

SYMBOL-SEQUENCE STATISTICS FOR MONITORING FLUIDIZATION

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

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

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

ABSTRACT

We propose that data-symbolization methods derived from nonlinear dynamics and chaos theory can be useful for characterizing and monitoring patterns in fluidized-bed measurement signals. Data symbolization involves the discretization of a measurement signal into a limited set of values. In this discretized form, the measurements can be processed very efficiently to detect dynamic patterns that signify various types of physical phenomena, including bubbling, slugging, and transitions between fluidization states. Besides computational efficiency, symbolic methods are also robust when noise is present. Using various types of measurements from experimental beds, we illustrate specific examples of how symbolization can be applied to fluidization diagnostics. We also suggest directions for future research.

NOMENCLATURE

H_S	Modified Shannon entropy
N_{obs}	Number of non-zero-frequency sequences
p_i	Observed relative frequency (probability) of sequence i
T	Euclidean norm
T_{irr}	Time irreversibility index, based on Euclidean norm of forward- and backward-time symbol-sequence histograms

INTRODUCTION

The detection of specific types of dynamic phenomena can be very important to the successful operation of fluidized-bed processes. For example, short-time-scale pressure-drop variations can be used to detect the onset of solids agglomeration, defluidization, or changes in fluidization state [1,2,3,4,5]. It has also been recognized that dynamic measurements provide a means for establishing scale-up between laboratory and plant fluidized beds [6,7].

One limiting factor in fully exploiting dynamic fluidized-bed data, especially in plant environments, is the high level of measurement noise introduced by extraneous sources such as electrical equipment, mechanical vibration, and low sensor resolution. Another limiting factor is that fluidized-bed dynamics is both highly complex and nonlinear [8,9], making analysis difficult even when no measurement noise is present. We began investigating symbolization to address these problems after discovering that symbolization has already been successfully applied to a number of other noisy nonlinear processes [10,11,12,13]. In addition, we had independently conceived of a related approach in studying fluidized-bed patterns [14].

The concept of symbolization has its roots in dynamical-systems theory, particularly in the study of nonlinear systems which can exhibit bifurcations and chaos [15]. As explained more fully elsewhere [16], the fundamental theory must be adapted for practical situations because of the effects of noise and because the underlying dynamical equations are usually not available. In practice, the essential step involves discretizing time-series measurements into a limited set of values (*i.e.*, the data range is divided with partitions into symbolic regions such that a specific symbol is assigned to data valued within a given region). Each finite-length dynamic pattern can then be represented as a unique sequence of symbols. With all possible dynamic sequences thus encoded, recognition of certain events or classification of different patterns can be accomplished efficiently. Such computational efficiency is highly desirable in real-time monitoring and control situations [17]. We expect that this approach can also be extended to the validation of computational models [13,18].

Our principal objective in this paper is to illustrate specific approaches for implementing symbolization with fluidized-bed measurements and interpreting the results. We use pressure and acoustic measurements from laboratory fluidized beds of Group B and Group D particles. In addition to illustrating the observed dynamic patterns, we briefly discuss rules for choosing appropriate symbolization parameters and methods for calculating quan-

tative statistics from the symbol patterns that can be used collectively to characterize fluidization states.

EXPERIMENTAL APPARATUS AND METHODS

We report on application of symbolic analysis to data from two different fluidized beds and two different particulate solids.

The first fluidized bed consisted of a 10.2-cm-diameter, 260-cm-long plexiglass tube. Regulated, pressurized house air was introduced into the plenum, where the air passed through a seven-tuyere distributor. The bed material was uniform 0.45-cm-diameter steel spheres with a particle density of 7.5 gm/mL. The static bed height was 23.5 cm. A Baratron differential pressure transducer measured the pressure drop wall taps located 11 and 23 cm above the distributor. The transducer output signal was analog bandpass filtered between 0.1 and 20 Hz. The filtered signals were recorded with a 12-bit digitizing oscilloscope at 200 Hz.

