

Estimation of Air Pollutant Transportation Equation in Surabaya using Kalman Filter Method

Didik Khusnul Arif, Helisyah Nur Fadhilah and Prima Aditya

Abstract—Surabaya is one of big cities in Indonesia. Since the number of industries in Surabaya is increasing, the level of air pollution in Surabaya is also increasing. To deal with this matter, procurement of measuring instruments has been carried out in several locations in Surabaya based on the level of emissions issued by motorized vehicles, but increasing the number of measuring instruments affects the cost significantly. Therefore, the estimation of air pollution will make it easier to determine the level of air pollution in Indonesia. Kalman filter a method to estimate the state variable by using a system model accompanied by a measurement model as the initial value that comprises the prediction stage and the correction stage. The results of this correction stage will be the estimation results. Then it will be compared with the data at some locations. The results obtained are quite accurate at the point of observation with a relatively small error.

Index Terms—Air pollutant, Kalman filter, state estimation.

I. INTRODUCTION

SURABAYA is the second largest city after Jakarta in Indonesia. Industrial development in Surabaya is increasing every year. The use of motorized vehicles also increases over time. Without realizing it, these things contribute a lot of air pollution in the city. Data from the Ministry of Environment shows that the pollution index in Surabaya in 2016 reached 89.57, up from the previous year. Air pollution or what is usually indicated by the PM10 symbol is caused by air emissions coming out from motorized vehicles. As a result of air pollution, this is very detrimental to living things, in breeding or growing.

Efforts have been made to prevent air pollution or mitigate it. What can be done by the Surabaya government is to place a measuring device at several observation points for an indication of the level of air pollution (PM10) in these places. From the observational data, it shows which locations have high concentrations, densely populated areas, around pollutant locations, projections, areas according to the control strategy, and overall observation sites. Based on the ISPU data recap from BLH in 2015, within a period of 4 years it was noted that based on the results of monitoring air quality, the data produced were not sufficient to describe Surabaya's air quality because it only had 3 active monitoring stations [1]. The impact of inadequate data is inaccurate information for the general public and environmental management policy making. In improving the availability of information on air quality data that is more accurate, technically and operationally easier, it

is necessary to provide alternative steps to provide good data, one of which is by determining the correct observation points and the minimum number of observers but able to represent the air quality of Surabaya.

Air pollution problems are represented in the form of mathematical models where air pollution is assumed to be a system. The system is formed from an air pollution model approach (Gaussian model, advection-diffusion model, box model). The system is a large-order system. The system is a combination of several components that work together to achieve a specific goal and the order is the number of positions specified. In this study, the model that has been obtained will be estimated by data assimilation techniques. Data assimilation is a combination of system models with measurement data. Unlike the statistical method, assimilation of data does not require a lot of data to draw conclusions from the results of the analysis.

One familiar method of data assimilation is the Kalman filter. Kalman filter was introduced by R.E. Kalman in the 1960s which the point was uses a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone, by estimating a joint probability distribution over the variables for each time-frame. Kalman filters are widely used in the world of applied mathematics such as controls on aircraft, controls on the motion of cars, etc. The algorithm works in a two-step process. In the prediction step, the Kalman filter produces estimates of the current state variables, along with their uncertainties. Once the outcome of the next measurement (necessarily corrupted with some amount of error, including random noise) is observed, these estimates are updated using a weighted average, with more weight being given to estimates with higher certainty. The algorithm is recursive. It can run in real time, using only the present input measurements and the previously calculated state and its uncertainty matrix; no additional past information is required. In its application in the case of air pollution, Kalman filter works well at the observation point with a small error while when it does not have data it will cause uncertainty because it cannot be compared with the original estimation [2]. Kalman filter and its extension have been widely developed in various problems such as estimation of radar tracking [3], in model reduction [4], estimation of state variable of stirred tank reactor [5], estimation of floodgate control [6], and estimation of multi-robot motion [7].

Manuscript received January 8, 2019; accepted November 28, 2019.

The authors are with the Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia. E-mail: didik@matematika.its.ac.id

II. MATHEMATICAL MODEL OF POLLUTANT TRANSPORTATION EQUATION IN SURABAYA

Estimation of the concentration of pollutants at certain time and location can be done by using a pollutant dispersion model. The spreading process of pollutants (carbon monoxide) in the air can be approached by the advection-diffusion model. Advection is a process of pushing pollutants because of their existence media movements. In this case, air movement is due to differences in pressure, while the diffusion process is the spread of pollutants because of their difference in concentration between one point and another around it. The model can be an analytical, statistical or numerical model. Each model has advantages and disadvantages, so for each problem, we need to choose the the most appropriate model. Determination of the type of model used depends on several things, including: modeling objectives, space scale, time scale, and available costs. In numerical modeling, there are two methods that can be used, namely finite element method and finite difference method. Difference method is selected for making the pollutant advection-diffusion models. Numerical finite difference method is one method that is widely used in solving engineering problems because it is easy to use.

