Fractional kalman filter to estimate the concentration of air pollution

Yessy Vita Oktaviana¹, Erna Apriliani¹, Didik Khusnul Arif¹
¹Department of Mathematics, Faculty of Mathematics, Computing, and Data Science, Sepuluh Nopember Institute of Technology, Indonesia

Email: yessyvitaoktaviana@gmail.com

Abstract. Air pollution problem gives important effect in quality environment and quality of human’s life. Air pollution can be caused by nature sources or human activities. Pollutant for example Ozone, a harmful gas formed by $NO_x$ and volatile organic compounds (VOCs) emitted from various sources. The air pollution problem can be modeled by TAPM-CTM (The Air Pollution Model with Chemical Transport Model). The model shows concentration of pollutant in the air. Therefore, it is important to estimate concentration of air pollutant. Estimation method can be used for forecast pollutant concentration in future and keep stability of air quality. In this research, an algorithm is developed, based on Fractional Kalman Filter to solve the model of air pollution’s problem. The model will be discretized first and then it will be estimated by the method. The result shows that estimation of Fractional Kalman Filter has better accuracy than estimation of Kalman Filter. The accuracy was tested by applying RMSE (Root Mean Square Error).

1. Introduction

Nowadays, many cities and their surrounding areas are facing serious air pollution problem due to the increasing of human’s population, motor vehicles, and industries. As the result, the number of cities with poor environment has grown. Moreover, air pollution problem which gives important effect on the quality of environment, further affects the process of human’s life.

Air pollution is the presence of one or more physical, chemical, or biological substances in the atmosphere in quantities that could harmful for human, animal and plant health and impair aesthetics and comfort, or damage property. Air pollution can be caused by natural sources or human activities. Source pollutants that pollute the air in the form of gas and smoke. The gas and smoke are derived from the incomplete combustion process of fuel produced by factory machinery, power plants and motor vehicles.

Many researchers observed pollutant gases such as Ozone ($O_3$), Nitrogen Oxide ($NO$), Nitrogen Dioxide ($NO_2$), Sulfur Dioxide ($SO_2$), Particulate Matter ($PM_{2.5}$ & $PM_{10}$) to determine the air quality in an area as well controlling the air quality stability. The research process is carried out by forming air quality modeling and subsequently formed a control device to control or reduce the levels of pollutant substances in the air. Because of the high cost of the control device and the amount of maintenance costs it can not be placed as much as possible a tool to measure the concentration of pollutants. Therefore, estimation of pollutant concentration is necessary in addition to predicting future pollutant...
concentrations as well as to estimate pollutant concentrations in areas not including the radius of the measuring instrument.

In 2010, Erna Apriliani has conducted research on “The Square Root Ensemble Kalman Filter to Estimate the Concentration of Air Pollution”. In that research, Square Ensemble Kalman Filter can be used to maintain the stability of computing time [2]. Furthermore, in 2013, Santanu Metia et al conducted an air pollution modeling study entitled “Environmental Time Series Analysis and Estimation with Extended Kalman Filtering”. In this study discussed estimation of air pollutant profile by using Extended Kalman Filter algorithm. The Kalman Filter Extended algorithm is used to estimate missing temporary data. It was found that the estimation result of Extended Kalman Filter method contributed significantly [7]. In 2014, Santanu developed his research on “Air Pollution Prediction Using Matern Function Based Extended Fractional Kalman Filtering”. In this study, the estimation results show an increase in air quality predictions in the observed region [8]. Moreover in 2014, Didik Khusnul Arif has also developed a research on Kalman Filtering entitled “Construction of the Kalman Filter Algorithm on the Model Reduction”. This research discussed about a construction of Kalman Filter algorithm on the reduced model. It aimed to obtain accurate estimation with short computing time on the reduced model [3]. In 2017, Risa Fitria and Didik Khusnul Arif had conducted research on “State Variable Estimation of Nonisothermal Continuous Stirred Tank Reactor using Fuzzy Kalman Filter”. Fuzzy Kalman Filter is a modification of Kalman Filter that combines with fuzzy theory. This method presented to estimate the state variable of Non-Isothermal CSTR. The estimation results shows that estimation using Fuzzy Kalman Filter had better accuracy than Ensemble Kalman Filter [4]. On the other side, in 2017 Prima Aditya and Didik Khusnul Arif also conducted a research about applied of Kalman Filter entitled “Modelling of Three-Dimensional Radar Tracking System and Its Estimation using Extended Kalman Filter”. In this study, Kalman Filter is developed into Extended Kalman Filter to estimate the measurement of nonliner model of three-dimensional radar tracking system. The results of the simulation show that the allegedly proposed formulation is very efective in the measurement calculation that is not linear [1].