The second fluidized bed consisted of a 7.6-cm-diameter, 68-cm-long acrylic tube. Air was drawn by an acoustically insulated blower through a perforated-plate distributor containing 185 0.15-cm-diameter holes. The bed material was millet seed with equivalent diameter of 0.35 cm and density of 0.97 gm/mL. The settled bed height was 35 cm. Sounds from the bed captured by a directional microphone were bandpass filtered (1600–6300 Hz) through a graphic equalizer and recorded with a DAT recorder at 44.1 kHz. The recorded sounds were later digitized by a computer sound card at 22.05 kHz. Additionally, a wire probe protruding through the bed wall measured particle impacts, much like a phonograph needle. The analog output signal from the wire probe was bandpass filtered and digitized and then numerically integrated using a moving-RMS filter to produce a time series of local particle-motion intensity. The microphone and wire probe were positioned 55 cm above the distributor on opposite sides of the bed.

The filter settings for the first bed were chosen to enhance signal features associated with bubble and slug events, which were expected to be primarily concentrated in frequencies less than 30 Hz. Such filtering also reduces 60-Hz contamination from nearby AC power systems and prevents signal aliasing. With the 7.6-cm bed system, there was significant acoustic background noise from the nearby blower and other sources, whereas the 10.2-cm bed system was relatively free of acoustic background noise. More details about the experimental apparatus, methods and comparison between acoustic and pressure measurements may be found in [19,20].

Figure 1 shows example time series for all sensor types at a range of fluidization conditions. In the plots, the scale of each time-series segment has been adjusted to have unit standard deviation. The steel-bed pressure-drop signals (Fig. 1(a)) show the fluidization transition from bubbling (series i) to very stable slugging (ii) to complex, interrupted slugging at high gas velocities

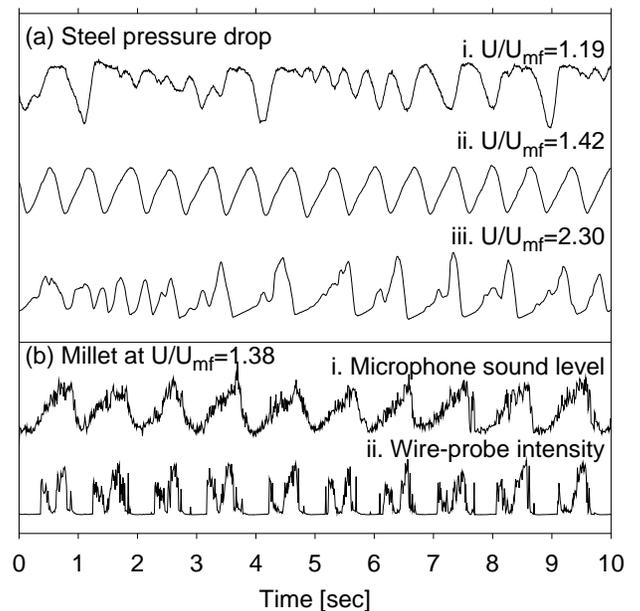


Figure 1: Example time series.

(iii). For the millet data (Fig. 1(b)), the time series for both sensors exhibit similar features. Because the wire probe was located above the bed static surface, when the bed was collapsed there were no appreciable oscillations in the probe, seen as the flat portions of series (ii). However, as the bed surface rose with a passing slug, evidenced by the rising intensity in the microphone signal (i), the wire probe measured particle passage.

A prominent feature in several of these signals is that the dominant rise and fall times are different. This difference reflects a characteristic time irreversibility; that is, we can distinguish whether we are observing this process in forward or reverse time. Physically, it reflects the fact that bed expansion and collapse occur by different mechanisms.

SYMBOLIZATION METHODS

We use an approach based on the work of Tang *et al.* [10]; an alternate but analogous approach may be found in [12]. Although this method is motivated by symbolic-dynamics theory, it is not completely equivalent, mainly because generating partitions are undefined in the presence of noise. The reader is referred to Crutchfield and Packard [15] for a detailed discussion of the latter issue.

