

Ice Shape Comparison Study and Survey

Ariane Vieira, Alessandro Donizetti

AeroTex GmbH; Politecnico di Milano

Presented 21 September 2026

CC BY-NC-ND 4.0

This work is licensed by the author(s) under a **Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence (CC BY-NC-ND 4.0)**. To view a copy of this licence, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Suggested citation: Vieira, A., Donizetti, A. (2026). *Ice Shape Comparison Study and Survey*. Presented at the 3rd AIAA Ice Prediction Workshop, online, 21 September 2026.



Ice Shape Comparison Study and Survey

Ariane Vieira — AeroTex GmbH

Alessandro Donizetti — Politecnico di Milano

3rd Ice Prediction Workshop

21/09/2026

Outline

Part 1: The ϵ similarity parameter (AeroTex) · ≈ 10 min

- How the score is built, its strengths and weaknesses

Part 2: The Jaccard-based comparison (Politecnico di Milano) · ≈ 15 min

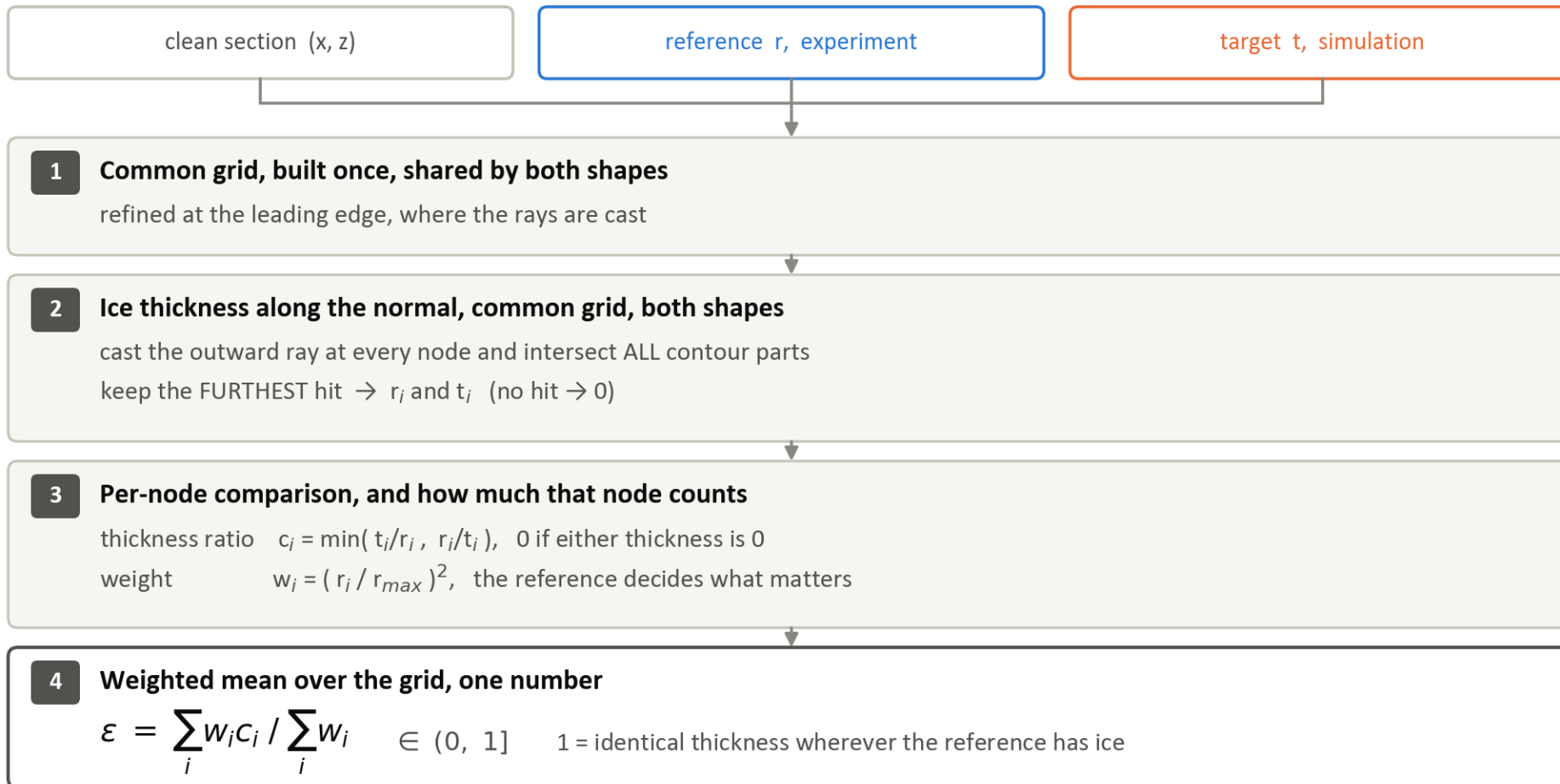
- Geometric comparison and local weighting

