Turning passive data into knowledge - a review of wastewater treatment monitoring techniques

Ll. Corominas[1*], M. Garrido-Baserba[2], K.Villez[3], G. Olsson[4], U. Cortés[5], M. Poch[6]

[1]ICRA, Catalan Institute for Water Research, Spain

[2]Department of Civil and Environmental Engineering, University of California, Irvine, US

[3]Department of Process Engineering, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Switzerland

[4]Industrial Electrical Engineering and Automation, Lund University, Sweden

[5] KEMLg, Universitat Politècnica de Catalunya, Barcelona, Spain

[6] LEQUIA. Institute of the Environment, University of Girona, Spain

*Corresponding author: *lcorominas@icra.cat*

Abstract: The aim of this paper is to describe the state-of-the art of computer-based methods and tools for data analysis to improve operation of wastewater treatment plants or support decision-making. A comprehensive review has been made which shows that EU has been leading computer-based method development with a presence in 61% of the reviewed papers. The most cited methods have been artificial neural networks, principal component analysis, fuzzy, clustering, independent component analysis and partial least squares regression. Even though there has been progress on tools related to the development of environmental decision support systems, knowledge discovery and management, the research sector is still far from delivering systems that smoothly integrate several types of knowledge and different methods of reasoning. Several limitations that currently prevent the application of computer-based techniques in practice are highlighted.

Keywords: sensors, instrumentation, monitoring, fault detection, state-of-the art

MOTIVATION

The automation and control of wastewater treatment plants (WWTPs) relies on instruments that generate a large number of signals. To enable a more effective and efficient wastewater treatment practice, it is essential that one can process and analyse these raw data properly. Unfortunately, the required effort to analyse real-world plant data in depth is often cost- and time-prohibitive, and potentially valuable information remains unknown and unused. How do we turn passive data into actionable knowledge or something compelling that helps or supports the operation of wastewater treatment plants? The aim of this paper is to describe the state-of-the art of computer-based methods and tools for data analysis and knowledge management as applied in the context of wastewater treatment operation. This critical review targets method developers by discussing the underlying features that lead to successful techniques. The paper can also help plant managers and software developers by identifying mature and proven techniques.

THE SELECTED COMPUTER-BASED TECHNIQUES

The increase in WWTP operational requirements that must be managed parallels the ever-increasing possibilities concerning data processing. In this review we include four levels of processing:

- Low-level data checking (for the handling of noise, delays, and communication failures, the recognition of missing data and outliers, simple consistency and sanity checks based on process knowledge and experience)

- Basic information extraction (remove both large measurement deviations - gross error detection- and random errors - data reconciliation)
- Advanced information extraction (visualizing the main sources of variations in collected datasets, identifying periods of normal and abnormal operation, predicting key variables which are difficult or impossible to measure online, assessing process states through visualization)
- Human-interpretable information extraction and knowledge management (supporting operators for solving day-to-day problems, structuring gained experience, case based reasoning, trend based reasoning, managing knowledge -ontologies-)

A subset of computer-based methods have been selected for this review including methods based on control charts, on mass balances, on regression models (multi-linear, partial least squares PLS), self-organizing maps (SOM), principal component analysis (PCA), independent component analysis (ICA), artificial neural networks (ANNs), clustering, fuzzy methods, support vector machines (SVMs), and methods for the recognition of qualitative features in data series. In addition, computer-based tools for information (EDSS) and knowledge management (ontologies) are included in the review. Mechanistic techniques/models were excluded from this review. A specific review of machine learning techniques applied to the field of water and wastewater is found in Hadjimichael et al. (2016).

THE REVIEW

We reviewed the most relevant methods/tools in WWTP operation belonging to the four types of data processing described before, using SCOPUS by applying searches including the names of the techniques (and the relevant variations) plus the term “wastewater treatment” and limiting the scope to papers published before 2015. The SCOPUS search after selection for relevance resulted in 340 papers. A majority of these papers discuss ANN (20%), PCA (13%) and fuzzy (12%).

RESULTS

Leadership in the field

EU is the leader region in this field with the largest number of contributions, with presence in 61% of the papers, followed by Asia-Oceania which contributed to 34% of the papers and North-America with a presence in 12% of the studies. A minority of studies (less than 4%) have been conducted by South-American or African research groups. For each of the techniques EU has the largest number of contributions. The Asia-Oceania region has largely contributed to ANN (38 studies), Fuzzy (17) and PCA (20). There are also 37 papers (12%) which are result of cooperation between research groups from different regions.

Trends

Figure 1 shows the sum of the citations per year for the reviewed papers, separately for each method/tool. For some methods/tools we observe a steadily increase of citations along the years. ANN and PCA are the methods that have generated more citations per year (more than 200 after 2010) followed by fuzzy method, clustering, ICA, and PLS, with around 100 citations after 2010). Wastewater treatment process improvements due to the application of this plethora of analysis tools may have been a driving force behind the steadily increasing interest in the field. Papers published on SOM, Regression, SVM and Qualitative features recognition receive between 50 and 100 citations per year. Control charts and mass balance computations receive less than 50 citations per year due to the fact that only a limited number of papers on these topics are published (one paper per year in the best case). In addition, the papers on knowledge generation and management (decision trees, rule induction, ontologies), receiving less than 20 citations per year, have retained the same level of interest through the years. The reason is either that they are highly specific to solving only select problems, that they lack novelty or that their potential cannot be efficiently exploited (e.g., ontologies, CBR). There is a slight increase in the application of ontologies but this is far less popular than in other sectors.

