ℓ_1 Trend Filtering-Based Change Point Detection for Pumping Line Balance of Deposition Equipment

Jeongsun Ahn, Duyeon Kim, Mingi Song, Jaehong Min, Jimin Hwang, Juhye Kwon, and Hyun-Jung Kim*

***Abstract*—One of the most significant manufacturing issues is how to monitor and diagnose the state of machines from various sensor data. Detecting machine state changes is very important because it can prevent machine breakdown or product quality deterioration. This work deals with a change point detection problem of semiconductor manufacturing equipment, especially for a deposition process, with real data. In a deposition machine, the pressures of two pumping lines, each of which absorbs a different gas (i.e., Zr and ozone), should be well-balanced so that the two gases are not mixed in a line. Otherwise, particles can be accumulated on the surface of wafers, which significantly affects the product quality. When one gas is pumped into another gas line, the balance of pumping lines is considered to be broken. We propose a ℓ_1 trend filtering-based change point detection method to identify such a pumping balance break in the deposition machine. This method is suitable for data having increasing or decreasing trends, such as pressure and temperature, and can be applied for the real-time detection. The proposed method shows an adequate true detection rate while effectively reducing the number of false alarms compared to other methods. We further develop a prediction model for estimating the pressure in the machine to improve the performance of the change point detection algorithm.**

***Index Terms*—Change point detection, ℓ_1 trend filtering, pumping balance, deposition equipment, multi-step ahead prediction model**

I. Introduction

IT is very important to detect machine state changes as early as possible to prevent machine breakdown and product quality degradation. Therefore, many manufacturing companies try to monitor real-time sensor data obtained from machines, analyze the machine state based on the data, and detect abnormal states. Since machine degradation can be recognized by analyzing data which have a different form from the normal state, many studies have been conducted on detecting such characteristics. Some of them use a control chart to monitor the real-time sensor data and classify machine states [1], [2], and some other studies detect abnormal states by forecasting the machine states with a prediction model [3], [4], [5].

There have also been many studies that use change point detection (CPD) methods to identify a change point in time series data [6], [7]. Most of the previous CPD studies have mainly focused on identifying the mean or variance changes in the given data without a trend [8], [9], [10], [11]. This study proposes a new CPD method using ℓ_1 trend filtering [12], which is suitable for data having an increasing or decreasing trend, to detect machine abnormal states in real time. The proposed algorithm is applied to detect pressure changes and identify an abnormal state in deposition machines for semiconductor manufacturing. This problem has been required from one of the semiconductor manufacturing equipment companies in Korea.

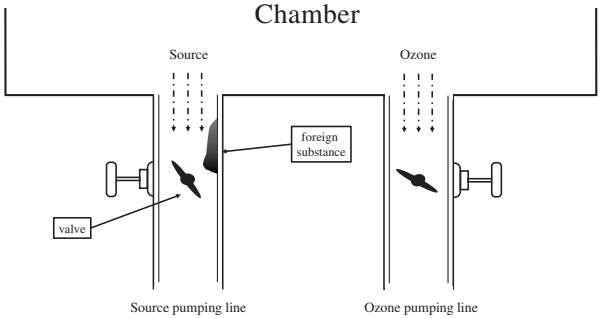


Fig. 1. An ALD machine example with source and ozone pumping lines. Foreign substances are being accumulated in the source pumping line.

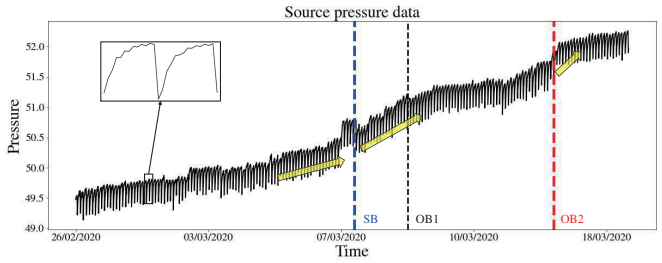
In an atomic layer deposition (ALD) process, a film is grown on a substrate by exposing its surface to alternate gaseous species (i.e., Zr or Hf), called 'source', and organic matter of the film is removed with the ozone gas. The source and ozone are not provided simultaneously and react with the surface of a substrate one at a time in a sequential manner (i.e., source $\rightarrow$ ozone $\rightarrow$ source $\rightarrow \cdots$). There are also many other types of ALD, and further details on ALD can be found in [13].

In the ALD machine, there are two pumping lines which absorb the source and ozone, respectively, as illustrated in Fig. 1. It is critical to pump the two gases out separately while keeping the pressure within the chamber in order to avoid particle generation on the surface of a wafer. The amount of the source and ozone absorbed to each line depends on the valve angle. The source line can adjust the valve angle independently, but the ozone line changes the angle according to the ratio of the source valve angle. As the number of wafers processed increases, foreign substances are accumulated on the source line, which decreases its absorption capacity, and the valve angle of the source line increases accordingly to maintain the pressure in the chamber. Then, the source pressure (SP) in the pumping line is increased, and the foreign substances

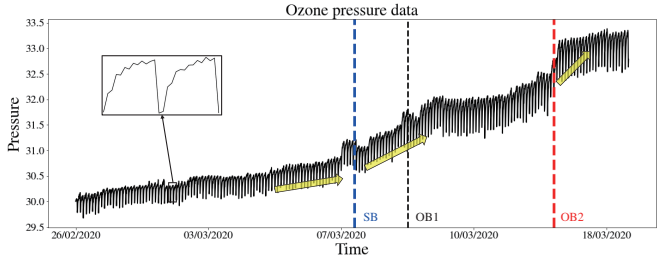
This work was supported by Wonik IPS. (*Corresponding author: Hyun-Jung Kim)

Jeongsun Ahn, Duyeon Kim and Hyun-Jung Kim are with the Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea. hyunjungkim@kaist.ac.kr

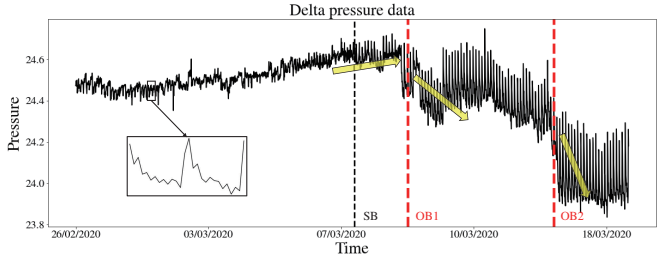
Minki Song, Jaehong Min, Jimin Hwang, and Juhye Kwon are with the Wonik IPS, Pyeongtaek 17840, South Korea.