Based on some previous research about Kalman Filter, the authors examine the problem of air pollution. The assessment was done by applying the modification of Kalman Filter namely Fractional Kalman Filter method using several different parameter values. So that we will get the estimated value approaching real system.

2. The Air Pollution Model with Chemical Transport Model (TAPM-CTM)

The TAPM-CTM air pollution model developed by Commonwealth Scientific and Industrial Research Organization (CSIRO). This model is a prognostic 3D equation to determine concentration of pollutant concentrations in the air [5].

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left( K_x \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left( K_y \frac{\partial C}{\partial y} \right) - \frac{\partial}{\partial z} \left( \frac{\partial C}{\partial z} \right) + \frac{\partial}{\partial \sigma} \left( \frac{\partial C}{\partial \sigma} \right) + \frac{\partial C}{\partial \sigma} \frac{\partial}{\partial \sigma} \left( \frac{wC'}{\sigma} \right) + S_C + R_C$$

with

$$\frac{wC'}{\sigma} = -K_\sigma \frac{\partial C}{\partial \sigma}$$

where $C$ is the concentration of pollutants, $K_x, K_y$, and $K_\sigma$ is diffusion coefficient in $x, y,$ and $\sigma$ direction, respectively, $S_C$ is pollutant emission, and $R_C$ is the effect of chemical reaction.

Suppose we assume the diffusion coefficient are constant, so that we can be rewritten the Eq. (1) as follows

$$\frac{\partial C}{\partial t} = K_x \frac{\partial^2 C}{\partial x^2} + K_y \frac{\partial^2 C}{\partial y^2} + K_\sigma \frac{\partial}{\partial z} \left( \frac{\partial C}{\partial \sigma} \right) + \frac{\partial C}{\partial \sigma} \frac{\partial}{\partial \sigma} \left( \frac{wC'}{\sigma} \right) + S_C + R_C$$

$$\frac{\partial C}{\partial t} = K_x \frac{\partial^2 C}{\partial x^2} + K_y \frac{\partial^2 C}{\partial y^2} + K_\sigma \left( \frac{\partial}{\partial \sigma} \right) \left( \frac{\partial^2 C}{\partial \sigma^2} + \frac{\partial C}{\partial \sigma} \frac{\partial}{\partial \sigma} \right) + S_C + R_C$$

(2)
3. Discretization of Model

Before we applied Fractional Kalman Filter to estimate the concentration of air pollution, we discretize Eq. (2) respect to space \( x, y, \sigma \) and time \( t \) and then we write in the state space form.

The mathematical model of the pollutant concentration model above is a dynamic system, so the time is continuous, while Fractional Kalman Filter method can only be implemented for the system with discrete time, so it must be discretized with change of state variable to time with finite difference method. To simplify, it is often written with index notation. The first and second subscript indexes as the third space and subscript variables as time variables. So, it can be written:

\[
c(x, y, \sigma, t) - c^n_{i,j,k} = c^n_{i,j,k-1} + c_n^n - c^n_{i,j,k} = c^n_{i,j,k} - c^n_{i,j,k-1}
\]

where \( t = n \Delta t \) and \( n = 0,1,2,3, \ldots \)

so the discretization in \( i, j, k \)

\[
c_t(x, y, \sigma, t) = \frac{c^n_{i,j,k} - c^n_{i,j}}{\Delta t}
\]

Forward difference is implemented to \( c_\sigma(x, y, \sigma, t) \) in \( i, j, k \) (at time \( n \)) is

\[
c_\sigma(x, y, \sigma, t) = \frac{c^n_{i,j,k+1} - c^n_{i,j,k}}{\Delta \sigma}
\]

On the same way, Center difference is implemented to \( c_{xx}, c_{yy}, \) and \( c_{\sigma \sigma} \) as follow:

\[
c_{xx}(x, y, \sigma, t) = \frac{c^n_{i+1,j,k} - 2c^n_{i,j,k} + c^n_{i-1,j,k}}{\Delta x^2}
\]

\[
c_{yy}(x, y, \sigma, t) = \frac{c^n_{i,j+1,k} - 2c^n_{i,j,k} + c^n_{i,j-1,k}}{\Delta y^2}
\]

\[
c_{\sigma \sigma}(x, y, \sigma, t) = \frac{c^n_{i,j,k+1} - 2c^n_{i,j,k} + c^n_{i,j,k-1}}{\Delta \sigma^2}
\]

