

INTISARI

METODE ALTERNATIF *AUTOGROUPING* DALAM SINGULAR *SPECTRUM ANALYSIS*

Oleh

GUMGUM DARMAWAN

17/420357/SPA/00626

Bermula dari fungsinya sebagai *tool* untuk dekomposisi, *Singular Spectrum Analysis* (SSA) sekarang mengalami perkembangan yang cukup pesat. SSA digunakan untuk berbagai keperluan, yang paling umum adalah untuk *forecasting* (peramalan), model-modelnya diantaranya *Linear Recurrent Formula* (LRF), *Vector Method* dan *Bootstrap*. Selain sebagai *tool* untuk *forecasting* dan *decomposition*, SSA juga digunakan sebagai *missing value imputation*, *filtering*, *linearity test*, *structure change detection* dan *structural identification*.

Pada disertasi ini, proses *autogrouping* pada SSA dilakukan secara otomatis dengan pendekatan *long memory*. Automatisasi pada proses *grouping* sangat diperlukan karena pada *basic* SSA pengelompokan pada tahap ini berdasarkan visual dapat mengakibatkan multi interpretasi. Dengan melakukan proses *autogrouping* memungkinkan analisis SSA menjadi otomatis secara keseluruhan (Auto SSA). Proses Automatisasi dilakukan juga pada model *hybrid* antara SSA dengan ARIMA, proses ini dibangun agar analisis dapat dilakukan secara lebih cepat dengan memperhitungkan berbagai kemungkinan model. Dalam disertasi ini juga penulis membangun algoritma untuk mengkonversi data harian menjadi data bulanan yang mengikuti kalender Hijriyah. Data runtun waktu dalam bentuk Hijriyah sangat diperlukan karena banyak fenomena di Indonesia yang mengikuti kalender Hijriyah.

Pada automatisasi *grouping*, vektor eigen hasil proses *Singular value decomposition* diidentifikasi tingkat stasioneritasnya (d) kemudian dilakukan pembedaan sebesar d . Statistik untuk melakukan pembedaan digunakan statistik Geweke dan Porter-Hudak. Statistik ini merupakan kelas regresi spektral. Pemilihan statistik ini dilakukan secara simulasi untuk menunjukkan kelayakan metode yang digunakan. Kelas regresi spektral yang dibandingkan adalah Metode Geweke dan Porter-Hudak (GPH), metode *Smoothed Periodogram Regression* (SPR), Metode Geweke and Porter-Hudak Trimming (GPHtr), Metode *Cosine-Bell Tapered Data* (GPHTa) dan Metode *Modified Geweke and Porter-Hudak* (MGPH).

Dalam simulasi, kelima metode ini digunakan untuk mendeteksi *true value* dari nilai koefisien pembeda (d) untuk berbagai tipe data. Tipe data yang disimulasikan adalah data *short memory*, data stasioner *long memory*, data non-stasioner *long memory* dan data *long memory* yang mengandung pencilan. Berdasarkan hasil simulasi, pada data yang tidak mengandung pencilan metode GPH, SPR dan MGPH memperlihatkan performansi yang baik (berdasarkan nilai MAPE). Sedangkan, pada data yang mengandung pencilan metode GPH menunjukkan hasil yang kurang baik dibandingkan dengan empat metode lainnya.

Kajian simulasi dilakukan juga pada proses *Autogrouping* SSA. Tiga pola data musiman dibangkitkan menggunakan bantuan Software R. Tiga pola musiman tersebut adalah pola musiman reguler, musiman aditif dan pola musiman multiplikatif. Ketiga pola tersebut masing-masing dilakukan analisis *autogrouping* dengan dua metode, yaitu *autogrouping* Alexandrov dan *autogrouping* alternatif. Hasil dari simulasi metode alternatif lebih akurat



hasil peramalannya (berdasarkan nilai MAPE) untuk horizon peramalan yang pendek (3 bulan dan 6 bulan) dibandingkan metode Alexandrov, sedangkan untuk horizon 12 bulan peramalan ke depan metode Alexandrov lebih akurat (berdasarkan nilai MAPE).