(a) Source pressure



(b) Ozone pressure



(c) Delta pressure

Fig. 2. Graphs for source, ozone, and delta pressures. The blue and red dotted lines represent the time at which the SB and OB occur, respectively, and significant slope changes can be observed.

are more accumulated. Then, source blockage (SB) can occur, especially when the pressure rises dramatically than before. After the SB, if the substances continue to pile up, the source starts to flow over to the ozone line from a certain point, which increases the ozone pressure (OP) in the line and causes ozone blockage (OB). When the SB or OB occurs, the pumping line balance is considered broken. Therefore, these two abnormal states of the deposition machine, the SB and OB, should be detected in advance by monitoring the SP and OP. In addition, the difference between the SP and OP, which is called the delta pressure (DP(=SP-OP)), should be monitored together. The slope of the SP is similar or steeper than that of the OP in general. If the DP is not monitored, the difference in increase rates between the SP and OP cannot be observed, i.e., it cannot be identified whether the SP rises steeper than the OP or vice versa. The DP decreases when the slope change in the OP is greater than the SP, which may imply the SB or OB. In addition, only monitoring the DP without the SP or OP, can also lead to false alarms because it is hard to detect the change points when both the SP and OP change at the same time. Therefore, the SP, OP, and DP should be monitored together. We note that there can be multiple SB and OB points.

Fig. 2 shows the SP, OP, and DP from Feb. 26, 2020 to Mar. 18, 2020, and the detailed plot of two batches, each of which consists of six wafers, is enlarged in each graph. Note that the data we show in this paper have been shifted slightly from the real data of the semiconductor manufacturing equipment company due to a confidential issue, but the overall trend is maintained. It is observed that there is an increasing trend in the SP and OP. The DP gradually increases and then decreases from a certain point. In the figure, there are one SB point and two OB points (OB1 and OB2), which are indicated with the dotted lines. The blue and red dotted lines are presented when the SB and OB points, respectively, are actually detected by analyzing the slope changes of the corresponding pressure.

In Fig. 2 (a), the slope of the SP becomes steep at the SB point. In Fig. 2 (b), it is also observed that the slope change of the OP at the SB point becomes large because of the dependency on the valve angle. The OB1 point can only be detected from the DP because it starts to decrease from the point, but there is no significant slope change in the SP or OP. The slopes of the SP and OP become again steep at the OB2 point. We can also see that the DP decreases significantly at the OB2 point in Fig. 2 (c). Note that there can be many slope changes observed in the SP, OP, and DP, but not all of them indicate the SB or OB. Therefore, it is important to set a proper threshold in the slope changes to derive potential SB and OB points for further investigation and to analyze the three pressures, SP, OP, and DP, comprehensively to detect the actual SB and OB.

The SB and OB can cause particles on the surface of a wafer and frequent valve replacement and increase the machine maintenance time. Once the SB or OB is identified, either the valve angle of the source or ozone can be adjusted or machine maintenance can be performed. The inaccurate and frequent alarms generated when detecting the SB or OB may require unnecessary maintenance operations, which degrades the productivity of the machine and increases cost. Therefore, a detecting method, which performs well in identifying an overall trend and is not sensitive to the variability of the data, is required. We address this pumping balance problem of a deposition process by proposing a ℓ_1 trend filtering-based CPD (ℓ_1CPD) method. ℓ_1 trend filtering is one of the popular trend estimation methods to find embedded trends in time series data. It is possible to adjust the degree of approximation without being sensitive to real-time data with ℓ_1 trend filtering. The method detects change points from the SP, OP, and DP, separately, and then determines the SB and OB by applying a classification rule with all of the change points. We also design the multi-step ahead prediction (MSP) model to predict the SP and OP for the next batch with the dual-attention recurrent neural network (DA-RNN) [14]. We apply the prediction model to the ℓ_1CPD method to advance the detection time of change points. The contributions of this paper are as follows.

- This is the first paper that handles the real pumping line balance problem of the deposition machine.
- We develop an effective CPD algorithm with ℓ_1 trend filtering for the first time.
- The proposed CPD algorithm detects true change points relatively well with a small number of false alarms.

- It is also shown that the proposed prediction model improves the performance of ℓ_1CPD.

This paper is organized as follows. Section II reviews related literature on CPD algorithms and MSP models. In Section III, we explain the data set and feature engineering, and Section IV describes the proposed method in detail. Experiment results are presented in Section V, and Section VI summarizes our work and discusses the future work.

II. Related Work

We review the previous studies on CPD algorithms and prediction models.

A. Change point detection

CPD can be classified into parametric and non-parametric approaches. Parametric methods assume a certain data distribution, whereas there is no assumption on the data with the non-parametric approaches [15].

The representative parametric methods include binary segmentation (BS), wild binary segmentation (WBS), and Segmented. BS [8] calculates a cumulative sum (CUSUM) of two intervals divided from each data point, and if the sum is larger than a given threshold, the data set is segmented on the point. WBS [9] is an improved version of BS by randomly selecting some intervals instead of computing the CUSUM of the entire data. Segmented [16] provides a piece-wise linear function by searching for coefficients of an approximation function representing data and then updating change point positions by using the coefficients. Although they are powerful CPD algorithms, they require information on the number of change points that are unknown in a practical case. In addition, it is not appropriate to use them if the data are not normally distributed. They can also be used only after all data gather, which means that they may not work well in real-time detection.

The non-parametric CPD can be conducted with a Lasso-based method, SeqTest, and narrowest-over-threshold (NOT). The Lasso-based method [10] searches for a candidate set of change points with the given maximum number of points by penalizing weight terms and finds an optimal point by computing the costs (squared difference from the mean). SeqTest [11] determines change points sequentially through test statistics that analyze the difference between existing data and incoming data. These two methods are especially useful when data present a piece-wise constant pattern. NOT [17] is a method which defines multiple contrast functions to detect change points in data with various features ranging from a piece-wise constant to piece-wise linear pattern. However, it may not be robust to be applied in real time because the smoothness cannot be adjusted, which may provide the excessive number of change points [18]. There are also other methods for CPD [15], [19].

ℓ_1 trend filtering identifies trend estimates by smoothing data points with time series data. ℓ_1 trend filtering can handle the degree of smoothness, which makes it possible to approximate general trends without being sensitive to the variability in the data collected in real time. Therefore, ℓ_1 trend filtering, which responds to overall trend changes rather than locally changing trends, effectively reduces the number of false alarms. In addition, it is suitable to use it in real time because ℓ_1 trend filtering does not require prior information on data, such as the number of change points. There have been many studies on using ℓ_1 trend filtering for approximating trends [20], [21], [22], [23], [24]. The SP and OP data we have from the deposition equipment have an increasing trend, and the SB and OB can be detected by analyzing the amount of the slope changes based on the SP, OP, and DP. Therefore, we adopt ℓ_1 trend filtering to consider the robust trend in the pressure data. The proposed method is compared to WBS, SeqTest, NOT, and Segmented in the experiment.