The basic process of data symbolization is illustrated in Fig. 2. In the so-called *static transformation* [11], the value of a given measurement is assigned one of n symbolic values (*e.g.*, 0 or 1 for $n = 2$). Typically, we define discretization partitions such that the individual occurrence of each symbol is equiprobable

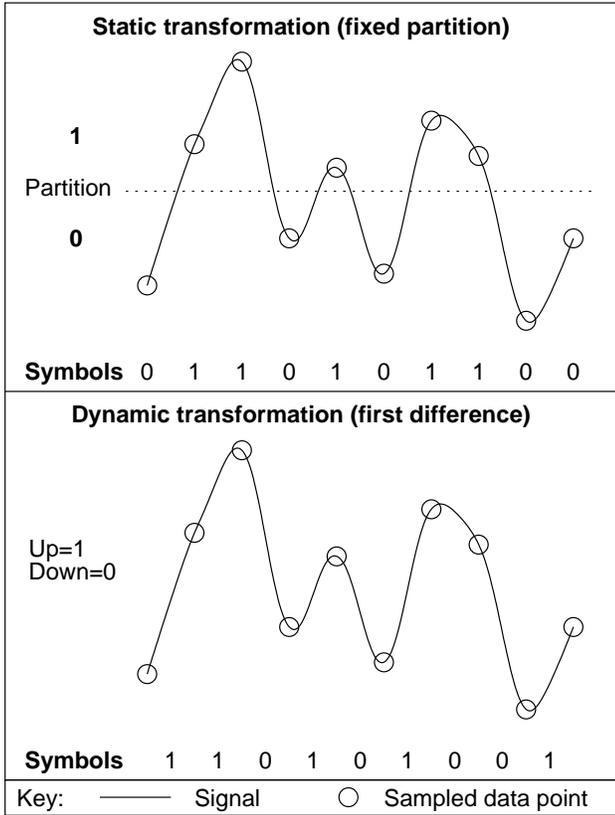


Figure 2: Static and dynamic transformations of a data series into a symbol series.

with all others. In the *dynamic transformation*, arithmetic differences of adjacent data points define the symbolic values. We symbolize a positive difference as 1 and a negative difference as 0. Such a differenced symbolization scheme is relatively insensitive to extreme noise spikes in the data.

Once a time series is symbolized, the relative frequencies of all possible symbol sequences in the data, defined by a symbol-sequence vector of length L , are evaluated. For example, if $L = 3$, the relative frequency of occurrence for each possible sequential combination of 3 symbols is computed. A simple way to index symbol-sequence frequencies is to assign a unique number to each possible sequence by evaluating the equivalent base-10 value of each base- n sequence; we call this number the *sequence code*. For example, a sequence of 101 occurring with $n = 2$ would have a sequence code of 5. Using this method of identifying each symbol sequence allows representation of the relative-frequency histograms, which we call *symbol-sequence histograms* (SSH), as two-dimensional plots. Because of our equiprobable partitioning rule, the relative frequency of each possible sequence for truly random data will be equal (subject to the availability of sufficient data). Thus any significant deviation from equiprobable sequences is indicative of time correlation and

deterministic structure.

From the above discussion, it can be seen that symbol sequences can be used to represent any possible variation over time, depending on the number of symbols used and the sequence lengths. This is a very powerful property because it does not make any assumption about the nature of the patterns (*e.g.*, it works equally well for linear and nonlinear phenomena). A detailed explanation of the “optimal” choice of number of symbols and sequence length is beyond the scope of this paper but is addressed in another paper presented in this session [16]. For practical purposes, we have empirically found that partitions of 2–8 symbols and sequence lengths of 2 to 6 are most useful for depicting the observed patterns in fluidized-bed data. We explicitly report the assumed number of symbols n and sequence lengths L in all of the examples shown below.

For continuous data, such as discussed here, there is also a sampling-rate issue. Specifically, our fluidized-bed signals could be digitized at any selected rate up to several thousand samples per second. At excessive sampling rates, a problem arises because we simply produce long sequences of the same repeated symbol. Thus it is important not to oversample the data. In the examples given below, we present a specific approach for selecting a meaningful sample interval. In fact, we show that this approach can be used to generate diagnostics about relevant timescales in the data.

