

Received June 25, 2019, accepted August 4, 2019, date of publication August 6, 2019, date of current version August 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933616

# Monitoring Influent Conditions of Wastewater Treatment Plants by Nonlinear Data-Based Techniques

TUOYUAN CHENG<sup>1</sup>, ABDELKADER DAIRI<sup>2</sup>, FOUZI HARROU<sup>3</sup>, (Member, IEEE),  
YING SUN<sup>3</sup>, AND TOROVE LEIKNES<sup>1</sup>

<sup>1</sup>Biological and Environmental Science and Engineering Division, Water Desalination and Reuse Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

<sup>2</sup>Signal, Image, and Speech Laboratory (SIMPA), Computer Science Department, University of Science and Technology of Oran-Mohamed Boudiaf (USTO-MB), Oran 31000, Algeria

<sup>3</sup>Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

Corresponding author: Tuoyuan Cheng (tuoyuan.cheng@kaust.edu.sa)

This work was supported by the King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR), under Grant OSR-2019-CRG7-3800.

**ABSTRACT** To operate wastewater treatment plants (WWTPs) with optimized efficiency, influent conditions (ICs) as initial states of inflow fed to WWTPs were monitored to identify potential anomalies that would trigger adverse events or system crash. To employ voluminous measurements for data-driven decisions, the non-linear, non-Gaussian, non-stationary, auto-correlated, cross-correlated, hetero-skedastic, case-specific nature of multivariate environmental datasets must be considered. This research proposed kernel machine learning models, the kernel principal components analysis based one-class support vector machine (KPCA-OCSVM) with various kernels, to learn anomaly-free training set then classify the testing set. A seven-years multivariate ICs time series was introduced with exploratory analysis performed to reveal temporal behaviors and statistical properties. KPCA with polynomial kernels sufficiently output representative features, based on which OCSVM with Gaussian kernels sensitively and specifically identified anomalies in ICs that were previously omitted by WWTP operators. The proposed kernel algorithms surpassed previous linear PCA-based K-nearest-neighbors models, and improved outcomes with limited increase in computation cost. Without requiring linear, Gaussian, stationary, independent, and homo-skedastic qualities from data, the proposed flexible environmental data science approach could be transferred, rebuilt, and tuned conveniently for ICs from different WWTPs.

**INDEX TERMS** Wastewater, influent condition, machine learning, kernel, support vector machines, data visualization.

## I. INTRODUCTION

Environmental systems are complex. Wastewater treatment plants (WWTPs) are environmental systems where physical, chemical, and biological unit processes are convoluted [1]. Modern WWTPs have to function unceasingly while accepting incoming wastewater of volatile quantity and quality. Industrial WWTPs have to manage wastewater with frequent abrupt changes in case-specific composition and temperature, while municipal counterparts often face impact from rainfall and snowmelt. Though having limited storage for them to be buffered, inflow cannot be abandoned or rejected. Nonlinear

The associate editor coordinating the review of this manuscript and approving it for publication was Changsheng Li.

dynamic multivariate processes, rising operational costs, and stringent discharge regulations altogether ask for optimal efficiency from practitioners [2].

Influent conditions (ICs) are initial states of inflow fed to WWTPs. ICs would affect system states, ongoing process mechanisms, and final product (*i.e* treated effluent) quality. Monitoring, detection, isolation, and diagnosis of potential anomalies or faults in ICs could, at an early stage, avoid unexpected system crash, maintain steady product quality, support efficient downstream processes, improve WWTPs reliability and reduce labor costs [3].

Conventionally, mechanistic model-based or analytical methods are developed for WWTP forecast monitoring. Utilizing first principals, such classical deterministic models

could theoretically explain process mechanism, but they were based on ideal hypotheses, requesting prior parameters for calibration and would be challenged by ill-conditioned problems and costly high-dimensional computation in real applications. Expert systems, as established on knowledge bases, rely on subjective judgments from individuals and may lead to biased diagnoses.

Environmental data science is an emerging interdisciplinary field, that effectively addresses the intricacy inside environmental systems, and could provide promising solutions for WWTP monitoring. Ever updating instrumentation, control, and automation in WWTPs are producing quantities of multivariate time series data, which are often unexploited. This “data-rich, information-poor” dilemma is attributed to the lack of methodology to select the right algorithm for a given case, the lack of standard prototypical data processing procedures, and the lack of trained environmental data scientist or data science expertise among environmental scientists [4]. Moreover, the nonlinear, non-Gaussian, non-stationary, auto-correlated, cross-correlated, heteroskedastic, and case-specific nature of multivariate environmental time series data are all adding difficulty for researchers to construct both suitable and flexible models.

Data-based methods can perform systematic and objective exploration, visualization, and interpretation of data [5], identify essential factors, features or patterns, and endorse then optimize data-supported decision-making [6]. Validated data-driven monitoring methods could be transferred and shared conveniently among domain experts, in virtue of the versatile nature of data science models [7].

Machine learning is a remarkable multidisciplinary field, where methods could be implemented for fault detection. Artificial neuron network (ANN) displayed good nonlinear projection quality, high fault tolerance, flexible self-adaptation, and parallel computing efficiency, which could deal with complexity and high-dimensionality of implied knowledge in WWTP anomalies. ANN-based predictive control has been introduced by [8] for WWTP monitoring and was tested on benchmark simulation data. Feedforward neural networks have been applied by [9] to predict wastewater effluent ammonia-nitrogen contents. Hybrid ANN has been developed by [3] to predict influent biochemical oxygen demand, which otherwise was expensive and difficult to measure with sensors. The requirements of suitable sample size and network architecture, the following intensive computation, overfitting, local optimal solutions, and inexplicability were hurdles to overcome.

Latent variable methods applied in statistical process control, including independent component analysis (ICA) [10] and principal component analysis (PCA) [11], could treat multi-dimensionality and auto or cross correlations in WWTP records. By dimension reduction, input data were projected onto feature spaces of lower dimension, where statistics could be built to reveal characteristics of interest and identify abnormal conditions. In such cases, non-stationarity, non-normality, and non-linearity were issues to resolve. ICA may

be further affected by the number and sequence of input signals or independent components.

To address the nonlinear nature of wastewater ICs data, kernel methods should be emphasized, where projection onto higher dimensions enables good reencoding of the data when it lies along a nonlinear manifold. Kernel PCA (KPCA) has been introduced by [12] to detect abrupt events in wastewater systems, including pollutions and sensor faults. Kernel ICA has been studied by [13] to detect exterior disturbance from rainfall, and has shown competitive accuracy, efficiency, and reliability. Support vector machine (SVM) has been proposed by [14] to improve oxygen uptake rate on-line measurement and control. Kernel methods demanded comparably larger storage during computation, where parallel computing could offer potential solutions. For ICs data, improved classification efficiency could be anticipated by adopting kernel techniques.

This paper proposed an effective monitoring strategy merging the desirable characteristics of KPCA modeling with an unsupervised one-class SVM (OCSVM) scheme to distinguish normal from abnormal measurements. In this regards, KPCA was used to account for nonlinearities in the multivariate ICs data. KPCA may discover relevant patterns in the data by transforming problems into higher dimensions via kernel functions, enabling non-linear relationships to be revealed as approximately linear. OCSVM may quantify the dissimilarity between normal and abnormal nonlinear features for detection without making assumptions on the underlying data distribution. Tests on real ICs data from a local coastal municipal WWTP showed the effectiveness of the proposed approach. Previous research employed PCA-based k-nearest neighbor (PCA-KNN) algorithms to connect dimension reduction with robust machine learning approach for fault detection, and introduced radial visualization for fault diagnosis, whose performances were partly merged and compared in this paper. Exploratory data analysis with data visualization were involved in uncovering and interpreting historical influent behaviors.