B. Multi-step ahead prediction model

Time series forecasting can be divided into one-step ahead prediction and MSP depending on the number of time steps to be predicted. When the observation recorded at time t is denoted as $x(t)$ and the value we want to obtain is $y(t)$, MSP provides $[y(t+1), \ldots, y(t+H)]$ using $[x(1), ..., x(t)]$ where H is the number of time steps to be predicted. MSP is of course useful but more difficult than one-step ahead prediction due to the accumulated errors, reduced accuracy, and increased uncertainty [25].

There have been some papers on MSP [26], [27], [28], [29], [30], [31], [32], [33]. Tran *et al.* [26] derived H different models to predict each of H time steps ahead for methane compressor conditions using the CART and neuro-fuzzy system. Niu and Yang [27] also predicted the condition of a methanol compressor by deriving a recursive model that takes the predicted values from the previous step as inputs to the next step with the Dempster–Shafer regression. Cheng *et al.* [28] compared the two modeling approaches, H independent models and the recursive one, with multiple linear regression (MLR), RNN, and hidden Markov model (HMM), and showed that there is a trade-off between error accumulation and complexity. Bao *et al.* [29] derived a multi-output model with support vector regression (SVR) for NN3 time series forecasting competition data.

MSP models can be built by using various methods from regression, such as MLR, and SVR, to neural networks. In particular, RNN and long short-term memory (LSTM) with iterative structures have been widely used for time series data. Yunpeng *et al.* [30] compared the performance of LSTM with the recursive structure according to data patterns. They showed that LSTM performed better than the autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN). Han *et al.* [31] and Zhou *et al.* [32] derived a multi-output LSTM to predict wind power and three indicators of air quality in Thailand. Aliabadi *et al.* [33] predicted the catalyst activity in a methanol manufacturing process by using the attention-based RNN.

We adopt the DA-RNN for predicting the SP and OP because it captures sequential features of time series data and exhibits dynamic behavior between consecutive batches. MSP is assumed for one batch prediction which takes 18 minutes on average. The proposed prediction model extends

TABLE I
STATISTICAL DOMAIN FEATURES USED FOR PREDICTION

Feature	Equation
Mean	$\frac{1}{n}\sum_{t=1}^{n} s_t$
Standard deviation	$\sqrt{\frac{\sum_{t=1}^{n}(s_t - mean(s))^2}{n-1}}$
Mean Absolute Deviation	$\frac{1}{n}\sum_{t=1}^{n} \lvert s_t^2 - mean(s)\rvert$
Kurtosis	$\frac{1}{n-1}\sum_{t=1}^{n}(s_t - mean(s)/std(s))^4 - 3$
Skewness	$\frac{1}{n-1}\sum_{t=1}^{n}(s_t - mean(s)/std(s))^3$
Interquartile range	Third quartile - first quartile
ECDF Percentile	x_{init}, x_{end}
ECDF Slope	$\frac{p_{end}-p_{init}}{x_{end}-x_{init}}$

the DA-RNN in [14] for obtaining multiple outputs and is compared to the independent DA-RNN and other methods, such as linear regression, RandomForest, Lasso, and Ridge, in the experiment.

III. Data Description and Feature Engineering

A. Description of the real data set

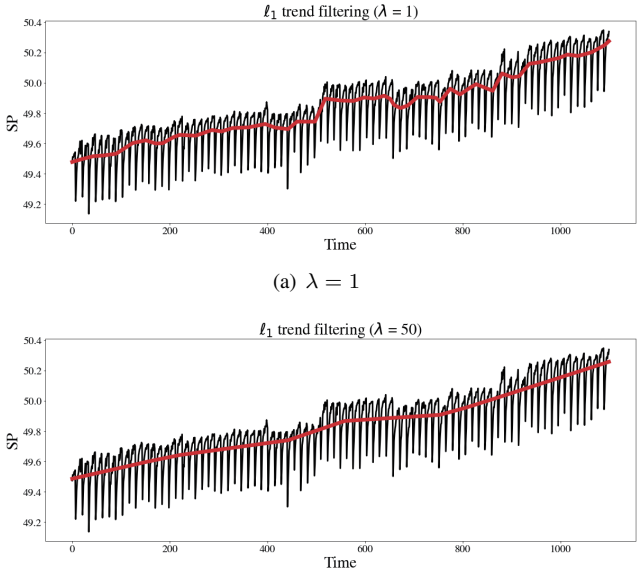
We obtained two data sets, each of which was generated from a different deposition machine. The two sets contain machine sensor data from Feb. 1 to June 19, 2020 and from Feb. 19 to June 3, 2020, respectively. During the data collection period, 1,092 and 1,039 batches were processed in each machine, and 819,000 and 778,500 data points gathered in total.

Two machine maintenance operations were performed in both sets. During the machine maintenance, the pumping lines were cleaned and the valve angles were adjusted for the pumping line balance. The machine then returned to the initial normal state. Therefore, each data set can be divided into three subsets, and six data sets in total are used in our study. There are 160 types of sensor data (features), including the SP, OP, temperature, vaporizer, rotation speed, etc., which were collected in seconds. The true SB and OB points were provided from engineers. Each data set contains at least one SB and OB, and the maximum number of SB and OB does not exceed five.

B. Feature extraction and selection

Even though there are 160 types of time series sensor data, some of them are not related to the pumping line balance. For example, the line heater temperature or stage heater temperature does not affect the SP and OP. As a result, 18 types of sensor data were selected based on the step-wise feature selection and the interviews with engineers, and they include the features related to chambers (ex. throttle valve, chamber pressure), mass flow controller, and source dry pump and ozone dry pump (ex. current, temperature, pressure, exhaust pressure). Then, data preprocessing of removing outliers and interpolating missing data was conducted. The average values of features in minutes are used to remove noise of data.

After that, feature extraction was performed to generate meaningful information for the construction of MSP from the 18 types of data. Table I shows the statistical features we used. Note that among the 17 statistical features in [34], only the eight features in Table I were selected from the preliminary test of various combinations. In Table I, s_t is the value of a feature at time t, and $mean(s)$ and $std(s)$ are the mean and standard deviation of a feature value, respectively. n is the total number of data points. p_{init} and p_{end} are the first and third quartile values, respectively. x_{init} and x_{end} are the empirical cumulative distribution values in the first and third quartile, respectively. Finally, step-wise feature selection was performed to select the best combination of the features we have. As a result, 57 and 45 features were selected for the two machines, respectively.



