NON-PARAMETRIC HIGH AND LOW EXTREMA FILTERING METHOD FOR FILTERING EXTREMA IN DYNAMIC DOMAINS

K. K. L. B. Adikaram[1,2,3], M. A. Hussein[1], M. Effenberger[2], T. Becker[4]

[1] Research Group of Bio-Process Analysis Technology, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany.
[2] Institute for Agricultural Engineering and Animal Husbandry, Bavarian State Research Center for Agriculture, Vöttinger Straße 36, 85354 Freising, Germany.
[3] Computer Unit, Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiy, Sri Lanka.
[4] Lehrstuhl für Brau- und Getränketechnologie, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany.

ABSTRACT

Parametric domain dependent properties such as magnitude of prominences (or crater) and the widths at half prominence (or crater) are two properties of a signal that used to filter peaks and valleys with high and low amplitude. However, parametric methods are not suitable for filtering signals in unknown domains and very dynamic systems. In this paper we introduce a novel non-parametric method for filtering high and low extrema. The ratio $R_{LH\_min} = n/(S_n + (1 - a_{min}) * n) \rightarrow 1$ and $R_{LH\_max} = n/((a_{max} + 1) * n - S_n) \rightarrow 1$ implies that the a_{min} is very close to the other points (low crater) and a_{max} is very close to other points (low prominence), respectively, where n, a_{min}, a_{max} and S_n is number of data points, minimum, maximum, and sum of terms in the considered window, respectively. Consequently, $R_{LH\_min} \rightarrow 0$ and $R_{LH\_max} \rightarrow 0$ implies that the extrema has high crater and high prominences, respectively. Results show that the robustness of the method in different window sizes.

Key words: Extrema filtering, Extrema finder, High and low extrema, Maxima and minima, Peaks and valleys.

1. INTRODUCTION

One of the main classifications existing in data analysis techniques is whether the method is parametric or non-parametric in its nature [1]. Most popular extrema filtering methods suffer from parametric concerns. Particularly, parametric methods use domain dependent value as detection criteria such as average, standard deviation, prominences of an extrema, and the widths at half prominence of an extrema depend on fewer number of underlying assumptions [1,2,3]. These criteria are based on domain dependent parameters and are therefore valid only for the considered data model or considered conditions in the domain. Thus, majorly parametric methods' accuracy depends on of underlying assumptions [4].

In reality, data capturing, especially within dynamic systems, such as biogas plants, data are produced with various alterations. The most common method for filtering extrema in such situations is to consider the magnitude of prominences (or crater) and the widths at half prominence (or crater). These values are domain dependent: if the domain conditions changes, the selected model generates incorrect results. Therefore, using parametric methods in dynamic processes is a not a reliable approach.

2. METHODOLOGY

Consider a series with n terms where $n-1$ terms agree with $y = c$. If the minimum term is not agree with $y = c$, this situation will create a minimum (valley). Then,

$$S_n = a_{max} * (n-1) + a_{min} \quad (01)$$

where a_{min}, a_{max} and S_n is minimum, maximum, and sum of terms in the series, respectively.

When a valley has very small crater, $a_{min} \approx a_{max}$. Then, eq. (01) can be expressed as:

$$S_n \approx a_{min} * (n-1) + a_{min}$$

$$S_n \approx a_{min} * n \quad ;(< a_{max} * n)$$

$$R_{LH\_min} = (a_{min} * n)/ S_n \ ; 0< R_{LH\_min} \leq 1 \quad (02)$$

If there are negative values, S_n can be zero and eq. (02) becomes invalid. This can be overcome by applying "Min-Max normalization" [5]. Then,

$$a_{i\_New} = a_i - a_{min} + k, \quad (03)$$

where $k > 0$. When $k=1$ eq. (03) becomes:

$$a_{i\_New} = a_i - a_{min} + 1 \quad (04)$$

From eq. (02) and (04),

$$R_{LH\_min} = ((a_{min} - a_{min} + 1) * n)/\sum_{i=1}^{n}(a_i - a_{min} + 1)$$

$$R_{LH\_min} = n/(S_n - a_{min} * n + n) \ ; 0<R_{LH\_min} \leq 1$$

$$R_{LH\_min} = n/(S_n + (1 - a_{min}) * n) \ ; 0<R_{LH\_min} \leq 1 \quad (05)$$

If $R_{LH\_min} \rightarrow 1$, it implies that the a_{min} is very close to the other points (low crater), consequently $R_{LH\_min} \rightarrow 0$ implies that the a_{min} is away from the other points (high crater).

