

PEMROSESAN AWAL DATA RUNTUN WAKTU HASIL PENGUKURAN UNTUK IDENTIFIKASI SISTEM TUNGKU SINTER DEGUSSA

Dede Sutarya

Pusat Teknologi Bahan Bakar Nuklir
Badan Tenaga Nuklir Nasional, Serpong, Banten, Indonesia, 15313
dedes@batan.go.id

ABSTRAK—Pemrosesan awal data merupakan langkah penting dan kritis serta memiliki dampak besar pada keberhasilan analisis atau penggunaan selanjutnya dari data. Data hasil pengukuran yang terbebas dari *noise* tidak pernah dapat diperoleh dalam pengukuran menggunakan sensor dilaboratorium proses kimia atau fisika karena adanya *noise* yang timbul dari efek termodinamika dan kuantum. Selain itu *noise* juga terjadi karena kesalahan transmisi, lokasi memori yang rusak, dan kesalahan timing pada konversi analog ke digital. Pada makalah ini dilakukan penelitian dan eksperimen untuk mencari parameter optimal penapisan *noise spike* (outlier) menggunakan *median filter* pada data runtun waktu input-output hasil akuisisi proses sintering. Dari hasil eksperimen diperoleh parameter optimal median filter untuk data runtun waktu input-output proses sintering adalah dengan lebar jendela pergeseran $N=25$. Parameter tersebut menghasilkan rasio sinyal terhadap *noise* (SNR) yang cukup tinggi dengan rerata 3,6685 dan rerata kesalahan kuadrat (MSE) rendah dengan rerata 0,0352 dengan tetap mempertahankan bentuk asli puncak sinyal. Dengan demikian data hasil pemrosesan awal data dapat digunakan pada proses selanjutnya yaitu identifikasi sistem menggunakan teknik cerdas dengan efisien dan akurat.

Kata Kunci – Pemrosesan awal data, runtun waktu, median filter, sintering

ABSTRACT—*Data preprocessing is an important and critical step and have a huge impact on the success of the analysis or the subsequent use of the data. Measurement data with free of noise can never be obtained from the sensor measurement in chemical or physical process laboratory due to the noise arising from thermodynamics and quantum effects. In addition, noise also occurs because of a transmission error, faulty memory location, and timing errors at the analog to digital conversion. In this paper carried out research and experiments to find the optimal parameters for filtering spikes noise (outliers) using a median filter on input-output time series data that obtained from the data acquisition of the sintering process. From the experimental results obtained that median filter optimal parameter is using moving windows size $N=25$. These parameters produce sufficiently high Signal to Noise Ratio (SNR) with the average of 3.6685 and a low Mean Square Error (MSE) with the average of 0.0352 while maintaining the shape of the original signal peaks on the data. Thus the results of data preprocessing can be used in the next step of the data usage i.e. for system identification using intelligent technique efficiently and accurately.*

Keywords – *Data preprocessing, time series, median filtering, sintering*

I. PENDAHULUAN

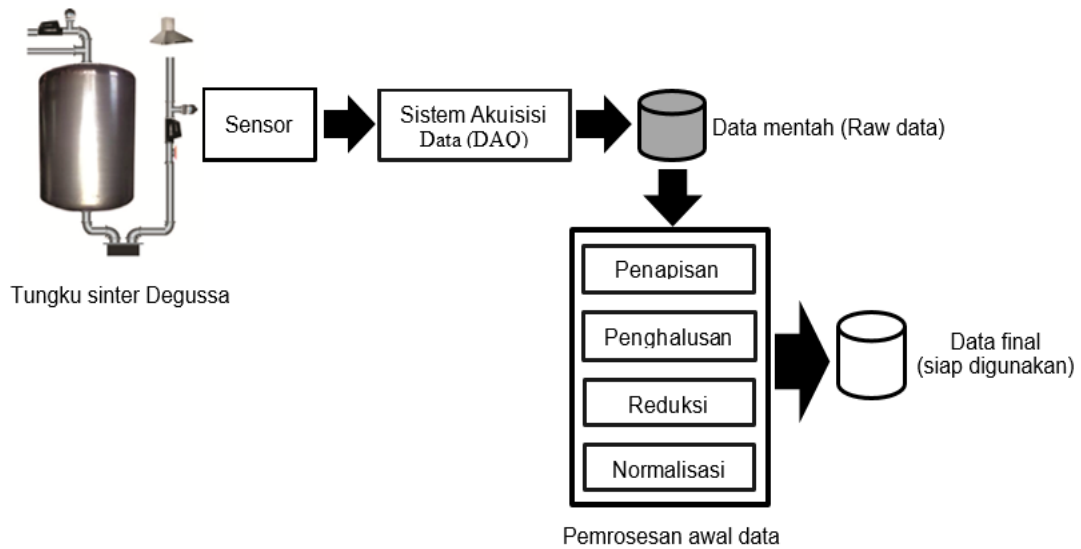
Identifikasi sistem adalah ilmu tentang bagaimana membangun model matematik dari sistem dinamis melalui pengamatan terhadap data input-output dari sistem. Tahapan awal dalam identifikasi sistem adalah merancang eksperimen yang tepat