(a) $\lambda = 1$

(b) $\lambda = 50$

Fig. 3. ℓ_1 trend filtering with a different λ value. The trend estimates are smoother with large λ.

IV. Solution Approach

A. ℓ_1 trend filtering

ℓ_1 trend filtering which provides piece-wise linear trend estimates has been proposed in [12] while minimizing the following:

$$\frac{1}{2}\sum_{t=1}^{n}(y_t - z_t)^2 + \lambda\sum_{t=2}^{n-1}|z_{t-1} - 2z_t + z_{t+1}| \qquad (1)$$

where y_t is an output (SP or OP) at time t, and n is the number of data points. The trend estimate z_t at time t that minimizes Eq. (1) is found with a specialized interior point method [12]. The first term in Eq. (1) measures the size of the residual, $y_t - z_t$, and the second term indicates the smoothness of the estimated trend [12]. The second term becomes 0 if there are three data points (z_{t-1}, z_t, z_{t+1}) on a line. The trade-off between the residual and smoothness can be controlled through the regularization term λ ($\lambda \geq 0$). Fig. 3 shows the trend estimates colored in red according to λ with the SP. The trend estimates are smoother with λ of 50 than that of 1. Note that

Algorithm 1 Algorithm for obtaining the threshold

Input: C^p, λ, l, w
Output: th_p

for p in P **do**
  for C_t^p in C^p **do**
    for $t = C_t^p, \cdots, C_t^p + l$ **do**
      Apply ℓ_1 trend filtering to the data points from time steps 1 to t using λ.
      for $i = t - w, \cdots, t - 1$ **do**
        Compute the absolute value of the difference of trend estimates between two consecutive points $(i, i + 1)$, denoted as d_i.
      end for
      $d_t^p = \max_i(d_i)$
    end for
    $d_{C_t^p} = \max_t(d_t^p)$
  end for
  $th_p = \min_{C_t^p \in C^p}(d_{C_t^p})$
end for

ℓ_1 trend filtering is applied to each of pressure types, SP, OP, and DP in the proposed method.

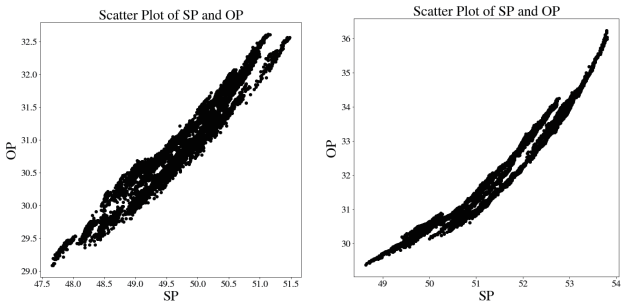
B. Threshold

When the absolute value of the difference between successive trend estimates is larger than a given threshold, it is considered to be a change point. A large threshold may not detect real change points, whereas a small threshold can provide too many change points (i.e., many false alarms). Therefore, it is important to set an appropriate threshold to find the SB and OB.

A data-driven approach for calculating the threshold is proposed in Algorithm 1. We denote p as a pressure type where $p \in P = \{SP, OP, DP\}$. C^p is used for a given set of change points in p from train data, and C_t^p indicates the change point at time t in C^p. For each C_t^p, ℓ_1 trend filtering is applied to the data from time steps 1 to t, and the slopes, from two consecutive trend estimates at time points i and $i + 1$ where $t - w \leq i \leq t - 1$ and w is a given parameter, are computed. Then the maximum absolute value among w slopes is denoted with d_t^p. ℓ_1 trend filtering is again applied to the data from time steps 1 to $t + 1$, and d_{t+1}^p, which is the maximum absolute value among w slopes between time steps $t + 1 - w$ to $t + 1$, is obtained. This procedure is repeated until time steps $t + l$ where l is another given parameter. We then obtain l d_i^p's where $t \leq i \leq t + l$, and the maximum one is denoted with $d_{C_t^p}$. The procedure is applied for all of the change points in C^p, and the minimum value among $|C^p|$ slopes is taken as th_p, the threshold for pressure p. We set l and w to 100 from the preliminary experimental result.

C. Dual-Attention RNN

A RNN is well-suited to time series data because it processes a time series step by step by taking data from the previous time steps as inputs to the current time step [35]. To predict the SP and OP, we build a RNN-based MSP model. The



(a) Process machine 1 (b) Process machine 2

Fig. 4. Scatter plots of the SP and OP in each machine. A high correlation is observed.

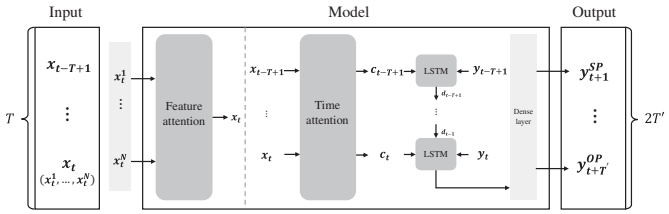


Fig. 5. Structure of MODA-RNN with feature and time attentions to predict the SP and OP for the next 18 minutes.

model provides the SP and OP for the next T' (i.e., 18 minutes) time steps based on the input features and the output values of the last T time steps (called a look-back size). We want to obtain $\hat{y}_{t+1}$ at time t, $\hat{y}_{t+1} = (y_{t+1}, y_{t+2}, \cdots, y_{t+T'}) \in \mathbb{R}^{T'}$ where y_t indicates the SP (y_t^{SP}) and OP (y_t^{OP}) at time t. The following model is used:

$$\hat{y}_{t+1} = F(y_{t-T+1}, \cdots, y_t, x_{t-T+1}, \cdots, x_t). \quad (2)$$

where $x_t = (x_t^1, \cdots, x_t^N)$ is a set of input variables at time t, N is the number of input features, and $F(\cdot)$ is a nonlinear mapping function.

The DA-RNN, which focuses on separate input features first and then on different time instances, was introduced in [14]. The DA-RNN solved the performance degradation issue occurring when the sequence length considered becomes long by adding attention mechanism to RNN. The DA-RNN consists of input attention for feature representation and temporal attention for time indication. This architecture is very helpful for taking key information from time series data. For more detailed explanation, refer to [14].