EXAMPLE RESULTS

Symbol-sequence histograms can be used to detect transitions in fluidization behavior and to indicate which features are predominant. In Fig. 3, we illustrate symbol-sequence histograms to compare bubbling, regular slugging, and higher-flow slugging conditions for the pressure-drop signal for steel particles. The original data were recorded at 200 samples per second, but the results shown in Fig. 3 were for data series symbolized every 20 samples (represented as the points in Fig. 4). The SSHs depicted in Fig. 3 compare bubbling with regular slugging (a) and regular slugging with higher-flow slugging (b). For these SSHs, four symbols were used with a sequence length of 3. Three characteristic sequences are highlighted on the SSHs: sequence 100 (sequence code 16), sequence 123 (sequence code 27), and sequence 112 (sequence code 22). Sequence 100 is more prominent in the bubbling time series (the filled-circle triplets in Fig. 4(a)) than in the regular slugging condition; this transition is visible in the SSH (Fig. 3(a)), as the peak at sequence code 16 diminishes as flow increases. Sequence 123 becomes the predominant feature of regular slugging (the filled-circle triplets in Fig. 4(b)), and this pattern shifts from very infrequent to dominant in the SSH (Fig. 3(a)). As gas flow increases beyond the regular-slugging state, interruptions or *stutters* [8] in the slugging become more frequent. These stutters are manifest in a collapse of the bed followed by successively increasing oscillations until

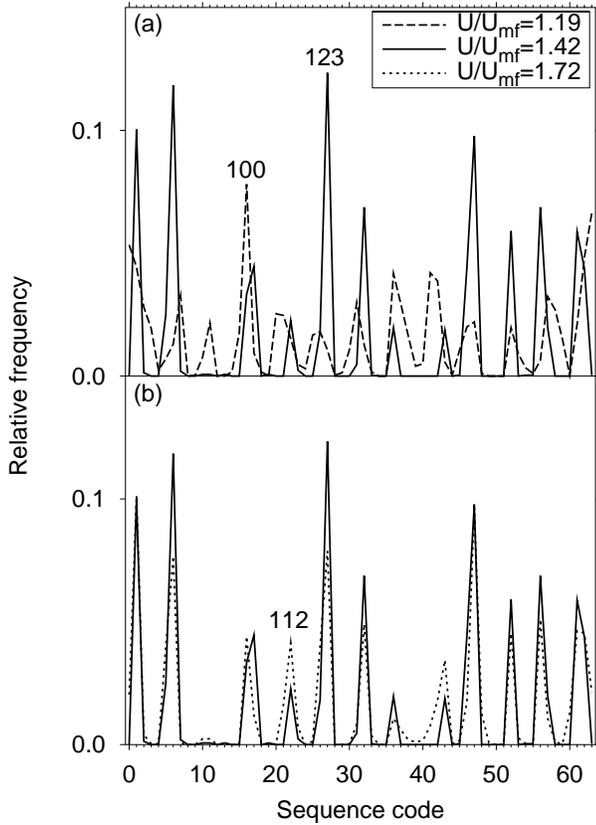


Figure 3: Example symbol-sequence histograms for fluidized steel spheres. Four symbols of sequence length 3 with symbolization interval 20 were used to construct the SSHs.

regular slugging is resumed; an example is seen in Fig. 4(c) between 1.5 and 3 sec elapsed time. Sequence 112 (the filled-circle triplets in Fig. 4(c)) reflects the increased occurrences of stutters as flow increases, which may be seen in the SSH (Fig. 3(b)). An even more telling feature is sequence 111 (sequence code 21), which goes from zero to significant relative frequency in the SSH (Fig. 3(b)).

The time-irreversible nature of the slugging process is seen in Fig. 4(b). Typically, between four and five of the resampled points are visible in the rising portion of each “cycle”, whereas two to three points are visible in the falling portion. In the SSH for $U/U_{mf} = 1.42$ (the solid lines in Fig. 3), the relative frequencies of sequences 331 (0.059 for sequence code 61) and 320 (0.069 for sequence code 56), representing a sharp fall in the pressure signal, are much greater than those for sequences 133 (0.005 for sequence code 31) and 023 (0.0007 for sequence code 11), representing a sharp rise in the pressure signal. Time irreversibility implies that the process described by the data may not be explained as a Gaussian linear random process [21]; al-