sehingga data yang diperoleh berisi informasi maksimum tentang proses. Selanjutnya data yang diperoleh mengalami beberapa teknik pemrosesan awal untuk menghilangkan efek gangguan yang tidak diinginkan, hal ini akan meningkatkan efisiensi dari proses identifikasi. Dalam paper ini akan dilakukan studi untuk tahapan awal dalam identifikasi sistem seperti yang di sebutkan diatas.

Untuk mendapatkan data input-output dari proses, maka pada tungku sinter Degussa ditambahkan sensor-sensor untuk mengukur parameter proses yang kemudian diakuisisi oleh sebuah modul sistem akuisisi data dan disimpan sebagai *database* proses. Data mentah hasil pengukuran yaitu parameter input dan output proses sintering terlebih dahulu melalui tahapan pemrosesan awal data (*data preprocessing*), sebelum digunakan sebagai bahan untuk memodelkan perilaku sistem dalam identifikasi sistem tungku sinter.

Data hasil akuisisi yang terbebas dari *noise (outlier)* tidak pernah dapat diperoleh dalam pengukuran menggunakan sensor dilaboratorium proses kimia atau fisika karena beberapa jenis *noise* yang timbul dari efek termodinamika dan kuantum tidak mungkin dihilangkan. Penggunaan luas komputer dalam instrumentasi proses kimia dan fisika serta fleksibilitas pemrograman perangkat lunak membuat penggunaannya untuk deteksi *noise* dan *filtering* [1], pembersihan data *outlier* [2-4], *median filtering* [5,6], dan penghalusan data (*smoothing*) [7,8] banyak di implementasikan dengan hasil yang cukup memuaskan. Oleh karena itu maka pemrosesan awal terhadap data runtun waktu (*time series*) hasil akuisisi data proses sintering akan diawali dengan proses *filtering* data *outlier* menggunakan *median filter*, kemudian dilanjutkan dengan melakukan normalisasi terhadap data. Hal ini dilakukan untuk menjaga akurasi dan ketepatan analisis pada proses identifikasi sistem.

Pemrosesan awal data (*data preprocessing*) merupakan teknik *data mining* yang melakukan perubahan data mentah menjadi format yang lebih dimengerti. Data *real* sering tidak lengkap, tidak konsisten, dan atau kurang dalam perilaku tertentu, dan kemungkinan mengandung banyak gangguan (*noise*). Data *preprocessing* menyiapkan data mentah untuk diproses lebih lanjut sesuai kebutuhan. Diagram umum pemrosesan awal data diperlihatkan pada Gambar 1. Walaupun demikian dalam praktek belum tentu semua tahapan dalam pemrosesan awal data digunakan, hal ini tergantung dari kebutuhan dan kondisi data mentah yang ada.

