

# Engineering Notes

## Ice Shape Characterization Using Self-Organizing Maps

Stephen T. McClain\*

Baylor University, Waco, Texas 76798-7356

Peter Tiño†

University of Birmingham,

Birmingham, England B15 2TT, United Kingdom

and

Richard E. Kreeger‡

NASA John H. Glenn Research Center at Lewis Field,  
Cleveland, Ohio 44524

DOI: 10.2514/1.C031209

### Nomenclature

$b$	=	codebook vector element
$h(i, j)$	=	neighborhood function of $i$ to $j$ codebook vectors
$j$	=	codebook vector index
$M$	=	number of codebook vectors
$R^d$	=	high-dimensional data space
$x$	=	element of data set
$\delta$	=	scaling parameter governing neighborhood size
$\eta$	=	learning rate

### I. Introduction

**D**URING the validation and verification of ice accretion codes, predicted ice shapes must be compared with experimental measurements of wind-tunnel or atmospheric ice shapes. Current methods for ice accretion code validation are based on ice shape features such as horn angle, horn thickness, stagnation thickness, and wrap limits, depicted in Fig. 1.

While the shape features noted by Wright and Chung [1] are important and can be related to airfoil performance, in many instances, distinct horns and other ice features are not apparent. Furthermore, horn and ice features can change considerably along the span of a two-dimensional airfoil. Consequently, most comparisons of computational predictions to ice shape measurements provide an ad hoc comparison of the predicted ice shapes to a typical two-dimensional ice shape trace. Most important, no definition of what makes the measured ice shape typical is provided, and no statistical data are provided to determine how well the predictions compare to the typical ice shape.

Most two-dimensional ice shapes on straight wings have considerable three-dimensional variations along the span of the

wing. Figure 2 shows a three-dimensional point cloud from an ice shape measured after a test in the NASA Icing Research Tunnel at the John H. Glenn Research Center at Lewis Field. The ice shape point cloud shows significant variation along the spanwise direction. This variation along the spanwise direction is evident when the point cloud is projected onto the chord-thickness plane, shown in Fig. 3. Figure 3 shows that the typical ice shape may have a deviation of as much as 0.1 in. in any direction on the horn. However, current ice shape comparison techniques do not consider this experimental uncertainty in the comparisons. More rigorous typical ice shapes and comparison methods are required to advance ice accretion codes.

Because of the spanwise variation of ice shape measurements, a new method is needed to determine what constitutes a typical ice shape. The following sections present the basics of a neural network concept called a self-organizing map (SOM) [2]. Combining SOMs or other nonlinear manifold modeling schemes with multidimensional statistical descriptions has the potential for enabling this more rigorous comparison of ice shape predictions to three-dimensional ice shape measurements.

### II. Self-Organizing Maps

Figure 4 shows three-dimensional data scattered about  $\beta$ , which is inherently one-dimensional (a manifold). In a signal-processing application, the data in Fig. 4 are noisy and require significant memory. If information about  $\beta$  could be obtained, the trend of the noisy data could be transmitted much less expensively than transmitting the complete noisy data set. The important question is: how can the nature of  $\beta$  be determined using the experimental data?

SOMs depend on the use of codebook vectors,  $\mathbf{b}$ , to represent clumps of data, depicted in Fig. 5. When a series of codebook vectors are connected, the one-dimensional manifold guiding the data in Fig. 5 is represented by the spline through the codebook vectors. To develop the codebook vectors, an initial shape or random distribution of the vectors is determined. The codebook vectors are then moved in the direction of the clump of points to which the codebook vector is closest. Like most neural network approaches, the SOM requires a learning or training process. Over iterative moves, the codebook vectors spread out and settle into their local clumps.

The SOM can be best described as a constrained clustering method [2]. Consider a data set of high-dimensional points aligned (up to some noise) along a lower dimensional manifold embedded in the high-dimensional data space  $R^d$ , as depicted in Fig. 4. In a SOM, such a data set is described through a collection of  $M$  codebook vectors  $b^j \in R^d$ ,  $j = 1, 2, \dots, M$ , living in the data space. Each codebook vector,  $b^j$ , represents the region of the data space around it (Voronoi compartment of  $b^j$ ), such that all data points in that region are closer to  $b^j$  than to any other codebook vector. Crucially, a topological neighborhood structure is imposed on the codebook using a neighborhood function  $h(i, j)$ ,  $i, j = 1, 2, \dots, M$ . Higher values of  $h(i, j)$  signify that codebook vectors  $b^i$  and  $b^j$  should be neighbors (e.g., lie close to each other in the data space). Smaller values of  $h(i, j)$  mean that no such requirement applies.

