision-making of every stage of the product lifecycle covering designing, manufacturing, maintenance, disassembling, recycling,

disposing, etc. The authors demonstrated the framework in detail and also validated the framework with real case studies showing significant improvement of material efficiency, energy efficiency, emission reduction, customer service improvement and economic benefits growth. Later, Liang et al. [104] adopted the Fruit Fly Optimisation and proposed the novel cyber physical system for machining optimisation during the manufacturing lifecycle. Another framework named the Intelligent Immune System is introduced in [105]. It was also designed for improving productivity and energy saving, employed artificial neural networks for identifying abnormal patterns and planted the re-scheduling algorithm for follow-up processing.

A special focus on the electrical and electronic equipment waste/recycling management with big data analytics is addressed in [106]. The authors proposed and illustrated the waste recovery/recycling framework based on the Internet of Things and big data technologies, and also discussed its economical and technological challenges of implementation.

### 3.6. Key Techniques for Big Data in Climate Change

The above sections have presented a selective review of applications and research in recent years, which are categorized into five main topics. In this section, from a technical perspective, we summarize the mainstream techniques for big data in climate change studies regardless of the topics of applications. It is of note that these techniques are not exclusive just for big data in climate change studies, but well established and widespread for applications in big data analyses on a broader range of subjects.

According to [107], big data is not a business solution but the underlying technology, where models are discovered through the algorithmic search process of data exploration [108]. There are a variety of big data tools and the most popular ones are summarized by [21], which are Apache Hadoop and related projects, S4, Storm, Mahout, R, Massive On-line Analysis (MOA), Waikato Environment for Knowledge Analysis (WEKA), and Vowpal Wabbit. Each tool/platform has its own merits and their advancements are rather rapidly fast along with the evolutions of corresponding infrastructures, but the underlying techniques for big data analysis are mainly grouped into four categories: clustering, classification, regression and association rule mining [20]. Moreover, it is also summarized by Wu et al. that the ten most influential data mining algorithms are C4.5, k-mean, SVM (support vector machine), Apriori, EM (expectation maximization), PageRank, AdaBoost, kNN (k-nearest neighbors), Navie Bayes and Classification and Regression Trees (CART), for which they have been incorporated in almost all commercial and open source big data analysis systems/platforms [12]. To serve as a one stop directory guide for readers who are interested, we also briefly summarize these widely applied techniques as follows (the overall structure can be found in Figure 3). It is also necessary to address that the advancements of those techniques and new algorithms are also developing on an almost daily basis along with today's rapid technological progress.

Clustering has the key objective of grouping similar or closely related data objects together through data exploration so that different objects are separated [31,109]. Its underlying fundamental concept is distance measure and it includes a variety of approaches: hierarchical clustering, k-means (partitioning clustering), high dimensional methods, density based clustering, co-occurrence based clustering and other evolutionary methods.

Classification, on the other hand, categorizes data objects into the predefined groups and it is one of the most fundamental big data analysis techniques with a collection of well established methods [110]. One of the most well known methods is decision tree, which applies a series of crafted questions so to achieve the classification tasks for attributes [111]. Support Vector Machines (SVM) [112] divided the objects into two classes by using an optimal separating hyperplane that was designed to assure the classification error is minimized. Another classifier named the Naive Bayes Rule applies Bayes Rule to calculate the probability of each class [113]. This method is based on the assumption that all attributes are

independent, and the prediction class can be determined for the one with the highest posterior probability. Neural Networks are inspired by the structure and mechanism of real neurons in gathering and processing information, and it enables the estimation of the posterior probabilities for completing the classification task [114]. The kNN approach seeks to identify the nearest neighbours of the observation object so to determine its class label [115].

