Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023
doi:10.20944/preprints202306.1026.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Estimating APC Model Parameter for Dynamic Intervals Determined using Change Point Detection in Petrochemical Continuous Process

Yoseb Yu [1], Minyeob Lee [1], Chaekyu Lee [1], Yewon Cheon [2], Seungyun Baek [3], Youngmin Kim [3], Kyungmin Kim [3], Heechan Jung [4], Dohyeon Lim [5], Hyogeun Byun [6] and Jongpil Jung [1,*]

[1] Department of Smart Factory Convergence, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Korea; mditm153@g.skku.edu(Y.Y.); voninlee7@g.skku.edu(M.L.); leechgyu@skku.edu(C.L.)
[2] Department of Statistics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea ; 1000daughter@g.skku.edu (Y.C.)
[3] Department of Chemical Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Korea; tmddbsqor@g.skku.edu(S.B.); kym990325@g.skku.edu(Y.K.); rudals267555@g.skku.edu(K.K.)
[4] Department of Mechanical Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Korea; bay06059@g.skku.edu(H.J.)
[5] Department of System Management Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Korea; hukxie@g.skku.edu(D.L.)
[6] Department of Computer Science and Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Korea; byun0424@g.skku.edu(B.H.)
* Correspondence: jpjeong@skku.edu(J.J.); Tel.: +82-31-299-4267

Abstract: Most petrochemical plants still maintain a Proportional-Integral-Differential controller (PID) system, which is a feedback control system. However, gradually, the PID system is being extended and introduced to the Advanced Process Controller (APC) system, which is an integrated control system of feedforward and feedback that predicts external influences in advance. In the process of conducting on-site plant tests and calculating APC model parameter for the application of APC systems, a problem arises that Model Parameter are implemented differently depending on the proficiency of APC engineers. To minimise this problem, a technique for estimating APC model parameter without a plant test is required. In order to estimate the APC model parameter, it is necessary to train on dynamic interval data. In this paper, we use statistical techniques such as PELT, Linear Kernel, and Radial Basis Function Kernel of Change Point Detection (CPD) to extract dynamic data with minimum Mean Absolute Error (MAE) from time series data of a real petrochemical plant. Then, the hyper parameter is fixed and the APC model parameter is estimated by learning the dynamic section data. By applying the estimated APC model parameter to the APC Model Tool and measuring the fitting rate, it was confirmed that it is possible to estimate the APC model parameter with excellent control performance without plant test.

Keywords: Petrochemical, Continuous Process, Advanced Process Control, Change Point Detection, Model Parameter Estimation

1. Introduction

Most petrochemical plants still maintain a PID (Proportional-Integral-Differential Controller) system for feedback control [28]. However, gradually, the PID system, which is a feedback control, is being expanded and introduced to APC (Advanced Process Controller), which is an integrated control system of feedforward and feedback that predicts external influences in advance. The model underlying these two systems is the response of the output variable to the input of the manipulated variable. In particular, most control systems used in the industrial field adopt a first-order time-delay model with a stepped response. The reason for the widespread use of first-order time-delay models is not only their simplicity, but also their ability to capture the essential dynamics of many industrial processes [11].

© 2023 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

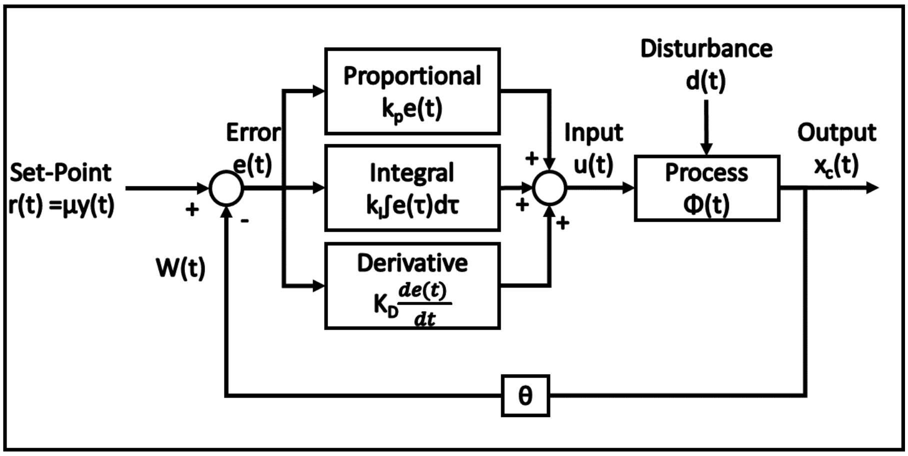

Figure 1. PID System Structure

The representative control method of classical control theory is PID control. As shown in Figure 1, the transfer function of a PID controller consists of three terms: proportional, integral, and derivative. The terms K_P, K_I, and K_D are called proportional, integral, and differential gains, respectively. As K_P increases, the overshoot increases, but the rise time decreases, approaching the target value faster and reducing the steady-state error. However, requiring a large amount of control can strain the system, and the settling time is not affected. As K_I increases, the overshoot increases due to the large change in the amount of control over the residual deviation from steady state, and since it is a fine control, the rise time decreases slightly and the steady state error, which is the goal of integral control, is eliminated, but the settling time increases additionally. Increasing K_D reduces the overshoot, reduces the rise time, and reduces the settling time because the error is corrected more quickly. However, it has no effect on steady-state error. The Laplace inversion of the transfer function G(s) given in Equation (1) yields the output equation in the time domain, as shown in Equation (2).

$$G(s) = K_P + \frac{K_I}{s} + K_D s \tag{1}$$

$$y(t) = K_p e(t) + K_I \int_0^t e(\tau)\, d\tau + K_D \frac{de(t)}{dt} \tag{2}$$

PID system is characterised by a 1:1 control method between output and input variables, and feedback control that compensates for the difference between predicted and actual values in real time. The control performance is determined by the PID tuning value.

Unlike PID systems, APC systems can be applied to processes with large time delays, unstable processes, and multi-variable processes. The characteristics of APC control are N:N control method, which can control multiple output variables with multiple input variables, and a mixed control method of feedback control, which controls by compensating the difference between the predicted value and the actual value in real time, and feedforward control, which considers the influence of external disturbance variables in advance and controls before the external influence [19]. The following figure presents a schematic diagram of the APC system.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

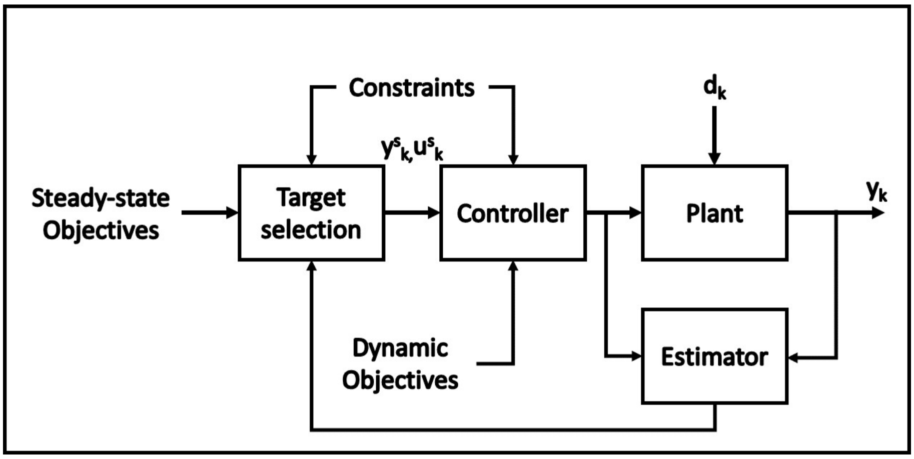

Figure 2. APC System Structure

A typical APC system consists of a system model, constraints, disturbance model, cost function, optimisation method, and control range, all of which can affect the performance of the APC System [1]. In general, the APC system used in industrial applications is modeled as a first-order time-delay control system, as shown in Equation (3). Where K = gain, TD = Delay, and τ = Time Constant. By performing the Laplace inverse transformation, the output equation can be obtained in the time domain of the first-order time-delay model as shown in Equation (4).

