

Applying big data in water treatment industry: A new era of advance



Djamel Ghernaout^{1, 2, 3, *}, Mohamed Aichouni^{2, 4}, Abdulaziz Alghamdi^{2, 5}

¹Chemical Engineering Department, College of Engineering, University of Ha'il, PO Box 2440, Ha'il 81441, Saudi Arabia

²Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry, College of Engineering, University of Ha'il, PO Box 2440, Ha'il 81441, Saudi Arabia

³Chemical Engineering Department, Faculty of Engineering, University of Blida, PO Box 270, Blida 09000, Algeria

⁴Industrial Engineering Department, College of Engineering, University of Ha'il, PO Box 2440, Ha'il 81441, Saudi Arabia

⁵Mechanical Engineering Department, College of Engineering, University of Ha'il, PO Box 2440, Ha'il 81441, Saudi Arabia

ARTICLE INFO

Article history:

Received 27 October 2017

Received in revised form

8 January 2018

Accepted 11 January 2018

Keywords:

Big data

Predictive analytics

Water (wastewater) treatment industry

Potable water

Process control

ABSTRACT

It is well-known that water is an invaluable natural resource and it is also obvious that demand is always going to augment and shortages become more frequent. On the other hand, the development of Big Data (BD), machine learning and artificial intelligence, is beginning to offer realistic opportunities to operate water treatment systems in more efficient manners. In fact, BD concerns all the data we now possess and transform it into knowledge that we may directly employ to manage treatment facilities in a better fashion. The right data, analytics, and decision framework may pilot water utilities to a well-optimized efficiency. Indeed, possessing too much data but not sufficiently comprehensible or ready for use, fine-tuning data collection and funneling it into an integrated data management system may be the manner to become more enterprising and make better decisions. However, employing BD in water treatment remains at its first initiating steps. As a future trend, pooling data and using analytical tools to predict where we should be heading to become more proactive will be a great stage towards the water industry advance.

© 2018 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

As a background, it is just to announce that water treatment plants in several communities, and also in industry, are not frequently highly developed (Thompson and Kadiyala, 2014a, 2014b; Sirkiä et al., 2017; Zhang et al., 2017; Beal and Flynn, 2015; KEMIRA, 2017; Deloitte, 2017). The operations usually date back to the 1970s or 80s, and are costly in terms of energy and chemical consumption. Process control is often performed via manual adjustments of parameters such as aeration pumps and chemical injection, which in turn are based on manual sampling and retrospective tests realized on methodical periods. To be on the secure place concerning the pollutant boundaries in effluent, overdosing of chemicals is more frequent (KEMIRA, 2017).

At a very basic level, Big Data (BD) just signifies we have a great quantity of data (Tracy, 2016; Herschel and Miori, 2017; Birgé et al., 2016; Rinaldo

et al., 2007). Water utilities consider data from supervisory control and data acquisition (SCADA) systems, comprising flow statistics, online controlling, dissolved oxygen (DO) measurements, and air flows, and also data from laboratory information management systems (LIMS) and computerized maintenance management systems (CMMS), to cite many examples (Shaw, 2017; Robinne et al., 2018; Gwenzi et al., 2017).

Such data is of use, and much of it has been around for decades. However, the manner data is assembled at treatment facilities is frequently broken into pieces. There are huge quantities of information in computer systems that are not frequently connected to each other. The Internet Age has guided in the capacity to centralize contrasting information into an elementary, relevant reservoir of data that lets water and wastewater treatment plant operators to comprehend, administer, and employ it to enhance plant dependability and efficiency. BD initiatives and new information administration implements make us able to transform all that data into comprehensible, functional information that aids us become more enterprising and make better resolutions concerning plant administration (Shaw, 2017; Big Data, 2017; Ler, 2016; Pylro et al., 2016;

* Corresponding Author.

Email Address: sdjamelandalus@hotmail.com (D. Ghernaout)

<https://doi.org/10.21833/ijaas.2018.03.013>

2313-626X/© 2018 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

De Mulder et al., 2016; Yang et al., 2017; Stewart et al., 2013; Hassani et al., 2017).

As an illustration, Black and Veatch provides ASSET360™ (2016) a smart analytics platform to give utilities, cities, and other entities a holistic, 360-degree understanding of their infrastructure-based systems. In spite of the fact that the center of attention on BD in the water industry is relatively recently developed, comprehensive data handling is not fresh for energy utilities; Black and Veatch has given asset analytic solutions to utility clients for more than 20 years and has managed a smart analytics monitoring and diagnostics service for more than 10 years. The company's utility analytics comprise operational intelligence and adaptive planning solutions (Shaw, 2017; Hampton et al., 2013).

This short review focuses on BD using in water treatment industry and tries to attract water treatment specialists on this promising tool for better efficiency in treating water. Also, this paper presents a brief history and the basic types of analytics useful to water utilities, and provides insights into methodologies for improved process monitoring and control and increased system reliability using predicative analytics.