During the training process, the codebook vectors get adapted to the data set so that the quantization error (resulting from representing each original data point  $x$  by the codebook vector  $b^{\text{win}(x)}$  closest to it) is minimized and, at the same time, the layout of codebook vectors  $b^j$  in the data space respects the neighborhood properties dictated by the neighborhood function  $h(i, j)$ . For a one-dimensional data structure (e.g., Fig. 4), one simply prescribes that the representative codebook vectors  $b^j$  must lie on a bicycle chain embedded in the data space. This corresponds to imposing a linear order on the codebook vector indexes  $1 < 2 < 3 < \dots < M$  and defining

Presented as Paper 2009-3865 at the 1st AIAA Atmospheric and Space Environments Conference, San Antonio, TX, 22–25 June 2009; received 9 August 2010; revision received 15 November 2010; accepted for publication 16 November 2010. Copyright © 2010 by the American Institute of Aeronautics and Astronautics, Inc. All rights reserved. Copies of this paper may be made for personal or internal use, on condition that the copier pay the \$10.00 per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923; include the code 0021-8669/11 and \$10.00 in correspondence with the CCC.

\*Assistant Professor, Mechanical Engineering Department, One Bear Place #97356. Member AIAA.

†Senior Lecturer, School of Computer Science.

‡Aerospace Engineer, Icing Research Branch, 21000 Brookpark Road. Senior Member AIAA.

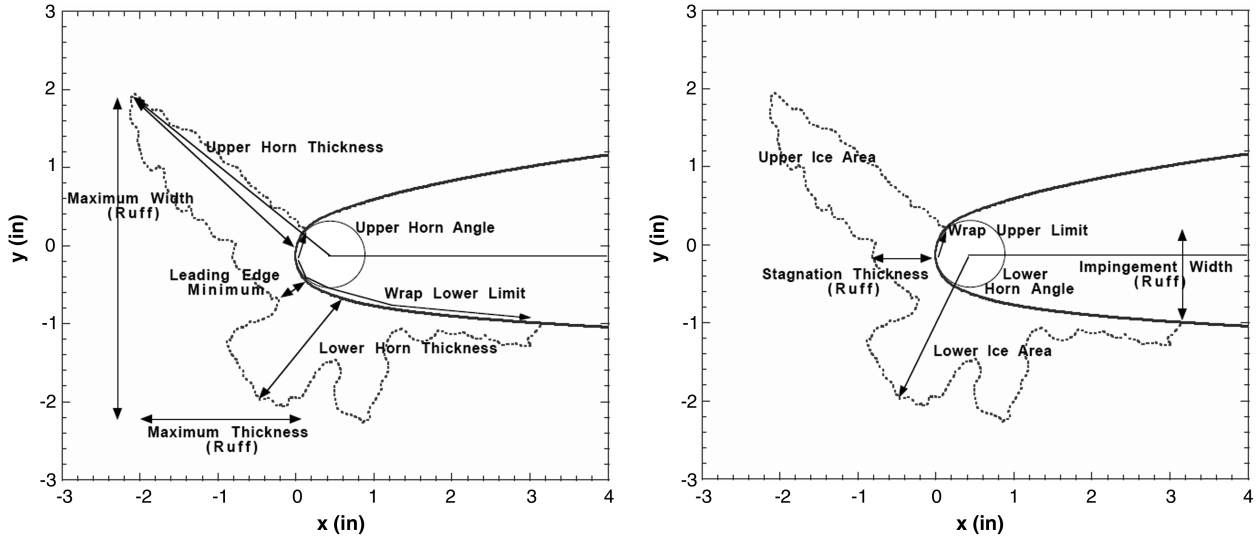


Fig. 1 Ice shape characterization parameters (reprinted from [1]).

$$h(i, j) = \exp\left(-\frac{(i - j)^2}{\delta^2}\right), \quad i, j = 1, 2, \dots, M$$

where  $\delta$  is the scale parameter governing the neighborhood size. During the training phase, data points are iteratively selected from the data set, and for each data point  $x$ , all codebook vectors are moved from their current positions closer to  $x$ . How much each codebook vector  $b^j$  gets moved depends on how close we want  $b^j$  to be to the principal representative  $b^{\text{win}(x)}$  of  $x$ . The closer  $b^j$  should be to  $b^{\text{win}(x)}$ , as measured by the value  $h(\text{win}(x), j)$  of the neighborhood function, the more it gets moved toward  $x$ . The update equation can be summarized as

$$b^j \leftarrow b^j + h[\text{win}(x), j] \cdot \eta \cdot (x - b^j), \quad j = 1, 2, \dots, M$$

where  $\eta$  is a positive real number, called the learning rate, modulating the proportions of codebook vector updates.

To ensure convergence of the algorithm, the learning rate,  $\eta$ , is made to decrease over time (e.g., exponentially) from some initial value to zero. It can be shown that, in order to preserve the neighborhood relations among the codebook vectors, it is recommended that the neighborhood scale parameter,  $\delta$ , decreases over time as well. Starting with a broader neighborhood (higher value of  $\delta$ ) enabling rough ordering of codebook vectors in the data space, the neighborhood size is gradually decreased, leading to a more selective codebook ordering.