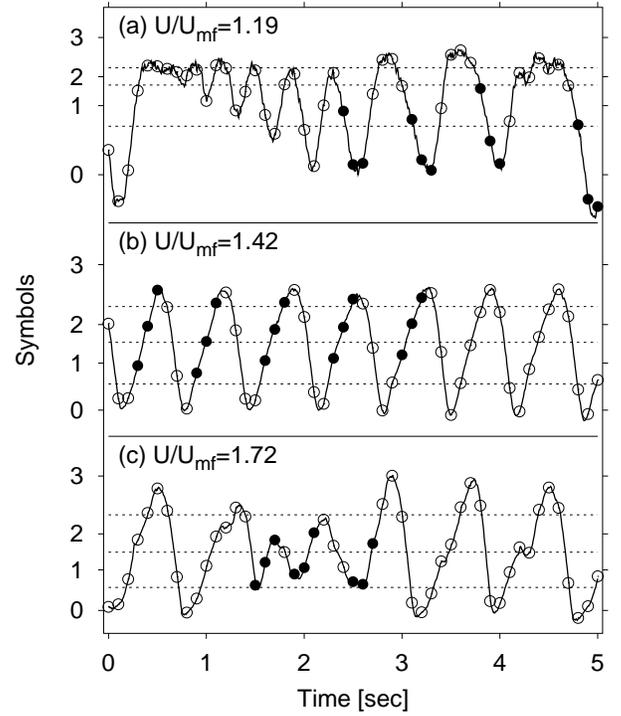


Figure 4: Illustration of relationship between specific physical events and symbol sequences. The data points (light and dark circles) are data resampled from the original digitized signal for symbolization, and the darkened point triplets correspond to specific sequences highlighted in Fig. 3, as described in the text.

though not an absolute test of nonlinearity, we use the degree of time reversibility as defined below as a good practical indicator of nonlinear structure.

Various statistics can be used to characterize symbol-sequence histograms. For example, we define a modified Shannon entropy as:

$$H_S(L) = -\frac{1}{\log N_{\text{obs}}} \sum_i p_{i,L} \log p_{i,L} , \quad (1)$$

where N_{obs} is the total number of sequences with non-zero frequency, i is a symbol-sequence index of sequence vector length L , and $p_{i,L}$ is the relative frequency of symbol sequence i . The only difference between Eq. 1 and the definition used by Tang *et al.* is that we use the number of non-zero-frequency sequences instead of the total number of possible sequences. This choice of N_{obs} reflects the fact that many possible sequences may not be realized because of finite data-set length. The result is to bias H_S upward when the number of possible sequences becomes large relative to the available data. For random data H_S should equal 1, whereas for nonrandom data it should be between 0 and 1.

To compare two symbol-sequence histograms A and B , we

use the Euclidean norm as a difference statistic:

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

where i is indexed over all possible sequence codes. Computing T is useful for comparing SSHs of different fluidization conditions or for testing for time reversibility by comparing the SSH for a time series with that of its time reverse. For more details about data symbolization for measurements of engineering systems, see [16,18,22].

As noted above, when dealing with time series which vary smoothly with time, such as the continuous, fast-time-sampled analog signals used in this study, some data treatment is necessary in order to obtain a good symbolic transformation of the data. In the coarse-grained sense, the measurement signal is oversampled, meaning that many consecutive data points would be symbolized in a long string of the same symbols. To obtain information about how the data change meaningfully in time, it is necessary to choose a *symbolization interval* (also termed *inter-symbol interval*) which defines the number of actual data points between successive symbols (the other data points within that interval are ignored).

The appropriate symbolization interval may be chosen on the basis of how symbol statistics vary as a function of that interval. Figure 5 shows how H_S and T_{irr} vary with symbolization interval for 3 symbols and sequence length 6 for the millet data. Here, T_{irr} refers to the difference statistic of SSHs computed from the (forward-time) time series and with the reverse-ordered (backward-time) time series. As seen in Fig. 5(a), at short symbolization intervals, H_S is low because the data are effectively oversampled, and the predominant symbol patterns are long sequences of the same repeated symbols and the transitions between symbols. As the symbolization interval increases, the data irregularity increases H_S . At an interval near the mean slugging period of 0.95 sec, H_S decreases; the symbol statistics can identify the predominant timescales in both the wire probe and microphone measurement data. Overall, H_S for the microphone sound level is lower than that for the wire-probe intensity. At very long symbolization intervals, the effects of aperiodicity and noise in the data reduce apparent deterministic structure, and H_S approaches unity.

