

# Analysis of Consistence Level Using New Method of Statistical Transformation Approach in Multi-Spectral Fluctuation Pattern

Melinda<sup>1,2</sup>, Agus Santoso Tamsir<sup>1</sup>, Basari<sup>1</sup>, Dadang Gunawan<sup>1</sup>  
Electrical Department, Engineering faculty, Universitas Indonesia<sup>1</sup>  
Depok, Jawa Barat, Indonesia, 16424  
Electrical Department, Engineering faculty, University of Syiah Kuala<sup>2</sup>  
Banda Aceh, Indonesia, 23111  
melinda41@ui.ac.id, tamsir@ee.ui.ac.id, basari.st@ui.ac.id, guna@eng.ui.ac.id

**Abstract**— In this study, quite new method of statistical approach which is known as TST (tamsir statistical transformation) is introduced. TST is applied in order to obtain a consistence level of multi-spectral fluctuation pattern of HHF (high high-frequency). This is done because the data from the measurement results have large amount of data and have not been consistent yet. Therefore, it is essential to have quite good method to treat the data that can bridge the data processing previously. There are several parameters of the data analysis, which will be discussed, such as: the value of Total-C (total value comparison), consistence of fluctuation (CF), consistence of variance to mean ratio (C-VMR) and consistence of value (CV). Besides, the data are broken down into several groups of data. This is done because it is fruitful to seek the best group that has preferable consistence compare to others. The results obtained show that the statistical approach can determine the consistent results of data grouping for large data size. Moreover, the new approach of TST can accommodate to compute the consistence level of multi-spectral fluctuation pattern.

**Index Terms**—fluctuation, consistence level, statistical approach

## I. INTRODUCTION

Statistical approach is needed to cope with large amounts of data processing. In addition, this approach is valuable to compare the performance of the variable that uses small data sample [1]. Massive research has been done in digital signal processing by using statistical approach, as will be described later. The statistical approach is used as an estimation method of noise level in the digital signal processing by applying the variance and the average value by [2]. This study showed the method for estimating the automatic noise based on a statistical approach to additive white Gaussian noise (AWGN). In addition, [3] developed statistical methods and wavelet as a mathematical tool to seek the impact of noise on the frequency of the vortex flow-meter.

Research conducted by [4] is to develop a comparison function of the average value and variance to quantize the actual changes that occur in the model output and the average value of these variants. It also conducted a comparison between

the classical variant comparison function using the results of 3D plot. Statistical models were also used by [5] for processing the signal processing in the pre-processing method, which was a classification of genomics non-random sample.

Statistical approach more widely used is the analysis of variance to mean ratio (VMR) as has been done by some scholars. Moreover, [6] has conducted research on the application of the mean square error to variant (RM2V) of the sample as a statistical tests and also detects the ability of spectrum in cognitive radio systems. Where, RM2V performance was evaluated by theoretical analysis using Monte Carlo simulation. This analysis found that the sensing RM2V is free of noise variance and complex computing compared to other methods. In addition, the calculation for the VMR has also been used in the computation of radar imaging [7], in order to detect signals that are in noise environment, which is applied fuzzy algorithm for computation.

This research is the developed study that has been done before by [8] and [9]. Research [8] shows the process of extraction feature by using 2D-DWT (two dimension discrete wavelet transform) to some fluctuation patterns, such as: MF (mean fluctuation), HF (high fluctuation), HHF (high high-fluctuation) and HMF (high mean-fluctuation). Then, the output of HHF after extracting by [8] is segmented, which is aims to see more details about the pattern of the fluctuations.

However, some of the processing that have been done before and has also been done by [10] does not gain the maximum results due to the large amount of data obtained. Therefore, researchers need to have an advance stage to obtain a consistence of level data. So, the data will have a high level of consistence level by having VMR values that meet the standards.

Quantitative observation is one of the observations that will be used in this study. Method of analysis used in this study is to examine the value of MF (mean fluctuation), HF (high fluctuation) and HHF (high-high fluctuation) and VMR (variance to mean ratio). In order to accommodate the data processing for obtaining consistence level of data, it is applied a new approach of statistics transformation that is known as

TST (Tamsir statistical transformation). One of reasons is the data used is a multi-spectral fluctuation pattern of HHF. There are some of consistence levels to be analyzed, namely: total-C (total value comparison), consistence of fluctuation (CF), consistence of VMR (C-VMR) and consistence of value (CV).