A. Pollutant Transportation Equation

Advection-diffusion equation or pollutant transport equation is a solution to find out the concentration of pollutants in a particular area with a velocity profile and certain wind direction [8]:

$$\frac{\partial C}{\partial t} = -v_x \frac{\partial C}{\partial x} - v_y \frac{\partial C}{\partial y} + D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + Q. \quad (1)$$

Using finite difference method approach to first-order differential terms:

$$\frac{\partial C}{\partial x} \cong \frac{C_{i+1,j} - C_{i,j}}{h} \quad \text{and} \quad \frac{\partial C}{\partial y} \cong \frac{C_{i,j+1} - C_{i,j}}{k} \quad (2)$$

$$\frac{\partial C}{\partial t} \cong \frac{C_{i,j}^{m+1} - C_{i,j}^m}{\Delta t} \quad (3)$$

while the second-order differential term is approximated by:

$$\frac{\partial^2 C}{\partial x^2} \cong \frac{C_{i+1,j} - 2C_{i,j} + C_{i,j-1}}{h^2} \quad \text{and} \quad \frac{\partial^2 C}{\partial y^2} \cong \frac{C_{i,j+1} - 2C_{i,j} + C_{i-1,j}}{k^2} \quad (4)$$

by applying a finite difference approach to the first-order and second-order differential terms above, (1) changes to

$$\begin{aligned} C_{i,j}^{m+1} - C_{i,j}^m &= -v_x \frac{\Delta t}{h} [C_{i+1,j}^m - C_{i,j}^m] - v_y \frac{\Delta t}{k} [C_{i,j+1}^m - C_{i,j}^m] + \\ &D_x \frac{\Delta t}{h^2} [C_{i+1,j}^m - 2C_{i,j}^m + C_{i-1,j}^m] \\ &+ D_y \frac{\Delta t}{k^2} [C_{i,j+1}^m - 2C_{i,j}^m + C_{i,j-1}^m] + Q_{i,j} \Delta t \end{aligned} \quad (5)$$

If we assumed the grid width used is the same, so ($h = k$), then the equation becomes

$$C_{i,j}^{m+1} - C_{i,j}^m = \frac{\Delta t}{h} \{ -v_x [C_{i+1,j}^m - C_{i,j}^m] - v_y [C_{i,j+1}^m - C_{i,j}^m] +$$

$$\begin{aligned} &D \frac{\Delta t}{h^2} [C_{i+1,j}^m + C_{i,j+1}^m - 4C_{i,j}^m + C_{i-1,j}^m \\ &+ C_{i,j-1}^m] + Q_{i,j} \Delta t \end{aligned} \quad (6)$$

reset the arrangement of the equations to

$$\begin{aligned} C_{i,j}^{m+1} &= C_{i,j}^m - v_x \frac{\Delta t}{h} [C_{i+1,j}^m - C_{i,j}^m] - v_y \frac{\Delta t}{k} [C_{i,j+1}^m - C_{i,j}^m] \\ &+ D_x \frac{\Delta t}{h^2} [C_{i+1,j}^m - 2C_{i,j}^m + C_{i-1,j}^m] \\ &+ D_y \frac{\Delta t}{k^2} [C_{i,j+1}^m - 2C_{i,j}^m + C_{i,j-1}^m] + Q_{i,j} \Delta t \end{aligned} \quad (7)$$

if

$$a = v_x \frac{\Delta t}{h}, \quad b = v_y \frac{\Delta t}{k}, \quad c = D_x \frac{\Delta t}{h^2}, \quad d = D_y \frac{\Delta t}{k^2} \quad (8)$$