Popularity of papers among scientists

The most influential method has been the ICA method with a citation rate (the ratio between number of citations and number of papers) of 63 (in 2015), followed by SVM with a rate of 51. PCA and CBR obtain a citation rate of 38. Most of the other techniques lie in the range between 20 and 40, except for control chart and mass balance that result in citation rates less than 20. We hypothise that "newer" techniques are cited more because they are new. In particular control charts and mass balances may be considered general knowledge and therefore not cited as much.

Bringing techniques into practice

With regards to basic and advanced information extraction wastewater treatment process improvements due to the application of this plethora of techniques may have been a driving force behind the steadily increasing interest in the field. Those methods that rise in popularity in academia, are often gaining interest for commercial implementation. Actually, only 9% of the publications clearly stated that the methods were implemented at full-scale (e.g. controller running at real-time). Still, it is difficult to keep track on which methods made it into practice, as this is out-of-the-scope of scientific literature, and this would require a targeted search through commercial products. The other publications remain as an academic exercise, even if full-scale data were used, which is the case for 46% of the papers.

Several limitations have been highlighted here that currently prevent the application of advanced and human-interpretable information extraction tools into practical applications. This may provide insight into the research and implementation challenges ahead. The most important ones are considered to be i) lack of validation, ii) lack of guidelines, iii) a gap between statistical analysis and engineering, iv) a lag in the adjustment of educational programs for a data-rich future and v) lack of proper knowledge generation and management.

CONCLUSIONS

The study demonstrates that EU has been leading computer-based method development with a presence in 61% of the reviewed papers. The most cited techniques are found to be artificial neural networks, principal component analysis, fuzzy methods, clustering, independent component analysis, and partial least squares regression. Even though there has been progress on techniques related to the development of environmental decision support systems, knowledge discovery and management, the research sector is still far from delivering systems that smoothly integrate several types of knowledge and different techniques of reasoning. We hope that our initial assessments serve as a discussion starter on training in, selection of, and desired features for computer-based methods/tools to transform passive data into a reliable and timely source of information.

REFERENCES

Hadjimichael, A., Comas, J., Corominas, L., 2016. Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector. AI Commun. 29, 747–756. doi:10.3233/AIC-160714.

FIGURES

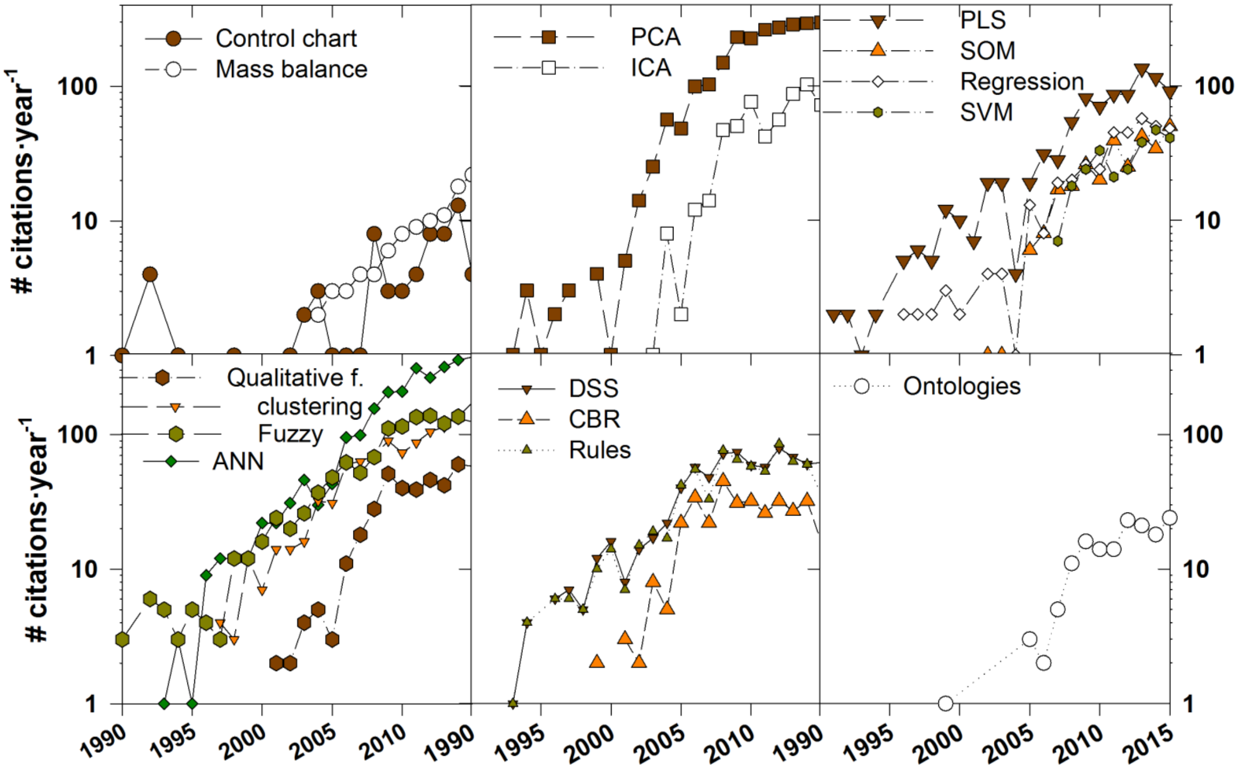

Figure 1. Number of citations per year and per technique