This study consists of several sections, starting from some application studies of statistical approaches, especially VMR that has been done by researchers. Followed by method of data analysis, this section is going to employ a new statistical approach transformation, which is called TST (Tamsir statistical transformation). Furthermore, the result and analysis are going to discuss in the next section. The last part is the conclusion derived from the research and the stage for future work.

## II. METHOD OF DATA ANALYSIS

In the theory of probability and statistics, the index of dispersion, coefficient of dispersion, relative variance, or variance-to-mean ratio (VMR), such as the coefficient of variation, is the normal size of the probability dispersion of distribution. Typically, it is used to measure whether set of observed event is clustered or dispersed as compared with the standard of statistical model.

The interpretations of dispersion index are shown as follows:

- Case 1: VMR = 1  
If the probability distribution has an average value and variants that provide dispersion index = 1, then this generally occurs in the Poisson distribution
- Case 2:  $0 < \text{VMR} < 1$   
If the index value is dispersion between the values of '0' and '1', then the distribution is called under Dispersed.
- Case 3: VMR > 1  
Usually the negative binomial distribution and geometric distribution have VMR greater than 1. In this case, the data set will be dispersed over.
- Case 4: VMR = 0  
If the constant random variable is in the data, then the distribution will not be dispersed at all.

This research uses TST as a new statistical approach for obtaining the consistency level of data. In the theory of probability and statistics, [1] variance-to-mean ratio (VMR) is a normal measured dispersion of the probability distribution. It is defined as the ratio of the variance to the mean value. However, in the TST, VMR value is the ratio of the square of the standard deviation with the average of value (mean) of fluctuations MF which is shown as follow:

$$\text{VMR} = \frac{\sigma^2}{\mu} \quad (1)$$

Stage of the data analysis can be seen in the schematic diagram below:

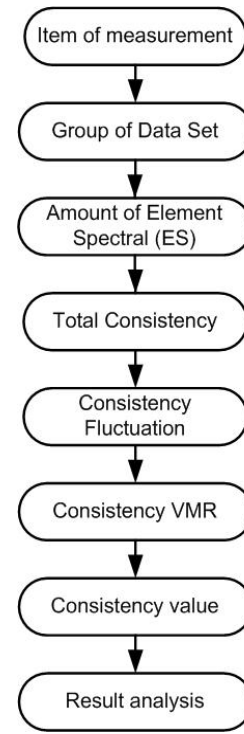


Fig. 1. Analysis Scheme of Consistence Level

### A. Item of measurement

Data sets of measurement items can be seen in Table 1, which shows some of the parameters used at measurement, such as: materials used, the amount of data, frequency and input voltage.

Table 1. Specification items of measurement

Items of parameter	Remarks
Item of measurement	H2O+HCl with (5 molar)
Frequency perturbation	0 & 50 KHz
Input voltage	26 Volt peak to peak
Data set	300 data set
Amount of element spectral	(<1 + >1)
Fluctuation Pattern	HHF

### B. Group of Data Set

Group of data sets are required for grouping data during processing. It aims to simplify the processing and see the similarities of the data measured by the value of consistency. If the material has a high consistency value then the data has a fairly close resemblance between data. In this experiment, the data will be divided into two data sets that are 75 and 100 data sets, for the same ingredients, as shown in Table 1.

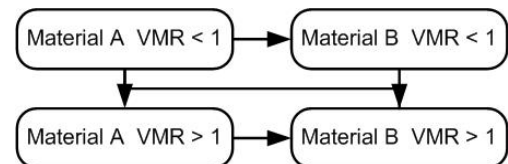


Fig. 2. The concept of a comparison between two similar and dissimilar materials

The concept of comparison between two similar and dissimilar materials is shown in fig.1. This concept is used to observe the data consistency. For example, material used data from 1 to 75 of data set, as well as the material and the same material data from 76 to 150 data sets and so on. In observation of consistency, the 300 data will be divided into 75 and 100 data sets. By using data set of B1 as a reference and then, it will be observed the comparison of data between the data 1 as B1 with others data sets e.g. (B2 and B3). In detail, it is going to discuss in the result and analysis section.