Equation becomes

$$\begin{aligned} C_{i,j}^{m+1} &= C_{i,j}^m - a[C_{i+1,j}^m - C_{i,j}^m] - b[C_{i,j+1}^m - C_{i,j}^m] + \frac{\Delta t}{h^2} \\ &c[C_{i+1,j}^m - 2C_{i,j}^m + C_{i-1,j}^m] \\ &+ d[C_{i,j+1}^m - 2C_{i,j}^m + C_{i,j-1}^m] + Q_{i,j} \Delta t \end{aligned} \quad (9)$$

can be rearranged into

$$\begin{aligned} C_{i,j}^{m+1} &= [1 + a + b - 2c - 2d]C_{i,j}^m + (c - a)C_{i+1,j}^m \\ &+ (d - b)C_{i,j+1}^m + aC_{i-1,j}^m + bC_{i,j-1}^m \\ &+ Q_{i,j} \Delta t \end{aligned} \quad (10)$$

if

$$e = c - a; \quad f = d - b; \quad g = (1 + a + b - 2c - 2d) \quad (11)$$

then the equation becomes

$$\begin{aligned} C_{i,j}^{m+1} &= gC_{i,j}^m + eC_{i+1,j}^m + fC_{i,j+1}^m + aC_{i-1,j}^m + bC_{i,j-1}^m \\ &+ Q_{i,j} \Delta t \end{aligned} \quad (12)$$

If we run the index i and j then we obtain:

$$\begin{bmatrix} g & e & 0 & \dots & 0 & f & 0 & 0 & 0 & 0 \\ a & g & e & 0 & 0 & 0 & f & 0 & \dots & 0 \\ 0 & a & g & e & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & a & g & e & 0 & 0 & 0 & f & 0 \\ 0 & 0 & 0 & \dots & \dots & \dots & 0 & 0 & 0 & f \\ b & 0 & 0 & 0 & \dots & \dots & \dots & 0 & 0 & 0 \\ 0 & b & 0 & 0 & 0 & \dots & \dots & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & a & g & e & 0 \\ 0 & \dots & 0 & b & 0 & 0 & 0 & a & g & e \\ 0 & 0 & 0 & 0 & b & 0 & 0 & 0 & a & g \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ C_n \end{bmatrix}^m = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ C_n \end{bmatrix}^{m+1}$$

Or the equation above can be described as model system with replacing m with k which denotes a time index, as follows:

$$A x_k = x_{k+1} \quad (13)$$

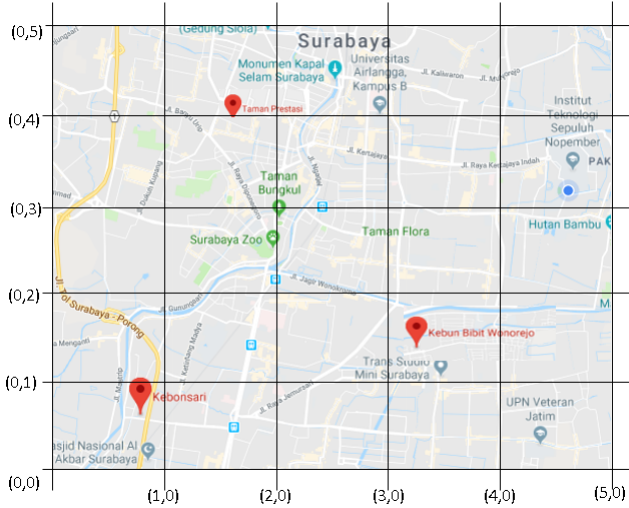


Fig. 1: Grid map of Surabaya city

B. Pollutant Transportation in Surabaya

Discretization in this discussion aims to obtain information/data at a certain time and distance so that the grid in Surabaya becomes a number of points as in Fig. 1 to facilitate the analysis process. In Fig. 1, the city map of Surabaya on the grid becomes 36 points with the determination of $N = M = 5$. These points represent the density of PM10 pollutants which are then needed in the analysis of determining the number of air quality observation points. Modeling of the air pollution system is approached by a 2D diffusion advection model according to (1) then discretization using the following finite difference method [3]. Based on data from the Surabaya City Environmental Service in this study, we set $\Delta t = 1800$ seconds, $\Delta x = 2700$ meter and $\Delta y = 2500$ meter. The wind speed and diffusion coefficient are given in Table I.

So, by running the index and input the parameters to the equation of pollutant transportation, we obtain matrix A , B , C , and D . Matrix A is of size 25×25 , B is of size 25×20 , C is of size 25×25 . Matrix B shows that the system input is 20 in accordance with the input variable u_k and also suitable with matrix A . The measurement matrix C is determined to indicate the observation points to be installed on the measuring instrument and the point is assumed to be valuable and if in this discussion the observation point is not a measuring instrument, meaning there is no measurement data in the position, then it is 0. In this research, we determine 18 observation points from 25 points of variable states. Information from 18 observation points can provide output y that is information about the concentration of PM10 pollutants at 25 points in the Surabaya area at the next time instant.

