A SYSTEMATIC APPROACH FOR QUANTIFYING AND IMPROVING CFD COMPUTATIONS OF COMPLEX FLOWS

Svetlana V. Poroseva, M. Yousuff Hussaini
School of Computational Science,
Florida State University
Tallahassee 32306-4120, USA
poroseva@csit.fsu.edu, myh@csit.fsu.edu

Stephen L. Woodruff
Center for Advanced Power Systems,
Florida State University
Tallahassee 32306-2734, USA
woodruff@caps.fsu.edu

ABSTRACT

When turbulence models are used in the simulation of a flow for which experimental results do not exist, there is as yet no reliable procedure for choosing a model or for quantifying the uncertainty of results. The present study develops an approach based on evidence theory for the resolution of these issues. The results of turbulence-model validations and of predictions using those models are fused to determine the intervals in which a flow quantity is likely to fall and a measure of confidence for each interval. The approach is tested in a subsonic flow around the RAE 2822 airfoil.

INTRODUCTION

CFD simulations have become a primary tool for the prediction of complex fluid flows of scientific and engineering interest. However, to draw meaningful conclusions from results of such simulations, information about their accuracy must be available. Only recently (AIAA, 1998) has the need for a systematic analysis of simulation accuracy received much attention. The present work is an investigation into the use of evidence theory as a mathematical foundation for such an analysis, including quantitative accuracy assessments and procedures for combining the results of different simulations to produce the best possible prediction.

Ideally, the assessment of simulation accuracy would involve the identification and quantification of all sources of error and uncertainty. The results of turbulence simulations are compromised by the failure of turbulence models to describe correctly the flow physics, by model parametric uncertainty, by uncertainties and errors in the code and by numerical errors, including grid discretization and convergence issues. To identify and quantify all sources of uncertainty is generally impossible, to say nothing of quantifying the interaction between them. As a result, there is presently no standard procedure to quantify numerically uncertainties in turbulence simulations, even for flows for which experimental data are known. When predictions are to be made for flows for which experimental data are not available, systematic procedures do not exist for choosing an appropriate model or quantifying the uncertainty of the results. Evaluation of model performance, even under controlled conditions, is thus very subjective.

Developing new turbulence models will not resolve these issues. Efforts directed towards developing reliable mathematical tools to assess simulation accuracy should be made. This is the goal of our study. With the current state of art in this field, we find that a realistic approach to the problem is to quantify the total uncertainty of simulation results obtained with a given turbulence model, computational grid and code. Then, the objectives are i) to find an appropriate measure to quantify this uncertainty and ii) to develop a procedure that uses this information to improve predictions made with those simulation tools. The statistical theory known as evidence theory (Shafer, 1976) provides a systematic framework within which these goals become attainable.

MATHEMATICAL BACKGROUND

Mathematically, two types of uncertainty are recognized: aleatory and epistemic. Aleatory uncertainty is due to stochastic influences (e.g., random noise) and cannot be reduced. It is well described by probability theory. Epistemic uncertainty is subjective and originates from incomplete knowledge at any stage of modeling and simulation. Increasing one's knowledge reduces epistemic uncertainty. Theories modeling epistemic uncertainty include possibility theory, fuzzy-set theory, and evidence theory. Uncertainty in turbulence simulations results from both types of uncertainty sources (Oberkampf et al., 2001). Thus, the statistical method chosen to quantify the simulations' uncertainty should be able to handle both.

Among the few statistical theories able to work with both types of uncertainty, evidence theory (Shafer, 1976) is the most developed mathematically. Moreover, the theory does not require separation of the two types' contributions, which is often an impossible anyway. It works with the limited information available for CFD problems and new data can be incorporated as it becomes available. Therefore, evidence theory provides the statistical foundation for the present approach. Notice, though, that engineering applications of evidence theory are few (Oberkampf & Helton, 2002) and differ considerably from CFD problems. Thus, significant extensions of existing evidence theory tools and concepts are made in order to benefit fully from evidence theory and maximize improvement in prediction quality.

