ITcon
www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

BIG DATA IN BUILDING DESIGN: A REVIEW

SUBMITTED: October 2017
REVISED: November 2018
PUBLISHED: November 2018 at https://www.itcon.org/2018/13
EDITOR: Amor R.

Mauricio Loyola
Princeton University
mloyola@princeton.edu

SUMMARY: *Compared with other industries, the use of big data is substantially less developed in the architecture, construction, engineering, and operation (AECO) industry. Available examples are mostly experimental, limited in scope, and without an obvious transfer to practice. Yet, a substantial number of academics and practitioners envision that the use of big data will have a significant impact on how decisions are made in the building design process, providing designers with an unprecedented amount of exceptionally detailed data on buildings and their occupants that is redefining what can be observed, modeled, simulated, predicted and measured in the industry. We first review the main concepts and technologies that shape this technological phenomenon, and then, based on the survey of around 100 cases of applications of big data and machine learning throughout all phases of building design we identify 12 clusters of applications where the technology seems to be more disruptive. The discussion then focuses on the transformative impacts of the technology on three niches where further research is more urgently needed: 1) understanding the design problem; 2) simulating and predicting the holistic performance of the design solution; 3) evaluating integrally and continuously the design product during its complete lifecycle.*

KEYWORDS: *Big Data, Building Design, Machine Learning, Architecture*

REFERENCE: *Mauricio Loyola (2018). Big data in building design: a review. Journal of Information Technology in Construction (ITcon), Vol. 23, pg. 259-284, http://www.itcon.org/2018/13*

1. INTRODUCTION

Many authors claim that big data is one of the most important and revolutionary scientific and industrial phenomenon in the last half century (Walker 2014). The range of applications is endless: from predicting consumer behavior to detecting bank fraud, and from scheduling flights to defining custom medical treatments. In fact, there are documented uses of big data in virtually every sphere of human activity: business (Chen et al. 2012; Minelli et al. 2012), management (Ju et al. 2013), finance (Fang & Zhang 2016), education (Tulasi 2013; Sin & Muthu 2015), manufacturing (Bi & Cochran 2014), healthcare (Bellazzi 2014; Groves et al. 2013), public administration (Maciejewski 2016; Martin 2014), tourism (Xiang et al. 2015; Fuchs et al. 2014), law enforcement (Glasgow 2015), and increasingly in all areas of scientific research (Bryant 2011; Szalay 2011; NIST 2015b).

Although the architecture, engineering, construction and operation (AECO) industry is no exception, its current level of development is comparatively much less advanced. In AECO, neither the same amount of data is collected nor is it used with the same intensity. There are very few concrete applications, mostly conducted with experimental purposes in academic settings. Their transferability to practice is still unclear. Yet, a significant number of academics and practitioners envision that the use of big data will have a significant impact on the building industry (Bilal et al. 2016; Deutsch 2015a; Deutsch 2015b). The details of this transformation are still to be seen, but even in this seminal phase, it is apparent that big data is enabling designers with the ability not only to inform design with an unprecedented level of detail and specificity, but also to model phenomena that traditionally are considered subjective, extremely complex or even unpredictable. From assessing the "character" of a neighborhood, or measuring the level of "selfie attractiveness" of a building façade, to identifying the behavioral patterns in window operation or predicting the occupancy schedule of office spaces, it seems that the deluge of data combined with machine learning algorithms are to redefine what can be understood, simulated, and measured in the building design process.

In this review, we make a comprehensive and systematic review of the technologies, applications, and opportunities of the use of big data in building design. First, we present a general review of the main concepts and technologies that shape this technological phenomenon. Then, we survey around 100 cases of applications of big data and machine learning in building design identifying 12 clusters of applications where the technology seems to be more disruptive. Finally, we discuss the transformative impacts of the technology on three incipient niches where more research is needed. First, how big data enables a more thorough and accurate definition of the design problem by providing detailed characteristics of the occupants, context, and other relevant agents. Second, the possibility of conducting complex simulations of the performance of the design solution based solely on statistical models developed through machine learning, including those aspect that today are considered unpredictable because of the lack of scientific explanatory models. Third, the opportunity to assess continuously the design product, both the project and the building, during its entire lifecycle, rethinking the concept of building performance to include more complex and multivariable metrics.

1.1 Methods

Around 700 abstracts of peer-reviewed conference and journal papers obtained from leading academic databases were examined (Google Scholar, Web of Science, Science Direct/Scopus, EBSCOhost, IEEE Xplore, CuminCAD, JSTOR). The generic descriptor "big data" was used in conjunction with other concepts associated with the field of study ("building design", "architecture", "architectural design", "structural design", "building construction", "building operation".). All abstracts were screened to evaluate their relevance, since the coincidence of terms between the construction industry and systems engineering (systems "architecture", network "design", etc.) makes the search difficult. About half of the papers came from engineering journals focusing on energy issues and building performance. On the other hand, papers from architecture journals and conferences (i.e. CuminCAD: ACADIA, eCAADe, CAAD Futures etc.) were notoriously less. Around 200 papers were finally selected as pertinent (i.e. related to the purpose of the review). The final number of papers was slightly extended through a selective snowball method, as the need to deepen the review in complementary topics arose. For the analysis, Mendeley was used as a reference manager and FreeMind was used for conceptual maps.

2. DEFINITIONS

2.1 Building Design

In this paper, *building design* refers to the conceptualization, formalization, and elaboration of all necessary information for the construction and subsequent operation of a building. Contrarily to what colloquial language suggests, building "design" is not an artistic and/or technical activity conducted exclusively by architects, but it is in fact an integrated and multidisciplinary activity that requires the participation of architects, structural engineers, MEP engineers, lighting designers, construction consultants, and many other specialists. We use the word "designer" to refer generically to the whole set of specialists.

Arbitrarily, we also have limited the scope of "buildings" to the most traditional meaning of the word, excluding works of civil engineering, landscape architecture, interior design, and urban design. Obviously, the use of big data goes far beyond building design, but circumscribing this paper to this domain facilitates its understanding and eventual extrapolation or contraposition to other design domains.

2.2 Data Driven Design

It is reasonable to say that the use of data in building design is not new (Oechslin 1993; Schmitt 1999). Client briefs, structural calculations, and building codes can be considered compendiums of data that restrict and/or drive design. Some activities, such as structural dimensioning, MEP engineering, and to a lesser extent, space programming, can be considered to be traditionally data-driven. Other activities, such as the election of the *parti pris* or the aesthetic criteria for a building, have always been strongly based on the intuition, experience, personal talent, and implicit knowledge of the designers (Binnekamp 2010). The line separating both realms is not clear, but in recent years a growing trend in the AECO industry has emerged promoting a greater use of objective, verifiable, and quantitative data in the decision-making process, including activities traditionally considered subjective or expert-based. Borrowing a concept used in other industries (Karakiewicz 2010; Yin et al. 2014), this trend has been called *data driven design* (Deutsch 2015a).

Maybe the most well-known data driven design approach is *evidence-based design* (EBD), which promotes the conscientious, explicit, and judicious use of current best evidence of research and practice in making critical decisions about the design of a project (Hamilton & Watkins 2009). It is a movement that was born in the context of the design of healthcare buildings, in direct relation to the trend of evidence-based medicine that emerged in the 1990s (Ulrich et al. 2008; Mccullough 2010), and that has progressively extended towards other areas: education, retail (Brown & Corry 2011), office buildings (Sailer et al. 2008), scientific buildings, historic preservation. EBD promotes the use of data understood as processed information in the form of traditional (i.e. academic) formal research from social and natural sciences (Brandt et al. 2010; CHD 2016).

All data-based design approaches share the basic underlying assumption that the use of more objective, quantifiable, and verifiable information in the design process allows to better fit the requirements and clients by increasing awareness of the constraints and opportunities, eliminating bias in selecting options, reducing risks when adopting new solutions, and, in general, minimizing decisions that are made without basis. The space for big data in building design must necessarily be understood as part of this larger conceptual umbrella. Only in this context does it make sense to favor methods that provide data to support the decision-making in design.

2.3 Big Data

There is no specific threshold that defines when exactly a dataset should be called big data. In commercial contexts, it is usually defined as a dataset that is impossible to analyze with traditional means within an acceptable range of time and resources. However, this is a slippery definition as "traditional means" are relative to time and place. Instead, academic definitions describe big data as extensive datasets that fulfill three characteristics (the "3 V"); *volume* (large quantities expressed in tera-, peta-, exa- or zillion- bytes), *variety* (different types, formats, origins, structures, etc.), and *velocity* (produced in real time, with low latency, forcing its continuous processing), requiring a scalable architecture for efficient storage, manipulation and analysis (Douglas 2001; NIST 2015a). Several authors have expanded this definition by adding as many "Vs" as linguistic creativity allows: *veracity* (reliability and trustworthiness), *virtual* (= digital), *value* (content-rich and useful for decision-making), *variability* (variation in data flow, not consistent), *validity* (correct and accurate for the intended use), *venue* (=distributed, heterogeneous data from multiple platforms), *volatility* (how long data should be stored), etc. (Alekhya et al. 2016; Laxmi et al. 2016; Gandomi & Haider 2015).

Despite of its fuzzy boundaries, it is consensual that the idea of big data encompasses two main areas of development: *big data engineering*, i.e. the computational infrastructure for collecting and storing very large datasets at reasonable cost and time; and *big data analytics*, i.e. the statistical and analytical processes for extracting meaningful knowledge from the data (De Mauro et al. 2015). These two areas are briefly presented next.

2.3.1 Big Data Engineering

Big data engineering (BDE) refers to the development of the technological scalable infrastructure necessary for the efficient capture, collection, storage, and preparation of large datasets (NIST 2015a). BDE is necessary because traditional data systems are not suitable for large, volatile, and unorganized datasets, which is the typical case with big data.

In a traditional database, the data is typically organized in table-based models with rows and columns called *relational database management systems* (RDBMS). To access and manages entries, a series of standardized commands are used, called *structured query language* (SQL). RDBMS and SQL work very well with structured data, being the basis of almost all modern commercial database systems since the 1980s. However, they do not work well with unstructured data and are not easily scalable for large datasets that need distributed computing (Hu et al., 2014)

In contrast, *non-relational databases* and *NoSQL systems* are much more flexible, with several options for storing data other than tabular format (key-value, document, graph, etc.). NoSQL databases are also more scalable and provide superior performance than RDBMS systems.

The most common contemporary big data infrastructure platform is *Apache Hadoop*, an open source framework technology based in Java. Hadoop allows managing large volumes of data in a distributed system at a low cost, with high degree of fault tolerance and high scalability. Very broadly, Hadoop consists of two main components: a storage part called *Hadoop Distributed File System* (HDFS) that allows storing and indexing large files across multiple machines; and a processing part called *MapReduce*, that breaks down complex data into small units of work that are processed in parallel. More detailed reviews of these technologies are found in Chen et al. (2014), Chen & Zhang (2014), and Bendre & Thool (2016).

