2004 International Symposium on Nonlinear
Theory and its Applications (NOLTA2004)
Fukuoka, Japan, Nov. 29 - Dec. 3, 2004

Reliable correlation dimension estimation method for analyzing two-phase flow

Akio Suzuki† and Masaki Misawa†

†National Institute of Advanced Industrial Science and Technology (AIST)
1-2 Namiki, Tsukuba, Ibaraki, 305-8564 Japan
Email: akio.suzuki@aist.go.jp

Abstract– The correlation dimension of two-phase flow is found to be of a relatively high dimension, typically greater than four. All the results for two-phase flows, however, were estimated using the method that was originally proposed in 1980's for a low dimensional chaotic system. We first introduce an optimal delay time reconstruction method and a point correlation dimension method into two-phase flow time series analysis. Using artificially generated quasiperiodic time series and real-world two-phase flow time series, we show that applying two methods together largely improve the reliability of analysis.

1. Introduction

The studies of low-dimensional chaos provide new analysis methods: nonlinear time series analysis [1]. The methods turned out to be powerful tools in understanding irregular time series that generated by a low-dimensional nonlinear system. Then it began to apply nonlinear time series analysis to more complex time series from real world, such as human electroencephalogram (EEG) data, financial data, two-phase flow data, and so on.

In the fluid engineering field, many studies have been devoted to achieve flow regime identification, scale-up design, validation of numerical simulation result, and flow control by nonlinear time series analysis [2-4]. However, the estimation of invariants is difficult and sometimes becomes subjective, and the reliability of estimated results have not been sufficiently shown. Therefore, there are some researchers who suspect the reliability of the analysis results. One of the reasons why the estimation becomes unreliable is in the used analysis methods. All the results for two-phase flows we know were estimated using the method that was originally proposed in 1980's for a low dimensional chaotic system. The theory of nonlinear time series analysis is still subject to research, and new techniques and methods have been proposed for a high dimensional system. Practical methods for noisy time series were also studying. Applying these methods, the analysis of two-phase flow time series will become more reliable.

We show that applying an optimal delay time reconstruction method [5] and a point correlation dimension method [6] greatly improves the reliability of analysis [7]. In this paper, the methods and the results of artificially generated time series and various two-phase flow time series are shown. Then it is discussed that what state of two-phase flows can be analyzed.

2. Analysis method and reliability of estimation

2.1. Method for low dimensional system [1]

Nonlinear time series analysis is based on the reconstruction of an orbit in phase space from time series. The most common reconstruction technique is the method of delays. State vectors $z(k)$ in a m-dimensional phase space are formed from the time delayed values of scalar time series $x(k)$:

$$z(k) = \left[x(k),\, x(k+\tau),\, \cdots,\, x(k+(m-1)\tau\right]^T, \tag{1}$$

where m and τ is called an embedding dimension and a delay respectively.

Correlation dimension is a kind of fractal dimension that characterize the geometrical complexity of an orbit. The method proposed by Grassberger and Procaccia (GP method) is widely used. The first step is the calculation of correlation sum $C(r)$.

$$C(r) = \frac{1}{N_{pair}} \sum_{\substack{i=1 \\ |i-j|>w}}^{N} \sum_{j=1}^{N} \Theta(r - \left|z(i) - z(j)\right| \ , \tag{2}$$

$$\Theta(x) = 0 \text{ for } x < 0\,,\ \Theta(x) = 1 \text{ for } x \ge 1\,,$$

where N denotes the overall number of reconstructed vector z in Eq.(1), N_{pair} is the actual number of pairs of points, Θ is the Heaviside step function, and w is the length of the Theiler window. For small distances r, $C(r)$ is expected to scale with a power of r. The scaling exponent defines correlation dimension D_2:

$$C(r) \propto r^{D_2} \text{ , for } r \to 0 \text{ and } N \to \infty\,. \tag{3}$$

Usually one cannot know whether the embedding dimension m is large enough, one must compute the $C(r)$ for increasing m and calculates related $D_2(m)$. If $D_2(m)$ is saturated, this gives the correlation dimension D_2.

2.2. Optimal delay for orbit reconstruction

In the almost all analysis of two-phase flow, the delay τ is determined by the relation between $x(t)$ and $x(t+\tau)$ such as autocorrelation or mutual information. The problem is the optimization between $x(t)$ and $x(t+\tau)$ does not mean the optimization of the other elements such as $x(t+2\tau)$, …, $x(t+(m-1)\tau)$. The more the embedding dimension m becomes large, the more the problem becomes serious. Many methods have been proposed for the estimation of the optimal delay time τ (reviewed in [8]). We evaluate some of them by artificially generated quasiperiodic time series, and decide to use the average displacement method [5] that gives the most robust results.