When the SOM method is applied to the two-dimensional projection of the ice shape data, the result is Fig. 6. Figure 6 presents a 30-point SOM representation of the ice shape. The SOM results demonstrate that the codebook vectors spread out and, since many

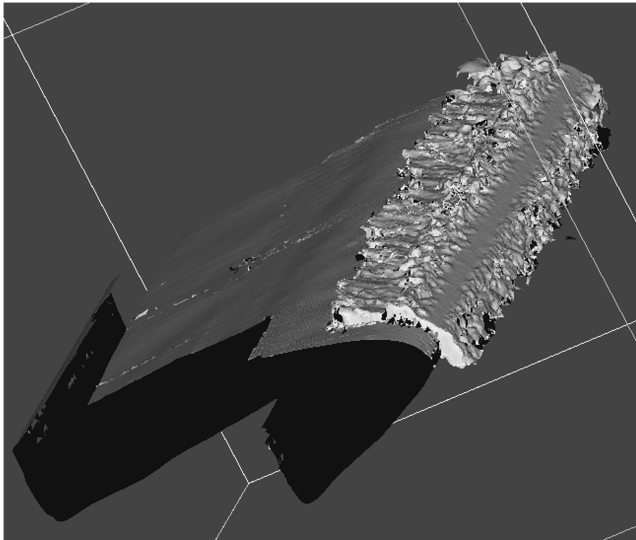


Fig. 2 Three-dimensional ice shape.

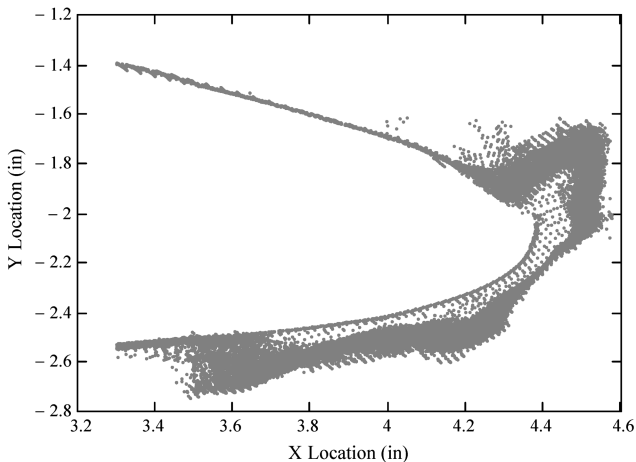


Fig. 3 Two-dimensional presentation of ice shape point cloud on straight airfoil.

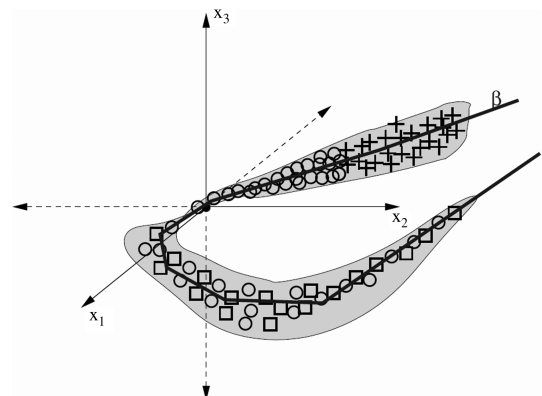


Fig. 4 Three-dimensional data scattered about a one-dimensional manifold [6].

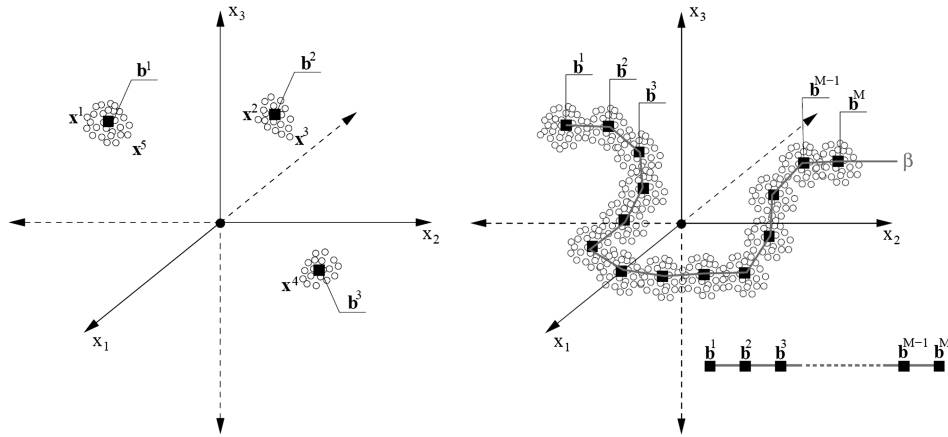


Fig. 5 Creation of codebook vectors about data clumps [6].

