Machine Learning Research
2017; 2(1): 10-18
http://www.sciencepublishinggroup.com/j/mlr
doi: 10.11648/j.mlr.20170201.12

Review Article

An Overview of Big Data Applications in Water Resources Engineering

Sirisha Adamala

Applied Engineering Department, Vignan's Foundation for Science, Technology and Research University (VFSTRU), Vadlamudi, Andhra Pradesh

Email address:
sirisha@agfe.iitkgp.ernet.in

To cite this article:
Sirisha Adamala. An Overview of Big Data Applications in Water Resources Engineering. *Machine Learning Research.* Vol. 2, No. 1, 2017, pp. 10-18. doi: 10.11648/j.mlr.20170201.12

Received: January 15, 2016; **Accepted**: February 3, 2017; **Published**: March 1, 2017

Abstract: One of the emerging challenges in the 21th century era is collecting and handling 'Big Data'. The definition of big data changes from one area to other over time. Big data as its name implies is unstructured data that is very big, fast, hard and comes in many forms. Though the applications of big data was confined to information technology before 21st technology, now it is of emerging area in almost all engineering specializations. But for water managers/engineers, big data is showing big promise in many water related applications such as planning optimum water systems, detecting ecosystem changes through big remote sensing and geographical information system, forecasting/predicting/detecting natural and manmade calamities, scheduling irrigations, mitigating environmental pollution, studying climate change impacts etc. This study reviewed the basic information about big data, applications of big data in water resources engineering related studies, advantages and disadvantages of big data. Further, this study presented some of review of literature which has been done on big data applications in water resources engineering.

Keywords: Big Data, Petabytes, Gigabyte, Water Resources, Climate, Artificial Neural Network, Remote Sensing

1. Introduction

Big data is a term that is used to describe data that is high volume, velocity, and variety; requires new technologies and techniques to capture, store, and analyze it; and is used to enhance decision making, provide insight and discovery, and support and optimize processes. Science is driven by data. Big data is growing rapidly, currently at a projected rate of 40% growth in the amount of global data generated per year versus only 5% growth in global Information Technology (IT) spending [5]. Around 90% of the world's digitized big data was captured over just the past two years [15]. Big Data has the potential to revolutionize not only research, but also science. Imagine a system in which one has an access to a huge database; where one collect every detailed measure of every event that occurs on the globe. Big Data does not arise out of a vacuum: it is recorded from some data generating source: scientific experiments and simulations can easily produce petabytes (PB) of data per day. Much of this data is of no interest, and it can be filtered and compressed by orders of magnitude. One of the main challenges is to define these filters in such a way that they do not discard useful information [32]. For example, suppose one sensor reading differs substantially from the rest: it is likely to be due to the sensor being faulty, but how can one be sure that it is not an artifact that deserves attention? In addition, the data collected by these sensors most often are spatially and temporally correlated. Furthermore, one requires "on-line" analysis techniques that can process such streaming data on the fly, since one cannot afford to store first and reduce afterward.

The demand for food is expected to double by 2050 as the world's population heads toward 9 billion people. Lack of water is a critical constraint to increasing food production, particularly as droughts and other consequences of climate change are making water scarcer. For water managers, big data is showing big promise to plan water systems optimally, to analyze climate change impact, to detect changes in ecosystem through remote sensing, to predict natural and

manmade calamities, to schedule irrigation plans, to mitigate environmental pollution, etc. Many have understood the importance of big data analysis in the above mentioned areas and they initiated their research. In fact, mobilizing Big Data to improve global water and food security was the subject of the 2014 Water for Food Global Conference hosted by the Robert B. Daugherty Water for Food Institute at the University of Nebraska in association with the Bill and Melinda Gates Foundation being held October 19-22 in Seattle. Further, IBM has started its research on big data applications for watershed management by capturing meteorological, surface, sub-surface and groundwater data, monitoring rain, snow, soil moisture, water turbidity, flow rates, temperature, and groundwater quality using different sensors. Therefore, the study of big data in real-life water related applications helps in preventing many natural and man-made disasters like floods, droughts, overflow of rivers/streams containing toxic wastes, siltation, erosion etc.

Greater data gathering and computing power is allowing researchers to develop drought-resistant crop breeds, better understand climate change and create models that help us to understand risks and opportunities moving forward, among other research goals. Climate change is a very real problem facing our planet. The term "climate change" can cover a great many things, some natural and some manmade, including global warming and loss of wildlife habitat. Each of these brings its own challenges but, increasingly, big data and analytics are being put to use to come up with new solutions and research methods. Climate change has been attracting a lot of attention for a long time due to the adverse effects of it is being felt everywhere. High spatial and temporal resolution data, which is of big in nature is required to study about the weather impacts in current and changing climate. Further, uncertainty in the climate studies is addressing through multi climate model ensemble as the data is big and models are complex.

