

Automated State Estimation for Summarizing the Dynamics of Complex Urban Systems Using Representation Learning

Maira Alvi¹, Tim French¹, Philip Keymer², Rachel Cardell-Oliver¹

¹Department of Computer Science and Software Engineering, The University of Western Australia, Australia

²Australian Centre for Water and Environmental Biotechnology, University of Queensland, Brisbane, Australia
maira.alvi@research.uwa.edu.au, tim.french@uwa.edu.au, p.keymer@uq.edu.au, rachel.cardell-oliver@uwa.edu.au

Abstract

Complex urban systems can be difficult to monitor, diagnose and manage because the complete states of such systems are only partially observable with sensors. State estimation techniques can be used to determine the underlying dynamic behavior of such complex systems with their highly non-linear processes and external time-variant influences. States can be estimated by clustering observed sensor readings. However, clustering performance degrades as the number of sensors and readings (i.e. feature dimension) increases. To address this problem, we propose a framework that learns a feature-centric lower dimensional representation of data for clustering to support analysis of system dynamics. We propose Unsupervised Feature Attention with Compact Representation (UFACR) to rank features contributing to a cluster assignment. These weighted features are then used to learn a reduced-dimension temporal representation of the data with a deep-learning model. The resulting low-dimensional representation can be effectively clustered into states. UFACR is evaluated on real-world and synthetic wastewater treatment plant data sets, and feature ranking outcomes were validated by wastewater treatment domain experts. Our quantitative and qualitative experimental analyses demonstrate the effectiveness of UFACR for uncovering system dynamics in an automated and unsupervised manner to offer guidance to wastewater engineers to enhance industrial productivity and treatment efficiency.

Introduction

Complex urban systems, including wastewater treatment plants (WWTPs), encompass various interconnected processes that span multiple time scales to provide an offset against environmental implications (Alvi et al. 2022). In order to maintain sustainable treatment objectives, these facilities monitor key system parameters through online sensor systems. However, these sensors only provide limited information about the global state of the systems. The development of techniques that can provide a global perspective of the underlying system (also known as state estimation) facilitates enhanced decision-making by wastewater engineers.

To date, state estimation methods in wastewater leverage basic statistical and physical modelling techniques. These

techniques couple critical treatment source’s stochastic behavior to determine the system’s dynamics. However, they suffer from the steep costs of developing high-fidelity algorithms. This involves the need for detailed descriptions of all physicochemical and biological factors affecting the dynamics of the different processes in a treatment system. Often, simplifying assumptions are made to allow desired mathematical tractability and computational efficiency. However, these simplifications potentially introduce errors and systematic bias in the state estimation (Jones, MacGregor, and Murphy 1989).

On the other hand, clustering can be a parallel choice to uncover hidden interesting dynamics/states of the underlying system. Only a handful of works have explored their potential in the wastewater domain. However, these works have focused on objectives including regression and classification (Sharghi et al. 2019; Zhao et al. 2016). This is partially because clustering approaches work well in lower-dimensional feature spaces. Thus, we require techniques to transform non-linear temporal data into a representative lower-dimension feature space.

This work proposes a novel approach to provide a high-level summary of WWTPs. At its core, it mitigates the higher dimensionality of feature space by an attention-based mechanism, i.e., features are weighted according to their saliency. Thus, it provides a lens that enables several critical insights into the system state. This includes extraction of state trajectories with Markov chain models, possible anomaly identification and discovery of salient recurrent patterns. Consequently, the insights equip field experts to have a better understanding of the system and allow them to make intelligent and informed decisions.

The contributions of this work are summarized below,

1. An unsupervised feature attention method (UFA) that leverages explainable artificial intelligence (XAI) to attend features.
2. A ‘Compact Representation’ (CR) method inspired by language embedding models to reduce feature dimensions and uphold temporal relationships.
3. Demonstration of key insights from simple clustering over compact representation. This includes anomaly detection, state trajectories, identification of repeating patterns, and validation from wastewater experts.

The next section provides a summary of existing literature before a discussion of the proposed methodology, experimental setup and results.

Literature Review

State estimation provides a holistic view of the underlying process dynamics that are not measurable directly with the online sensors. At its core, it effectively bridges the gap of limited observability and provides a comprehensive insight into the intricate internal dynamics of wastewater treatment systems. Therefore, it could be useful for process monitoring, fault detection, and diagnosis. Existing work for state estimation is normally based on using principles techniques such as statistical or physical process modelling.