### C. Element Spectral

Amount of element spectral (ES) represents the amount of the data that can be extracted from the spectral frequency. ES is divided into two categories of VMR, namely:  $\text{VMR} < 1$  and  $\text{VMR} > 1$ . Furthermore, ES is required to obtain the final value (FV) of three types of fluctuation patterns used: MF, HF and HHF. Some of the equations used to obtain the final value (FV) for fluctuations HF, MF and HHF and VMR using spectral element (ES) that are described as follow:

$$\text{FV}_{\text{HF}} = \frac{1}{n} \left( \frac{\sum \text{ES}_{\text{HF-C}}}{\sum \text{ES}_{\text{HF-R}}} \right) \quad (2)$$

$$\text{FV}_{\text{HHF}} = \frac{1}{n} \left( \frac{\sum \text{ES}_{\text{HHF-C}}}{\sum \text{ES}_{\text{HHF-R}}} \right) \quad (3)$$

$$\text{FV}_{\text{MF}} = \frac{1}{n} \left( \frac{\sum \text{ES}_{\text{MF-C}}}{\sum \text{ES}_{\text{MF-R}}} \right) \quad (4)$$

### D. Total Consistency

The results for main total consistency consist of some values consistency, namely: consistency of fluctuation (CF), consistency of value (CV) and consistency of VMR (C-VMR). Those values are obtained once we get the results of data. For more details, then the calculation is explained as follows:

- R represents the reference data. Reference data that we use is the first data, which is 1 to 75 data set or 1-100 group of data set R is described as a percentage and the spectral element is within VMR under and above 1.
- C represents the data that is compared to comparable data. Similarly, the value of VMR in C is also less than 1 and above 1.
- Total C determines the total percentage of all data contained in the reference comparison data.
- Consistency of VMR is consistency value of variance to mean ratio of HF pattern.
- Consistence value (CV) is a combination of consistency and consistency of VMR fluctuations.

The total value of C is a combination of the final value for the fluctuation HF, MF and HHF referring to the equation 5, while the value equation for consistency of VMR is shown by equation 6.

$$\text{Total-C} = \frac{|100 - \text{FV}_{\text{MF}}| + |100 - \text{FV}_{\text{HF}}| + |100 - \text{FV}_{\text{HHF}}|}{3} \quad (5)$$

$$\text{Consistency-VMR} = |100 - \text{VMR}| \quad (6)$$

After that, total value of C of fluctuations HF, MF and HHF with equations 5 will be obtained and the total final value for the VMR in Equation 7.

$$\text{CF} = |100 - \text{Total-C}| \quad (7)$$

In order to acquire a comparison between the value of total final value for the data A and B, equation 8 is used as a consistence value (CV), which refers to the value of R and C, as follow:

$$\text{CV} = \frac{\text{Total-C(AB)} + \text{Total-C(BA)}}{2} \quad (8)$$

## III. RESULT AND ANALYSIS

In this study, MATLAB 2013a is used as simulation for data processing and parameters used are referred to Table. 1. The data sets are divided into several sets of data, 75 and 100 from 300 of all data and with perturbation frequency of 0 KHz and 50 KHz. The results obtained by applying STS approach that can be seen in Table 2,3 and 4.

Based on calculations of data, there are some concern of areas, which is the number of spectral elements ( $\mu \cdot \sigma$ ) with value of ( $<1$  and  $>1$ ). Furthermore, the values of the spectral elements are divided into two groups that are comparative value (C) and also reference value (R). These values are applied by two different sets of data but still in the same material. Further calculation is to obtain the total value of several categories, such as: Total-C, the value of consistency for consistence of fluctuation (CF), consistence of VMR (C-VMR) and consistence of value (CV).

Table 2 shows the results of data processing for the consistency of the material ( $\text{H}_2\text{O} + \text{HCl}$ ) per 75 data from a total of 300 data. There are six groups of data, as presented in Table 2. The mean value, standard deviation and VMR value are obtained for all categories. Obviously, for a percentage of the value of total-C, group data of (B1-B2) has the score with 86.20%. Moreover, 99.39% is the greatest of CF values for the data group of (B1-B4). Furthermore, the data group of (B2-B3) has the highest percentage of the C-VMR value with 97.21%, while the highest percentage of CV value is 98.45% is in the group data of (B2-B4).

Similarly, Table. 3 also shows the results of TST statistical data processing, which is similar to Table. 2. However, it has quite different parameters in the grouping of data that is 100 data from a total of 300 data and has three grouping of data as shown in Table 3. In addition, the results for some of the parameters analyzed have different values as shown in Table. 2. The top percentage of total-C values is in the group data of (B2-B3) with 86.16%. Then, percentage of 99.40% is the greatest CF value for the group data of (B2-B3), while the highest percentage of the C-VMR value is in the group data of (B2-B3) with 97.96%.