Part 3: The community survey (AeroTex) · ≈ 15 min

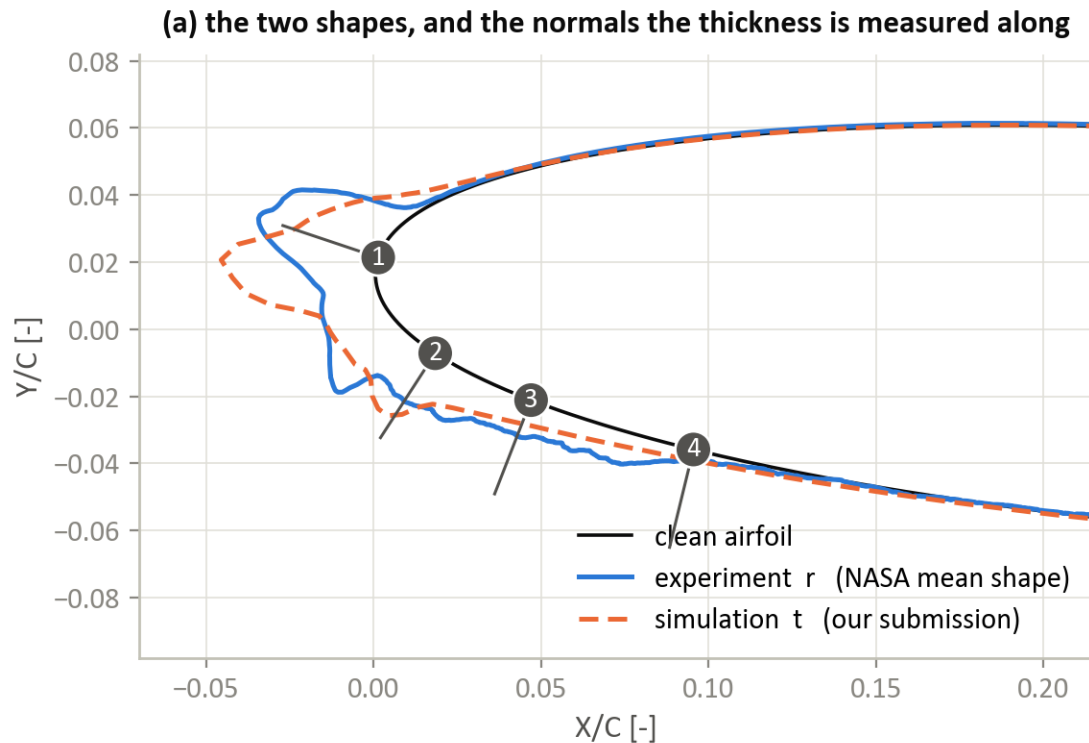
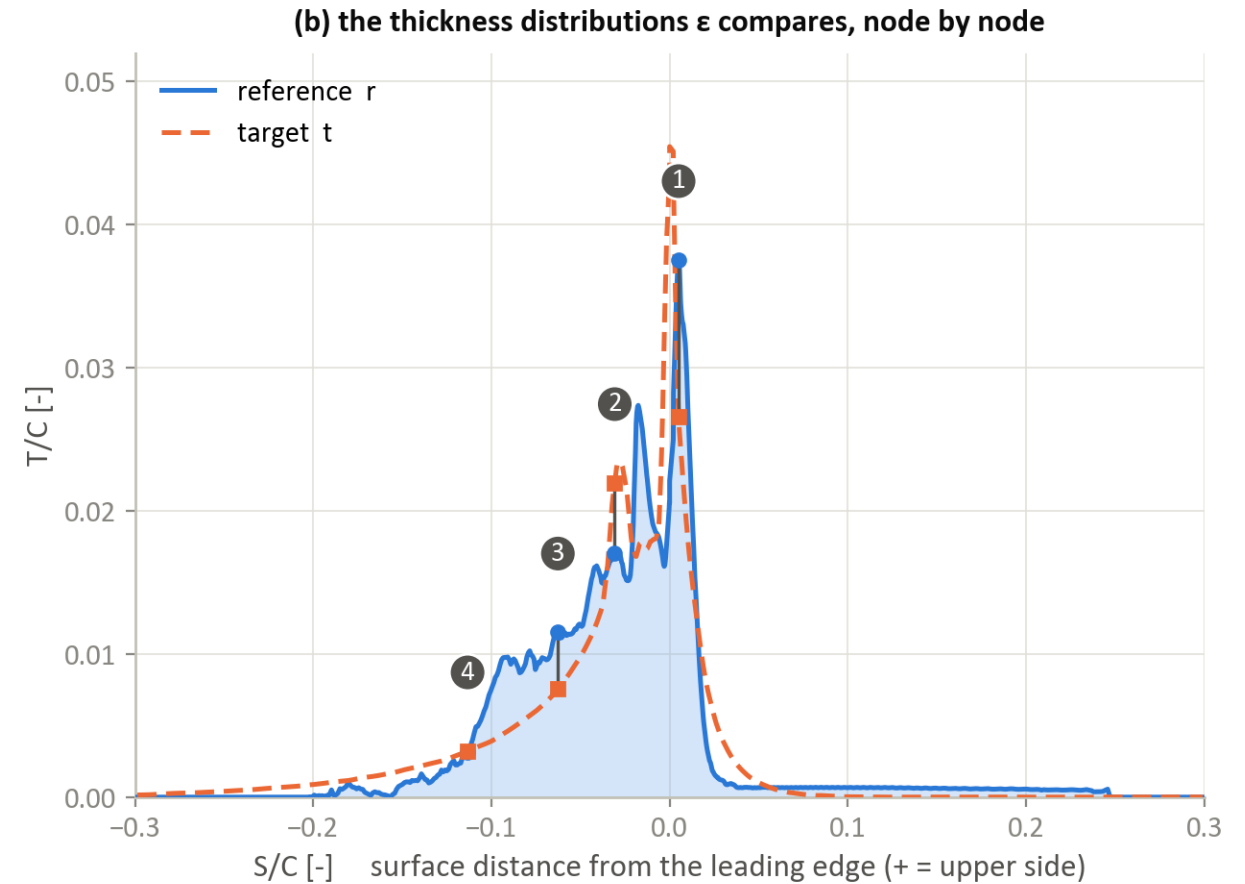
- Do the numbers agree with the community? A proposed calibration for ϵ