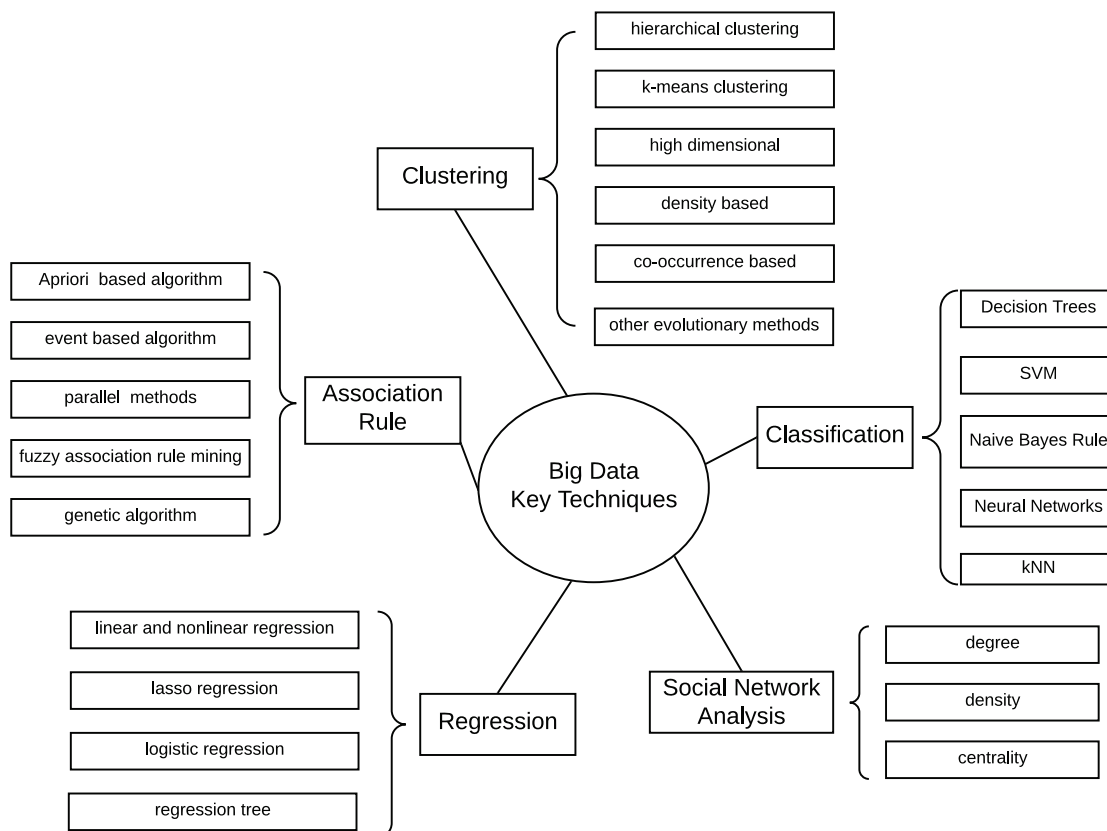


Figure 3. The structure of big data key techniques.

Association Rule Mining was firstly proposed in [116] on the supermarket data set and aimed to investigate into co-occurrences among data objects. According to [117], it is a technique for identifying the simultaneous occurrence that happens more frequently than the average co-occurrence frequency existing in the data set.

Regression is also considered an important big data analysis technique due to its capabilities in dimension reduction, information extraction, estimation and prediction. The fundamental concept of regression is to investigate the relationship between two or more variables so to assist forecasting and decision-making. There are a variety of well established regression techniques like linear and nonlinear regression, lasso regression, logistic regression and regression tree, etc.

Social Network Analysis (SNA) is based on the principle of graph theory and it is a relatively new big data analysis technique. It investigates the connections and contents among objects in a massive stack of information in order to structure a social network. The basic social network is formed by nodes and its related nodes connected by links (namely edges) [118,119]. The most frequently used measurement techniques to analyze the patterns consisted of the social network including degree, density and centrality.

#### 4. Conclusions and Future Research

Researchers have generally realized the big impact of data intensive research. Guo et al. [120] have summarized the milestones of Big Earth Data development and the big challenges that follow. In this paper, we present the most up-to-date overview of big data analytics in climate change covering over 100 research papers. It is observed that energy efficiency remains the most popular “practice court” followed by smart farming/agriculture and natural disaster assessment. Although there is a lot of existing literature on big data and smart cities, the sustainable aspect that links to climate change still stays in the early stage of its development. The topic of disease management also lacks thorough exploration. Moreover, there are some non-mainstream applications discovered in Section 3.5 that covers regional sustainable development strategy recommendation, product lifecycle management, and electronic waste management.

It is noted that the focus of big data analytics application in climate change seems to be unbalanced, and topics like waste/recycle management with valuable potential are neglected. The reasons may be a lack of corresponding data resources, a lack of economical profit and research funding [121], poor communication between groups of researchers with different skills, uncertainty information [122], etc. This review also identifies a few mainstream limitations of the current research: the uniformity and uncertainty of big climate data, the capability and efficiency of learning/analyzing big data compared to the speed of continuous data explosion, the dilemma of time, spatial and scale complexity, etc. (more details can be found in [3,123,124]).

One of the research trends identified among the recent applications is cloud computing, which provides a better solution for big data storage, transmitting and computational requirements [125]. It is also worth noting the popularity of integrating the Internet of Things [126], the architecture of a platform that enables efficient, real-time, in memory cloud computing and storage of complex big data with the most advanced analytics or data mining techniques, as well as integrating the Internet of Things is the key to future research.