The PCA and KPCA models were briefly presented in Section II. The proposed KPCA-OCSVM monitoring schemes were discussed in Section III. The performances of the recommended approaches were evaluated via real data, together with exploratory analysis and visualization in Section IV. Conclusions were drawn in Section V.

## II. LINEAR AND KERNEL PRINCIPAL COMPONENTS ANALYSIS MODELS

This section provided an overview of the linear PCA and the derived KPCA models.

### A. LINEAR PRINCIPAL COMPONENTS ANALYSIS (PCA) MODEL

Principal Components Analysis, as a dimension reduction approach, was a popular modeling framework to learn crucial features from multivariate data. By projecting the process variables into lower-dimensional subspace, PCA

enabled revealing the inherent cross-correlation among process variables [15].

Let us consider  $\mathbf{X} = [\mathbf{x}_1^T, \dots, \mathbf{x}_n^T]^T \in \mathbb{R}^{n \times m}$  be a scaled data matrix of recorded inflow conditions with  $n$  observations and  $m$  variables. Based on the PCA model, the data matrix  $\mathbf{X}$  can be expressed as a sum of the approximated matrix,  $\hat{\mathbf{X}}$ , and residual data,  $\mathbf{E}$ .

$$\mathbf{X} = \mathbf{T}\mathbf{W}^T = \sum_{i=1}^k \mathbf{t}_i \mathbf{w}_i^T + \sum_{i=k+1}^m \mathbf{t}_i \mathbf{w}_i^T = \hat{\mathbf{X}} + \mathbf{E} \quad (1)$$

where  $\mathbf{T} \in \mathbb{R}^{n \times m}$  represented a matrix of the principal components (PCs) and  $\mathbf{W} \in \mathbb{R}^{m \times m}$  was the loading matrix. In the presence of cross-correlation, the original multivariate data  $\mathbf{X}$  could be sufficiently preserved and approximated by the first ‘ $k$ ’ PCs (where  $k < m$ ). One necessary step here was to select the number of PCs. For this purpose, the cumulative percentage variance procedure was adopted due to its simplicity and accuracy.

The loading matrix was frequently calculated using singular value decomposition of the covariance matrix  $\mathbf{S}$  of the design matrix  $\mathbf{X}$ :

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T \quad \text{with} \quad \mathbf{Q} \mathbf{Q}^T = \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_m \quad (2)$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_m) = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$  was a diagonal matrix with eigenvalues of  $\mathbf{S}$  filled in decreasing order. The eigenvalue  $\lambda_i$  was equal to  $\sigma_i^2$ , the variance of the corresponding  $i$ -th PC  $\mathbf{t}_i$ .

### B. KERNEL PRINCIPAL COMPONENTS ANALYSIS (KPCA) MODEL

Kernel PCA, as an extension of the linear PCA with kernel tricks, allowed learning and revealing of nonlinear relationships among process variables. This non-linear dimension reduction algorithm was applied to feed further one-class SVM scheme, aiming at improved fault detection performance over traditional linear PCA-based monitoring methods. Core principles were sketched in Figure 1. By KPCA, firstly the input space was transformed via nonlinear mapping into a high-dimensional feature space, in which data were more linear. Then for dimension reduction, principal components were extracted by applying kernel tricks with inner products of nonlinear functions. Hence, procedures from

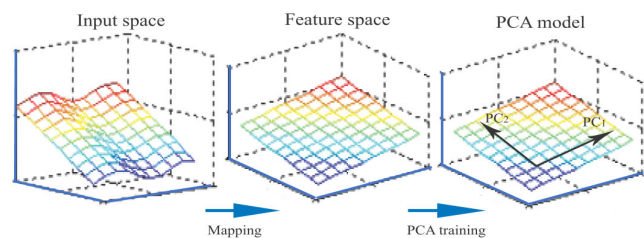


FIGURE 1. Space mapping in kernel principal components analysis (KPCA).

linear PCA could be largely inherited, and nonlinear optimization with high computation cost was saved. Moreover, hyper-parameters were less involved here compared to other nonlinear methods such as artificial neuron networks where architecture designing could be empirical or metaphysical.

Let us consider the original training dataset  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^m$ , where  $n$  was the sample number,  $m$  was the number of process variables. The feature space was constructed by using a nonlinear mapping:  $\mathbb{R}^m \xrightarrow{\Phi(\cdot)} \mathbb{F}^h$ , where  $\Phi(\cdot)$  was a nonlinear mapping function and  $h$ , as a huge positive integer, was the dimension of the feature space. Similar to PCA, the covariance matrix in the feature space  $\mathbb{F}$ ,  $\Sigma_{\mathbb{F}}$  could be calculated as:

$$\Sigma_{\mathbb{F}} = \frac{1}{n} \sum_{i=1}^n \left[ \Phi(\mathbf{x}_i) - \mathbf{m}_{\Phi} \right] \left[ \Phi(\mathbf{x}_i) - \mathbf{m}_{\Phi} \right]^T \quad (3)$$

where  $\mathbf{m}_{\Phi} = \sum_{i=1}^n \Phi(\mathbf{x}_i) / n$  was the sample mean in the feature space. Let  $\bar{\Phi}(\mathbf{x}_i)$  denote a mapped point after centering with the corresponding mean as  $\bar{\Phi}(\mathbf{x}_i) = \Phi(\mathbf{x}_i) - \mathbf{m}_{\Phi}$ . To find the principal components, we solved the eigenvalue decomposition problem in the feature space such that:

$$\lambda \mathbf{v} = \Sigma_{\mathbb{F}} \mathbf{v} = \frac{1}{n} \sum_{i=1}^n \left[ \bar{\Phi}(\mathbf{x}_i)^T \mathbf{v} \right] \bar{\Phi}(\mathbf{x}_i) \quad (4)$$

where  $\lambda$  and  $\mathbf{v}$  denote eigenvalue and eigenvector of the covariance matrix  $\Sigma_{\mathbb{F}}$ , respectively. By multiplying  $\bar{\Phi}(\mathbf{x}_i)$  from left in Eq.(4), and defining the kernel matrix  $\mathbf{K}$  or  $\bar{\mathbf{K}} \in \mathbb{R}^{n \times n}$  such that:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (5)$$

$$\bar{\mathbf{K}}(\mathbf{x}_i, \mathbf{x}_j) = \bar{\Phi}(\mathbf{x}_i)^T \bar{\Phi}(\mathbf{x}_j) \quad (6)$$

$$\bar{\mathbf{K}} = \mathbf{K} - \mathbf{K}\mathbf{E} - \mathbf{E}\mathbf{K} + \mathbf{E}\mathbf{K}\mathbf{E} \quad (7)$$

together with  $\boldsymbol{\alpha} \in \mathbb{R}^n$  to span the kernel PCs by feature space training samples, satisfying:

$$\mathbf{v} = \sum_{i=1}^n \alpha_i \bar{\Phi}(\mathbf{x}_i) \quad (8)$$

Then the eigenvalue decomposition problem was reformulated as [16]:

$$n\lambda \boldsymbol{\alpha} = \bar{\mathbf{K}} \boldsymbol{\alpha} \quad (9)$$

The eigenvectors identified in the feature space  $\mathbb{F}$ , namely nonlinear kernel PCs, would characterize nonlinear processes. Since the number of eigenvectors was the same as the number of samples, it was multiple times more than the number of linear PCs that could be offered by the conventional PCA.