Comprehensive expositions of the foundations of evidence theory and its relation to other uncertainty theories have been given by Shafer (1976) (also see his later publications; Oberkampf et al., 2001). Following is a brief review of those aspects of evidence theory required for the present application.

We wish to find the true value (meaning the value or the range of values that would be measured in an experiment) of a flow quantity, in this case, the mean velocity. We denote this mean velocity by U and the set of its possible values by $\mathbf{U}$. The fundamental goal of evidence theory is to determine the degree of confidence, or *support*, that may be attributed to a proposition. In the present application, propositions are of the form "the true value of U is in ΔU", where ΔU is a subset of $\mathbf{U}$.

Whenever ΔU is interpreted as a proposition, its complement $\overline{\Delta U}$ (the set of all elements of $\mathbf{U}$ not in ΔU) must be interpreted as the proposition's negation. The set of all subsets of $\mathbf{U}$, the power set, includes the empty set $\emptyset$ (corresponding to a necessarily false proposition, since the true value cannot lie in $\emptyset$) and the entire set $\mathbf{U}$ (corresponding to a necessarily true proposition, since the true value is assumed to be in $\mathbf{U}$).

In probability theory, evidence supports either a proposition or its negation. The degree of support of a proposition is the probability; probabilities of all propositions sum to one. In evidence theory, evidence does not need to support a proposition or its negation. For instance, it can support the total set of propositions without supporting each of them separately. As in probability theory, the total support is distributed over all subsets of the power set and is equal to one:

$$\sum_{\Delta U \subset \mathbf{U}} m(\Delta U) = 1.$$

The quantity $m(\Delta U)$ is the basic probability assignment of a subset ΔU. It is the support committed exactly to ΔU. This support does not carry over to subsets of ΔU nor is it the total support of this subset. The total support of ΔU includes the basic probabilities of all proper subsets of ΔU:

$$S(\Delta U) = \sum_{\delta U \subset \Delta U} m(\delta U).$$

That is, the support committed to one proposition is committed to any subset containing it. A subset ΔU is called a focal element of a support function S over $\mathbf{U}$ if $m(\Delta U) > 0$. The union of all focal elements of a support function is called its core.

Note that the probability-theory equality

$$P(A \cup B) = P(A) + P(B)$$

for $A \cap B = \emptyset$ does not hold in evidence theory. Instead,

$$m(A \cup B) \neq m(A) + m(B)$$

and

$$S(A \cup B) \geq S(A) + S(B).$$

Also, $S(A) + S(\overline{A}) \leq 1$, in contrast to probability theory. In general, it can be said that probability theory is a special case of evidence theory for which there is no uncertainty in evidence and evidence is available in unlimited amounts.

The specification of $\mathbf{U}$ is epistemic; it is based on our current, limited knowledge, which results from any type of information or evidence: experimental data, theoretical results, expert opinion, etc. It can be changed as new evidence becomes available.

The main tool of evidence theory is Dempster's rule for the combination of support functions (Shafer, 1976). Though there is a controversy on its use in engineering problems (Oberkampf et al., 2001), we find it applicable to our case. Given several support functions over the same set $\mathbf{U}$, a composite support function is computed as their orthogonal sum. In the simplest case of two support functions S_1 and S_2 with basic probability assignments m_1 and m_2, the normalized orthogonal sum is

$$m(C) = \frac{\sum\limits_{\substack{i,j \\ A_i \cap B_j = C}} m_1(A_i) m_2(B_j)}{1 - \sum\limits_{\substack{i,j \\ A_i \cap B_j = \emptyset}} m_1(A_i) m_2(B_j)},$$

where $A_1, ..., A_k$ and $B_1, ..., B_k$ are focal elements of S_1 and S_2, respectively. The core of the support function given by m is equal to the intersection of the cores of S_1 and S_2.