The characteristics of T_{irr} at different symbolization intervals (Fig. 5(b)) are similar to those for H_S but with new information. The mean slugging periods are identified, seen strikingly in the microphone data as sharp troughs at multiples of one-half the slugging period. However, the maximal degree of time irreversibility is seen in the microphone data at a symbolization interval of 0.3 sec, which corresponds to the approximate sound-level collapse timescale seen in Fig. 1(b)(i). Whereas H_S highlighted the dominant slugging period, T_{irr} highlights the minimal time interval necessary to distinguish the time series and its time inverse. Physically, such a measure identifies the distinctive

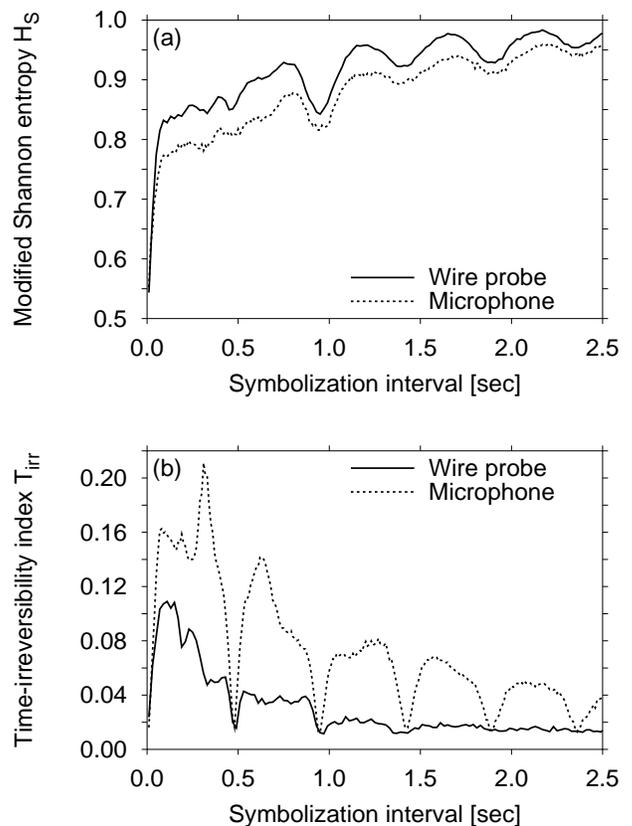


Figure 5: Variation of the symbolization time interval reveals deterministic and time-irreversible bubble features in the wire probe and microphone measurements. The symbolization parameters were 3 symbols of sequence length 6.

feature of these data, that slug growth and bed expansion occur uniquely compared with the bed collapse.

It is important to note that data nonstationarity results in time irreversibility. The data sets reported in this paper have been tested for stationarity to verify that the macroscopic statistics (measured over 10s to 100s of slugging cycles) do not change significantly throughout the data set (*e.g.*, the running mean and variance are fairly constant). The degrees of and timescales of time irreversibility reported in Fig. 5 result from the asymmetry of the signal during the slug cycle; any statistical nonstationarity in the data occurs on very long timescales.

Symbol-sequence histograms are compact representations of the overall dynamics of time series and may be used to determine the approximate fluidization condition of a test signal when compared with a library of reference SSHs. To illustrate this point, we compare SSHs from two very similar fluidization conditions against a library of 12 fluidization conditions for the steel pressure-drop signal. We expect that the difference statistic between SSHs should be minimal at approximately similar fluidiza-