2. Getting MAD to be smart

No matter what specific services or instruments water and wastewater utilities select to employ, it is crucial to elaborate a management plan, pull all significant information together, and take advantage of dashboards and smart screens that employ that data to do calculations and recognize trends. Therefore, utility managers can break the information down to answer questions such as "Where am I using energy or spending my dollars?" or "What am I spending on energy and chemicals in different parts of the facility?" (Shaw, 2017; Thames Water, 2016).

Moreover, utility staff with such information in hand may proactively distinguish likely difficulties before they appear rather than react to something like a broken pump. Even though SCADA systems have real-time capabilities, revealing current status and instantly warning of problems, being good enough to predict a likely difficulty by means of using smart analytic platforms is a game changer. The upcoming phase — collecting information and employing analytical instruments to predict where we must be heading to become more proactive — is a great one for the water industry (Shaw, 2017).

In order to achieve such success, nevertheless, needs concentration on both the quality and the quantity of our data, moving our attention from BD to bad data. If sensors (measurement devices) are not cleaned, calibrated, or properly used, as an illustration, it doesn't matter what we perform with the resulting data. The starting point is to make certain you have good primary measurements (Shaw, 2017; Ler, 2016; Kato et al., 2017).

Ingildsen and Olsson (2016) examined what utilities require to perform to be smart. They boiled it down to a simple, yet very useful, framework and suggested that water utilities have to be "MAD" to be smart (Shaw, 2017). As they explained, M is for MEASURE, since we must concentrate on possessing good measurements in the right place; A is for ANALYTICS, since we must comprehend and examine the data we gather; and D concerns the DECISION-making operation. Employing what we have knowledge to take valid decisions may be an automated procedure in several situations. It can be useful to split BD into these three parts (Thompson and Kadiyala, 2014a; 2014b; Shaw, 2017; Zhang et al., 2016; Ahmad et al., 2017; Chen and Han, 2016).

In terms of accuracy, the instrumentation that we possess at this time is better than ever, while measurement devices were previously a weak point. People usually comprehend the requirement to clean and calibrate instruments; however, it may remain a significant starting step (Shaw, 2017, Ler, 2016; Gourbesville, 2016, Kato et al., 2017; Imen et al., 2015).

Enhanced analytics are more the focal point nowadays, with the advantages and requirements explained above. Decisions will be the following focal point, and fairly soon, as evidenced by research now underway (Shaw, 2017; Ler, 2016; Gourbesville, 2016; Kato et al., 2017; Imen et al., 2015).

Smart analytics — named Smart Integrated Infrastructure (SII) at Black and Veatch — have been used to power stations for several years. In SII, a rigorous question is "How efficient is the plant as a whole?" With the ability to zoom in on specific pieces and ask questions such as "How many dollars per hour does it cost us not to have this part of the plant operating as well as it could?" utilities and cities can use smart analytics to make smarter decisions by proactively identifying and prioritizing improvements (Shaw, 2017). Black and Veatch has developed tools specifically for combined heat and power, membranes, and activated sludge. Black and Veatch are working with the city of Lawrence, KS, to refine tools to enable the city's plant managers to optimize operations. Initially these tools will be used at the wastewater treatment plant, but eventually they will also be extended to the city's water treatment facility.

Plant operators are already seeing the benefit of being able to visualize information by pulling all operations data together in a consolidated database (Shaw, 2017).

3. BD basics

Regardless of the present-day center of attention for several on enhancing analytics and/or decisions, there is also a lot to be said for making sure our foundations are sound. Below are five keys to making BD work and avoiding the pitfalls of bad data (Shaw, 2017).

3.1. Data quality rather than quantity

Not even the most advanced analytics may prevail over measurement errors, whether that's noise, drift, or interferences (Fig. 1). If you are not confident in your main measurement devices and analyzers, you may have a lot of bad data that is worthless, no matter what you do with it. As an illustration, a Water Environment Research Foundation (now the Water Environment and Reuse Foundation) decision support system project needed

a research team member to realize data analytics to discern anomalies that might show toxins in plant influent; however, differentiating anomalies attributed to toxins from anomalies affected to measurement issues confirmed to be a big obstacle (Shaw, 2017, Ler, 2016; Gourbesville, 2016).

Confidence in sensors and analyzers may be reached by following three fundamental steps. Table 1 gives an abstract of these basic stages (Shaw, 2017; Ler, 2016; Gourbesville, 2016).



