SYMBOL STATISTICS: A NEW TOOL FOR UNDERSTANDING MULTIPHASE FLOW PHENOMENA

C.S. Daw
Oak Ridge National Laboratory
Oak Ridge TN 37831-8088

C.E.A. Finney, **K. Nguyen**
University of Tennessee
Knoxville TN 37996-2210

J.S. Halow
Federal Energy Technology Center
Morgantown WV 26507-0880

ABSTRACT

We discuss data symbolization as a tool for identifying temporal patterns in complex measurement signals. We describe the basic concepts involved and illustrate their application for the analysis of gas-bubble injection data. Specific issues addressed include selection of symbolization parameters, construction of symbol-sequence histograms, and statistical characterization and comparison of these histograms. We demonstrate that symbol-sequence statistics can reveal unique information about deterministic patterns. Such information may be useful for developing flow diagnostics and comparing computational models with experiments.

NOMENCLATURE

H_S	Modified Shannon entropy
$N_{\rm obs}$	Number of non-zero-frequency sequences
p_i	Observed relative frequency (probability) of sequence i
T	Euclidean norm
$T_{\rm irr}$	Time irreversibility index, based on Euclean norm of forward- and backward-time symbol-sequence histograms
χ^2	Modified chi-square statistic
$\chi^2_{\rm irr}$	Time irreversibility index, based on chi-square difference of forward- and backward-time symbol-sequence histograms

INTRODUCTION

Experimental dynamic measurements of multiphase flows can be extremely complex. Analysis of such measurements can be problematic, especially when only limited amounts of data are available or when the data are contaminated with high levels of noise. In this context, well-known tools for temporal data analysis such as correlation functions and Fourier transforms are often inadequate for distinguishing different dynamic states.

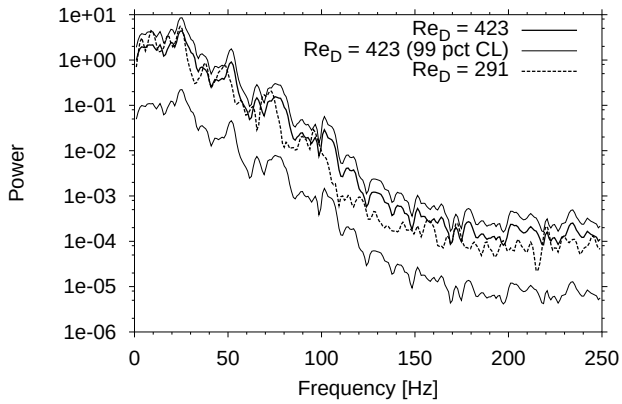

Figure 1: Fourier power spectra for two injection-rate conditions. Upper and lower curves indicate 99% confidence limits.

Figure 1 illustrates a comparison of Fourier power spectra for two different flow conditions in a bubble-injection experiment. In both cases, the pressure fluctuations in an injection nozzle were measured with a high-speed transducer when the nozzle was immersed in water at ambient conditions. Details of the apparatus and measurement process are described in the Experimental section below. Our main point here is to note that although the gas-injection rates differed by more than 30%, the Fourier spectra of the pressure fluctuations are indistinguishable within measurement uncertainty, indicated by the error bands for the lower-flow case. The principal difficulty in using Fourier analysis under such circumstances is the characteristic spectral broadening of turbulent or near-turbulent conditions. Thus we are motivated to find other analytical tools that are less affected by such broadening.

Our main focus in this paper is to consider a new approach

for analyzing complex measurements known as data symbolization. Briefly, data symbolization transforms an original series of measurements into a limited number of discrete symbols. The resulting symbol series is then analyzed for nonrandom temporal patterns. For our purposes, we are specifically interested in identifying and measuring repeating unstable patterns which continue to come and go even when the flow parameters (*i.e.*, the system boundary conditions) are kept fixed. Such unstable patterns are the source of spectral broadening. In the language of nonlinear dynamics, we use symbol statistics to identify characteristic unstable periodicities.

Even though it seems counter-intuitive that discretization can lead to improved information extraction, symbolization helps to separate unstructured small-amplitude noise from significant large-amplitude trends. Several other investigators have demonstrated successful application of this basic approach to complex time-series data [1,2,3,4,5,6]. We build on this previous work and describe our own unique methods for visualizing and characterizing the resulting information (see also [7], this session). We begin by describing our symbolic transformation process and how we quantify the resulting patterns. After defining our symbolization procedures, we illustrate the usefulness of our methods using pressure fluctuation measurements from a bubble-injection experiment.

SYMBOLIC TRANSFORMATION AND SYMBOL-SEQUENCE HISTOGRAMS

Transformation of the original measurements into symbolic form is based on partitioning the measured variable range into a discrete number of regions, as illustrated in Fig. 2. For illustration purposes, we confine ourselves here to the simplest possible partition involving a division of the data range into two parts (a binary partition). Original measurements which are above the designated cutoff value are assigned a symbolic value of 1, and measurements below the cutoff are assigned a symbolic value of 0. For the moment, we postpone discussion of how the cutoff value is selected and arbitrarily choose a cutoff located near the mean observed value. Clearly, the detailed features of the resulting symbol series will depend on the specific choice of cutoff location.