$$G(s) = K\frac{e^{-Ds}}{1+\tau s} \tag{3}$$

$$y(t) = K(1 - e^{-\frac{t}{\tau}}) \cdot x(t - D) \tag{4}$$

The performance of an APC system is determined by the APC model parameter values calculated from the formula above. APC systems require regular APC model parameter maintenance over time due to changes in production grades, equipment obsolescence, and replacement. Programs that help APC engineers easily obtain Model Parameter (MATLAB, Model-ID, etc.) are commercially available, but in the process of conducting on-site plant tests and calculating APC model parameter, APC model parameter are implemented differently depending on the proficiency of APC engineers. This leads to a decrease in the accuracy of the APC model and the control performance of the APC system. Therefore, a method for estimating APC model parameter without a plant test is needed to minimise the difference in the proficiency of APC engineers.

In order to estimate the APC model parameter, it is necessary to train the data in the dynamic interval to know the correlation between the output variable (CV) and the manipulated variables (MV, DV). However, it is not easy to determine the dynamic interval due to the complexity of process dynamics [14]. In this paper, we use the time series data of a real petrochemical plant to extract the dynamic intervals of the time series data using various statistical techniques of Change Point Detection (CPD). In order to estimate the accurate APC model parameter, a comparison of PELT-based Learning, Linear Kernel-based Learning, and Radial Basis Function Kernel-based Learning of CPD is performed to find the hyper parameter of the dynamic section with the smallest MAE (Mean Absolute Error). Then, the APC model parameter is estimated using the Levenberg-Marquardt algorithm in the dynamic range of the fixed hyper parameter. Finally, the estimated APC model parameter is applied to the APC control program to verify the accuracy by presenting the fitting rate results in three random sections of evaluation data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

By comparing the fitting rate of the estimated APC model parameter in three random sections of evaluation data, the average fitting rate of 86.09% for Plant A and 79.94% for Plant B was found. This shows that it is possible to estimate APC model parameter with good control performance without plant tests.

2. Background and Methodology

2.1. APC Model Design Flow

Considering the inherent characteristics of the process, such as time delays, mutual interference, back-reaction, and process constraints, the APC system is introduced because the PID system alone has limitations in performing optimised operation. To carry out an APC project, the following steps should be taken [39].

(1) Functional Design
By analysing the flow and operation purpose of the target process, the PID loop performance of the MV to be used by the APC system is identified, and the CV, MV, and DV to configure the APC system for optimal operation are set [39].

(2) Plant Test
The APC system is a model-based control system. To obtain the model, a plant test is required to check the movement of CV by changing the expected MV and DV with the desired amplitude at the appropriate period. At this time, the actual process is moved arbitrarily, so when performing the plant test, sufficient consultation should be made with the person in charge of the site, and care should be taken to prevent process problems from occurring [39]. Figure 3 below shows an example of a plant test.

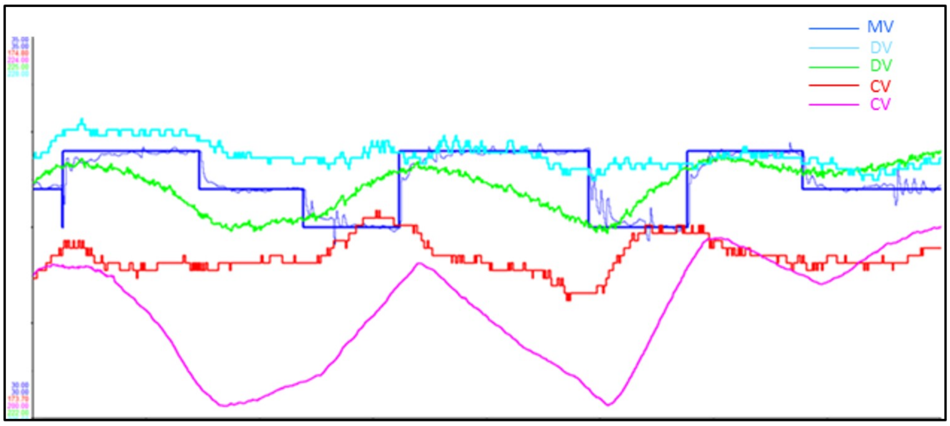

Figure 3. Plant Test

(3) Detail Design
Using the data obtained through the plant test, the process variables such as CV, MV, and DV selected in the basic design are finally determined, and a dynamic model is built to predict and verify the control performance of the APC with an offline controller. The dynamic model has a structure such as FIR (Finite Impulse Resp), Laplace, etc. and is mostly designed using a programme provided by the APC system supplier as shown in Figure 4 [39]. Offline verification can be achieved by using an APC simulator to mimic the actual process and apply the APC system to verify its performance. This can reduce the implementation error when building an APC system online [Figure 5].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