2.3.2 Big Data Analytics

In general terms, *(big) data analytics* refers to the overarching process of using large raw databases to discover and obtain meaningful knowledge for decision making. For this reason, this process is also called *knowledge-discovery in databases* (KDD) (Bi & Cochran 2014; Pyne et al. 2016; Chong & Shi 2015).

Depending on the industry, type of analysis, and knowledge goal, the big data analysis process may require multiple and complex skills, giving rise to a new field called *data science* (Haider 2015). For human resources experts, this interdisciplinary field is one with the largest predicted growth for the coming years (Davenport & Patil 2012; Van der Aalst 2014). Several educational institutions have responded to this demand by offering new programs typically mixing computer science, statistics, machine learning, network engineering, research methods, business, data visualization and communication. (Provost & Fawcett 2013; O'Neil & Schutt 2015).

In other industries other concepts are used that overlap in meaning. In business, *business intelligence* describes the process of analyzing large datasets of market data and internal corporate data to assist corporate executives to make informed business decisions (Chen et al. 2012). In marketing and finance, *predictive analytics* refers to the creation of a statistical model of future or unknown events, particularly consumer or market behavior (Siegel 2013). In computer science, *data mining* refers to the specific activity of examining large datasets to discover unexpected and previously unrecognized patterns, usually involving advanced computational and statistical methods (Han et al, 2011).

2.4 Machine Learning

Machine learning is generally understood as a type of artificial intelligence that provides computers with the ability to learn to do something without being explicitly programmed to do it. Machine learning algorithms examine large datasets and used statistical tools to discover underlying patterns or relationships between data, without having prior knowledge of them in the form of a formal model of relationships (Alpaydin 2014.; Shalev-Shwartz & Ben-David 2014; Marsland 2015).

Although it is generalized the idea that big data analysis requires the use of machine learning algorithms, this is not completely correct. If the data are structured and the researcher/designer has sufficient previous knowledge about them to define a formal model of relationships between them, then traditional statistical techniques are more than adequate. However, it is true that in most cases of big data these conditions are not met and therefore it is necessary to resort to methods that can deal with unstructured and unknown data.

According to the type of interaction with the data (Ayodele 2010), machine learning algorithms can be classified as: *supervised learning*, when the objective is to analyze a set of labeled data mapped as input → output pairs and to infer a general function that allows mapping previously unseen values; *unsupervised learning*, when the objective is to analyze a set of unlabeled data and to infer a function to describe its structure, typically groups or clusters; and *reinforcement learning*, when the objective is to automatically determine an ideal value or behavior within a specific context based on the maximization of an external reward feedback.

Some of the most common and most influential algorithms used in building design (Wu et al. 2008; Lee & Siau 2001) are the following:

Decision trees are supervised learning algorithms that construct a model (in the form of a decision tree) that predicts the value of output value (leaves) based on several inputs values (branches). When the output value is a class, the tree is called classification tree, and when the output value is a real number, it is called regression tree. Therefore, decision trees algorithms are also called classification and regression trees (CART). Most decision tree algorithms are constructed following a top-down approach, choosing at each step (or branch) the variable that best divides the data set. Algorithms vary in the metrics used to evaluate the effectiveness of each division. In certain problems, in order to compensate for the habit of decisions trees to overfit the training set, multiple but different decision trees are simultaneously utilized, and then, the mode or mean are taken as the final value. This technique is called *random forests* (Rokach & Maimon 2014).

Artificial neural networks (ANNs) are growingly popular algorithms, with many applications in structured and unstructured learning. Inspired by biological neural networks, they are series of interconnected "artificial neurons" (i.e. weighted composition of mathematical functions) that process and analyze data in a non-linear fashion, constantly updating their weights based on iterative results (i.e. learning). Multilayer perceptrons are probably the most popular ANN in use today. ANNs are used in a wide variety of cases: to model complex relationships between inputs and outputs, to find unexpected patterns or structures in data. When an ANN has multiple layers of artificial neurons (as opposed to a simple single-level ANN with only input→output neurons) it is called *deep learning ANN*. A typical application of deep learning is computer vision (Yegnanarayana 2009). However, as ANN models increase in complexity, they become a sort of "black box" with mathematical processes that are uninterpretable by humans. Their results, although valuable, become less explainable. This is a key disadvantage in cases where it is needed to understand how the systems arrive to a certain result, for example, when trying to predict occupants' actions.

Support vector machines (SVMs) are a set of related supervised learning algorithms used for classification and regression. They are linear classificators, i.e. algorithms that separate group based on the value of a linear combination of their similar features (Ayodele 2010). From a labeled dataset, a SVM obtains an optimal n-dimensional hyperplane which is used to categorize new cases. They are frequently used to detect if a case is similar to a particular class, for which, notably, they do not require training examples of all classes, but only that which is required to evaluate.

Cluster analysis, or simply *clustering*, as its name suggests, is a set of unsupervised learning methods that classifies a set of values into classes or clusters according to a certain criterion. The variety of criteria gives rise to different clustering algorithms. One of the most common algorithms is k-means, which divides a space into k clusters in which each case belongs to the cluster with the closest mean, in a sort of multidimensional Voronoi cells. This algorithm only describes convex clusters, although it can be paired with other algorithms to cope with this limitation. One of its main weaknesses is that it is very sensitive to outliers.

A *genetic algorithm (GA)* is, technically, a metaheuristic search algorithm based on the process of natural selection. Given a population and an optimization problem (expressed as an objective function or "fitness"), the algorithm samples a random population from cases that are evaluated, and then, stochastically, recombine them based on their basic parameters (or "genes"). The new generation is used to extract a new random sample. The process is repeated for a number of iterations ("generations") or until a critical value of objective function is reached. GA are very simple to implement, but they are very limited since they do not assure to obtain a global optimum, and even more, they tend to converge to local optima. They are often used to find approximate solutions within very wide search spaces of which their structure is unknown. In building design, they are very popular for the optimization of geometric problems.

2.5 Visualization

The process of data analysis is of a quantitative nature, which often creates a cultural barrier for users unfamiliar with these methods, particularly in design contexts. Therefore, a very important step in the process of working with big data is the communication and visualization of data in order to facilitate its interpretation. In the field of building design, and especially in terms of spatial information, data visualization plays a crucial role. In fact, the process of visualization can be itself a central part of the analysis when the way of presenting the data is so enlightening that leads to discover new meaningful knowledge not previously identified with solely quantitative methods.

This approach combining data visualization, data analytics, and human-computer interaction to facilitate reasoning, understanding, and decision making is called visual analytics (Keim et al. 2010).

There are a large number of digital visual analytics tools. Some are open-source, but require high-level expertise. The programming language R has become one of the most commonly used environments for data analysis, particularly in conjunction with dedicated graphical packages (e.g. ggplot). Other software solutions are commercial and very easy to use (e.g. Tableau, Spotfire, QlikView), but are usually limited in the flexibility of their analysis (Zhang et al. 2012).

3. APPLICATIONS

In this section, we review about 100 applications of big data in building design. The review is organized in 5 subsections following the traditional project development phases: pre-design, schematic design, design development & construction documents, construction, and operation. It should be noted, however, that this distinction of phases is a simplification that is typically due to administrative purposes. The design process is naturally more erratic and less defined. Design is not a self-contained activity performed only in the two "design" phases, but an interrelated activity performed throughout the entire process, thus, the need to examine all phases.

3.1 Pre-Design

Despite of its name, the *pre-design* phase truly represents the first activity of the design process. During this phase, the design problem is defined in terms of what is required (scope, goals, needs), and on what conditions (resources available, limitations, restrictions) (Hershberger 2015). Two important pre-design activities where big data is useful are *programming* and *site analysis*.

3.1.1 Programming

Programming is the identification and study of the needs and requirements underlying a design project. The scope of this task depends on the magnitude and complexity of the project, but in its most traditional form, it comprises the definition of spaces needed, building features, budget, timeframe, and construction conditions (Kumlin 1995). The main source of data for this phase is typically the self-perception of needs by the client, which is not necessarily an objective, precise, or sufficient source. Therefore, numerous programming guides strongly recommend designers collecting additional information: studying facility users, activities, and schedules (Cherry & Petronis 2016), investigate cultural and human values and issues through interviews and surveys (Hershberger 1999), or even conduct scientific experiments (Duerk 1993). Although undoubtedly this information would be extremely useful, it is often prohibitively difficult and expensive to obtain.

However, there are several sources of big data —maybe collected for other purposes— that can provide this information and much more. For example, data from social networks and geolocalized devices can be used to discover behavioral patterns and/or space use preferences, both at a personal and group level.

For instance, Davis (2016) shows two case studies where data gathered from building inhabitants through online apps was analyzed to define requirements for the design of new offices meeting spaces, and Sokmenoglu et al. (2010) investigated patterns and trends of socio-spatial activities of specific student communities in Istanbul by mining data from previous public surveys. Chen & Chiu (2006) proposed a systems to collect about pedestrian navigating patterns related to street design and urban furniture. Yan (2006) developed a system based on automated video tracking to study behavioral patterns of people in different architectural spaces. Ugarte & Leef (2016) described a clustering algorithm based on Instagram images to qualitatively assess the notion of proximity relations in a given area. Friedrich (2013) shows how to use social networks and data gathered at a big scale to massively aggregate and analyze opinions from future building users looking for trends that inform design. In fact, it is predicted that online networks and big data systems will be a catalyzer for crowd-sourced design approaches by enabling the use of participatory design methods over very large populations (Williams & Sotta 2015).

In fact, in other industries one of the main applications of big data is, precisely, to study and characterize relevant individuals (e.g. customers, patients). Aggregated transactional data from credit cards or from public online networks are widely used to segment markets and create personalized marketing campaigns (Grigsby 2016.; Adomavicius & Tuzhilin 2001; Linoff & Berry 2001; Artun & Levin 2015). A similar approach can be used to in building design to understand the preferences and needs of clients and occupants.

3.1.2 Site Analysis

Another important task during the pre-design phase is the study of the building site. The more information designers have, the better they can assess the constraints and opportunities of a building site. Typical information collected by designers include physical data (i.e. topographical surveys, geotechnical studies), environmental data (i.e. climate charts), legal antecedents (i.e. zoning, regulatory restrictions), infrastructural data (i.e. public services, road accessibility, traffic considerations), and even non-technical site data, such as prevailing views from the site or special cultural considerations of the surroundings (LaGro 2013; Zimmerman 2000). This information is usually dispersed among different sources, unorganized, incomplete, and, in the end, difficult to find and use.

Designers find in the combination of big data and *geographical information systems* (GIS) a convenient solution to collect, organize, and interpret substantially large volumes of geospatial information about a site, expanding the analysis of GIS far beyond traditional sources.