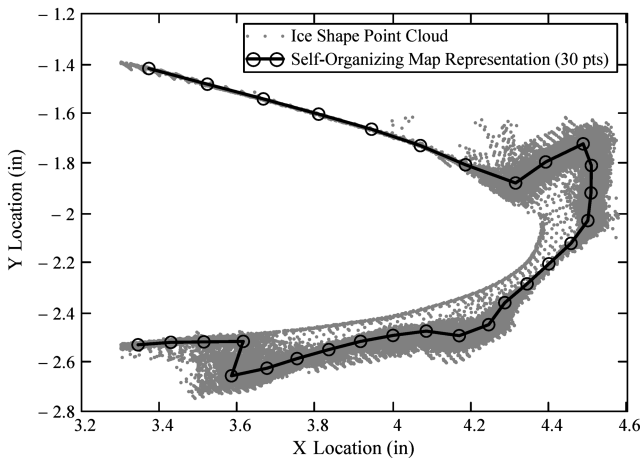


Fig. 6 Application of SOM to ice shape point cloud.

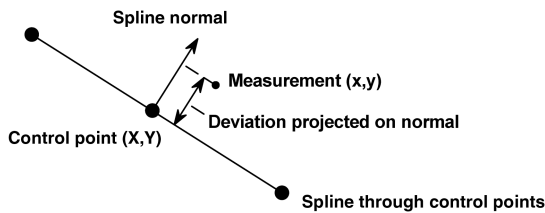


Fig. 7 Statistics of local point distribution (clump) about control point or codebook vector.

points are usually packed in feature changes (like horn tips), that most of the important shape features are captured well by the SOM representation.

### III. Application to Ice Shape Characterization

The nature of the SOM method and the positioning of the codebook vectors enable a more rigorous validation method for ice accretion codes when combined with multidimensional statistical descriptors. Since the clumps of points are distributed about the codebook vectors, the deviations of the point measurements in the clumps can be used to evaluate the coverage statistics and uncertainty of the codebook vector representation. Figure 7 demonstrates how each surface measurement is used to determine a deviation from the spline surface through the control points or codebook vectors.

When the deviation of all of the points in a clump are used to calculate the spatial standard deviation about the codebook vectors, coverage limits that contain a set percentage of the points may be determined. An example of the generation of 95% coverage limits is presented in Fig. 8. Figure 9 shows the resulting 95% coverage limits when the method is applied to the three-dimensional ice shape using the central difference estimate for local derivatives at a codebook vector, as opposed to the true spline normals.

Figure 9 demonstrates how the coverage limits provide a mechanism for comparison of the measured ice shapes to predicted ice shapes. If the predicted ice shape is within the coverage limits or uncertainty bands, then the predicted ice shape agrees with the experimental measurements.

Recognizing that the ultimate goal of ice shape characterization is aerodynamic performance prediction, the parameters presented by Wright and Chung [1] may still be used to describe mean ice shapes characterized by the proposed methods. The benefit of the proposed

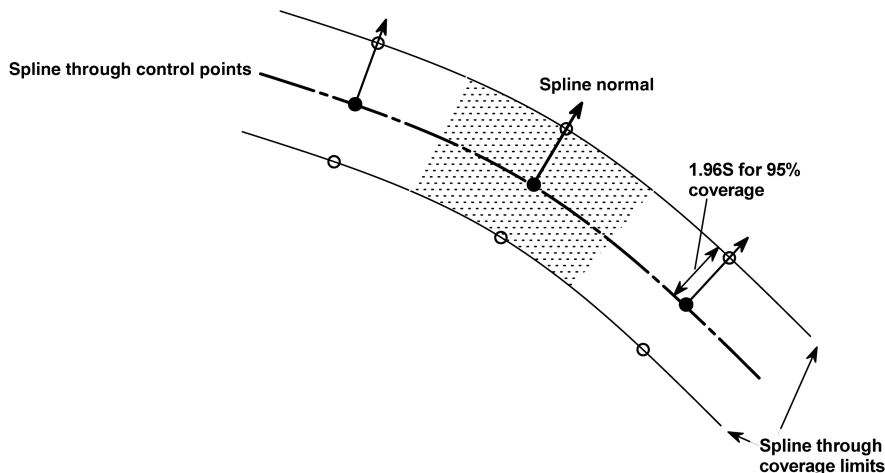
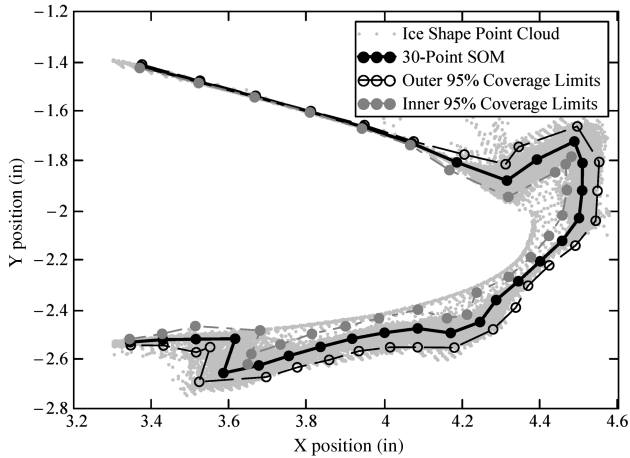


Fig. 8 Generation of coverage limits using local point statistics.