Various kernel functions were available in literature [17], including polynomial kernels:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d, \quad d \in \mathbb{Z}^+ \quad (10)$$

cosine kernels:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\langle \mathbf{x}_i, \mathbf{x}_j \rangle}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} \quad (11)$$

and radial basis functions (RBF):

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\delta^2}} = e^{-\gamma \|x_i - x_j\|^2}, \quad \gamma \in \mathbb{R}^+ \quad (12)$$

where  $2\delta^2 = \frac{1}{\gamma}$  was the width of the Gaussian kernel.

### III. THE PROPOSED KPCA-BASED ONE CLASS SUPPORT VECTOR MACHINE (OCSVM) MODELS

#### A. OCSVM MODELS

One-class support vector machine was an unsupervised machine learning classifier and a special case of multi-class SVM. OCSVM learned decision rules from anomaly-free data, and then applied to classify new test data as either comparable or different from the learned anomaly-free case [18]. OCSVM would project the input data via kernel functions into a high dimensional space where a hyper-plane could be built for classification. This transformation permitted the projected data to be relatively linear, which made it easier to categorize. To infer the properties or features of normal cases, the pre-defined hyperplane would help to perform classification, where the decision function,  $f(x)$  would determine whether a new observation lies within the hyperplane side as inliers (normalcies,  $f(x) = +1$ ) or outliers (abnormalcies,  $f(x) = -1$ ).

We let  $x_1, \dots, x_j \in \mathcal{D}$  and  $j \in [1, k]$  be the training dataset. OCSVM mapped input data into the high-dimensional feature space  $\mathcal{F}$  via kernels such as the radial basis function (RBF) (Eq. 12). As illustrated in Figure 2(a), the decision rule  $f(x)$  aimed to maximize the Euclidean distance between the origin and the separating hyperplane  $\mathcal{H}$ , which separated the training data in the features space  $\mathcal{F}$ . Therefore we have the objective function  $f(x)$  expressed as:

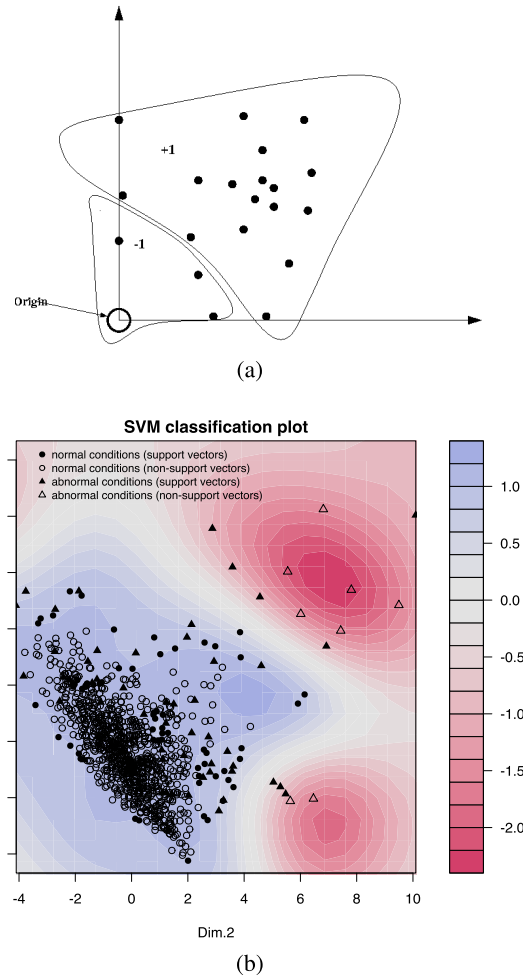
$$f(x) = \text{sign}(\langle w, \Psi(x) \rangle - \rho) \quad (13)$$

where  $w$ ,  $\rho$  and  $\Psi$  represented respectively a weight vector, an offset, and a feature mapping  $\mathcal{D} \rightarrow \mathcal{F}$ . Here,  $w$  and  $\rho$  could be determined by solving the following quadratic optimization problem.

$$\begin{aligned} \min_{w \in \mathcal{F}, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \quad & \frac{1}{2} \|w\|^2 + \frac{1}{\nu l} \sum_i^l \xi_i - \rho, \\ \text{s.t.} \quad & \langle w, \Psi(x) \rangle \geq \rho - \xi_i, \quad \xi_i \geq 0, \end{aligned} \quad (14)$$

where  $\nu \in [0, 1]$  represented the parameter that defined the solution.

For illustration purpose, a simplified SVM classification was plotted in Figure 2(b), where an SVM was built based only on the first two PCs for ICs classification. The diverging color reflected model prediction. Normal conditions as in triangles were intensely centered around the bottom left red corner. Anomalies as in circles were generally outside the red region and scattered among blue areas. The white band around zero delineated decision boundaries formed in this model. The capability of SVM to tackle nonlinearity by forming nonlinear hyperplanes via kernel functions could be witnessed.



**FIGURE 2. (a) One-class support vector machine (OCSVM). The hyperplane was formed amid two classes, with maximized distance from the origin; (b) Simplified SVM for illustration, predicting anomalies from only first two PC scores, trained by radial basis function (RBF) kernels. Support vectors functionally affecting the decision boundaries were marked solid while trivials were hollow.**

#### B. KPCA-OCSVM FOR ENHANCED NONLINEAR INFLUENT CONDITIONS (ICS) MONITORING

The proposed KPCA-based OCSVM approach for detecting abnormal changes in multivariate IC time-series was briefly described in this section and illustrated in Figure 3. KPCA possessed simplicity and capability to extract relevant features in multivariate time series data appropriately, and therefore was considered as an efficient approach for capturing important features in high-dimensional nonlinear processes. Features (loadings on PCs) extracted from multivariate IC time series by KPCA could be offered to train OCSVM in an unsupervised manner, which would further tackle nonlinear and non-Gaussian properties flexibly, for not assuming any distribution of the underlying data. By introducing radial basis kernel function in the projection of OCSVM, we could access nonlinear hyperplanes as decision boundaries to differentiate between normal observations and anomalies.

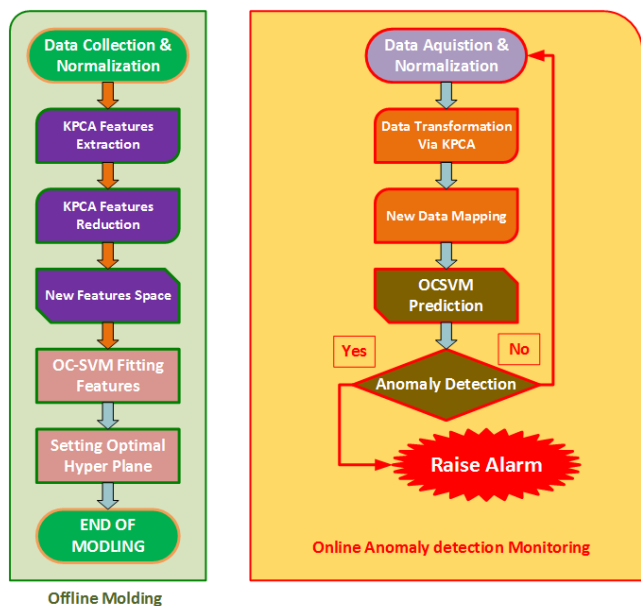


FIGURE 3. KPCA-OCSVM for improved nonlinear process monitoring.