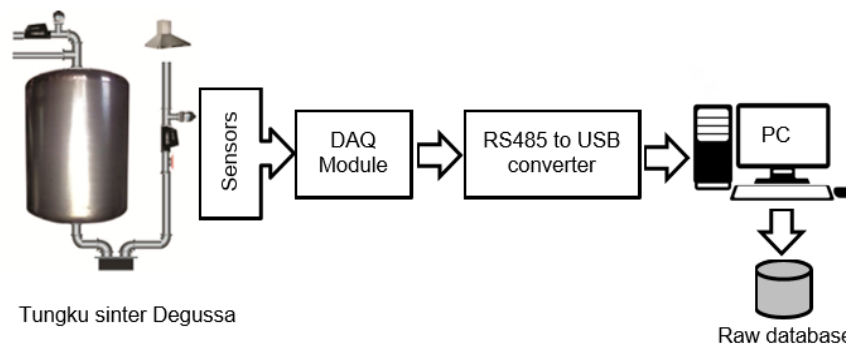


Gambar 1. Diagram umum pemrosesan awal data runtun waktu

II. METODE AKUISISI DAN PRE-PROCESSING DATA

A. Data akuisisi proses sintering

Antar muka instrumen pengukuran analitis menggunakan komputer (PC) untuk akuisisi data saat ini telah menjadi fasilitas standar laboratorium modern. Untuk mendapatkan data I/O (input-output) dari proses, maka pada tungku sinter harus ditambahkan sensor-sensor untuk mengukur parameter proses yang kemudian diakuisisi oleh sebuah modul sistem akuisisi data, seperti yang diperlihatkan pada Gambar 2. Data yang diperoleh kemudian disimpan sebagai *database* proses. Selanjutnya *database* tersebut akan digunakan sebagai bahan untuk memodelkan perilaku sistem dalam identifikasi sistem tungku sinter menggunakan teknik cerdas.



Gambar 2. Diagram sistem akuisisi data untuk mengukur I/O proses sintering

Pada awal proses sintering, setelah pelet uranium dioksida (UO_2) dimasukan ke tungku sinter dan setting parameter pada komputer kendali dilakukan, proses pemanasan dimulai dengan mengalirkan gas nitrogen terlebih dahulu sampai temperatur *auto-ignition* hidrogen terlewati (diatas $580^{\circ}C$). Setelah itu kemudian hidrogen dialirkan kedalam ruang bakar tungku sampai temperatur $1700^{\circ}C$, kemudian dilakukan penahanan pada temperatur tersebut selama 3-4 jam dan dilanjutkan dengan proses pendinginan hingga $580^{\circ}C$, gas atmosfer proses kemudian diganti kembali dengan nitrogen.

Karena adanya perangkat-perangkat yang kemungkinan menjadi sumber *noise* serta reaksi kimia dan fisika dalam proses sintering sehingga dibutuhkan sensor-sensor yang memiliki karakteristik khusus untuk penggunaan gas berbahaya seperti hidrogen. Selain itu instalasi sistem akuisisi data terutama sensor yang bersinggungan langsung dengan hidrogen perlu pengujian dan kalibrasi khusus menyangkut tingkat kebocoran dan kekuatan material terhadap gas hidrogen. Pemasangan sensor sensor pada sistem pemipaan gas hydrogen memerlukan isolasi yang memadai dengan tingkat kebocoran tidak melebihi 1×10^{-9} lpm [9] dengan demikian pengujian tingkat kebocoran pada sambungan perlu dilakukan dengan teliti setelah instalasi sistem akuisisi data selesai dilakukan.

Program akuisisi data di kembangkan dengan software LabVIEW, komputer kendali ditempatkan sekitar 20m dari tungku sinter menggunakan komunikasi RS-485 untuk menghindari bahaya yang mungkin terjadi selama proses eksperimen.

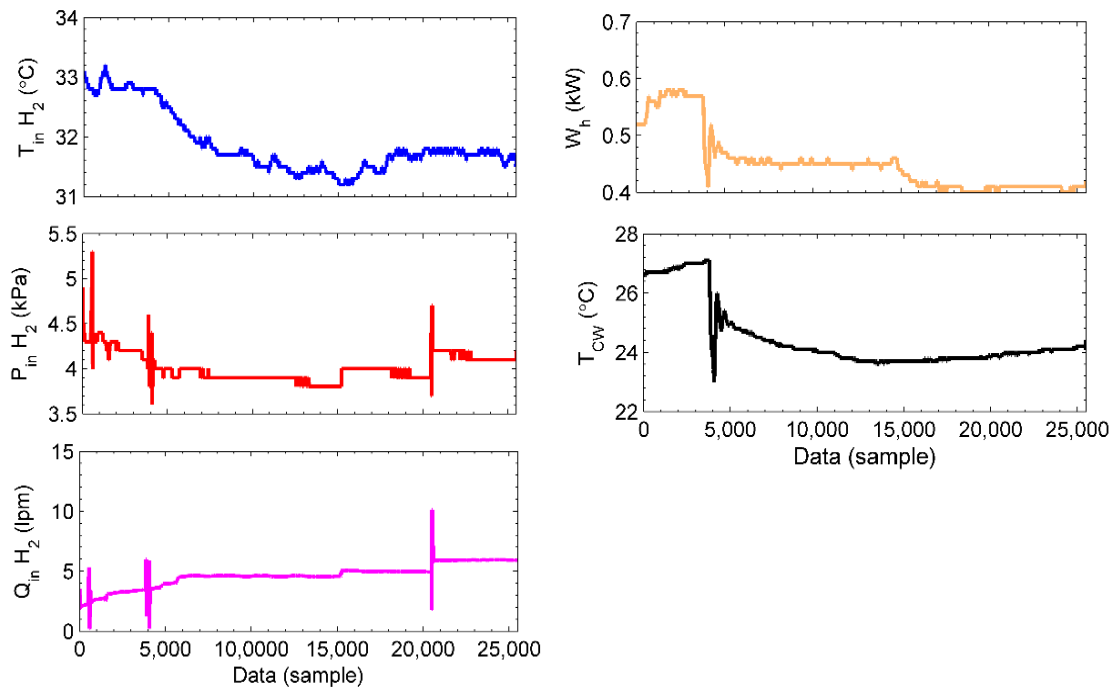
Penempatan sistem akuisisi data pada panel kendali sinter memanfaatkan ruang yang telah disediakan oleh pabrikan untuk pengembangan sistem, sehingga desain dan sistem keselamatannya sudah dipertimbangkan dengan matang oleh pabrikan. Penggunaan kabel dan komponen sistem dengan kualitas dan isolasi terhadap *noise* yang memadai sangat penting, selain untuk kualitas data yang akan diperoleh tetapi juga terkait dengan keselamatan selama proses pengambilan data yang cukup lama yaitu sekitar 36 jam untuk satu siklus proses.