**Fig. 9 Application to ice shape point cloud using control point central difference derivatives.**

approach is that each of the aerodynamic performance criteria noted by Wright and Chung [1] may be determined with quantifiable uncertainties and using either the mean ice shape or the 95%-outer coverage limits. Furthermore, given the fact that the difference between the mean ice shape and the 95%-outer coverage limits is twice the local root-mean-square ice roughness, distinguishing between the mean ice shape and the 95%-outer coverage limits enables the ability to discern the difference between ice form aerodynamics effects and ice roughness aerodynamic effects, given a situation where the resolution and accuracy of the wind-tunnel balances used to measure performance are sufficient to discern the difference.

It should be noted that the SOM approach is one of several different dimensionality reduction or manifold learning strategies that could be used for ice shape characterization. For example, Laplacian eigenmaps [3] or generative topographic mapping [4] could also be used for ice shape characterization in a similar manner. The important aspect of the manifold learning strategy for ice shape characterization is that local neighborhoods of the point cloud are associated with each codebook vector as part of the learning process. These associated regions or neighborhoods of the point cloud lend themselves to the evaluation of local neighborhood statistics. This process allows the manifold to be characterized, representing the mean ice shape and the distribution of the experimental surface measurements about the manifold (i.e., the measurement noise caused by the roughness of the ice shape) to be characterized as well. Thus, the local data neighborhoods may be used to generate coverage limits that capture the characteristics of the point cloud in a statistically meaningful way.

Many questions remain regarding the repeatability of SOM application methods and the related manifold learning strategies. For example, the most important is: how many codebook vectors are needed to fully characterize a shape? Others are: if an initially random set of codebook vectors is used, is the method repeatable, and can initial shapes be used that provide repeatable and quick ice shape characterization?

In the next sections, the SOM will be applied to a rime ice shape, to a glaze ice shape formed on an airfoil at an angle of attack (AOA) (depicted in Fig. 3), to a bimodal glaze ice shape, and to a glaze ice shape with multiple upper and lower horns. Initial observations based on the ice shape characterizations are presented, and improvements and future explorations are discussed.

#### IV. Application to Selected Ice Shapes

Four ice shapes were examined using SOMs. The ice shapes were generated in the Icing Research Tunnel at NASA Glenn Research Center in Cleveland, OH. Castings were made of the ice shapes to enable better reflectivity of the laser off the ice accretion surfaces. The castings were scanned with different laser systems and used to

create point clouds of the ice shapes. Each of the point clouds consisted of well over 100,000 point measurements. Because of the memory limits with the SOM tool used with more than 30 codebook vectors, only the 50% spanwise-center section and approximately 20% of the chordwise airfoil measurements closest to the leading edge (approximately 20,000 points for each ice shape) were analyzed.

The shortened point clouds were then analyzed using the JAVA applet BSOM1 [5]. Initially, BSOM1 places a random distribution of codebook vectors in the ice shape domain. Running BSOM1 with a random initial placing of the codebook vectors produces a final ice shape characterization where the codebook vectors adequately capture the ice shape, but the points are out of order along the arclength of the manifold (nonbicycle-chained). To create ice shape characterizations with ordered points along the arclength of the ice shape, the initial points were manually placed in order along a shape that resembled the first 15% of a NACA airfoil. BSOM1 was then set to automatic learning and run until significant movement of the codebook vectors could not be detected. The resulting codebook vectors from BSOM1 and the point clouds were then read into Mathcad to generate the coverage limits.

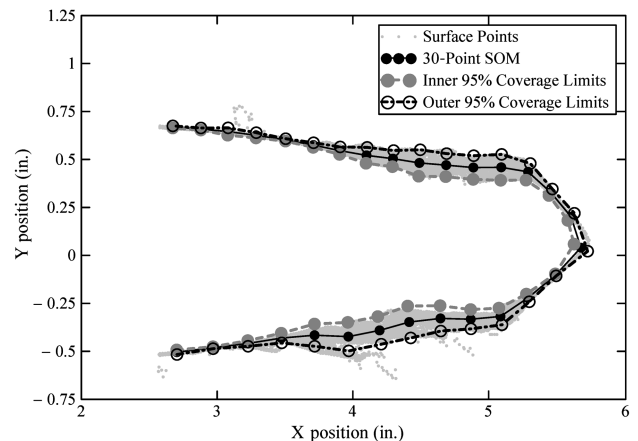
The next sections describe the results of the SOM and statistical coverage limits for the four icing shapes. Those ice shapes include 1) a rime ice shape, 2) a glazed ice shape created with the airfoil positioned with a nonzero AOA, 3) a bimodal glaze ice shape, and 4) a multihorn glaze ice shape. The ice shapes and corresponding SOM representations explore issues with the rapid and repeatable application of SOM methods to ice shapes.