For example, CitysDK is an online map of all 9.8 million buildings in the Netherlands, with details year of construction, function, and area, built based on large public governmental sources (WAAG & Span 2016). Less comprehensive but more specialized, the Historic Quest DC is a map of 127,000 historic buildings in DC that includes historical and technical data for each building (HPO 2016). The NYC Parks Department has a map featuring every tree in New York City (+683,000), including specie, trunk diameter, photo, closest address, and ecological benefits (NYC Parks 2016). The MIT has a similar resource for 10 global cities (MIT SCL 2016).

In fact, the combination of big data GIS allows the analysis of almost any geolocalizable attribute: from crime (Igarapé Institute 2016), to household income of each census tract (ESRI 2016), and from property value (Petroutsos 2014) to most common languages spoken per street (Hubley 2016). A good compendium is MIT GeoWeb, an online tool to search, view, and download a repository of 2000+ layers of international GIS data from various public and academic sources (MIT GIS Services 2016). Using meteorological data from Hong Kong, Lou et al. (2016) build a model to predict the horizontal sky-diffuse solar irradiance for any specific site conditions.

This vast amount of geospatial big data allows designers to analyze a building site in ways that otherwise would be impossible, even with traditional GIS. For example, "street attractiveness" or "urban character" are two metrics that would be usually left to subjective expertise of a designer. Yet, with big data, more systematic, data-based methods could be utilized. Strelka (2016) shows a Russian study that analyzed 16,000 photos of Moscow posted in social networks to determine which streets are most chosen for walks around the city center and are most attractive for photos at various times of the day. With data from the U.S. National Trust for Historic Preservation, the Atlas of ReUrbanism maps the neighborhood "character", based on a series of objective metrics: median age of buildings, the diversity of age of the buildings, and the size of buildings and parcels (PGL 2016). Lopes et al. (2015) propose a method for the integration and synchronic multidimensional analysis and classification of public open spaces.

3.2 Schematic design

During the *conceptual* or *schematic design* phase, the general characteristics of the design are established, including scale, form, the size and organization of space, the general image of the building, and a preliminary estimate of construction costs (AIA 2014). Although this is a highly intuitive and expert-based task, several researchers have proposed data-driven methods that might assist the iterative process of generating, evaluating, and informing preliminary conceptual designs.

3.2.1 Generative Design Exploration

The early steps of the design process often involve the rapid generation of multiple design alternatives that are quickly evaluated and re-generated in an iterative process that helps designers to better understand the design problem and to find unexpected solutions.

Numerous researchers have faced the challenge of creating generative design tools that may assist this crucial but time-consuming task. Approaches attempted combinatorial solutions with search algorithms, but the results are not easily scalable, since the search space of all possible design options —even for the simplest design problem— is too large and makes the analysis computationally unfeasible (Evins 2013).

A different approach is to use big data analysis tools to analyze very large sets of combinatorial designs (but not all), statistically learn from them, and then generate or search for new solutions that meet certain characteristics learned. Massing studies, layout schemes, and circulation diagrams are three specific areas where good results have been achieved with this approach.

For example, Merrell et al. (2010) propose a system that, given a set of high-level requirements, synthesize a list of spaces and relationships between them using a Bayesian network trained on real-world data, and then generate a floorplan obtained through stochastic optimization. Michalek et al. (2002) proposed a method to generate automated layouts for rectangular single-story apartments by searching over the 2x106 designs using an evolutionary algorithm. Reynold et al. (2015) show a method to use machine learning (polynomial regression and decision trees) to assess early massing studies and provide immediate feedback to expedite design iterations. Mehanna (2013) proposed a system for conceptual structural design based on machine learning and evolutionary optimization. Erhan et al. (2014) proposed and approach for conceptual design exploration based on a similarity-based search algorithm that iteratively reduces and collapses design spaces into manageable scales.

3.2.2 Design Validation

Together with generating multiple design options, designers must ensure that designs fulfill the requirements stated in the programming phase, a process called *design* or *program validation*. For complex projects with many conflicting requirements, this is a daunting task that may be assisted by *automated rule checking systems*, i.e. programs that assess designs based on a predefined set of rules about the configuration of its objects, their relations, or attributes. These systems are widely used for code compliance analysis (Eastman et al. 2009), and increasingly used for checking several other requirements: space validation, circulation issues (wayfinding), ergonomic studies, constructability conditions. (Solihin & Eastman 2015; Nawari 2012; Pauwels et al. 2011).

One of the main difficulties in automated rule checking is obtaining and formalizing the rules themselves, even in cases where the "rules" are explicit, such as building codes. Typically, they need to be hard-coded based on the knowledge of expert practitioners, which is an extremely slow and restraining process. Just as a reference, in the US there are more than 60,000 different applicable building codes and standards (MADCAD 2016), a number than overwhelms any human capacity. Examining such amount of codes and extract interpretable rules is a task well suited for big data analysis techniques, particularly machine learning algorithms. For example, Niemeijer et al. (2014) shows a method for the analysis of natural language input from building code regulations and its automated transformation into computable design rules applicable to BIM models. A similar work was done by Yang & Xu (2004).

The sources of data do not need to be as structured as building codes, but, quite the contrary, completely unstructured. For example, rules can be extracted from validated designs (=labeled sets), or even natural language from designers' conversations. Krijnen & Tamke (2015) show a system to extract meaningful knowledge from sets of BIM models using supervised and unsupervised learning models. Fernando et al. (2010) proposed a system combining evolutionary algorithms and parametric design that extract a designer's intent based on their design history. Lee et al. (2005) proposed a system that extract circulation patterns of people within a space from mining past location data.

3.2.3 Precedents Analysis

An important activity conducted throughout the entire building design process, but especially intense during schematic design, is the study of *precedents*, which here refer to any project or building that share any characteristic/feature with the required building and whose study might serve as a guide for design (Clark & Pause 2012). Apart from their own professional experience, the main source of data for precedents are specialized journals and websites. For example, *ArchDaily*, the largest architectural online database, contain more than 300,000 images of about 20,000 buildings, and has more than 10 million visitors every day (Morales 2014; ArchDaily 2016). *Structurae*, an online database of structural engineering projects, features data about 68,000 structures (Structurae 2016). However, despite of these web databases are rich in content, they are poor in organization. In ArchDaily, for example, only text descriptors and basic categories (country, architect or building type) can be used to search precedents. Filters such as form, organization, program, or phenomenological qualities —so important during the early stages of design— are not available. Searching for relevant precedents is a slow and labor-intensive process that often must be prematurely shortened because of resources constraints.

Several researchers have proposed big data-based systems that, to a greater or lesser extent, assist the process of analyzing buildings, categorizing them, and extracting meaning information. Because a large part of building design information is stored in image format (e.g., floorplans, drawings, photographs), these systems are typically framed as computer vision methods for detecting meaningful features: spatial organization, spatial structure, or architectural style, to name a few.

For example, Lin (2011, & Chiu 2010) developed a system that uses k-means clustering methods to automatically and massively classify patterns of spatial topology of floor plans. Strobbe et al. (2016) addressed the problem of automatic detection of architectural designs belonging to a particular architectural style. They proposed the use of one-class SVMs with graph kernels to classify an unobserved floor plan as similar or different from a previously learned style. Ahmed et al. (2014) presented a method for automating the extraction of semantics from existing floor plans with the aim of developing a massive floor plan repository.

Similarly, but using other unstructured data, Mathias et al. (2012) proposed an algorithm which automates the classification of architectural styles from facade images. They used a SVM with a Gaussian radial basis kernel function to classify building in one of three predefined architectural styles. Römer & Plümer (2010) used SVMs with a radial basis function as kernel to classify low-level 3D models of buildings as consistent or not consistent with a learned style. Kuo (2003) proposed a scheme for analyzing Chinese gardens designs by mining textual data from traditional Chinese design books, and Reffat (2008) discussed how to mine public data from municipalities to identify the patterns of contemporary architecture in Saudi Arabia.

The study of precedents is not exclusive of architectural design. All building designers always instinctively look for precedents when face a new project. In engineering, previous cases are equally important for determining benchmarks, documenting successes and failures, and extracting proved solutions. For example, in energy analysis, Park et al. (2016) analyzed 1,072 office buildings in South Korea using correlation analysis, decision trees, and traditional statistical analysis to develop a new energy benchmark for improving the operational rating system of office buildings. Similarly, Yalcintas (2006; & Aytun Ozturk 2007) developed an energy benchmarking model based on ANNs for different cases. Ma & Cheng (2016) analyzed 1,000 projects certified by LEED-NC v3 in order to understand or discover patterns in the achievement of LEED credits. Mathew et al. (2015) discussed technical considerations in compiling a large database of building energy use, based on the pilot experience of analyzing data from 750,000 residential and commercial buildings from the DOE Buildings Performance Database.

3.3 Design Development & Construction Documentation

During the design development and construction documents phases is when most of project information is produced. However, only a minor fraction is properly collected, organized, and stored. Most information is non-systematized, extremely scattered, erroneous, or simply lost. The lack of usable data is, indeed, one of the urgent problems for developing data-driven approaches in the AECO industry.

3.3.1 Managing Building Data

The main data management technology in the AECO industry is *building information modeling* (BIM), i.e. integrated and object-oriented databases of all physical and functional characteristics of a building throughout its complete lifecycle (Eastman 2011). As such, it is per se the epitome of a structured database for building design and has the greatest potential today to become the basis of all building data-driven processes. However, in practice, BIM technology is mostly underutilized in relation to its potential uses. It is widely accepted that the largest benefits of BIM are achieved when all participants use the models as integrated repositories of all building data throughout the complete lifecycle of the project. This type of advanced use has been documented in only a few countries, notably United Kingdom, United States, Singapore, Finland, Norway, and Denmark. In most other countries, BIM is still at its infancy, with uses focused on visualization, production of construction documents, basic systems coordination, and quantity take-off (McGraw-Hill Construction 2009, 2010, 2012a, 2012b, 2014a, 2014b; Loyola 2016, 2018; NBS 2018)

Several researchers have proposed frameworks and methods based on big data and machine learning to enhance BIM models with more data, transforming into true whole building databases.

Krijnen & Tamke (2015) showed how to assess implicit knowledge in BIM models with supervised and unsupervised machine learning methods based on IFC geometrical data. Lan & Chiu (2005) applied data mining techniques to discover the design semantic patterns in communication information within teams working on collaborative design projects. Koch & Firmenich (2011) proposed a process-oriented data management that could take advantage of the knowledge produced during the design process but that is not necessarily encapsulated in the final design outcome. Ekholm (2001) and Simeone et al. (2013) proposed to integrate BIM with a knowledge base (as semantic ontologies) in order to represent not only building components and spaces, but also users, activities and the relations among them. Chen et al. (2016), Amarnath et al. (2011), Chuang et al. (2011), and Jiao et al. (2013) have explored different approaches and possibilities for interconnecting BIM models hosted online, forming a large collaborative information system with continuous real-time dynamic data from sensors.