Higher levels of symbolization can be constructed by using more partitions. Greater degrees of partitioning can be useful where one is concerned with observing measurement details which are small relative to the overall range. It is also possible to base the symbol transformation on differences between successive measurements. This latter approach can be useful when the observed process appears nonstationary or has very long-time-scale variations. We do not use the difference-based symbolization here but refer the reader to Kurths *et al.* [2] for further discussion.

Once the original data have been converted to symbolic form,

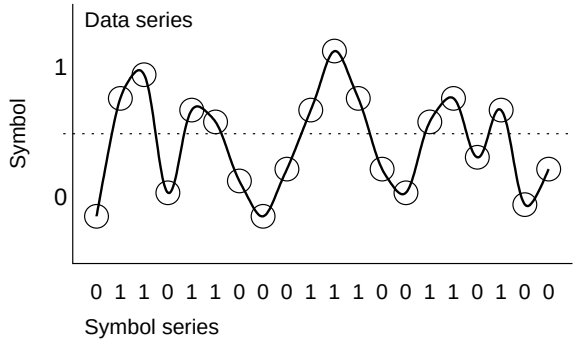

Figure 2: Illustration of symbolization. Each value in the symbol series at the bottom corresponds to the point directly above in the data series.

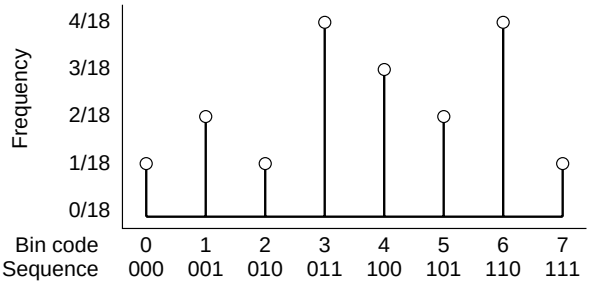

Figure 3: Example symbol-sequence histogram for the symbol series shown in Fig. 2. Note that sequences 011 and 110 are most frequent.

our objective is to look for characteristic temporal patterns. One particularly useful approach is to select a standard number of sequential measurements as comprising a symbol sequence ("word", in the symbolic-dynamics literature), and then observe the relative frequency of occurrence for all possible sequences as one indexes the symbol series through time. For example, if we consider a binary symbol series and look for the occurrence of symbol sequences consisting of 3 consecutive measurements, there will be 8 (2^3) possible distinct sequences which can be observed. By counting the number of times each sequence occurs and dividing by the total number of observed sequences, the observed dynamics can be described in terms of a symbol-sequence histogram.

We find it convenient to depict symbol-sequence histograms as illustrated in Fig. 3. On the horizontal axis, we designate each possible sequence as a single number which is the base-10 equivalent of the binary pattern and is termed the sequence code. For example, the symbol sequence 101 is represented by the number 5. We extend this concept to different partitions and sequence lengths by transforming in the same fashion from the appropriate base numbering system (*e.g.*, we use base 4 when a partition

of 4 is used). This convention makes it very convenient to refer to any specific sequence for purposes of visualization or computation, no matter what the sequence length or level of partition. With the vertical axis in Fig. 3 indicating frequency, we can now directly view the relative importance of all possible symbol patterns in the measured signal simultaneously in two dimensions. This kind of simplification can be useful for high-dimensional processes where phase-space depictions of the dynamic trajectory are not visually resolved in even three or four dimensions.

CHARACTERIZING AND COMPARING SYMBOL-SEQUENCE HISTOGRAMS

In addition to providing a visual representation of the dynamic patterns, symbol-sequence histograms provide the basis for quantitative statistics. Three statistics we have found to be especially useful are a modified Shannon entropy, the Euclidean norm (T statistic), and a modified χ^2 statistic.

We define the modified Shannon entropy as

$$H_S = -\frac{1}{\log N_{\rm obs}} \sum_i p_i \log p_i \ , \qquad (1)$$

where p_i is the probability (normalized observed frequency) of the ith symbol sequence (that is, the symbol sequence designated by index i), and $N_{\rm obs}$ is the number of possible sequences which are actually observed in the data. Note that Eq. 1 is different from the standard definition of Shannon entropy because it has been normalized with the Shannon entropy for a completely random process (one in which all sequences are equally probable). Thus, when employing this statistic, it is useful to define the partition cutoff such that it divides the observed data into regions of equal probability (for a binary partition, this is the data median). This type of normalization means that the modified Shannon entropy will converge to 1 as the data approach true randomness. Likewise, for data that exhibit complex deterministic structure (*e.g.*, unstable or noisy periodicities), this statistic will be less than 1. The motivation for using $N_{\rm obs}$ rather than the total number of possible sequences in the summation is so that we do not bias our statistics because of the finite size of our data sets; that is, we do not observe some sequences simply because we did not observe long enough. Assuming that those frequencies are truly zero will bias H_S to be lower than it truly is.

Both the Euclidean norm and modified χ^2 statistics are useful for comparing different histograms. The Euclidean norm is defined as

$$T_{AB} = \sqrt{\sum_i (A_i - B_i)^2} \ , \qquad (2)$$

and the modified χ^2 is defined by

$$\chi^2_{AB} = \sum_i \frac{(A_i - B_i)^2}{(A_i + B_i)} \ , \qquad (3)$$

where A_i and B_i are the individual sequence probabilities for sequence i for histograms A and B, as above.