**Author Contributions:** All authors have equally contributed substantially to the work reported.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Hassani, H.; Huang, X.; Silva, E. Digitalisation and Big Data Mining in Banking. *Big Data Cogn. Comput.* **2018**, *2*, 18. [CrossRef]
2. Hassani, H.; Huang, X.; Silva, E. Big-Crypto: Big Data, Blockchain and Cryptocurrency. *Big Data Cogn. Comput.* **2018**, *2*, 34. [CrossRef]
3. Faghmous, J.H.; Kumar, V. A big data guide to understanding climate change: The case for theory-guided data science. *Big Data* **2014**, *2*, 155–163. [CrossRef] [PubMed]
4. Götz, M.; Richerzhagen, M.; Bodenstern, C.; Cavallaro, G.; Glock, P.; Riedel, M.; Benediktsson, J.A. On scalable data mining techniques for earth science. *Procedia Comput. Sci.* **2015**, *51*, 2188–2197. [CrossRef]
5. Wu, J.; Guo, S.; Li, J.; Zeng, D. Big data meet green challenges: Greening big data. *IEEE Syst. J.* **2016**, *10*, 873–887. [CrossRef]
6. Almgren, K.; Alshahrani, S.; Lee, J. Weather Data Analysis using Hadoop to Mitigate Event Planning Disasters. Available online: <https://scholarworks.bridgport.edu/xmlui/handle/123456789/1105> (accessed on 1 February 2019).
7. Oury, D.T.M.; Singh, A. Data Analysis of Weather Data Using Hadoop Technology. In *Smart Computing and Informatics*; Springer: Singapore, 2018; pp. 723–730.
8. Guo, H.D.; Zhang, L.; Zhu, L.W. Earth observation big data for climate change research. *Adv. Clim. Chang. Res.* **2015**, *6*, 108–117. [CrossRef]

9. Faghmous, J.H.; Kumar, V. Spatio-temporal data mining for climate data: Advances, challenges, and opportunities. In *Data Mining and Knowledge Discovery for Big Data*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 83–116.
10. Sun, Q.; Miao, C.; Duan, Q.; Ashouri, H.; Sorooshian, S.; Hsu, K.L. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.* **2018**, *56*, 79–107. [[CrossRef](#)]
11. Schnase, J.L.; Lee, T.J.; Mattmann, C.A.; Lynnes, C.S.; Cinquini, L.; Ramirez, P.M.; Webster, W.P. Big data challenges in climate science: Improving the next-generation cyberinfrastructure. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 10–22. [[CrossRef](#)]
12. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2016.
13. Hassani, H.; Huang, X.; Silva, E.S.; Ghodsi, M. A review of data mining applications in crime. *Stat. Anal. Data Min. ASA Data Sci. J.* **2016**, *9*, 139–154. [[CrossRef](#)]
14. Shi, G.R.; Yang, X.S. Optimization and data mining for fracture prediction in geosciences. *Procedia Comput. Sci.* **2010**, *1*, 1359–1366. [[CrossRef](#)]
15. Yaqoob, I.; Hashem, I.A.T.; Gani, A.; Mokhtar, S.; Ahmed, E.; Anuar, N.B.; Vasilakos, A.V. Big data: From beginning to future. *Int. J. Inf. Manag.* **2016**, *36*, 1231–1247. [[CrossRef](#)]