3.3.2 Energy Studies

By far, energy studies represent the area of building design where big data has been most successfully used. Given the considerable extent of work, there are many reviews available that offer an overview of the current state:

Magoulès & Zhao (2016) provided a comprehensive review of machine learning methods for building energy analysis, with an emphasis of ANNs and SVMs. Foucquier et al. (2013) reviewed the state of the art in building modeling and energy performances prediction, comparing physical model and machine learning models. Zhao & Magoulès (2012) reviewed recently developed statistical models and artificial intelligence methods for the prediction of building energy consumption. Yu et al. (2016) provided an overview of the use of predictive and descriptive studies using big data in energy studies. Yu et al. (2013) reviewed commonly used data mining methods for extracting knowledge from building energy data. Borgstein et al. (2016) provided a review of available methods for analyzing, classifying, benchmarking, rating and evaluating energy performance in non-domestic buildings.

Kalogirou (2000, 2001) reviewed the applications of artificial neural networks in the design and evaluation of renewable energy systems.

Specifically regarding the use of big data for design optimization, Evins (2013) reviewed computational optimization methods applied to sustainable building design including common heuristics and evolutionary algorithms. Kalani et al. (2016) reviewed different studies on optimization of building energy consumption through data mining, and Machairas et al. (2014) reviewed several algorithms for optimization of building design.

In general terms, big data and machine learning in energy analysis are increasingly being used in the analysis, simulation, and prediction of phenomena in which there are not established physical models that can be used. Big data and machine learning operate on the basis of statistical models, so there are no need thermal equations, lighting behavior assumptions, acoustic properties data, or geometrical parameters. Therefore, it is well suited to modeling energy demands, predicting user behavior, or optimizing complex multi-variable designs where creating a traditional physical model might be too complex, imprecise, impractical, or even unfeasible.

For example, Castelli et al. (2015) proposed a model for predicting energy needs for residential buildings using a type of evolutionary algorithms called geometric semantic genetic programming. Capozzoli et al. (2015) analyzed data from around 90,000 dwellings in order to detect heating and domestic hot water primary energy demands and to identify the most influencing factors. Tsanas & Xifara (2012) developed a statistical machine learning framework based on random forests to study and predict the effect of physical variables on heating and cooling load of residential buildings. Kalogirou (2000) utilized ANNs to predict the energy consumption of a passive solar building, based on easily measurable physical qualities easily measurable. Yu et al. (2010) proposed a decision tree to estimate residential building energy performance indexes by modeling building energy use intensity levels. Li et al. (2011) proposed a hybrid system between a genetic algorithm and an adaptive network to predict the energy consumption in a case study in China. Hyunjoo Kim et al. (2011) proposed a data mining technique for discovering interrelationships between different systems to improve building design. Su & Yan (2014) used genetic algorithms to optimize layout and daylighting performance, effectively reducing computing time in comparison to the traditional approach. Feng et al. (2014) proposed a model that use cluster analysis and association rules to identify valid window behavioral/operational patterns that were later associated to different natural ventilation strategies and building design recommendations.

Advanced computational energy simulations have high computational requirements and may pose a bottleneck in the design process. Chatzikonstantinou & Sariyildiz (2016) explored an alternative approach, building a prediction model for visual comfort of office space based feed-forward networks, SVMs and random forests, trained with simulation-derived data. De Wilde et al. (2013) assessed the use of building simulation data to deliver accurate SVMs used in energy management of actual buildings.

Because the use of big data and machine learning is a relatively recent approach, there is still little clarity about the strengths or weaknesses of each method. A number of important studies have been devoted to comparing different. Mateo et al. 2013) compared a wide selection of analysis methods (autoregressive models, robust multiple linear regression, multilayer perceptron, extreme learning machine, non-linear autoregressive exogenous, k-means, fuzzy c-means, cumulative hierarchical tree, DBSCAN) to classic techniques in building temperature prediction. Neto & Fiorelli (2008) contrasted a ANN-based model to a physical model in the prediction of building energy consumption. Fonseca et al. (2013) compared the use of ANNs to multivariate linear regression to predict the impact of daylighting on building final electric energy requirements in office buildings.

3.3.3 Structural Design

Structural design is recognized as a data-driven activity and, as such, the use and application of big data analysis methods appears more natural than in other building design domains.

One application gaining exponential notoriety is *structural health monitoring* (SHM), i.e. the use of sensor devices in actual structures to collect data to detect and characterize damage. A complete review covering SHM from a machine learning perspective is found in Farrar & Worden (2012).

Similarly to energy studies, big data is remarkably helpful to analyze phenomena in which there is not complete physical understanding based on statistical learning over very large datasets of real and simulated structures. For example, several researchers have developed algorithms to model complex indeterminate structures. Mehanna (2013) explored the use of big data and ANNs to develop a system capable of teaching lightweight, flexible, and unanchored structures to self-rectify after falling through their interactions with their environment. Yamamoto et al. (2011) proposed a genetic algorithm to assist designers to obtain tensegrity structures with less design variables, requiring only a minimal knowledge of the initial structure configuration.

To a lesser extent, structural design shares the same problems of lack of systematization of expert knowledge than other building design disciplines, and big data can also play a role in solving this issue. For example, Dolšak & Novak (2011) proposed a framework for organizing structural knowledge and develop an intelligent decision support system. Freischlad et al. (2006) proposed a genetic algorithm for the data-driven extraction of fuzzy rules for the design of reinforced concrete structural members.

3.4 Construction

Although initially design and construction appear as two separate consecutive phases, they are, in fact, two closely related and are mutually dependent activities. Buildings undergo constant design modifications informed by field data during their construction, in a bidirectional and continuous process that only stabilizes when the building is delivered to the client.

3.4.1 Constructability Analysis

The most iconic application of construction data in building design is represented by constructability reviews.

Constructability (also *buildability*) is a concept coined in the 1980s to refer to the extent to which a [building] design facilitates its construction, subject to all requirements of the client (CIRIA 1983, CII 1986). In order to improve the constructability of a project, designers must consider all the particularities of the construction process, which is usually a type of knowledge that resides exclusively on builders or very experienced designers. Constructability guides are normally structured as general guidelines or rules-of-thumb which are either too generic to be useful or too specific to be widely applicable (Adams 1989; Lam et al. 2006).

Several researchers have proposed systems that take advantage of big data and machine learning methods to acquire construction knowledge from unstructured sources and generate structured constructability rules. For example, Skibniewski et al. (1997) conducted a feasibility study of automated constructability knowledge acquisition in the form of decision rules using as database of 31 construction example projects. Freischlad & Schnellenbach-Held (2005) proposed a system for the acquisition and representation of structural design knowledge using linguistic fuzzy modeling and multi-objective evolutionary algorithms.

3.4.2 Construction Engineering

Slightly tangent to building design, construction engineering is a good example of the wide variety of applications of big data and machine learning in context of rapidly mutating data, ranging from monitoring construction sites to predict worker's injuries. A very good review is provided by Bilal et al. (2016).

For example, Lu et al. (2015) examined more than 2 million waste disposal records from thousands of projects in two consecutive years to build waste generation rate indicators and benchmarks for different categories of projects. Asadi et al. (2015) used a decision tree a Naive Bayes algorithm for predicting delays in construction logistics. Yang et al. (2015) reviewed several approaches to construction monitoring using big data and machine learning. Dimitrov & Golparvar-Fard (2014) used SVMs to classify unordered site image collections to build a tool of vision-based material recognition system for automated monitoring of construction progress. Li et al. (1999) used machine learning and genetic algorithms to solve time-cost trade-off problems. Milion et al. (2016) proposed a method for estimating the consumption of electrical materials in construction, based on ANNs by using information from early project stages. Tixier et al. (2016) presented an application of random forests to predict construction injuries. Cheng & Wu (2009) combined genetic algorithms and SVMs to propose an evolutionary support vector machine inference model used to solve problems in construction management. Kim & Teizer (2014) proposed a method for automated data-driven design of scaffolding.

3.5 Operation

Data collected after construction, i.e. during the operation of the building, are, perhaps, the least used in design. Perhaps the two most formally established methods for collecting data rigorously from operative buildings are *building commissioning* and *post occupancy evaluations*.

Building commissioning is the process of verifying that "all" technical subsystems operate as designed and achieve the project owner's requirements (Grondzik 2009). In practice, the commissioning process has usually a much narrower scope, limited to building systems, energy consumption, and overall operating costs (Katipamula & Brambley 2005; Djuric & Novakovic 2009). A post-occupancy evaluation (POE) is a more comprehensive approach, but much less common. Is formally defined as the set of systematic procedures for collecting data on the ground to measure the performance of a building and to construct it with the design simulations. POE comprises from checking systems performance to keep operating costs under control, to the complete monitoring, evaluation and validation of long-term design intentions (Zimring & Reizenstein 1980; Preiser 1995; Zimmerman & Martin 2001). Nevertheless, in both cases, the primary goal is to gather enough data for purposes of immediate problem solving or troubleshooting any potential issue. Very rarely the objective is to document successes and failures to feedback designers and increase their body of knowledge for future projects. As a matter of fact, the cost of collecting representative data for operating buildings can be so prohibitively expensive and complex, that may be not worth any meaningful knowledge obtained.

Big data technology comes to challenge this situation. Two applications where this is most evident is *building performance monitoring* and *occupancy modeling*.

3.5.1 Building performance monitoring

The emergence of *Internet of Things* (IoT) technologies offers major opportunities to expand these types of assessments. The lower costs of sensors, the availability of wireless networks, the expansion of mobile devices, the use of open data standards, and the incorporation of embedded digital technologies in buildings (for different purposes) have substantially reduced the complexity and cost of data collection. In fact, a lot of data is already being collected for use in other contexts (e.g. by Building Management Systems, BMS, or Building Automation Systems, BAS), but have not yet been used by building designers. These continuously operating sensors are producing a stream of data that, when is properly analyzed, is admittedly much more valuable than its current limited applications (Dibley et al. 2011).

For example, Peña et al. (2016) proposed a rule-based system to detect energy efficiency anomalies in smart buildings, using data mined from a full set of building sensors. Capozzoli et al. (2015) proposed a data mining system for fault detection in office buildings. Kim et al. (2011) showed a method of analysis of an energy efficient building design through data mining approach. Reffat & Gero (2005) proposed a virtual environment integrating database management systems with object-based CAD systems and 3D virtual environments to mine operational data to support decision making in building maintenance.

3.5.2 Occupancy modeling

Most building monitoring systems focus on infrastructure, building systems, and energy performance. Much less are concerned with analyzing how people interact with spaces. Yet, understanding occupancy patterns is key in predicting building performance. Birt & Newsham (2009) showed how designers are usually wrong about their assumptions of the behavior of the occupants.