If the maximum is not agree with $y = c$, this situation will create a maximum (peak). Then,

$$S_n = a_{min} * (n-1) \quad + a_{max} \qquad (06)$$

When the prominence of a peak is very small, $a_{max} \approx a_{min}$. Then, eq. (06) can be expressed as,

$$S_n \approx a_{max} * (n-1) \quad + a_{max}$$

$$S_n \approx a_{max} * n \qquad ;(> a_{min} * n)$$

$$R_{LH\_max} = (a_{max} * n)/ S_n \ ; >0 \qquad (07)$$

In eq. (07), R_{LH_max} has no upper limit. Assume $a_{i\_New} = (a_{max} + a_{min}) - a_i$. (08)

Eq. (08) transforms values into their complements (e.g.: the maximum value into the minimum, vice versa). If the R_{LH_max} is the corresponding ratio in relation with high and low peaks identification, then, from eq. (07) and (08),

$$R_{LH\_max} = (((a_{max} + a_{min}) - a_{max}) * n)/ \sum_{i=1}^{n}((a_{max} + a_{min}) - a_i)$$

$$R_{LH\_max} = (a_{min} * n)/ \sum_{i=1}^{n}((a_{max} + a_{min}) - a_i)$$

From eq. (04),

$$R_{LH\_max} = ((a_{min} - a_{min} + 1) * n)/ \sum_{i=1}^{n}((a_{max} - a_{min} + 1 + a_{min} - a_{min} + 1) - (ai - amin + 1))$$

$$R_{LH\_max} = n/((a_{max} + 1) * n - S_n) \ ; 0 < R_{LH\_max} \leq 1 \qquad (09)$$

$R_{LH\_max} \rightarrow 1$ implies that the a_{max} is very close to other points (low prominence). Consequently $R_{LH\_max} \rightarrow 0$ implies that the a_{max} is away from the other points (high prominence).

The ratios intimation with eq. (05) and eq. (09) are totally depend on *n*. Therefore, selecting suitable window size will determine the *n*. Advancing the selected window by one point while applying the detection method (for a selected detection criteria) allows to filter the extrema.

The algorithm was implemented using C++ in .Net 2008 platform and tested with biogas data which were collected online form a biogas plant using NIR spectroscopy for a period of seven months with a frequency of twelve data points per day. Data of each month was considered as a segment, where each segment consists of 350 – 400 data points. Among the different parameters, the H_2 content measured in ppm was selected, which has considerable amount of variations during the process. The proposed detection methods were applied on each segment with different criteria.

3. RESULTS

Figure 1 shows a selected result in relation with one data set for two different detection criteria. In plot (a) and (b), the selected window size is 15 (n = 15) and R_{LH_min} = R_{LH_max} = 0.1 and R_{LH_min} = R_{LH_max} = 0.04, respectively. When the detection criteria are near to zero, it is possible to filter the extrema with low crater or prominence. Since, the criteria are depending on the window size (n), the detection is non-parametric and not depend on any assumption.

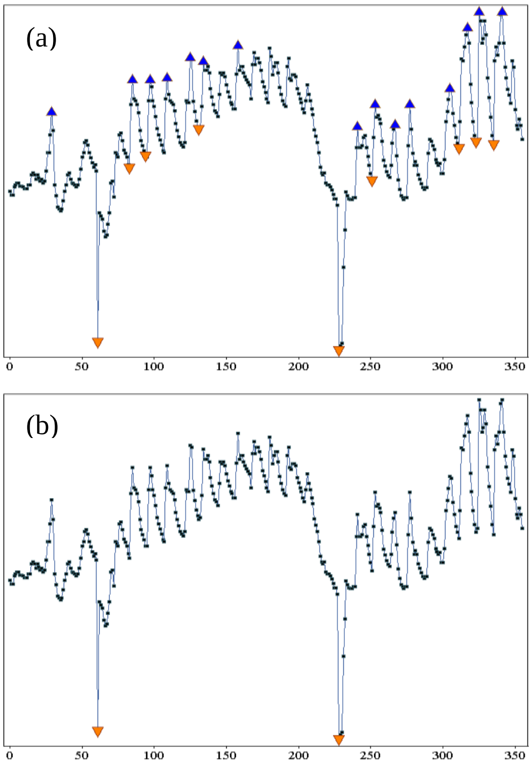

Figure 1: Filtering of low and high extrema using proposed method. Plot (a) shows the extrema filtering for n = 15 and R_{LH_min} = R_{LH_max} = 0.1. Plot (b) shows the extrema filtering for n = 15 and R_{LH_min} = R_{LH_max} = 0.04. For the same window size (n), when the detection criterion is small, extrema with low crater or prominence were filtered.

4. CONCLUSION

The proposed extrema filtering method is can be considered as a non-parametric extrema filtering method which is not depend on domain dependent parameters such as height and width of an extremum. Results prove that the detection is capable of identifying all the extrema with 0%

error. Thus, the proposed method is useful for filtering extrema in signals with very dynamic nature.

5. REFERENCES

[1] L. Györfi, M. Kohler, A. Krzyzak and H. Walk, *"A Distribution-Free Theory of Nonparametric Regression"*, Springer New York. pp. 9-14, 2006:

[2] L. Wasserman, *"All of Nonparametric Statistics"*,Springer New York. pp. 1-2, 2006.

[3] H. Gilgen, "*Univariate Time Series in Geosciences: Theory and Examples",* Springer Berlin Heidelberg, pp. 390-393, 2006.

[4] S.J. Roberts, *"Parametric and non-parametric unsupervised cluster analysis",* Pattern Recognition*"*, vol. 30, no. 2, pp. 261-272, 1997.

[5] J. Han, M. Kamber and J. Pei, "*Data Mining, Southeast Asia Edition: Concepts and Techniques",* Elsevier Science, pp. 71-7, 2006.