In the proposed approach, we expect the KPCA would sufficiently learn and accurately describe the anomaly-free-only multivariate time series, the training IC dataset, and produce scores enabling the training process of OCSVM. Then the OCSVM would sensitively and specifically capture the presence of anomalies in the testing IC dataset. Several kernel functions were examined with different numbers of PCs in the KPCA step, while OCSVM was trained using the most popular flexible RBF kernel with fixed  $\nu = 0.01$  and  $\gamma = 0.1$  in all scenarios. By connecting the two models, we hypothesized the original non-Gaussian, linearly inseparable influent conditions data could be treated well.

#### IV. RESULTS AND DISCUSSION

##### A. EXPLORATIVE ANALYSIS AND VISUALIZATION OF A HISTORICAL MULTIVARIATE IC DATASET

To verify the hypothesis, one dataset from full-scale WWTP was engaged for training and testing. In this dataframe, operators have kept seven-years records of twenty-one variables including various flow quantity and water quality values. Measurements were performed on samples taken from the headwork of WWTP in order to maintain compliance with local regulations and standards (Figure 4). The monitored plant was receiving municipal wastewater, which was representative of cases in the literature. Though maintained with laborious effort and duly examination by local technicians, more than a hundred abnormal influent conditions were not detected. All those anomalies have caused negative effects on the process due to various reasons as reported by the practitioners, which let the classification and soft sensing research feasible and necessary. R package Amelia was imported to impute a few missing data (132 out of 63950, less than 1%) during the preprocessing step [19]. To explore the dataset,

descriptive analysis and visualization were conducted, during which RStudio with R packages (kernlab, cluster, factoextra, ggplot2, ggTimeSeries, gplots) were employed [20]–[26].

To facilitate visualization, anomaly-free-only dataset grouped by clustering large applications (CLARA) algorithm was further played together with dimension reduction (PCA), as shown in Figure 5(a) with unit aspect ratio. Normal ICs were clustered to five typical classes by which similar ICs were grouped within each category and hence could be represented by their corresponding medoid. ICs of each medoid and the average values of anomalies were shown in Table 1. Clustered data were further submitted to PCA via which scores on the first two PCs were plotted. It was shown in the axes of Figure 5(a) that the first two PCs from linear PCA accounted for only around 40% variation, implying the high dimensionality and complexity of the dataset. The envelope or outline of all 2401 normal ICs as a whole was far from the elliptical shape, demonstrating the non-Gaussian characteristic. Though not separated well in the PCA plot, the fifth cluster was located distinctly from others, which could be attributed to their high conductivity, total dissolved solids (TDS), magnesium hardness, chloride content but low fat-oil-grease (FOG) and nitrate as listed in Table 1.

To investigate temporal behavior, the calendar plot in Figure 5(b) was employed, where entries from the ICs were sequentially vertically aligned, circulated by weeks and months, and colored according to their clusters. Interestingly, clusters showed temporal agglomerative behavior. The first cluster was popular from 2011 to 2013, especially from June to October. Later on, the second cluster replaced the first cluster from 2014 to 2016 - both of the above represented typical cases in summer. The third cluster presented from 2012 to 2017, primarily from November to April, could represent cases in winter. The fourth cluster was intensively concentrated in 2011 from February to June, and surrounded most of the anomalies, with least inflow compared to others. At the beginning of 2012, there was a gathering of consecutive anomalies that were recorded as ‘total alkalinity over the limit’ by the operators. Such agglomerative phenomenon (*i.e* lack of complete randomness in different variables on different dates) reflected the existence of non-Gaussian, non-stationary, auto-correlated (seasonal recurrent), cross-correlated, and heteroskedastic characteristics in environmental datasets.

To summary ICs as univariates, wastewater quality index (WWQI) given by [27] was introduced, of which higher values represented better qualities. Calendar plot of WWQI was displayed in Figure 6. WWQI of medoids and the averaged value of anomalies were listed in Table 1. A significant low WWQI region was observed in 2011, coinciding with the fourth cluster, whose medoid owned WWQI even lower than anomalies. The fifth medoids showed highest WWQI. ICs in summer (first and second cluster) possessed higher WWQI than ICs in winter (third cluster). Probably higher usage of water in summer inflated the inflow and diluted its pollutant contents to some extent. The methods above



FIGURE 4. (a) Overview of the wastewater treatment plant; (b) Fine screen inside the headwork, where influent was filtered.

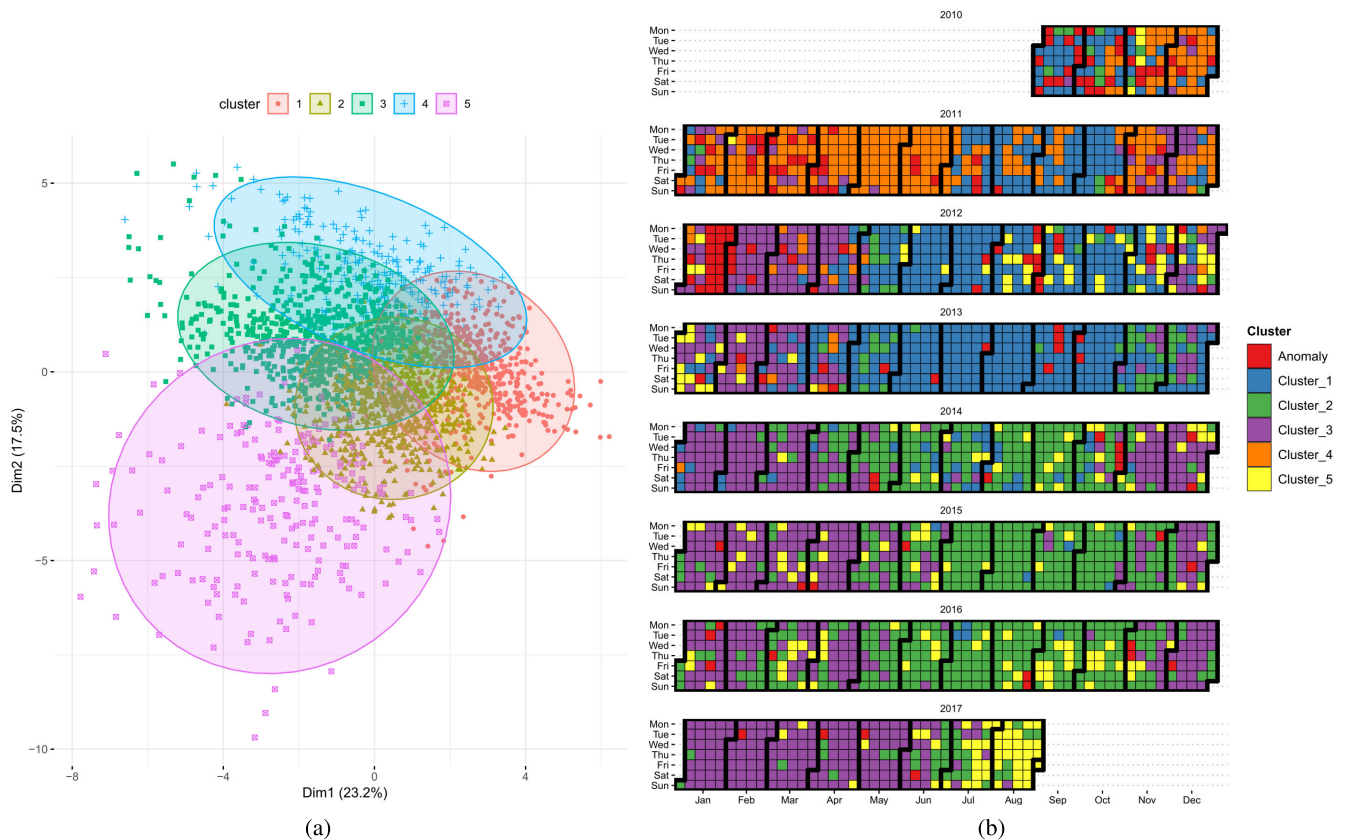


FIGURE 5. (a) Scatterplot of the dataset via PCA and clustering large applications (CLARA) algorithm; (b) Calendar plot of temporal CLARA result.

supported efficient analysis of ICs and yielded concise information. Moreover, practitioners could gain experience by understanding such visualizations. They would be vigilant during the winter or when the second cluster appeared, since the conditional probability of anomaly was higher in those cases.