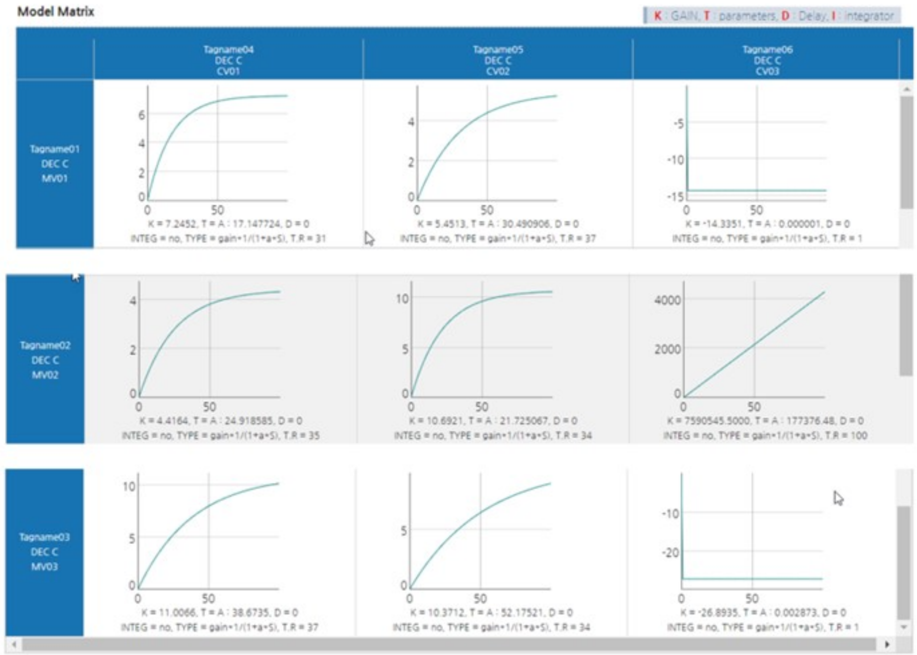

Figure 4. Calculate a Dynamic Characteristic Model of the APC

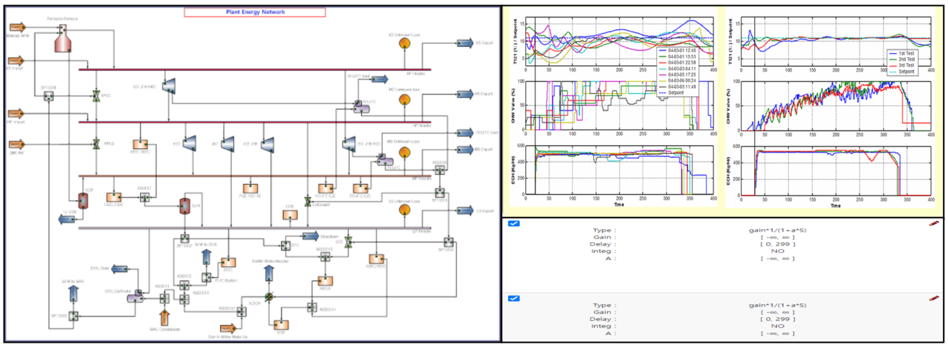

Figure 5. Offline Validation of APC system

(4) Commissioning and Performance Analytics

Commissioning is the stage where the actual APC system is brought online and optimised for operation. Once the APC model parameter are tuned, the actual benefits are calculated by analysing the performance of the optimised operation by comparing the performance before and after the application of the APC system as shown in Figure 6 [39].

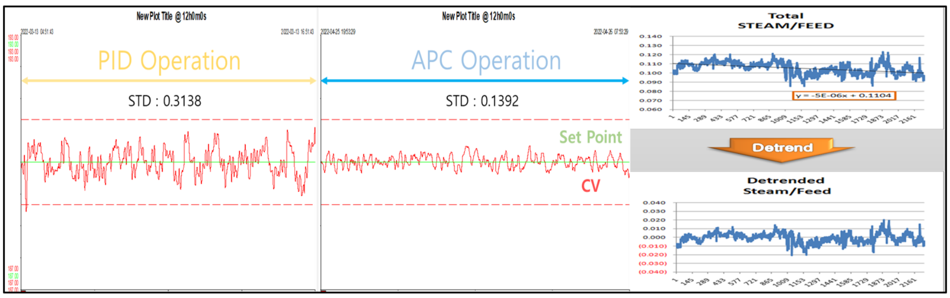

Figure 6. APC Model Commissioning & Performance Analytics

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

2.2. Change Point Detection

Change Point Detection (CPD) is a statistical technique that literally looks for points of trend change in time series data, as shown in the figure below. In other words, it is a matter of finding points in the time series data where the time series characteristics such as mean, standard deviation, and slope change rapidly. Below [Figure 7] is an example of CPD, and the vertical lines marked with blue dots in the figure below represent the points where trend changes are detected.

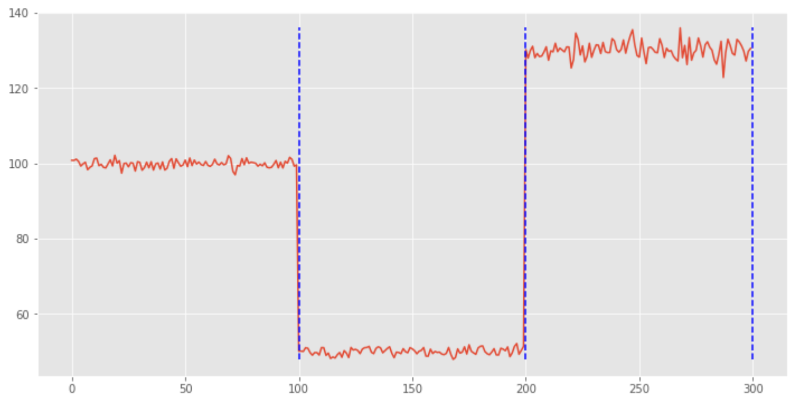

Figure 7. Example of Applying Change Point Detection

CPD is performed by dividing the time series data into bins and minimising the sum of the costs for each bin. In other words, CPD can be viewed as a kind of partial time series clustering problem that bins time series data with similar characteristics. In other words, the starting point of each bin can be called a change point. The change detection problem for the time series y=(y_1,y_2 $\cdots$,y_t) can be formally defined as follows.

$$S^* = \arg\min_{S} \sum_{s \in S} C(s) \tag{5}$$

Where S is the set of bins, C is the cost function for the bins, and S^* is the optimal set of bins.

(1) PELT(Pruned Exact Linear Time)

The PELT algorithm is a change detection algorithm that finds the optimal interval in linear time when the number of change points is unknown, and consists of the following steps [17,30].

(a) Input: time series y, cost function C, penalty β.
(b) Step 1: Initialise
    - Initialise z as an empty array of size $T + 1$.
    - Initialise with $Z[0] = -\beta$
    - Initialise with $L[0] = \varnothing$
    - Initialise with $x = \{0\}$
    - Initialise with $t = 1$
(c) Step 2: Update $\hat{t}$, $Z[t]$, $L[t]$, and χ as follows
    - $\hat{t} \leftarrow \arg\min_{\tau \in \chi}(Z[\tau] + C(y_{\tau:t}) + \beta)$
    - $Z[\hat{t}] \leftarrow Z[t] + C(y_{\hat{t}:t}) + \beta$
    - $L[t] \leftarrow L[\hat{t}] \cup \{\hat{t}\}$
    - $\chi \leftarrow \{\tau \in \chi : Z[\tau] + C(y_{\tau:t}) \leq Z[t]\} \cup \{\hat{t}\}$
(d) Step 3: Terminate the algorithm if $t = \tau$, otherwise increment t by 1 and return to step 2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

(2) Kernel Change Point Detection

Kernel Change Point Detection is a method for dividing bins based on the change in the mean of each bin [5]. In KCP, data is projected into a high-dimensional space through a measurable function, Kernel, and then change points are detected by comparing the homogeneity of each sequence [3,15]. It is characterised by the fact that individual points are mapped using a mapping function ϕ, i.e., the cost for a set of intervals S is defined as

$$C(S) = \sum_{s \in S} \sum_{y_t \in S} \|\phi(y_t) - \bar{s}\|^2 \tag{6}$$

where $\bar{s}$ is the average of every value in the interval s for every element in the interval s. During the mapping process, we can use the following kernel functions

$$\|\phi(y_t)\|^2 = K(y_t, y_t) \tag{7}$$

$$\phi(y_t) \cdot \phi(y_\tau) = K(y_t, y_\tau) \tag{8}$$

Where K represents a kernel function, the most commonly used kernel functions are Linear Kernel, RBF Kernel (Radius Basis Function Kernel), etc.

3. Research Methods Overall Process

3.1. Overall Architecture

In this section, we first describe the overall structure of the study to estimate the APC model parameter and propose the expected results of the study in terms of the estimated APC model parameter. Next, as described in the introduction, it is very important to collect data of dynamic intervals to estimate the APC model parameter, so we describe the research methodology of CPD used to find data of dynamic intervals, and the definition and purpose of the evaluation metric Mean Absolute Error (MAE). We also explain the rationale for the scope of the hyper parameter Grid used to find the MAE. First, the overall process of this study can be briefly described as follows, which consists of four steps, as shown in Figure 8.