These include classical observers, the Kalman Filter (KF), the extended KF, and Luenberger which are developed with the perfect knowledge of the process kinetics. However, wastewater processes are subject to substantial external dynamic disturbances, strong non-linearities, widely varying system time constants for different processes, and treatment targets (Burton et al. 2013). Thus, constraining with known process kinetics leads to uncertainty and significant bias in estimating the unmeasured states (Dochain 2003).

Another class of first principle methods for state estimation is mechanistic models. They leverage the mass balance and equilibrium relationships of the underlying biochemical and chemical processes. Fundamentally, there are behaviors in wastewater treatment systems that cannot be (with reasonable computational resources) represented using mechanistic models. To minimize the computational cost, a simplified representation is often selected for numerical and mathematical tractability that makes their predictive ability susceptible to the limitations of these assumptions (Alvi et al. 2023). Therefore, there is room to design alternative solutions for state estimation of these intricate non-linear processes.

Clustering is a possible alternate choice to estimate the system’s state without imposing any assumption about underlying dynamics. It has been widely utilized for exploratory analysis in various fields and can yield interesting patterns (Aghabozorgi, Shirkhorshidi, and Wah 2015). It relies on the similarity or the distance between data points. However, distance metrics disperse their discriminative ability for high dimensional feature space and, as a result, yield ineffective clusters. To circumvent this problem, dimensionality reduction is used as a pre-processing step. This includes Principle Component Analysis (PCA) (Choi and Park 2001) and correlation analysis (Tomperi et al. 2017). The correlation-based analysis overlooks the causality of the system, which is essential for any dynamical system. Similarly, PCA struggles to capture time-evolving and time-shifted patterns (Li and Prakash 2011).

To infer faithful representations of data, deep clustering can be employed. Unlike formerly discussed techniques, the proposed methodology belongs to this category (i.e., deep clustering). Motivated by the fact that the clustering algorithm’s performance often relies on effective features (Li and Prakash 2011), we propose a feature-centric representation learning framework. It pays attention to the salient features in the data under the lens of explainable AI (Adadi

and Berrada 2018) and employs a seq2seq encoder-decoder model to learn a compact embedded space that generally lies on many non-linear manifolds. The learned representation aids the clustering algorithm to yield interesting system dynamics. See the results section.

Proposed Methodology

This work proposes a novel method to extract discrete states of a time-varying wastewater treatment plant. In contrast with earlier proposed approaches, it pays attention to salient features before learning a compact representation.

Attention mechanisms have demonstrated performance boosts in several machine learning tasks, including language processing. However, for system analysis, defining attention is not trivial. This work proposes attention-based feature weighting. It first performs a clustering with the help of K-Means algorithm (MacQueen et al. 1967). Here, the contribution of individual features to clustering decisions is leveraged to determine respective saliency weights for feature attention. Finally, the attended features are utilized to learn a compact representation in an unsupervised manner with the sequence-to-sequence encoder decoder-based network. A holistic view of the proposed algorithm is provided in Fig. 1. The compact representation demonstrates improved clusters as indicated qualitatively and quantitatively with the help of different metrics.

The identified discrete states are further cast in a symbol sequence to study state transitions and event mining to unveil interesting process dynamics and thus enrich knowledge discovery. For a concise representation of the proposed scheme contribution, first, a formal description of the attending feature is provided, followed by a temporal representation learning algorithm. The overall approach is summarized in Algorithm 1.

Unsupervised Feature Attention (UFA)

Let’s say we have a time-series dataset $\mathbf{X} \in \mathbb{R}^{m \times n}$ with ‘ m ’ time steps and consistent of ‘ n ’ co-evolving features. With the broader aim to identify useful states in the system. Algorithm 1 (line 1) first clusters the data using K-means to get clusters with assigned ‘pseudo-labels’ s^* . Here, the number of clusters (k) was identified with the standard elbow method. To identify the saliency of features, the algorithm proposes a proxy task of classification. The computed pseudo-labels ‘ s^* ’ are used to train a classifier $\mathcal{K}(\cdot)$ (line 2). For simplicity, the experiments utilized a Random Forrest Classifier. However, the algorithm is generalizable to any classifier. The classifier thus essentially maps the individual samples, \mathbf{x}_i , to respective pseudo-label as, $\mathcal{K}(\cdot) : \mathbf{x}_i \rightarrow s_i^*$. Finally the trained classifier $\mathcal{K}(\cdot)$ is used to compute feature-attribution scores ‘ $\rho \in \mathbb{R}^n$ ’ for all possible classes. For this purpose, algorithm 1 (line 3) leverages a model agnostic technique SHAP (SHapley Additive exPlanations) (Lundberg and Lee 2017) to yield the feature’s relative significance in clustering task. The relative importance score is then normalized for numerical consistency as $\rho = \rho / \|\rho\|$ (line 4).