It is evident from the many studies that the change in climate is happening in a very fast way. For example: Over the past century, the global sea level has risen by about 10 to 25 cm, Arctic ice sheets were shrinking at about 13% a decade (according to NASA), glaciers are melting, cities are experiencing recurring floods and deforestation is on the rise, and Earth's average temperature has risen by 8.6 °F, and is projected to rise another 0.5 to 8.6 °F over the next hundred years [8]. To handle this issue, countries need a good action plan, which should be made on the basis of accurate, real-time or near real-time analytics. Big data and predictive analytics can potentially provide accurate, real-time or near real-time analytics. Given the rate at which climate is changing, one need to respond fast. Big data and predictive analytics technologies have enabled stakeholders to process huge volumes of data fast and generate accurate insights. Sensors are collecting data on various variables such as rain, soil, and forest cover and helping establish correlations between datasets. It is clear that big data and predictive analytics is, and will be one of the most important tools researchers will be using while they find ways to mitigate effects of climate change. More frequent and intense weather events can be predicted and managed with greater accuracy using big climatic data.

Big Data analysis and modelling can reach even farmers in poor, rural areas of developing countries through cell-phones, providing access to weather forecasting and market information to make better decisions and thereby helping improve livelihoods as well as local water and food security. Developing countries are taking on their own Big Data projects to better plan for the future. Sri Lanka, for example, recently began mapping many of its primary river basins to develop a comprehensive flood and drought mitigation plan. And 'Kerala Water Authority' uses IBM Big Data and analytics technology for seamless water distribution in the city of Thiruvananthapuram with more than 3.3 million inhabitants.

Big data, if done responsibly, can deliver significant benefits and efficiencies in water resources analysis, scientific research, environment, and other specific areas. Archive of big datasets collected during historical events that can be shared among researchers would be helpful in enabling quantitative analysis between different models i.e., the use of a common dataset for evaluations and validations of models would help researchers in developing models with improved quality. Especially, in the water resources area, there should be a well-established system of depositing collected data into a public repository, and also of creating public databases for use by other scientists/researchers. In fact, there should be an entire discipline of water resources that is largely devoted to the curation and analysis of such data.

Large amount of remote sensing data are now freely available from the NASA Open Government Initiative. Only one of NASA archives, the Earth Science Data and Information System (ESDIS), holds 7.5 petabytes (PB) of data with nearly 7,000 unique datasets and 1.5 million users in 2013 [31].

2. Review Studies on Big Data Applications in Water Resources

As we know, there is a flood of big data available at present era, but the applications of these big data are limited in many areas. It needs some time to get advance in the area of analyzing and applications of big data rather than collection and storing. Very few limited studies are available in reviewing the application of big data analysis in diverse areas other than information technology, which are mentioned below. [11] reported the various techniques to analyze big datasets and they have addressed various big data application in finance and business fields.

[2] performed both the biophysical and economic analysis and generated 6800 fully big and detailed future climates per emissions scenario. These climates were designed to represent the range of possible climate outcomes for South Africa by 2050. [25] reviewed about what is Big Data? Which technologies were used to build a Big Data infrastructure and

how it can be useful in the development of nation with prospect to improve decision-making in critical development areas such as health care, employment, economic productivity, crime and security, natural disaster and resource management? [34] introduced the water resources and hydropower cloud GIS platform based on development of service oriented big data architecture. The developed platform managed the various and massive data efficiently based on the construction of big data framework of survey, design, construction, environment, immigrant and equipment and supplies. [12] addressed about some of the real world examples of big data analytics and they differentiated how the big data analytics could be different from most supervisory control and data acquisition (SCADA) system.

[1] described the applicability of Big Data and analyzed the Big Data process model, which consists of storage system, handling process, and analysis mechanism. Further, they suggested that in future, the interest in and use of big data platforms will continue and expand and the applicable area too will go beyond Information Technology (IT) and be expanded to every possible sector. [24] introduced Big Data platform for environmental sciences and water resources management. This Platform was designed to provide effective tools that allow water system managers to solve complex water resources systems, water modeling issues and help in decision making. The Platform brings a variety of information technology tools including stochastic aspects, high performance computing, simulation models, hydraulic and hydrological models, grid computing, decision tools, Big Data analysis system, communication and diffusion system, database management, geographic information system (GIS) and Knowledge based expert system. The operators' objectives of this Big Data Open Platform were to solve and discuss water resources problems that are featured by a huge volume of collected, analyzed and visualized data, to analyze the heterogeneity of data resulting from various sources including structured, unstructured and semi-structured data, also to prevent and/or avoid a catastrophic event related to floods and/or droughts, through hydraulic infrastructures designed for such purposes or strategic planning.