##### A. Rime Ice Shape

Figure 10 presents a rime ice shape point cloud and its 30-point SOM representation. A significant observation about the SOM representation is the inability of the SOM to capture the sharp leading edge of the ice shape. Because the codebook vectors are automatically moved in the direction of the closest clump, they do not seem to capture abrupt changes in the manifold (discontinuous manifold derivatives), such as the leading edge on the rime ice shape where there is a low level of roughness. Otherwise, the codebook vectors capture the ice shape very well, qualitatively. Figure 10 also presents the 95% SOM coverage limits of the rime ice shape. While the SOM does not capture the location of the point of the leading edge of the airfoil, the computed coverage limits enclose the point cloud.

##### B. Glaze Ice Shape with Angle of Attack

Figures 3, 6, and 9 present a glaze ice shape formed on an airfoil placed in the wind tunnel with an AOA. Inspecting Figs. 3, 6, and 9, the far spanwise sections of the ice shape are bare airfoil surfaces; i.e., the edges of the airfoil have no ice accumulation. While the codebook vectors seem to represent the ice shape adequately, Fig. 9



**Fig. 10 Rime ice shape point cloud with SOM representation and coverage limits.**

demonstrates that the inclusion of these non-iced airfoil points significantly increases the coverage limits on the bottom surface of the airfoil. Since the measurements on the clean airfoil regions create non-Gaussian distributions of measurements around the codebook vectors, the coverage limits include less than 95% of the points in each of local data clump. Ensuring the coverage limits include the proper percentage of the local neighborhood points is an important test of the combined SOM-statistical point cloud representation. Since the coverage limits were constructed to be 95% coverage limits, if more or less than 95% of the local data neighborhood are included within the limits, then there are two inferences:

- 1) The data is distributed about the codebook vector in a non-Gaussian manner.
- 2) More codebook vectors should be included to represent the data.

To investigate the question of how many codebook vectors are needed to represent an ice shape, the shortened section of the ice shape of Fig. 3, without the clean airfoil sections, was analyzed using a 40-point SOM and a 50-point SOM. Figure 11 presents the 40-point SOM representation and coverage limits of a glaze ice accretion formed on an airfoil at an AOA. Figure 11 demonstrates that increasing the number of codebook vectors to 40 continues to capture the form of the ice shape, qualitatively. Figure 11 also demonstrates that eliminating the clean airfoil points that were included in the original ice shape file greatly improves the coverage limits captured by the SOM. That is, not more or not less than 5% of the shortened point cloud measurements appear outside the spline through the coverage limit points.

Figure 12 presents the results of a 50-point SOM for the shortened point cloud from the ice shape of Fig. 3. Figure 12 demonstrates that increasing the number of codebook vectors does not necessarily improve the point cloud representation. Figure 12 shows that, on the downstream section of the top horn, the codebook vectors begin to deviate about the perceived manifold in a meandering (i.e., zigzag) pattern. Figure 12 identifies a natural limiting factor of quality of SOM representation. The limiting case being that the minimum arclength between two codebook vectors should not be less than twice the local standard deviation of the data points from the manifold in order to represent the original manifold.

The fact that there is an upper bound or limitation on the number of codebook vectors that can be used to represent a noisy data set distributed about a smooth manifold initially appears to be a limitation of the combined SOM-statistical approach. However, this artifact actually reduces some of the subjectivity of the approach. The limitation on the proximity of codebook vectors to the local neighborhood standard deviation means is that, although more codebook vectors could be used for a given data set, once the distance between the codebook vectors approaches twice the standard deviation of the point cloud from the manifold, adding codebook vectors does not improve the representation coverage. This is, in

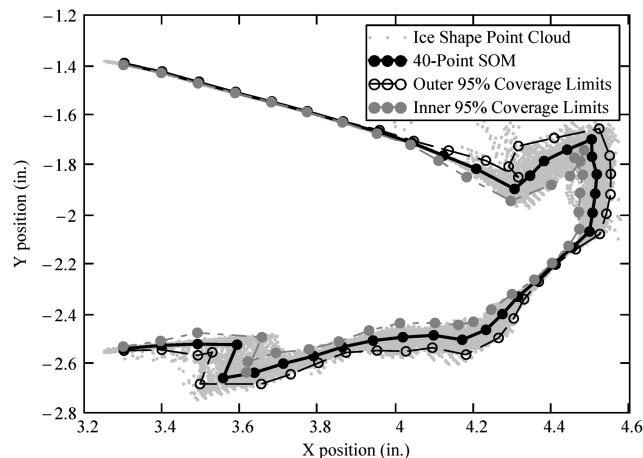


Fig. 11 Reduced glazed AOA ice shape with coverage limits of 40-point SOM Representation.