Dari data mentah (raw data) hasil akuisisi proses sintering yang terdiri dari 73.000 data, kemudian diambil sebanyak 25.000 sampel data proses yang menggunakan atmosfer hydrogen untuk dilakukan pre-processing. 25.000 data I/O tersebut diperlihatkan pada Gambar 3 dan Gambar 4.

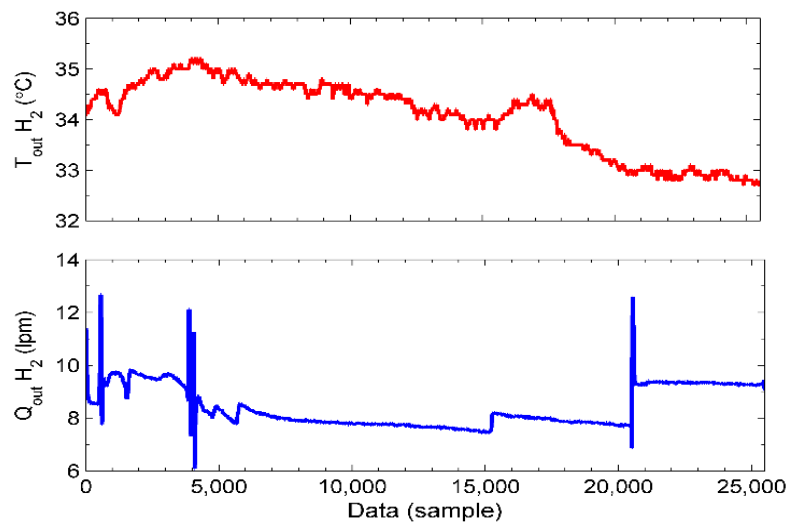
Parameter input proses sintering yaitu; Temperatur input hidrogen (T_{in}), Tekanan input hidrogen (P_{in}), Laju alir input hidrogen (Q_{in}), Daya pemanas (W_h), dan

Temperatur air pendingin (T_{CW}). Sedangkan parameter output proses yaitu; Temperatur output hidrogen (T_{out}) dan Laju alir output hidrogen (Q_{out}).

Data mentah parameter I/O hasil pengukuran proses sintering terlebih dahulu melalui tahapan pemrosesan awal data (*data preprocessing*). Dengan melihat kondisi data mentah (*raw data*) yang diperoleh, pemrosesan awal yang dibutuhkan adalah proses penapisan atau filtering untuk menghilangkan outlier dan normalisasi pada data sebelum digunakan.



Gambar 3. Data parameter input hasil pengukuran sebanyak 25.000 sampel.



Gambar 4. Data parameter output hasil pengukuran sebanyak 25.000 sampel.