To be used in Dempster's rule, support functions should satisfy some conditions (Shafer, 1976): they do not flatly contradict each other and they are based on separate sources of evidence. Independence of evidence sources is extremely important, but its definition is highly subjective (Shafer, 1990). In the present work, results from simulations obtained with different turbulence models and at different flow positions are assumed independent.

QUANTIFICATION OF SIMULATION UNCERTAINTY

The support function introduced above as a tool of evidence theory can be used to quantify the total uncertainty in turbulent flow simulations. The procedure is illustrated for the case of an RAE-2822 airfoil for which experimental data is available.

Test case

As a test flow, the subsonic flow around a RAE-2822 airfoil is chosen. Flow conditions correspond to those of Case 1 in AGARD (1979). Rather than attempt to correct explicitly for wall influences, they will be considered to be an additional uncertainty in the simulations.

Two standard turbulence models: $k - \varepsilon$ and $k - \omega$, are used in the present simulations. All calculations were made with ISAAC (Morrison, 1992), a second-order finite-volume code for solving the Favre-averaged Navier-Stokes equations. An upwind scheme based on Roe's flux-splitting is used for the convective terms and central differencing is applied to the viscous terms. Iterations are performed using an implicit diagonalized spatially split approximate factorization scheme. The grid is a nonuniform C-mesh with 257 mesh points in the wrap-around direction (with 177 points on the airfoil surface) and 97 points in the wall-normal direction. The grid extends approximately 18 chords from the airfoil. The same grid was used in all computations. More details on the code and turbulence models may be found in Morrison (1992).

Measure of result uncertainty

The accuracy with which the simulations reproduce reality is assessed through comparison with experimental data. This process is called *validation* (AIAA, 1998). In the example considered, velocity profiles calculated with turbulence models are compared with experimental data available in the y-direction, normal to the airfoil surface at two positions $x/c = 0.75$ and 0.95 along the airfoil chord (c). These positions are called *validation points*. As the metric for judging the quality of the computed results, the relative error (or deviation) is chosen. If the error ΔU_e of experimental data is known, the deviation is defined as $Dev = (U_e \pm \Delta U_e - U_m)/U_\infty$, where the choice between *plus* or *minus* signs depends on which of them returns the minimum absolute Dev-value. Unfortunately, information about experimental error is not always offered. In such cases, the deviation value is defined as $Dev = (U_e - U_m)/U_\infty$.

Obviously, the more experimental data we have, the more confidence we have in the support function we construct. It is not uncommon, though, that data are sparse. For instance, in the example we consider there are only 28 experimental values along the y-direction at $x/c = 0.95$ and 20 at $x/c = 0.75$. To reduce the spreading of limited evidence over a wide range and, thus, to increase confidence in the results, we will use the absolute value of the relative error as the measure of simulation accuracy:

$$Dev = \left| \frac{U_e - U_m}{U_\infty} \right| .$$

This introduces additional uncertainty (insensitivity to the sign of Dev) and will affect prediction quality; this is the price paid for improving the statistics.

At a validation point x/c, the deviation varies along the y-coordinate. Let N be the total number of points in the y-direction at which comparisons of calculated and experimental data are made. Dividing the deviation space $(-\infty, +\infty)$ into intervals ΔDev, one can construct at each validation point a deviation distribution based on the frequency n with which Dev-values along the y-coordinate fall in each interval ΔDev ($\sum_{\Delta Dev} n(\Delta Dev) = N$). The deviation distribution so constructed is normalized by dividing by the total number of points N: $S(\Delta Dev) = (n(\Delta Dev))/N$. At this point, the total body of evidence sums to unity:

$$\sum_{\Delta Dev} S(\Delta Dev) = 1.$$

The different number of experimental points available at different validation points is accounted for by assuming that a total support of unity corresponds to 28 points and, when fewer points (20) are available, the total support applied to specific intervals is 20/28, with the remainder (8/28) committed to the entire set of supported Dev-intervals as a whole.