16. Ahmed, E.; Yaqoob, I.; Hashem, I.A.T.; Khan, I.; Ahmed, A.I.A.; Imran, M.; Vasilakos, A.V. The role of big data analytics in Internet of Things. *Comput. Netw.* **2017**, *129*, 459–471. [[CrossRef](#)]
17. Zhu, W.; Cui, P.; Wang, Z.; Hua, G. Multimedia big data computing. *IEEE Multimed.* **2015**, *22*, 96–c3. [[CrossRef](#)]
18. Fan, W.; Gordon, M.D. The power of social media analytics. *Commun. ACM* **2014**, *57*, 74–81. [[CrossRef](#)]
19. Khan, N.; Yaqoob, I.; Hashem, I.A.T.; Inayat, Z.; Ali, M.; Kamaleldin, W.; Alam, M.; Shiraz, M.; Gani, A. Big data: Survey, technologies, opportunities, and challenges. *Sci. World J.* **2014**. [[CrossRef](#)] [[PubMed](#)]
20. Chen, C.P.; Zhang, C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf. Sci.* **2014**, *275*, 314–347. [[CrossRef](#)]
21. Fan, W.; Bifet, A. Mining big data: Current status, and forecast to the future. *ACM SIGKDD Explor. Newsl.* **2013**, *14*, 1–5. [[CrossRef](#)]
22. Pickard, B.R.; Daniel, J.; Mehaffey, M.; Jackson, L.E.; Neale, A. EnviroAtlas: A new geospatial tool to foster ecosystem services science and resource management. *Ecosyst. Serv.* **2015**, *14*, 45–55. [[CrossRef](#)]
23. Yang, C.; Yu, M.; Hu, F.; Jiang, Y.; Li, Y. Utilizing cloud computing to address big geospatial data challenges. *Comput. Environ. Urban Syst.* **2017**, *61*, 120–128. [[CrossRef](#)]
24. Schnase, J.L.; Duffy, D.Q.; Tamkin, G.S.; Nadeau, D.; Thompson, J.H.; Grieg, C.M.; McNerney, M.A.; Webster, W.P. MERRA analytic services: Meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. *Comput. Environ. Urban Syst.* **2017**, *61*, 198–211. [[CrossRef](#)]
25. Hu, F.; Yang, C.P.; Duffy, D.; Schnase, J.L.; Li, Z. ClimateSpark: An In-memory Distributed Computing Framework for Big Climate Data Analytics. *Comput. Geosci.* **2018**, *115*, 154–166. [[CrossRef](#)]
26. Manogaran, G.; Lopez, D. Spatial cumulative sum algorithm with big data analytics for climate change detection. *Comput. Electr. Eng.* **2018**, *65*, 207–221. [[CrossRef](#)]
27. Manogaran, G.; Lopez, D.; Chilamkurti, N. In-Mapper combiner based MapReduce algorithm for processing of big climate data. *Future Gener. Comput. Syst.* **2018**, *86*, 433–445. [[CrossRef](#)]
28. Wu, X.; Zhu, X.; Wu, G.Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 97–107.
29. Chen, M.; Mao, S.; Liu, Y. Big data: A survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [[CrossRef](#)]
30. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [[CrossRef](#)]
31. Hassani, H.; Huang, X.; Ghodsi, M. Big Data and Causality. *Ann. Data Sci.* **2017**, *5*, 1–24. [[CrossRef](#)]
32. Hassani, H.; Silva, E.S. Big Data: A big opportunity for the petroleum and petrochemical industry. *OPEC Energy Rev.* **2018**, *42*, 74–89. [[CrossRef](#)]
33. Zhou, K.; Fu, C.; Yang, S. Big data driven smart energy management: From big data to big insights. *Renew. Sustain. Energy Rev.* **2016**, *56*, 215–225. [[CrossRef](#)]