The average displacement method measures the average distance $S(\tau)$ of the reconstructed vector $z(k)$ in Eq.(1) from the main diagonal of the m dimensional space.

$$S(\tau)=\frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{j=1}^{m-1}\left(x(i+j\cdot\tau)-x(i)\right)^{2}} \tag{4}$$

The main problem of this method is that there is no theoretical reason for the τ determining rule. Rosensteine *et al.* choose τ as suitable time delay, where the slope of $S(\tau)$ has dropped to 40% of the value at $\tau \to 0$. We examine the Rosensteine's criterion using quasiperiodic time series. The results show that it tends to give rather lower dimensions. We examined some criterions, and found that the better result is obtained at first minima of the slope of $S(\tau)$.

Importance using proper delay time is shown in Fig.1. This figure shows the local slopes of $\log C(r)$ of quasiperiodic time series. The power law property in Eq.(3) appears a straight line parallel to the x-axis, the so-called scaling region. The left panels are the plots using τ determined by the mutual information. The right panels are the plot using constant τ determined by the average displacement method and our proposed criterion. If the reconstructed orbit is low-dimensional, both methods work well (Fig.1a). However, if the orbit is high-dimensional, the problem appears clearly (Fig.1b). In the case of using constant τ, the local slope becomes oscillational. The scaling regions are narrow and unclear, so that it is difficult to do reliable estimation. In the case of using optimal τ, the scaling regions are wide and clear, and also the saturation of the local slopes for increasing m can be seen. Using optimal time delay improves the reliability of the correlation dimension analysis of a high dimensional orbit.

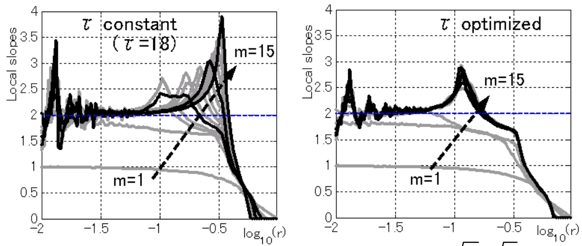
2.3. Point correlation dimension method

Point correlation dimension is a locally defined fractal dimension at a specific point on an orbit. This method was proposed by Skinner *et al.* [6] for the analysis of physiological time series such as EEG. The procedure is same as the GP method except to calculate a local correlation sum $C_i(r)$ at a reference point $z(i)$.

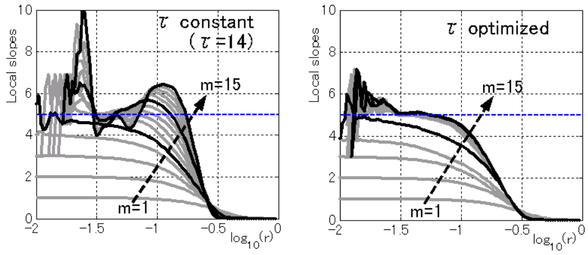
$$C_i(r)=\frac{1}{N_{pair}}\sum_{\substack{j=1\\ |i-j|>w}}^{N}\Theta\left(r-\left|z(i)-z(j)\right|\right)\ , \tag{5}$$

$$C_i(r)\propto r^{PD_2(i)}\ \text{, for } r\to 0 \text{ and } N\to\infty . \tag{6}$$

The scaling exponent of $C_i(r)$ gives the local point correlation dimension $PD_2(i)$ at each reference point, and in practice may vary over the entire orbit (see Fig.2d, right panel is the distribution of $PD_2(i)$). Averaging $PD_2(i)$ over all reference point yields the (averaged) point correlation dimension PD_2, which gives the global description of the geometrical structure of an orbit.



(a) 2-quasiperiodic time series ($\omega=\sqrt{2},\sqrt{3}$)