Part 1: The ϵ similarity parameter

How ε is computed

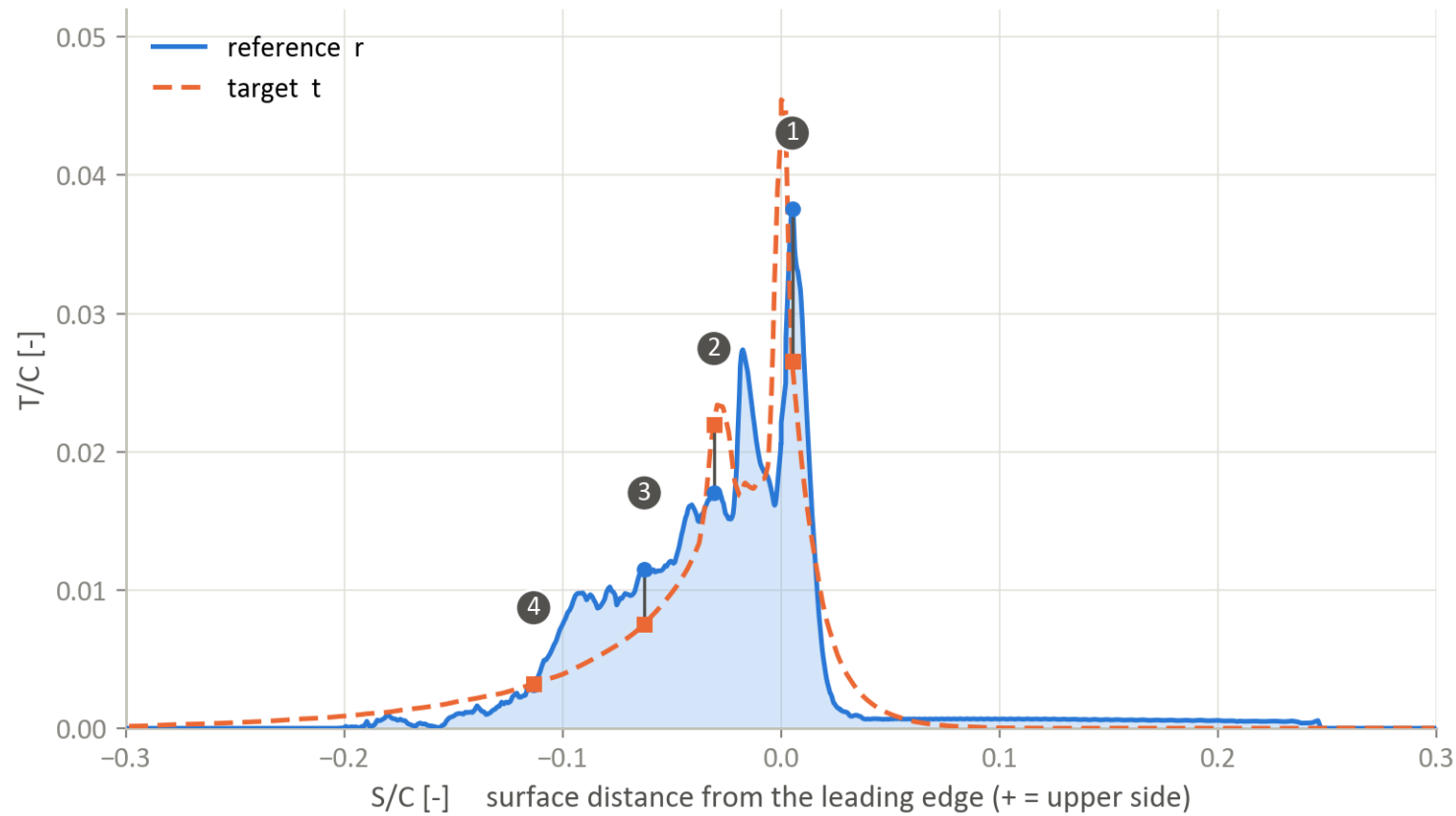


ϵ example: AE3932 vs Exp. Mean Shape



ε example: AE3932 vs Exp. Mean Shape

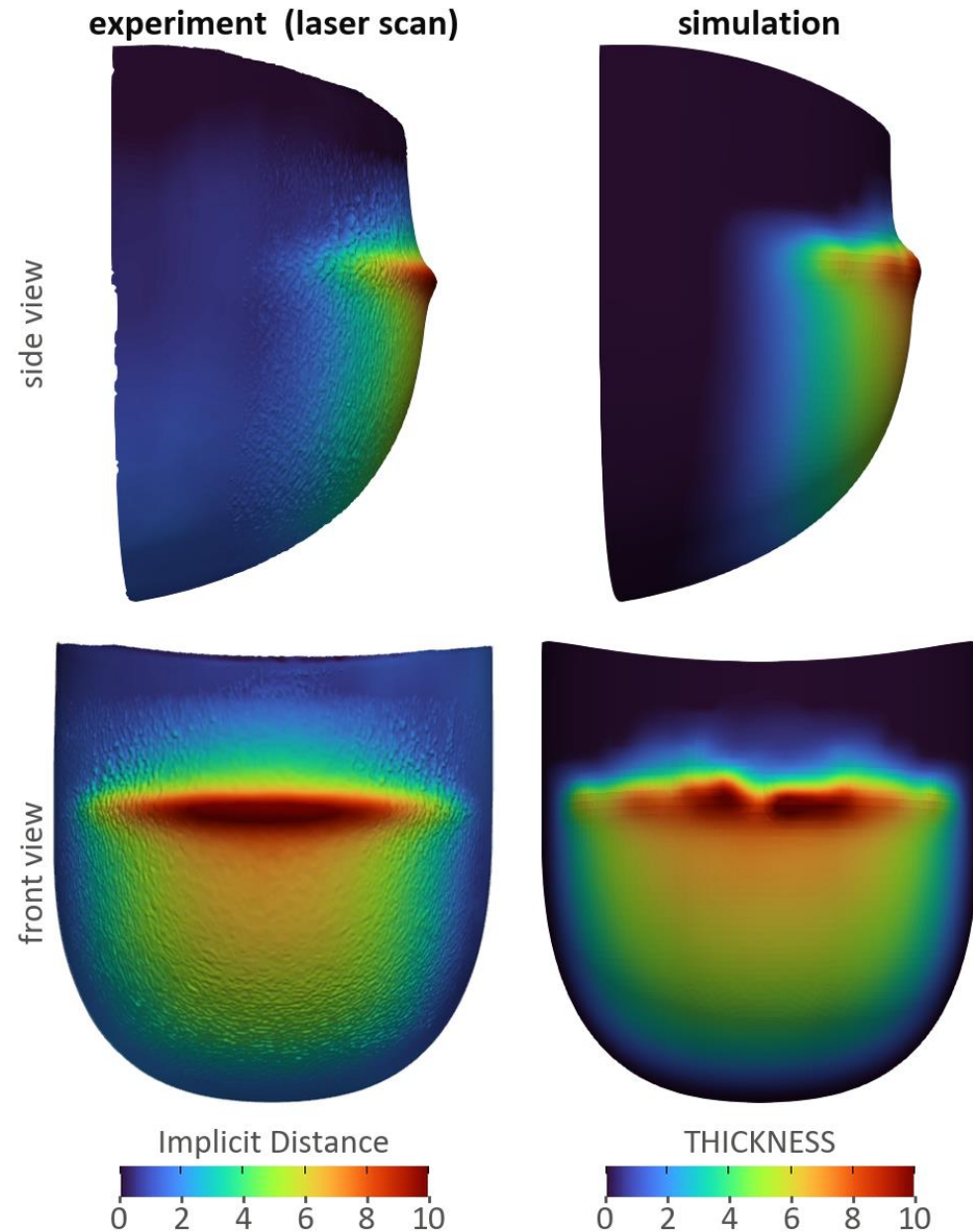
AE3932: our own submission (002) vs the NASA mean experimental shape



$\varepsilon = 0.731$ over all 2001 nodes

ε example: 3D

- Same algorithm applies
 1. Common base mesh
 2. Per-node thickness values
 3. Per-node ratios
 4. Final calculation
- Example: 3D-ICESIM radome
 - $\varepsilon = 0.81$



Strengths and weaknesses

Strengths

- Threshold-free and deterministic: no tolerance band, no tunable constant, the same rule for every case
- Easy to compute and reproduce

Weaknesses

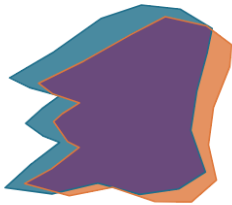
- One-way: the reference sets the grid and the weights; ice outside the reference footprint is not penalized (see later slides for alternative)
- Known from the first survey: ϵ reads higher where agreement is poor compared to experts