Fig. 1: Modern instruments are more reliable than they were in the past, but they still need to be cleaned and taken care of (Shaw, 2017)

Table 1: Three stages for gaining confidence in sensors and analyzers (Shaw, 2017)

Three stages for better manipulating sensors and analyzers	
Stage #1: Cleaning them	Wastewater treatment is an especially fouling environment and not the best place to put scientific equipment. Operators frequently underestimate how quickly sensors become fouled. Go for auto-cleaning whenever possible and avoid installing anything in raw sewage or primary effluent unless you really need the measurement because both areas are particularly prone to fouling. Mixed liquor is an easier place to take measurements, and final effluent is the easiest place of all. Water treatment systems usually are less fouling, but sensors still need periodic cleaning.
Stage #2: Calibrating them	This is generally understood, although the frequency of calibration, particularly for sensors that tend to drift, typically is shorter than ideal.
Stage #3: Validating them	This may be the action overlooked by most instrumentation suppliers. Analytics to validate the measurements, particularly during calibration, frequently need more attention.

3.2. Measuring useful items

What will you really employ to best run the plant? Many treatment plants need significant and fundamental measurements (such as DO in the aeration basins, airflow to each aeration zone, and electricity use by blowers); however, we must be careful in our enthusiasm not to swing to the other extreme and take measurements that are not particularly helpful. You may spend serious money measuring ammonia and nitrate all over a treatment plant; however, unless you are really employing it for monitoring, the measurements will probably be ignored and the instruments neglected. It is best to have a handful of good instruments, positioned in locations where you are really measuring something you may monitor, and to try to keep those

measurement devices functioning perfectly (Shaw, 2017, Ler, 2016; Gourbesville, 2016).

3.3. Dynamics rather than steady state

A lot of the design and operational guidance in textbooks and training materials has easy equations into which you plug a single number to get your answer (such as sludge age calculation or removal efficiency). Comparably, influent and effluent samples are frequently flow-weighted or time-averaged composites. It is usual to think and talk about average daily conditions. Nevertheless, the reality is that our treatment plants see significant daily variations in flows and concentrations, and consequently we have to consider them as dynamic systems. As an illustration, an online phosphate

analyzer taking measurements at the end of the aeration basin just prior to the clarifiers might reveal daily phosphate peaks of 1 or 2 mg/L every afternoon for just an hour or so; however, the effluent composite sample measurements could be consistently below 0.2 mg/L. In order to comprehend our treatment systems, we require quantifying and analyzing their dynamics (Shaw, 2017).

3.4. Different timescales

Hand in hand with dynamics is the need to think about different timescales: diurnal (daily) variations, weekly trends (especially weekend versus weekday differences), and seasonal shifts. For each of these, the data analytics requirements are quite different and have to be carefully considered (Fig. 2). For diurnal variations, it may be helpful to compare one day to the following by overlaying the dynamic data. For weekly trends, we may perform something similar over a seven-day horizon. Moreover, for seasonal shifts, it is frequently interesting to plot and compare long-term trends to temperature and maybe rainfall shifts (Shaw, 2017).



Fig. 2: Tools such as Black and Veatch's ASSET360 system help water utility managers follow in the footsteps of their energy utility colleagues to harness data for improved decisions and operations (Shaw, 2017)

3.5. Handling outliers and extraordinary events

In data analytics, it is usual to identify and eliminate outliers, supposing they are either bad measurements or not typical and thus something to ignore. However, experience proves that a lot of what is performed at water and wastewater treatment plants is trying to keep the process stable in response to abnormal events, like upsets from shock loads or toxins, or, more usually, responding to wet weather for wastewater plants or major line breaks or droughts for water treatment. We have to identify outliers; however, rather than throw them away, we should decide how to respond (Shaw, 2017).

BD, concluded (Shaw, 2017), concerns taking all the data we now possess at our fingertips and turning it into knowledge that we may implement to run our treatment facilities in a better fashion. The

right data, analytics, and decision framework may guide water (and energy) utilities to optimal performance.

4. KEMIRA story with BD

Recently, KEMIRA (2017) launched a program to survey the options BD might provide to assist aging plants in both augmenting operating performance and also satisfying increasing demand for water. During the time that BD may signify several things and enclose several various subjects like smart technology, machine learning, artificial intelligence and other new areas, KEMIRA (2017) decided from the beginning that any BD program should be aimed at tangible problem solving.

4.1. Exploring the challenges and needs

KEMIRA (2017) performed a series of three increasingly-targeted interview programs with dozens of water treatment operators. National and regional water regulatory agencies were also implicated in the interviews. Not amazingly, the interviews showed that ~80% of challenges in water/wastewater treatment plants (WWTPs) are linked to poor plant running. Predominantly a lack of comprehending of the chemistry implicated is considered as part of the issue (Fig. 3) (Oyebamiji et al., 2017). The scale transition from a bacterium or cellular level (microscale – < micrometer size) to the floc and biofilm aggregates (mesoscale – millimeter size) to the macroscopic bulk WWTP operation as well as floc and biofilm interactions (macroscale – meter size). The emulator is linking the microscopic (bacterium/cell) to the mesoscopic (biofilm/floc) and, ultimately, to the macroscopic bulk operational parameters (Oyebamiji et al., 2017)