Several researchers have proposed methods for detecting occupancy patterns using big data collected from building sensors. Most of them have energy performance motivations. For example, D'Oca & Hong (2015) used building big data streams collected through 2 years to obtain patterns of occupancy schedules that can be used in energy simulations. Edwards et al. (2012) assessed seven different machine learning algorithms applied to big datasets of building measurements with the objective of determining which techniques are most successful for predicting next hour residential building consumption. Zhao et al. (2014) developed an "indirect" practical data mining approach using office appliance power consumption data to learn the occupant "passive" behavior. Liang et al. (2016) proposed a model for predicting occupancy schedules in office buildings.

Other researchers have focused on discovering more complex occupancy patterns. Simeone & Kalay (2012) used agent-based modeling techniques derived from videogames to simulate human behavior. Williams et al. (2014) combined methods and techniques from sensor networks, signal processing, data mining, network theory, and information visualization to propose a framework that facilitates the understanding social behaviors in indoor environments. Tomé et al. (2015) proposed a method for the analysis of the space–use interactions patterns obtained by data fusion of video plus RFID inputs, linking their method to wayfinding studies.

Understanding occupancy may be useful to understand how building are being utilized, even in cases in which there is no previous knowledge about the building. *Indoor positioning systems*, for example, may be an indirect path to analyze buildings. Alzantot (2012; & Youssef 2013) presented a crowdsourcing-based system for the automatic construction of buildings floorplans based on the aggregated use of smartphones sensors of building occupants and standard decision tree and clustering algorithms. Shin et al. (2012) proposed a similar work, automatically constructing indoor floor plans for anonymous buildings using odometry tracing from smartphones inertial sensors.

4. DISCUSSION

Compared to other design disciplines, the use of big data in building design is still in its infancy. Yet the dozens of examples reviewed in this survey show that, very slowly, the technology is beginning to produce changes in the AECO industry. In the next paragraphs, we enunciate the transformative impact of big data in the AECO industry on three promising niches of application. In general, the definition of these niches responds mainly to the greater number of reviewed papers focused on these issues, and more specifically, the number of papers that suggested areas and topics where more research is urgently needed. The three niches are organized in a similar way to how the paper is structured, that is, based on the traditional design phases as understood in the industry: pre-design (i.e. using big data to understand the design problems); during schematic design and design development (i.e. using big data to simulate the design options and get feedback), and during construction and operation (i.e. using big data to evaluate continuously the design products). However, it is important to emphasize that this division is a simplification made only for consistency and clarity purposes, but in reality, the design activity is naturally more complex and less compartmentalized.

4.1 Understanding Design Problems

One of the most important factors to consider when designing a building is the needs of future occupants. And yet, ironically, it is perhaps one of the topics on which less information is available to designers. Data about the site, climate, legal requirements, construction costs, or even current architectural trends are typically easier to obtain, and therefore more abundant, that than data on the specific characteristics of future users. The main source of data for understanding the requirements that define the design problem is typically the client's own self-perception of needs, which is not necessarily an objective, precise, or sufficient source. Not only that, previous research has shown that both clients and designers are usually wrong about their own assumptions.

In the most common scenario, designers are forced to resort to generic data that is "adjusted" or "tailored" to each case to the best of their expertise and intuition. From room sizes (i.e. graphic standards) to comfort levels (i.e. ASHRAE standards), building design heavily relies on descriptors of "average users". For an industry whose main value is, precisely, the production of "customized" works, this represents a vital contradiction.

Big data technology can provide abundant data about future occupants and their habits and preferences, allowing designers to understand their needs with an unparalleled level of detail. In fact, the fine-grained characterization of relevant individuals ("customers" in retail, "patients" in healthcare) is one of the most common uses of big data in other industries. It is reasonable to think that the same approach could be used in the AECO industry, where our "relevant individuals" are the building occupants.

For example, data from social networks and geolocalized devices can be used to discover behavioral patterns of movement and permanence. Data from buildings embedded sensors can be used to extract space use preferences. Data can be aggregated to a larger scale to exact information that is not representative of only one case, but of a given subset of people, or generalizable to larger populations.

Commercial buildings can be designed based on customer preferences and consumer behavior. Schools can be designed based on the analysis of spatial conditions that are most strongly associated with good academic performance metrics. Meeting spaces can be designed based on the cultural differences of different social groups. Comfort conditions can be adjusted to the specific thermal preferences of a target group. Even an ordinary person who wants a particular building (e.g. a home) can learn from himself (of his habits, his preferences), more objectively than mere self-perception.

However, it is still not clear how this user characterization process should be carried out. There is too much data, much more than what is needed, and we do not know what data are the most useful in building design. Other industries have already identified what raw data work best for them. There are isolated technologies that can be used, from sentiment analysis to indoor positioning systems, but there are no frameworks or methodologies to combine them into a unitary system for understanding user spatial preferences. As evidenced in this review, current applications are narrow in scope with little practical applicability. Finally, there are also legitimate concerns about the security and privacy of personal data used for these purposes.

4.2 Simulating design solutions

According to Lawson (1990), "the best test of most designs is to wait and see how well they work in practice". Indeed, while the are very accurate simulation systems in specific engineering domains (e.g. energy or structural performance), there are few systems capable to simulate a more holistic building performance.

Big data opens up the opportunity for new, more complex simulations. Coupled with the ability to obtain accurate information about the design problem, there is the ability to use that information to simulate the performance of the design solution.

Perhaps the most powerful application today is the combination between building analytics and people analytics for simulating human behavior in buildings. This has been successfully done in contexts where human behavior is highly predictable, such as the simulation of emergency evacuations during a fire. In other cases where human behavior is much more stochastic and there are no models for predicting it, machine learning methods stand out.

By crossing data from hundreds or thousands of cases, it is possible to create statistical models of people movement and permanence within a space. In this review, we have seen examples with office spaces and courthouses that outperform greatly the more widespread agent-based simulations. However, the same principle could be extended to more complex situations such as public spaces, large-scale buildings, or social gathering spaces. The simulations could be adjusted to specific segments of users, and detect, for example, patterns of space use differentiated according to social, cultural characteristics or environmental conditions. Designers could learn what design features make spaces subutilized or overutilized and could maximize the efficiency of space use for their designs. The value of these simulations extends beyond sole space planning and be used to simulate energy demands, determine structural overloads, predict acoustic conditions, or among several other applications.

Some designers are reluctant to this type of simulations because they consider they represent a deterministic approach to design. Quite the contrary. The construction industry is conservative and risk-minimizing when facing the uncertainty of how a new design will perform, and regularly tends towards known solutions. A holistic simulation system would provide the client with objective arguments to prefer an innovative design over a known but less efficient solution.

To achieve this goal, there are still many research obstacles to be overcome. One particularly challenging is the lack of robust standards to integrate data from different sources into unified models for simulations. Notable advances have been made with Industry Foundation Classes (IFC) and Green Building XML (gbXML), although they are limited when manipulating really large and rich datasets. Developing a robust standard for big data is challenging because addressing requires not only an AECO industry-wide effort, but also the participation of electronics and IT industries. More importantly, it is not clear how to create open simulation processes that are useful and meaningful to different types of projects, and can be appropriately compared across different contexts. It is also necessary to create metrics and protocols to evaluate the quality of the simulations in a way that does not need the building in operation as point of comparison. There are also doubts on how data collected in one context may be generalized or transferred to another context, or how massive or specific the data collection must be in order to have results that are not only significate but meaningful for design.

4.3 Evaluating continuously design products

Designers are especially persuasive when making promises or predictions about their designs. Architects, for example, often use "photorealistic" images to illustrate how a proposed design will look and perform. Yet it is much less common for such images and predictions to be effectively tested in real life. Current building assessment practices, such as post-occupancy evaluations, are limited to mainly engineering performance indicators. Other systems to analyze more comprehensively the performance of a building –in the widest sense of the term– are normally too complex to be feasibly implemented.

With big data technologies, these assessments can go beyond traditional energy, structural, or operational cost concerns. For example, clients could assess space use efficiency, user comfort levels, average circulation traffic speed, or even, very complex metrics like social interaction or emotional comfort. In this review we have seen experimental systems that display the overall "mood" of occupants while they move around using surveillance video cameras, systems that can track individual comfort levels throughout the day for every worker in an office building using institutional apps, and systems to assess the performance of air conditioning systems in gyms using data from users' sport wristbands. In fact, one of the main drivers of the use of big data technologies to evaluate buildings is the increasing availability of data provided by electronics already embedded in buildings for other reasons (security control, environmental systems automation, domotics, etc.). The data are available and waiting to be used.

The possibility of collecting so many different types of data necessarily leads to the problem of redefining the concept of building performance. Different data enable the creation of different metrics. In fact, one of the biggest problems with design promises that are never evaluated in practice and with designers who are not held accountable, is simply the lack of agreed metrics and protocols to assess complex and multivariable architectural concepts.

The wide range of data available is a patent opportunity to rethink new broader and deeper assessments that are not feasible with traditional means. Concepts that have traditionally considered subjective, such as façade attractiveness or room coziness, can be operationalized based on observable indicators, from the number of selfie pictures with a given façade to the median permanence time in a room. Although this is undoubtedly a fascinating and crucial topic for the discipline and the profession –that certainly triggers antagonistic positions–, the research in this area is surprisingly scarce.

The assessment of a design product can not only be conducted to test the actual building against the original design intent, but also during the design process as a method to inform design iterations, and also after the delivery of the building as a continuous evaluation process during the entire lifecycle of a facility.

As a matter of fact, once a building is delivered to the client, the relationship between designers and design products is typically terminated. The feedback of information from the occupants or clients to designers is minimal. In practice, designers not only are unaware of how their designs performed, but they also waste a valuable opportunity to learn what they could have been done better.

As the availability of building electronics steadily increases, continuous and permanent building monitoring and assessment become feasible, which in turn, leads to buildings that could be regularly improved based on the analysis of continuous streams of information. Buildings may be understood not as static "completed projects", but as dynamic objects that constantly evolve and change in time informed by data. The relationship between designers and operational buildings can be explicit and permanent. On the one hand, designers could learn from building in a systematic and unbiased manner, and on the other hand, clients and occupants could be benefited with designs and spaces that better fit their constantly changing needs.

5. CONCLUSIONS

In this paper, we have reviewed the current state and future directions of big data technologies in building design.

The survey shows that the range of applications covers the entire spectrum of activities throughout the design process, from pre-design to building operation. The greatest progress and number of examples today exist in those activities that are traditionally understood as data-driven, such as energy analysis or structural design. Yet, it is precisely those activities that rely not on data but on intuition or tacit expert knowledge that have space for more interesting and disruptive opportunities to use big data. The implicit promise of big data analytics is to inform design with such unparalleled level of specificity, objectivity, and insightfulness, that they can de facto dismantle the dependency on generic data and anecdotal evidence that distinguishes building design. Big data does not exclude intuition and expert knowledge, but inform intuition and expert knowledge.