Part 2: The Jaccard-based comparison

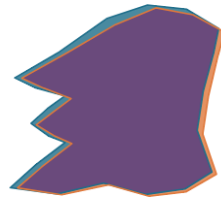
The Jaccard index



Poor match
 $J=0.38$

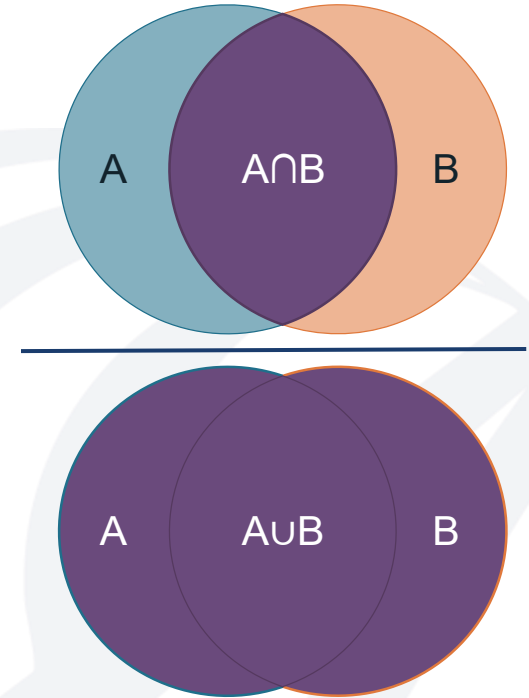


Good match
 $J=0.68$



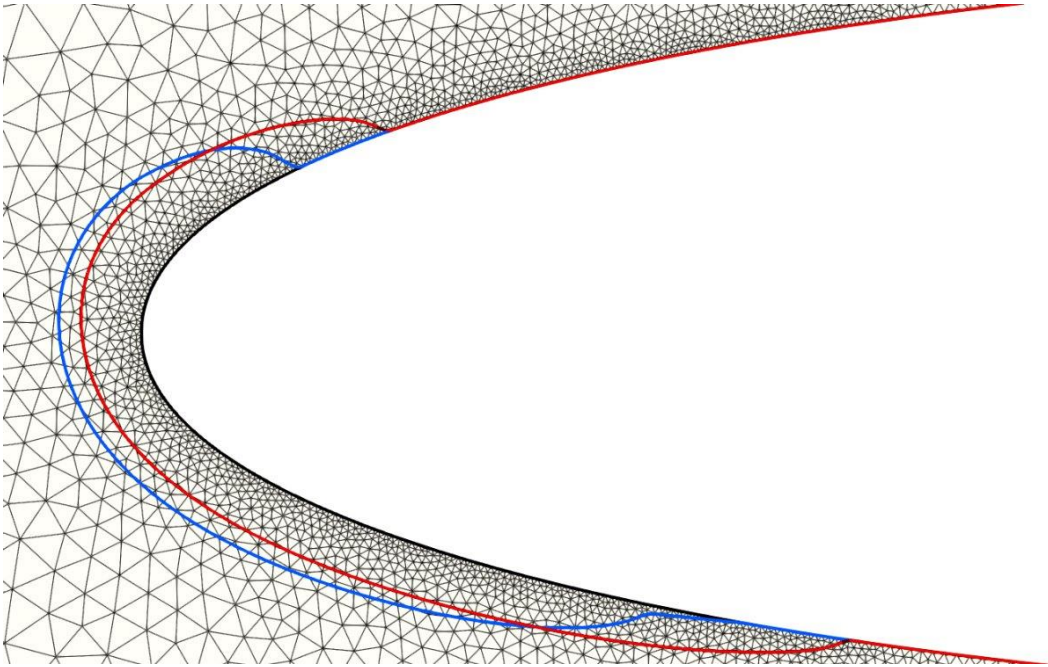
Excellent match
 $J=0.9$

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

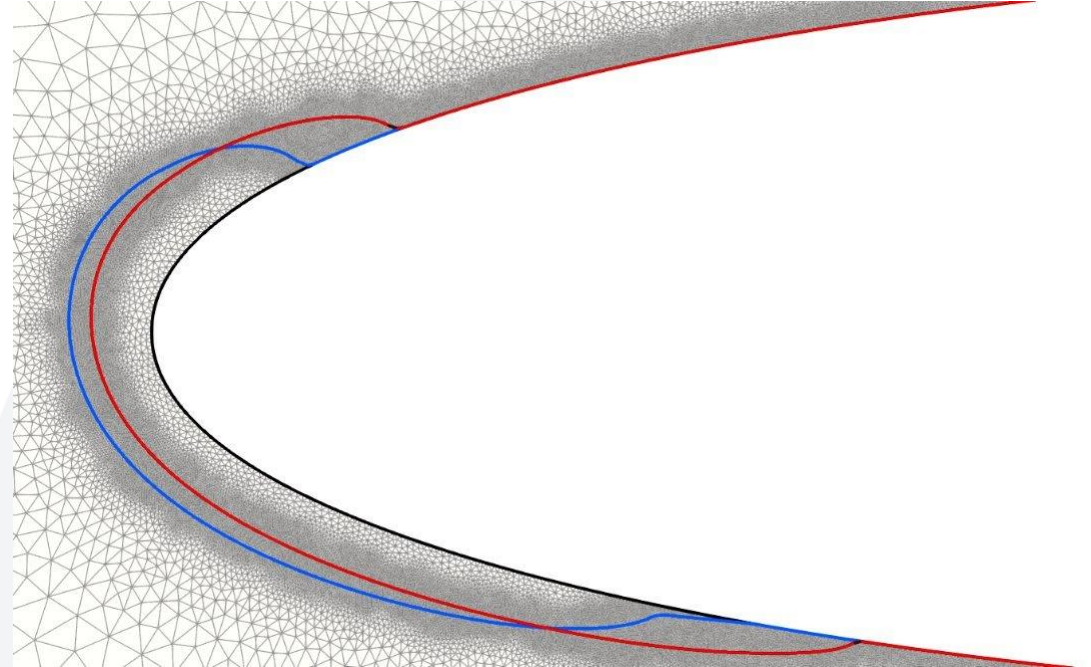


The **Jaccard index** is a statistic used for gauging the **similarity** and **diversity** of sample sets. It is defined in general taking the ratio of two sizes (areas or volumes), the intersection size divided by the union size. It is also called **intersection over union (IoU) index**.

Background adaptive mesh



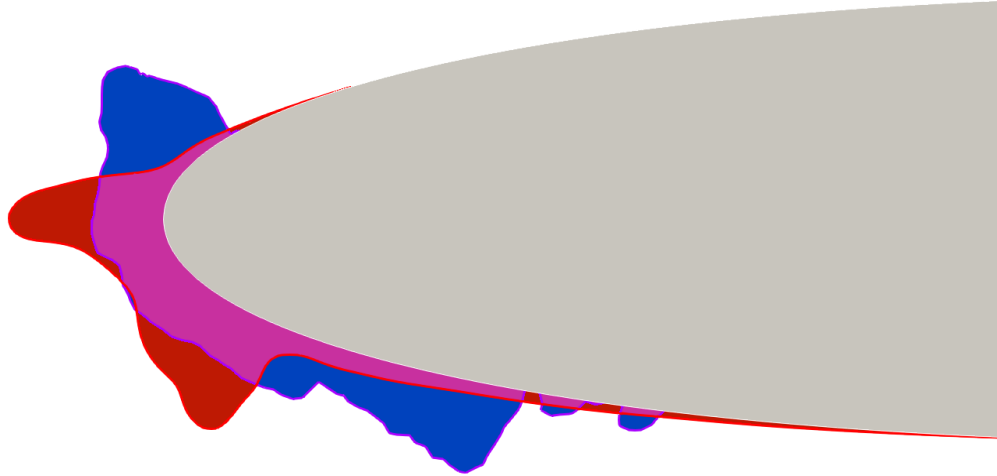
Background mesh



Refined mesh

Higher mesh resolution near the ice-shape boundary improves the accuracy of the discretized intersection/union areas/volumes used for the computation of the Jaccard index.

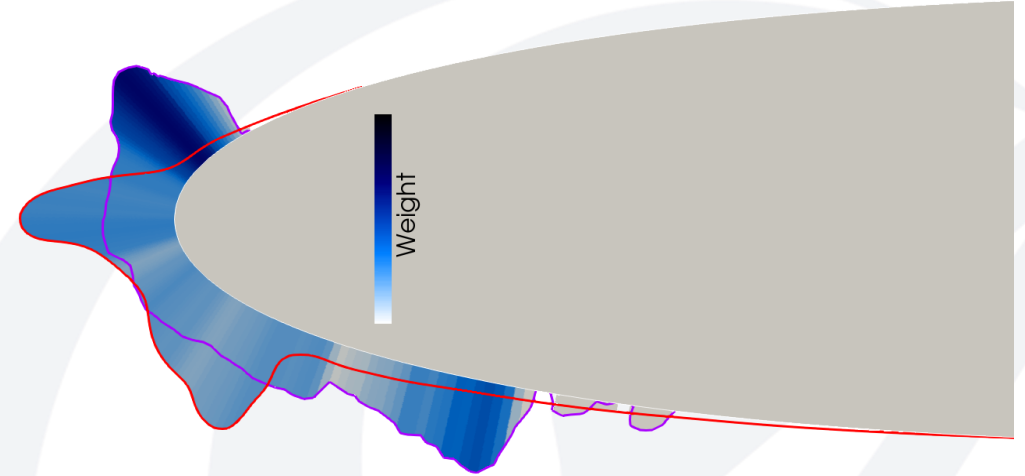
Local Weighting



$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum_{j=1}^N a_{j11}}{\sum_{j=1}^M a_{j01} + \sum_{j=1}^L a_{j10} + \sum_{j=1}^N a_{j11}}$$