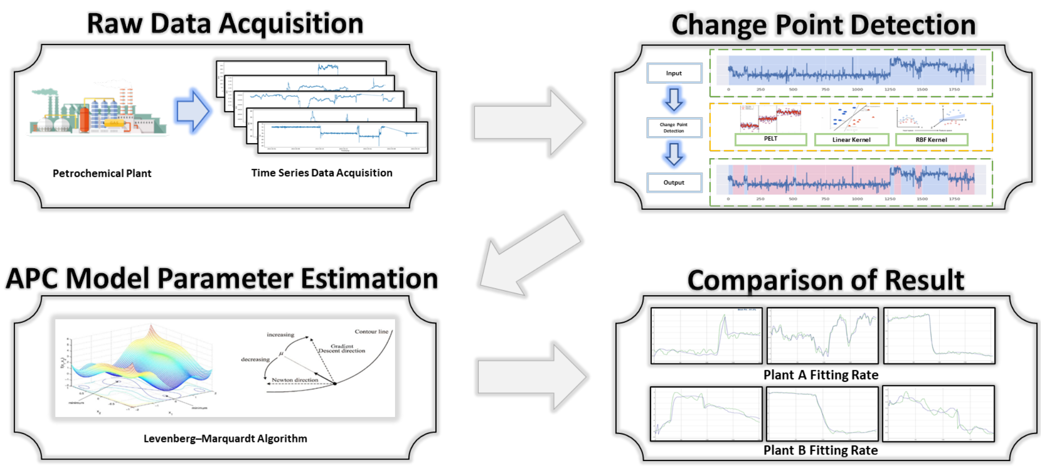

Figure 8. The Overall Research Process

(1) Raw Data Acquisition
(2) Change Point Detection
(3) APC Model Parameter Estimation
(4) Comparison of Result

The data used in this study are time series data of actual petrochemical plants in Korea, collected through OPC DA (Data Access) connected to DCS (Distributed Control System),

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

and the names of each plant are anonymised as Plant A and Plant B in consideration of data security issues, and output variables (CV) and manipulation variables (MV, DV) are separated. In order to estimate APC model parameter without plant test, which is the purpose of this study, it is essential to cluster only the dynamic part of the data. CPD algorithm is used to find the dynamic section of the data. Among the CPD algorithms, PELT, Linear Kernel, and RBF Kernel techniques are used to find the hyper parameter of the dynamic region with the minimum MAE (Mean Absolute Error). First, the CPD algorithm is used to find the hyper parameter of the dynamic section with the minimum MAE, and then the value of the hyper parameter is fixed to estimate the APC model parameter, K (Gain) and T (Time Constant), through the Levenberg-Marquardt algorithm. Finally, the estimated APC model parameter are applied to the APC control programme to verify the accuracy by comparing the fitting rate of the predicted and actual values. As the MAE value of the manipulated variables (MV, DV) for each output variable (CV) becomes smaller through CPD, the fitting rate of the estimated APC model parameter will show high accuracy, and the fitting rate is expected to increase as the number of manipulated variables (MV, DV) required to predict and control the output variable (CV) increases.

3.2. Designing a Change Point Detection Model

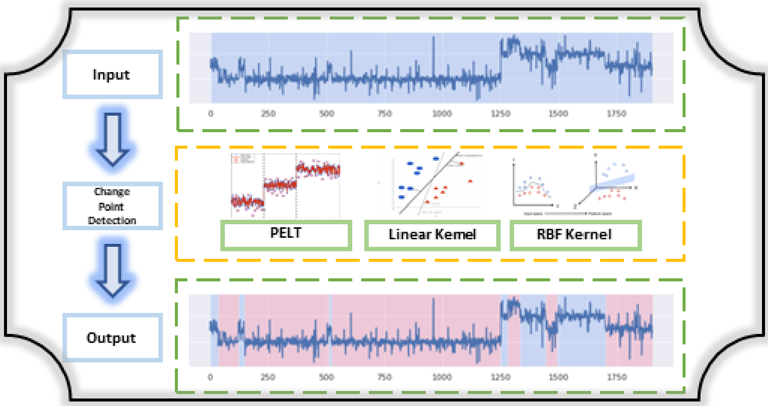

Figure 9. Change Point Detection Techniques

In order to estimate APC model parameter from time series data, it is necessary to consider the dynamic nature of the process, so it is first necessary to accurately identify the dynamic regions of the data. In this study, CPD algorithms such as PELT, Linear Kernel, and RBF Kernel are used to identify fluctuating data as shown in Figure 9. However, even with CPD techniques, it is difficult to perfectly identify when the trend starts to change. In order to perfectly detect when the trend starts to change, we need to be very sensitive to detecting the trend change, which is problematic because we will consider most of the points as change points. Therefore, for any point with a trend change found using CPD, we can define the fluctuating interval as -a to r+a where a is calculated by the following algorithm.

(1) Calculate the slope θ_1 of a simple linear regression model using the data from r-1 to r+1.
(2) Initialise the variable a which represents the length of the interval, to 2.
(3) Let the independent variable be t and the dependent variable be r-a to r+a Compute the slope θ_a of a simple linear regression model with data from r-a to r+a.
(4) $\theta_{a-1} \times \theta_a < 0$ or the rate of change of the intercept of θ_{a-1} relative to the intercept of θ_a, $\frac{|\theta|_a - |\theta|_{a-1}|}{|\theta|_{a-1}} > \varepsilon$, return a-1. Otherwise, increase a by 1 and revert to (3).

3.3. Defining and Purpose of Performance Indicators

This section explains what Mean Absolute Error (MAE) is and why it is used as a metric in CPD. MAE is a popular metric used to determine if a regression model has been

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

trained properly. It can be derived by converting the difference between the actual correct answer and the predicted value into absolute value and then averaging it, and the smaller the value, the better the performance of the model. The formula for MAE is as follows

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{9}$$

In the formula, y_i is the actual value and $\hat{y}_i$ is the predicted value. Since MAE takes an absolute value, it is difficult to determine whether the regression model predicted higher or lower than the actual answer. However, it is easy to interpret the results because it has the same unit as the actual answer value and the predicted value. Also, since it takes absolute values, the results can be interpreted intuitively. In this paper, we need an intuitive indicator not to evaluate the correlation of the model, but to detect the change point of the data and indicate the difference between the predicted value and the actual value. Therefore, we use MAE as a performance indicator for detecting change points.

3.4. Defining and Designing Hyper Parameters

Hyper parameter refers to a variable that is manually set by the user in the model to achieve an optimal training model. In this study, we compare the MAE of the three change point detection techniques mentioned above, Pruned Exact Linear Time (PELT), Linear kernel-based technique, and Radial Basis Function (RBF) kernel-based technique, to identify the dynamic interval with the smallest MAE value and find the hyper parameter. To effectively detect changes in time series data, Delay (D) and Minimum Segment Size (MS) are applied as hyper parameters. Since there is no rule or method to design the hyper parameter Grid perfectly mathematically, the range is set and tuned empirically. In general, it is suggested to set the range from 0 to 10 for hyper parameters and 5 to 20 for MS because they rarely have values above 10 when affected by variables in the same process. D, If the hyper parameter Grid of MS has a narrow spacing, you can find a hyper parameter with good performance, but the computation time will increase considerably.

4. Experimental Results

4.1. Experimental Environments

Specify the experimental environment of the study for reproducibility of the experiment.