The distribution $S(\Delta Dev)$ constructed in such a manner is not a probability density function in the sense of probability theory, but rather a support function in the sense of evidence theory. The difference is not only in the way we introduce uncertainty in the data into the S-distribution, but also in the way we divide the Dev-space into intervals supported by evidence. This discretization may be nonuniform and deviation intervals may intersect or may be even nested, resulting in different support functions. When new data becomes available, the support function will change further. Support functions are not unique. Here, we choose a uniform discretization of the deviation space, with Dev-intervals intersecting only at boundaries, for the sake of simplicity. The interval size is selected to produce support functions possessing the desirable properties discussed next.

Properties of the support function

Intuitively, the most favorable support function would be one focused on the smallest Dev-interval around zero. The minimum interval size would correspond to the computational grid size. Such a Dev-distribution would indicate the most accurate simulations possible with a given code, grid and turbulence model. While simulation accuracy at the validation points does not guarantee predictive quality, this is clearly a desirable feature of support functions. The availability of experimental error ranges is helpful here, because more calculation data will fall inside the error range and therefore, the interval including zero deviation will have increased support. Unfortunately, error ranges are rare in CFD applications and we do not rely on them here.

Another desirable support-function feature is *compactness*, in the sense that all available evidence is located (focused) inside a single interval. In reality, Dev-values are scattered due to errors and uncertainties in both calculated and experimental data. Compactness of a support function means that the evidence supports one interval over all others. However, different uncertainty sources can favor different deviation ranges and there is no reason to believe that some Dev-intervals will be preferred over others or that focal intervals will be close to each other. Notice, though, that the number of possible outcomes from simulations is finite, implying that the range of deviations is finite as well and so it is always possible to extract at least a single finite interval, which includes all Dev values.

One can, consequently, always construct a simple and compact support function completely focused on this interval. The price to pay for this is the discretization step, ΔDev, which for this support function can be large. The size of the discretization step becomes important when support functions are going to be used to improve results of simulations in a flow for which experimental data are not available (as will be seen later). If the influence of different uncertainty sources is not equal and the Dev-values they produce are not very scattered, a compact support function may be determined without increasing ΔDev too much.

It will be required here that support functions have a single maximum. While desirable, this feature may not be found in real cases and the restriction may be weakened in future work.

To summarize: ΔDev is chosen to be as small as possible but still yield a support function which is compact and has only one maximum. For the present example, $\Delta Dev = 0.04$ for support functions for the $k - \varepsilon$ model at both validation points and for the $k - \omega$ model at $x/c = 0.75$. For the $k - \omega$ model at $x/c = 0.95$, the only way to get a compact support function is to set ΔDev be equal to the entire range of Dev-values ($\Delta Dev = 0.28$).

Finally, two support functions for the $k - \varepsilon$ model at $x/c = 0.75$ and 0.95 and two support functions for the $k - \omega$ model at the same points are constructed. They are shown in Fig. 1. The reason for constructing two support functions for each turbulence model is that below we will need support functions for different turbulence models at different x/c-positions to be able to use another tool of evidence theory – Dempster's rule.

PREDICTION

Support functions corresponding to flow simulations with different turbulence models, grids and codes can be used by themselves to describe quantitatively the uncertainty introduced in results of simulations by varying simulation tools. Here, however, we will focus on exploring perspectives of the support function application to improve the results of flow prediction.

Prediction is defined in the AIAA Guide (1998) as the use of a CFD model to foretell the state of a physical system under conditions for which the CFD model has not been validated. Usually, it means that experimental data for a flow (or a part of the flow) are not available. There is as yet no reliable procedure for choosing a model for such simulations or for quantifying the uncertainty of the results. The approach developed here allows one to fuse information from several experts, such as different turbulence models (or grids, codes etc.), instead of making a subjective choice between them. It is expected that the overall credibility of predictions can be increased in this way (Shafer, 1976; Hemsch, 2002).