Finally, the learned saliency of features is used to give attention to the entire data set and produces $\mathbf{X}^* \in \mathbb{R}^{m \times n}$ as

Algorithm 1: UFA-CR

Input: Data set $\mathbf{X} \in \mathbb{R}^{m \times n}$, number of clusters k
Output: Cluster results \mathbf{s}
 1: $\mathbf{s}^* \leftarrow \text{Kmeans}(\mathbf{X}, k)$
 2: $\mathcal{K}(\cdot) \leftarrow \text{Classifier}(\mathbf{X}, \mathbf{s}^*)$
 3: $\boldsymbol{\rho} \leftarrow \text{SHAP}(\mathbf{X}, \mathcal{K}(\cdot))$
 4: $\boldsymbol{\rho} = \boldsymbol{\rho} / \|\boldsymbol{\rho}\|$
 5: $\mathbf{X}^* \leftarrow \mathbf{X} * \boldsymbol{\rho}$
 6: $\mathbf{Z} \leftarrow \text{Encoder}(\mathbf{X}^*) \in \mathbb{R}^{m \times o}$
 7: $\mathbf{s} \leftarrow \text{Kmeans}(\mathbf{X}, k)$
 8: return \mathbf{s}

shown on line 5. It is worth noting that while the dimensionality of the data stays the same, attention boosts the important features and diminishes others. This intuitively allows the algorithm to learn a representative and yet compact representation of original features by a sequence-to-sequence deep learning model discussed in the next subsection.

Compact Representation (CR)

Clustering techniques are known to extract distinguishable patterns relatively well in lower dimensional vector spaces (Alelyani, Tang, and Liu 2013). Thus, the second part of proposed algorithm includes learning a compact and representational feature space, $\mathbf{Z} \in \mathbb{R}^{m \times o}$ from attended features $\mathbf{X}^* \in \mathbb{R}^{m \times n}$ where ‘ o ’ is much less than ‘ n ’. Reducing the dimension of non-linear and temporal features is challenging for algorithms that assume linear relationships like PCA. The algorithm proposes a deep learning pipeline that learns a compact representation and upholds the temporal relationship. The developed pipeline is essentially a sequence-to-sequence model that logically consists of an encoder and decoder blocks and is trained in a completely unsupervised manner.

The algorithm defines non-linear mappings as encoder : $\mathbf{X}_i^w \rightarrow \mathbf{z}_i$, and decoder : $\mathbf{z}_i \rightarrow \tilde{\mathbf{X}}_i^{w^*}$ where $w, w^* \in \mathbb{N}^+$ is window sizes of past observation and future steps respectively. Here, $\mathbf{z}_i \in \mathbb{R}^o$, presents the learnt representation of ‘ i_{th} ’ sample. The encoder consists of two LSTM layers followed by a dense fully connected layer. The LSTM layers enable it to learn temporal relationships, while the dense layer assists in dimensional compression. Similarly, the decoder simply consists of two LSTM layers that produce a sequence of the desired length from a given compact vector from an encoder. The parameter weights of the encoder-decoder network are learnt through optimization of reconstruction loss $\mathcal{L}(\cdot)$ of Mean Square Error and thus learn a compact representation of input data. The entire pipeline is summarized in Fig. 1. The first phase performs unsupervised feature attention to prepare a weighted feature representation that afterwards enables the sequence-to-sequence model to learn a compact representation. The learned representation enables better clustering. It is qualitatively validated by the t-SNE (Van der Maaten and Hinton 2008) plots in Fig. 2 that the proposed method can learn a feature-centric representation of data that aids the clustering algorithm to generate meaningful clusters, in contrast to the existing approaches.