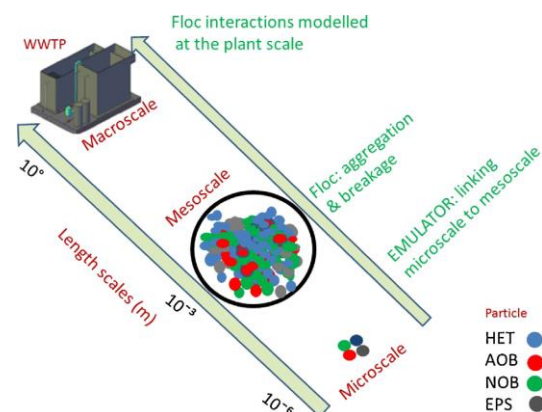


Fig. 3: Schematic of various length scales for multiscale modelling of an activated sludge process based WWTP

4.2. Rapid, accurate prediction of sludge properties

KEMIRA (2017) realized a thorough effort which conducted to a fresh tool that may very minutely and quickly, within seconds indeed, predict the properties of sludge at a WWTP. This is not a future scenario; it is what one may already perform

nowadays. By combining existing operational data, historical process data, machine data, chemical data and site data, and then applying newly-developed advanced analytics to benchmark the customer against similar sites, one may obtain a very accurate prediction of sludge properties. The main advantage of this tool (KEMIRA, 2017) is that it gives operators tangible ways to decrease operating costs via better identification and comprehending of the key process conditions and chemical properties which influence the sludge dryness. This also allows smoothing out of plant operations. The algorithms are now being moreover developed to perform even more, enabled by artificial intelligence and machine learning (KEMIRA, 2017).

4.3. Pioneering work continues

For several years, KEMIRA (2017) has been a pioneer and thought leader in the use of real-time sensors to optimize water treatment. In the long run, one may imagine an operation where all water inflow and outflow is measured constantly in real-time by sensors for variables such as pH, DO, nutrients, phosphates, nitrates, sludge dryness, pathogens, etc. With the right algorithms, this data may be employed to continuously optimize pumping and aerating energy consumption as well as chemical injection. In this fashion, operations can be smoothed out and operating costs trimmed, while still remaining safely within legal limits (KEMIRA, 2017).

5. Predictive analytics

One of the most well-known features of BD is predictive analytics. Far from the latest business buzzword, predictive analytics is a set of techniques that have become crucial to the business strategies of many household name brand firms, such as Netflix, Google, and Amazon. These firms, and many others, dominate their respective markets, due in large part to the important use of predictive analytics. Predictive analytics is a form of business intelligence gathering, the strategic business use of which is powerful enough to upend an industry. Driven by the tremendous-revenue generating potential of predictive analytics, more firms are investing in the necessary infrastructure, such as data storage and processing hardware and software and both database administrators and data analysts. As they do so, predictive analytics tools and techniques, grow in sophistication and refinement. Moreover, as more firms adopt predictive analytics, and incorporate it into their existing strategies, they fuel its widespread adoption, as competitors must adopt it or risk losing significant market share.

6. BD for better water management

All of the necessary technology exists to expand water resources and ensure its delivery to end-users.

Data acquisition has expanded dramatically in recent years with low-cost sensors and widespread adoption of geospatial analysis. These new technologies have increased our ability to find and monitor water stores. And infrastructure implemented to existing sensors allows for cloud computing and increased visibility of data across systems. Those wide-deployment technologies combined with unstructured data like social media, web content, and crowdsourcing, increases visibility and the amount of useable, useful water data (Tracy, 2016; Ler, 2016; Gourbesville, 2016).

BD analytics can continue to optimizing the balance between performance and reliability when it comes to farming. It may also prevent man-made disasters, such as sudden drops in water quality, which may not be detected until after the full affects are realized (Tracy, 2016).

BD such as these can help water utilities understand trends in land use and climate that will influence key decisions about planning an adaptive and responsive water system. BD and modeling can also help water utilities and land use planners collaborate to assess what amount of water will be needed and are available for different city growth scenarios (Tracy, 2016; Irvine et al., 2016; Karjalainen et al., 2017).

7. Roadmap, from now until then

The technology is in place, and is continually being optimized and made more affordable for use in farming. In fact, a lot of data has already been gathered. Still, there are barriers that need to be broken for the data we have and will continue to gather to be put to good use. Incentives for data sharing, such as financial gain that is actionable beyond regulatory enforcement, may help unlock sources of data. As a first step, a baseline set of water standards, indicators and measurements should be defined that reflect the core data on the state of our water system. Data standardization will enable integration of data collected for different purposes (Tracy, 2016):

- The rise of BD and new measurement technologies can transform the way that water is managed in the coming decades.
- However, water data must be synthesized more rapidly than government agencies' current pace of analysis.
- A national water data policy is needed that standardizes data integration and storage for more effective water management across sectors.
- Overcoming privacy constraints would help to maximize the potential of water data.
- Accurate assessments of private sector water risk require better matched data sources and data analytics across industry (Tracy, 2016).