TABLE 2. The results of data processing for the consistency of the material H<sub>2</sub>O + HCl per 75 Data With disturbing frequency of 0 KHz

Item of measurement	Group of data Set	Amount of element spectral (<1+>1)	Element spectral<1 (%)		Element spectral>1 (%)		TOT-C (%)	CF (%)	C-VMR (%)	CV (%)
	B=75		R	C	R	C				
H <sub>2</sub> O+HCl (Perturbation Freq. 0 KHz)	B1-B2	11.113	31,96	27,95	68,04	58,25	86,20	99,18	95,36	98,22
	B1-B3	10.854	32,19	25,73	67,81	55,20	80,93	98,76	95,09	97,84
	B1-B4	10.884	32,17	25,36	67,83	55,29	80,65	99,39	94,26	98,11
	B2-B3	11.003	32,20	26,56	67,80	54,63	81,19	98,42	97,21	98,12
	B2-B4	11.032	32,18	25,70	67,82	54,38	80,08	99,35	95,76	98,45
	B3-B4	10.773	32,42	26,70	67,58	56,44	83,14	99,15	93,12	97,64
	Mean	10.942	32,19	26,33	67,81	55,70	82,03	99,04	95,13	98,13
	Stadev	127.19	0,14	0,95	0,14	1,44	2,29	0,38	1,38	0,32
	VMR	1.48	0,0006	0,0343	0,0003	0,0372	0,064	0,0014	0,0201	0,0011

TABLE 3. The results of processing the data for consistency Material H<sub>2</sub>O + HCl per 100 Data with disturbing frequency of 0 KHz

Item of measurement	Group of data Set	Amount of element spectral (<1+>1)	Element spectral<1 (%)		Element spectral>1 (%)		TOT-C (%)	CF (%)	C-VMR (%)	CV (%)
	B=100		R	C	R	C				
H <sub>2</sub> O+HCl (Perturbation Freq. 0 KHz)	B1-B2	11.209	31,97	27,22	68,03	57,06	84,28	98,69	96,92	98,24
	B1-B3	11.196	31,75	25,75	68,25	55,25	81,00	99,26	96,28	98,51
	B2-B3	10.900	32,46	27,58	67,54	58,58	86,16	99,40	97,96	99,04
	Mean	11.101	32,06	26,85	67,94	56,96	83,81	99,12	97,05	98,13
	Stadev	174.77	0,36	0,97	0,36	1,67	2,61	0,38	0,85	0,32
	VMR	2.75	0,0041	0,035	0,0019	0,0488	0,0814	0,0014	0,0074	0,0011

TABLE 4. The results of processing the data for consistency Material H<sub>2</sub>O + HCl per 75 Data with disturbing frequency of 50 KHz

Item of measurement	Group of data Set	Amount of element spectral (<1+>1)	Element spectral<1 (%)		Element spectral>1 (%)		TOT-C (%)	CF (%)	C-VMR (%)	CV (%)
	B=75		R	C	R	C				
H <sub>2</sub> O+HCl (Perturbation Freq. 50 KHz)	B1-B2	10.917	31.12	24.81	68.88	55.88	80.68	98.92	95.86	98.15
	B1-B3	10.673	27.28	20.54	72.72	58.86	79.40	61.49	74.21	64.67
	B2-B3	10.802	27.51	19.25	72.49	56.96	76.21	63	73.44	65.61
	Mean	10.797	28.64	21.53	71.36	57.23	78.76	74.47	81.17	98.13
	Stdev	122.07	2.15	2.91	2.15	1.51	2.30	21.19	12.73	0.32
	VMR	1.38	0.162	0.393	0.065	0.039	0.067	6.028	1.996	0.0011

Moving onto the percentage of CV value, the greatest value is also in the group data of (B2-B3) with 99.04%. Surprisingly, the group data of (B2-B3) dominates the highest percentage value for the major parameter of criteria assessment.

As highlighted, Table 4 shows the results of the experimental data that presents the result parameter of data consistency but using a different perturbation frequency of 50 KHz. However, the different set of this perturbation frequency is for obtaining a significant phenomenon that occurs in the parameter level of data consistency. Here, the data is run by 75 data among 300 data. The highest percentage of the consistency value is CF value of 95.86%, followed by 98.15% of CV value. VMR consistency value was 95.86% and the highest percentage of C-total value is 80.68%.