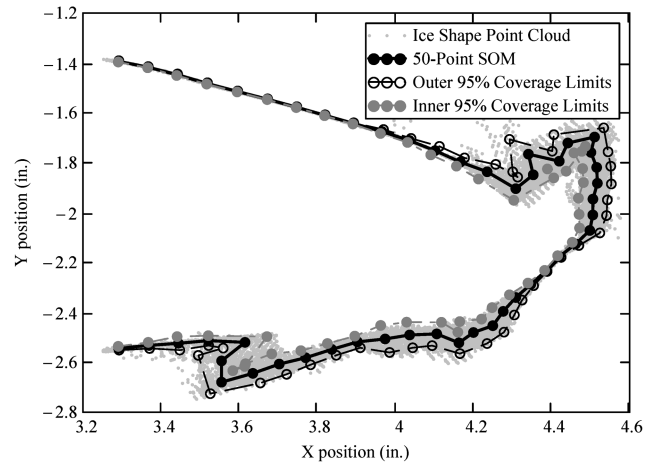


Fig. 12 Reduced glazed AOA ice shape with coverage limits of 50-point SOM representation.

essence, saying that once the distance between the codebook vectors approaches the local statistical coverage limits, both the mean ice shape and the ice roughness are characterized in a statistically meaningful way given the noisy data set.

### C. Bimodal Glaze Ice Shape

Some ice shapes have a significant variation in form, as opposed to smaller scale changes referred to as roughness, across the span of the airfoil and ice shape. In some instances, two distinct modes of ice shape can be identified within one ice shape point cloud. Figure 13 presents the SOM results and coverage limits for a bimodal ice shape. Figure 13 shows that, for the shortened point cloud from the ice shape, the outer mode is dominant. The codebook vectors most closely follow the outer mode, but the codebook vectors are slightly skewed inward because of the points following the inner mode of the ice shape. Figure 13 also demonstrates that inclusion of the inner mode points inflates the coverage limits such that the outer limit points are far outside the true scatter of the point cloud outside of the codebook vectors.

For a bimodal ice shape, a meaningful representation of the ice shape would be obtained by characterizing both the inner and outer modes separately. For example, Fig. 14 presents a 50-point SOM representation of the bimodal ice shape with most of the points following the inner mode removed from the point cloud. While a few of the points begin to deviate from the outer mode shape, the new SOM representation and coverage limits represent the outer mode manifold very well.

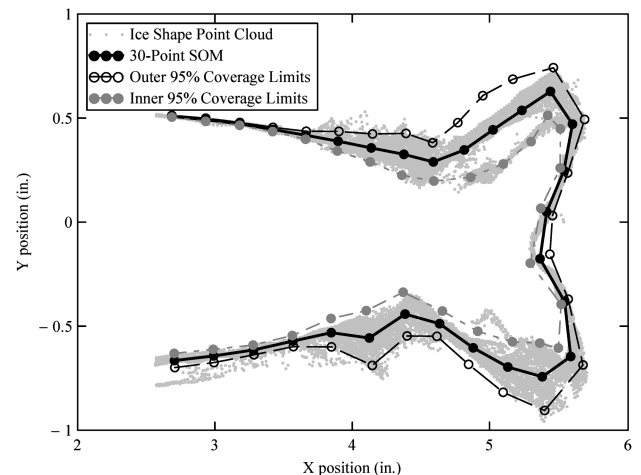


Fig. 13 Bimodal ice shape point cloud with 30-point SOM representation and coverage limits.

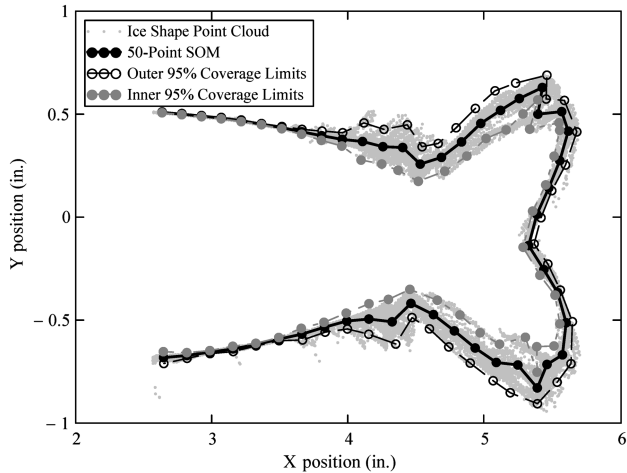


Fig. 14 Reduced bimodal ice shape with 50-point SOM representation and coverage limits.