34. Schuelke-Leech, B.A.; Barry, B.; Muratori, M.; Yurkovich, B.J. Big Data issues and opportunities for electric utilities. *Renew. Sustain. Energy Rev.* **2015**, *52*, 937–947. [[CrossRef](#)]
35. Tardioli, G.; Kerrigan, R.; Oates, M.; James, O.D.; Finn, D. Data driven approaches for prediction of building energy consumption at urban level. *Energy Procedia* **2015**, *78*, 3378–3383. [[CrossRef](#)]
36. Ma, Z.; Xie, J.; Li, H.; Sun, Q.; Si, Z.; Zhang, J.; Guo, J. The role of data analysis in the development of intelligent energy networks. *IEEE Netw.* **2017**, *31*, 88–95. [[CrossRef](#)]
37. Tu, C.; He, X.; Shuai, Z.; Jiang, F. Big data issues in smart grid—A review. *Renew. Sustain. Energy Rev.* **2017**, *79*, 1099–1107. [[CrossRef](#)]
38. Reka, S.S.; Dragicevic, T. Future effectual role of energy delivery: A comprehensive review of Internet of Things and smart grid. *Renew. Sustain. Energy Rev.* **2018**, *91*, 90–108. [[CrossRef](#)]
39. Wen, L.; Zhou, K.; Yang, S.; Li, L. Compression of smart meter big data: A survey. *Renew. Sustain. Energy Rev.* **2018**, *91*, 59–69. [[CrossRef](#)]
40. Koseleva, N.; Ropaita, G. Big data in building energy efficiency: Understanding of big data and main challenges. *Procedia Eng.* **2017**, *172*, 544–549. [[CrossRef](#)]
41. Molina-Solana, M.; Ros, M.; Ruiz, M.D.; Gómez-Romero, J.; Martín-Bautista, M.J. Data science for building energy management: A review. *Renew. Sustain. Energy Rev.* **2017**, *70*, 598–609. [[CrossRef](#)]
42. Fan, C.; Xiao, F.; Li, Z.; Wang, J. Unsupervised Data Analytics in Mining Big Building Operational Data for Energy Efficiency Enhancement: A Review. *Energy Build.* **2018**, *159*, 296–308. [[CrossRef](#)]
43. Wei, Y.; Zhang, X.; Shi, Y.; Xia, L.; Pan, S.; Wu, J.; Han, M.; Zhao, X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1027–1047. [[CrossRef](#)]
44. Tian, W.; Heo, Y.; De Wilde, P.; Li, Z.; Yan, D.; Park, C.S.; Feng, X.; Augenbroe, G. A review of uncertainty analysis in building energy assessment. *Renew. Sustain. Energy Rev.* **2018**, *93*, 285–301. [[CrossRef](#)]
45. Fan, C.; Xiao, F.; Yan, C. Research and Applications of Data Mining Techniques for Improving Building Operational Performance. *Curr. Sustain. Renew. Energy Rep.* **2018**, *5*, 181–188. [[CrossRef](#)]
46. Schmidt, M.; Ahlund, C. Smart buildings as Cyber-Physical Systems: Data-driven predictive control strategies for energy efficiency. *Renew. Sustain. Energy Rev.* **2018**, *90*, 742–756. [[CrossRef](#)]
47. Linder, L.; Vionnet, D.; Bacher, J.P.; Hennebert, J. Big Building Data—A Big Data Platform for Smart Buildings. *Energy Procedia* **2017**, *122*, 589–594. [[CrossRef](#)]
48. Terroso-Saenz, F.; González-Vidal, A.; Ramallo-González, A.P.; Skarmeta, A.F. An open IoT platform for the management and analysis of energy data. *Future Gener. Comput. Syst.* **2019**, *92*, 1066–1079. [[CrossRef](#)]
49. Zakovorotnyi, A.; Seerig, A. Building energy data analysis by clustering measured daily profiles. *Energy Procedia* **2017**, *122*, 583–588. [[CrossRef](#)]
50. Fan, C.; Xiao, F. Mining Gradual Patterns in Big Building Operational Data for Building Energy Efficiency Enhancement. *Energy Procedia* **2017**, *143*, 119–124. [[CrossRef](#)]
51. Ashouri, M.; Haghghat, F.; Fung, B.C.; Lazrak, A.; Yoshino, H. Development of building energy saving advisory: A data mining approach. *Energy Build.* **2018**, *172*, 139–151. [[CrossRef](#)]
52. Jiang, R.; Lu, R.; Choo, K.K.R. Achieving high performance and privacy-preserving query over encrypted multidimensional big metering data. *Future Gener. Comput. Syst.* **2018**, *78*, 392–401. [[CrossRef](#)]
53. Marinakis, V.; Doukas, H.; Tsapelas, J.; Mouzakitis, S.; Sicilia, A.; Madrazo, L.; Sgouridis, S. From big data to smart energy services: An application for intelligent energy management. *Future Gener. Comput. Syst.* **2018**. [[CrossRef](#)]
54. Liu, G.; Yang, J.; Hao, Y.; Zhang, Y. Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering. *J. Clean. Prod.* **2018**, *183*, 304–314. [[CrossRef](#)]
55. Eggimann, S.; Mutzner, L.; Wani, O.; Schneider, M.Y.; Spuhler, D.; Moy de Vitry, M.; Beutler, P.; Maurer, M. The potential of knowing more: A review of data-driven urban water management. *Environ. Sci. Technol.* **2017**, *51*, 2538–2553. [[CrossRef](#)] [[PubMed](#)]
56. Stewart, R.A.; Nguyen, K.; Beal, C.; Zhang, H.; Sahin, O.; Bertone, E.; Vieira, A.S.; Castelletti, A.; Cominola, A.; Giuliani, M.; et al. Integrated intelligent water-energy metering systems and informatics: Visioning a digital multi-utility service provider. *Environ. Model. Softw.* **2018**, *105*, 94–117. [[CrossRef](#)]