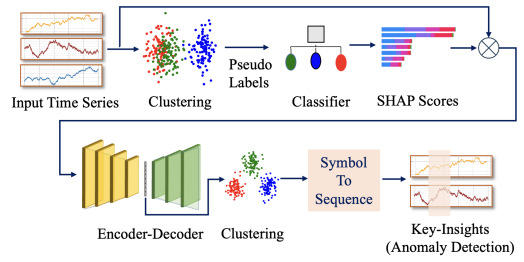


Figure 1: Illustration of the proposed approach.

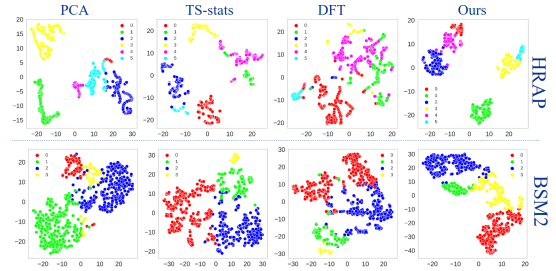


Figure 2: t-SNE visualization of the clustering performance of different approaches as indicated on the top, on HRAP and BSM2 data sets.

The empirical evaluation reaffirms the effectiveness and utility of the proposed method.

Experimental Setup

This section first discusses the utilized data sets, different comparison methods, their implementation details and respective evaluation metrics.

Data Sets

High Rate Algae Pond (HRAP): This data set was collected from continuous algae raceways with a surface area of $3.5m^2$ and a weather station at the Luggage Point sewage treatment plant in Pinkenba, Queensland (provided by Urban Utilities). It is a two-phase secondary wastewater treatment process that removes organics in an up-flow anaerobic sludge blanket reactor and then further treats the effluent in algae raceways. Weather station data and online measurements were sampled at 15-minute intervals, and offline measurements were collected daily through laboratory analysis. Daily samples captured between 25 July 2020 to 30 July 2021 were selected for this work so that online measurements and weather station data had corresponding laboratory data.

Benchmark Simulation Model 2 (BSM2): BSM2 is a publicly available Simulink model that provides an environment to define plant layout, influent loads, performance criteria, and test procedures (Jeppsson et al. 2007). It is a plant-wide representative model of activated sludge (ASM1 (Henze et al. 1987)) and anaerobic digestion (ADM1(Batstone et al. 2002)) process. It incorporates daily, weekly, and seasonal variations to mimic the dynamics of real WWTPs with high fidelity. To generate effluent data, the MATLAB/Simulink

implementations are run in two stages, in open-loop mode with constant influent to reach a steady initial state. Afterward, in a closed-loop with dynamic influent to generate the required process data for 609 days.

Comparison Methods

The proposed method is compared with the following popular techniques.

1. **Principal Component Analysis (PCA):** is a linear dimensionality reduction technique that endeavors to preserve maximum variation. It transforms the input variables into uncorrelated variables called Principal components that are ranked in accordance with the amount of variance they can explain. In our experiments, K-means clustering was performed on the top p principal components of data using PCA, where p is set to keep 85% variance of the data.
2. **Discrete Fourier Transform (DFT):** is another popular technique that reduces dimensions of time series data using the frequency domain (Agrawal, Faloutsos, and Swami 1993). For real-valued signals, the coefficients follow Hermitian symmetry, therefore half of them are considered redundant. Following this strategy, in this work empirical evaluation, a rolling window of 5-time steps was used. This effectively produced 5 Fourier coefficients from which the last two were discarded. PCA then processes these Fourier coefficients with a variance of 85% and finally, K-means identifies the clusters.
3. **Time-series Statistics:** Time series statistical information is known to provide an alternative description of data. This work explores the prominent statistics including skewness, trend (slope), kurtosis, mean, and maximum value. The empirical results indicated the best performant variables to be maximum value and trend. Therefore the included results have only considered this subset.

Implementation Details

For fair analysis, same feature set was employed as an initial step for all the discussed techniques. All experiments were run in Python 3.8 on Ubuntu 18.04 (Bionic Beaver). For PCA (Pedregosa et al. 2011), DFT (Harris et al. 2020), and TS-stat (Virtanen et al. 2020) libraries were employed.