Proskuryakova et al. (2018) used a combination of Foresight methods including scenario analysis to

develop long term trajectories and discuss strategies for the Russian water sector with a 20-year time horizon. The methodology was designed based the

key phases of the Foresight for Science, Technology and Innovation (ForSTI) process, proposed by Miles et al. (2016) with a Forstar model (Fig. 4).

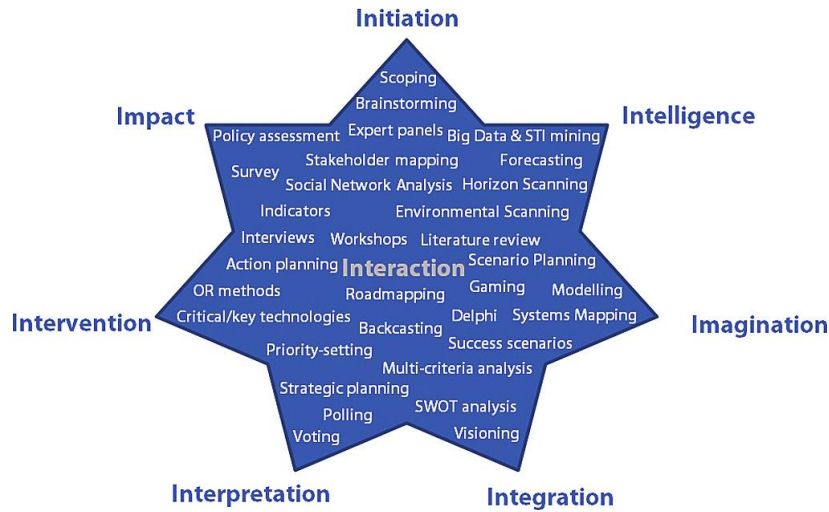


Fig. 4: ForSTI process and methods (Miles et al., 2016)

8. Modelling potable water production process with integrated life cycle assessment

Water treatment technologies and their operating conditions must be able to adapt to different types of raw waters (river, groundwater, sewage, and seawater) and their properties, to fluctuations in raw water quality, to the different technical, economic, and environmental objectives, and also to different available networks for raw material and energy resources (Ahmadi et al., 2016).

In this context and based on the EVALEAU tool, an improved PM-LCA-NET platform was developed and used for a potable water plant, by combining Process Modelling (PM), Life Cycle Assessment (LCA), and different raw material and energy resource

networks (NET) (Fig. 5). The left side of Fig. 5 shows the structure of the PM-LCA-NET, where a library of water treatment unit process modules (Python scripts) to generate process inventories with a high degree of specifications are coupled with several libraries and databases for water quality characterization, water chemistry, and drinkability criteria. The modules are also linked to Ecoinvent datasets (2.2) (Weidema et al., 2015) via the Umberto® (v5.6) software in order to include the necessary background processes and to complete the system's life cycle (Fig. 5, left bottom).

The PM-LCA-NET platform is therefore thought to fulfil the requirements listed in Table 2 (Ahmadi et al., 2016).

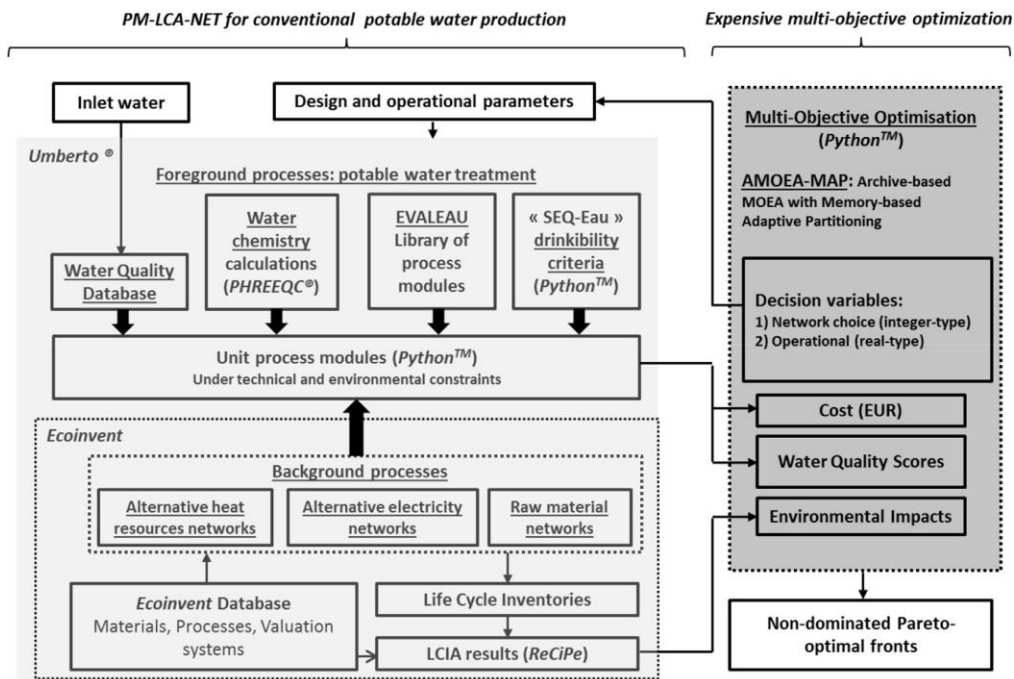


Fig. 5: Principle of the PM-LCA-NET tool combined with appropriate expensive multi-objective optimization platform (Ahmadi et al., 2016)