Both of the above statistics are generated by differencing the frequencies of individual sequences for different data sets. When the frequency differences are large, the resulting statistics will also be large. Likewise, large values for the statistics imply that the dynamic patterns in the data sets are different. The Euclidean norm statistic is based on the idea that each symbol sequence histogram can be thought of as a vector in multi-dimensional space, where the number of dimensions is the number of possible unique sequences. Thus the magnitude of the vector difference between histograms should provide a comparison of the histograms. The modified χ^2 statistic is derived from the standard χ^2 statistic, where univariate measurement frequencies have been replaced with sequence frequencies.

SELECTING SYMBOLIZATION PARAMETERS

There is not yet a general rule for "optimally" selecting an appropriate symbolizing partition. Rigorously, it is recognized that the output of a deterministic system can only be transformed into a unique symbol series by using a so-called generating partition. However, choosing generating partitions is almost always impractical because generating partitions do not exist when any noise is present [8]. Based on trial experience, we suggest that a generally good choice is to partition the data range into 2–10 equiprobable regions. Symbol alphabets greater than 10 tend to capture too much detail at the expense of global patterns. Binary partitions are useful for extracting deterministic patterns when high noise is present, but these can miss distinctive intermediate-scale features useful for diagnostics. Equiprobable partitioning allows distinction between stochastic and deterministic structure because nonrandom patterns show up as "peaks" in the symbol sequence histogram.

In general, construction of symbol sequences also requires selection of an inter-symbol interval and an overall sequence length. In some sense, choosing these parameters is analogous to choosing an appropriate embedding dimension and lag for time-delay embedding [9,10]. Inter-symbol interval is specifically the fixed time interval between successive symbols in each symbol sequence. For measurements that are naturally discretized in time (*e.g.*, when we record events such as the time between bubbles), it is usually sufficient to use one natural period for inter-symbol interval.

Continuously measured signals are more difficult because one has to establish an inter-symbol interval which does not result in excessive repetitions of the same symbol and which also is not so large that it misses important high frequencies. We have found empirically that an inter-symbol interval between 0.1 – 0.5 of an average cycle time for the time series of interest is generally a good choice. By average cycle time, we refer to the average time interval between crossings of the mean value of the time

series (being consistent to consider only upward or downward crossings).

The number of successive symbols (sequence length) combined with the inter-symbol interval determines the sequence time span. In the example experimental data considered here, there are two important considerations affecting our choice of sequence length. For sequence time spans that are less than or equal to an average cycle time, we are typically observing the detailed patterns of individual bubbles. Of course the details are enhanced if we use smaller inter-symbol intervals. Longer time spans reveal patterns associated with successive bubble trains. Thus by choosing larger inter-symbol intervals and long sequence lengths, we can focus more on the dynamics of bubble interactions.

We caution that the above procedures are based on heuristic arguments and empirical experience. It is likely that in some contexts other approaches would be more fruitful. We simply offer them as suggestions in the absence of more rigorous guidelines.

EXPERIMENTAL

Our air injection experiment consists of an injection nozzle immersed in a 27-cm tall by 29-cm diameter cylinder filled to a depth of 22-cm with ambient-temperature water. The experimental apparatus is pictured in Fig. 4. Compressed air is supplied from a cylinder and passed through a rotameter through a 0.15-cm diameter line before flowing into either a 0.14- or 0.45-cm-diameter tube and straight-bore injection nozzle. Air flow is metered at rates between 10 and 2000 mL/min and held constant for each data set.

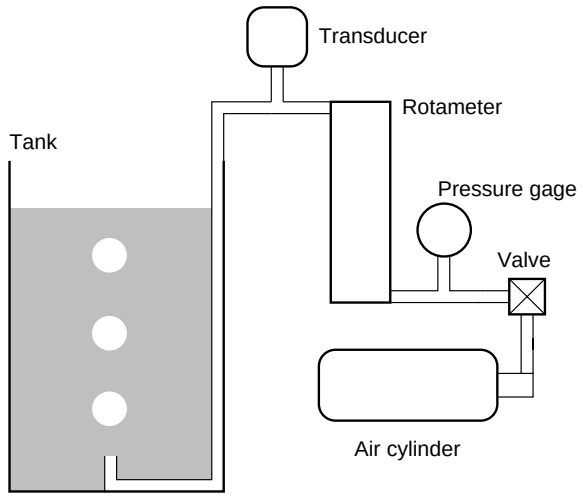

Figure 4: Schematic of experimental apparatus.

High-speed pressure measurements were made at selected air flows using a sensor attached directly to the air line going to the nozzle. Pressure signals were high-pass filtered at 0.1 Hz, anti-alias filtered, and then digitally recorded at 500 Hz. We always waited a minimum of 30 seconds after flow adjustment before starting to record. For the results presented here, we typically considered contiguous record segments of 6000 points.