57. Lokers, R.; Knapen, R.; Janssen, S.; van Randen, Y.; Jansen, J. Analysis of Big Data technologies for use in agro-environmental science. *Environ. Model. Softw.* **2016**, *84*, 494–504. [[CrossRef](#)]
58. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big data in smart farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [[CrossRef](#)]
59. Kamilaris, A.; Kartakoullis, A.; Prenafeta-Boldu, F.X. A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* **2017**, *143*, 23–37. [[CrossRef](#)]
60. Kamilaris, A.; Prenafeta-Boldu, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
61. Pham, X.; Stack, M. How data analytics is transforming agriculture. *Bus. Horiz.* **2018**, *61*, 125–133. [[CrossRef](#)]
62. Popović, T.; Latinović, N.; Pešić, A.; Zečević, Ž.; Krstajić, B.; Djukanović, S. Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: A case study. *Comput. Electron. Agric.* **2017**, *140*, 255–265. [[CrossRef](#)]
63. Franklin, J.; Serra-Diaz, J.M.; Syphard, A.D.; Regan, H.M. Big data for forecasting the impacts of global change on plant communities. *Glob. Ecol. Biogeogr.* **2017**, *26*, 6–17. [[CrossRef](#)]
64. Song, M.; Cen, L.; Zheng, Z.; Fisher, R.; Liang, X.; Wang, Y.; Huisingh, D. How would big data support societal development and environmental sustainability? Insights and practices. *J. Clean. Prod.* **2017**, *142*, 489–500. [[CrossRef](#)]
65. Li, L.; Hao, T.; Chi, T. Evaluation on China's forestry resources efficiency based on big data. *J. Clean. Prod.* **2017**, *142*, 513–523. [[CrossRef](#)]
66. Li, T.; Cui, Y.; Liu, A. Spatiotemporal dynamic analysis of forest ecosystem services using “big data”: A case study of Anhui province, central-eastern China. *J. Clean. Prod.* **2017**, *142*, 589–599. [[CrossRef](#)]
67. Zhang, M.; Wang, K.; Liu, H.; Zhang, C.; Yue, Y.; Qi, X. Effect of ecological engineering projects on ecosystem services in a karst region: A case study of northwest Guangxi, China. *J. Clean. Prod.* **2018**, *183*, 831–842. [[CrossRef](#)]
68. Xie, H.; He, Y.; Xie, X. Exploring the factors influencing ecological land change for China's Beijing–Tianjin–Hebei region using big data. *J. Clean. Prod.* **2017**, *142*, 677–687. [[CrossRef](#)]
69. Xiao, Y.; Wu, X.Z.; Wang, L.; Liang, J. Optimal farmland conversion in China under double restraints of economic growth and resource protection. *J. Clean. Prod.* **2017**, *142*, 524–537. [[CrossRef](#)]
70. Hughes, S. A meta-analysis of urban climate change adaptation planning in the US. *Urban Clim.* **2015**, *14*, 17–29. [[CrossRef](#)]
71. Hashem, I.A.T.; Chang, V.; Anuar, N.B.; Adewole, K.; Yaqoob, I.; Gani, A. The role of big data in smart city. *Int. J. Inf. Manag.* **2016**, *36*, 748–758. [[CrossRef](#)]
72. Kitchin, R. The real-time city? Big data and smart urbanism. *GeoJournal* **2014**, *79*, 1–14. [[CrossRef](#)]
73. Albino, V.; Berardi, U.; Dangelico, R.M. Smart cities: Definitions, dimensions, performance, and initiatives. *J. Urban Technol.* **2015**, *22*, 3–21. [[CrossRef](#)]
74. Al Nuaimi, E.; Al Neyadi, H.; Mohamed, N.; Al-Jaroodi, J. Applications of big data to smart cities. *J. Internet Serv. Appl.* **2015**, *6*, 25. [[CrossRef](#)]
75. Iqbal, R.; Doctor, F.; More, B.; Mahmud, S.; Yousuf, U. Big data analytics: Computational intelligence techniques and application areas. *Technol. Forecast. Soc. Chang.* **2018**. [[CrossRef](#)]
76. Haarstad, H. Constructing the sustainable city: Examining the role of sustainability in the ‘smart city’ discourse. *J. Environ. Policy Plan.* **2017**, *19*, 423–437. [[CrossRef](#)]
77. Kudva, S.; Ye, X. Smart Cities, Big Data, and Sustainability Union. *Big Data Cogn. Comput.* **2017**, *1*, 4. [[CrossRef](#)]
78. Bibri, S.E.; Krogstie, J. Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustain. Cities Soc.* **2017**, *31*, 183–212. [[CrossRef](#)]
79. Bibri, S.E. The IoT for Smart Sustainable Cities of the Future: An Analytical Framework for Sensor-Based Big Data Applications for Environmental Sustainability. *Sustain. Cities Soc.* **2017**, *38*, 230–253. [[CrossRef](#)]
80. Kaur, G.; Tomar, P.; Singh, P. Design of Cloud-Based Green IoT Architecture for Smart Cities. In *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*; Springer: Cham, Switzerland, 2018; pp. 315–333.
81. Lim, C.; Kim, K.J.; Maglio, P.P. Smart cities with big data: Reference models, challenges, and considerations. *Cities* **2018**, *82*, 86–99. [[CrossRef](#)]