Table 2: Requirements fulfilled by the PM-LCA-NET platform (Goedkoop et al., 2009)

Requirement	Nine requirements fulfilled by the PM-LCA-NET platform
1	Provide an appropriate inlet water quality database using an external excel library grouping together, different water qualities from various origins (river, groundwater, sewage, and seawater).
2	Make the water chemistry calculations by PHREEQC® in order to model the chemical reactions in aqueous solutions.
3	Combine unit process modules in different treatment chains in order to assess a variety of different technologies.
4	Evaluate the local efficiency, technical performance, and cost evaluations.
5	Resolve complex flow networks using Umberto® 5.6.
6	Evaluate drinkability criteria via the SEQ-EAU model.
7	Perform the LCA calculations for different water treatment processes according to ISO 14040-44 standards, using conventional LCI databases for the background processes and recognized LCIA methods.
8	Assess the available alternative networks for raw materials and energy resources as background processes (Ecoinvent).
9	Account for the influence of design (integer-type) and operational (real-type) parameters on the LCI in order to evaluate the LCA results for different available networks and different working conditions of the plant.

9. Mining web-based data to assess public response to environmental events

Cha and Stow (2015) examined how the analysis of web-based data, such as Twitter and Google Trends, may be employed to evaluate the social importance of an environmental accident. The concept and methods were applied in the shutdown of drinking water supply at the city of Toledo, Ohio, USA. Toledo's notice, which persisted from August 1 to 4, 2014, is a high-profile event that directly influenced approximately half a million people and received wide recognition. The notice was given when excessive levels of microcystin, a by-product of cyanobacteria blooms, were discovered at the drinking water treatment plant on Lake Erie. Twitter

mining results illustrated an instant response to the Toledo incident, the associated collective knowledge, and public perception. The results from Google Trends, on the other hand, revealed how the Toledo event raised public attention on the associated environmental issue, harmful algal blooms, in a long-term context. Therefore, when jointly applied, Twitter and Google Trend analysis results offer complementary perspectives (Fig. 6). Web content aggregated through mining approaches provides a social standpoint, such as public perception and interest, and offers context for establishing and evaluating environmental management policies (Cha and Stow, 2015).

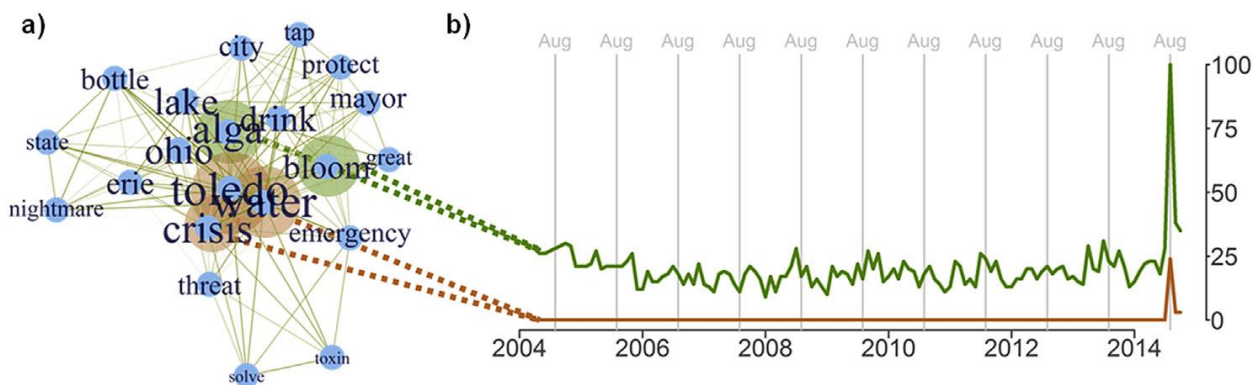


Fig. 6: a) A network of frequently associated terms with a keyword, “toledo water”, in tweets and b) monthly patterns of the number of searches as a percent of the maximum number of searches given in any month between Jan 2004 and Oct 2014 in response to querying “toledo water crisis” (brown line) and “algae (or algal) bloom” (green line) in Google Trends. In a), the thickness of an edge and font size of a term indicate the degree of correlation with the keyword or the level of interest the term receives (Cha and Stow, 2015)