Then the discretization will be implemented to the model (2), it can be:

\[
c^n_{i,j,k+1} = c^n_{i,j,k} + K_x \frac{\Delta t}{\Delta x^2} \left( c^n_{i+1,j,k} - 2c^n_{i,j,k} + c^n_{i-1,j,k} \right) + K_y \frac{\Delta t}{\Delta y^2} \left( c^n_{i,j+1,k} - 2c^n_{i,j,k} + c^n_{i,j-1,k} \right) + K_\sigma \frac{\Delta t}{\Delta \sigma} \left( c^n_{i,j,k+1} - 2c^n_{i,j,k} + c^n_{i,j,k-1} \right) + \Delta t S_C + \Delta t R_C
\]

The discretization in Eq. (3) can be written by using tridiagonal matrix as follow:

\[
\begin{pmatrix}
  c^n_{i,j,1} \\
  c^n_{i,j,2} \\
  \vdots \\
  c^n_{i,j,i,j+1,1} \\
  c^n_{i,j+1,i,j+1,2} \\
  c^n_{i,j,i,j-1,2} \\
  c^n_{i,j,i,j-1,1} \\
\end{pmatrix}
= \begin{pmatrix}
  F & B & D & E & 0 & A_1 & 0 & 0 & 0 \\
  B & F & 0 & D & E & 0 & A_1 & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  0 & D & 0 & B & F & 0 & 0 & 0 & 0 \\
  A_1 & 0 & D & 0 & B & 0 & \cdots & \cdots & \cdots \\
  0 & A_1 & 0 & 0 & B & F & \cdots & \cdots & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{pmatrix}
+ \begin{pmatrix}
  c^n_{i,j,1} \\
  c^n_{i,j,2} \\
  \vdots \\
  c^n_{i,j,i,j+1,1} \\
  c^n_{i,j+1,i,j+1,2} \\
  c^n_{i,j,i,j-1,2} \\
  c^n_{i,j,i,j-1,1} \\
\end{pmatrix}
+ \begin{pmatrix}
  \Delta t S_C \\
  \Delta t R_C \\
\end{pmatrix}
\]

Generally, The mathematic model in discrete time in Eq. (4) can be written as the discrete state space system of air pollution modelling as follow:

\[
x_{n+1} = Ax_n + Bu_n
\]

The TAPM-CTM model on air pollution problems is used to determine the concentration of a pollutant in the air pollution process. But in the process of spreading often found problems that contain elements of uncertainty or known probabilistic. The model is not exactly the same with the real system, we take some assumption that cannot be written in the model so that we write the stochastic system as

\[
x_{n+1} = Ax_n + Bu_n + w_n
\]

where \( x_{n+1} \) is the state variable in time \( n + 1 \), \( x_n \) is the state variable in time \( n \), \( u_n \) is system input, \( A \) is matrix whose its element is the coefficient of variable state \( x_n \) in model system, \( B \) is matrix whose its element is the coefficient of variable state \( u_n \), \( w_n \) in this case is system noise namely Gaussian white noise, respectively, \( w_n \sim N(0, Q_n) \).

To make correlation between the state which we will estimate and the measurement data, we define the measurement equation as follow:

\[
x_{n+1} = Ax_n + Bu_n + w_n
\]
\[ z_n = Hx_n + v_n \]  \hspace{1cm} (6)

where \( z_n \) is measurement data, \( H \) is measurement matrix, \( x_n \) state variable in time \( n \), and \( v_n \) measurement Gaussian white noise where \( v_n \sim N(0, R_n) \).

Noise system and noise measurements that are assumed normally distributed and zero mean values are generated through computer aids. In general the system noise variance is expressed by \( Q_n \) and the measurement noise variance is expressed by \( R_n \). Both depend on time and value assumed to be constant.

4. Fractional Kalman Filter Method

Suppose, we has a dynamic stochastic system Eq. (5) and measurement Eq. (6). By using Kalman filter we estimate the state variables of Eq. (1) based on the data measurement Eq. (6). Kalman filter is one of data assimilation method, because in Kalman filter, we combine the model of system Eq. (5) and the data measurement Eq. (6). Fractional Kalman filter is one of Kalman filter modifications.

Kalman Filter is used in network systems for fusion data or error detection in the system. But this method has limitations in estimating unknown state variables to find out the accuracy at a higher level. On the other hand the Fractional Kalman Filter algorithm has been used to estimate unknown state variables in complex systems where fractional derivatives can be used to describe good accuracy.