No weight (pure geometrical overlap)

$J = 0.38$



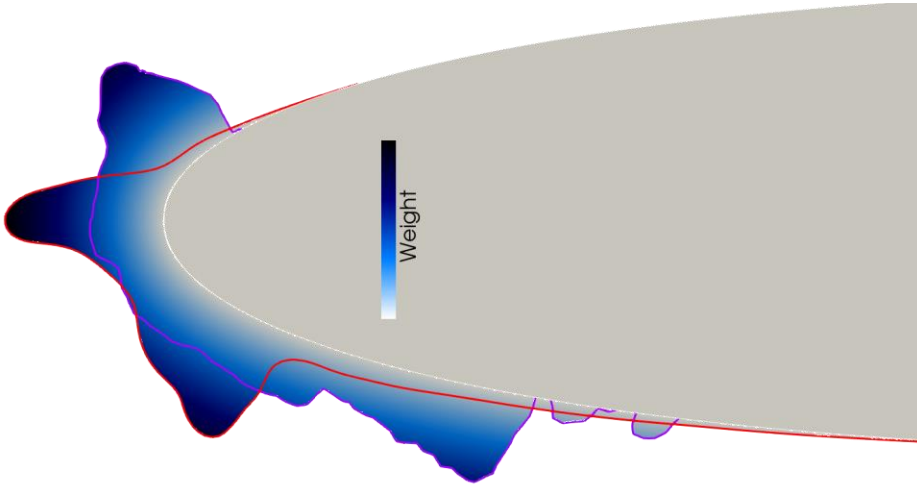
$$J = \frac{\sum_{j=1}^N a_{j11} x_j}{\sum_{j=1}^M a_{j01} x_j + \sum_{j=1}^L a_{j10} x_j + \sum_{j=1}^N a_{j11} x_j}$$

Weight = reference ice thickness

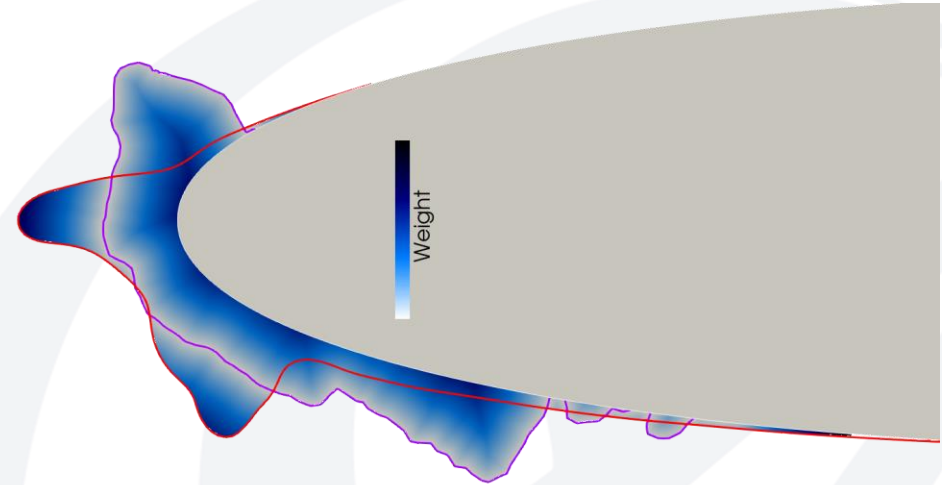
$J = 0.36$

Local Weighting

$$J = \frac{\sum_{j=1}^N a_{j11} x_j}{\sum_{j=1}^M a_{j01} x_j + \sum_{j=1}^L a_{j10} x_j + \sum_{j=1}^N a_{j11} x_j}$$



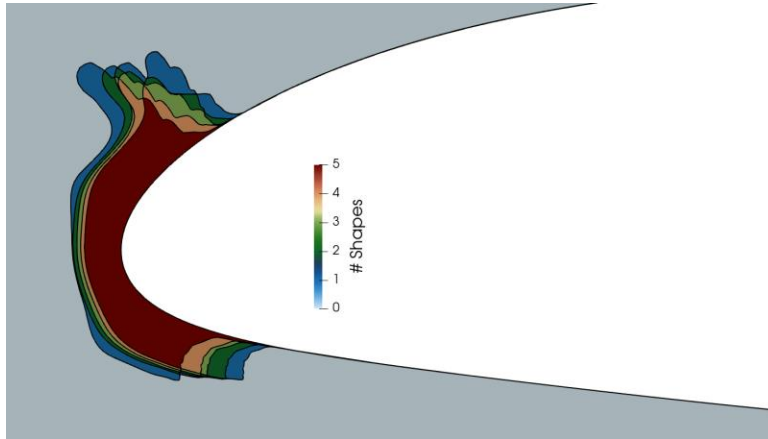
Weight = distance from clean surface
 $J = 0.21$



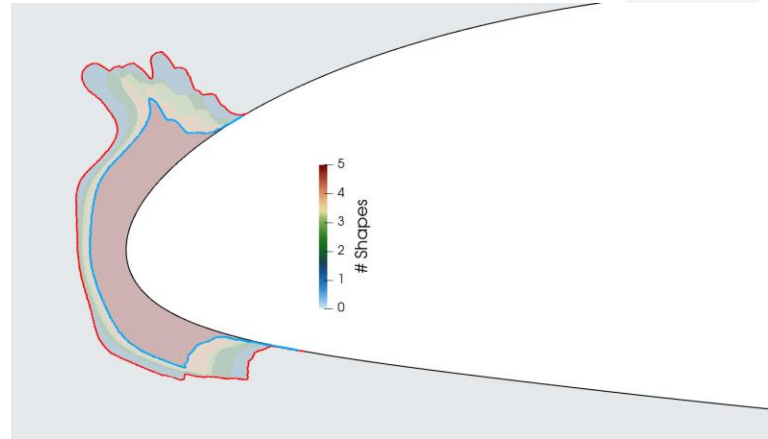
Weight = distance from reference shape
 $J = 0.44$

The choice of local weighting can significantly change the score J . Different weights can be chosen for different applications (Aerodynamic impact, IPS efficiency, etc...)