Based on the result analysis that is indicated in Table. 2, 3 and 4, it is perceptible to state that Table. 3 has preferable parameters of consistency level and also dominates the consistence value compared to the result in Table. 2. Furthermore, there are other important things of interest that can be described is by grouping more data, it provides data more consistent compared to a smaller grouping of data.

In addition, the percentages of the greatest value of C-VMR also contain in Table. 3. Nevertheless, the data in Table. 2 looks less so consistent, which is each parameter of consistency data is at different sets of data. In contrast, the data in the Tabel.3 exhibits the percentage of consistency, which has very consistent parameters that all of the highest percentage values are in the same group data indeed.

Hereafter, the analysis for each percentage value of the mean, standard deviation and VMR have qualified specified statistics. Here, the percentages of VMR value are in appropriate values on of all the results that are quite close to 1 or 100 percent. Furthermore, based on Tables 2, 3 and 4, the top mean values are in the consistency of CF value in Table 3 that is equal to 99.12%. Then, the standard deviation percentage value can be the best value if it has the smallest percentage of consistency level that is represented in Table 3. Lastly, VMR values shown in Table 2, 3 and 4 are under 1. This occurrence indicates that the application of statistical methods is meet the new statistical approach of TST transformation.

#### IV. CONCLUSION

There are some fruitful points that can be concluded in this study. Firstly, in order to analyze the experimental data of HHF fluctuation, it is applied a quite new method of statistical approaches that is known as TST (tamsir statistical transformation). Moreover, there are several criteria parameters used to assess the level consistency data such as: the value of total-C (total value comparison), consistence of Fluctuation (CF), consistence of VMR (C-VMR) and consistence of value (CV). In addition, the data is carried out into group of data set, the reason is to make it easier to be analyzed the consistency of level data. Whereas, the result represents essential point that is the larger of the group data is made, the more consistent of parameter value will be gained. Lastly, the percentage value of the mean, standard deviation and VMR have fulfill statistics standard.

#### ACKNOWLEDGMENT

This work was supported by the Ministry of Research, Technology and Higher Education of Indonesia under grant No. 1140/UN2.R12/HKP.05.00/2016.

#### REFERENCES

- [1] Z. Bai, K. Wang, and W. Wong, "The mean – variance ratio test — A complement to the coefficient of variation test and the Sharpe ratio test," *Stat. Probab. Lett.*, vol. 81, no. 8, pp. 1078–1085, 2011.
- [2] P. Jiang and J. Zhang, "Fast and reliable noise level estimation based on local statistic ☆," *Pattern Recognit. Lett.*, vol. 78, pp. 8–13, 2016.
- [3] H. Sun and T. A. Zhang, "Digital signal processing based on wavelet and statistics method for vortex flowmeters," in *Machine Learning and Cybematics*, 2004, vol. 9, no. August, pp. 26–29.
- [4] P. Wei, Z. Lu, W. Ruan, and J. Song, "Regional sensitivity analysis using revised mean and variance ratio functions," *Reliab. Eng. Syst. Saf.*, vol. 121, pp. 121–135, 2014.
- [5] A. H. T. Abdullah K. Alqallaf, "Signal Processing Techniques And Statistics For The Analysis Of Human Genome Associated With Behavior Abnormalities," in *IEEE*, 2007, pp. 36–38.
- [6] Y. Yinghui and L. Guangyue, "Blind spectrum sensing based on the ratio of mean square to variance," *J. China Univ. Posts Telecommun.*, vol. 23, no. 1, pp. 42–48, 2016.
- [7] S. Lo and E. Paso, "Computing the Range of Variance-to-Mean Ratio under Interval and Fuzzy Uncertainty," in *IEEE Proceeding*, 2011.
- [8] Melinda, A S Tamsir, D Gunawan, D Sujana, "Consistency Analysis of Mapping System of Noise Spectral Fluctuations in Multi-Frequency using Two-Dimension Discrete Wavelet Transform (2D-DWT)," in *Asea-Uninet 2016*, 2016, pp. 48–62.
- [9] Melinda, Agus, Dodi, D Gunawan "Implementation of Segmentation Scheme Based On Wavelet Transform In Multi-Spectral Fluctuation Patterns," *Matec Web Conf.*, 2016.
- [10] H Hasan, Melinda, A S Tamsir, D Gunawan "Analysis on Impedance Influence on Multifrequency Capacitive Sensor," in *IC-STAR*, 2016, pp. 2–6.