

A Wavelet Based Algorithm without a Priori Knowledge of Noise Level for Gross Errors Detection

Paolo Mercorelli

Institute of Product and Process Innovation
Leuphana University of Lueneburg
Volgershall 1, D-21339 Lueneburg, Germany
Tel.: +49-(0)4131-677-5571, Fax: +49-(0)4131-677-5300
mercorelli@uni.leuphana.de

Abstract

This paper deals with Gross Error Detection using a signal-based approach and proposes an algorithm to be applied in industrial processes. The developed algorithm is used in some industrial software platforms to detect sensor outliers. A validation of this algorithm through computer simulations is shown. At the end of the paper results using real sensor measurements from industrial processes are presented.

Keywords: Fault detection, Wavelets, Industrial applications

1. Introduction and motivation

The use of soft sensing, data reconciliation, and also parameter estimation are examples of applications that require "clean data" as for instance in [1]. A method of de-noising which is based on methods already introduced in [2] is presented in [3]. This method is based on the comparison of an information theory criterion which is the "description length" of the data. In [4] and [5] this theoretical approach is applied. In

[6], [7] and [8]) the authors have developed a robust version of the Hampel filter where a lower threshold may be set in order to avoid the undesired detection (removal) of noise. The algorithm assumes knowledge of the noise level of the signal, which is often unknown. The aim of this paper is to develop an approach that addresses outlier detection without de-noising that may be applied in real-time. In [9] a procedure based on a median function is presented and it is shown with a first version of a wavelet based algorithm. This contribution emphasizes the quality of the wavelet based algorithm and shows, shortly at the end, some possible misclassifications. The proposed algorithm does not require a priori knowledge of the noise level and for that it is able to work autonomously. This represents an advantage because of the fact that noise level algorithms are often very complex and require a relatively large amount of data, see [2], [5] and [3]. The paper is organized as follows. Section 2 presents the wavelet based algorithm together with some basic outliers definitions and validation tests. Section 3 presents some more structured valida-

tion tests of the algorithm. Moreover, the results are compared with another existing algorithm. At the end, a real industrial application is shown. The conclusions close the paper.

2. The Proposed Algorithm and its Validation

The proposed algorithm is based on the estimation of the variance of the local Lipschitz constant of the signal over a receding time horizon. The fault (outlier) is recognized if the local Lipschitz constant lies outside the computed boundary. Parameters c_1 and c_2 indicate two confidential constants.

Structure of the algorithm

- **Step 1** The signal is stored in a register and the standard deviation σ of the local Lipschitz (L) constant of its first consecutive 7 samples is calculated using the scalar product between two consecutive samples and Haar functions consisting of two samples. Afterwards, the local Lipschitz constant between the 7th and 8th sample is calculated in the same way.
- **Step 2a** If the local Lipschitz constant calculated between the 7th and 8th sample of the signal is less than constant $c_1\sigma$ (where c_1 is equal to 2), then the considered element of the sequence is not an outlier and the local Lipschitz constant is added to σ .
- **Step 2b** If the local Lipschitz constant calculated between the 7th and 8th sample is greater than

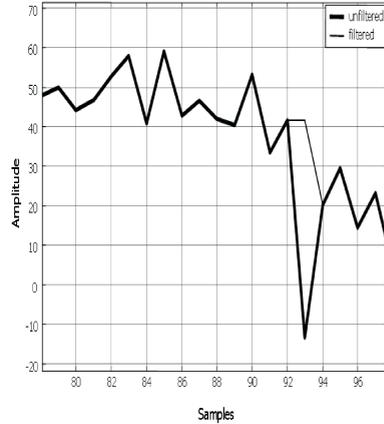


Fig. 1: An isolated outlier

constant $c_1\sigma$, then the considered element of the sequence is an outlier, see Fig. 1, and its local Lipschitz constant is not stored!!

- **Step 3** The local Lipschitz constant is stored and the next step of the sequence is considered.
- **Step 4** In case the stored local Lipschitz constant value is more than $c_1\sigma$, then the sign of the last two calculated local Lipschitz constants is checked. If they do not have opposite signs, then multi-outliers occur, see Fig. 2 or Fig. 3.

If they have opposite signs, then the stored local Lipschitz constant value is checked and if this value is not less than $c_2\sigma$ (with c_2 equal to 3), then it is a double inverse multi-outlier, see Fig. 4.