By further examining Table 1, it was exhibited that anomalies averagely had smaller WWQI and lower temperature

as expected, since abnormal events were observed more in winter and were related to worse influent qualities. Moreover, they were reported to exhibit high pH, conductivity, TDS, calcium hardness, magnesium hardness, total alkalinity, nitrate, phosphate, chloride, and boron contents. Such phenomena coincided with reports from local operators that majority anomalous events occurred in this western-coastal-desert municipal WWTP were brackish/saline water

TABLE 1. Descriptive statistics of medoids.

Cluster	Anomaly	Medoid_1	Medoid_2	Medoid_3	Medoid_4	Medoid_5	Explanation
WWQI	53.04	56.63	56.29	54.67	48.73	57.58	Wastewater quality index
InFlow-Total	4184.17	4207.00	4716.00	4385.00	3300.00	4403.00	Inflow, total, in m <sup>3</sup> /day
InFlow-LS1	3810.23	4048.00	4471.00	4153.00	3062.00	4180.00	Inflow, municipal, in m <sup>3</sup> /day
InFlow-LS8	227.55	148.00	212.00	185.00	227.00	145.00	Inflow, from a desalination plant, in m <sup>3</sup> /day
InFlow-DP	50.01	11.00	33.00	47.00	11.00	51.00	Inflow, recycled from WWTP itself, in m <sup>3</sup> /day
Temp	29.53	32.03	30.55	28.65	30.40	31.73	Temperature, in Celsius
pH	7.91	7.31	7.48	7.36	7.28	7.36	Potential of hydrogen, unitless
Conductivity	1574.20	568.00	852.00	808.00	568.00	1577.00	Conductivity, in $\mu$ S/cm
TDS	1102.36	382.00	582.00	547.00	381.00	1104.00	Total dissolved solid, in mg/L
TSS	68.65	42.00	81.00	79.00	63.00	61.00	Total suspended solid, in mg/L
CaHardness	101.76	52.00	48.00	56.00	64.00	64.00	Calcium hardness, in mg/L
MgHardness	96.15	28.00	32.00	24.00	28.00	88.00	Magnesium hardness, in mg/L
TotalAlkalinity	187.79	92.00	108.00	136.00	120.00	128.00	Total alkalinity, in mg/L
BOD <sub>5</sub>	97.72	65.00	110.00	134.00	96.00	107.00	5-day biochemical oxygen demand, in mg/L
COD	153.39	93.00	155.00	194.00	173.00	164.00	Chemical oxygen demand, in mg/L
FOG	31.14	22.90	11.40	11.40	80.00	5.70	Fat, oils and grease, in mg/L
TKN	14.67	12.40	14.30	17.60	16.80	15.80	Total Kjeldahl nitrogen, in mg/L
NH <sub>3</sub> N	9.08	7.90	9.80	13.00	10.60	12.00	Ammonia nitrogen, in mg/L
NO <sub>3</sub> N	4.76	2.90	2.60	2.80	4.30	1.50	Nitrate nitrogen, in mg/L
PO <sub>4</sub> P	10.72	9.50	4.70	9.90	7.80	5.40	Phosphate phosphorus, in mg/L
Cl	369.88	94.00	157.00	130.00	82.00	353.00	Chloride, in mg/L
Boron	1.95	0.90	1.20	1.10	1.00	1.10	Boron, in mg/L

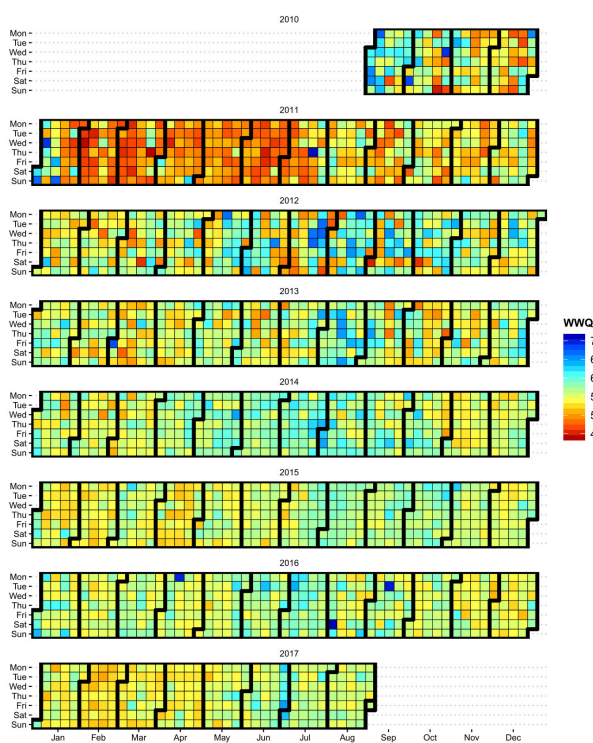


FIGURE 6. Calendar plot of temporal wastewater quality index.

intrusion, whether from groundwater surrounding lift pump station, from upstream desalination plant emergent discharge, or from rare rainfall streamflow.

To show relationships among measured variables and temporal records, double hierarchical clustered heatmap with density plot based on quarterly averages were delineated

in Figure 7. Z scores were normalized values calculated by row averages and summarized as probability density distribution via kernel density estimation, as shown in the density plot. Though quarterly averages were taken for smoothing, summary, and visualization, asymmetric unimodal distribution with positive skewness and positive kurtosis was evidently observed. This point coincided with the previous PCA plot to imply extreme non-Gaussian property of this environmental dataset. Temporal records were reordered in columns by clusters. Similar to the CLARA results, the unique year 2011 was isolated from others.

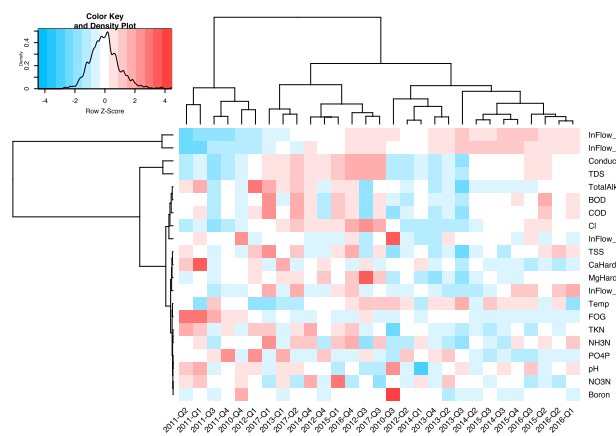


FIGURE 7. Double hierarchical clustered heatmap with density plot.