- Hardware platform architecture: GPU-enabled laptop
- Laptop configuration: CPU Core i5-8250, quad-core processor, 8GB RAM.
- Operating system: Window 10.

For an optimal development environment that is fast and collaborative, we used the cloud-based Google Collaboratory. Google Collaboratory provides a free Jupyter Notebook environment and is available on the cloud without installation. Google Collaboratory enables high-performance development, sharing, and computing resources. In particular, you can process large amounts of data.

4.2. Experimental Datasets Design

The primary objective of the experiment is to determine the data in the dynamic section in order to estimate the correct APC model parameter. Finally, the objective is to train the model with data from only the dynamic bins and estimate the APC model parameter with good performance. The data used in this study are time series data obtained from two different factories, Plant A and Plant B, at two different times, which are referred to as A-1, A-2, B-1, and B-2, respectively. The sampling time of each data is 1 minute, and the experiment is conducted on 5 days of data. The range of each data is as follows.

- A-1: 23 April 2023 00:00:00 to 27 April 2023 23:59:00
- B-1: 26 March 2023 00:00:00 to 30 March 2023 23:59:00

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

- A-2: 30 April 2023 00:00:00 to 04 May 2023 23:59:00
- B-2: 19 April 2023 00:00:00 to 23 April 2023 23:59:00

Let A-1 and B-1 be the training data and A-2 and B-2 be the respective test data. In addition, all four data described above consist of one control variable (CV) and two manipulation variables (MV, DV). In addition, the experimental trend change detection technique and its hyper parameter grid are shown in the following Table 1.

Table 1. Change Point Detection Techniques과 Hyper Parameter Grid

Change Point Detection Technique	Hyper Parameter Grid
PELT	MS: {5, 10, 15, 20} D: {0, 3, 5, 10}
Kernel-based detection ['Linear', 'RBF']	MS: {5, 10, 15, 20} D: {0, 3, 5, 10}

Otherwise, the hyper parameter Grid evaluates for the following sections

- ε : 0.001, 0.01, 0.05, 0.1

The specific process using the dataset is as follows
[Step 1]
To objectively evaluate the proposed Plant A and Plant B data, divide the dataset into a training dataset and a test dataset. Specifically, the first 5 days data (50%) of Plant A and Plant B are used as training data, and the last 5 days data (50%) are used for testing.
[Step 2]
Hyper parameters are divided into MS and MS, and dynamic bins are detected and identified using PELT, Linear kernel-based technique, and RBF kernel-based technique. We use the above three methods because they are the most commonly used methods for anomaly detection in time series data in CPD's previous studies [17,30,35-36].
[Step 3]
The data trained by the PELT, Linear kernel-based, and RBF kernel-based methods are used to determine the accuracy of the dynamic segments through the MAE metric. Here, the hyper parameter of the algorithm with the smallest MAE is fixed, and the APC model parameter estimation proceeds.
[Step 4]
The APC model parameter trained with the above proposed metrics is learned by applying the following equation to the Levenberg-Marquardt algorithm.

$$y(t) = K(1 - e^{-\frac{t}{\tau}}) \cdot x(t - D) \tag{10}$$

Here, each variable means the following
t : Elapsed Time
y(t): Value of the CV (output variable) for time t
x(t): Value of the MV (manipulation variable) for time t (may be replaced by DV)
K : Gain
τ : Time Constant
D : Delay
[Step 5]
Verify the accuracy of the control performance by comparing the fitting rate of the predicted and actual values with the APC model parameter estimates obtained in Step 3 and Step 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

4.3. Results

(1) Experiment 1 Results: Plant A For the manipulated variable MV, the 10 models and hyper parameters with the smallest MAE are shown in Table 2.

Table 2. Plant A - Parameter tuning results for the manipulated variable MV

Rank	Algorithms	D	MS	ε	MAE
1	Linear Kernel	10	10	0.05	8.642452
2	Linear Kernel	10	15	0.05	8.651746
3	Linear Kernel	10	5	0.05	8.656836
4	Linear Kernel	10	20	0.05	8.674863
5	Linear Kernel	5	10	0.05	8.682052
6	Linear Kernel	5	5	0.05	8.691545
7	Linear Kernel	3	10	0.05	8.693691
8	Linear Kernel	5	15	0.05	8.699134
9	Linear Kernel	3	5	0.05	8.700352
10	Linear Kernel	3	15	0.05	8.703656

For the manipulated variable DV, the 10 models and hyper parameters with the smallest MAE are shown in Table 3.

Table 3. Plant A - Parameter tuning results for the operational variable DV

Rank	Algorithms	D	MS	ε	MAE
1	Linear Kernel	10	20	0.01	10.73935
2	Linear Kernel	5	20	0.01	10.739671
3	Linear Kernel	3	20	0.01	10.739682
4	Linear Kernel	0	20	0.01	10.740037
5	Linear Kernel	0	15	0.01	10.741053
6	Linear Kernel	0	5	0.01	10.741164
7	Linear Kernel	0	10	0.01	10.741403
8	Linear Kernel	0	5	0.01	10.741504
9	Linear Kernel	10	5	0.01	10.742831
10	Linear Kernel	5	5	0.01	10.743456

From the two tables, we can see that kernel-based detection with a linear kernel performs well. The average MAEs for D, MS, and ε are shown in Tables 4, 5, and 6, respectively. Depending on the manipulated variables, we can see that the distribution of performance across parameters is significantly different. For example, for the manipulation variable MV, the larger the value, the better the performance, while for DV, the smaller the value, the better the performance. This suggests that it is very important to tune the appropriate parameters according to the data.

Table 4. Average MAE according to D: Plant A

Manipulated Variables	D	Mean MAE
MV	0	9.934820
	3	9.914994
	5	9.905677
	10	9.820333
DV	0	11.741451
	3	12.000861
	5	11.999641
	10	11.997169

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

Table 5. Average MAE according to MS: Plant A

Manipulated Variables	MS	Mean MAE
MV	5	9.876135
	10	9.904980
	15	9.927402
	20	9.867307
DV	5	11.864946
	10	12.038425
	15	11.882389
	20	11.953361

Table 6. Average MAE according to ε: Plant A

Manipulated Variables	ε	Mean MAE
MV	0.01	9.898091
	0.05	9.579578
	0.10	10.204200
DV	0.01	11.386267
	0.05	12.186671
	0.10	12.231404

A comparison of the average MAE across algorithms is shown in Table 7 below

Table 7. Average MAE according to Algorithms: Plant A

Manipulated Variables	Algorithms	Mean MAE
MV	Linear Kernel	9.399081
	RBF Kernel	9.507844
	PELT	10.774943
DV	Linear Kernel	11.443224
	RBF Kernel	12.540117
	PELT	11.821000

As shown in Table 7, we can see that the linear kernel-based change detection algorithm performs well for both manipulated variables. Based on the best performing linear kernel-based hyper parameter of = 10, MS = 10, and ε = 0.05 for the manipulated variable MV, the Levenberg-Marquardt algorithm yielded a model parameter estimate of MV for CV of K = 15.3188 and τ = 0.3221. Also. Based on the best performing Linear kernel-based Hyper Parameter for the manipulated variable DV, = 5, MS = 10, and ε = 0.01, the Model Parameter estimates for DV for CV resulted in K = 22.85, τ = 0.0309. The graphical representation of the APC model parameter estimated based on the best performing model for the operational variables MV and DV and the result of measuring the fitting rate through the APC Model Tool is shown in the figure below [10,11,12]. The intervals were randomly selected from the evaluation data intervals of Plant A. Three intervals of about 200 minutes in length were selected.