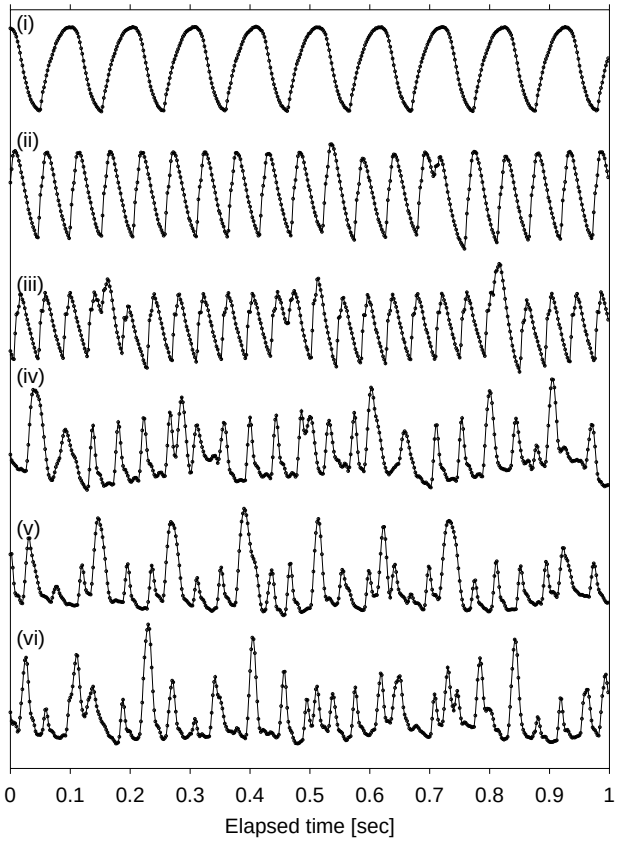

Figure 5: Example time series for the pressure transducer at six injector flows (Re_D): (i) 13, (ii) 63, (iii) 100, (iv) 216, (v) 291 and (vi) 423.

Typical time series for the transducer in the nozzle itself are illustrated in Fig. 5. Flows are reported Reynolds number based on the injector nozzle diameter, and amplitude scales have been adjusted to display similar ranges for all segments. At very low flows the pressure signal oscillates in a purely periodic pattern as each bubble is formed and then released. We visually observe at these low flows that the pressure signal reaches a local minimum just at the moment of bubble release. The pressure then rises sharply as the meniscus collapses back into the mouth of the nozzle and reaches a peak as the next bubble begins to form. The pressure falls continuously as the new bubble grows until the moment of release, and then the cycle repeats. At very low flows the pressure rise and fall appear to be essentially symmetric, that is, they are mirror images.

As the air flow is increased, we observe bubbles whose pressure rise and fall are more asymmetric, suggesting that the processes for bubble growth and release are becoming distinctly different. Occasionally, some forming bubbles become unstable and break apart into two bubbles. This "stutter" then leads to several successive bubbles whose appearance deviates from the norm until the oscillations damp out and the original pattern resumes. Stuttering continues to increase as flow increases. Based on previous studies, we expect that these stutter patterns include bifurcations and the onset of deterministic chaos [11,12,13,14]. However we have not yet studied this system in detail and cannot ascribe the stutters to a specific type of instability (such as a period-two bifurcation or a Pomeau-Manneville intermittency). We conjecture that the behavior in water will be more complex because of the relatively low viscosity (low damping) compared with liquids such as glycerine.

EXPERIMENTAL SYMBOL-SEQUENCE HISTOGRAMS

To begin our symbolization examples, we consider symbol-sequence histograms for the same two measurement series that produced the Fourier power spectra illustrated above. Fig. 6 illustrates the corresponding histograms produced using a binary partition (with equiprobable division), a sequence length of 6, and an inter-symbol interval of half the average cycle time (as defined above). Note the larger peaks at decimal sequence numbers 21 and 42 for the lower-flow replicate data sets (*i.e.*, from two nonconsecutive repeat experiments). In binary form, these two peaks are 010101 and 101010, respectively. As explained below, this oscillation pattern represents the frequency of occurrence of three successive small bubbles. It appears that this pattern is significantly more common at the lower flow than at the higher flow. Likewise, several peaks associated with irregular bubble sequences such as 010111 (decimal number 23) are less common for the lower flows.

Figure 7 illustrates another comparison of the same data sets using a higher-resolution, four-level partition. In this case the peaks associated with the sequences 030303 (decimal number 819), 020303 (decimal number 563), and 303030 (decimal number 3276) are especially prominent for the lower flows.

To better appreciate the physical meaning of the above histogram features, it is helpful to plot example time-series segments as shown in Fig. 8. Here we see two example segments corresponding to occurrences of the 030303 sequence (bounded by vertical dashes) and measurements just prior to and just following this pattern. The target sequence itself appears to represent a succession of three nearly identical (but still distinctive) bubbles. Interestingly, an unusually large bubble typically precedes the target sequence, and a succession of small bubbles with gradually increasing pressure occurs afterward. This collective chain of large and small bubbles clearly represents a deterministic pattern that is continually revisited (there were 53 occurrences of the pattern in this data set alone). The dominance of this pattern also suggests that it reflects important aspects of the physics behind bubble formation and release.

From a nonlinear-dynamics perspective, we interpret repeating patterns like the above as indicating characteristic unstable orbits (unstable fixed points on a Poincaré map). This interpretation is supported by the remarkable degree to which the sequence details are repeated in each occurrence and by the fact that the precursor and postcursor patterns are so consistent. In regard to the latter, we suspect that the precursor events represent the action of a stable manifold, and the postcursor events reflect the action of an unstable manifold.