B. Membersihkan outlier data menggunakan median filter

Outlier adalah data hasil pengamatan yang memiliki penyimpangan secara signifikan dari sebagian besar pengamatan. *Outlier* tersebut mungkin disebabkan oleh lonjakan sinyal (*spike*), kesalahan pengukuran sensor, *noise* yang diakibatkan peralatan proses, degradasi peralatan instrumen atau berhubungan dengan kesalahan manusia. *Outlier* pada data akan menyebabkan analisa terhadap data menjadi tidak berguna karena *outlier* dapat menyebabkan ketidaksesuaian spesifikasi model, estimasi parameter menjadi bias bahkan menghasilkan kesimpulan analisa yang salah.

Oleh karena itu untuk mengatasi masalah data *outlier* tersebut digunakan *median filter* dalam mendeteksi dan menghapuskan *outlier* yang didasarkan pada jendela data yang bergerak (*moving window*) pada data runtun waktu (*time series*). Penjelasan tentang prinsip kerja median filter adalah sebagai berikut.

Dengan asumsi bahwa pembangkitan *outlier* dapat dijelaskan oleh model *outlier* aditif yang populer pada *robust-time series analysis* [2,10] sebagai berikut

$$y_k = x_k + o_k. \quad (1)$$

dimana y_k adalah urutan data hasil pengukuran, x_k adalah urutan data nominal yang kita inginkan dan O_k menggambarkan urutan data yang berisi *outlier*. Nilai-nilai urutan O_k di asumsikan menjadi nol kecuali untuk kasus waktu sesaat dengan nilai O_k yang jauh lebih besar dibandingkan dengan variasi nominal yang terlihat dalam data. Pencarian nilai perkiraan x_k didasarkan pada pengamatan data saat ini dan sebelumnya y_{k-j} untuk $k \geq j \geq 0$. Secara khusus data, y_k , $N-1$ dan y_{k-j} disimpan dalam jendela data W_k dengan lebar N .

$$W_k = \{y_{k-N+1}, y_{k-N+2}, \dots, y_k\}. \quad (2)$$

Nilai data dalam jendela kemudian disusun berdasarkan peringkat untuk mendapatkan

$$R_k = \{y_{(1)}^k \leq y_{(2)}^k \leq \dots \leq y_{(N)}^k\}, \quad (3)$$

dan median y_k^m dari urutan R_k dihitung sebagai

$$y_k^m = \begin{cases} y_{((N+1)/2)}^k & \text{untuk } N \text{ ganjil} \\ (y_{(N/2)}^k + y_{(N/2+1)}^k) / 2 & \text{untuk } N \text{ genap} \end{cases} \quad (4)$$

Nilai median y_k^m menyediakan referensi data nominal sebagai kompensasi data saat ini y_k , kemudian di evaluasi dengan menentukan jarak d_k antara y_k dan y_k^m .

$$d_k = |y_k^m - y_k| \quad (5)$$

Jika jarak ini melebihi ambang batas yang ditentukan yaitu $T_k \geq 0$, maka y_k dinyatakan sebagai *outlier* dan menggantinya dengan nilai prediksi y_k^m untuk mendapatkan urutan data setelah difilter f_k dengan persamaan berikut

$$f_k = \begin{cases} y_k & \text{jika } d_k \leq T_k \\ y_k^{pred} & \text{jika } d_k > T_k \end{cases} \quad (6)$$

C. Normalisasi data

Normalisasi merupakan tahapan terakhir dalam pemrosesan awal data. Tujuan normalisasi data adalah untuk memastikan bahwa distribusi statistik nilai untuk setiap I/O dalam jaringan saraf tiruan yang akan digunakan untuk identifikasi sistem mendekati seragam. Selain itu, nilai-nilai harus diskala untuk menyesuaikan dengan masukan *neuron* dalam jaringan saraf tiruan (neural networks).

Metode normalisasi data yang digunakan dalam penelitian ini adalah skala *linear* dimana data akan diubah dalam range -1 sampai 1 atau 0 sampai 1 sesuai dengan fungsi aktivasi yang digunakan pada *neuron* input. Fungsi normalisasi linear didefinisikan sebagai berikut.

$$X_{i, 0 \text{ sampai } 1} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \quad (7)$$

$$X_{i, -1 \text{ sampai } 1} = \frac{X_i - \left(\frac{X_{\max} + X_{\min}}{2} \right)}{\left(\frac{X_{\max} - X_{\min}}{2} \right)} \quad (8)$$

dimana X_i adalah titik data ke- i yang akan dinormalisasi, X_{\max} merupakan nilai terbesar dari seluruh data, dan X_{\min} merupakan nilai terkecil dari seluruh data.