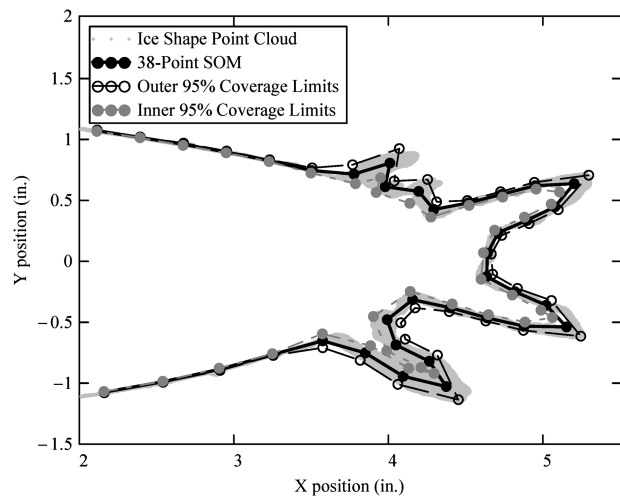


Fig. 15 Multihorn ice shape point cloud with SOM representation and 95% coverage limits.

#### D. Multihorn Ice Shape

Finally, the SOM approach was applied to a glaze ice shape that presented multiple horns on both the upper and lower surfaces of the airfoil. A 38-point SOM representation and the resulting coverage limits for an ice shape with multiple upper and lower surface horns are presented in Fig. 15. Figure 15 demonstrates that the 38-point SOM representation effectively captures the overall manifold shape and the two dominant horn features; however, the smaller features of the surface are smoothed. Also, the coverage limits on the secondary-bottom horn are somewhat inflated, because the feature is actually a spanwise-developing feature. That is, the form of the secondary-bottom horn is not constant across the shortened point cloud, but it is actually changing along the spanwise distance across the shortened point cloud.

## V. Conclusions

The neural network concept called a SOM was presented for the identification of typical ice shapes for comparison of measured wind-tunnel ice shapes with the results from ice accretion prediction codes. In addition to the ice shape characterization, a method was presented for the generation of coverage limits for the SOM representation using the statistics of the surface points around each point's winning codebook vector.

The results of the SOM characterizations and coverage limits for four ice shapes were presented. The SOMs provided captured the significant trends of each of the ice shapes studied. While significant trends were captured, observations and issues noted during the application of the SOM and coverage limit methods to the four ice shapes included the following:

- 1) There is an upper limit on the number of points able to capture the true form of the manifold (ice shape). This limit is set by the deviation of the point cloud about the manifold and the arclength along the manifold between two codebook vectors.
- 2) Errors in the SOM representation or coverage limit inflation may be caused by spurious points, noise shape artifacts (such as bare airfoil points), and bimodal ice shapes.
- 3) When low noise (roughness) is present in the ice shape, the SOM methods may smooth regions with discontinuous slopes on the manifold, such as at the tip of a rime ice shape.

Based on this initial investigation, the SOM process represents a promising approach to rigorous geometric characterization of three-dimensional ice shape point clouds. The generation of statistical coverage limits based on the SOM representation, which are the equivalent of uncertainty regions for the experimentally measured ice shapes, also constitutes a significant advancement in the validation and verification of ice accretion codes and modeling tools. Because the ultimate goal of ice shape characterization is aerodynamic performance prediction of iced vehicles, which goes beyond the geometric characterization presented in this work, future efforts are suggested regarding 1) the evaluation of aerodynamically important ice shape parameters, noted by Wright and Chung [1], using the SOM ice shape representations, and 2) the relationship of the identified parameters to iced airfoil and vehicle aerodynamic performance.

## References

- [1] Wright, W., and Chung, J., "Correlation Between Geometric Similarity of Ice Shapes and the Resulting Aerodynamic Performance Degradation: A Preliminary Investigation Using WIND," NASA CR 1999-209417; also AIAA Paper 2000-0097, 2000.
- [2] Kohonen, T., *Self-Organizing Maps*, 3rd ed., Springer-Verlag, Berlin, 2001.
- [3] Belkin, M., and Niyogi, P., "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, Vol. 15, No. 6, 2003, pp. 1373–1396. doi:10.1162/089976603321780317
- [4] Bishop, C. M., Svensen, M., and Williams, C. K. I., "The Generative Topographic Mapping," *Neural Computation*, Vol. 10, No. 1, 1998, pp. 215–234. doi:10.1162/089976698300017953
- [5] Utsugi, A., "BSOM1: Bayesian Self-Organizing Map Ver. 1.0," A. Utsugi, Ibaraki, Japan, 1996.
- [6] Tiño, P., "Intelligent Data Analysis: Topographic Maps of Vectorial Data," COMP/06-20233/LM Intelligent Data Analysis, Course Notes, Univ. of Birmingham, England, U.K., 2006, <http://www.cs.bham.ac.uk/~pxt/IDA/topmap.pdf> [retrieved 2010].