[16] analyzed the challenges and opportunities that big data bring in the context of remote sensing applications. Furthermore, they described the most challenging issues in managing, processing, and efficient exploitation of big data for remote sensing problems. They had demonstrated the two case studies discussing the use of big data in the aforementioned aspects. In the first test case, big data were used to detect marine oil spills automatically using a large archive of remote sensing and social media data together. In the second test case, content-based information retrieval was performed using high performance computing to extract information from a large database of remote sensing images, collected after the terrorist attack on the World Trade Center in New York City on Sept. 11, 2001. [4] studied the opportunities, challenges and benefits of incorporating big data applications for smart cities. The results reveals that several opportunities were available for utilizing big data in smart cities; however, there were still many issues and challenges to be addressed to achieve better utilization of this technology. [10] demonstrated a full implementation of a travel demand model utilizing mobile phone big data in the form of calls, Global Position System (GPS) traces, or real time traffic monitoring system as an input. They have mapped the flows of people within cities on to transportation infrastructure

[35] developed a learning algorithm named 'Deep Big Networks (DBN)' from a big data set and applied it to large water distribution system (WDS) containing a dozen tanks and several pump stations. A total of 13 Artificial Neural Networks (ANNs) have been trained, with each producing one output, for representing the whole system operation. In addition, before training the ANNs, sensitivity analysis was performed to figure out what inputs are sensitive to the desired output. Results suggest that DBN can eliminate the requirement for sensitivity analysis and also avoid use of multiple ANNs to represent the whole system. Further, comparison results suggest that the DBN model outperformed the conventional ANN and they are efficient and require less computational units than the shallow ANNs. [19] reviewed the 180 articles related to the opportunities and threats of Big Data Analytics for international development. The advent of Big Data delivers a cost-effective prospect for improved decision-making in critical development areas such as healthcare, economic productivity and security.

[18] reviewed the significance of big data requirement in modelling integrated urban hydrology, which comprises of land-use change modeling, urban drainage modeling, rainfall–runoff modeling, and urban water quality modeling. [13] used the big data in refining the geospatial targeting of new drought-tolerant (DT) maize varieties in Malawi, Mozambique, Zambia, and Zimbabwe (southern Africa). Results indicated that more than 1.0 million hectares (Mha) of maize in the study countries was exposed to a seasonal drought frequency exceeding 20% while an additional 1.6 Mha experience a drought occurrence of 10–20%. Spatial modeling indicated that new DT varieties could give a yield advantage of 5–40% over the commercial check variety across drought environments while crop management and input costs are kept equal. [21] summarize the strengths, weaknesses, and opportunities associated with big water data. He developed a tool named "Water Quality Risk Assessment Tool". The tool is a model for how agencies and private developers can create value by identifying specific data sources, processing data with analytics, and visualizing data to address a specific user need. Furthermore, the tool shows how data value is increased with the user can interact with the information in an intuitive, functional interface.

[14] assessed the physical quantity and value of forest ecosystem services using oxygen release, soil and water conservation, carbon sequestration, and air purification as input data in Anhui provinces, central eastern China from 2009 to 2014 using big data. Further, optimization of output was done to identify the spatial and temporal heterogeneity during the study period. [29] evaluated the efficiency of

China's forest resources in terms of economic, social and ecological indices by utilizing the big data, which was collected from 31 inland provinces and municipalities of China from 2005 to 2013. [30] proposed and tested a theoretical framework to explain resilience in supply chain networks for sustainability using unstructured Big Data, based upon 36,422 items gathered in the form of tweets, news, Facebook, WordPress, Instagram, Google+, and YouTube, and structured data, via responses from 205 managers involved in disaster relief activities in the aftermath of Nepal earthquake in 2015. The authors have used Big Data analysis, followed by a survey, which was analyzed using content analysis and confirmatory factor analysis (CFA).

[9] addressed the big data challenges of climate science, using a notion of Climate Analytics-as-a-Service (CAaaS), which is enabled by Cloud Computing. A subset of cloud-enabled CAaaS 'Modern-Era Retrospective Analysis for Research and Applications (MERRA)' was developed, which integrates observational data with numerical models to produce a global temporally and spatially consistent synthesis of 26 key climate variables. It represents a type of data product that is of growing importance to scientists doing climate change research and a wide range of decision support applications. [20] documented the improved natural resource management in supporting the societal development and environmental sustainability of China's ecosystem using big data support. They have also proposed policies and decision-support models to provide optimization of sustainable technologies. [6] explored the factors, which influences the ecological land change using relevant big data sets, including spatial land data, soil data, DEM, climatic data, and socio-economic data during the period of 2000–2005 in China's Beijing–Tianjin–Hebei Region. The results showed that the factors influencing different types of ecological land change have substantial differences.