The proposed approach utilized data with a train/test split of 90:10, in contrast to other discussed techniques. For UFA, Random Forrest classifier was employed using library (Pedregosa et al. 2011), with 100 trees, random-state of 42 and maximum features set to 'sqrt', and for feature-attribution score SHAP (Lundberg and Lee 2017) was employed with the default setting. For CFR seq2seq model is employed with an encoder-decoder architecture trained in Keras, a high-level API of TensorFlow (Abadi et al. 2016), using NVIDIA GeForce RTX 3080 GPU. For HRAP and BSM2, encoder is composed of 2 Long Short-term Memory (LSTM) layers with 100 cells and return sequence set to True and 'tanh' activation. The output of second LSTM layer is passed through a Dense layer with 6 units to get a compact representation of features. It is then processed by Decoder (i.e., two LSTM

layers with 100 cells) to predict two time step future sequence. It is important to note that, encoder takes 3 past steps while decoder predicts two steps ahead. To make our analysis fair, we employed a 5 day window for other comparative techniques. The Adam optimizer is used with learning rate of 10^{-3} and decay factor of 0.9 and mean squared error loss is used. We set the batch size 256, number of epochs 300. For clustering, all the discussed techniques employed K-means (Pedregosa et al. 2011) and t-distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton 2008) to map the learned representation into a 2D plot.

Evaluation Metrics

To validate clustering performance, we employ the standard intrinsic evaluation metrics, namely; Silhouette, Davies-Bouldin (DB), and Calinski-Harabasz (CH) score (Liu et al. 2010), and performed qualitative analysis with field experts.

Results & Discussion

In this section, we first provide qualitative analysis followed by the quantitative analysis of the proposed approach.

Qualitative Analysis

HRAP: The process data collected from the algal wastewater treatment system was clustered into six homogeneous groups as shown in Fig. 2. In the figure, three clusters are distinctly separate from the other three, indicating significant differentiation between the two groups. This contrast was confirmed by wastewater engineers, who confirmed that these two groups correspond to two distinct operational conditions triggered by changes in the system's feed and loading rate. The representative example of such dynamics captured by the proposed approach is depicted in Fig. 3. The first half of the time series shows a system with poor algal treatment performance, characterized by elevated ammonia levels and low pH, and consequently, the dominant microbiological activity in the system was bacterial during this time period. Meanwhile, the second half of the time series shows a healthy algal treatment process, which is evident from the high pH levels and diminished ammonia concentrations. The algal-dominated state of the system exhibits optimal treatment performance regarding nutrient removal. We turn next

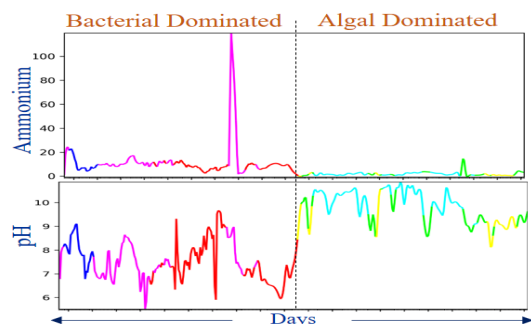


Figure 3: Illustration of two operating conditions, i.e., bacterial and algal dominated system. Color scheme is consistent with t-SNE visualization.

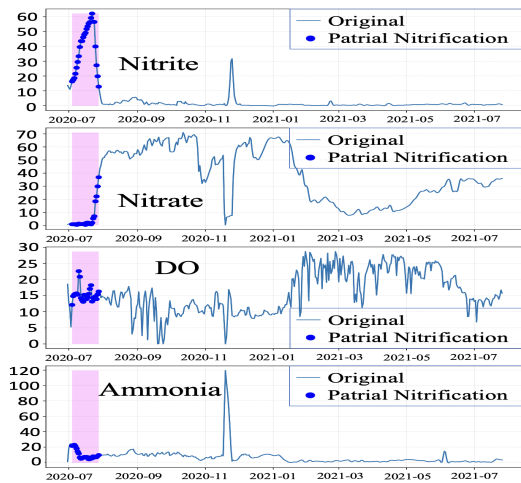


Figure 4: Representative results of the proposed method for partial nitrification. As illustrated, in contrast with time-series clustering, we cluster the entire system based on the co-evolving feature’s dynamics.