Measured variables were reordered in rows by clusters. Similar palette composition in rows signified positive correlations, while the reverse for negative ones. The inflow from lift station one clearly contributed the majority of total

inflow. The inflow from lift station eight (connected to a desalination plant) was shown negative to chloride, chemical oxygen demand (COD), and biochemical oxygen demand (BOD), and alkalinity, but its trend was similar to boron contents. This showed utterly different compositions of industrial discharge and municipal wastewater in our case. The inflow from WWTP inside ('Inflow DP') was shown closer to hardness and suspended solids (TSS), probably due to chemicals dosed for membrane cleaning and sludge processing events. The temperature was displayed moving negatively to almost all water quality variables but positively to major municipal water quantity variables, which confirmed previous statements on seasonal wastewater qualities and flow quantities. The overall agglomerative phenomenon demonstrated the non-linear, non-Gaussian, non-stationary, auto-correlated, cross-correlated, and heteroskedastic characteristics. Those empirical historical observations would challenge traditional regression techniques that were typically requiring linear, independent (no autocorrelation or multicollinearity), homoskedastic, and Gaussian distributions. Therefore, versatile preprocessing treatment of environmental data such as kernel techniques could be promising.

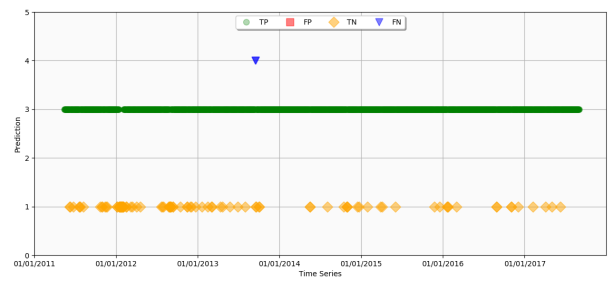
**B. DETECTION PERFORMANCE OF THE PROPOSED KPCA-OCSVM APPROACH**

To assess quantitatively the detection efficiency of the proposed procedures, the following metrics were employed: true-positive rate (TPR, or recall), false-positive rate (FPR), area under the receiver operating characteristic curve (AUC), accuracy, precision, and F1Score. A confusion matrix with equations for frequently visited metrics in assessing classification algorithm performances was summarized in Figure 8. Detection results were delineated in Figure 9 in time series. Derived performance metrics were computed and listed in Table 2. Previous research on PCA-based methods (Table 3) were revisited as comparison [11]. Since previously we defined normal/abnormal as +1/ - 1, consistently here they were taken as positive/negative classes.

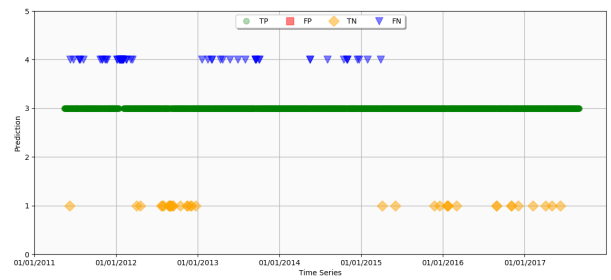
		Real Class		
		Positive	Negative	
Predicted Class	Positive	True Positive (TP) Detected normal	False Positive (FP) False-alarm	$TPR = Recall = TP/(TP+FN)$ $FPR = FP/(FP+TN)$ $Accuracy = (TP+TN)/(TP+TN+FP+FN)$ $Precision = TP/(TP+FP)$ $F1Score = 2TP/(2TP+FP+FN)$
	Negative	False Negative (FN) Missed-detection	True Negative (TN) Detected abnormal	

**FIGURE 8. Confusion matrix and associated performance metrics in this study.**

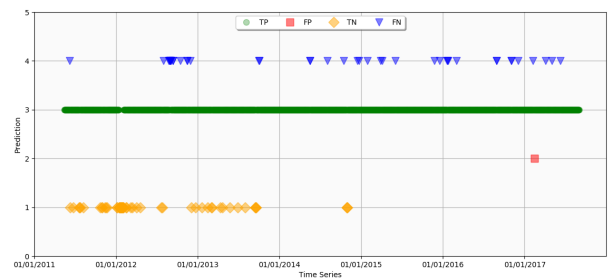
For series in Figure 9, true positives were in green circles, representing normal ICs that were classified as normal, trivial cases. False negatives were in blue down triangles, representing normal ICs that were classified as abnormal, raising false-alarms, and would cause unnecessary operating costs. True negatives were in yellow diamonds, representing successfully captured abnormalities. False positives were in red squares,



(a)



(b)



(c)

**FIGURE 9. Anomaly detection results of KPCA-OCSVM algorithms. KPCA kernels were selected as (a) polynomial, with 5 PCs; (b) cosine, with 5 PCs; (c) RBF, with 40 PCs.**

representing missed detection, were most dangerous and may lead to system failures. Therefore, from model evaluation aspect, we were expecting no red squares, less blue triangles, or as numerically in Table 2, minimized FPR, higher TPR, AUC, accuracy, precision, and F1Score.

In Table 2, various kernels for KPCA were compared with increasing numbers of PCs. With more PCs, the anomaly-free training dataset would be learned and approximated more precisely. However, excessive PCs would over-express the original dataset and lead to ill-structured classification problem for later OCSVM processing. Thereby with rising counts of PCs, generally both TPR and FPR would rise, while precision would fall. AUC, as a weighted combination of TPR and FPR, together with F1Score as the harmonic mean of TPR and precision, showed optimal maximum values when tuning. With polynomial KPCA, perfect detections were displayed with around 10 PCs, showing stronger abilities to learn the training dataset.



**TABLE 2.** Anomaly detection performance of KPCA-OCSVM algorithms with various kernel functions and different numbers of selected PCs in the KPCA.

Kernel	num.PCs	TPR	FPR	AUC	Accuracy	Precision	F1Score
Poly	5	0.9991	0.0000	0.9995	0.9991	1.0000	0.9995
Poly	10	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000
Poly	15	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000
Poly	20	1.0000	0.0097	0.9951	0.9996	0.9995	0.9998
Poly	25	1.0000	0.0467	0.9766	0.9978	0.9977	0.9989
Poly	30	1.0000	0.0727	0.9636	0.9965	0.9964	0.9982
Poly	40	1.0000	0.1570	0.9215	0.9917	0.9914	0.9957
Cosine	5	0.9717	0.0000	0.9859	0.9722	1.0000	0.9857
Cosine	10	0.9852	0.0000	0.9926	0.9857	1.0000	0.9926
Cosine	15	0.9955	0.0108	0.9924	0.9952	0.9995	0.9975
Cosine	20	0.9982	0.0200	0.9891	0.9974	0.9991	0.9986
Cosine	25	0.9982	0.0200	0.9891	0.9974	0.9991	0.9986
Cosine	30	0.9982	0.0200	0.9891	0.9974	0.9991	0.9986
Cosine	40	0.9982	0.0200	0.9891	0.9974	0.9991	0.9986
RBF	5	0.9565	0	0.9783	0.9566	1	0.9778
RBF	10	0.9826	0.0308	0.9759	0.9822	0.9991	0.9908
RBF	15	1	0.1639	0.9180	0.9913	0.9909	0.9954
RBF	20	0.9594	0.0000	0.9797	0.9596	1.0000	0.9793
RBF	25	0.9636	0.0000	0.9818	0.9639	1.0000	0.9815
RBF	30	0.9696	0.0000	0.9848	0.9700	1.0000	0.9846
RBF	40	0.9883	0.0130	0.9877	0.9883	0.9995	0.9939

**TABLE 3.** Anomaly detection performance of PCA-based algorithms, from prior research. PCA reconstructions formed statistics including univariate residuals, squared prediction error (SPE), T2, and K-nearest neighbor distances (Euclidean or Manhattan), for which parametric/nonparametric thresholds were set to detect anomalies.