III. KALMAN FILTER ALGORITHM

Kalman filter estimates one process through a feedback control mechanism: the filter estimates the state of the process and then gets feedback in the form of a noisy measurement. The equation for Kalman filters is grouped in two parts: time update equations and measurement update equations. The time update equation is assigned to get the pre-estimated value for

TABLE I: Diffusion coefficient and wind speed

(i, j)	D_x	D_y	v_y
(1,1)	2296.5424	3479.2617	0.0000
(1,2)	2245.7357	3351.7605	0.0000
(1,3)	2313.3385	3522.0579	0.0000
(1,4)	2445.2976	3869.6835	0.0000
(1,5)	2660.4993	4481.6110	0.9300
(2,1)	1821.8245	2396.6101	0.0800
(2,2)	1756.7835	2266.2508	0.5050
(2,3)	1833.1890	2419.8094	0.5050
(2,4)	2046.6221	2879.5973	1.0050
(2,5)	2330.0661	3565.0011	0.0000
(3,1)	1323.7715	1497.1855	0.0800
(3,2)	1756.7835	2266.2508	0.0800
(3,3)	1341.0658	1524.7918	0.6970
(3,4)	1659.0770	2077.9940	1.0050
(3,5)	2035.8624	2855.2971	1.0050
(4,1)	817.0584	790.5040	0.0800
(4,2)	530.2510	469.2721	0.0800
(4,3)	861.7808	845.4070	0.5800
(4,4)	1362.5688	1559.4600	1.0800
(4,5)	1619.2778	2003.8563	0.0000
(5,1)	663.6289	612.1976	0.0000
(5,2)	204.5575	163.5437	0.0800
(5,3)	730.9625	688.5667	1.0800
(5,4)	1306.3945	1469.6938	1.0800
(5,5)	1795.1762	2342.7049	0.0000

TABLE II: Dimensions of discrete time system variables

Variable	Description	Dimension
x	state vector	$n_x \times 1$
y	output vector	$n_y \times 1$
u	input vector	$n_u \times 1$
w	process noise vector	$n_x \times 1$
v	measurement noise vector	$n_y \times 1$
A	system matrix - state	$n_x \times n_x$
G	system matrix - input	$n_x \times n_u$
H	observation matrix	$n_y \times n_x$

the next time instant. The measurement update equation is responsible for feedback needs, such as integrating the latest measurement results with pre-estimated values to obtain a better post-estimation value. The time update equation is also called the prediction equation, while the measurement update equation is called the correction equation. The Kalman estimation algorithm filters resemble predictive-correction algorithms to solve numerical problems. Kalman filter is designed to operate on systems in linear state space format i.e. [9]

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + w_k \\ y_k &= H_k x_k + v_k \end{aligned}$$

where the variable definitions and dimensions are detailed in Table II.

The state vector x are the values that will be estimated by the filter, in this case x means pollutant level in Surabaya. Kalman filter uses a prediction then continued by a correction in order

to determine states of the filter. This is called predicted-updated. The main idea is that using information about the dynamics of the state, the filter will project forward and predict what the next state will be. Starting from some initial estimation of state, \hat{x}_0 , and initial state error covariance matrix, P_0 , the predicted-updated format is applied recursively at each time step, e.g. using a loop. First, the state vector is predicted from state dynamic equation using

$$\hat{x}_{k+1|k} = A_k \hat{x}_k + B_k u_k \quad (14)$$

where $\hat{x}_{k+1|k}$ is the predicted state vector, \hat{x}_k is the previous estimated state vector. Next, the state error covariance matrix must also be predicted using

$$P_{k+1|k} = A_k P_k A_k^T + Q \quad (15)$$

where $P_{k+1|k}$ represents the predicted state error covariance matrix, P_k is the previous estimated state error covariance matrix, and Q is the process noise covariance matrix. Once the predicted values are obtained, the Kalman gain matrix K_{k+1} is calculated by

$$K_{k+1} = P_{k+1|k} H_k^T (H_k P_{k+1|k} H_k^T + R_k)^{-1} \quad (16)$$

where H is a matrix necessary to define the output equation and R is the measurement noise covariance. The state vector is then updated which is the difference between the measurement of the output z_k , and the predicted output $H_k \hat{x}_{k+1|k}$, by calculated Kalman gain matrix in order to correct the prediction by the appropriate amount, as in

$$\hat{x}_{k+1} = \hat{x}_{k+1|k} + K_{k+1} (z_{k+1} - H_k \hat{x}_{k+1|k}) \quad (17)$$

Similarly, the state error covariance is updated by

$$P_{k+1} = (I - K_{k+1} H_k) P_{k+1|k} \quad (18)$$

where I is an identity matrix.