Model yang telah dibangun digunakan pada data real berupa data covid-19 (*confirmed cases*, *recovered cases* dan *death cases*), data curah hujan, CO₂, *AirPassengers* dan *LakeHuron*. Data CO₂, *AirPassengers* dan *LakeHuron* tersedia dalam Software R, ketiga data tersebut merupakan data internal dalam Software R. Untuk data curah hujan digunakan data curah hujan Kota Banda Aceh, data curah hujan ini dikonversi ke dalam dua jenis kalender yaitu kalender Masehi (Gregorian) dan kalender Hijriyah. Konversi dari data harian menjadi data bulanan kalender Hijriyah dilakukan secara otomatis dengan menggunakan Software R. Dalam disertasi ini penulis menggunakan data curah hujan dalam membandingkan kalender Hijriyah dan kalender Masehi. Analisis autokovarian memberikan hasil yang berbeda periode musiman antara data yang menggunakan kalender Masehi dengan kalender Hijriyah. Berdasarkan Kalender Hijriyah periode musiman 20 bulan sedangkan dengan menggunakan kalender Masehi periode musimannya 12 bulan.

Kata Kata Kunci: Peramalan, Kalender Hijriyah, Geweke and Porter-Hudak, *Singular Spectrum Analysis*, Covid-19, *Regresi Spektral*, *Long Memory*.



ABSTRACT

***AUTOGROUPING ALTERNATIVE METHOD ON SINGULAR
SPECTRUM ANALYSIS***

By

GUMGUM DARMAWAN
17/420357/SPA/00626

Starting from its function as a decomposition tool, Singular Spectrum Analysis is now experiencing reasonably rapid development. SSA is used for various analyzes; the most common is forecasting, which includes models such as the Linear recurrent formula (LRF), the Vector Method, and the Bootstrap. Apart from being a tool for forecasting and decomposition, SSA is also used for missing value imputation, filtering, linearity test, structure change detection, and structural identification.

In this dissertation, the auto grouping process in SSA is done automatically with a long memory approach. Automation in the grouping process is essential because in basic SSA the grouping is based on visuals; this can lead to multiple interpretations. By performing the auto grouping process, it is possible to fully automate the SSA analysis (Auto SSA). The automation process is also carried out on the *hybrid* model between SSA and ARIMA, this process is built so that analysis can be carried out more quickly by taking into account various possible models. In this dissertation, the author also builds an algorithm to convert daily data into monthly data that follows the Hijri calendar. Time series data in the form of hijri is necessary because a lot of phenomena in Indonesia are based on the Hijri calendar.

In grouping automation, the eigenvectors resulting from the Singular value decomposition process were identified for their stationarity (d) and then differentiated by d . Statistics Geweke and Porter-Hudak, which is a spectral regression class, is used to make the statistical difference. The decision to use this type of statistics was made after conducting a simulation study that aimed to compare and test the feasibility of different methods of spectral regression classes. These were the Geweke and Porter-Hudak (GPH) method, the Smoothed Periodogram Regression (SPR) method, the Geweke and Porter-Hudak Trimming (GPHtr) method, the Cosine-Bell Tapered Data (GPHTa) method, and the Modified Geweke and Porter-Hudak method. (MGPH).

In the simulation, these five methods were used to detect the true value of the differentiating coefficient (d) for various data types. The simulated data types were short memory data, long memory stationary data, long memory non-stationary data, and long memory data containing outliers. Based on the simulation results, the GPH, SPR and MGPH methods showed good performance (based on the MAPE value) for data that did not contain outliers. Meanwhile, the GPH method showed less accurate results compared to the other four methods for data containing outliers.



Simulation studies were also carried out on the SSA Autogrouping process. Three seasonal data patterns were generated using Software R. The three seasonal patterns were regular, additive, and multiplicative seasonal patterns. These patterns were analyzed by auto grouping using two methods: Alexandrov auto grouping and Alternative auto grouping. The results from the simulation of the Alternative method were more accurate in forecasting results (based on MAPE values) for short forecasting horizons (3 months and 6 months) than the Alexandrov method. In contrast, for horizon 12 months, the forecasting results of the Alexandrov method were more accurate (based on MAPE values).

An application that had been designed was then used in the forecasting of actual data such as covid-19 (confirmed, recovered, and death cases), rainfall data, Co₂, AirPassengers, and LakeHuron. It was carried out on two types of calendars for rainfall data, namely the Gregorian calendar and the Hijri calendar. Conversion from daily data to monthly Hijri calendar data was done automatically using R software; the complete procedure can be seen in chapter III. In this dissertation, the author uses rainfall data in comparing the Hijri Calendar and the Gregorian Calendar. The auto covariant analysis provided different results between the data using the Gregorian Calendar and the Hijri Calendar. Based on the Hijri calendar, the seasonal period was 20 months, while using the Gregorian calendar, the seasonal period was 12 months.

Keywords: Forecasting, Hijri Calendar, Geweke and Porter-Hudak, Singular Spectrum Analysis, Covid-19, Spectral Regression, Long Memory Time Series.