Figures 9 and 10 are examples of other unstable repeating patterns we have observed. In Fig. 9 we see a 202031 sequence (4 partition) associated with large bubbles that come both before and after two small bubbles. We observe this sequence occurring relatively frequently among the three-bubble sequences at the same flow condition used for Fig. 8. In Fig. 10 we see a 021030 sequence associated with a distorted bubble which momentarily disrupts the more periodic pattern at low air flow. This type of pressure fluctuation appears to be caused by a bubble splitting before it is released from the nozzle. Such splitting appears to contribute toward destabilizing the originally periodic bubble regime encountered at low flow.

SYMBOL-SEQUENCE STATISTICS

Using Eqs. 1–3, we can quantitatively compare the symbol-sequence histograms in Fig. 6. Results for Eq. 2 are summarized in Table 1.

The T statistic provides a quantitative method for comparing overall differences between symbol-sequence histograms (*i.e.*, differences in the frequencies of all possible patterns are considered). In Table 1 we observe that whereas there is some variation between the flow replicates, such variations are significantly smaller than the variation between flows. These statistics appear to confirm that symbol sequences can effectively discriminate between truly different flows.

Used in a different way, the T and modified χ^2 statistics can be used to characterize an even more unique feature of our bubble signals. Specifically, we can characterize the time irreversibility in a measured signal by observing difference in the symbol-sequence histograms for the forward-time and backward-time realizations. These differences can be quantified by

$$T_{\mathrm{irr}} = \sqrt{\sum_i (F_i - B_i)^2} \quad , \tag{4}$$

and

$$\chi^2_{\mathrm{irr}} = \sum_i \frac{(F_i - B_i)^2}{(F_i + B_i)} \quad , \tag{5}$$

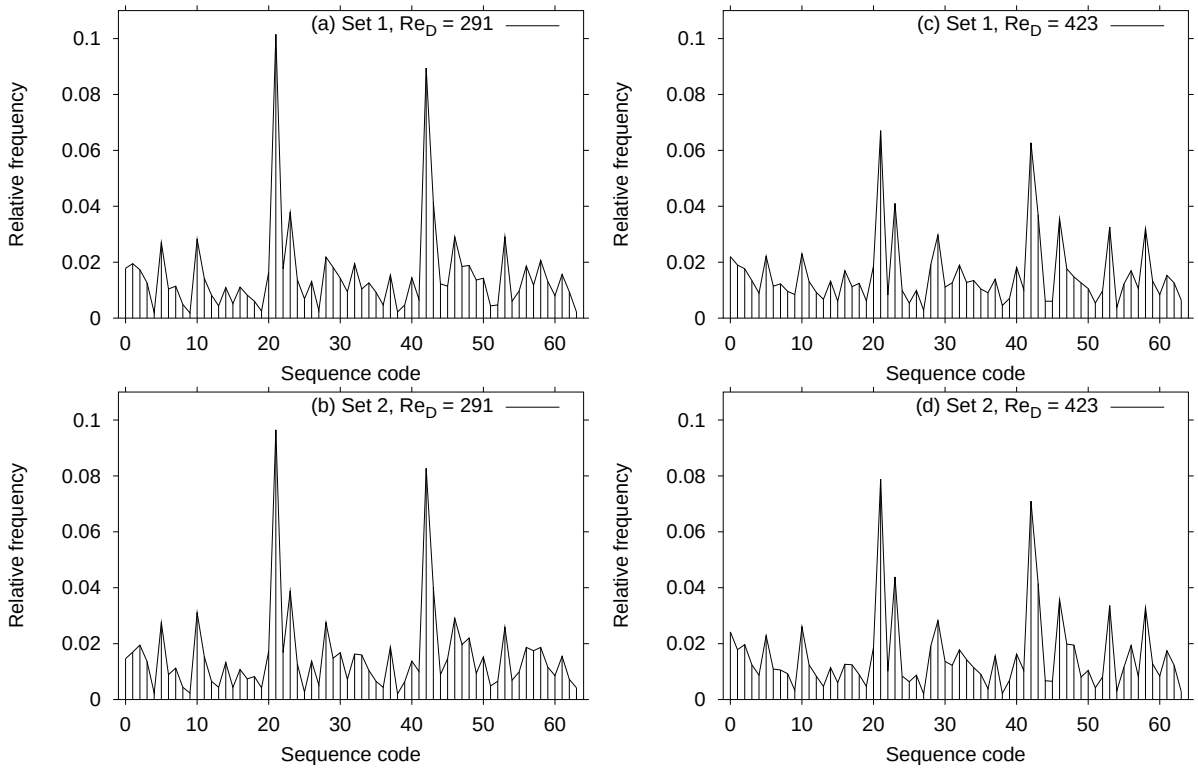

Figure 6: Symbol-sequence histograms for two replicate pairs (repeat experiments at the same flow conditions). Two symbols of sequence length 6 and inter-symbol interval of 0.02 sec were used.