10. Conclusion

The main important points drawn from this review may be listed as:

In the water industry, BD analysis is a relatively new concept and is not widely utilized. However, due to the development and deployment of rapid water quality sensors, there is more data than ever for operators, engineers, plant managers and other stakeholders to sift through and analyze. This data is often stored in various platforms (PI database, Excel spreadsheets, notebooks, etc.) and ultimately underutilized. There is value in harnessing the power of historical and real-time data to compliment traditional operational decision support systems. Advancements in computerized data analysis allow

for the immense amount of data that is generated to be transformed into informative insights and decision support in a fraction of the time it would take a human. Shortening analysis time allows operators to identify problems earlier, allowing them to be proactive rather than reactive in their response. BD driven decision support tools can also be used to provide real-time treatment process optimization resulting in increased energy efficiency and reduced waste. Ultimately, BD in the water industry is about using the data we already generated to increase the performance and resiliency of our infrastructure.

Growing concerns over water supply scarcity have prompted an increased global interest in conservation, reuse, and alternative water sources.

Globally, water utilities are looking for more efficient ways to manage their resources. The use of analytics is not new to many in the businesses field, but is an unfamiliar topic to many on the operations side of water utilities. "BD" is a common catchphrase in the water industry and one that can be useful to municipalities if properly harnessed and utilized. One such 'smart water utility', making use of analytics is highlighted. Perhaps the largest potential value in analytics lies in its applications to potable reuse. Improved technological processes may be coupled with analytical instruments that give real-time monitoring in order to advise operators of water potential water risks so production can be stopped and/or supply diverted. The continued improvement of online analytical instrumentation will be a key contributor to a safer, more reliable finished quality water quality. Combining stable and reliable online instrumentation with operational analytics provide real-time predictive performance. Predictive analytics provide operations new tools by which to proactively protect public health.

References

- Ahmad A, Khan M, Paul A, Din S, Rathore MM, Jeon G, and Choi GS (2017). Toward modeling and optimization of features selection in Big Data based social Internet of Things. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2017.09.028>
- Ahmadi A, Tiruta-Barna L, Benetto E, Capitanescu F, and Marvuglia A (2016). On the importance of integrating alternative renewable energy resources and their life cycle networks in the eco-design of conventional drinking water plants. *Journal of Cleaner Production*, 135: 872-883.
- ASSET360™ (2016). Smart water analytics applications available in FATHOM Store. Available online at: <https://www.wateronline.com/doc/asset-smart-water-analytics-applications-available-in-fathom-store-0001>
- Beal CD and Flynn J (2015). Toward the digital water age: Survey and case studies of Australian water utility smart-metering programs. *Utilities Policy*, 32: 29-37.
- Big Data (2017). EU FP7 DANSE: Integrated water treatment systems. Available online at: <http://www.lboro.ac.uk/research/avrrc/research/currentprojects/bigdata/eu-fp7-danse-integrated-water-treatment-systems.html>
- Birgé HE, Allen CR, Garmestani AS, and Pope KL (2016). Adaptive management for ecosystem services. *Journal of Environmental Management*, 183(2): 343-352.
- Cha Y and Stow CA (2015). Mining web-based data to assess public response to environmental events. *Environmental Pollution*, 198: 97-99.
- Chen Y and Han D (2016). On big data and hydroinformatics, *Procedia Engineering*, 154: 184-191.
- De Mulder C, Van Hoey S, Van Hulle S, Agathos SN, Cauwenberg P, Mergen P, and Schowanek D (2016). Pressing topics in the Belgian water sector anno 2015. *Sustainability of Water Quality and Ecology*, 7: 32-36.
- Deloitte (2017). *Water Tight 2.0; The top trends in the global water sector*. Deloitte Company, New York, USA.
- Goedkoop M, Heijungs R, Huijbregts M, De Schryver A, Struijs J, and Van Zelm R (2009). A life cycle impact assessment method which comprises harmonised category indicators at the midpoint and the endpoint level. 1st Edition, ReCiPe.
- Report I: Characterisation. Available online at: <http://www.pre-sustainability.com/reports>
- Gourbesville P (2016). Key challenges for smart water. *Procedia Engineering*, 154: 11-18.
- Gwenzi W, Chaukura N, Noubactep C, and Mukome FN (2017). Biochar-based water treatment systems as a potential low-cost and sustainable technology for clean water provision. *Journal of Environmental Management*, 197: 732-749.
- Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, and Porter JH (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11(3): 156-162.
- Hassani H, Silva ES, and Al Kaabi AM (2017). The role of innovation and technology in sustaining the petroleum and petrochemical industry. *Technological Forecasting and Social Change*, 119: 1-17.
- Herschel R and Miori VM (2017). Ethics and big data. *Technology in Society*, 49: 31-36.
- Imen S, Chang NB, and Yang YJ (2015). Developing the remote sensing-based early warning system for monitoring TSS concentrations in Lake Mead. *Journal of environmental management*, 160: 73-89.
- Ingildsen P and Olsson G (2016). *Smart water utilities: complexity made simple*. IWA Publishing, London, UK.
- Irvine K, Weigelhofer G, Popescu I, Pfeiffer E, Păun A, Drobot R, and Habersack H (2016). Educating for action: Aligning skills with policies for sustainable development in the Danube river basin. *Science of the Total Environment*, 543(Part A): 765-777.
- Karjalainen T, Hoeveler A, and Draghia-Akli R (2017). European Union research in support of environment and health: Building scientific evidence base for policy. *Environment International*, 103: 51-60.
- Kato N, Hirano N, Okazaki S, Matsushita S, and Gomei T (2017). Development of an all-solid-state residual chlorine sensor for tap water quality monitoring. *Sensors and Actuators B: Chemical*, 248: 1037-1044.
- KEMIRA (2017). Big data applications and advances for the water treatment industry. Chemical Industry Company, Helsinki, Finland. Available online at: <http://www.kemira.com/en/newsroom/whats-new/pages/Big-Data-applications-and-advances-for-the-water-treatment-industry.aspx>
- Ler LG (2016). Analysis of current ICT solutions in water business processes. *Procedia Engineering*, 154: 3-10.
- Miles I, Saritas O, and Sokolov A (2016). *Foresight for science, technology and innovation*. Springer International Publishing, Heidelberg, Berlin, Germany.
- Oyebamiji OK, Wilkinson DJ, Jayathilake PG, Curtis TP, Rushton SP, Li B, and Gupta P (2017). Gaussian process emulation of an individual-based model simulation of microbial communities. *Journal of Computational Science*, 22: 69-84.
- Proskuryakova LN, Saritas O, and Sivaev S (2018). Global water trends and future scenarios for sustainable development: The case of Russia. *Journal of Cleaner Production*, 170: 867-879.
- Pyro VS, Mui TS, Rodrigues JL, Andreote FD, and Roesch LF (2016). A step forward to empower global microbiome research through local leadership. *Trends in Microbiology*, 24(10): 767-771.
- Rinaldo A, Bertuzzo E, Blokesch M, Mari L, and Gatto M (2017). Modeling key drivers of cholera transmission dynamics provides new perspectives for parasitology. *Trends in Parasitology*, 33(8): 587-599.
- Robinne FN, Bladon KD, Miller C, Parisien MA, Mathieu J, and Flannigan MD (2018). A spatial evaluation of global wildfire-water risks to human and natural systems. *Science of the Total Environment*, 610: 1193-1206.