Three application niches where more research is urgently needed are 1) understanding and defining the design problem by mining particularized data of the client, occupants, site, and other design conditions; 2) simulating and predicting the holistic performance of complex design solutions without the need of explanatory physical models by using statistical models obtained with machine learning methods; 3) evaluating integrally and continuously the design product during its entire lifecycle with new building metrics that take advantage of advanced data collecting and analysis methods.

6. REFERENCES

Adams, S., 1989. Practical buildability. CIRIA.

Adomavicius, G. & Tuzhilin, A., 2001. Using Data Mining Methods to Build Customer Profile. Computer, 34(3), pp.74–82.

Ahmed, S. et al., 2014. Automatic analysis and sketch-based retrieval of architectural floor plans. Pattern Recognition Letters, 35(1), pp.91–100.

Alekhya, G.S.S.L.A., Lydia, D.E.L. & Challa, D.N., 2016. Big Data Analytics: A Survey. International Journal of Application or Innovation in Engineering & Management, 5(10), pp.2319–4847.

Alpaydin, E. 2014. Introduction to machine learning. MIT press.

Alzantot, M. & Youssef, M., 2013. Demonstrating CrowdInside: A system for the automatic construction of indoor floor-plans. In 2013 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2013. pp. 321–323.

Amarnath, C.B., Sawhney, A. & Uma Maheswari, J., 2011. Cloud computing to enhance collaboration, coordination and communication in the construction industry. In 2011 World Congress on Information and Communication Technologies. IEEE, pp. 1235–1240.

American Institute of Architects [AIA], 2014. The Architect's Handbook of Professional Practice,

ArchDaily, 2016. ArchDaily, the world's most visited architecture website. Available at: http://www.archdaily.com/content/about

Artun, O., & Levin, D. 2015. Predictive marketing: Easy ways every marketer can use customer analytics and big data. John Wiley & Sons.,

Asadi, A., Alsubaey, M. & Makatsoris, C., 2015. A machine learning approach for predicting delays in construction logistics. International Journal of Advanced Logistics, 4(2), pp.115–130.

Ayodele, T.O., 2010. Types of Machine Learning Algorithms. In Y. Zhang, ed. New Advances in Machine Learning. pp. 19–49.

Bellazzi, R., 2014. Big data and biomedical informatics: a challenging opportunity. Yearbook of Medical Informatics, 9, pp.8–13.

Bendre, M.R. & Thool, V.R., 2016. Analytics, challenges and applications in big data environment: a survey. Journal of Management Analytics, 3(3), pp.206–239.

Bi, Z. & Cochran, D., 2014. Big data analytics with applications. Journal of Management Analytics, 1(4), pp.249–265.

Bilal, M. et al., 2016. Big Data in the construction industry: A review of present status, opportunities, and future trends. Advanced Engineering Informatics, 30(3), pp.500–521.

Binnekamp, R., 2010. Preference-Based Design in Architecture 1st ed., Amsterdam: Delft University Press - IOS Press.

Binnekamp, R., Van Gunsteren, L.A. & Van Loon, P.-P., 2006. Open Design, a Stakeholder-oriented Approach in Architecture, Urban Planning, and Project Management, Amsterdam: Delft University Press.

Birt, B.J. & Newsham, G.R., 2009. Post-occupancy evaluation of energy and indoor environment quality in green buildings: a review. Proceedings of the 3rd International Conference on Smart and Sustainable Built Environments, Delft, the Netherlands.

Borgstein, E.H., Lamberts, R. & Hensen, J.L.M., 2016. Evaluating energy performance in non-domestic buildings: A review. Energy and Buildings, 128, pp.734–755.

Brandt, R.M., Chong, G.H. & Martin, W.M., 2010. Design informed : driving innovation with evidenced-based design, John Wiley & Sons.

Brown, R.D. & Corry, R.C., 2011. Evidence-based landscape architecture: The maturing of a profession. Landscape and Urban Planning, 100(4), pp.327–329.

Bryant, R.E., 2011. Data-intensive scalable computing for scientific applications. Computing in Science and Engineering, 13(6), pp.25–33.

Capozzoli, A. et al., 2015. Discovering knowledge from a residential building stock through data mining analysis for engineering sustainability. Energy Procedia, 83, pp.370–379.

Capozzoli, A., Lauro, F., & Khan, I. 2015. Fault detection analysis using data mining techniques for a cluster of smart office buildings. Expert Systems with Applications, 42(9), 4324-4338.

Castelli, M. et al., 2015. Prediction of energy performance of residential buildings: A genetic programming approach. Energy and Buildings, 102, pp.67–74.

Center for Health Design [CHD], 2016. Knowledge Repository. Available at: https://www.healthdesign.org/knowledge-repository

Chatzikonstantinou, I. & Sariyildiz, S., 2016. Approximation of simulation derived visual comfort indicators in office spaces a comparative study in machine learning. Architectural Science Review,, 59(4), pp.307–322.

Chen, C.H. and Chiu, M.L., 2006. Space Tags and User Behavior Modeling-Applying agents to detect navigational patterns in urban streets. Education and Research in Computer Aided Architectural Design in Europe ecaade (2006).

Chen, C.L.P. & Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, pp.314–347.

Chen, H., Chiang, R. & Storey, V., 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly.

Chen, H.M., Chang, K.C. & Lin, T.H., 2016. A cloud-based system framework for performing online viewing, storage, and analysis on big data of massive BIMs. Automation in Construction, 71, pp.34–48.

Chen, M., Mao, S. & Liu, Y., 2014. Big data: A survey. In Mobile Networks and Applications. pp. 171–209.

Cheng, M.-Y. & Wu, Y.-W., 2009. Evolutionary support vector machine inference system for construction management. Automation in Construction, 18(5), pp.597–604.

Cherry, E. & Petronis, J., 2016. Architectural Programming. WBDG Whole Building Design Guide..

Chong, D. & Shi, H., 2015. Big data analytics: a literature review. Journal of Management Analytics, 2(3), pp.175–201.

Chuang, T.H., Lee, B.C. and Wu, I.C., 2011. Applying cloud computing technology to BIM visualization and manipulation. In 28th International Symposium on Automation and Robotics in Construction (Vol. 201, No. 1, pp. 144-149).

CIRIA, 1983. Buildability: An Assessment., London.

Clark, R.H. & Pause, M., 2012. Precedents in architecture : analytic diagrams, formative ideas, and partis, John Wiley & Sons.

Construction Industry Institute [CII], 1986. Constructability: A Primer, Austin:

D'Oca, S. & Hong, T., 2015. Occupancy schedules learning process through a data mining framework. Energy and Buildings, 88, pp.395–408.

Davenport, T.H. & Patil, D.J., 2012. Data scientist: the sexiest job of the 21st century. Harvard business review, 90(10).

Davis, D. 2016. Evaluating Buildings with Computation and Machine Learning.

De Mauro, A., Greco, M. & Grimaldi, M., 2015. What is big data? A consensual definition and a review of key research topics. AIP Conference Proceedings, 1644, pp.97–104.

Deutsch, R., 2015a. Data-driven design and construction : 25 strategies for capturing, analyzing and applying building data 1st ed., Hoboken, New Jersey Published: John Wiley & Sons, Inc.

Deutsch, R., 2015b. Leveraging data Across the Building Lifecycle. Procedia Engineering, 118, pp.260–267.

Devisch, O.T.J., Timmermans, H.J.P., Arentze, T.A. and Borgers, A.W.J., 2006. Modelling Residential Search and Location Choice-Framework and Numerical Experiments. Progress in Design & Decision Support Systems in Architecture and Urban Planning, Eindhoven: Eindhoven University of Technology.

Dibley, M.J. et al., 2011. Towards intelligent agent based software for building related decision support. Advanced Engineering Informatics, 25(2), pp.311–329.

Dijkstra, J. and Timmermans, H., 2002. Towards a multi-agent model for visualizing simulated user behavior to support the assessment of design performance. Automation in construction, 11(2), pp.135-145.

Dimitrov, A. & Golparvar-Fard, M., 2014. Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. Advanced Engineering Informatics, 28(1), pp.37–49.

Djuric, N. & Novakovic, V., 2009. Review of possibilities and necessities for building lifetime commissioning. Renewable and Sustainable Energy Reviews.

Dolšak, B. & Novak, M., 2011. Intelligent decision support for structural design analysis. Advanced Engineering Informatics, 25(2), pp.330–340.

Douglas, L., 2001. 3d data management: Controlling data volume, velocity and variety. Application Delivery Strategies, 6, p.4.

Duerk, D.P., 1993. Architectural Programming: Information Management for Design, Wiley.

Eastman, C. et al., 2009. Automatic rule-based checking of building designs. Automation in Construction, 18(8), pp.1011–1033.

Eastman, C.M., 2011. BIM handbook : a guide to building information modeling for owners, managers, designers, engineers and contractors, Wiley.

Edwards, R.E., New, J. & Parker, L.E., 2012. Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, 49, pp.591–603.

Ekholm, A., 2001. Modelling of user activities in building design. Proceedings of the 19th Conference on Education in Computer Aided Architectural Design in Europe, pp.67–72.

Ellis, G. and Mansmann, F., 2010. Mastering the information age solving problems with visual analytics. In Eurographics (Vol. 2, p. 5).

Erhan, H., Wang, I. and Shireen, N., 2014. Interacting with thousands: A parametric-space exploration method in generative design. In Proceedings of the 2014 Conference of the Association for Computer Aided Design in Architecture. ACADIA.

ESRI, 2016. Wealth Divides. Available at: http://storymaps.esri.com/stories/2016/wealth-divides/index.html

Evins, R., 2013. A review of computational optimisation methods applied to sustainable building design. Renewable and Sustainable Energy Reviews, 22, pp.230–245.

Fang, B. & Zhang, P., 2016. Big data in finance. In Big Data Concepts, Theories, and Applications. pp. 391–412.

Farrar, C.R. and Worden, K., 2012. Structural health monitoring: a machine learning perspective. John Wiley & Sons.

Feng, W. et al., 2014. A Data-mining Approach to Discover Patterns of Window Opening and Closing Behavior in Offices. Journal of Building and Environment, 82(May), pp.726–739.

Fernando, R., Drogemuller, R., Salim, F. and Burry, J., 2010. Patterns, heuristics for architectural design support: making use of evolutionary modelling in design. In New Frontiers: Proceedings of the 15th International Conference on Computer-Aided Architectural Design Research in Asia (pp. 283-292). Association for Research in Computer-Aided Architectural Research in Asia (CAADRIA.

Fonseca, R.W. da, Didoné, E.L. & Pereira, F.O.R., 2013. Using artificial neural networks to predict the impact of daylighting on building final electric energy requirements. Energy and Buildings, 61, pp.31–38.