In the application to the test case, the goal is to improve the mean velocity prediction at a prediction point using i) information from the validation of the turbulence models at two validation points ($x/c = 0.75$ and 0.95) and ii) results of calculations by both models at the prediction point ($x/c = 0.9$). Experimental data for the velocity profile $U(y)$ are assumed not to be known at the prediction point. The grid and the code were the same in all simulations. This is not necessary, though.

The procedure is the following. At each y, the space of all possible velocity values $(-\infty, +\infty)$ is divided into intervals ΔU to which are assigned the degrees of support obtained at the validation points. A deviation of zero in a deviation support function is aligned with a calculated velocity value U_m at each y. Because the sign of the deviation is lost when the support functions are constructed, the deviation support function is applied symmetrically on both sides of U_m. The resulting support function is called the velocity support function. It is emphasized that when sufficient experimental data is available to allow one to preserve the deviation sign while constructing a support function, both the deviation and velocity support functions would coincide. Because for each turbulence model there are two deviation support functions corresponding to two different validation points, we obtain two different velocity support functions for each U_m.

As an example, a result of application of the support function obtained for the $k - \varepsilon$ model at the validation point $x/c = 0.75$ to the velocity profile calculated with the same model (black solid line) at $x/c = 0.9$ is shown in Fig. 2. Areas inside the band of colors have various non-zero degrees of support; that is, the true velocity profile is expected to occur there. Different colors correspond to different degrees of support.

Supported areas (or areas where the true velocity profile is expected to be found) obtained for two turbulence models at the prediction point, may be combined using Dempster's rule of evidence theory. There are two velocity support functions at each y-position for each model; we overcome this ambiguity in the following manner. To fuse results of simulations with two models, Dempster's rule requires support functions for each model to be from independent sources. Therefore, we apply this rule twice. One solution ($R1$) is obtained by combining the velocity support function for the $k - \varepsilon$ model obtained using the deviation support function from $x/c = 0.75$ and the velocity S-function for the $k - \omega$ model obtained using the deviation S-function from $x/c = 0.95$. This solution is shown in Fig. 3a. Solution $R2$ uses the deviation support function for the $k - \varepsilon$ model from $x/c = 0.95$ and the deviation S-function for the $k - \omega$ model from $x/c = 0.75$ (Fig. 3b). The bands of colors in Fig. 3 show the areas of nonzero support. It is not expected that the true velocity profile will be found outside these bands.

Statistically, solutions $R1$ and $R2$ are equally likely. Then, they are combined so as to average the information they carry. Details on the averaging procedure can be found in Poroseva et al. (2005). Here, we will only present the averaged solution $R12$ (Fig. 4).

From the solution $R12$, one can select at each position y the single interval with the maximum degree of support. Connecting such intervals along the y-direction, the path of maximum support can be extracted. This path is the most probable candidate to include the true velocity profile. This path should be smooth, but quite possibly will not be. For instance, the path obtained in this manner from the $R12$-solution is not smooth (Fig. 5a). Also, the true velocity value can only be inside a single velocity interval at each y, while the averaged solution can have several extrema (Fig. 4). One of the reasons is again the fact that the sign of deviation was lost during constructing the support functions and therefore, resulting velocity support functions are symmetrical. Both observations suggest applying additional smoothing procedure (Poroseva et al., 2005) to the averaged solution prior extracting the path of maximum support. The smoothed $R12$-solution (not shown here) was obtained in two iterations. The path of maximum support extracted from the smoothed $R12$-solution is given in Fig. 5b.

This is the final prediction of the approach considered in this study. In the figure, the path is compared with the velocity profiles calculated by the $k - \omega$ model (dashed line) and by the $k - \varepsilon$ model (solid line). Also, experimental data (black squares) are shown in the figure to assess the quality of the prediction. It is seen that the $k - \omega$ model result is far from the experimental values, whereas the $k - \varepsilon$ model is in very good agreement with the experiment. Our approach combines results of both models and yet, its prediction is also in very good agreement with experiment. It shows a good potential of the approach to "weight" correctly differing experts' opinions. Also, in contrast to the $k - \varepsilon$ model result, our approach produces not just a single line, the accuracy of which cannot be estimated in the absence of experimental data, but zones with well-defined degrees of support. This is an obvious advantage of the present method in comparison with a single model prediction.