IV. NUMERICAL RESULTS

Simulation of pollutant transportation Equation on (1) can be presented in state-space form as on 13. By using Kalman filter, the estimation results are obtained at several points around the known point of the real data. Based on Fig. 1, it can be seen that the point or area that has real data is the area with the red mark, Kebonsari, Wonorejo, and Taman Prestasi (three locations of Surabaya grid). The data used in the simulation is data from the Department of Environment of the city of Surabaya on February 28, 2018. In this paper, we use the hourly data.

From the real data, estimates can be made around the area based on the coordinates in Fig. 1. Estimated results for the coordinates (1,3), (1,4), (2,3), and (3,2) are presented in Fig. 2-5.

Based on Fig. 2-5, it can be seen that the estimation of pollutant distribution in February 2018 has the same characteristics as real data.

The absolute error of the estimated pollutant distribution with Kalman Filter on real data can be formulated as follows

$$Error = |x_{est} - x_{real}| \quad (19)$$

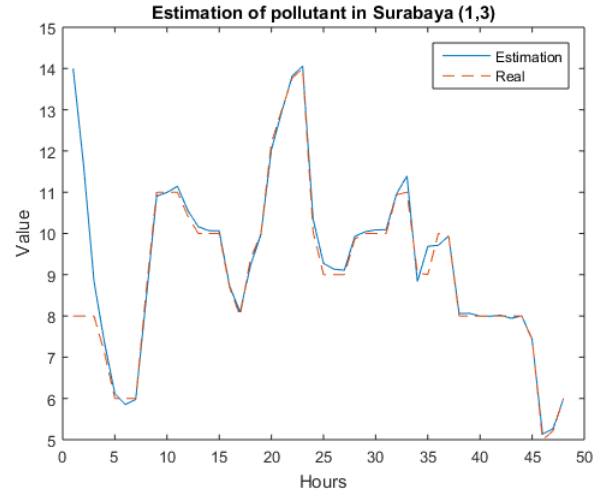


Fig. 2: Estimation of pollutant in coordinate (1,3) or in Kebonsari.

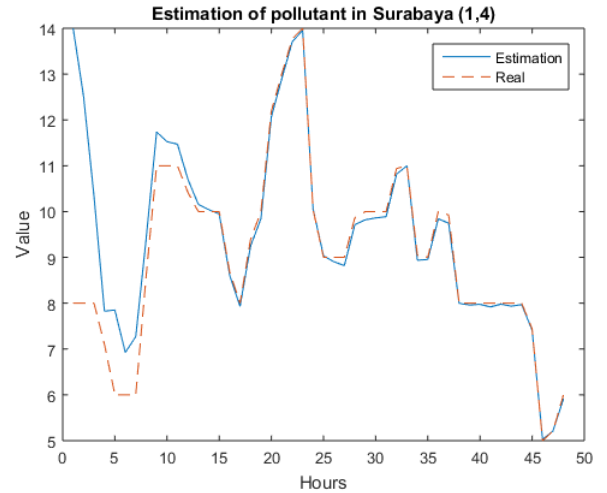


Fig. 3: Estimation of pollutant in coordinate (1,4) or in Taman Prestasi.

Absolute error can be only calculated on points that have real data. In this case, the points are (1,3) and (3,2) which is Kebonsari and Wonorejo. From (19), we can calculate value of estimation error of pollutant distribution using Kalman filter on (1,3) as 0.93631 and on point (3,2) is 0.146192. At the observation point (1,4) namely Taman Prestasi and (2,3) in Taman Bungkul, we did not have the data so the error was still quite large above 1 so we did not display it in the paper.