3. Theoretical Consideration

Big data is characterized by five features: Volume, Velocity, Variety, Veracity, and Value defined as five "V" dimensions.

- The term, 'Volume' means amount or quantity of data that has been created from all the sources (eg. TeraBytes (TB = 1024 GigaBytes), PetaBytes (PB = 1024 TB), and ExaBytes (EB=1024 PB)).
- 'Velocity' represents the rate at which data is created, stored, analyzed and processed. For example: the velocity of big data in remote sensing involves not only generation of data at a rapid growing rate, but also efficiency of data processing and analysis. In other words, the data should be analyzed in a (nearly) real or a reasonable time to achieve a given task, e.g., seconds can save hundreds or thousands of lives in an earthquake.
- 'Variety' indicates that there are various types of data, and they could be classified to structured, semi-structured, and unstructured data sets depending on the sort of structure. For example, big remote sensing data consist of multisource (laser, radar, optical, etc.), multi-temporal (collected on different dates), and multi-resolution (different spatial resolution) data, as well as data from different disciplines depending on several application domains.
- 'Veracity' can be described as the big noise in big data, which refers to the accuracy and truthfulness of the captured data and the meaningfulness of the results generated from the data for certain problems.
- 'Value' refers to the possible advantage that a big data can offer based on good data collection, management, and analysis. Some more Vs of big data are 'Volatility, which refers to the retention policy of the structured data implemented from different sources. Also there is 'Validity' that refers to the correctness, accuracy, and validation of the data.

3.1. Big Data Analysis

Recently due to the increased use of internet, digital devices, mobile phones, sensors, social networking sites, Satellite images, etc., large amount of big data (more than petabytes) is produced per minute or second. One can utilize these data for decision making, trend finding, weather forecast, planning optimum resources management, etc. For this purpose the big data present in the world has to be stored and analyzed.

Big data is creating a new generation of decision support data management. A key to deriving value from big data is the use of analytics. Collecting and storing big data creates little value; it is only data infrastructure at this point. The collected big data should be analyzed to derive a value or decision before drawing any results. Big data and analytics are intertwined, but analytics is not new. Many analytic techniques, such as regression analysis, simulation, and machine learning, have been available for many years. Analysis of Big Data can be helpful in improving the decision-making in critical development areas such as, environment, remote sensing, artificial intelligence, natural disaster and resource management, etc. The challenge of big data in water resources involves not only dealing with high volumes of data. In particular, challenges on data acquisition, storage, management and analysis are also related to water resources problems involving big data. By analyzing this data, organizations are able to learn trends about the data they are measuring, as well as the people generating this data. The hope for this big data analysis are to provide more customized service and increased efficiencies in whatever industry the data is collected from.

Analyzing big data helps in answering below questions:

- What happened?
- Why did it happen?
- What will happen?
- How can we make it happen?

The enormous volumes of data require automated or semi-automated analysis techniques to detect patterns, identify anomalies, and extract knowledge. Again, the "trick" is in the software algorithms - new forms of computation, combining statistical analysis, optimization, and artificial

intelligence, are able to construct statistical models from large collections of data and to infer how the system should respond to new data.

Maintaining and analyzing a sheer size of the data is a major challenge and finding right analytical tools and smart systems to sit on top of the data and interpreting it in order to assist management in decision-making processes is another challenge. The analysis of Big Data involves multiple distinct phases, each of which introduces challenges. Many people unfortunately focus just on the analysis/modeling phase: while that phase is crucial, it is of little use without the other phases of the data analysis. Even in the analysis phase, which has received much attention, there are poorly understood complexities in the context of multi-tenanted clusters where several users' programs run concurrently. Many significant challenges extend beyond the analysis phase. For example, Big Data has to be managed in context, which may be noisy, heterogeneous and not include an upfront model. Doing so raises the need to track provenance and to handle uncertainty and error: topics that are crucial to success, and yet rarely mentioned in the same breath as Big Data.

3.2. Big Data Analyzing Techniques

A/B testing or split testing or bucket testing: It is a technique in which control group is compared with a variety of test groups in order to determine what treatments (i.e., changes) will improve a given objective variable, e.g., crop yield [3].

Association rule learning or Fuzzy learning: It is a technique consisting variety of algorithms for discovering interesting relationships, i.e., "association rules," among variables in largeig or b databases.