to other interesting dynamics of the system encapsulated in cluster 2 (blue see Fig. 2), which identifies an initial operational state in the system whereby the system was not performing nitrification efficiently as seen by the accumulation of nitrite while nitrate levels remained low (see Fig. 4), i.e. partial nitrification. The partial nitrification state is not a typical operational condition and is indicative that the microbial communities in the system have not yet developed the appropriate biological populations to allow for the full conversion of ammonia to nitrate. An intriguing observation emerges from the data as the nitrite peak occurs three days prior to the conclusion of the dark blue series, signifying the shift in the microbial populations in the bacterially dominant system (also validated in Fig. 3). This is supported by the accumulation of nitrite (NO_2^-) (partial nitrification), which results from the activity of Ammonia oxidising bacteria in the system. This behavior aligns with the earlier explanation regarding a shift in the system’s feed. Eventually, the culture developed a population of nitrite oxidising bacteria allowing for complete nitrification, evidenced by nitrate (NO_3^-) accumulation, also indicating that identification is not occurring. Due to the relatively slow nature of the upfront UASB treatment system, the change in the treated effluent quality of the two-stage treatment system was observed several days after the change in the system feed conditions. This is primarily due to the acclimatisation of the methanogenic population to the changed feed composition.

Furthermore, clusters 0 (red) and 4 (magenta) were mapped to an operating condition in which the system was fed with a high-strength influent related to a bacteria-dominated system, with cluster 4 mapping to conditions with higher biomass concentrations than cluster 0. The bacterial dominance is apparent from the high nitrate conditions and low dissolved oxygen and pH levels (refer to Fig. 4). Conversely, clusters 1, 3 and 5 (green, yellow and cyan) mapped to periods with a lower strength influent, where the system

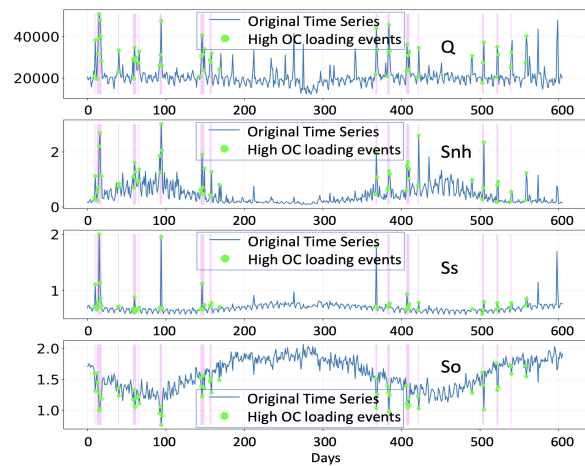


Figure 5: Illustration of High Organic Carbon (OC) loading events. High flow (Q) increases organic carbon loads and causes an increase in discharge organic carbon concentration (Ss). This consumes dissolved oxygen (So) that reduces oxidation of ammonia and results in higher levels of ammonia (Snh).

was dominated by algal activity, evidenced by the lower nitrate concentration and elevated dissolved oxygen and pH levels. Of the clusters mapping to the algal-dominated system, cluster 3 (yellow) appears to indicate conditions where the algal system performs sub-optimally where both the DO and pH levels are reduced, which is typically associated with overcast days.

In summary, the proposed methodology successfully identified conditions under which the system performed sub-optimally, as shown in cluster 1. Similarly, for the algal system, the distinction between two operational modes (high vs low strength influent) could be identified by grouping several clusters, and an abnormal condition (partial nitrification) was identified by one cluster. This shows promise in the ability of the system to identify potentially abnormal conditions which will be of interest to the wastewater treatment utilities operating these treatment systems.

BSM2: The proposed approach successfully identified 4 operational states of the generated wastewater treatment process. Three clusters represent the normal operating conditions of the process during different seasons, namely, normal operation during summer (cluster 0), normal operation during winter (cluster 2) and normal operation during summer-winter transitions (cluster 3). It is worth noting that high-level states represented by three clusters are validated by wastewater engineers, ensuring their accuracy in capturing significant system dynamics throughout different seasons. Due to the page limit, the illustrative examples of these three clusters are not included herein.

Building upon the validation of clusters, one cluster (i.e., Cluster 1) is of significant interest, as it effectively identifies the states where the wastewater discharge from the process exceeds target levels of ammonia (Snh) and organics (Ss), during conditions of increased flow (Q), with an associ-

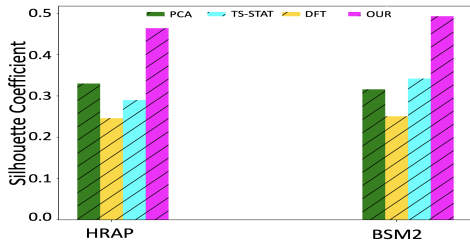


Figure 6: Comparison of silhouette coefficients.