In the Kalman filter algorithm, we give just one vector for initial estimation \( x_0 \), and the covariance of prediction step is determined from the equation [6].

But in the Fractional Kalman Filter, we use a definition of fractional discrete derivative, Grünwald-Letnikov definition to get fractional order difference [9].

**Definition 1.** The fractional order Grünwald-Letnikov difference is given by the following equation

\[ \Delta^r x_n = \frac{1}{h^r} \sum_{s=0}^{n} (-1)^s \binom{r}{s} x_{n-s} \]  \hspace{1cm} (7)

where \( r \) is a fractional order and \( h \) is a sampling time later equal to 1, \( k \) is a number of sample for which the derivative is calculated. The factor \( \binom{r}{s} \) can be obtained from

\[ \binom{r}{s} = \begin{cases} 1 & \text{for } s = 0 \\ r(r-1)...(r-s+1)/s! & \text{for } s > 0 \end{cases} \]

According to this definition, it is possible to obtained a discrete equivalent of the derivative (when \( r \) positive), a discrete equivalent of integration (when \( r \) negative), or when \( r \) equals 0, the origin function [9].

Now the generalization of the discrete state space model for fractional first order can be obtained by using Definition 1.

\[ \Delta^1 x_{n+1} = x_{n+1} - x_n \]

Using formula stochastic system in Eq.(5) and Eq. (6), we can get the linier fractional first order stochastic discrete state-space system as follow:

\[ \Delta^1 x_{n+1} = Ax_n + Bu_n + w_n \]
\[ x_{n+1} = \Delta^1 x_{n+1} + x_n \]
\[ z_n = Hx_n + v_n \]

where \( \Delta^1 x_{n+1} \) is the first order difference for state \( x_{n+1}, u_n \) is a system input, \( w_n \) is a system noise, \( A = A - I \) and \( I \) is an identity matrix.

Variables \( w_n \sim N(0, Q) \) and \( v_k \sim N(0, R) \) are assumed white noise (normally distributed with mean 0), not correlated to each other nor with the initial estimation value \( x_0 \).

The algorithm of the Fractional Kalman Filter is built from Definition 1.

**First Step: Initialization**

Initial value :

\[ \hat{x}_0 = x_0 \]

covariance

\[ P_0 = E[(\hat{x}_0 - x_0)(\hat{x}_0 - x_0)^T] \]
Second Step: Prediction
We get prediction value from $x_n$ from the previous step by adding noise system $w_n$
\[ \hat{x}_{n+1} = A_d \hat{x}_n + Bu_n \]
with error covariance
\[ P_{n+1} = A_d P_n A_d^T + Q_n \]

Third Step: Correction
Kalman Gain:
\[ K_{n+1} = P_{n+1}^{-1} H^T (H P_{n+1}^{-1} H^T + R_{n+1})^{-1} \]
Estimation:
\[ \hat{x}_{n+1} = \hat{x}_{n+1}^- + K_{n+1} (z_{n+1} - H \hat{x}_{n+1}^-) \]
Error covariance:
\[ P_{n+1} = (I - K_{n+1} H) P_{n+1}^- (I - K_{n+1} H)^T + K_{n+1} R_{n+1} K_{n+1}^T \]

5. Main Result
Actually, it is difficult to get real data, the concentration of pollution, because the measurement tools are limited, therefore, we concern on the algorithm Kalman Filter, and the Fractional Kalman filter, so that the measurement data is generated from MATLAB program that represents Eq (6).

The simulation results will be evaluated by performing several simulations with many different iteration. And we can know the computing time of the program that has been run. The parameter value used diffusion coefficient which assumed constant with $x_0 = 0.02$, $K = 0.5$ and $K_x = K_y = K_\sigma = \min(10, K)$ and $\Delta t = 0.01$.

The simulation is done several times by changing the number of iterations to see the movement condition of the pollution dispersion graph. Noise system ($w_k$) and noise measurement ($v_k$) are searched for the most appropriate.

Here we compare the accuracy of estimation by using Kalman Filter, with by using Fractional Kalman Filter. The simulation results of the pollutant concentration estimate on the pollution process are shown in the following figure:

\[ \text{Figure 1. The concentration of pollutant when 50 iteration.} \]
Figure 1 shows real and estimation graphic of concentration pollutant with 50 iteration using Fractional Kalman filter first order. It shows the fluctuation of concentration pollutant. It shows that the concentration pollutant is smaller by the time, so it can give influence to the quality of the air. The graphic shows that the highest concentration happens at 10 iteration with the concentration value is about $0.03 - 0.035 \frac{kg}{m^3}$.