Fig. 4 shows the scatter plots of the SP and OP in each machine. It can be seen that the SP and OP have a very high correlation. Therefore, the DA-RNN we design has 36 outputs from $t + 1$ to $t + 18$ of the SP and OP, so that it can learn their relationship. The method is called multi-output DA-RNN (MODA-RNN). Fig. 5 shows the overall structure of MODA-RNN.

D. Overall Method

The overall procedure of the ℓ_1CPD is presented in Fig. 6. A threshold is determined based on train data from Algorithm

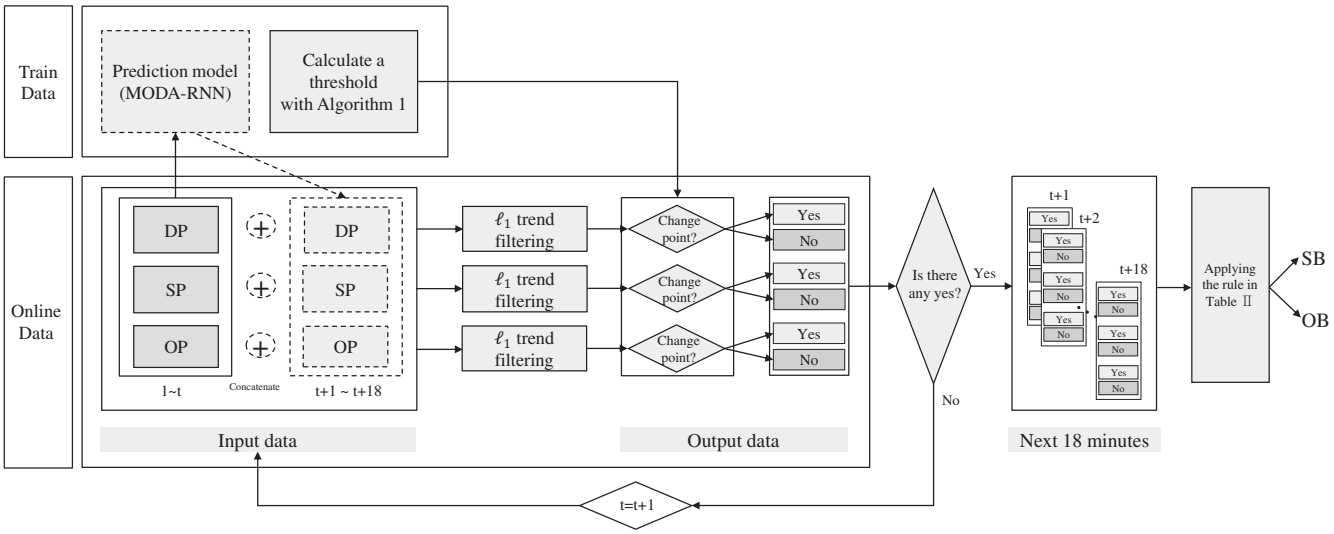


Fig. 6. Overall procedure of (P-) ℓ_1CPD. l_1 trend filtering is applied to the predicted DP, SP and OP from MODA-RNN, and SB or OB is detected by analyzing the change points.

1. ℓ_1 trend filtering is applied to calculate the trend estimates from time steps 1 to t for each of pressure types, SP, OP, and DP, at current time t. It always calculates trend estimates from time step 1 instead of using a fixed sliding window size to observe the overall trend of the data. If the amount of changes, the absolute value of the difference between two consecutive trend estimates, is larger than the given threshold, it is indicated with a change point. If the change point is detected on a certain pressure type, the method further checks whether change points in other pressure types occur within the next 18 minutes to include the process of the next one batch. Then an alarm is generated by following the rule in Table II which classifies the change points into the SB and OB. When there is a change point only in the SP, then an alarm labeled with the SB is generated, which corresponds to case 1 in Table II. When the change points in both the SP and OP are detected as in case 3, an alarm labeled with the SB is first provided and then the OB is considered to occur from the next time. Note that this rule has been derived by analyzing the given data and interviews from the engineers.

The proposed ℓ_1CPD can also be applied with the estimated future pressure from the MODA-RNN model. This step is represented by a dotted line in Fig. 6, and it is called P-ℓ_1CPD.

TABLE II
CLASSIFICATION RULE OF THE SB AND OB. SB OR OB IS DETERMINED DEPENDING ON WHICH PRESSURE HAS THE CHANGE POINT.

Case	SP	OP	DP	Class
1	✓			SB
2	✓		✓	
3	✓	✓		SB or OB
4		✓		OB
5		✓	✓	
6			✓	
7	✓	✓	✓	

Algorithm 2 P-ℓ_1CPD

1: Construct a prediction model with MODA-RNN and obtain a threshold value from Algorithm 1 by using train data.
2: For data collected in real time, the SP and OP from time steps $t+1$ to $t+18$ are predicted at current time t with the prediction model. The DP is calculated with SP-OP.
3: For each pressure (SP, OP, DP), calculate the trend estimates from time steps 1 to $t+18$ with ℓ_1 trend filtering.
4: Compute the absolute value of the difference, called 'the amount of changes', between two consecutive estimation values from time steps 1 to $t+18$.
5: **if** a change point whose amount of changes is larger than the threshold is newly found **then**
6: Check whether a change point occurs at other pressures during the next 18 minutes.
7: An alarm is generated at time $t+18$ based on the rule in Table II.
8: Go to Step 2.
9: **else**
10: Go to Step 2.
11: **end if**

At current time t, the SP and OP for the next $t+18$ time steps are predicted with the MODA-RNN, and the DP is calculated by SP-OP. Then ℓ_1 trend filtering is applied to each of pressure types, the SP, OP, and DP, from time steps 1 to $t+18$. It then works the same as ℓ_1CPD. Algorithm 2 summarizes the overall procedure of P-ℓ_1CPD.