Method	HRAP		BSM2	
	DB	CH	DB	CH
PCA	1.24	124	1.4	181
TS-stat	1.27	109	1	409
Our	0.88 <u>+29</u>	850 <u>+585</u>	0.76 <u>+24</u>	1175 <u>+187</u>

Table 1: Intrinsic measure DB and CH score for the top three techniques over HRAP and BSM2 data set. The best results are highlighted in bold, where DB (lower) and CH (higher) are better, and performance gain (%) over the closest competitor is underlined.

ated reduction in dissolved oxygen in the bioreactor (So), as shown in Figure 5. Although many of the exceedance events in the data set are short-lived, cluster 1 (high organic carbon (OC) loading event) successfully identified points prior to the event, showing that this may be able to provide an early indicator of the exceedance event.

Quantitative Analysis

Figure 6 summarizes the comparative results of the discussed techniques. It is evident that the proposed methodology categorically outperforms other methods with a performance gain of $\frac{0.46-0.33}{0.33} \times 100 = 39\%$ on HRAP and 44% on BSM2 over the closest competitor. Furthermore, we validate clustering performance using two other intrinsic measures. Results are summarized in Table 1.

Broader Application

The proposed approach has established its efficacy both in qualitative and qualitative aspects, as discussed above. We take a step further by leveraging these high-level states of the system to conduct event mining. We cast the cluster labels into a sequence of symbols and analyze events using various window sizes to uncover motifs, anomalies, and state transitions. Note the size of the window depends on the objective of process modeling. We use different sizes of windows to uncover different process dynamics and visualize their frequency using a histogram that revealed some interesting insights into the system, such as some combination of 3-day windows lasting longer, highlighting it captures seasonal changes and the system is quite stable. It also unveils some hidden patterns that only occurred once, which could be attributed to anomalous behaviors and, in some cases, to no worry state of the system. The six high-level states expanded to 19 unique states for delving deeper into the system’s dynamics. Figure 7 illustrates the state transitions matrix of

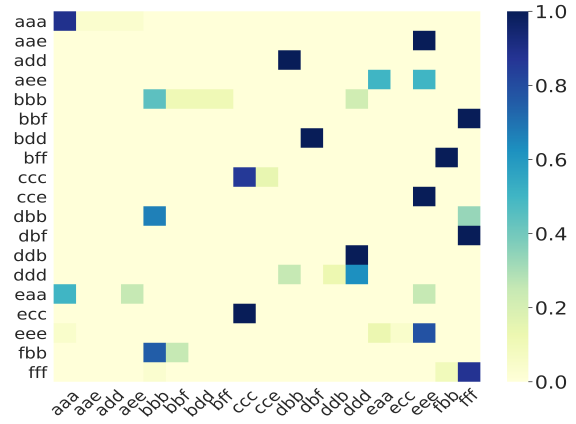


Figure 7: Probability state transition matrix of HRAP system.

the underlying system on a 3-day window. It is worth noting that we employ the Markov chain to compute the probability transition matrix. This summarization or additional states offers a comprehensive overview of the system. Leveraging this information, proactive decisions can be made based on the system’s probabilities, proving particularly advantageous in remote wastewater treatment processes where anticipation of upcoming events is crucial. Moreover, for sites attended infrequently by operators, this approach can bring significant benefits in terms of system management, particularly in highlighting times at which the process may move to a suboptimal state, allowing for intervention or rectification actions to be taken and thus minimising impacts of the suboptimal operational states.

Conclusion

This work presented a novel state estimation method to summarize the underlying system behaviour of wastewater processes in an unsupervised manner. In essence, it is a transformative paradigm for enhanced decision support of these complex non-linear systems. The developed system offers several key insights, including the identification of normal, anomalous, seasonal, under-performance, and optimal operating states of the system, as validated by industry experts. The detailed empirical evaluation over real-world and synthetic data sets validates its generalization. These insights are of great practical value and can facilitate wastewater engineers and experts in several ways. It can assist in the optimization of processes to achieve desired target outcomes. Similarly, having a better understanding of the boundary states of a system can better inform operational management to reduce the likelihood of detrimental (environmental and/or financial) failures.