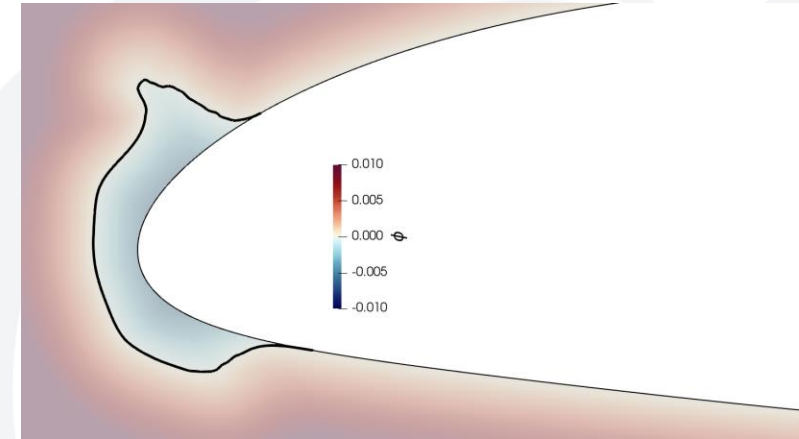
Characterization of ice shapes



Multiple ice shapes

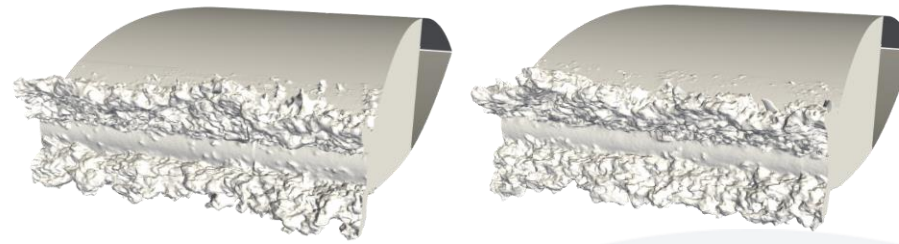


Minimum combined cross section (mCCS)
Maximum combined cross section (MCCS)

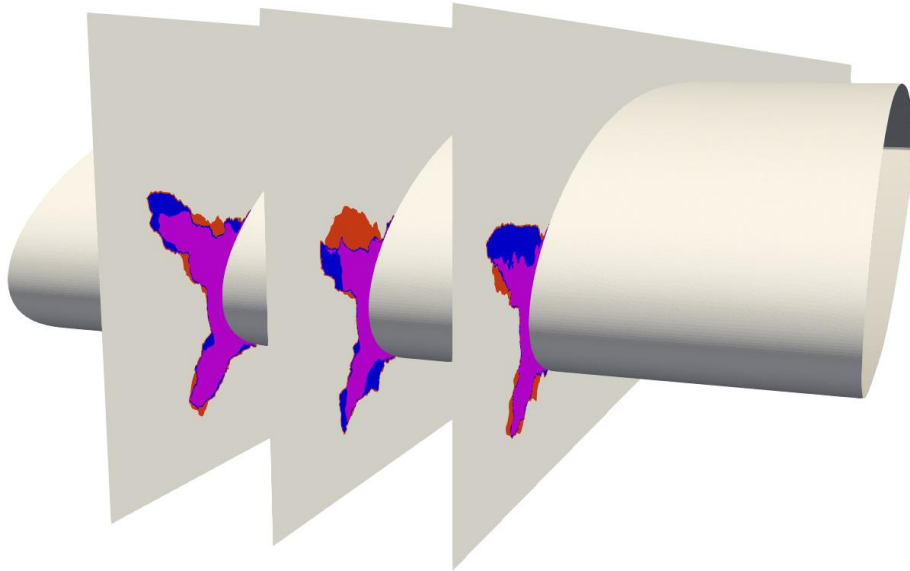


mean Shape

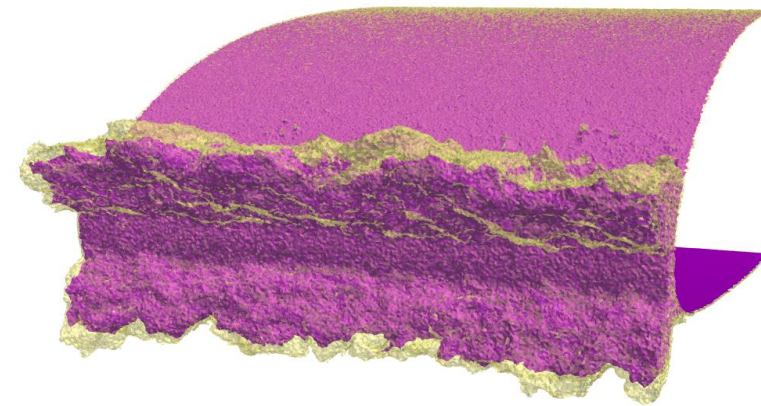
3D Extension



Case 242 from 1st IPW



$J = 0.67$



3D minimum and maximum Combined Shape (mCS and MCS)

Strengths and weaknesses

Strengths

- Robust and deterministic, local weighting can be chosen on specific test cases
- Every ice shape can be analysed, no clean-iced surface mapping required
- Can be employed to characterize ice shapes (mCCS and MCCA) and compute mean and standard deviation

Weaknesses

- Needs a closed ice shape (closed loop in 2D / watertight surface in 3D)
- Needs the creation of a background mesh (currently BF based on pointwise scripts, coming soon a IBM open-source version with AMR)

Part 3: The community survey

Comparison of Ice Shapes

- Code-vs-experiment comparison has previously been done either by visual inspection or individual parameters
- Our proposed ϵ comparison parameter is fast and simple to compute, but tends to be too generous on lower scorings¹
- Jaccard method proved to align better on the lower scorings, but demands extra calculations¹
- But can the community itself reach any consensus on a single-number evaluation?
 - → A survey was generated to investigate this aspect

• ¹Geometrical Comparison and Characterization of 2D and 3D Ice Shapes. Donizetti et al., AIAA 2026-4534 (Eq. 1)

Survey Responses

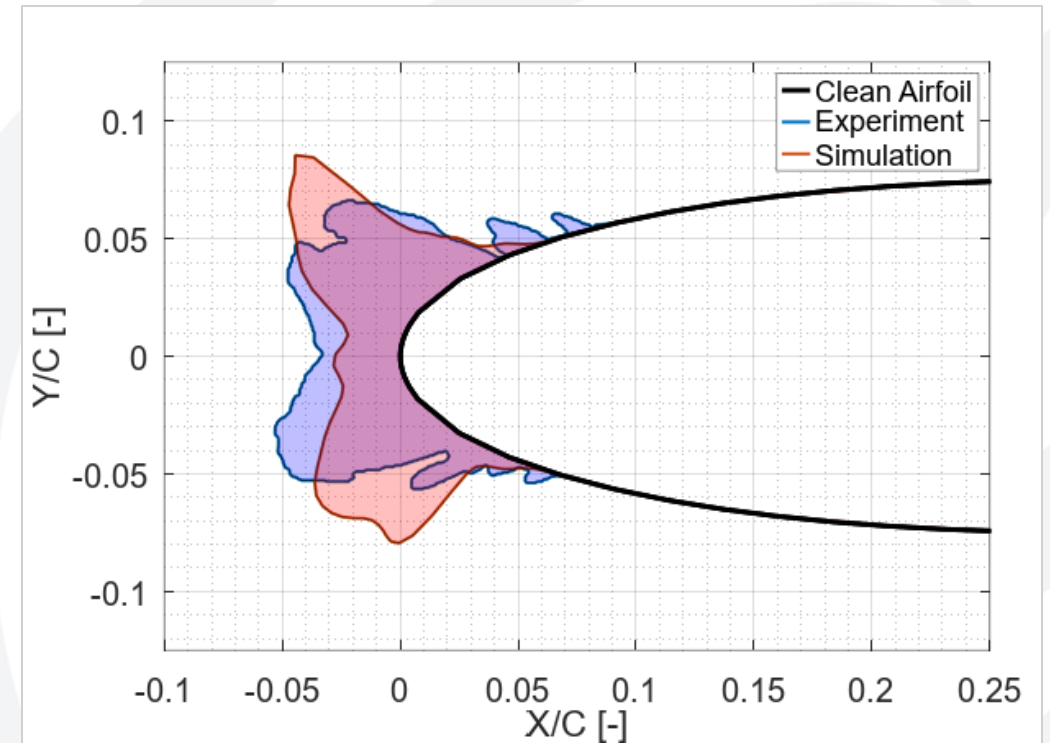
57 comparison cases from 5 databases

- LEWICE validation · IPW1 · IPW2 · AIAA 2017 · JOICE
- rime · mixed · double horn · glaze runback · protected surface (thermal anti-ice)
- cases chosen to span the whole range: clearly good and clearly poor matches were included