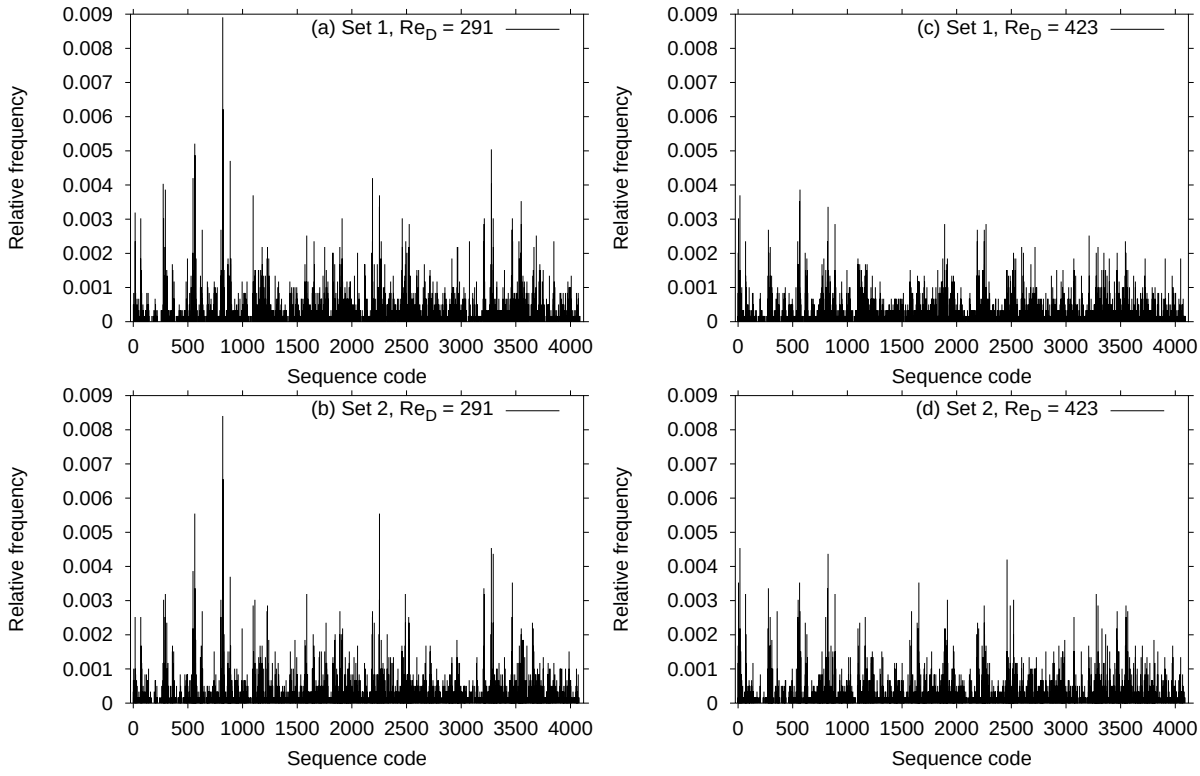

Figure 7: Symbol-sequence histograms for two replicate pairs (repeat experiments at the same flow conditions). Four symbols of sequence length 6 and inter-symbol interval of 0.02 sec were used.

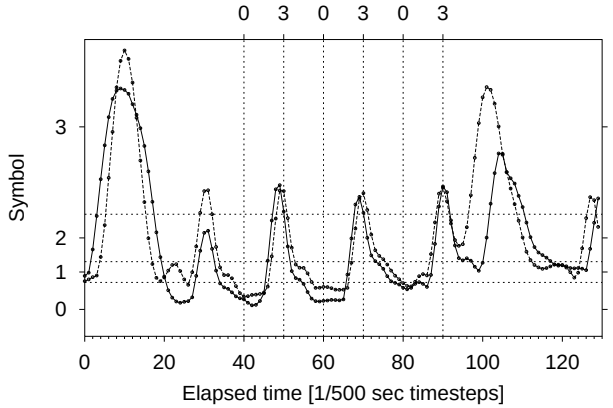

Figure 8: Repeating trajectory segments for set 1, $\mathrm{Re}_D = 291$. Four symbols of sequence length 6 and inter-symbol interval of 0.02 sec were used to find sequence 030303 in both segments. The second segment (dashed line) occurred 1.6 sec after the first segment.

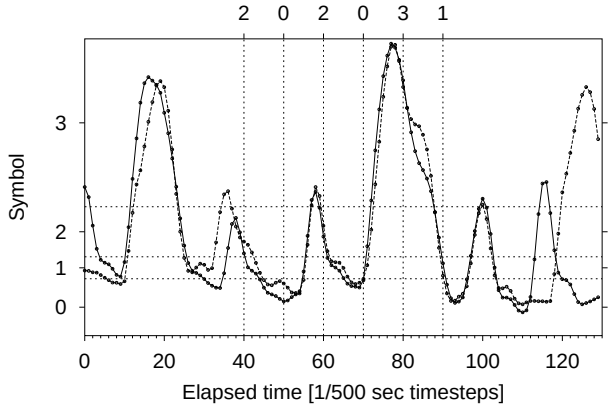

Figure 9: Repeating trajectory segments for set 1, $\mathrm{Re}_D = 291$. Four symbols of sequence length 6 and inter-symbol interval of 0.02 sec were used to find sequence 202031 in both segments. The second segment (dashed line) occurred 5.9 sec after the first segment.

where F is the forward-time and B the backward-time symbol-sequence histogram.

The modified entropy (Eq. 1) provides an indication of the relative complexity of the signals, and as we might expect, the higher flow replicates tend to have modified entropy values closer to 1. A detailed explanation of the importance of time irreversibility is beyond the scope of this paper, but we refer the reader to [15] for an in-depth discussion. It is sufficient to point out here that confirmation of time irreversibility implies that Gaussian linear models will not adequately characterize such measurements. Thus Fourier decomposition and autocorrelation analysis of such signals are likely to lose important information.