Acknowledgements

This research was supported by the Cooperative Research Centre Projects (CRC-P) ‘Transforming wastewater treatment in regional Australia with robust technology for multiple benefits’ under Grant CRCPSIX000079.

References

- Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 265–283.
- Adadi, A.; and Berrada, M. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, 6: 52138–52160.
- Aghabozorgi, S.; Shirkhorshidi, A. S.; and Wah, T. Y. 2015. Time-series clustering—a decade review. *Information systems*, 53: 16–38.
- Agrawal, R.; Faloutsos, C.; and Swami, A. 1993. Efficient similarity search in sequence databases. In *Foundations of Data Organization and Algorithms: 4th International Conference, FODO'93 Chicago, Illinois, USA, October 13–15, 1993 Proceedings 4*, 69–84. Springer.
- Alelyani, S.; Tang, J.; and Liu, H. 2013. Feature selection for clustering: a review. *Data clustering: algorithms and applications*, 29(1).
- Alvi, M.; French, T.; Cardell-Oliver, R.; Batstone, D.; and Akhtar, N. 2023. Enhanced Deep Predictive Modelling of Wastewater Plants With Limited Data. *IEEE Transactions on Industrial Informatics*.
- Alvi, M.; French, T.; Cardell-Oliver, R.; Keymer, P.; and Ward, A. 2022. Cost Effective Soft Sensing for Wastewater Treatment Facilities. *IEEE Access*, 10: 55694–55708.
- Batstone, D. J.; Keller, J.; Angelidaki, I.; Kalyuzhnyi, S.; Pavlostathis, S.; Rozzi, A.; Sanders, W.; Siegrist, H.; and Vavilin, V. 2002. The IWA anaerobic digestion model no 1 (ADM1). *Water Science and technology*, 45(10): 65–73.
- Burton, F.; Tchobanoglous, G.; Tsuchihashi, R.; Stensel, H.; and Metcalf and Eddy Inc. 2013. *Wastewater Engineering: Treatment and Resource Recovery*. McGraw-Hill Education.
- Choi, D.-J.; and Park, H. 2001. A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process. *Water research*, 35(16): 3959–3967.
- Dochain, D. 2003. State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of process control*, 13(8): 801–818.
- Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; and Oliphant, T. E. 2020. Array programming with NumPy. *Nature*, 585(7825): 357–362.
- Henze, M.; Grady, C. L.; Gujer, W.; Marais, G.; Matsuo, T.; et al. 1987. Activated sludge model n 1. *IAWQ Scientific and Technical Report*, 1.
- Jeppsson, U.; Pons, M.-N.; Nopens, I.; Alex, J.; Copp, J.; Gernaey, K.; Rosén, C.; Steyer, J.-P.; and Vanrolleghem, P. 2007. Benchmark simulation model no 2: general protocol and exploratory case studies. *Water Science and Technology*, 56(8): 67–78.
- Jones, R.; MacGregor, J.; and Murphy, K. 1989. State estimation in wastewater engineering: application to an anaerobic process. *Environmental monitoring and assessment*, 13: 271–282.
- Li, L.; and Prakash, B. A. 2011. Time series clustering: Complex is simpler! In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 185–192.
- Liu, Y.; Li, Z.; Xiong, H.; Gao, X.; and Wu, J. 2010. Understanding of internal clustering validation measures. In *2010 IEEE international conference on data mining*, 911–916. IEEE.
- Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- MacQueen, J.; et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, 281–297. Oakland, CA, USA.
- Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85): 2825–2830.
- Sharghi, E.; Nourani, V.; Ashrafi, A.; and Gökçekub, H. 2019. Monitoring effluent quality of wastewater treatment plant by clustering based artificial neural network method. *Desalination and Water Treatment*, 164: 86–97.
- Tomperi, J.; Koivuranta, E.; Kuokkanen, A.; and Leiviskä, K. 2017. Modelling effluent quality based on a real-time optical monitoring of the wastewater treatment process. *Environmental technology*, 38(1): 1–13.
- Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
- Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3): 261–272.
- Zhao, Y.; Guo, L.; Liang, J.; and Zhang, M. 2016. Seasonal artificial neural network model for water quality prediction via a clustering analysis method in a wastewater treatment plant of China. *Desalination and Water Treatment*, 57(8): 3452–3465.