CONCLUSIONS

A new approach for quantifying uncertainty of results in turbulence modeling and for using this information to improve the quality of prediction in untested conditions is developed. The approach is an alternative to a subjective choice of a turbulence model to simulate flows in situations where experimental validation is not possible. The approach relies on the mathematical tools of evidence theory, which appear to be effective in this application. Results in the application of the approach to a subsonic flow around the RAE 2822 airfoil are encouraging, but this work should be considered as an initial step in testing the approach. In the future, we are planning to apply the approach to other cases of the RAE 2822 flow, introduce a mathematical description of uncertainty originating from the distance between validation and prediction points and consider the prediction of other flow parameters, as well as predict a flow around an airfoil using results of validation of turbulence models in flows around other types of airfoils.

REFERENCES

AGARD, 1979, "Experimental Data Base for Computer Program Assessment," AGARD AR-138.

AIAA, 1998, "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations", American Institute of Aeronautics and Astronautics, AIAA-G077-1998, Reston, VA.

Hemsch, M., 2002, "Statistical analysis of CFD solutions from the Drag Prediction Workshop," AIAA 2002-0842.

Morrison, J. H., "A comprehensive Navier-Stokes solver with two-equation and Reynolds -stress turbulence closure models," NASA CR 4440, 1992.

Oberkampf, W. L., Helton, J. C., and Sentz, K., 2001, "Mathematical Representation of Uncertainty," AIAA 2001-1645.

Oberkampf, W. L. and Helton, J. C., 2002, "Investigation of Evidence Theory for Engineering Applications," AIAA 2002-1569.

Poroseva, S. V., Hussaini, M. Y., and Woodruff, S. L., 2005, "On Improving the Predictive Capability of Turbulence Models Using Evidence Theory," AIAA-2005-1096.

Shafer, G., 1976, "A Mathematical Theory of Evidence," Princeton, NJ: Princeton University Press.

Shafer, G., 1990, "Perspectives on the Theory and Practice of Belief Functions," *International Journal of Approximate Reasoning*, Vol. 4, pp. 323-362.