Classification: It consists some set of techniques to identify the categories in which new data points belong, based on a training set, which containing data points that have already been categorized. These techniques have been divided as supervised learning and unsupervised learning. Supervised learning is a set of machine learning technique that infer a function or relationship from a set of training data. Unsupervised learning is a set of machine learning technique that finds hidden structure in unlabeled data.

Cluster analysis: A statistical method for classifying objects that splits a diverse group into smaller groups of similar objects, whose characteristics of similarity are not known in advance. This is a type of unsupervised learning because training data are not used. This technique is in contrast to supervised learning.

Crowd sourcing: A technique for collecting data submitted by a large group of people or community (i.e., the "crowd") through an open call, usually through networked media such as the Web.

Data fusion and data integration: A set of techniques that integrate and analyze data from multiple sources in order to develop insights in ways that are more efficient and potentially more accurate than if they were developed by analyzing a single source of data. Signal processing techniques can be used to implement some types of data fusion. One example of an application is sensor data from the Internet of Things being combined to develop an integrated perspective on the performance of a complex distributed system such as an oil refinery.

Data mining: A set of techniques to extract patterns from large datasets by combining methods from statistics and machine learning with database management. These techniques include association rule learning, cluster analysis, classification, and regression [35].

Ensemble learning: Using multiple predictive models (each developed using statistics and/or machine learning) to obtain better predictive performance than could be obtained from any of the constituent models. This is a type of supervised learning.

Genetic algorithms: A technique used for optimization that is inspired by the process of natural evolution or "survival of the fittest." In this technique, potential solutions are encoded as "chromosomes" that can combine and mutate. These individual chromosomes are selected for survival within a modeled "environment" that determines the fitness or performance of each individual in the population. Often described as a type of "evolutionary algorithm," these algorithms are well-suited for solving nonlinear problems.

Machine learning: A subspecialty of computer science (within a field historically called "artificial intelligence") concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data.

Natural language processing: A set of techniques from a subspecialty of computer science (within a field historically called "artificial intelligence") and linguistics that uses computer algorithms to analyze human (natural) language.

Neural networks: Computational models, inspired by the structure and workings of biological neural networks (i.e., the cells and connections within a brain), that find patterns in data. Neural networks are well-suited for finding nonlinear patterns. They can be used for pattern recognition and optimization. Some neural network applications involve supervised learning and others involve unsupervised learning.

Network analysis: A set of techniques used to characterize relationships among discrete nodes in a graph or a network. In social network analysis, connections between individuals in a community or organization are analyzed, e.g., how information travels, or who has the most influence over whom.

Optimization: A portfolio of numerical techniques used to redesign complex systems and processes to improve their performance according to one or more objective measures (e.g., cost, speed, or reliability). Genetic algorithms are an example of an optimization technique.

Pattern recognition: A set of machine learning techniques that assign some sort of output value (or *label*) to a given input value (or *instance*) according to a specific algorithm. Classification techniques are an example.

Predictive modeling: A set of techniques in which a

mathematical model is created or chosen to best predict the probability of an outcome.

Regression: A set of statistical techniques to determine how the value of the dependent variable changes when one or more independent variables are modified. Often used for forecasting or prediction.

Sentiment analysis: Application of natural language processing and other analytic techniques to identify and extract subjective information from source text material. Key aspects of these analyses include identifying the feature, aspect, or product about which a sentiment is being expressed, and determining the type, "polarity" (i.e., positive, negative, or neutral) and the degree and strength of the sentiment.

Signal processing: A set of techniques from electrical engineering and applied mathematics originally developed to analyze discrete and continuous signals, i.e., representations of analog physical quantities (even if represented digitally) such as radio signals, sounds, and images. This category includes techniques from signal detection theory, which quantifies the ability to discern between signal and noise. Sample applications include modeling for time series analysis or implementing data fusion to determine a more precise reading by combining data from a set of less precise data sources (i.e., extracting the signal from the noise).

Spatial analysis: A set of techniques, some applied from statistics, which analyze the topological, geometric, or geographic properties encoded in a data set. Often the data for spatial analysis come from geographic information systems (GIS) that capture data including location information, e.g., addresses or latitude/longitude coordinates.

Statistics: The science of the collection, organization, and interpretation of data, including the design of surveys and experiments. Statistical techniques are often used to make judgments about what relationships between variables could have occurred by chance (the "null hypothesis"), and what relationships between variables likely result from some kind of underlying causal relationship (i.e., that are "statistically significant"). Statistical techniques are also used to reduce the likelihood of Type I errors ("false positives") and Type II errors ("false negatives"). An example of an application is A/B testing to determine what types of marketing material will most increase revenue.