Ketika jaringan saraf tiruan selesai diuji dengan data *testing* pada tahapan identifikasi sistem, output yang dihasilkan harus di de-normalisasi. Jika normalisasi dapat sepenuhnya dikembalikan ke nilai asalnya dengan sedikit atau tidak ada kehilangan dalam akurasi, maka hal tersebut tidak menjadi masalah. Namun, jika normalisasi terganggu oleh nilai yang tidak diinginkan dalam data (*outlier*), maka output nilai yang sama dengan *outlier* akan dianggap sebagai nilai sebenarnya, namun jika nilai tersebut dapat diterima untuk aplikasi yang diinginkan, maka metode normalisasi dapat dikatakan cukup reversibel.

III. HASIL DAN PEMBAHASAN

Proses untuk menghapus *outlier* pada data runtun waktu, input dan output hasil eksperimen proses sintering dilakukan dengan memvariasikan ukuran jendela N yang bertujuan untuk mencari ukuran jendela yang tepat atau optimal yang menghasilkan peningkatan rasio sinyal terhadap gangguan (SNR) yang terbaik. Untuk mencapai hal tersebut variasi ukuran jendela yang digunakan adalah 15, 25, 35, 65 dan 115.

Pada penelitian ini bentuk sinyal asli terutama puncak sinyal sangat penting sehingga akan dicari lebar jendela filter yang optimal dengan menentukan kriteria rata-rata kuadrat kesalahan (MSE) terendah dan rasio sinyal terhadap *noise* (SNR) yang paling tinggi. MSE dan SNR dihitung menggunakan persamaan berikut.

$$MSE = \frac{1}{N} \sum_{n=0}^N (V(n) - V_R(n))^2 \quad (9)$$

$$SNR = \log_{10} \frac{\sum_{n=0}^N V_R^2(n)}{\sum_{n=0}^N S_R^2(n)} \quad (10)$$

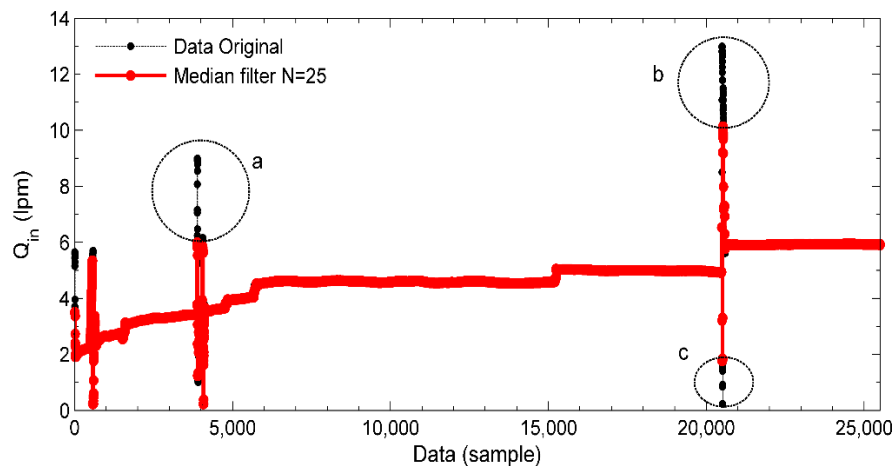
dimana N adalah jumlah data runtun waktu, $V(n)$ adalah sinyal asal sebelum difilter, $V_R(n)$ adalah sinyal hasil rekonstruksi setelah difilter, dan $S_R(n)$ merupakan perbedaan antara sinyal asal dan sinyal hasil rekonstruksi setelah difilter.

Hasil perhitungan MSE dan SNR untuk data input dan output setelah difilter untuk menghilangkan data *outlier* diperlihatkan pada Tabel 1.