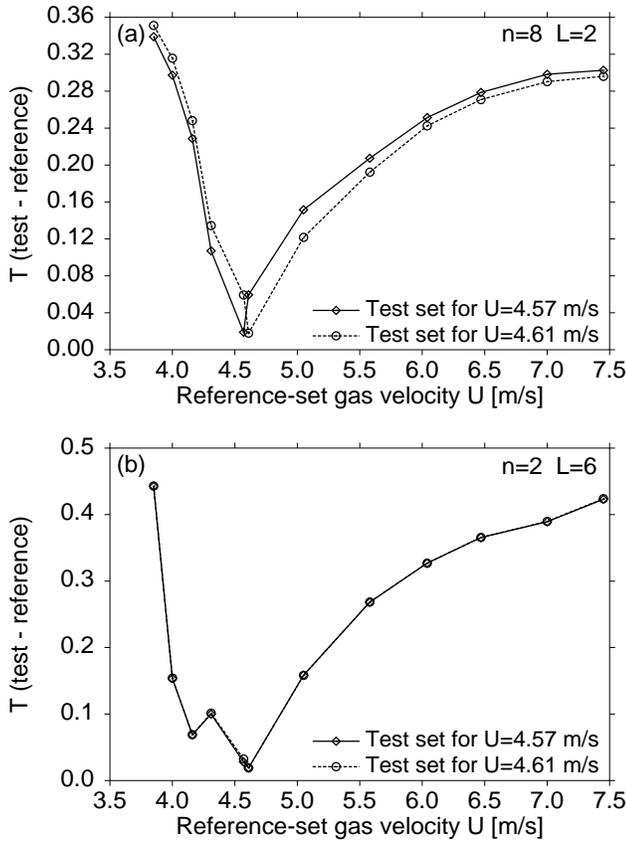


Figure 6: Comparison of symbol-sequence histograms of two test fluidization conditions with a reference library of known fluidization conditions. The error function T reaches a minimum for the reference set corresponding to each test set. The finer symbolization (a) detects subtler differences between the two very similar test conditions than does the coarser symbolization (b).

tion conditions. As seen in Fig. 6, the symbol statistics correctly identify the approximate fluidization of the two test sets with respect to the reference library sets. The eight-symbol sequence of length 2 (Fig. 6(a)) provides greater discrimination than the two-symbol sequence of length 6 (Fig. 6(b)) – even though the fluidization conditions are very similar, subtle differences are measured by a coarse (8-symbol) data description. In actual applications to monitoring fluidization, the degree of symbolization would depend on the desired degree of precision with respect to the reference library sets.

STATISTICAL CONFIDENCE

Our estimates of symbol-sequence statistics will be uncertain to some extent because we use finite data samples and because of experimental variability. For comparative purposes, it is im-

portant to establish confidence limits for the symbol-sequence statistics. This issue is complicated by the fact that symbol-sequence statistics 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.

The sampling variability (that is, the normal variations in the symbol-sequence histogram produced by having finite-length data sets) may be estimated empirically by taking a subset of randomly selected samples from the symbol sequences observed in a single data set and repeatedly constructing histograms from each sample set. 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 have been determined, it is then possible to apply this scaling rule to determine how large a data set must be to characterize the SSH 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.

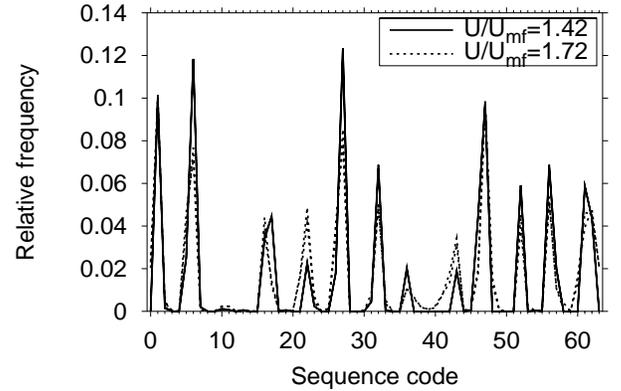


Figure 7: Experimental repeatability and symbol statistics for two flow conditions depicted in Fig. 3.

It is also appropriate to compare replicate experimental data sets to evaluate the variability of the experiment itself. Fig. 7 shows SSHs from replicate sets depicted in Fig. 3(b). The replicate sets were measured at the same flow rates within a few minutes of each other. The two SSHs for $U/U_{mf} = 1.42$ are nearly indistinguishable, and those for $U/U_{mf} = 1.72$ have only minor differences. The differences between the SSHs from the two flow rates are because of real, dynamical differences in fluidization and not because of experimental variability. Furthermore, the increased frequency of sequence 112 for the higher-flow data was shown (Fig. 4) to arise from the stutters resulting from in-

creased slugging instability at higher gas flows. The reader is referred to [22] for treatment of bootstrapped confidence intervals, particularly in distinguishing data patterns from symbol statistics arising from random processes.

SUMMARY

Data symbolization shows promise as a new approach to analyzing dynamic fluidized-bed data from a variety of sensors. Symbol statistics can provide new information about relevant timescales or data patterns using a low-precision approach which is less adversely affected by the presence of noise and which does not depend significantly on the nature of the data (linear or nonlinear). The computational simplicity of symbolic data analysis suggests that it could be useful in real-time monitoring of fluidized-bed measurement signals.