69 respondents scored each case 1–10

- experience: none (5) · 0–3 yrs (14) · 3–10 yrs (22) · >10 yrs (28)

3,922 individual judgments



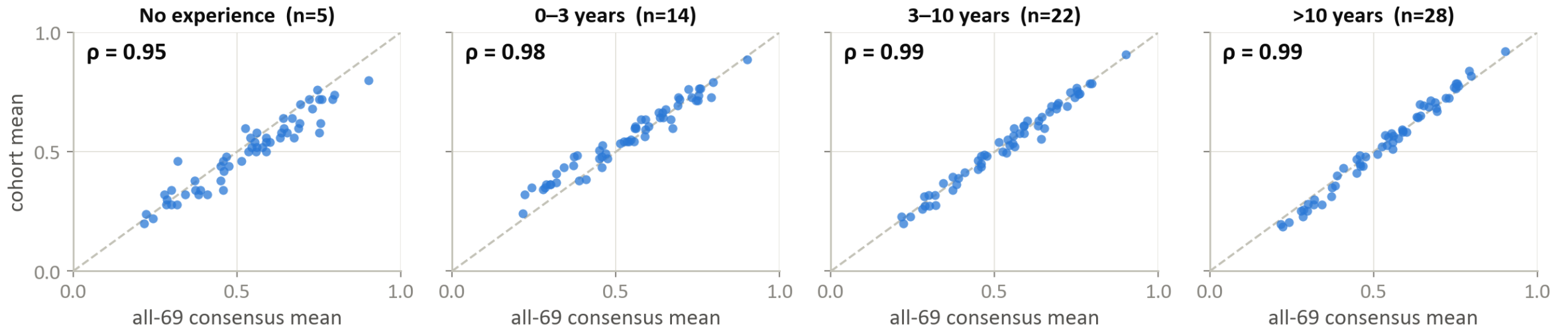
One of the 57 case pairs, as it was shown in the survey

Sources — LEWICE validation database: Wright, NASA/CR-1999-208690 (1999) · AIAA 2017: Trontin et al., “Description and assessment of the new ONERA 2D icing suite IGLOO2D”, AIAA 2017-3417 · JOICE: Breitfuß et al., “Comparison of Numerical Simulations with Experimental Data for an Electrothermal Ice Protection System in Appendix O Conditions”, SAE 2023-01-1396

Ranking of Results

The community agrees on the ranking order

Per-case survey means by experience cohort vs the all-69 consensus — 57 cases

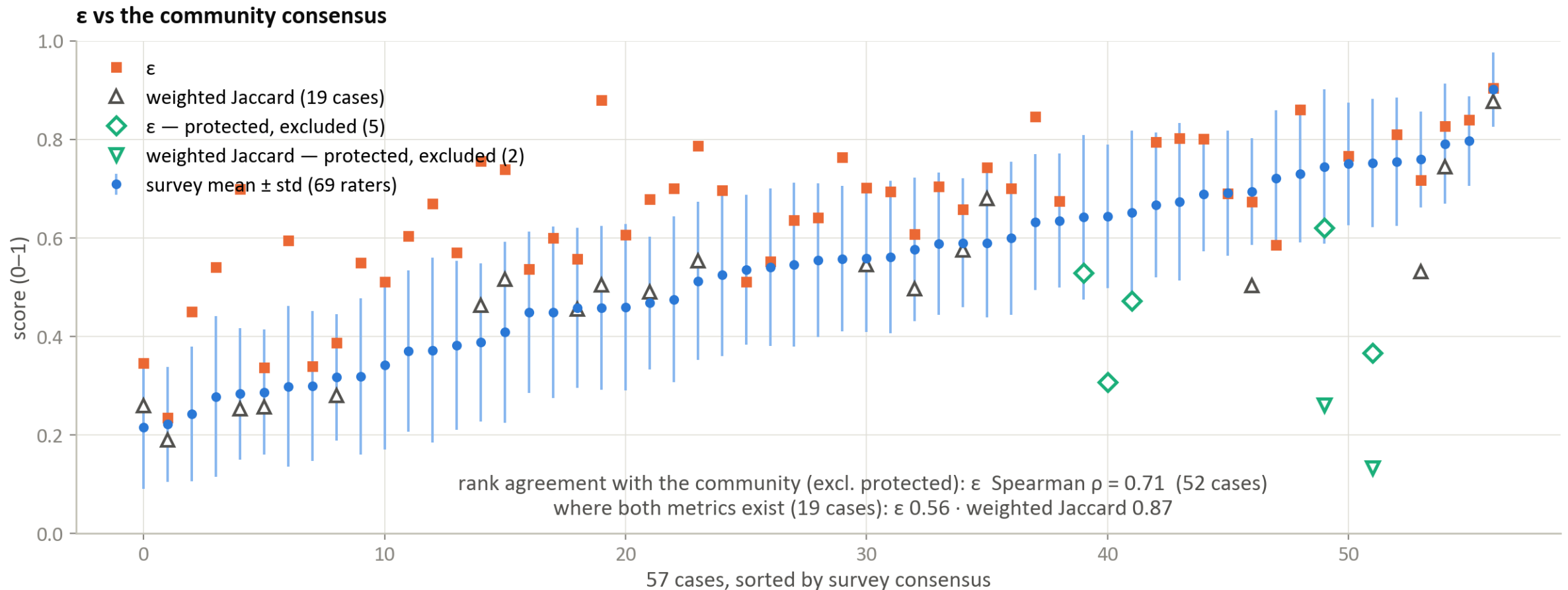


Kendall's $W = 0.69$: how similar are 69 independent rankings of the 57 cases? (0 = unrelated, 1 = identical)

