

ABSTRACT

Missing values can be caused by broken observation tools or incomplete responses given by the observation subject (incomplete surveys data). One of the most common data that is prone to missing values is time series data. Time series is a sequence of data observed in a consistent interval of time. Missing values occurred in a consecutive manner for several points can affect the performances of an estimation method.

One of the developed and implemented methods to handle missing values on time series data is Centroid Decomposition (CD). The method is used to estimate the missing values that already interpolated by linear interpolation using the correlated series in which have the same length with the data that have missing values. The missing values themselves have several missing points and occur in a consecutive manner. This research is meant to find out the influence of smoothing method implementation on the correlated data. Hence, it can improve the estimation accuracy of CD method. This research uses Savitzky-Golay as a smoothing method to be applied on one time series data with several scenarios of evaluation.

The results of this research show that the implementation of smoothing method to the correlated data on the CD estimation process has a good influence in terms of accuracy improvement. The error measurement is done using Root Mean Square Error (RMSE). The evaluation result shows that the smoothed correlated data using Savitzky-Golay significantly has lower RMSE than unsmoothed correlated data on CD estimation process with a confidence level of 95% with an error reduction rate of 0.09% for 5 MV, 0.14% for 10 MV, and 0.08% for 20 MV.

Keywords : Missing values, Estimation, Centoroid Decomposition, Savitzky-Golay, Smoothing

INTISARI

Missing values disebabkan oleh kerusakan alat observasi atau ketidaklengkapan data observasi pada saat pengambilan data (subjek observasi tidak bersedia melengkapi data). Salah satu jenis data yang sering mengalami *missing values* adalah data runtun waktu (*time series*). *Time series* atau data runtun waktu merupakan kumpulan data yang terobservasi selama rentang waktu tertentu. *Missing values* yang terjadi pada sebuah *dataset* dengan jumlah yang besar dan berurutan dapat mempengaruhi performa sebuah metode estimasi.

Metode estimasi yang telah dikembangkan dalam penanganan *missing values* pada data runtun waktu adalah *Centroid Decomposition*. Metode tersebut digunakan untuk melakukan estimasi menggunakan data berkorelasi dari interval yang sama dengan data yang mengalami *missing values* terhadap data *missing values* yang telah diinterpolasi menggunakan interpolasi linier. Penelitian ini bertujuan untuk mengetahui pengaruh penggunaan metode penghalusan terhadap data yang berkorelasi dengan data yang mengalami *missing values* pada beberapa titik berurutan. Penerapan metode penghalusan diharapkan dapat memperbaiki akurasi estimasi terhadap *missing values*. Metode penghalusan yang digunakan pada penelitian ini adalah Savitzky-Golay (SG) yang diterapkan pada sebuah data runtun waktu dengan beberapa skenario pengujian.

Hasil penelitian ini menunjukkan bahwa penerapan metode penghalusan terhadap data acuan pada proses estimasi memberi pengaruh pada akurasi estimasi metode CD. Pengukuran metode akurasi dilakukan dengan menghitung nilai *Root Mean Square Error* (RMSE). Hasil evaluasi menunjukkan RMSE yang diperoleh dengan menerapkan metode penghalusan Savitzky-Golay pada estimasi *consecutive missing values* lebih rendah secara signifikan dibandingkan dari proses estimasi CD tanpa metode penghalusan pada tingkat kepercayaan 95% dengan penurunan tingkat *error* sebesar 0,09% untuk 5 MV, 0,14% untuk 10MV, dan 0,08% untuk 20MV.

Kata kunci – Estimasi, Penghalusan, Savitzky-Golay, *Missing Values*, *Centroid Decomposition*