In the future, we wish to continue research on finding optimal symbolic representations of data from a wide variety of measurement types and fluidization conditions. Of particular practical interest is the development of reference libraries based on the SSHs to compare the dynamical similarity of unknown fluidization conditions with the known conditions. Implementation of the library system could include more sophisticated comparisons (e.g., using neural networks) which could lead to the avoidance of previously observed adverse conditions in process operations. Additionally, we foresee the utility of data symbolization in model validation and parameter selection.

REFERENCES

1. Shuster W., Kisliak P. (1952). "The measurement of fluidization quality", *Chemical Engineering Progress* **48**:9, 455–458.
2. Sadasivan N., Barreteau D., Laguerie C. (1980). "Studies on frequency and magnitude of fluctuations of pressure drop in gas-solid fluidised beds", *Powder Technology* **26**:1, 67–74.
3. Bai D., Bi H.T., Grace J.R. (1997). "Chaotic behavior of fluidized beds based on pressure and voidage fluctuations", *AIChE Journal* **43**:5, 1357–1361.
4. Marsaly A., Martens A., Morterol F.R.M.M., Raufast C. (1989 August 15). "Detection of anomalies in gas fluidized bed polymerization", U.S. Patent No. 4,858,144.
5. Daw C.S., Hawk J.A. (1995 July 25). "Fluidization quality analyzer for fluidized beds", U.S. Patent No. 5,435,972.
6. Glicksman L.R., Hyre M.R., Farrell P.A. (1994). "Dynamic similarity in fluidization", *International Journal of Multiphase Flow* **20**, Supplement, 331–386.
7. Schouten J.C., Stappen M.L.M. vander, Bleek C.M. van den (1996). "Scale-up of chaotic fluidized bed hydrodynamics", *Chemical Engineering Science* **51**, 1991–2000.
8. Daw C.S., Finney C.E.A., Vasudevan M., Goor N.A. van, Nguyen K., Bruns D.D., Kostelich E.J., Grebogi C., Ott E., Yorke J.A. (1995). "Self-organization and chaos in a fluidized bed", *Physical Review Letters* **75**:12, 2308–2311.
9. Stappen M.L.M. vander (1996). **Chaotic Hydrodynamics of Fluidized Beds** (Delft University Press, ISBN 90-407-1375-8) (Ph.D. Thesis, Delft University of Technology (Netherlands)).
10. 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.
11. 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.
12. Lehrman M., Rechester A.B., White R.B. (1997). "Symbolic analysis of chaotic signals and turbulent fluctuations", *Physical Review Letters* **78**:1, 54–57.
13. Tang X.Z., Tracy E.R., Brown R. (1997). "Symbol statistics and spatio-temporal systems", *Physica D* **102**, 253–261.
14. Halow J.S., Daw C.S. (1994). "Characterizing fluidized-bed behavior by decomposition of chaotic phase-space trajectories", *AIChE Symposium Series* **90**:301, 69–91.
15. Crutchfield J.P., Packard N.H. (1983). "Symbolic dynamics of noisy chaos", *Physica D* **7**, 201–223.
16. Daw C.S. (1998). "Symbol statistics: A new tool for understanding multiphase flow phenomena", to be presented at the *1998 ASME International Congress & Exposition* (Anaheim, California USA; 1998 November 15–20).
17. Tang X.Z., Tracy E.R. (1998). "Data compression and information retrieval via symbolization", accepted for publication in *Chaos*.
18. Daw C.S., Kennel M.B., Finney C.E.A., Connolly F.T. (1998). "Observing and modeling nonlinear dynamics in an internal combustion engine", *Physical Review E* **57**:3, 2811–2819.
19. Halow J.S., Daw C.S., Finney C.E.A., Nguyen K. (1996). "Interpretation of acoustic signals from fluidized beds", *Proceedings of the 5th World Congress of Chemical Engineering* (San Diego, California USA; 1996 July 14–18), VI-291–VI-296.
20. Finney C.E.A., Daw C.S., Halow J.S. (1998). "Measuring slugging bed dynamics with acoustic sensors", to appear in *KONA: Powder and Particle*.
21. 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.
22. Finney C.E.A., Green J.B. Jr., Daw C.S. (1998). "Symbolic time-series analysis of engine combustion measurements", SAE Paper No. 980624.