We can also find characteristic time scales by observing how the symbol statistics change with sequence length. For example, in Fig. 11 we depict T_{irr} for one of the above bubble flows

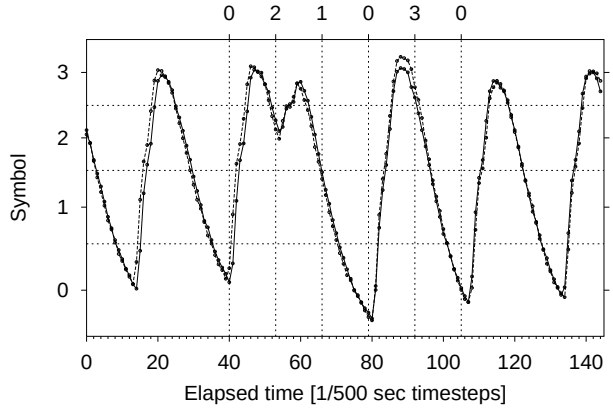

Figure 10: Repeating trajectory segments for $\mathrm{Re}_D = 63$. Four symbols of sequence length 6 and inter-symbol interval of 0.026 sec were used to find sequence 021030 in both segments. The second segment (dashed line) occurred 5.3 sec after the first segment.

Table 1: Symbol statistics for comparisons of two replicate pairs (repeat experiments). Two symbols of sequence length 6 and inter-symbol interval of 0.02 sec were used.

Data set A		Data set B		T_{AB}
Comparisons of similar flows				
Set 1	$\mathrm{Re}_D = 291$	Set 2	$\mathrm{Re}_D = 291$	0.0284
Set 1	$\mathrm{Re}_D = 423$	Set 2	$\mathrm{Re}_D = 423$	0.0392
Comparisons of different flows				
Set 1	$\mathrm{Re}_D = 291$	Set 1	$\mathrm{Re}_D = 423$	0.0606
Set 1	$\mathrm{Re}_D = 291$	Set 2	$\mathrm{Re}_D = 423$	0.0497
Set 2	$\mathrm{Re}_D = 291$	Set 1	$\mathrm{Re}_D = 423$	0.0583
Set 2	$\mathrm{Re}_D = 291$	Set 2	$\mathrm{Re}_D = 423$	0.0466

Table 2: Time-reversibility statistics for two replicate pairs (repeat experiments). Two symbols of sequence length 6 and inter-symbol interval of 0.02 sec were used.

Data set	H_S	T_{irr}	χ^2_{irr}
Set 1, $\mathrm{Re}_D = 291$	0.900	0.0604	703
Set 2, $\mathrm{Re}_D = 291$	0.908	0.0679	721
Set 1, $\mathrm{Re}_D = 423$	0.930	0.0258	110
Set 2, $\mathrm{Re}_D = 423$	0.910	0.0390	286

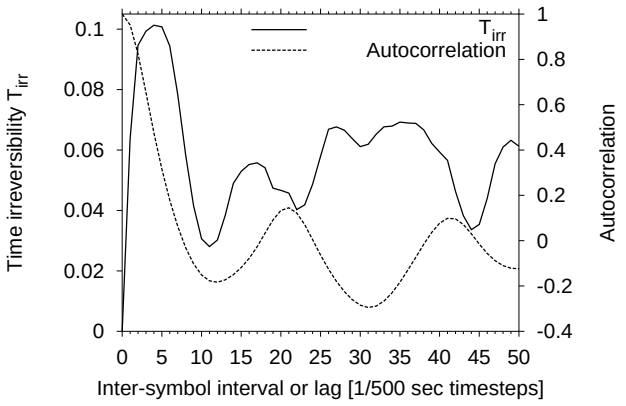

Figure 11: Time-irreversibility statistic as a function of inter-symbol interval and autocorrelation for $\mathrm{Re}_D = 291$. Eight symbols of sequence length 2 were used to calculate differences between forward-time and reverse-time symbol-sequence histograms.

as a function of sequence length for a two-level, eight-partition symbolization (thus defining sequence length based on the time interval between the two symbols). The autocorrelation function for the same data is also shown for comparison. It is apparent that the time irreversibility profile is richer in time-scale features than the autocorrelation. Among other things, we observe that the strongest time irreversibility occurs at time scales smaller than the average bubble period, reflecting the "warping" effect in the bubble profile described above. Also, other time-irreversibility peaks are evident at longer periods, indicating asymmetries in the formation of bubble sequences.

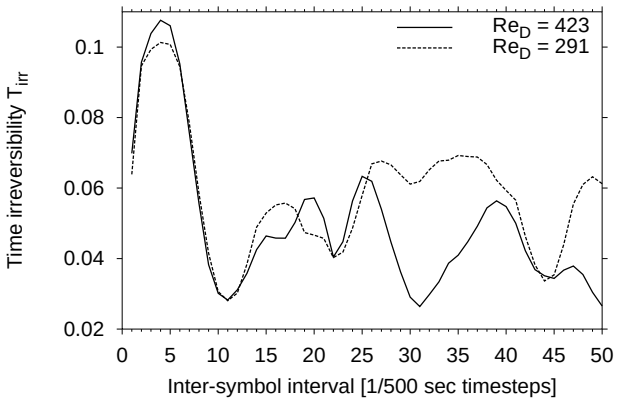

Figure 12: Time-irreversibility statistic as a function of inter-symbol interval for two of the data sets used in Fig. 1.