Figure 5 shows an example where a multi-outlier is misclassified. In Fig. 6 a flow chart of the algorithm is shown

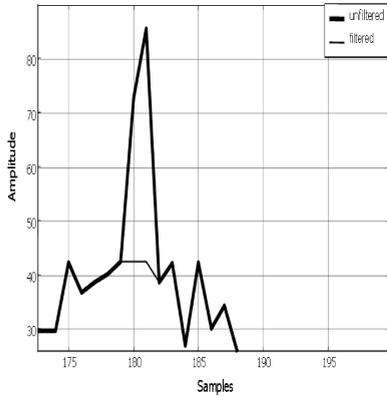


Fig. 2: A multiple outlier

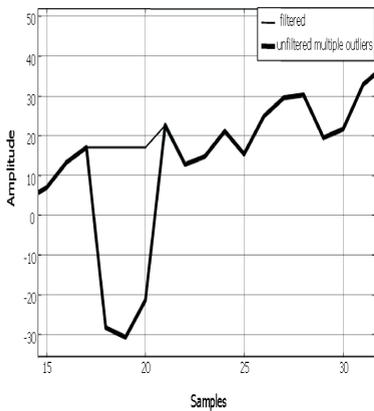


Fig. 3: A multiple (3) outlier

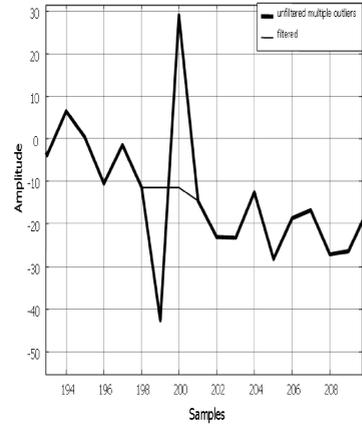


Fig. 4: A single inverse outlier

in which the main function and steps are indicated.

3. Results

The procedure is validated by using artificial data where the positions of the outliers are known. Fig. 7 shows the results obtained using a median filter of which the noise level as a priori knowledge is required to be known. It can be seen that 94.68% of outliers are correctly detected in data. Fig. 8 shows the simulated results obtained using the proposed wavelet algorithm without having a priori knowledge of the noise. It can be seen that almost 100% of outliers are correctly detected in the data. The results shown in Fig. 8 indicate that the misclassifications of the outliers are mostly located in the first part of the data. The percentage of the incorrect outliers detected in the data increases in this first part of the classification. The reason of that is that the initial standard deviation of the local Lipschitz constant is very small at the

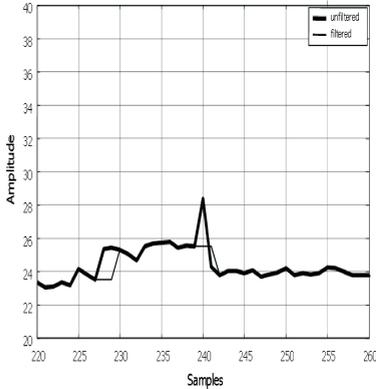


Fig. 5: Incorrect detection of multi-outliers (on sample number 227 and 228)

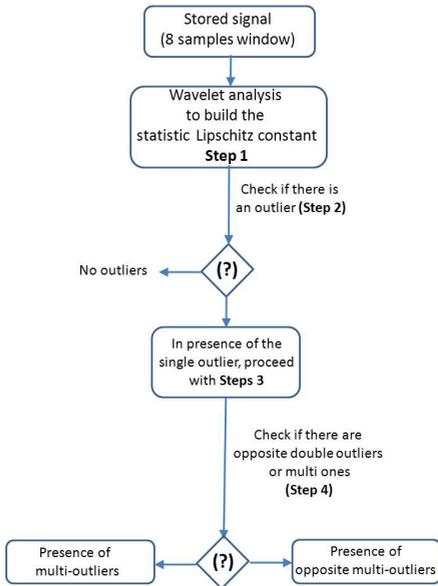


Fig. 6: Flow chart of the proposed algorithm

algorithm start. The initialization usually can take some samples, typically 15–20 samples. This aspect represents the weak point of the procedure which is a stochastic one and in some cases can be not robust. One of the most critical cases is when the standard deviation of the local Lipschitz constant results to be small. In Fig. 8 this effect can be seen in the first 20 samples. In Fig. 9 real application data are shown.

4. Conclusion

This paper deals with Gross Error Detection using a wavelet signal based approach. More specifically, this work presents a signal-based algorithm for applications in industrial processes. The proposed algorithm does not need a priori knowledge of the noise level. Validation through computer simulations together with industrial real cases were shown.

ACKNOWLEDGMENT

The author would like to thank ABB (Asea Brown Boveri), Ladenburg (Germany) for the technical support. This work would not have been accomplished without them.