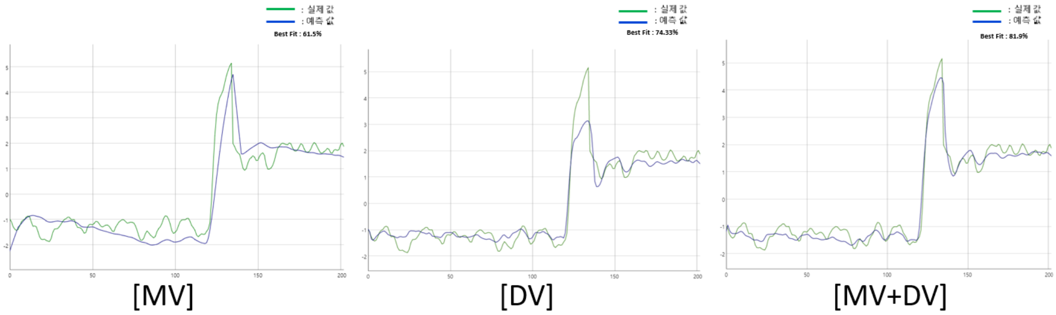

Figure 10. (Section 1) Fitting rate of CV with Estimated APC Model Parameter

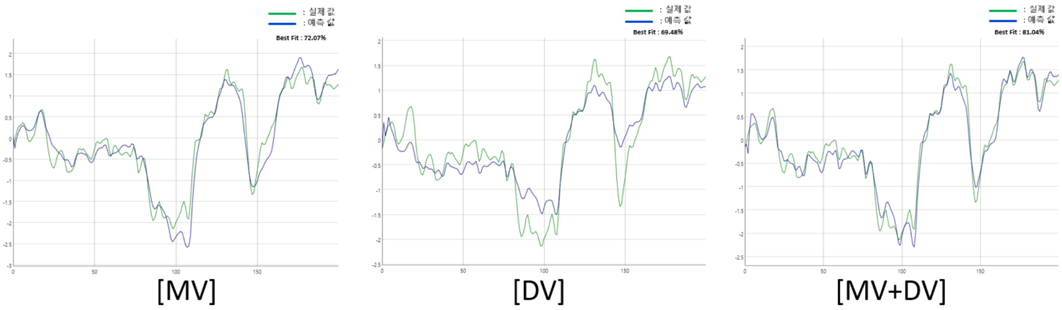

Figure 11. (Section 2) Fitting rate of CV with Estimated APC Model Parameter

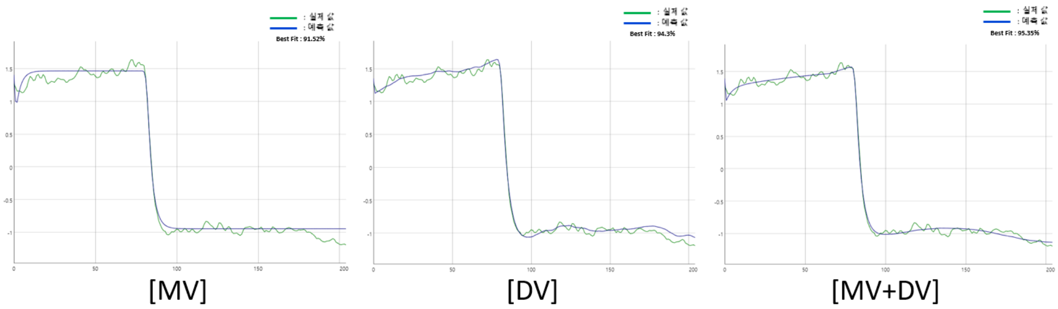

Figure 12. (Section 3) Fitting rate of CV with Estimated APC Model Parameter

We measured the fitting rate of the estimated APC model parameter to the predicted and actual values using the APC Model Tool, and found the following results.

- (Section 1) Fitting Rate of CV with Estimated APC Model Parameter(MV+DV): 81.9%
- (Section 2) Fitting Rate of CV with Estimated APC Model Parameter(MV+DV): 81.04%
- (Section 3) Fitting Rate of CV with Estimated APC Model Parameter(MV+DV): 95.35%

(2) Experiment 2 Results: Plant B For the manipulated variable MV, the 10 models and hyper parameters with the smallest MAE are shown in Table 8.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

Table 8. Plant B - Parameter tuning results for the manipulated variable MV

Rank	Algorithms	D	MS	ε	MAE
1	RBF Kernel	0	10	0.01	45.629945
2	RBF Kernel	0	5	0.01	45.647986
3	RBF Kernel	0	20	0.01	45.813426
4	Linear Kernel	0	5	0.1	46.103089
5	Linear Kernel	0	5	0.05	46.117504
6	RBF Kernel	0	15	0.01	46.118124
7	RBF Kernel	0	5	0.1	46.121928
8	RBF Kernel	10	10	0.01	46.134523
9	RBF Kernel	5	10	0.01	46.136264
10	RBF Kernel	0	15	0.1	46.147291

For the manipulated variable DV, the 10 models and hyper parameters with the smallest MAE are shown in Table 9.

Table 9. Plant B - Parameter tuning results for the operational variable DV

Rank	Algorithms	D	MS	ε	MAE
1	RBF Kernel	3	20	0.01	4.110627
2	RBF Kernel	5	20	0.01	4.192519
3	RBF Kernel	0	20	0.01	4.19666
4	RBF Kernel	10	20	0.01	4.200493
5	PELT	0	20	0.1	4.216778
6	Linear Kernel	0	10	0.01	4.231443
7	Linear Kernel	3	10	0.01	4.234592
8	PELT	3	20	0.1	4.269921
9	Linear Kernel	5	10	0.01	4.27022
10	PELT	5	20	0.1	4.274902

The averages of MAE according to D, MS, and ε are shown in Tables 10, 11, and 12, respectively. We can see that the parameterized performance distributions differ significantly depending on the manipulated variables.

Table 10. Average MAE according to D: Plant B

Manipulated Variables	D	Mean MAE
MV	0	47.014508
	3	47.716971
	5	47.703097
	10	47.672040
DV	0	4.402237
	3	4.417269
	5	4.420273
	10	4.431750

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

Table 11. Average MAE according to MS: Plant B

Manipulated Variables	MS	Mean MAE
MV	5	47.090598
	10	47.331090
	15	47.803609
	20	47.881317
DV	5	4.425821
	10	4.406654
	15	4.43916
	20	4.404137

Table 12. Average MAE according to ε: Plant B

Manipulated Variables	ε	Mean MAE
MV	0.01	47.603724
	0.05	47.531555
	0.10	47.444681
DV	0.01	4.339782
	0.05	4.431148
	0.10	4.482717

A comparison of the average MAE across algorithms is shown in Table 13 below.