m	2	3	4	5	6	7	8	9	10	11	12	13	14	15
τ	36	20	22	18	14	12	11	9	9	8	7	7	6	6

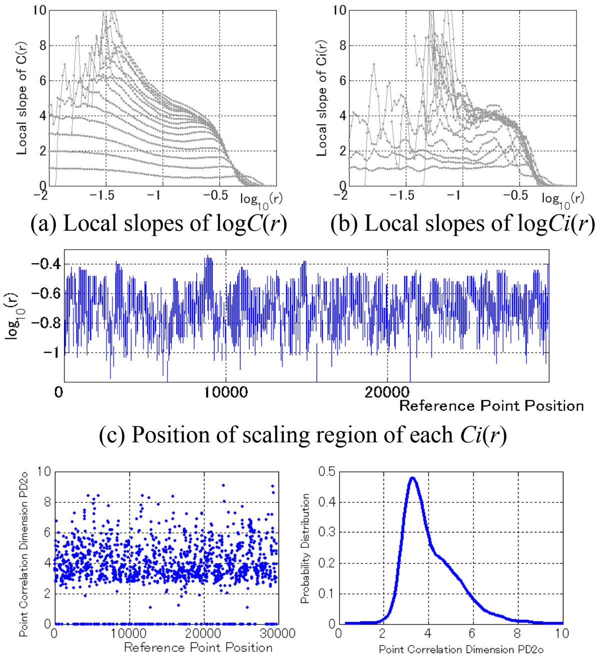
(b) 5-quasiperiodic time series ($\omega=\sqrt{2},\sqrt{3},\sqrt{5},\sqrt{7},\sqrt{11}$)

Fig.1 Local slopes of correlation sums by τ constant and τ optimized reconstruction (N=300000)

We have failed to analysis two-phase flow time series by the GP method many times, even if optimal time delay is used. However, many of those failed time series can be analyzed by the point correlation dimension method. In order to explain this, we investigate scaling regions using two-phase flow time series. Fig.2a, 2b is the local slopes of $\log C(r)$ and $\log C_i(r)$ for a void fraction time series. In the local slope of $\log C(r)$, clear flat region is not appeared and slopes do not saturate with increasing m (Fig.2a). On the other hand, many of the $\log C_i(r)$ show a clear and saturating scaling region. A good example is shown in Fig.2b. The positions of the scaling region of $C_i(r)$ are shown in Fig.2c. The positions and lengths of the scaling regions are different where the reference points are located on an orbit. This is the reason of why PD_2 well manage to real world time series. Because $C(r)$ is the sum of $C_i(r)$, the scaling region appears only in the overlapped part of each $C_i(r)$'s scaling region. If an orbit is high dimensional and its structure is complex, the variance of scaling regions becomes large and the overlap of scaling regions becomes narrow.

2.4. Reliability of Estimation

Fig.3 shows estimated dimensions for quasiperiodic time series (dimension D=1~5) versus number of data points N. Fig.3a is the estimation results using the common method for low dimensional chaos: constant time delays were determined by first minima of mutual information and the GP method was used. Fig.3b is the results using the proposed method: optimal delay reconstruction and point correlation method. For low dimensional time series (orbit dimension D=1, 2, and 3), both procedures give good results. However, for high dimensional time series (D>3), the common method gives the results with large error especially when the number of data point N is small. On the other hand, the proposed method gives the results with relatively small error. Estimated dimensions converge rapidly, so that one can obtain a proper dimension if N is small. Even if N is sufficiently large, correct dimensions could not be obtained. Steady errors remain. It is seems to be caused by the geometrical effect of a reconstructed orbit structure (the constant delay reconstructed orbit of quasiperiodic time series is not homogeneous at the scale of the scaling region). For example, if the reference point located at the edge of an orbit, scaling property is disturbed and underestimation is induced. If the neighbor orbits of the reference point are strongly distorted or folded, overestimation is induced. The GP method with optimal time delay reconstruction gives better results than the standard method, however, not better than the proposed method.

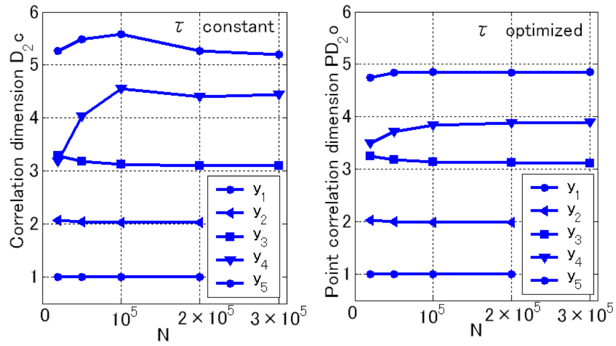


(a) Local slopes of log$C(r)$ (b) Local slopes of log$Ci(r)$

(c) Position of scaling region of each $Ci(r)$

(d) Estimated PD2(i)

Fig.2 Local slopes of $C(r)$ and $Ci(r)$, and estimated PD2(i) (Void fraction time series of slug flow, N=30000)



(a) D_2c: correlation dimension by constant τ (b) PD_2o: Point correlation dimension by optimized τ

Fig.3 Estimated dimensions for quasiperiodic time series y_d (d =1～5) versus number of data points N