References

- [1] J.S. Albuquerque, L.T. Biegler: Data Reconciliation and Gross-Error Detection for Dynamic Systems. *AIChE Journal* 42: 2841–2856, 1996.
- [2] R.R. Coifman, M.V. Wickerhauser: Entropy based algorithm for best basis selection. *In IEEE Transactions on Information Theory, IEEE*, 38:712–718, 1992.
- [3] R. Shao, F. Jia, E.B. Martin, A.J. Morris: *Wavelets and non linear*

principal component analysis for process monitoring. Control Engineering Practice, 7:865–879, Elsevier, 1999.

- [4] S. Beheshti, M.A. Dahleh, *On denoising and signal representation.* proceeding of the 10th Mediterranean Conference on Control and Automation (MED2002, July 9–12, Lisbon, (Portugal), 2002.
- [5] S. Beheshti, M.A. Dahleh, *Noise variance and signal denoising.* proceeding of 2003 IEEE International Conference on Acoustic Speech, and Signal Processing (ICASSP). pages 185–188, April 6–10, 2003.
- [6] P.H. Menold, R.K. Pearson, F. Allgöwer, *Online outlier detection and removal.* proceeding of the 7th Mediterranean Conference on Control and Automation (MED1999), June 28–30, Heifa, (Israel), 1999.
- [7] R.K. Pearson: Exploring process data. *Journal of Process Control*, 11:179–194, Elsevier, 2001.
- [8] R.K. Pearson: *Outliers in Process Modelling and Identification.* IEEE Transactions on Control Systems Technology, IEEE, 10: 55–63, 2002.
- [9] P. Mercorelli, A. Frick *Industrial Applications Using Wavelet Packets for Gross Error Detection.* in *Computational Intelligence in Information Assurance and Security.* Edited by N. Nadjah, A. Abraham and L.d.M. Mourelle, Springer Publisher, 89-127, 2007.

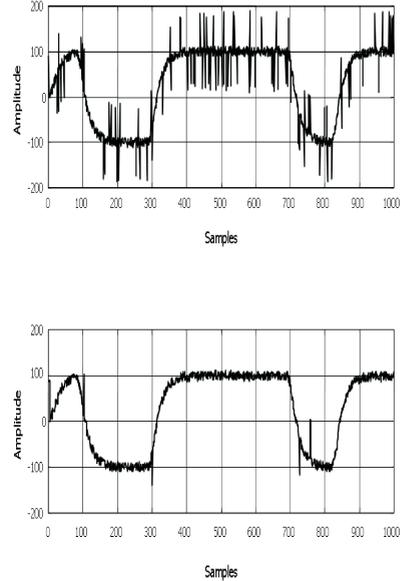


Fig. 7: Simulation using median filter (Algover’s Algorithm) with a priori knowledge on the noise

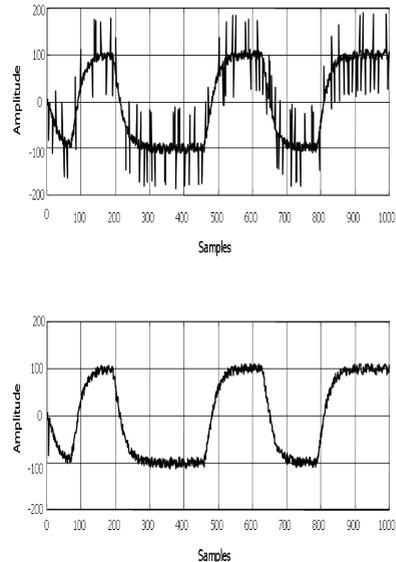


Fig. 8: Simulations by using wavelet algorithm without a priori knowledge on the noise

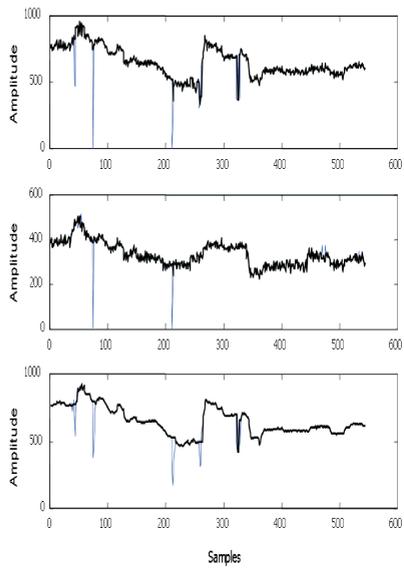


Fig. 9: The wavelet based algorithm applied to the Mining data. The fine line is the original data, and the bold is the filtered one