Table 13. Average MAE according to Algorithms: Plant B

Manipulated Variables	ε	Mean MAE
MV	Linear Kernel	47.092989
	RBF Kernel	46.742333
	PELT	48.744638
DV	Linear Kernel	4.399071
	RBF Kernel	4.498206
	PELT	4.356370

As shown in Table 13, the appropriate algorithm depends on the manipulated variables. We can also see that the difference in performance across manipulated variables is nearly an order of magnitude. This suggests that the performance difference can be large depending on which variable is used to predict the control variable. Based on D = 0, MS = 5, and ε = 0.05, the linear kernel-based hyper parameters that performed best for the manipulated variable MV, the model parameter estimation of MV for CV using the Levenberg-Marquardt algorithm was K = 2.644706, τ = 0. 038512. Based on D = 0, MS = 15, ε = 0.1, the best performing hyper parameter based on the PELT technique for the manipulated variable DV, the model parameter estimate of DV for CV was K = 0.9556, τ = 0.0337. The graphical representation of the APC model parameter estimated based on the best performing model for the operational variables MV and DV and the fitting rate measured by the APC Model Tool is shown in the figure below [13,14,15]. The intervals were randomly selected from the evaluation data intervals of Plant B. Three intervals of about 200 minutes in length were selected.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

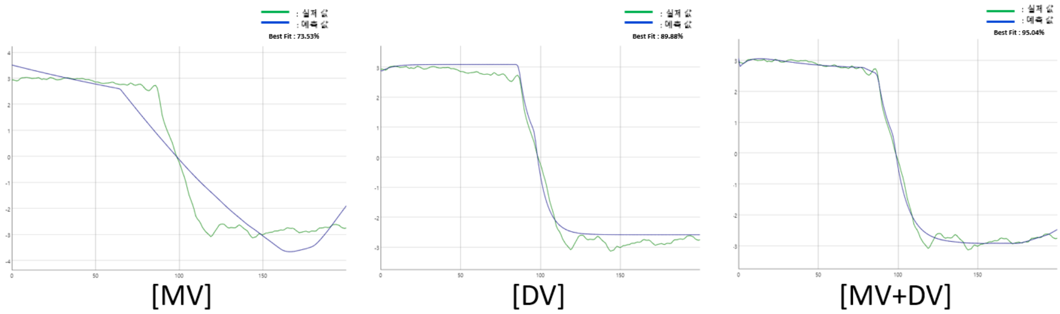

Figure 13. (Section 1) Fitting rate of CV with Estimated APC Model Parameter

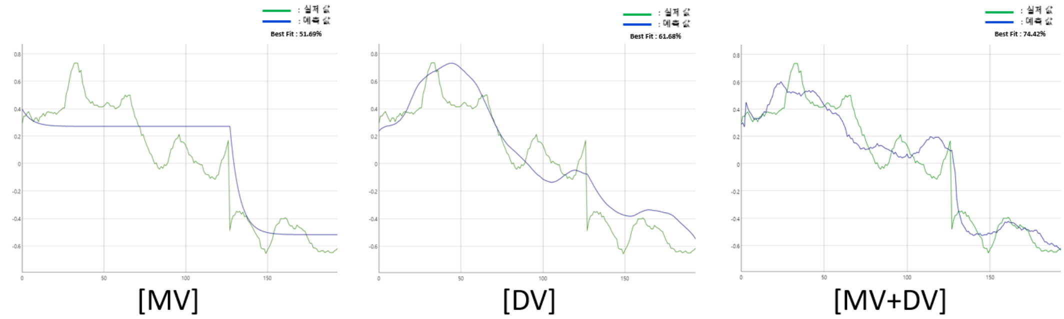

Figure 14. (Section 2) Fitting rate of CV with Estimated APC Model Parameter

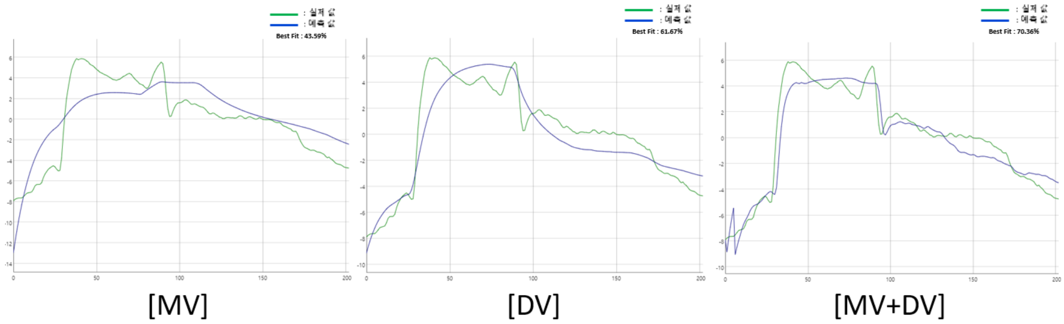

Figure 15. (Section 3) Fitting rate of CV with Estimated APC Model Parameter

We measured the fitting rate of the estimated APC model parameter to the predicted and actual values using the APC Model Tool, and found the following results.

- (Section 1) Fitting Rate of CV with Estimated APC Model Parameter(MV+ DV): 95.04%
- (Section 2) Fitting Rate of CV with Estimated APC Model Parameter(MV+ DV): 74.42%
- (Section 3) Fitting Rate of CV with Estimated APC Model Parameter(MV+ DV): 70.36%

5. Conclusion

APC model parameter play a key role in APC control. Most of the previous studies have been conducted in various industries such as semiconductor and bio, but few papers have been applied to the petrochemical industry. Since it is essential to maintain the APC system over time, it is very important to obtain dynamic interval data to estimate the APC model parameter. In this paper, PELT, Linear kernel-based, and RBF kernel-based techniques were applied for Change Point Detection as described in Chapter 3 to evaluate the MAE of the dynamic section. The results show that the Linear kernel-based method is the best for MV and DV of Plant A, the RBF kernel-based method is the best for MV of Plant B, and the PELT method is the best for DV. Since the variables can be current values, set values, or valve values of flow, pressure, temperature, etc. in petrochemical processes, it suggests that the performance of the model can be significantly different depending on

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

which variables are used to predict the control variables. As shown in the experimental results in Chapter 4, the CV control method that considers both MV and DV has the highest fitting rate, rather than controlling CV with MV or DV alone. As a result, by fixing the hyper parameter in the dynamic interval with the minimum MAE, the estimated APC model parameter was measured for the fitting rate of the predicted value and the actual value through the APC Model Tool, and the fitting rate was found to be 86.09% on average for Plant A and 79.94% on average for Plant B. Therefore, it is possible to estimate the APC model parameter with good control performance without plant test. In the future, it is necessary to increase the reliability of the results through extended experiments with more process data. In this experiment, only MAE was used as an evaluation metric, but it is necessary to expand evaluation metrics such as MSE (Mean Squared Error), RMSE (Root Mean Squared Error), and MAPE (Mean Absolute Percentage Error) in addition to MAE to improve the reliability of evaluation metrics. Also, it is necessary to analyze how to reduce the MAE of each variable through approaches other than PELT-based Learning, Linear Kernel-based Learning, and Radial basis function Kernel-based Learning applied in this experiment.

Author Contributions: Conceptualization, Y.Y. and M.L.; validation, Y.Y., M.L. and J.J.; formal analysis, Y.C. and S.B.; investigation, Y.C., S.B., Y.K., H.J., D.L. and K.K.; methodology, Y.Y.; software, Y.Y. and M.L.; data curation, J.J.; original draft preparation, Y.Y.; review and editing, Y.Y., M.L. and J.J.; funding acquisition, J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by Sungkyunkwan University and Infotrol Technology Co.