Simulation: Modeling the behavior of complex systems, often used for forecasting, predicting and scenario planning. Monte Carlo simulations, for example, are a class of algorithms that rely on repeated random sampling, i.e., running thousands of simulations, each based on different assumptions. The result is a histogram that gives a probability distribution of outcomes.

Time series analysis: Set of techniques from both statistics and signal processing for analyzing sequences of data points, representing values at successive times, to extract meaningful characteristics from the data.

Visualization: Techniques used for creating images, diagrams, or animations to communicate, understand, and improve the results of big data analyses.

3.3. Applications of Big Data

Nowadays Big Data are becoming a popular topic and a comparatively new technological concept focused on many different disciplines like agriculture, remote sensing, climate change, water resources, environmental science, etc. Everyone is in eager of research to find how the captured data and analysis of it is transforming the management of our most precious resource: water. A main goal in big data applications is to identify the right data to solve the problems at hand, which are difficult to be addressed or mostly cannot be manipulated by traditional data. In general, big data are of no value until they are utilized for applications. There are many applications of big data, which are mentioned below.

- Automated sensor and monitoring systems are providing large amounts of real time flow data. For example, automated sensors in irrigation systems are producing various climate (temperature, radiation, wind speed, and humidity), crop (crop height, plant density, leaf area index etc.), and soil (moisture content, infiltration etc.) data on the order of seconds, minutes, and other can be a matter of hours. These data can be stored and analyzed to understand and automate an irrigation water source to either on or off. The data generated by sensors need to be processed in real-time for immediate action. However, the development and validation of models using real-time data is a challenging task.
- There is a tremendous amount of geospatial data that created by cell phones, GPS, space borne or air borne sensors and that can be used for many applications like to find exact latitude and longitude of remote location, natural hazard monitoring, global climate change, urban planning, etc.
- A Global Climate Model (GCM) needs to analyse a variety of atmospheric and ocean models (big data) with an integration of many heterogeneous big data sources and software tools to achieve global weather prediction.
- The major threat faced from nature in developing countries is flood, drought, earthquakes or any other natural disaster. Big Data can be proved fruitful in many instances to tackle with the unpredicted nature. Application which can give correct statistics about the rainfall, humidity, correct measure about water resources, snow fall etc. can be developed, which analyses the big data gathered historically, from satellite images or through digital medium.
- Sensors, robotics and computational technology in terms of big data can be used to track river and estuary ecosystems, which help officials to monitor water quality and supply through the movement of chemical constituents and large volumes of underwater acoustic big data that tracks the behavior of fish and marine mammal species [26].
- Big Data can be useful in finding new sources of natural resources like coal, gold, iron ore etc. Sometimes, big data applications will help the government to make correct estimation for the use of natural resources,

inform the coastal area people about high tides, informing farmers about the rainfall, which crop to cultivate, which species of animal or plant to save, which natural resources to save etc.

- The information processed by big data applications should not be personal: data generated by sensors for monitoring natural or atmospheric phenomena like the weather or pollution should have a global access.
- Crop assessment and yield forecasting [27, 17] requires big data analysis.
- Big data should be aggregated from diverse sources in a huge volume and imported to a model which allows decision-making in minutes rather than weeks or months. This is a big challenge for PetaByte level or larger volume of data inputs, for instance in applications related with hazard monitoring.
- Big data techniques have been successfully used for different applications, such as agriculture (e.g., food security monitoring, pasture monitoring), oceanic (e.g., ship detection, oil spill detection), Urban planning, management, and sustainability [33, 23], human settlements (both urban and rural), grape productivity [22], food security monitoring, water quality monitoring, energy assessment, population of disease, ecosystem assessment, global warming, global change, global forest resources assessment, ancient site discovery (For instance, a hidden relic site can be found by high resolution remote sensing data in a dense forest without modern infrastructure, which is an incredible barrier for field archaeologists to penetrate), Land development and use [28] and so on.
- With the availability of big data, one can predict the trend of any even (Example: Flood) over hourly basis unlike days, weeks and years.
- Helps in preventing man-made disasters, such as sudden drops in water quality, which may not be detected until after they are reported in the media or after the outbreak of a contagious disease.
- Provide weather information that will lead to improving the country's agriculture, better informing people of possible hazardous conditions, and better management of energy utilization by providing more accurate predictions on demand.
- Big Data can help in all four phases of disaster management: prevention, preparedness, response, and recovery. Two major sources of big data, dedicated sensor networks (e.g., earthquake detection using seismometers) and multi-purpose sensor networks (e.g., social media such as Twitter using smart phones), both have demonstrated their usefulness in disasters such as the Earthquake.
- Big data generated from the seismometers and other sources in terms of Seismic intensity maps, wind velocity maps, Tsunami estimation maps, radiation distribution maps and rainfall maps are continuously fed into geological models that can find earthquake epicenters precisely within seconds of the detection of earthquake vibrations.
- Understanding the environment requires collecting and analyzing data from thousands of sensors monitoring air and water quality and meteorological conditions. These measurements can then be used to guide simulations of climate and groundwater models to create reliable methods to predict the effects of long-term trends, such as increased CO_2 emissions and the use of chemical fertilizers.
- Data feeds from pumping stations, sewage plants, and reservoirs deliver information to the central control systems on a range of measurements. This helps in spotting the flash floods by monitoring water levels in a reservoir timely and a GPS system assists in tracking the location flash flood site to follow up warning and rescue operations.
- Helps in improving the quality of precision farming which, reduce water requirements, allow for an increase in both crop volume and nutritional value of food grown.