Tabel 1. Hasil perhitungan MSE dan SNR setelah difilter

Data	Kriteria	Lebar jendela median filter (M)					
		5	15	25	35	65	115
T_{in}	MSE	1.5e-03	1.8e-03	2.0e-03	2.1e-03	2.4e-03	2.8e-03
	SNR	5.8172	5.7498	5.7152	5.6897	5.6276	5.5611
P_{in}	MSE	4.2e-03	1.3e-02	1.1e-02	1.2e-02	1.3e-02	1.3e-02
	SNR	3.5832	3.0973	3.1670	3.1278	3.0826	3.0915
Q_{in}	MSE	2.3e-02	1.1e-01	1.5e-01	1.7e-01	1.3e-01	1.2e-01
	SNR	2.9895	2.3323	2.1777	2.1166	2.2470	2.3323
W_h	MSE	1.3e-05	1.5e-05	1.7e-05	2.0e-05	2.9e-05	5.0e-05
	SNR	4.2084	4.1291	4.0774	4.0216	3.8602	3.6201
T_{cw}	MSE	4.7e-03	5.4e-03	6.1e-03	7.1e-03	1.2e-02	2.4e-02
	SNR	5.1087	5.0460	4.9906	4.9245	4.7045	4.4015
T_{out}	MSE	2.2e-03	2.8e-03	3.3e-03	3.6e-03	4.4e-03	5.3e-03
	SNR	5.7314	5.6146	5.5516	5.5063	5.4223	5.3423
Q_{out}	MSE	1.2e-02	4.3e-02	7.4e-02	8.9e-02	1.1e-01	1.2e-01
	SNR	3.7693	3.2229	2.9847	2.9048	2.8099	2.7497

Pada Tabel 1 terlihat bahwa variasi lebar jendela pada *median filter* semakin besar maka MSE naik yang diikuti oleh penurunan SNR dengan kata lain semakin besar ukuran jendela maka kinerja *filter* semakin progresif tetapi deformasi akan semakin besar terhadap sinyal aslinya. Walaupun demikian tergantung dari karakter sinyal aslinya seperti pada data input P_{in} (lihat Tabel 1), MSE dan SNR fluktuatif pada perubahan lebar jendela *filter*. Oleh karena itu dalam pemilihan ukuran jendela pengamatan *visual* dan *engineering judgment* penting dilakukan untuk melihat sinyal setelah proses *filtering*.

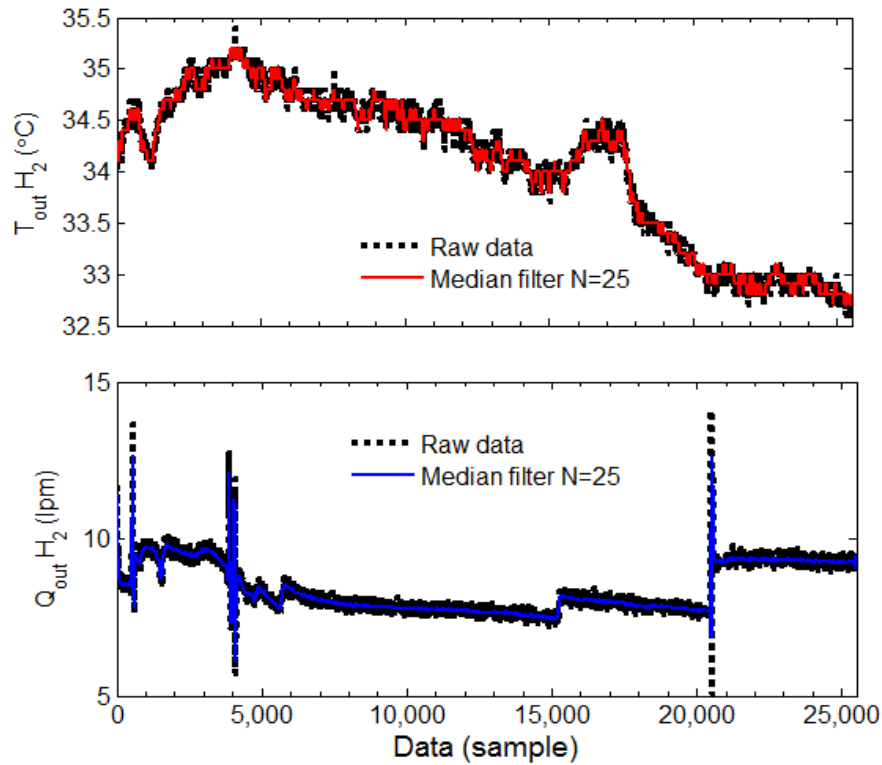


Gambar 5 *Scater plot* data laju alir hidrogen, data original dan setelah difilter dengan ukuran jendela $N=25$. Bagian a, b, c merupakan *outlier* dalam data.