Data Availability Statement: The data used to support the findings of this study were provided by the corresponding author upon request (jpjeong@skku.edu).

Acknowledgments: This study was supported by Sungkyunkwan University and Infotrol Technology Co.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. AFRAM, Abdul; JANABI-SHARIFI, Farrokh. Theory and applications of HVAC control systems–A review of model predictive control (MPC). *Building and Environment*, **2014**, *72:* 343-355.
2. AHN, Gilseung, et al. Development of indicator of data sufficiency for feature-based early time series classification with applications of bearing fault diagnosis. *Processes*, **2020**, *8.7:* 790.
3. AMINIKHANGHAHI, Samaneh; COOK, Diane J. A survey of methods for time series change point detection. *Knowledge and information systems*, **2017**, *51.2:* 339-367.
4. ANG, Kiam Heong; CHONG, Gregory; LI, Yun. PID control system analysis, design, and technology. *IEEE transactions on control systems technology*, **2005**, *13.4:* 559-576.
5. ARLOT, Sylvain; CELISSE, Alain; HARCHAOUI, Zaid. A kernel multiple change-point algorithm via model selection. *Journal of machine learning research*, **2019**, *20*. 162.
6. BORASE, Rakesh P., et al. A review of PID control, tuning methods and applications. *International Journal of Dynamics and Control*, **2021**, *9:* 818-827.
7. CARTER, A.; IMTIAZ, S.; NATERER, G. F. Review of Interpretable Machine Learning for Process Industries. *Process Safety and Environmental Protection*, **2022**
8. CAO, Liang, et al. Soft Sensor Change Point Detection and Root Causal Analysis. In:*2022 61st Annual Conference of the Society of Instrument and Control Engineers (SICE)*, **IEEE 2022**. p, 242-247.
9. CHEVALIER, Michael, et al. Design and analysis of a proportional-integral-derivative controller with biological molecules. *Cell systems*, **2019**, *9.4:* 338-353. e10.
10. DEIVACHILAI, Rakesh; OATES, Tim. On-Line Signature Verification Using Symbolic Aggregate Approximation (SAX) and Sequential Mining Optimization (SMO). In: *2014 13th International Conference on Machine Learning and Applications*, **IEEE, 2014**. p. 195-200.
11. FEDELE, Giuseppe. A new method to estimate a first-order plus time delay model from step response. *Journal of the Franklin Institute*, **2009**, *346.1:* 1-9.
12. FURAT, Murat; EKER, İlyas. Computer-aided experimental modeling of a real system using graphical analysis of a step response data. *Computer Applications in Engineering Education*, **2014**, *22.4:* 571-582.
13. GAVIN, Henri P. The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems. *Department of Civil and Environmental Engineering, Duke University*, **2019**, *19*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2023 doi:10.20944/preprints202306.1026.v1

14. HA, Dae-Geun, et al. APC Technique and Fault Detection and Classification System in Semiconductor Manufacturing Process. *Journal of Institute of Control, Robotics and Systems*, **2015**, *21.9:*, 875-880
15. HARCHAOUI, Zaid; CAPPÉ, Olivier. Retrospective mutiple change-point estimation with kernels. In: *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, **IEEE, 2007**. p. 768-772.
16. HOLKAR, K. S.; WAGHMARE, Laxman M. An overview of model predictive control. *International Journal of control and automation*, **2010**, *3.4:* 47-63.
17. JIANG, Feng; ZHU, Qiannan; TIAN, Tianhai. An ensemble interval prediction model with change point detection and interval perturbation-based adjustment strategy: A case study of air quality. *Expert Systems with Applications*, **2023**, *222:119823.*
18. KELLY, Jeffrey D.; HEDENGREN, John D. A steady-state detection (SSD) algorithm to detect non-stationary drifts in processes. *Journal of Process Control*, **2013**, *23.3:* 326-331.
19. KILLICK, Rebecca; FEARNHEAD, Paul; ECKLEY, Idris A. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, **2012**, *107.500:* 1590-1598.
20. KIM, Minsung, et al. Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner. *International journal of refrigeration*, **2008**, *31.5: 790-799.*
21. LOU, Zhijiang; SHEN, Dong; WANG, Youqing. Two-step principal component analysis for dynamic processes monitoring. *The Canadian Journal of Chemical Engineering*, **2018**, *96.1:160-170.*
22. MESBAH, Ali. Stochastic model predictive control: An overview and perspectives for future research. *IEEE Control Systems Magazine*, **2016**, *36.6: 30-44.*
23. NGUYEN-TRUONG, Hieu T.; LE, Hung M. An implementation of the Levenberg–Marquardt algorithm for simultaneous-energy-gradient fitting using two-layer feed-forward neural networks. *Chemical Physics Letters*, **2015**, *629: 40-45.*
24. QIN, S. Joe; BADGWELL, Thomas A. A survey of industrial model predictive control technology. *Control engineering practice*, **2003**, *11.7: 733-764.*
25. ROMERO-GÜIZA, M. S., et al. Enhancement of sewage sludge thickening and energy self-sufficiency with advanced process control tools in a full-scale wastewater treatment plant. *Water Research*, **2022**, *222: 118924.*
26. RUDER, Sebastian. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, **2016**.
27. SERALE, Gianluca, et al. Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities. *Energies*, **2018**, *11.3: 631.*
28. SHA'ABAN, Yusuf Abubakar; LENNOX, Barry; LAURÍ, David. PID versus MPC performance for siso dead-time dominant processes. *IFAC Proceedings Volumes*, **2013**, *46.32: 241-246.*
29. SUMAYLI, Abdulrahman. Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models. *Arabian Journal of Chemistry*, **2023**, *16.7: 104833.*
30. TRUONG, Charles; OUDRE, Laurent; VAYATIS, Nicolas. Selective review of offline change point detection methods. *Signal Processing*, **2020**, *167: 107299.*
31. WANG, Qing-Guo; GUO, Xin; ZHANG, Yong. Direct identification of continuous time delay systems from step responses. *Journal of Process Control*, **2001**, *11.5: 531-542.*
32. WU, Jianguo, et al. Online steady-state detection for process control using multiple change-point models and particle filters. *IEEE Transactions on Automation Science and Engineering* **2015**, *13.2: 688-700.*
33. XIE, Sen, et al. A data-driven adaptive multivariate steady state detection strategy for the evaporation process of the sodium aluminate solution. *Journal of Process Control* **2018**, *68: 145-159.*
34. XU, He, et al. An RFID indoor positioning algorithm based on Bayesian probability and K-nearest neighbor. *Sensors* **2017**, *17.8: 1806.*
35. XU, Xiaojie. Short-run price forecast performance of individual and composite models for 496 corn cash markets. *Journal of Applied Statistics* **2017**, *44.14: 2593-2620.*
36. XU, Xiaojie; ZHANG, Yun. Corn cash price forecasting with neural networks. *Computers and Electronics in Agriculture* **2021**, *184: 106120.*
37. YAN, Qinglong, et al. Automatic identification of near-stationary traffic states based on the PELT changepoint detection. *Transportation Research Part B: Methodological* **2018**, *108: 39-54.*
38. ZHANG, Nan, et al. Sustaining high energy efficiency in existing processes with advanced process integration technology. *Applied Energy* **2013**, *101: 26-32.*
39. 안성모; 서승택. [지상강좌] 화학 공정의 최적화 솔루션 소개. *News Information for Chemical Engineers* **2012**, *30.1: 53-62.*