Methods	Thresholds	TPR	FPR	AUC	Accuracy	Precision	F1Score
PCA-KNN	Eucl-np	0.9277	0.1275	0.9001	0.9253	0.9937	0.9596
PCA-KNN	Eucl-p	0.7627	0.0490	0.8569	0.7711	0.9970	0.8643
PCA-KNN	Manh-np	0.9332	0.1176	0.9078	0.9309	0.9942	0.9627
PCA-KNN	Manh-p	0.7650	0.0294	0.8678	0.7741	0.9982	0.8662
PCA-Residual	np	0.4673	0.2353	0.6160	0.4805	0.9772	0.6322
PCA-SPE	np	0.9432	0.1471	0.8981	0.9392	0.9928	0.9674
PCA-SPE	p	0.9991	0.8039	0.5976	0.9635	0.9640	0.9812
PCA-T2	np	0.9950	0.5980	0.6985	0.9687	0.9729	0.9838
PCA-T2	p	0.9409	0.3431	0.7989	0.9283	0.9834	0.9617

Cosine KPCA showed comparable performance with 15 PCs. RBF KPAs were similarly poor with fewer PCs but functioned better with more, which would involve intensive computation or potential overfitting and thus not suitable for online process monitoring or knowledge sharing among different WWTP ICs.

By comparison between Table 2 and Table 3, we could conclude that the nonlinear attribute of the dataset was generally better learned by the proposed kernel approaches than directly applying traditional PCA-based methods. Though KNN was applied as a local lazy learning algorithm in PCA-KNN to tackle nonlinear behavior and outperformed other counterparts, outputs from PCA were not sufficiently revealing data nature. Probably the data nature was nonlinear to some extent and could be explained better by polynomial KPCA and summarized by RBF-based OCSVM than linear PCA, but

not “too” nonlinear that undermined outcomes would arrive when applying RBF in KPCA.

With proper training adopting local data, the proposed data-driven KPCA-OCSVM models could sufficiently learn and approximate nonlinear, non-Gaussian, non-stationary, auto-correlated, cross-correlated, and heteroskedastic features from anomaly-free local WWTP ICs data by KPCA, then effectively identify real anomalies from the test dataset with competitive performances by OCSVM, in contrast to previous linear PCA-KNN algorithms. Kernel techniques successfully improved model outcomes with limited increase in computation cost or model complexity. Without assuming the underlying data structure, the investigated flexible environmental data science approach could be transferred, rebuilt, and tuned for ICs from different WWTPs.

## V. CONCLUSION

To operate wastewater treatment plants with optimized efficiency, influent conditions as initial states of inflow fed to WWTPs were monitored to identify potential anomalies that would trigger adverse events or system crash. To involve voluminous ICs data for data-driven decisions, the non-linear, non-Gaussian, non-stationary, auto-correlated, cross-correlated, and heteroskedastic nature of environmental dataset must be considered. This research introduced kernel machine learning models, KPCA-OCSVM with various kernels, to learn anomaly-free training set then classify the testing set of ICs. Exploratory analysis with data visualization was performed to reveal temporal behaviors and statistical properties of multivariate IC time series. KPCA sufficiently output representative features, based on which OCSVM sensitively and specifically identified anomalies in ICs that were previously omitted by operators. The proposed kernel algorithms surpassed previous linear PCA-based KNN models, and improved outcomes with limited increase in computation cost. Without requiring linear, Gaussian, stationary, independent, and homoskedastic qualities from data, the proposed flexible environmental data science approach could be transferred, rebuilt, and tuned conveniently for ICs from different WWTPs. Future research would further investigate data from full-scale wastewater treatment process and apply kernel machine learning techniques for effective whole process monitoring.

## REFERENCES

- [1] L. Corominas, M. Garrido-Baserba, K. Villez, G. Olsson, U. Cortés, and M. Poch, "Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques," *Environ. Model. Softw.*, vol. 106, pp. 89–103, Aug. 2018.
- [2] H. Haimi, M. Mulas, F. Corona, and R. Vahala, "Data-derived soft-sensors for biological wastewater treatment plants: An overview," *Environ. Model. Softw.*, vol. 47, pp. 88–107, Sep. 2013.
- [3] J.-J. Zhu, L. Kang, and P. R. Anderson, "Predicting influent biochemical oxygen demand: Balancing energy demand and risk management," *Water Res.*, vol. 128, pp. 304–313, Jan. 2018.
- [4] K. Gibert, J. S. Horsburgh, I. N. Athanasiadis, and G. Holmes, "Environmental data science," *Environ. Model. Softw.*, vol. 106, pp. 4–12, Aug. 2018.
- [5] T. Cheng, C.-H. Wei, and T. Leiknes, "Polishing of anaerobic secondary effluent by *Chlorella vulgaris* under low light intensity," *Bioresour. Technol.*, vol. 241, pp. 360–368, 2017.
- [6] K. B. Newhart, R. W. Holloway, A. S. Hering, and T. Y. Cath, "Data-driven performance analyses of wastewater treatment plants: A review," *Water Res.*, vol. 157, pp. 498–513, Jun. 2019.
- [7] K. Gibert, J. Izquierdo, M. Sánchez-Marré, S. H. Hamilton, and I. Rodríguez-Roda, and G. Holmes, "Which method to use? An assessment of data mining methods in Environmental Data Science," *Environ. Model. Softw.*, vol. 110, pp. 3–27, Dec. 2018.
- [8] H.-G. Han, L. Zhang, and J.-F. Qiao, "Data-based predictive control for wastewater treatment process," *IEEE Access*, vol. 6, pp. 1498–1512, 2017.
- [9] G. Han, Q. Cheng, X. Sun, L. Li, and W. Di, "A biological mechanism based structure self-adaptive algorithm for feedforward neural network and its engineering applications," *IEEE Access*, vol. 7, pp. 25111–25122, 2019.
- [10] J.-M. Lee, C. K. Yoo, and I.-B. Lee, "Statistical process monitoring with independent component analysis," *J. Process Control*, vol. 14, no. 5, pp. 467–485, Aug. 2004.
- [11] T. Cheng, F. Harrou, Y. Sun, and T. O. Leiknes, "Monitoring influent measurements at water resource recovery facility using data-driven soft sensor approach," *IEEE Sensors J.*, vol. 19, no. 1, pp. 342–352, Jan. 2019.
- [12] J. Ni, C. Zhang, L. Ren, and S. X. Yang, "Abrupt event monitoring for water environment system based on KPCA and SVM," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 4, pp. 980–989, Apr. 2012.
- [13] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, Jan. 2002.
- [14] H. Ye, F. Luo, and Y. Xu, "Dissolved oxygen generic model control of wastewater treatment process based on immune optimization least squares support vector machine," in *Proc. IEEE 8th World Congr. Intell. Control Automat.*, Jul. 2010, pp. 5082–5086.
- [15] A. Perera, N. Papamichail, N. Barsan, U. Weimar, and S. Marco, "On-line novelty detection by recursive dynamic principal component analysis and gas sensor arrays under drift conditions," *IEEE Sensors J.*, vol. 6, no. 3, pp. 770–783, Jun. 2006.
- [16] S. W. Choi, C. K. Lee, J.-M. Lee, J. H. Park, and I.-B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics Intell. Lab. Syst.*, vol. 75, no. 1, pp. 55–67, Jan. 2005.
- [17] B. Schölkopf, A. Smola, and K. Müller, "Kernel principal component analysis," in *Artificial Neural Networks—ICANN*. Berlin, Germany: Springer, 1997, pp. 583–588.
- [18] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [19] J. Honaker, G. King, and M. Blackwell, "Amelia II: A program for missing data," *J. Statist. Softw.*, vol. 45, no. 7, pp. 1–47, 2011.
- [20] *R: A Language Environment for Statistical Computing*, R Found. Stat. Comput., Vienna, Austria, 2019. [Online]. Available: <https://www.R-project.org/>
- [21] *RStudio: Integrated Development Environment for R*, RStudio, Inc., Boston, MA, USA, 2015. [Online]. Available: <http://www.rstudio.com/>
- [22] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "Kernlab—An S4 package for kernel methods in R," *J. Stat. Softw.*, vol. 11, no. 9, pp. 1–20, 2004.
- [23] M. Mächler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, and K. Hornik, "Cluster: Cluster analysis basics and extensions," *R Package Version*, vol. 1, no. 2, p. 56, 2012.
- [24] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, vol. 1. Scotts Valley, CA, USA: CreateSpace, 2017.
- [25] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. New York, NY, USA: Springer-Verlag, 2016.
- [26] G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, and B. Venables. (2019). *gplots: Various R Programming Tools for Plotting Data. R Package Version 3.0.1.1*. [Online]. Available: <https://CRAN.R-project.org/package=gplots>
- [27] M. Ebrahimi, E. L. Gerber, and T. D. Rockaway, "Temporal performance assessment of wastewater treatment plants by using multivariate statistical analysis," *J. Environ. Manage.*, vol. 193, pp. 234–246, May 2017.