- Shaw A (2017). Understanding Big Data in the water industry. Available online at: <https://www.wateronline.com/doc/understanding-big-data-in-the-water-industry-0002>
- Sirkiä J, Laakso T, Ahopelto S, Ylijoki O, Porras J, and Vahala R (2017). Data utilization at finnish water and wastewater utilities: Current practices vs. state of the art. *Utilities Policy*, 45: 69-75.
- Stewart RA, Giurco D, and Beal CD (2013). Age of intelligent metering and big data: Hydroinformatics challenges and opportunities. *International Association for Hydro-Environment Engineering and Research*, 2: 107-110.
- Thames Water (2016). TWIST: How 'BIG' data and neuroscience will revolutionise the water industry. Water Services Company, Reading, UK.
- Thompson K and Kadiyala R (2014a). Protecting water quality and public health using a smart grid. *Procedia Engineering*, 70: 1649-1658.
- Thompson K and Kadiyala R (2014b). Leveraging big data to improve water system operations. *Procedia Engineering*, 89: 467-472.
- Tracy P (2016). How big data can save our depleting water supply. Available online at: <https://enterpriseiotinsights.com/20160802/channels/fundamentals/big-data-water-management-tag31-tag99>
- Weidema B, Hirschier R, Althaus HJ, Bauer C, Doka G, Dones V, Frischknecht R, Jungbluth N, Nemecek T, Primas A, and Wernet W (2015). Code of practice (Final report Ecoinvent datasets). Swiss Centre for Life Cycle Inventories, Duebendorf, Switzerland.
- Yang T, Long R, Cui X, Zhu D, and Chen H (2017). Application of the public-private partnership model to urban sewage treatment. *Journal of Cleaner Production*, 142(part 2): 1065-1074.
- Zhang F, Xue HF, and Zhang JC (2017). Multi-source big data dynamic compressive sensing and optimization method for water resources based on IoT. *Sustainable Computing: Informatics and Systems*. <https://doi.org/10.1016/j.suscom.2017.08.003>
- Zhang J, Li WY, Wang F, Qian L, Xu C, Liu Y, and Qi W (2016). Exploring the biological stability situation of a full scale water distribution system in south China by three biological stability evaluation methods. *Chemosphere*, 161: 43-52.