V. EXPERIMENTS

A. Performance comparison of the CPD

We test the performance of our ℓ_1CPD with the real data. The four algorithms, WBS, SeqTest, NOT, and Segmented, are selected for the performance comparison. Table III shows the

TABLE III
CONTROL PARAMETERS OF CPD METHODS USED FOR COMPARISON

Method	Parameters to use
WBS	Threshold$\in \{50, 200\}$ Number of random interval $M = 10$
SeqTest	Type of statistical test$\in$ {Mood test, MW test}
NOT	Threshold$\in \{0.01, 0.05\}$ Number of random interval $M = 10$
Segmented	Threshold $\in \{0.001, 0.005\}$ Number of change points = 4

TABLE IV
EXPERIMENTAL RESULT OF CPD ALGORITHMS

Method	TDR	FAR	TCR	TDS
ℓ_1CPD (λ=20)	0.80 (0.30)	0.4 (0.26)	0.73 (0.29)	0.28
ℓ_1CPD (λ=30)	0.63 (0.42)	0.26 (0.20)	0.60 (0.40)	0.31
ℓ_1CPD (λ=50)	0.45 (0.24)	0.45 (0.28)	0.45 (0.24)	0.49
WBS (th=50)	0.46 (0.39)	2.41 (0.69)	0.28 (0.23)	1.14
WBS (th=200)	0.22 (0.25)	1.10 (0.51)	0.08 (0.20)	0.92
SeqTest (Mood)	0.72 (0.31)	4.25 (0.55)	0.31 (0.40)	1.09
SeqTest (MW)	0.86 (0.22)	5.55 (0.67)	0.57 (0.22)	0.88
NOT (th=0.01)	0.77 (0.25)	4.03 (0.58)	0.63 (0.42)	0.96
NOT (th=0.05)	0.25 (0.22)	0.85 (0.36)	0.08 (0.14)	0.79
Segmented (th=0.001)	0.68 (0.40)	4.05 (0.52)	0.33 (0.37)	1.13
Segmented (th=0.005)	0.33 (0.20)	1.06 (0.41)	0.25 (0.22)	0.84

parameters used for each method, and Table IV presents the experimental result. When the SB or OB is detected within a certain range from the true SB or OB point, it is considered to be found. The range, also called a decision interval, is set to two batches from the engineers.

In Table IV, four indicators, true detection rate (TDR), false alarm rate (FAR), true classification rate (TCR), and total detection score (TDS) are used to analyze the results. The experiments are conducted on six data-sets, and the mean and standard deviation in parenthesis for each indicator are presented in the table. The TDR represents the ratio of the number of alarms, which are detected as true alarms, without differentiating the SB and OB, from the proposed algorithm, to the total number of the actual SB and OB points. The FAR is computed with the ratio of the number of false alarms, which do not indicate the actual SB or OB, to the number of false alarms that are allowable. The denominator of the FAR is 10 which is the allowable number of alarms determined from the engineers. Whenever a change point is detected, the engineers need to check the machine states with the sensor data and to decide whether a further maintenance action is required. Therefore, it is desirable to reduce the number of false alarms. If the FAR is less than 1, it is assumed that the acceptable number of false alarms occurs; otherwise, too many false alarms occur. The TCR is the ratio of the number of true classification points (SB and OB) with the proposed method to the total number of the actual SB and OB. It indicates the effectiveness of the classification rule in Table II. Finally, we use the TDS to comprehensively consider the detection performance and the degree of occurrence of false alarms with $\sqrt{(1-TDR)*FAR}$. The performance of the algorithm is considered to be good when the TDR and FAR are close to 1 and 0, respectively. Therefore, the TDS close to 0 indicate that the algorithm performs well when considering the both aspects. The learning time of the algorithms is not presented because all the methods take less than 1 minute.

SeqTest (MW) has the largest TDR of 0.86, but the FAR is also large, which implies that it generates many unnecessary false alarms. Similarly, NOT (th=0.01) has a very high TDR of 0.77 but the FAR is 4.03. The FAR values greater than four indicate that approximately 40 alarms are produced. Considering that the length of one data set is 34 days, multiple checks per day are required, which is not practical.

NOT (th=0.05) has the FAR less than 1, but the TDR is 0.25. It implies that NOT (th=0.05) provides a small number of false alarms but there is a lack of ability to detect true alarms. On the other hand, the proposed ℓ_1CPD with λ of 20 have the FAR less than 1 and the large TDR of 0.80. We can see that the proposed method has the TDS of 0.28 which is the smallest compared to others. Therefore, it can be claimed that ℓ_1CPD detects the real SP and OP relatively well with a small number of false alarms. In addition, it is observed that ℓ_1CPD with λ of 20 effectively classifies the SB and OB from the TCR. It has the largest TCR of 0.73. Finally, the standard deviation of ℓ_1CPD (λ=20) has a value between 0.26 and 0.30, regardless of the indicator, meaning that the performance is more consistent compared to others.

Fig. 7 shows the alarms generated by a model that has the lowest TDS for each algorithm in Table IV. Fig. 7 (a) shows the true SB and OB points which are indicated with the blue and red dotted lines, respectively. The black line in other figures represents the false alarm. In Fig. 7 (b), we can see that ℓ_1CPD generates a small number of false alarms and detects the SB and OB relatively well. The WBS in Fig. 7 (c) has a low detection rate, and SeqTest in Fig. 7 (d) has a large number of false alarms. NOT in Fig. 7 (e) and Segmented in Fig. 7 (f) classify the blockage situation incorrectly.

B. Result of MSP

Now, the proposed MODA-RNN is compared with five other models, linear regression, RandomForest, Lasso, Ridge, and original DA-RNN. The MODA-RNN uses T' of 18, the look-back parameter T of 20, and the batch size (the number of samples propagated over the network) of 128 determined from the preliminary experiments.

The comparison methods are designed with 36 independent models, each of which predicts a pressure at a certain point in the source and ozone. For instance, the first model predicts the SP at $t+1$, and the last model predicts the OP at $t+18$. We use α of 1 and 0.1 for Ridge and Lasso, respectively. The number of trees in RandomForest is 200, the maximum depth has no limit, and the minimum number of samples is 1. The DA-RNN uses the batch size of 128 and the look-back parameter T of