Berdasarkan hasil perhitungan MSE dan SNR seperti diperlihatkan Tabel 1, serta pengamatan *visual* terhadap data asli dan hasil filtering maka diputuskan bahwa ukuran jendela pergeseran optimal adalah $N=25$ untuk pemrosesan awal data proses sinter. Dengan parameter tersebut menghasilkan SNR yang cukup tinggi dengan rerata 3,6685 dan MSE rendah dengan rerata 0,0352 dengan tetap mempertahankan bentuk asli puncak sinyal. Untuk menggambarkan potensi *outlier* pada data, Gambar 5 memperlihatkan potensi outlier pada data laju alir hidrogen setelah difilter dengan ukuran jendela $N=25$.

Gambar 6 memperlihatkan data mentah (raw data) hasil pengukuran dan data setelah difilter dengan median filter optimal untuk output temperatur dan laju alir hidrogen.

Dengan menggunakan lebar jendela pergeseran optimal median filter yang diperoleh maka semua data I/O hasil pengukuran akan difilter dengan parameter tersebut dan dilanjutkan dengan normalisasi data menggunakan persamaan (7) atau (8) sesuai dengan fungsi aktivasi yang akan digunakan pada tahapan proses identifikasi sistem menggunakan teknik cerdas.



Gambar 6 Scatter plot data temperatur dan laju alir hidrogen, data original dan setelah difilter dengan ukuran jendela $N=25$.

IV. KESIMPULAN

Pemrosesan awal data runtun waktu hasil akuisisi data proses sintering telah dilakukan. Pemrosesan awal data yang dilakukan adalah panapisan atau penghapusan spike noise (outlier) pada data runtun waktu menggunakan median filter. Dari hasil eksperimen untuk mencari parameter optimal median filter diperoleh ukuran jendela pergeseran optimal adalah $N=25$. Dengan menggunakan ukuran jendela $N=25$ diperoleh nilai rerata SNR yang cukup tinggi yaitu 3,6685 dan MSE rendah dengan rerata 0,0352 dengan tetap mempertahankan bentuk asli puncak sinyal. SNR tinggi dan MSE rendah menunjukkan bahwa deformasi yang terjadi pada data rendah, sehingga yang difilter benar-benar merupakan spike noise (outlier).

Noise (outlier) pada data akan menyebabkan estimasi parameter menjadi bias bahkan menghasilkan kesimpulan analisa yang salah. Dengan menggunakan parameter optimal yang diperoleh data hasil pemrosesan awal data dapat digunakan pada proses selanjutnya yaitu identifikasi sistem menggunakan teknik cerdas dengan efisien dan akurat.

UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih kepada Agus Sartono D.S., Sugeng Rianto, Mugiyono dan Ade Mahpudin atas dukungan teknis selama pelaksanaan eksperimen dan akuisisi data dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Bhausahab Shinde, A.R. Dani," Noise Detection and Removal Filtering Techniques in Medical Images", International Journal of Engineering Research and Applications, Vol. 2, Issue 4, p.311-316, July-August 2012
- [2] P.H. Menold, R.K. Pearson, F. Allgower," Online outlier detection and removal", Proceedings of the 7th Mediterranean Conference on Control and Automation (MED99) Haifa, Israel - June 28-30, p.1110-1133, 1999.
- [3] R. Ganguli," Noise and outlier removal from jet engine health signal using weighted FIR median hybrid filter", Mechanical Systems and Signal Processing, 16(6), p.967-978, 2002.
- [4] Hancong Liu, Sirish Shah, Wei Jiang," On-line outlier detection and data cleaning", Elsevier, Computers and Chemical Engineering, 28, p.1635-1647, 2004.
- [5] Lin Yin, Ruikang Yang," Weighted Median Filters: A Tutorial", IEEE Transactions on Circuit and System: Analog and Digital Signal Processing, Vol. 43, No. 3, p. 157-192, March 1996.
- [6] Ruikang Yang, dkk," Optimal Weighted Median Filterin Under Structural Constraints", IEEE Transactions on Signal Processing, Vol. 43, No. 3, p. 591-604, March 1995.
- [7] H. Hwang and R. A. Haddad," Adaptive Median Filters: New Algorithms and Results", IEEE Transactions on Image Processing, Vol. 4, No. 4, p. 499-502, April 1995.
- [8] Ronald W. Schafer," What Is a Savitzky-Golay Filter", IEEE Signal Processing Magazine, p.111-117, July 2011.
- [9] NFPA Standard, NFPA 2: Hydrogen Technologies Code, 2016
- [10] Martin, R. and V. Yohai,"Influence Functionals for Time Series," Ann. Statist., 14, p. 781-818,1986