Foucquier, A., Robert, S., Suard, F., Stéphan, L. and Jay, A., 2013. State of the art in building modelling and energy performances prediction: A review. Renewable and Sustainable Energy Reviews, 23, pp.272-288.

Freischlad, M. & Schnellenbach-Held, M., 2005. A machine learning approach for the support of preliminary structural design. Advanced Engineering Informatics, 19(4), pp.281–287.

Freischlad, M., Schnellenbach-held, M. & Pullmann, T., 2006. Evolutionary Generation of Implicative Fuzzy Rules for Design Knowledge Representation. In Intelligent Computing in Engineering and Architecture: 13th EG-ICE Workshop 2006, Ascona, Switzerland. Springer Berlin Heidelberg, pp. 222–229.

Friedrich, P., 2013 Web-based co-design Social media tools to enhance user-centred design and innovation processes.

Fuchs, M., Hopken, W. & Lexhagen, M., 2014. Big data analytics for knowledge generation in tourism destinations - A case from Sweden. Journal of Destination Marketing and Management, 3(4), pp.198–209.

Gandomi, A. and Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), pp.137-144.

Glasgow, K., 2015. Chapter 4 - Big Data and Law Enforcement: Advances, Implications, and Lessons from an Active Shooter Case Study. In Application of Big Data for National Security. pp. 39–54.

Grigsby, M., 2016. Advanced Customer Analytics: Targeting, Valuing, Segmenting and Loyalty Techniques. Kogan Page Publishers.

Grondzik, W.T., 2009. Principles of Building Commissioning, John Wiley & Sons.

Groves, P. et al., 2013. The “big data”revolution in healthcare. McKinsey Quarterly, (January), p.22.

Haider, M., 2015. Getting Started with Data Science: Making Sense of Data with Analytics 1st ed., IBM Press.

Hamilton, D.K. & Watkins, D.H., 2009. Evidence-based design for multiple building types, John Wiley & Sons.

Han, J., Pei, J. and Kamber, M., 2011. Data mining: concepts and techniques. Elsevier.

Hanna, K.C. & Culpepper, R.B., 1998. GIS and Site Design: New Tools for Design Professionals, Wiley .

Hensel, M., 2013. Performance-oriented architecture : rethinking architectural design and the built environment, Wiley.

Hershberger, R., 2015. Architectural programming and predesign manager. Routledge.

Hu, H., Wen, Y., Chua, T-S., Li, X. 2014. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access, 2, 652-687.

Hubley, J., 2016. Languages of NYC. Available at: http://jillhubley.com/project/nyclanguages/#close.

Igarapé Institute, 2016. CrimeRadar. Available at: https://rio.crimeradar.org/.

Jiao, Y. et al., 2013. A cloud approach to unified lifecycle data management in architecture, engineering, construction and facilities management: Integrating BIMs and SNS. Advanced Engineering Informatics, 27(2), pp.173–188.

John Walker, S., 2014. Big data: A revolution that will transform how we live, work, and think.

Ju, H. et al., 2013. Management in the big data & IoT Era: A report on APNOMS 2012. In Journal of Network and Systems Management. pp. 517–524.

Kalani, N. et al., 2016. Optimization of Building Energy Consumption through Data Mining Using Modern Science. International Journal of Advanced Biotechnology and Research, 7, pp.600–608.

Kalogirou, S., 2000. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy, 25(5), pp.479–491.

Kalogirou, S.A., 2001. Artificial neural networks in renewable energy systems applications: a review. Renewable & Sustainable Energy Reviews, 5(4), pp.373–401.

Karakiewicz, J., 2010. Data Driven Urban Design. In SIGRADI 2010 / Disrupción, modelación y construcción: Diálogos cambiantes.

Katipamula, S. & Brambley, M., 2005. Review article: Methods for fault detection, diagnostics, and prognostics for building systems—a review, part II. HVAC&R Research.

Kim, H., Stumpf, A. & Kim, W., 2011. Analysis of an energy efficient building design through data mining approach. Automation in Construction, 20(1), pp.37–43.

Kim, K. & Teizer, J., 2014. Automatic design and planning of scaffolding systems using building information modeling. Advanced Engineering Informatics, 28(1), pp.66–80.

Koch, C. & Firmenich, B., 2011. An approach to distributed building modeling on the basis of versions and changes. Advanced Engineering Informatics, 25(2), pp.297–310.

Krijnen, T. & Tamke., M., 2015. Assessing implicit knowledge in BIM models with machine learning. Modelling behaviour. Design Modelling Symposium, Copenhagen, pp.397–406.

Kumlin, R.R., 1995. Architectural Programming - Creative Techniques for Design Professionals, McGraw-Hill Education.

Kuo, C.-J., 2003. Spatial analysis of chinese garden designs with machine learning. CAADRIA 2003: Proceedings of the 8th International Conference on Computer Aided Architectural Design Research in Asia, Bangkok, Thailand, pp.541–552.

LaGro, J.A., 2013. Site analysis : informing context-sensitive and sustainable site planning and design, Wiley.

Lam, P.T.I., Wong, F.W.H. & Chan, A.P.C., 2006. Contributions of designers to improving buildability and constructability. Design Studies, 27(4), pp.457–479.

Lan, J.H. and Chiu, M.L., 2005. Information Mining to Enhance Shared Understanding in Collaborative Architectural Design. In Proceedings of the 10th Conference CAADRIA (pp. 83-93).Lawson, B., 1990. How designers think, Butterworth Architecture.

Laxmi, L.E. et al., 2016. A Literature Inspection on Big Data Analytics. International Journal of Innovative Research in Engineering & Management, (35), pp.2350–557.

Lee, S.J. & Siau, K., 2001. A review of data mining techniques. Industrial Management & Data Systems, 101(1), pp.41–46.

Lee, Y., Choi, J.W. and Lertlakkhanakul, J., 2005. Developing a user location prediction model for ubiquitous computing. In Proceedings of CAAD Futures (pp. 215-224).

Li, H., Cao, J. & Love, P.E.D., 1999. Using machine learning and GA to solve time-cost trade-off problems. Journal of Construction Engineering and Management, 125(5), pp.347–353.

Li, K., Su, H. & Chu, J., 2011. Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study. Energy and Buildings, 43(10), pp.2893–2899.

Liang, X., Hong, T. & Shen, G.Q., 2016. Occupancy data analytics and prediction: A case study. Building and Environment, 102, pp.179–192.

Lin, C., 2011. Topology Pattern Mining: A visual approach for representing and retrieving design patterns of spatial topology in a case library. In CAADRIA 2011 Circuit Bending, Breaking and Mending: Proceedings of the 16th International Conference on Computer-Aided Architectural Design Research in Asia, The University of Newcastle, Australia. pp. 535–543.

Linoff, G. & Berry, M.J.A., 2001. Mining the Web : transforming customer data into customer value, John Wiley & Sons.

Lopes, J. V. et al., 2015. Multidimensional Analysis of Public Open Spaces. Urban Morphology , Parametric Modelling and Data Mining. Real Time: Proceedings of the 33rd eCAADe Conference, Vienna, Austria., 1, pp.351–360.

Lou, S. et al., 2016. Prediction of diffuse solar irradiance using machine learning and multivariable regression. Applied Energy, 181, pp.367–374.

Loyola, M. (2016). National BIM Survey 2016 Chile: Summary Report. University of Chile: Faculty of Architecture and Urbanism.

Loyola, M., & López, F. (2018). An evaluation of the macro-scale adoption of Building Information Modeling in Chile: 2013-2016. Revista de la Construcción. Journal of Construction, 17(1), 158-171.

Lu, W. et al., 2015. Benchmarking construction waste management performance using big data. Resources, Conservation and Recycling, 105, pp.49–58.

Ma, J. & Cheng, J.C.P., 2016. Data-driven study on the achievement of LEED credits using percentage of average score and association rule analysis. Building and Environment, 98, pp.121–132.

Machairas, V., Tsangrassoulis, A. and Axarli, K., 2014. Algorithms for optimization of building design: A review. Renewable and Sustainable Energy Reviews, 31, pp.101-112.

Maciejewski, M., 2016. To do more, better, faster and more cheaply: using big data in public administration. International Review of Administrative Sciences.

MADCAD, 2016. MADCAD US Local Codes. Available at: http://www.madcad.com/buildingcodes/

Magoulès, F.M. and M.L. in B.E.A. & Zhao, H.-X., 2016. Data Mining and Machine Learning in Building Energy Analysis: Towards High Performance Computing, Hoboken, NJ: John Wiley & Sons, Inc.

Marsland, S., 2015. Machine learning: an algorithmic perspective. CRC press.

Martin, L., 2014. Deployment of Big Data Analytics Approaches in the Public Administration and Comparison of IT Performance Indicators. Current Trends in Public Sector Research.

Mateo, F. et al., 2013. Machine learning methods to forecast temperature in buildings. Expert Systems with Applications, 40(4), pp.1061–1068.

Mathew, P.A., Dunn, L.N., Sohn, M.D., Mercado, A., Custudio, C. and Walter, T., 2015. Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. Applied Energy, 140, pp.85-93.

Mathias, M., Martinovic, A., Weissenberg, J., Haegler, S. and Van Gool, L., 2011. Automatic architectural style recognition. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3816, pp.171-176.

McGraw-Hill Construction (2014a). The business value of BIM in Australia and New Zealand: How building information modelling is transforming the design and construction industry. SmartMarket Report. Bedford, Massachusetts: McGraw Hill Construction.

McGraw-Hill Construction. (2009). The business value of BIM. Smart Market Report. Bedford, Massachusetts: McGraw Hill Construction

McGraw-Hill Construction. (2010). The business value of BIM in Europe: Getting Building Information Modelling To The Bottom Line The United Kingdom, France And Germany. Smart Market Report. Bedford, Massachusetts: McGraw Hill Construction

McGraw-Hill Construction. (2012a). The business value of BIM in North America: multi-year trend analysis and user ratings (2007-2012). Smart Market Report. Bedford, Massachusetts: McGraw Hill Construction

McGraw-Hill Construction. (2012b). The Business Value of BIM in South Korea. SmartMarket Report. Bedford, Massachusetts: McGraw Hill Construction

McGraw-Hill Construction. (2014b). The Business Value of BIM in Major Global Markets. SmartMarket Report. Bedford, Massachusetts: McGraw Hill Construction

Mccullough, C., 2010. Evidence-Based Design for Healthcare Facilities 1st ed., Indianapolis, IN: Sigma Theta Tau International.

Mehanna, R., 2013. Resilient Strucctures Through Machine Learning and Evolution. In ACADIA 2013 Adaptive Architecture: Proceedings of the 33rd Annual Conference of the Association for Computer Aided Design in Architecture, Cambridge, Ontario, Canada. pp. 319–326.

Merrell, P., Schkufza, E. & Koltun, V., 2010. Computer-generated residential building layouts. ACM Transactions on Graphics, 29(6), p.Article No. 181.