82. Zhu, C.; Zhou, H.; Leung, V.C.; Wang, K.; Zhang, Y.; Yang, L.T. Toward big data in green city. *IEEE Commun. Mag.* **2017**, *55*, 14–18. [[CrossRef](#)]
83. Huang, L.; Bi, S.; Qian, L.P.; Xia, Z. Adaptive Scheduling in Energy Harvesting Sensor Networks for Green Cities. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1575–1584. [[CrossRef](#)]
84. Zhang, Y.; Mi, Z. Environmental benefits of bike sharing: A big data-based analysis. *Appl. Energy* **2018**, *220*, 296–301. [[CrossRef](#)]
85. Walther, G.R.; Post, E.; Convey, P.; Menzel, A.; Parmesan, C.; Beebee, T.J.; Fromentin, J.M.; Hoegh-Guldberg, O.; Bairlein, F. Ecological responses to recent climate change. *Nature* **2002**, *416*, 389. [[CrossRef](#)]
86. Van Aalst, M.K. The impacts of climate change on the risk of natural disasters. *Disasters* **2006**, *30*, 5–18. [[CrossRef](#)]
87. Helmer, M.; Hillhorst, D. Natural disasters and climate change. *Disasters* **2006**, *30*, 1–4. [[CrossRef](#)] [[PubMed](#)]
88. Goswami, S.; Chakraborty, S.; Ghosh, S.; Chakrabarti, A.; Chakraborty, B. A review on application of data mining techniques to combat natural disasters. *Ain Shams Eng. J.* **2016**, *9*, 365–378. [[CrossRef](#)]
89. Akter, S.; Wamba, S.F. Big data and disaster management: A systematic review and agenda for future research. *Ann. Oper. Res.* **2017**, 1–21. [[CrossRef](#)]
90. Wang, X.; Song, L.; Wang, G.; Ren, H.; Wu, T.; Jia, X.; Wu, H.; Wu, J. Operational climate prediction in the era of big data in China: Reviews and prospects. *J. Meteorol. Res.* **2016**, *30*, 444–456. [[CrossRef](#)]
91. Miyoshi, T.; Kunii, M.; Ruiz, J.; Lien, G.Y.; Satoh, S.; Ushio, T.; Bessho, K.; Seko, H.; Tomita, H.; Ishikawa, Y. “Big Data Assimilation” revolutionizing severe weather prediction. *Bull. Am. Meteorol. Soc.* **2016**, *97*, 1347–1354. [[CrossRef](#)]
92. Ofli, F.; Meier, P.; Imran, M.; Castillo, C.; Tuia, D.; Rey, N.; Briant, J.; Millet, P.; Reinhard, F.; Parkan, M.; et al. Combining human computing and machine learning to make sense of big (aerial) data for disaster response. *Big Data* **2016**, *4*, 47–59. [[CrossRef](#)]
93. Papadopoulos, T.; Gunasekaran, A.; Dubey, R.; Altay, N.; Childe, S.J.; Fosso-Wamba, S. The role of Big Data in explaining disaster resilience in supply chains for sustainability. *J. Clean. Prod.* **2017**, *142*, 1108–1118. [[CrossRef](#)]
94. Nguyen, D.T.; Ofli, F.; Imran, M.; Mitra, P. Damage assessment from social media imagery data during disasters. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, Sydney, Australia, 31 July–3 August 2017; pp. 569–576.
95. Kim, K.; Yoon, S. Assessment of Building Damage Risk by Natural Disasters in South Korea Using Decision Tree Analysis. *Sustainability* **2018**, *10*, 1072. [[CrossRef](#)]
96. Li, S.; Cui, Y.; Liu, M.; He, H.; Ravan, S. Integrating Global Open Geo-Information for Major Disaster Assessment: A Case Study of the Myanmar Flood. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 201. [[CrossRef](#)]
97. Cian, F.; Marconcini, M.; Ceccato, P. Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data. *Remote Sens. Environ.* **2018**, *209*, 712–730. [[CrossRef](#)]
98. Ramo, R.; Garcia, M.; Rodríguez, D.; Chuvieco, E. A data mining approach for global burned area mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 39–51. [[CrossRef](#)]
99. Hay, S.I.; George, D.B.; Moyes, C.L.; Brownstein, J.S. Big data opportunities for global infectious disease surveillance. *PLoS Med.* **2013**, *10*, e1001413. [[CrossRef](#)] [[PubMed](#)]
100. Traore, B.B.; Kamsu-Foguem, B.; Tangara, F. Data mining techniques on satellite images for discovery of risk areas. *Expert Syst. Appl.* **2017**, *72*, 443–456. [[CrossRef](#)]
101. Manogaran, G.; Lopez, D. Disease surveillance system for big climate data processing and dengue transmission. *Int. J. Ambient Comput. Intell.* **2017**, *8*, 88–105. [[CrossRef](#)]
102. Zhang, N.; Chen, Z. Sustainability characteristics of China’s Poyang Lake Eco-Economics Zone in the big data environment. *J. Clean. Prod.* **2017**, *142*, 642–653. [[CrossRef](#)]
103. Zhang, Y.; Ren, S.; Liu, Y.; Sakao, T.; Huisingh, D. A framework for Big Data driven product lifecycle management. *J. Clean. Prod.* **2017**, *159*, 229–240. [[CrossRef](#)]
104. Liang, Y.C.; Lu, X.; Li, W.D.; Wang, S. Cyber Physical System and Big Data enabled energy efficient machining optimisation. *J. Clean. Prod.* **2018**, *187*, 46–62. [[CrossRef](#)]
105. Wang, S.; Liang, Y.C.; Li, W.D.; Cai, X.T. Big Data enabled Intelligent Immune System for energy efficient manufacturing management. *J. Clean. Prod.* **2018**, *195*, 507–520. [[CrossRef](#)]