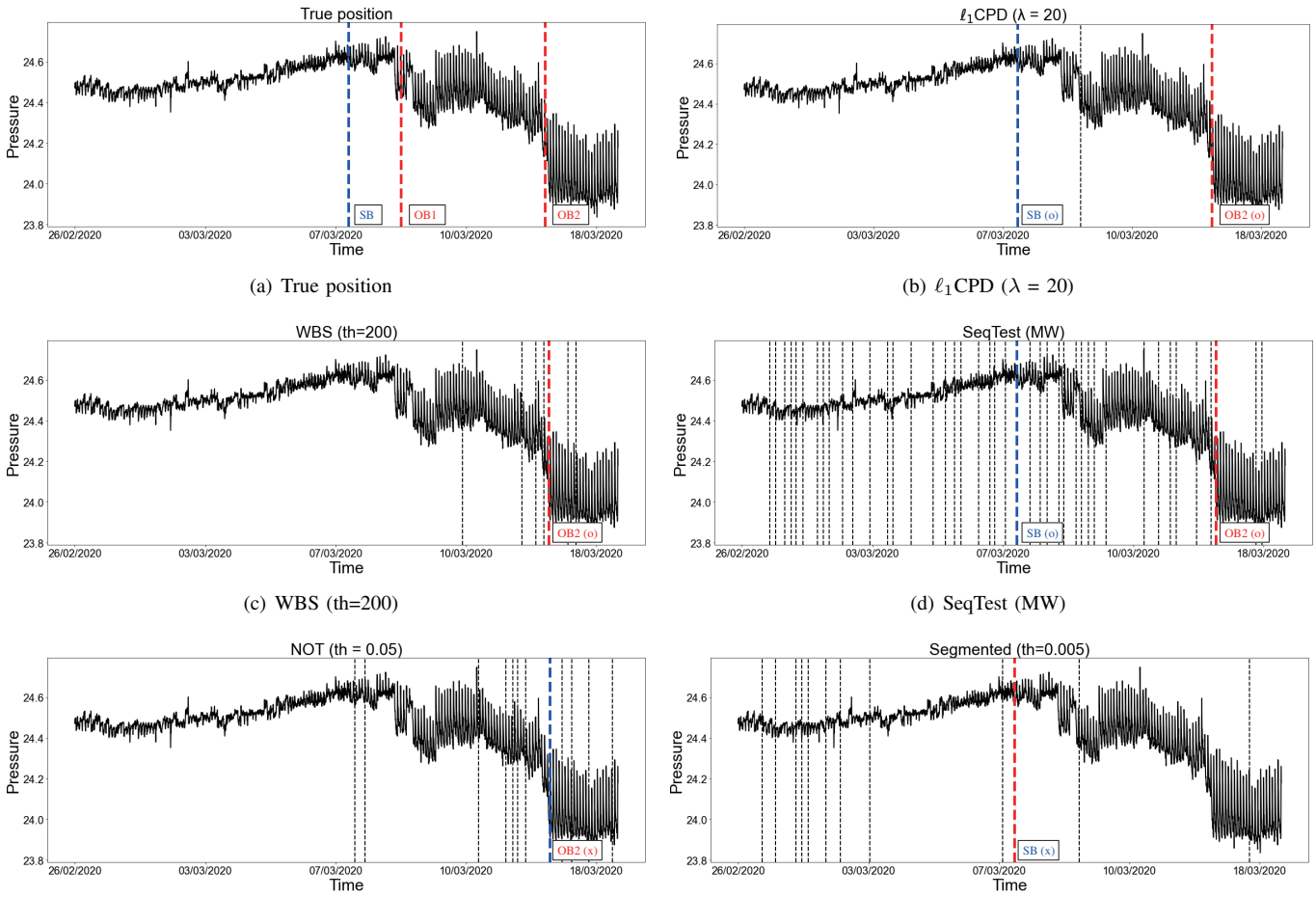


Fig. 7. An alarm generated by a model that has the lowest TDS. Fig. 7 (a) shows the true position of one SB and two OBs, which are indicated with the blue and red dotted lines, respectively. The black dotted lines in other figures are false alarms, and the blue and red dotted lines in Figs. 7 (b)-(f) are the SB and OB, respectively, detected with the corresponding model. The (o) and (x) mean the correct and incorrect alarm, respectively.

TABLE V
MAE OF PREDICTION MODELS

Model	SP	OP	DP	Average
linear regression	0.087 (0.016)	0.117 (0.036)	0.102 (0.065)	0.102 (0.039)
RandomForest	0.111 (0.017)	0.130 (0.022)	0.116 (0.040)	0.119 (0.026)
Lasso	0.148 (0.023)	0.195 (0.064)	0.085 (0.035)	0.143 (0.040)
Ridge	0.077 (0.019)	0.102 (0.032)	0.055 (0.018)	0.078 (0.023)
DA-RNN	0.067 (0.010)	0.099 (0.019)	0.082 (0.020)	0.083 (0.016)
MODA-RNN	0.022 (0.006)	0.025 (0.008)	0.015 (0.004)	0.021 (0.006)

20. Note that these parameters have been determined from the preliminary experiments with grid search.

Since there are three data sets for each machine, two of them are used as train data and the other one is used for test data (real-time data). The average mean absolute error (MAE) of five repetitions is presented in Table V. In the table, the value in parenthesis indicates the standard deviation of the MAE. We can see that the average MAE value of the proposed MODA-RNN is 0.021, whereas the DA-RNN has the MAE of 0.083. Ridge provides the MAE of 0.078, and other methods perform worse than Ridge. In addition, the standard deviation of the MAE of the MODA-RNN is 0.006, which is significantly smaller than those of other methods. The result shows that our MODA-RNN model has the best performance compared to other models.

C. Performance of the P-ℓ_1CPD

Table VI shows the performance of P-ℓ_1CPD. we can see that P-ℓ_1CPD performs better than ℓ_1CPD with the same λ in Table IV. When λ is 50, the TDR of ℓ_1CPD is relatively low as 0.45 because the trend estimate is very smooth. However, it is increased to 0.63 with the prediction model, which shows the effectiveness of forecasting the future SP and OP. Also, when λ is 20, P-ℓ_1CPD provides the TDS of 0.20, which is very small. We note that the SB and OB found with P-ℓ_1CPD are 18 minutes ahead of the points found with ℓ_1CPD on average.

TABLE VI
EXPERIMENTAL RESULT OF P-ℓ_1CPD

Method	TDR	FAR	TCR	TDS
P-ℓ_1CPD (λ=20)	0.86 (0.22)	0.31 (0.23)	0.82 (0.21)	0.20
P-ℓ_1CPD (λ=30)	0.69 (0.24)	0.26 (0.25)	0.62 (0.19)	0.28
P-ℓ_1CPD (λ=50)	0.63 (0.28)	0.43 (0.18)	0.45 (0.24)	0.39

VI. CONCLUSION

We have developed the ℓ_1 trend filtering-based CPD algorithm for detecting slope changes in data with an increasing trend. Since ℓ_1 trend filtering is useful to control the smoothness of data, it can be applied effectively in real time. We therefore have used the method for the pumping line balance problem of deposition machines faced with one of the semiconductor equipment companies in Korea. The ℓ_1CPD method has generated a relatively small number of false alarms and provided the accurate SB and OB compared to other four methods. Also, the proposed MODA-RNN model has performed better than the DA-RNN, linear regression, RandomForest, Lasso, and Ridge, and improved the performance of ℓ_1CPD. The proposed approach has been confirmed with engineers and is being implemented in the real machine. In the future, we can further improve the threshold algorithm and detection rule. In addition, CPD can be conducted using deep learning models.