In Fig. 12 we see a comparison of the time-irreversibility profiles for the two different flow conditions discussed earlier. The differences suggest that it should be possible to distinguish the two conditions based on these profiles. It also appears that increasing the flow resulted in an increased warping of the bubbles (higher time irreversibility at short time scales) and a shift to more symmetrical bubble sequences (reduced time irreversibility at long time scales).

CONFIDENCE LIMITS

We recognize that our estimates of symbol-sequence histograms will be uncertain to some extent because we use finite data samples for their construction and because of experimental variability. Thus at some point it is important to establish confidence limits for the symbol-sequence statistics. That is, how do we assess when the differences between estimated symbol-sequence histograms are large enough to warrant saying that the data sets being compared are truly different within some specified level of confidence? This issue is complicated by the fact that symbol-sequence histograms are typically constructed from data which involve strong temporal correlations, which means that the assumption of independence used to establish confidence limits for standard statistical tests does not hold.

It is possible to estimate the sampling variability (that is, the normal variations in the symbol-sequence histogram produced by having finite-length data sets) by taking random samples from the sequences observed in a single data set and repeatedly constructing histograms from each sample. This is frequently referred to as "bootstrapping" confidence limits. We have empirically observed that the sample-to-sample histogram variability scales as $1/\sqrt{n}$, where n is the random-sample population size. Once the variability of samples from a single data set has been determined, it is then possible to apply this scaling rule to determine how big a data set must be to characterize the symbol-sequence histogram to a specific level of confidence. It is important to note, however, that the resulting confidence limits will only apply to data of similar type. An assessment of data from another process or very different operating conditions will require another evaluation of random samples to account for the different temporal correlations. Likewise, it is also appropriate to compare replicate experimental data sets to evaluate the variability of the experiment itself. The reader is referred to a companion paper [7] for other treatment of this topic.

CONCLUSIONS

Through the above examples we have demonstrated how symbolization can be used to discriminate complex dynamic patterns and provide unique information about the physics of bubble flows. We expect that this analysis method will find numerous other applications in the field of multiphase flows in the near future. For algorithm development, critical areas which still need attention include optimal selection of symbolization parameters,

a priori establishment of statistical confidence limits, and analysis of multiple simultaneous measurements.

REFERENCES

1. Tang X.Z., Tracy E.R., Boozer A.D , deBrauw A., Brown R. (1995). "Symbol sequence statistics in noisy chaotic signal reconstruction", *Physical Review E* **51**:5, 3871–3889.
2. Kurths J., Schwarz U., Witt A., Krampe R.Th., Abel M. (1996). "Measures of complexity in signal analysis", in *Chaotic, Fractal, and Nonlinear Signal Processing, AIP Conference Proceedings* **375** (3rd Technical Conference on Nonlinear Dynamics and Full Spectrum Processing (Mystic, Connecticut USA; 1995 July 10–14)) Woodbury, New York 1996, 33–54.
3. Lehrman M., Rechester A.B., White R.B. (1997). "Symbolic analysis of chaotic signals and turbulent fluctuations", *Physical Review Letters* **78**:1, 54–57.
4. Tang X.Z., Tracy E.R., Brown R. (1997). "Symbol statistics and spatio-temporal systems", *Physica D* **102**, 253–261.
5. Finney C.E.A., Green J.B. Jr., Daw C.S. (1998). "Symbolic time-series analysis of engine combustion measurements", SAE Paper No. 980624.
6. Tang X.Z., Tracy E.R. (1998). "Data compression and information retrieval via symbolization", accepted for publication in *Chaos*.
7. Finney C.E.A., Nguyen K., Daw C.S., Halow J.S. (1998). "Symbol-sequence statistics for monitoring fluidization", to be presented at the *1998 International Mechanical Engineering Congress & Exposition* (Anaheim, California USA; 1998 November 15–20).
8. Crutchfield J.P., Packard N.H. (1983). "Symbolic dynamics of noisy chaos", *Physica D* **7**, 201–223.
9. Abarbanel H.D.I. (1996). *Analysis of Observed Chaotic Data*, Springer-Verlag, ISBN 0-387-94523-7.
10. Kantz H., Schreiber T. (1997). *Nonlinear Time Series Analysis*, Cambridge University Press, ISBN 0-521-55144-7.
11. Tritton D.J., Egdell C. (1993). "Chaotic bubbling", *Physics of Fluids A* **5**, 503–505.
12. Chakka P. (1994). "An experimental study of chaotic behavior of bubbles in liquid columns", *M.S. Thesis*, University of Tennessee, Knoxville.
13. Mittoni L.J., Schwartz M.P., La Nauze R.J. (1995). "Deterministic chaos in the gas inlet pressure of gas-liquid bubbling systems", *Physics of Fluids* **7**:4, 891–893.
14. Nguyen K., Daw C.S., Chakka P., Cheng M., Bruns D.D., Finney C.E.A., Kennel M.B. (1996). "Spatio-temporal dynamics in a train of rising bubbles", *Chemical Engineering Journal*, **64**:1, 191–197.
15. Diks C., Houwelingen J.C. van, Takens F., DeGoede J. (1995). "Reversibility as a criterion for discriminating time series", *Physics Letters A* **201**, 221–228.