V. CONCLUSIONS

From the experiments that have been done, it can be concluded that the mathematical model of pollutant equations can be applied in Surabaya by dividing it into several observation grids. The observational data held determines the accuracy of the estimation using the Kalman filter. Kalman filters work well or accurately at the point that has observational data while working less optimally at points that do not have observational

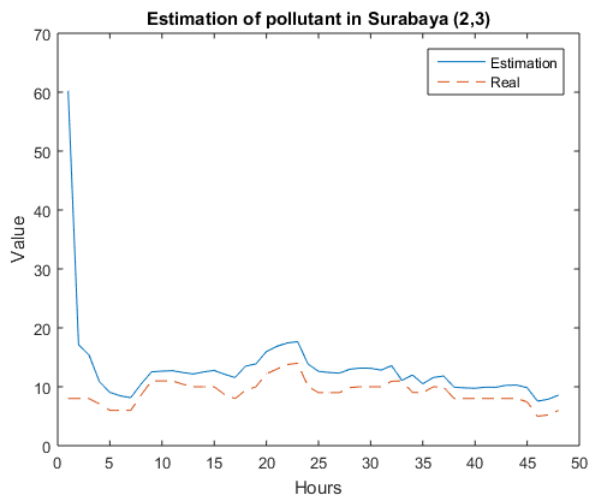


Fig. 4: Estimation of pollutant in coordinate (2,3) or in Taman Bungkul.

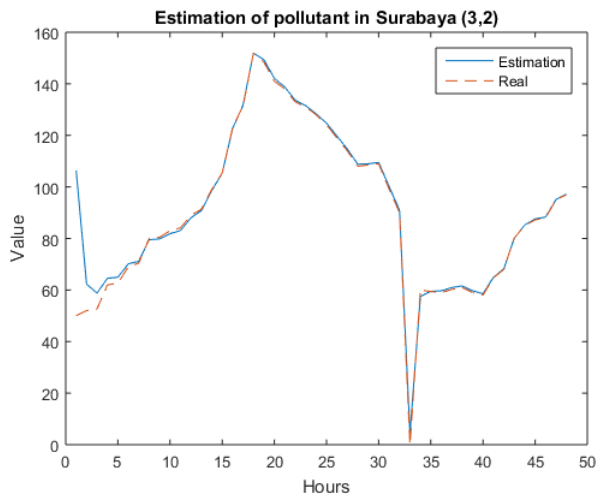


Fig. 5: Estimation of pollutant in coordinate (3,2) or in Wonorejo.

data. Error between estimation and original data generated is between 0.146192 up to 0.963631.

REFERENCES

- [1] K. Aisyiah, S. Sutikno, and I. Latra, "Pemodelan konsentrasi partikel debu (pm10) pada pencemaran udara di kota surabaya dengan metode geographically-temporally weighted regression," *Jurnal Sains dan Seni ITS*, vol. 3, no. 2, pp. D152–D157, 2014.
- [2] Y. Oktaviana, E. Apriliani, and D. Arif, "Fractional kalman filter to estimate the concentration of air pollution," in *Journal of Physics: Conference Series*, vol. 1008, no. 1, 2018, p. 012008.
- [3] D. Arif, D. Adzkiya, P. Aditya, F. Winata, D. Agustin, M. Habibi, A. Ririsati, and R. Prasyanto, "Modeling of three-dimensional radar tracking system and its estimation using extended kalman filter," in *5th International Conference on Instrumentation, Control, and Automation (ICA)*, 2017, pp. 51–55.
- [4] V. Rachmawati, D. Arif, and D. Adzkiya, "Implementation of kalman filter algorithm on models reduced using singular perturbation approximation method and its application to measurement of water level," in *Journal of Physics: Conference Series*, vol. 974, no. 1, 2018, p. 012018.

- [5] R. Fitria and D. Arif, "State variable estimation of nonisothermal continuous stirred tank reactor using fuzzy kalman filter," *International Journal of Computing Science and Applied Mathematics*, vol. 3, no. 1, pp. 16–20, 2017.
- [6] E. Purnaningrum and E. Apriliani, "Auto floodgate control using enk-mpc method," *International Journal of Computing Science and Applied Mathematics*, vol. 2, no. 1, pp. 14–19, 2016.
- [7] P. Aditya, E. Apriliani, G. Zhai, and D. Arif, "Formation control of multi-robot motion systems and state estimation using extended kalman filter," in *International Conference on Electrical Engineering and Informatics (ICEEI)*, 2019, pp. 99–104.
- [8] D. Arif, D. Adzkiya, E. Apriliani, and I. Khasanah, "Model reduction of non-minimal discrete-time linear-time-invariant systems," *Malaysian Journal of Mathematical Sciences*, vol. 11, no. 3, pp. 377–391, 2017.
- [9] D. Arif, Widodo, Salmah, and E. Apriliani, "Construction of the kalman filter algorithm on the model reduction," *International Journal of control and Automation*, vol. 7, no. 9, pp. 257–270, 2014.