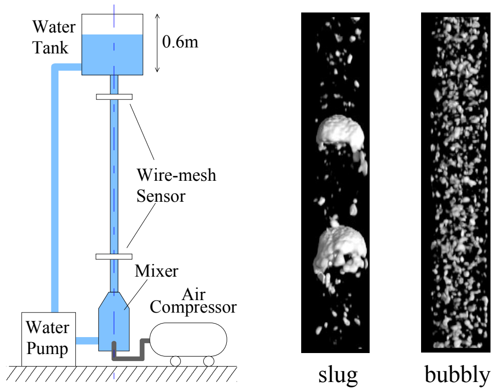


Fig.4 Schematic of experimental system and bubble distribution

3. Two Phase Flow Time Series Analysis

Analysis results of air-water two-phase flow in a vertical pipe are shown. The reliability of the estimated results is tested by a surrogate data method. We choose the null hypothesis of the test that the original time series is generated from a linear stochastic process possibly undergoing a nonlinear static transform. We then randomly shuffled the original time series and create surrogate data sets that have same spectrum and distribution as the original time series (we use free nonlinear time series analysis package TISEAN [1]).

3.1. Experimental system and measurements

The schematic of the experimental system is shown in Fig.4 [4]. Gas provided by an air compressor is controlled by a regulator, and water flow is controlled by a pump. Air-water two-phase flow in a circular tube is measured by a wire-mesh sensor installed at the exchangeable vertical test section. The wire-mesh sensor measures cross-sectional void fraction [9]. A 42mm inner diameter and 2.1m length test section was used. Void fractions were recorded for 30 seconds at sampling rate of 1000Hz. (number of data point N=30000). In order to remove sensor noise, FIR low pass filter is applied to the time series (typical cut-off frequency is about 100Hz).

3.2. Analysis Results

3.2.1 Comparison of analysis results

Analysis results for slug flow in various flow conditions and their surrogate test results (significance level α=0.05) are shown in Table.1. PD_{2o} is a point correlation dimension using optimal delay reconstruction, and D_{2c} is a correlation dimension using constant delay reconstruction. Almost all D_{2c} fail the surrogate test, however, all PD_{2o} pass the test. These results show that the newly introduced methods largely improve the reliability of real-world high dimensional time series analysis.

3.2.2 Bubbly flow and slug flow

Analysis results of bubbly and slug flow and their surrogate test results (α=0.05) are shown in Table.2. PD_{2o} of slug flow pass the test, except for the one case. This indicates that the degree of freedom sufficiently lowered and the dynamics of slug flows is ruled by the motion of huge bubbles. On the other hand, all PD_{2o} of bubbly flow fail the test. This indicates that the degree of freedom in bubbly flow does not sufficiently lowered because of the motions of small bubbles.

3.2.3 Measurement position

Slug flow time series measured at the different position of the test section were analyzed. The wire-mesh sensors are installed at 0.4m (upstream sensor) and 1.7m (downstream sensor) above the air/water mixer. PD_{2o} and their surrogate test results (significance level α=0.1) are shown in Table.3. PD_{2o} of the upstream time series are greater than PD_{2o} of the downstream time series. Only half of PD_{2o} of the upstream time series are pass the surrogate tests, while all PD_{2o} of the downstream time series pass. These results imply that the order of flow has not yet generated at upstream sensor position and the degree of freedom of flow is not sufficiently lowered. In other words, if a reliable PD_{2o} is estimated, it is evidence that deterministic order rules the flow.

4. Conclusion

We newly introduce two methods into the two-phase flow analysis in order to obtain a reliable estimation result: an optimal delay time reconstruction [5] and a point correlation dimension method [6].

In practical nonlinear time series analysis, the quality of a reconstructed orbit has not been sufficiently paid attention. Because one always obtains only a finite number of noisy time series from real world, an improper delay time embedding yields unreliable estimation results. Therefore, optimal delay time selection for each embedding dimension is essential for the analysis.

A point correlation dimension method is a practical analysis method. Disadvantage is point correlation dimension is estimated from only N interpoint distances (N is a time series length), while the GP method uses N^2 points. Therefore, its estimation becomes more difficult. In fact for hyperspheres whose dimension is greater than 7, our analysis results showed large errors.

Judd has proposed the new model of correlation sum and maximum-likelihood estimator [10]. The Judd's method has the ability to estimate correct correlation dimension even if a number of time series is small or attractor is high-dimensional [8]. Applying the Judd's method to local correlation sums, it will be possible to obtain more correct point correlation dimensions.