**Figure 2.** The concentration of pollutant when 100 iteration.

Figure 2 shows real and estimation graphic of concentration pollutant with 100 iteration. It shows the fluctuation of concentration pollutant. The graphic shows that distance between real value and estimation value is adequate small. The highest concentration happens at 10 iteration with the concentration value is about $0.04 - 0.048 \frac{kg}{m^3}$.

**Figure 3.** The concentration of pollutant when 200 iteration.

Figure 3 shows real and estimation graphic of concentration pollutant with 200 iteration using Fractional Kalman filter method. The graphic shows density because the iteration is more than before and the distance between real value and estimation value is adequate small.
Figure 4 shows the error value of concentration pollutant using Kalman Filter and Fractional Kalman Filter is small. But from many iteration we can know that error value of Fractional Kalman Filter is smaller than error value of Kalman Filter.

The accuracy of the method of each experiment is presented in the table as follows:

| RMSE         | KF            | Fractional KF | KF            | Fractional KF | KF            | Fractional KF |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|
| N=50         | 0.017289      | 0.0167837     | 0.0171377     | 0.164392      | 0.0170269     | 0.0161744     |
| N=100        | 0.839224      | 1.2665177     | 0.8826203     | 1.3403171     | 0.9632163     | 1.3830469     |
| N=200        | 1.2665177     | 0.8826203     | 1.3403171     | 0.9632163     | 1.3830469     |

It can be seen from Table 1 that the simulation is done with predetermined initial value parameters, and performed with different iterations to see the error values of each different iteration time. Table 1 is the RMSE average of 10 experiments. From the average of RMSE value, it can be seen that Fractional Kalman Filter first order is better than Kalman Filter where the average of RMSE value is smaller than Kalman Filter. For more iteration time, it shows that the average of error value generated smaller. But the computation time of Fractional Kalman Filter first order method tends longer than Kalman Filter method. This is because the Fractional Kalman Filter has a slightly longer step than the Kalman Filter.

6. Conclusions

Based on the analysis and discussion that has been done then can be drawn conclusion as follows:
1. The Fractional Kalman Filter Method can be applied in the estimation of state variables on the model of air pollution process.
2. Based on the estimation results, it is shown that the estimation using Fractional Kalman Filter has better accuracy than Kalman Filter.
3. The more iterations are done to see the error value of pollutant and computational time. It shows that for more iteration time, the average of error value is adequate smaller.

References

[1] Aleksandrovich KA, Mohan SH and Trujillo Juan J 2006 Theory and Applications of Fractional Differential Equations Elsevier 204
[2] Aditya P and Arif D K 2017 Modelling of Three-Dimensional Radar Tracking System and Its Estimation using Extended Kalman Filter Proceedings of the International Conference on Instrumentation, Control, and Automation

[3] Apriliani E, Arif D K and Sanjoyo B A 2010 The Square Root Ensemble Kalman Filter to Estimate the Concentration of Air Pollution International Conference on Mathematical Applications in Engineering (ICMAE’10), Kuala Lumpur Malaysia

[4] Arif D K 2014 Construction of the Kalman Filter Algorithm on the Model Reduction International Journal Control and Automation 7 9 257-270

[5] Fitria R and Arif D K 2017 State Variable Estimation of Nonisothermal Continuous Stirred Tank Reactor using Fuzzy Kalman Filter International Journal of Computing Science and Applied Mathematics 3 1 16-20

[6] Hurley PJ, Physick WL and Luhar AK 2005 TAPM: a practical approach to prognostic meteorological and air pollution modelling. Environmental Modelling & Software 20 6737-752

[7] Lewis F L Optimal estimation: with an introduction to stochastic control theory New York et al: Wiley; 1986 Apr

[8] Lewis JM, Lakshmivarahan S and Dhall S 2006 Dynamic data assimilation: a least squares approach Cambridge University Press

[9] Metia S, Oduro SD and HaQP 2013 Environmental time series analysis and estimation with extended kalman filtering IEEE First International Conference on Artificial Intelligence, Modelling & Simulation 202–207

[10] Metia S, Oduro SD and HaQP 2014 Air Pollution Prediction Using Matern Function Based Extended Fractional Kalman Filtering International Conferences and Control 758-763

[11] Sierociuk D and Dzieliński A 2006 Fractional Kalman Filter Algorithm for The States, Parameters, And Order of Fractional System Estimation International Journal Applied Mathematics Computation Science 16 1 129-140