REFERENCES

[1] L. Kang and S. L. Albin, "On-line monitoring when the process yields a linear profile," *J. Qual. Technol.*, vol. 32, no. 4, pp. 418–426, 2000.

[2] Y.-S. Jeong, B. Kim, and Y.-D. Ko, "Exponentially weighted moving average-based procedure with adaptive thresholding for monitoring nonlinear profiles: Monitoring of plasma etch process in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 40, no. 14, pp. 5688–5693, 2013.

[3] W. Puyati, A. Khawne, M. Barnes, B. Zwan, P. Greer, and T. Fuangrod, "Predictive quality assurance of a linear accelerator based on the machine performance check application using statistical process control and ARIMA forecast modeling," *J. Appl. Clin. Med. Phys.*, vol. 21, no. 8, pp. 73–82, 2020.

[4] F. Kang, J. Liu, J. Li, and S. Li, "Concrete dam deformation prediction model for health monitoring based on extreme learning machine," *Struct. Control Health Monit.*, vol. 24, no. 10, p. e1997, 2017.

[5] P. Kang, D. Kim, H.-J. Lee, S. Doh, and S. Cho, "Virtual metrology for run-to-run control in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2508–2522, 2011.

[6] S. Guo, W. G. Guo, A. Abolhassani, and R. Kalamdani, "Nonparametric, real-time detection of process deteriorations in manufacturing with parsimonious smoothing," *IISE Transactions*, vol. 53, no. 5, pp. 568–581, 2021.

[7] S. Guo, W. G. Guo, A. Abolhassani, R. Kalamdani, S. Puchala, A. Januszczak, and C. Jalluri, "Manufacturing process monitoring with nonparametric change-point detection in automotive industry," *J. Manuf. Sci. E.*, vol. 141, no. 7, 2019.

[8] A. Sen and M. S. Srivastava, "On tests for detecting change in mean," *Ann. Statist.*, pp. 98–108, 1975.

[9] P. Fryzlewicz, "Wild binary segmentation for multiple change-point detection," *Ann. Statist.*, vol. 42, no. 6, pp. 2243–2281, 2014.

[10] Z. Harchaoui and C. Lévy-Leduc, "Catching change-points with Lasso," in *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2007, pp. 617–624.

[11] G. J. Ross, D. K. Tasoulis, and N. M. Adams, "Nonparametric monitoring of data streams for changes in location and scale," *Technometrics*, vol. 53, no. 4, pp. 379–389, 2011.

[12] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "ℓ_1 trend filtering," *SIAM Rev.*, vol. 51, no. 2, pp. 339–360, 2009.

[13] C. S. Hwang, *Atomic Layer Deposition for Semiconductors*. New York, USA: Springer, 2014.

[14] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," 2017. [Online]. Available: http://arXiv.org/abs/1704.02971

[15] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowl. Inf. Syst.*, vol. 51, no. 2, pp. 339–367, 2017.

[16] V. M. R. Muggeo, "Estimating regression models with unknown break-points," *Statist. Med.*, vol. 22, no. 19, pp. 3055–3071, 2003.

[17] R. Baranowski, Y. Chen, and P. Fryzlewicz, "Narrowest-over-threshold detection of multiple change points and change-point-like features," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 81, no. 3, pp. 649–672, 2019.

[18] P. L. Davies, R. Fried, and U. Gather, "Robust signal extraction for on-line monitoring data," *J. Stat. Plan. Inference.*, vol. 122, no. 1-2, pp. 65–78, 2004.

[19] C. Truong, L. Oudre, and N. Vayatis, "Selective review of offline change point detection methods," *Signal Process.*, vol. 167, p. 107299, 2020.

[20] H. Yamada and L. Jin, "Japan's output gap estimation and ℓ_1 trend filtering," *Empirical Economics*, vol. 45, no. 1, pp. 81–88, 2013.

[21] H. Yamada, "Estimating the trend in US real GDP using the ℓ_1 trend filtering," *Appl. Econ. Lett.*, vol. 24, no. 10, pp. 713–716, 2017.

[22] S. Alawnah, S. Mukhopadhyay, and A. Sagahyroon, "Universal adaptive stabilization-based trend filtering for impending battery voltage collapse detection," *IEEE Contr. Syst. Lett.*, vol. 5, no. 2, pp. 635–640, 2020.

[23] A. Nadkarni and S. A. Soman, "Applications of trend-filtering to bulk PMU time-series data for wide-area operator awareness," in *Power Systems Computation Conference (PSCC)*, Dublin, Ireland, 2018, pp. 1–7.

[24] S. Selvin, S. G. Ajay, B. G. Gowri, V. Sowmya, and K. P. Soman, "ℓ_1 trend filter for image denoising," *Procedia Comput. Sci.*, vol. 93, pp. 495–502, 2016.

[25] A. S. Weigend, *Time Series Prediction: Forecasting The Future And Understanding The Past*. Evanston, IL, USA: Routledge, 2018.

[26] V. T. Tran, B.-S. Yang, and A. C. C. Tan, "Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems," *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9378–9387, 2009.

[27] G. Niu and B.-S. Yang, "Dempster–Shafer regression for multi-step-ahead time-series prediction towards data-driven machinery prognosis," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 740–751, 2009.

[28] H. Cheng, P.-N. Tan, J. Gao, and J. Scripps, "Multistep-ahead time series prediction," in *Proc. 10th Pacific-Asia Conf. Advances Knowl. Discovery Data Mining*, Berlin, Heidelberg, 2006, pp. 765–774.

[29] Y. Bao, T. Xiong, and Z. Hu, "Multi-step-ahead time series prediction using multiple-output support vector regression," *Neurocomputing*, vol. 129, pp. 482–493, 2014.

[30] L. Yunpeng, H. Di, B. Junpeng, and Q. Yong, "Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network," in *Proc. 14th Web Inf. Syst. Appl. Conf. (WISA)*, Liuzhou, China, 2017, pp. 305–310.

[31] L. Han, R. Zhang, X. Wang, A. Bao, and H. Jing, "Multi-step wind power forecast based on VMD-LSTM," *IET Renew. Power Gener.*, vol. 13, no. 10, pp. 1690–1700, 2019.

[32] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, and Y.-S. Wang, "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts," *J. Clean. Prod.*, vol. 209, pp. 134–145, 2019.

[33] M. M. Aliabadi, H. Emami, M. Dong, and Y. Huang, "Attention-based recurrent neural network for multistep-ahead prediction of process performance," *Comput. Chem. Eng.*, vol. 140, p. 106931, 2020.

[34] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.