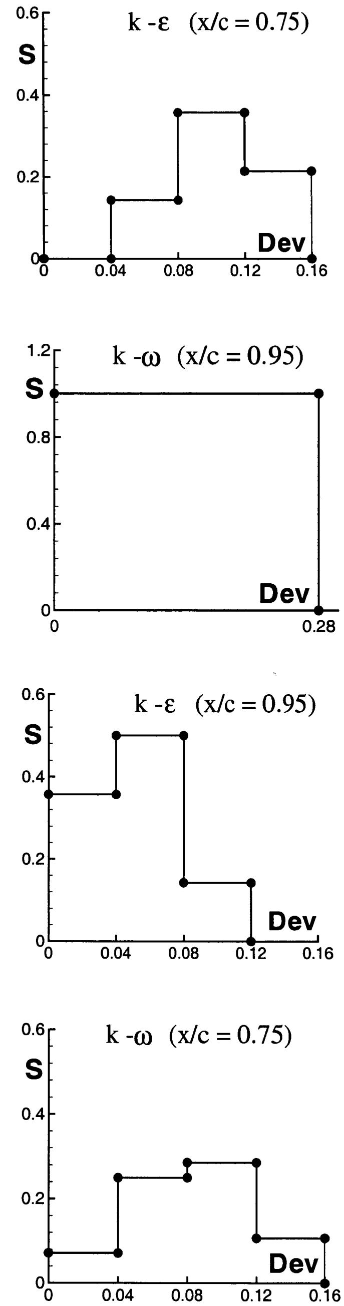

Figure 1: Support functions for turbulence models at two validation points

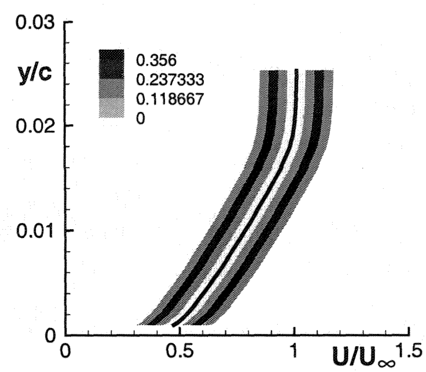

Figure 2: Area of possible true velocity values with nonzero support built around the $k-\varepsilon$ model profile at $x/c = 0.9$.

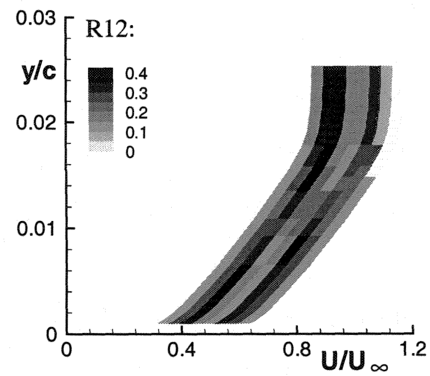

Figure 4: Averaged solution.

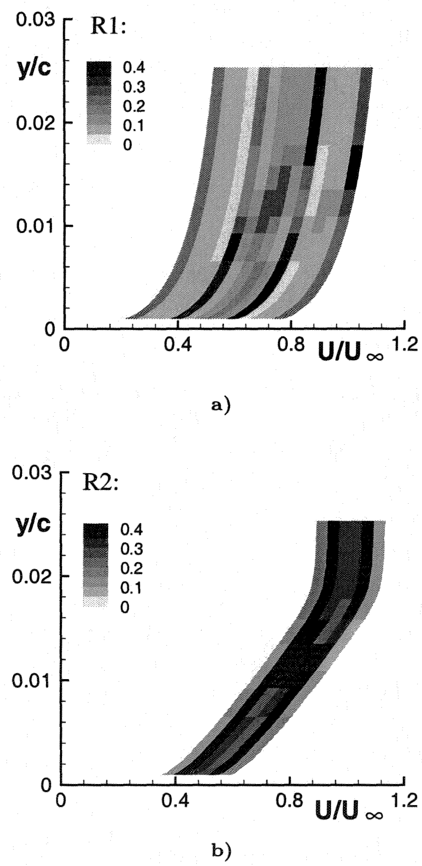

Figure 3: Two equally-likely supported velocity areas resulting from Dempster's rule application.

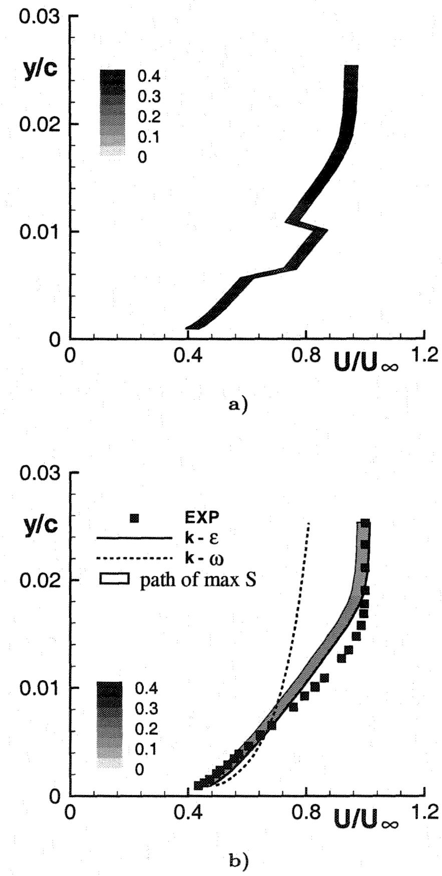

Figure 5: Path of maximum support: a) before smoothing the R12-solution, b) after smoothing.