3.4. Disadvantages of Big Data Analysis

- Inaccurate information drawn from big data analytics result in the distribution of rescuers and supplies to wrong places, wasting limited resources. False information could also misguide the public, increasing their stress level.
- The quantity of information now available to individuals and organizations is unprecedented in human history, and the rate of information generation continues to grow exponentially. Yet, the sheer volume of information is in danger of creating more noise than value, and as a result limiting its effective use.
- Disasters affect every country on Earth and effective disaster management is a global challenge. This is particularly the case of large-scale disasters that affect many countries (e.g., the 2004 Indian Ocean earthquake and tsunami) and multi-hazards such as the Earthquake and landslides. Tools that can be used by many countries will have significant broad impact in helping the world population as well as many government agencies and non-governmental organizations. At the same time, the importance of cultural, social, and linguistic differences among countries in emergency response and recovery, this will have impact on the big data used for disaster management.
- Having the ability to analyze Big Data is of limited value if users cannot understand the analysis. Ultimately, a decision-maker, provided with the result of analysis, has to interpret these results. This interpretation cannot happen in a vacuum. Usually, it involves examining all the assumptions made and retracing the analysis.
- There are many possible sources of error: computer systems can have bugs, models almost always have assumptions, and results can be based on erroneous data. Man power/human mind should be always there to understand and verify, the results produced by the computer. This is particularly a challenge with Big Data

due to its complexity.
- Privacy of information and security of some of data constitutes a constraint.

4. Summary and Conclusions

This paper has successfully studied the definition of big data, size of big data, various previous studies related to big data towards water sector, techniques to analyze big data, advantages and disadvantages in using this data, applications in various fields. Though an extensive review on big data applications in various water resources related subjects have been studied here, still much have to learn about how to use and analyze the big data.

References

[1] B. T. Chun, S. H. Lee, A study on big data processing mechanism and applicability. International Journal of Software Engineering and Its Applications, 8 (8) (2014) 73-82.

[2] C. A. Schlosser, X. Gao, K. Strzepek, A. Sokolov, C. E. Forest, S. Awadalla, W. Farmer. Quantifying the likelihood of regional climate change: A hybridized approach. J Clim, 26 (10) (2012) 3394-3414.

[3] C. Thota, G. Manogaran, D. Lopez, V. Vijayakumar, Big data security framework for distributed cloud data centers. In Cybersecurity Breaches and Issues Surrounding Online Threat Protection, IGI Global (2017) 288-310.

[4] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, J. Al-Jaroodi, Applications of big data to smart cities. Journal of Internet Services and Applications. 6 (1) (2015) 1-15.

[5] H. D. Guo, L. Zhang, L. W. Zhu, Earth observation big data for climate change research. Advances in Climate Change Research, 6 (2) (2015) 108-117.

[6] H. Xie, Y. He, X. Xie, Exploring the factors influencing ecological land change for China's Beijing–Tianjin–Hebei region using big data. Journal of Cleaner Production, 142 (2017), 677-687.

[7] I. Becker-Reshef, C. Justice, B. Doorn, C. Reynolds, A. Anyamba, C. Tucker, S. Korontzi, NASA contribution to the group on earth observation (GEO) global agricultural monitoring system of systems. The Earth Obs. 21 (2009) 24-29.

[8] IPCC, In: Solomon S (ed) Climate change: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, (2007).

[9] J. L. Schnase, D. Q. Duffy, G. S. Tamkin, D. Nadeau, J. H. Thompson, C. M. Grieg, ... W. P. Webster, MERRA analytic services: Meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. Computers, Environment and Urban Systems, 61 (2017), 198-211.

[10] J. L. Toole, et al. The path most traveled: Travel demand estimation using big data resources. Transport. Res. Part C, http://dx.doi.org/10.1016/j.trc.2015.04.022, (2015).

[11] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers, Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, (2011).