The proposed method applied to various two-phase flow time series. The comparison of estimated results and their surrogate test results demonstrates the high reliability of this newly applied analysis method. Many unsolved problems still remain: minimum data point number requirement, the stability of results (whether almost constant dimensions are obtained from a fixed flow condition), and so on. We plan to further investigation.

Table 1 Estimated dimensions of slug flow time series

Jl (m/s)	Jg (m/s)	D2c	test	PD2o	test	Jl (m/s)	Jg (m/s)	D2c	test	PD2o	test
0.00	1.20	4.32	*fail*	4.37	pass	2.41	1.20	5.88	*fail*	5.71	pass
0.00	2.41	8.80	*fail*	5.50	pass	2.41	2.41	6.87	*fail*	5.87	pass
0.00	3.61	9.58	*fail*	4.87	pass	2.41	3.61	7.10	*fail*	6.23	pass
0.00	4.81	8.57	*fail*	4.93	pass	2.41	4.81	9.99	*fail*	5.92	pass
0.00	6.02	9.82	*fail*	5.60	pass	2.41	6.02	9.23	*fail*	5.76	pass
0.60	0.12	3.82	*fail*	3.67	pass	4.81	1.20	7.76	*fail*	6.02	pass
0.60	0.36	4.20	pass	4.09	pass	4.81	2.41	9.75	*fail*	6.30	pass
0.60	0.60	4.63	pass	4.37	pass	4.81	3.61	7.03	*fail*	6.46	pass
0.60	1.20	5.41	*fail*	5.02	pass	4.81	4.81	9.47	*fail*	6.61	pass

Table 2 Slug and bubbly flow

flow type	Jl (m/s)	Jg (m/s)	PD2o	test
slug	0.00	0.05	2.59	pass
slug	0.00	0.30	3.61	pass
slug	0.60	0.05	2.22	pass
slug	0.60	0.30	3.85	*fail*
bubbly	0.00	0.05	7.12	*fail*
bubbly	0.00	0.30	6.34	*fail*
bubbly	0.60	0.05	7.58	*fail*
bubbly	0.60	0.30	7.03	*fail*

Table 3 Measurement position

		downstream		upstream	
Jg (m/s)	Jl (m/s)	PD2o	test	PD2o	test
0.4	1.2	3.1	pass	5.5	*fail*
0.7	1.2	3.7	pass	6.1	*fail*
1.1	1.2	3.7	pass	6.5	*fail*
1.5	1.2	4.1	pass	6.5	*fail*
1.8	1.2	4.5	pass	6.1	pass
2.2	1.2	5.1	pass	6.0	pass
2.6	1.2	4.9	pass	6.0	pass

Acknowledgments

This study was supported by Institute of Applied Energy (IAE) for fiscal years 2001 to 2003.

References

[1] H. Kantz and T. Schreiber, "Nonlinear Time Series Analysis", Cambridge University Press, 1997.

[2] C. M. Van den Bleek, M. O. Coppens, J. C. Schouten, "Application of chaos analysis to multiphase reactors", Chem. Eng. Sci., Vol.57, pp.4763-4778, 2002.

[3] F. Johnsson, *et.al.*, "Characterization of fluidization regimes by time-series analysis of pressure fluctuations", Int. J. of Multiphase Flow, Vol.26, pp.663-715, 2000.

[4] M. Misawa, *et.al.*, "Nonlinear Characteristics of Gas-Liquid Two-phase Flow and Verification of Extended Two-Fluid Model", Proceedings of International Conference on Multiphase Flow, 2004.

[5] M. T. Rosenstein, *et.al.*, "Reconstruction expansion as a geometry-based framework for choosing proper delay times", Physica D, Vol.73, pp.82-98, 1994.

[6] J. E. Skinner, *et.al.*, "The point correlation dimension: performance with nonstationary surrogate data and noise", Integrative Physiological and Behavioral Science, Vol. 29, pp.217-234, 1994.

[7] A. Suzuki, *et.al.*, "Reliable estimation of dynamical invariant of two-phase flow: Point correlation dimension analysis of void fraction time series", Proceedings of International Conference on Multiphase Flow, 2004.

[8] A. Galka, "Topics in Nonlinear Time Series Analysis", World Scientific Publishing, 2000.

[9] H. M. Prasser, *et.al.*, "A new electrode-mesh tomograph for gas-liquid flows", Flow Measurement and Instrumentation, Vol. 9, pp.111-119, 1998.

[10] K. Judd, "Estimating dimension from small samples", Physica D, Vol.71, pp.421-429, 1994