106. Gu, F.; Ma, B.; Guo, J.; Summers, P.A.; Hall, P. Internet of things and Big Data as potential solutions to the problems in waste electrical and electronic equipment management: An exploratory study. *Waste Manag.* **2017**, *68*, 434–448. [[CrossRef](#)]
107. Hassani, H.; Gheitanchi, S.; Yeganegi, M.R. On the Application of Data Mining to Official Data. *J. Data Sci.* **2010**, *8*, 75–89.
108. Hassani, H.; Saporta, G.; Silva, E.S. Data mining and official statistics: The past, the present and the future. *Big Data* **2014**, *2*, 34–43. [[CrossRef](#)] [[PubMed](#)]
109. Pang-Ning, T.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Pearson: London, UK, 2014.
110. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2017**, *160*, 3–24.
111. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: Burlington, MA, USA, 1992.
112. Cortes, C.; Vapnik, V. Support Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
113. Langley, P.; Iba, W.; Thompson, K. An analysis of Bayesian classifiers. In Proceedings of the 10th National Conference on Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992; pp. 223–228.
114. Zhang, G.P. Neural networks for classification: A survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2000**, *30*, 451–462. [[CrossRef](#)]
115. Han, E.H.S.; Karypis, G.; Kumar, V. Text categorization using weight adjusted k-nearest neighbor classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 53–65.
116. Agrawal, R.; Imieliński, T.; Swami, A. Mining Association Rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 26–28 May 1993; pp. 207–216.
117. Yun, H.; Ha, D.; Hwang, B.; Ryu, K.H. Mining Association Rules on significant rare data using relative support. *J. Syst. Softw.* **2003**, *67*, 181–191. [[CrossRef](#)]
118. Mena, J. *Investigative Data Mining for Security and Criminal Detection*; Butterworth-Heinemann: Oxford, UK, 2003.
119. Fard, A.M.; Ester, M. Collaborative mining in multiple social networks data for criminal group discovery. In Proceedings of the International Conference on Computational Science and Engineering, Vancouver, BC, Canada, 29–31 August 2009; pp. 582–587.
120. Guo, H.; Liu, Z.; Jiang, H.; Wang, C.; Liu, J.; Liang, D. Big Earth Data: A new challenge and opportunity for Digital Earth’s development. *Int. J. Digit. Earth* **2017**, *10*, 1–12. [[CrossRef](#)]
121. Ford, J.D.; Tilleard, S.E.; Berrang-Ford, L.; Araos, M.; Biesbroek, R.; Lesnikowski, A.C.; MacDonald, G.K.; Hsu, A.; Chen, C.; Bizikova, L. Opinion: Big data has big potential for applications to climate change adaptation. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10729–10732. [[CrossRef](#)]
122. Merchant, C.J.; Paul, F.; Popp, T.; Ablain, M.; Bontemps, S.; Defourny, P.; Hollmann, R.; Lavergne, T.; Laeng, A.; de Leeuw, G.; et al. Uncertainty information in climate data records from Earth observation. *Earth Syst. Sci. Data* **2017**, *9*, 511. [[CrossRef](#)]
123. Lee, J.G.; Kang, M. Geospatial big data: Challenges and opportunities. *Big Data Res.* **2015**, *2*, 74–81. [[CrossRef](#)]
124. Wang, H.; Xu, Z.; Fujita, H.; Liu, S. Towards felicitous decision making: An overview on challenges and trends of Big Data. *Inf. Sci.* **2016**, *367*, 747–765. [[CrossRef](#)]
125. Khan, S.; Shakil, K.A.; Alam, M. Cloud-Based Big Data Analytics—A Survey of Current Research and Future Directions. In *Big Data Analytics*; Springer: Singapore, 2018; pp. 595–604.
126. Maksimovic, M. Greening the future: Green Internet of Things (G-IoT) as a key technological enabler of sustainable development. In *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*; Springer: Cham, Switzerland, 2018; pp. 283–313.