[12] J. Neemann, D. Roberts, P. Kenel, A. Chastain-Howley and ScottStallard. Manager to manager: Will data analytics change the way we deliver water? Journal (American Water Works Association), 105 (11) (2013) 25-27.

[13] K. Tesfaye, K. Sonder, J. Caims, C. Magorokosho, A. Tarekegn, G. T. Kassie, ... O. Erenstein, Targeting drought-tolerant maize varieties in southern Africa: a geospatial crop modeling approach using big data, (2016).

[14] L. Li, T. Hao, T. Chi, Evaluation on China's forestry resources efficiency based on big data. Journal of Cleaner Production, 142 (2017a), 513-523.

[15] M. Bernard, Big data: using smart big data, analytics and metrics to make better decisions and improve performance. Chichester: John Wiley & Sons, (2015).

[16] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, Y. Zhu, Big data for remote sensing: challenges and opportunities. Proceedings of the IEEE, (2015).

[17] M. D. Steven, J. A. Clark, Applications of remote sensing in agriculture. Butterworth-Heinemann, (1991).

[18] M. G. Hutchins, S. J. McGrane, J. D. Miller, A. Hagen-Zanker, T. R. Kjeldsen, S. J. Dadson, C. S. Rowland, Integrated modeling in urban hydrology: reviewing the role of monitoring technology in overcoming the issue of 'big data' requirements. Wiley Interdisciplinary Reviews: Water, 4 (1) (2017).

[19] M. Hilbert, Big data for development: A review of promises and challenges. Development Policy Review, 34 (1) (2016), 135-174.

[20] M. Song, L. Cen, Z. Zheng, R. Fisher, X. Liang, Y. Wang, D. Huisingh, How would big data support societal development and environmental sustainability? Insights and practices. Journal of Cleaner Production, 142 (2017), 489-500.

[21] M. Ziman, Data intelligence for improved water resource management. Masters project: Nicholas School of the Environment of Duke University, (2016).

[22] N. Dokoozlian, Big data and the productivity challenge for wine grapes. In Agricultural Outlook Forum (No. 236854). United States Department of Agriculture, (2016).

[23] P. Gamba, M. Herold, Global mapping of human settlement: Experiences, datasets, and prospects. CRC Press, (2009).

[24] R. Chalh, Z. Bakkoury, D. Ouazar, M. D. Hasnaoui, Big data open platform for water resources management. In Cloud Technologies and Applications (CloudTech), International Conference on IEEE, (2015) 1-8.

[25] R. Panicker, Adoption of big data technology for the development of developing countries. In Proceedings of National Conference on New Horizons in IT-NCNHIT, (2013) 219.

[26] S. L. LaDeau, B. A. Han, E. J. Rosi-Marshall, K. C. Weathers, The next decade of big data in ecosystem science. Ecosystems, (2017) 1-10.

[27] S. Liaghat, S. Balasundram, A review: The role of remote sensing in precision agriculture. American Journal of Agricultural and Biological Sciences, 5 (1) (2010) 50-55.

[28] S. Martinuzzia, W. A. Goulda, O. M. R. Gonzaleza, Land development, land use, and urban sprawl in Puerto Rico integrating remote sensing and population census data. Landscape and Urban Planning, 79 (2007) 288-297.

[29] T. Li, , Y. Cui, A. Liu, Spatiotemporal dynamic analysis of forest ecosystem services using “big data”: A case study of Anhui province, central-eastern China. Journal of Cleaner Production, 142 (2017b), 589-599.

[30] T. Papadopoulos, A. Gunasekaran, R. Dubey, N. Altay, S. J. Childe, S. Fosso-Wamba,. The role of Big Data in explaining disaster resilience in supply chains for sustainability. Journal of Cleaner Production, 142 (2017), 1108-1118.

[31] T. W. Crowther, et al. Mapping tree density at a global scale. Nature, 525 (2015) 201-205.

[32] V. Mayer-Schonberger, K. Cukier, Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, (2013).

[33] X. Deng, J. Huang, S. Rozelle, E. Uchida, Growth, population and industrialization, and urban land expansion of China. J. Urb. Econ, 63 (2008) 96-115.

[34] X. Wang, Z. Sun, The design of water resources and hydropower cloud GIS platform based on big data. In Geo-Informatics in Resource Management and Sustainable Ecosystem, Springer Berlin Heidelberg, (2013) 313-322.

[35] X. Wu, X. Zhu, G. Q. Wu, W. Ding, Data mining with big data. IEEE transactions on knowledge and data engineering, 26 (1) (2014) 97-107.

[36] Z. Y. Wu, M. El-Maghraby, S. Pathak, Applications of deep learning for smart water networks. Procedia Engineering, 119 (2015) 479-485.