Michalek, J. & Papalambros, P., 2002. Interactive design optimization of architectural layouts. Engineering Optimization, 34(5), pp.485–501.

Michalek, J., Choudhary, R. & Papalambros, P., 2002. Architectural layout design optimization. Engineering Optimization, 34(5), pp.461–484.

Milion, R.N., Paliari, J.C. & Liboni, L.H.B.B., 2016. Improving consumption estimation of electrical materials in residential building construction. Automation in Construction, 72, pp.93–101.

Minelli, M., Chambers, M. & Dhiraj, A., 2012. Big data, big analytics: emerging business intelligence and analytic trends for today's businesses, John Wiley & Sons.

MIT GIS Services, 2016. MIT GeoWeb. Available at: https://arrowsmith.mit.edu/mitogp

MIT Senseable City Lab, 2016. Treepedia. Available at: http://senseable.mit.edu/treepedia.

Morales, C., 2014. ArchDaily: Los dos chilenos que conquistaron el mundo -. Forbes Mexico. Available at: http://www.forbes.com.mx/archdaily-los-dos-chilenos-que-conquistaron-el-mundo/

Nawari, N.O., 2012. BIM-Model Checking in Building Design. In Structures Congress 2012. Reston, VA: American Society of Civil Engineers, pp. 941–952.

NBS, 2018. The National BIM Report. The NBS, United Kingdom. Retrieved from: https://www.thenbs.com/knowledge/the-national-bim-report-2018

Neto, A.H. & Fiorelli, F.A.S., 2008. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy and Buildings, 40(12), pp.2169–2176.

Niemeijer, R.A., de Vries, B. & Beetz, J., 2014. Freedom through constraints: User-oriented architectural design. Advanced Engineering Informatics, 28(1), pp.28–36.

NIST, 2015a. NIST Big Data Interoperability Framework: Volume 1, Definitions.

NIST, 2015b. NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements

NYC Parks, 2016. New York City Street Tree Map. Available at: https://tree-map.nycgovparks.org/

O'Brien, D.T. & Montgomery, B.W., 2015. The Other Side of the Broken Window: A Methodology that Translates Building Permits into an Ecometric of Investment by Community Members. American Journal of Community Psychology, 55(1–2), pp.25–36.

Oechslin, W., 1993. CAAD und Geschichte — computus et historia. In Architectura et Machina. Wiesbaden: Vieweg+Teubner Verlag, pp. 14–23.

Park, H.S. et al., 2016. Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques. Applied Energy, 173, pp.225–237.

Pauwels, P. et al., 2011. A semantic rule checking environment for building performance checking. Automation in Construction, 20(5), pp.506–518.

Peña, M. et al., 2016. Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. Expert Systems with Applications, 56, pp.242–255.

Preiser, W.F., 1995. Post-occupancy evaluation: how to make buildings work better. Facilities, 13(11), pp.19-28.

Preservation Green Lab, 2016. Atlas of ReUrbanism. National Trust for Historic Preservation - Preservation Leadership Forum .

Provost, F. & Fawcett, T., 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. Data Science and Big Data, 1(1), pp.51–59.

Pyne, S., Rao, B.L.S.P. & Rao, S.B., 2016. Big Data Analytics Methods and Applications, NewDelhi.

Reffat, R.M. and Gero, J.S., 2005, November. A virtual mining environment for providing dynamic decision support for building maintenance. In the Proceedings of the 23rd eCAADs Conference on Digital Design: The Quest for NewParadigms (pp. 589-596).

Reffat, R.M., 2008. Investigating Patterns of Contemporary Architecture using Data Mining Techniques. In Proceedings of the 26th eCAADe Conference. pp. 601–608.

Reynolds, D., Ghantous, K. & Otani, R. 2015. Performance Measures from Architectural Massing Using Machine Learning. Proceedings of the International Association for Shell and Spatial Structures (IASS) Symposium 2015, Amsterdam

Rokach, L. and Maimon, O., 2014. Data mining with decision trees: theory and applications. World scientific.

Römer, C. & Plümer, L., 2010. Identifying Architectural Style in 3D City Models with Support Vector Machines. Photogrammetrie - Fernerkundung - Geoinformation, 2010(5), pp.371–384.

Sailer, K. et al., 2008. Evidence-Based Design: Theoretical and Practical Reflections of an Emerging Approach in Office Architecture. In Undisciplined! Proceedings of the Design Research Society Conference 2008. Sheffield, UK. July. Sheffield, UK, pp. 1–16.

Schmitt, G.N., 1999. Information Architecture: Basis and Future of CAAD, Birkhäuser--Publishers for Architecture.

Schutt, R. and O'Neil, C., 2013. Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.".

Shalev-Shwartz, S. and Ben-David, S., 2014. Understanding machine learning: From theory to algorithms. Cambridge university press.

Shin, H., Chon, Y. & Cha, H., 2012. Unsupervised construction of an indoor floor plan using a smartphone. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 42(6), pp.889–898.

Siegel, E., 2013. Predictive analytics the power to predict who will click, buy, lie, or die, John Wiley & Sons, Inc.

Simeone, D. & Kalay, Y.E., 2012. An Event-Based Model to Simulate Human Behaviour in Built Environments. In Digital Physicality: Proceedings of the 30th eCAADe Conference, Prague, Czech Republic. pp. 525–532.

Simeone, D., Schaumann, D., Kalay, Y. and Carrara, G., 2013, September. Adding users' dimension to BIM. In EAEA-11 conference (Track 3) Conceptual Representation: exploring the layout of the built environment (pp. 483-490).

Sin, K. & Muthu, L., 2015. Application of big data in education data mining and learning analytics-A literature review. Ictact Journal on Soft Computing: Special Issue on Soft Computing Models for Big Data, 5(4), pp.1035–1049.

Skibniewski, M., Arciszewski, T. & Lueprasert, K., 1997. Constructability Analysis: Machine Learning Approach. Journal of Computing in Civil Engineering, 11(1), pp.8–16.

Sokmenoglu, A., Burak, C. & Akgul, C.B., 2010. Exploring the Patterns and Trends of Socio-spatial Activities of Architecture Student Community in Istanbul by Data Mining. In Future Cities: Proceedings of the 28th eCAADe Conference, Zurich, Switzerland. pp. 143–150.

Solihin, W. & Eastman, C., 2015. Classification of rules for automated BIM rule checking development. Automation in Construction, 53, pp.69–82.

Strelka, 2016. What Do 16 000 Photos Say About Moscow. Available at: http://www.strelka.com/en/magazine/2015/11/26/16-photos-of-moscow

Strobbe, T. et al., 2016. Automatic architectural style detection using one-class support vector machines and graph kernels. Automation in Construction, 69, pp.1–10.

Structurae/ 2016. About Structurae. Available at: https://structurae.net/about/

Su, Z. & Yan, W., 2014. Improving Genetic Algorithm for Design Optimization Using Architectural Domain Knowledge. In ACADIA 2014 Design Agency: Proceedings of the 34th Annual Conference of the Association for Computer Aided Design in Architecture. pp. 653–660.

Szalay, A., 2011. Extreme Data-Intensive Scientific Computing. Computing in Science & Engineering, 13(6), pp.34–41.

Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B. and Bowman, D., 2016. Application of machine learning to construction injury prediction. Automation in construction, 69, pp.102-114.

Tomé, A. et al., 2015. Space-use analysis through computer vision. Automation in Construction, 57, pp.80–97.

Tsanas, A. & Xifara, A., 2012. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings, 49(February), pp.560–567.

Tufte, E.R., 1983. The visual display of quantitative information,

Tulasi, B., 2013. Significance of Big Data and Analytics in Higher Education. International Journal of Computer Applications, 68(14), pp.21–23.

Ugarte, J. P. and M. Leef. 2016. Digital Geo-Plexus: Instagram as a tool for re-evaluating notions of proximity. Proceedings of the 21st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2016) / Melbourne 30 March–2 April 2016, pp. 395-404

Ulrich, R.S. et al., 2008. A review of the research literature on evidence-based design. Health Environments Research and Design Journal, 1(3), pp.61–125.

Van der Aalst, W. M. 2014. Data scientist: The engineer of the future. In Enterprise Interoperability VI (pp. 13-26). Springer, Cham.

WAAG & Span, B., 2016. CitysDK: All buildings in the Netherlands. Waag Society.

Williams, M., Burry, J. & Rao, A., 2014. Understanding Social Behaviors in the Indoor Environment. In ACADIA 2014 Design Agency: Proceedings of the 34th Annual Conference of the Association for Computer Aided Design in Architecture, Los Angeles, California, US. pp. 671–680.

Wu, X. et al., 2008. Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), pp.1–37.

Xiang, Z. et al., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management, 44, pp.120–130.

Yalcintas, M. & Aytun Ozturk, U., 2007. An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database. International Journal of Energy Research, 31(4), pp.412–421.

Yamamoto, M. et al., 2011. A genetic algorithm based form-finding for tensegrity structure. In Procedia Engineering. Elsevier, pp. 2949–2956.

Yan, W. 2006. Integrating Video Tracking and Virtual Reality in Environmental Behavior Study.Proceedings of the 25th Annual Conference of the Association for Computer-Aided Design in Architecture] pp. 483-488

Yang, J. et al., 2015. Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. Advanced Engineering Informatics, 29(2), pp.211–224.

Yang, Q.Z. & Xu, X., 2004. Design knowledge modeling and software implementation for building code compliance checking. Building and Environment, 39(6), pp.689–698.

Yegnanarayana, B., 2009. Artificial neural networks. PHI Learning Pvt. Ltd.

Yin, S., Wang, G. & Karimi, H., 2014. Data-driven design of robust fault detection system for wind turbines. Mechatronics. Yu, Z. et al., 2010. A decision tree method for building energy demand modeling. Energy and Buildings, 42(10), pp.1637–1646.

Yu, Z. (Jerry) et al., 2016. Advances and challenges in building engineering and data mining applications for energy-efficient communities. Sustainable Cities and Society, 25, pp.33–38.

Yu, Z., Fung, B.C.M. & Haghighat, F., 2013. Extracting knowledge from building-related data - A data mining framework. Building Simulation, 6(2), pp.207–222.

Zhang, L. et al., 2012. Visual analytics for the big data era - A comparative review of state-of-the-art commercial systems. In IEEE Conference on Visual Analytics Science and Technology 2012, VAST 2012 - Proceedings. pp. 173–182.

Zhao, H.X. & Magoulès, F., 2012. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews.

Zhao, J. et al., 2014. Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. Energy and Buildings, 82, pp.341–355.

Zimmerman, A. and Martin, M., 2001. Post-occupancy evaluation: benefits and barriers. Building Research & Information, 29(2), pp.168-174.

Zimring, C.M. and Reizenstein, J.E., 1980. Post-occupancy evaluation: An overview. Environment and Behavior, 12(4), pp.429-450.