**TUOYUAN CHENG** received the bachelor's degree in water and wastewater engineering from the Harbin Institute of Technology. He is currently pursuing the Ph.D. degree in environmental science and engineering with the King Abdullah University of Science and Technology. His current research interests include the application of anomaly detection, data-driven prediction and optimization in water/wastewater treatment plant, and water distribution systems.



**ABDELKADER DAIRI** received the engineer's degree from the University of Oran 1 Ahmed Ben Bella, Algeria, in 2003, the Magister degree from the National Polytechnic School of Oran, Algeria, in 2006, and the Ph.D. degree from Ben Bella Oran 1 University, in 2018, all in computer science. From 2007 to 2013, he was a Senior Oracle Database Administrator (DBA) and Enterprise Resource Planning (ERP) Manager. He has over 20 years of programming experience in different

languages and environments. He is currently an Assistant Professor with the Signal, Image, and Speech Laboratory (SIMPA), Computer Science Department, University of Science and Technology of Oran-Mohamed Boudiaf (USTO-MB), Oran, Algeria. His research interests include deep learning approach for autonomous robot navigation, computer vision, image processing, and mobile robotics.



**FOUZI HARROU** received the M.Sc. degree in telecommunications and networking from the University of Paris VI, France, and the Ph.D. degree in systems optimization and security from the University of Technology of Troyes (UTT), France. He was an Assistant Professor with UTT for one year and with the Institute of Automotive and Transport Engineering, Nevers, France, for one year. He was also a Postdoctoral Research Associate with the Systems Modeling and Dependability Laboratory, UTT, for one year. He was a Research Scientist with the Chemical Engineering Department, Texas A&M University at Qatar, Doha, Qatar, for three years. Since 2015, he has been a Postdoctoral Fellow with the Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology. He is the author of more than 100 refereed journals and conference publications and book chapters. His current research interests include statistical decision theory and its applications, fault detection and diagnosis, and deep learning.



**YING SUN** received the Ph.D. degree in statistics from Texas A&M, in 2011. She held a two-year postdoctoral research position at the Statistical and Applied Mathematical Sciences Institute and the University of Chicago. She was an Assistant Professor with Ohio State University for a year before joining KAUST, in 2014. At KAUST, she established and leads the Environmental Statistics research group, which works on developing statistical models and methods for complex data to

address important environmental problems. She has made original contributions to environmental statistics, in particular in the areas of spatiotemporal statistics, functional data analysis, visualization, computational statistics, with an exceptionally broad array of applications. She received two prestigious awards: The Early Investigator Award in Environmental Statistics presented by the American Statistical Association and the Abdel El-Shaarawi Young Research Award from The International Environmetrics Society.



**TOROVE LEIKNES** is currently a Professor with the Environmental Engineering, Biological and Environmental Science and Engineering (BESE), Water Desalination and Reuse Center (WDRC), King Abdullah University of Science and Technology (KAUST), Saudi Arabia. He joined KAUST, in 2013, and served as the Director of WDRC, from 2014 to 2017. Prior to KAUST, he was a Professor of environmental engineering with the Department of Hydraulic and Environmental Engineering, Faculty Engineering Science and Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. His main research interest includes the application of membrane technology in environmental engineering and has more than 20 years of working experience in this area. Research topics include coupling membranes with physical/chemical and biological processes for advanced water and wastewater treatment, solid-liquid separation technologies, hybrid and integrated solutions for energy efficient water treatment, and development of advanced water treatment systems for sustainable strategies targeting water reuse. At NTNU, he was involved in various European Union (EU) projects and a Coordinator of the EUROMBRA project, targeting the development of membrane bioreactor (MBR) technology for municipal wastewater treatment. He has been involved in several sponsored projects on integrating membrane technology in drinking water treatment for removal of NOM and as a hygienic barrier, wastewater treatment on ships, treatment of produced water (PW) in the oil and gas industry, and advanced water treatment in recycling aquaculture systems (RAS). At KAUST, his main research activities are focused on the development of novel membrane technologies for the low-energy treatment of waters with impaired water quality for various reuse purposes, using wastewater as a resource for high-value product formation, new concepts of water treatment in RAS, and membrane filtration of PW for alternative reuse strategies. Developing integrated technologies for water reclamation and use in urban agriculture is a new research area of interest. A key component is fundamental studies of biofouling in membrane systems by applying in situ nondestructive techniques.

neering, Faculty Engineering Science and Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. His main research interest includes the application of membrane technology in environmental engineering and has more than 20 years of working experience in this area. Research topics include coupling membranes with physical/chemical and biological processes for advanced water and wastewater treatment, solid-liquid separation technologies, hybrid and integrated solutions for energy efficient water treatment, and development of advanced water treatment systems for sustainable strategies targeting water reuse. At NTNU, he was involved in various European Union (EU) projects and a Coordinator of the EUROMBRA project, targeting the development of membrane bioreactor (MBR) technology for municipal wastewater treatment. He has been involved in several sponsored projects on integrating membrane technology in drinking water treatment for removal of NOM and as a hygienic barrier, wastewater treatment on ships, treatment of produced water (PW) in the oil and gas industry, and advanced water treatment in recycling aquaculture systems (RAS). At KAUST, his main research activities are focused on the development of novel membrane technologies for the low-energy treatment of waters with impaired water quality for various reuse purposes, using wastewater as a resource for high-value product formation, new concepts of water treatment in RAS, and membrane filtration of PW for alternative reuse strategies. Developing integrated technologies for water reclamation and use in urban agriculture is a new research area of interest. A key component is fundamental studies of biofouling in membrane systems by applying in situ nondestructive techniques.

...