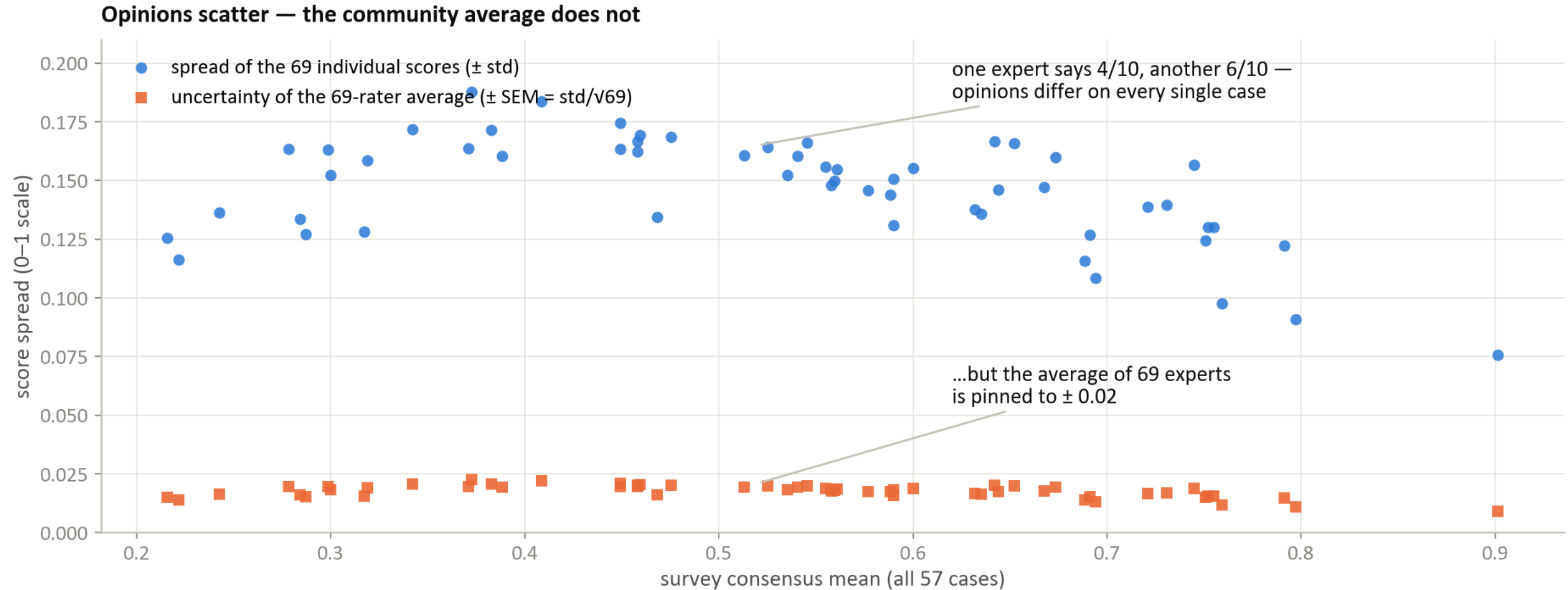
any two experts' orderings agree at $\rho \approx 0.7$ · the consensus mean is extremely stable ($\alpha = 0.99$)

experience does not matter: 0 of 57 cases differ across cohorts

ϵ /Jaccard vs. the community survey



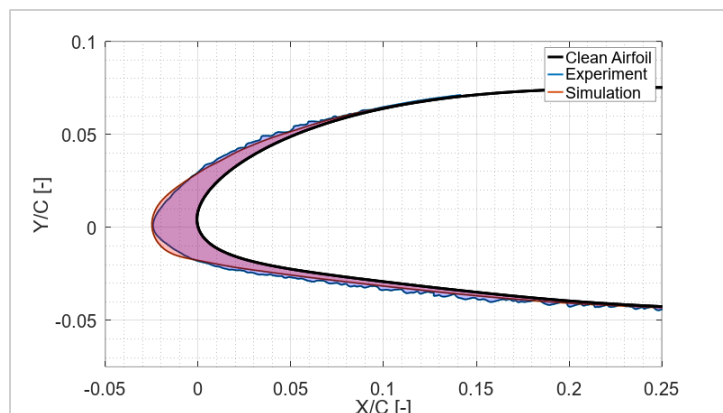
Error and Data Spread



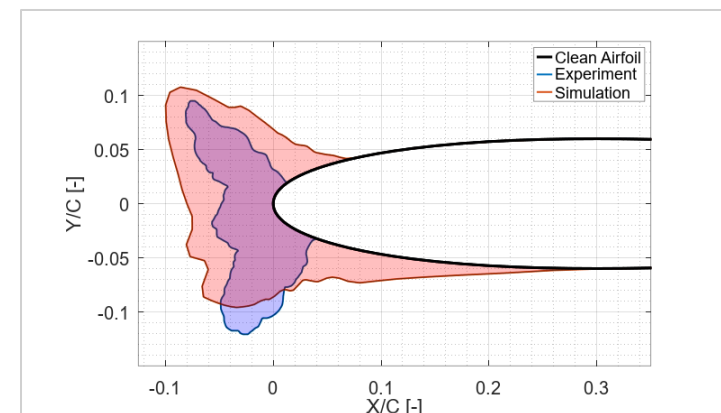
per-case scores spread ± 0.15 (std): opinions genuinely differ · the 69-rater average is precise to ± 0.02 (SEM = $\text{std}/\sqrt{69}$)

What agreement and disagreement looks like

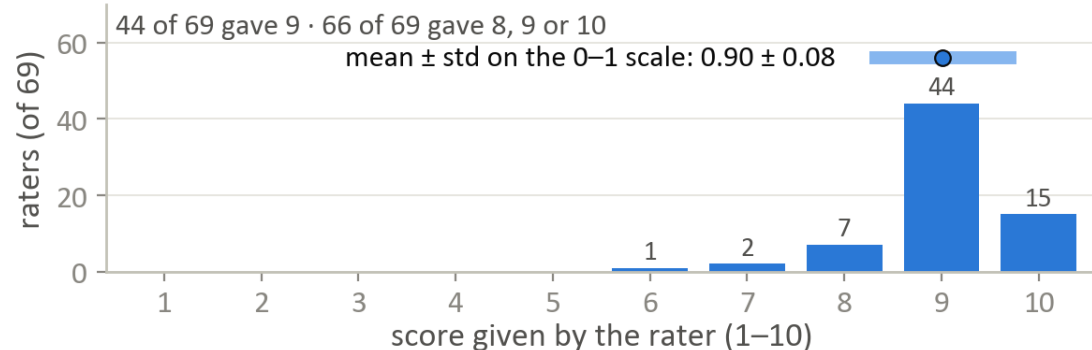
IPW1 Run 241 · rime · survey 0.90 ± 0.08



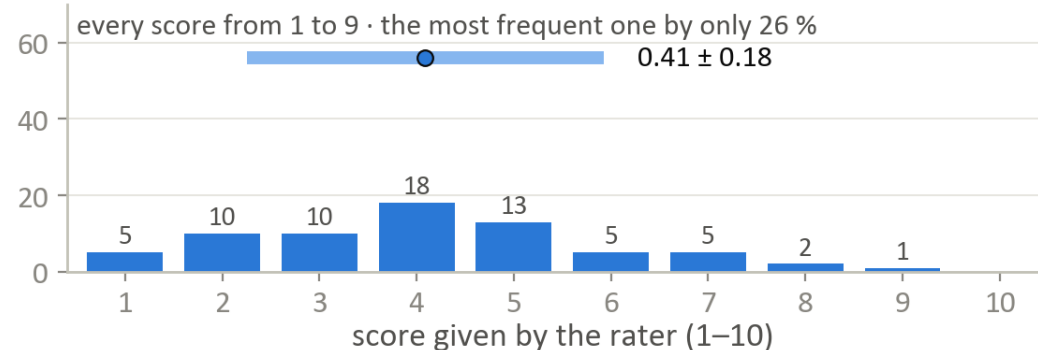
AIAA 2017 Run 1 · double horn · survey 0.41 ± 0.18



IPW1 Run 241 (rime): the community agrees



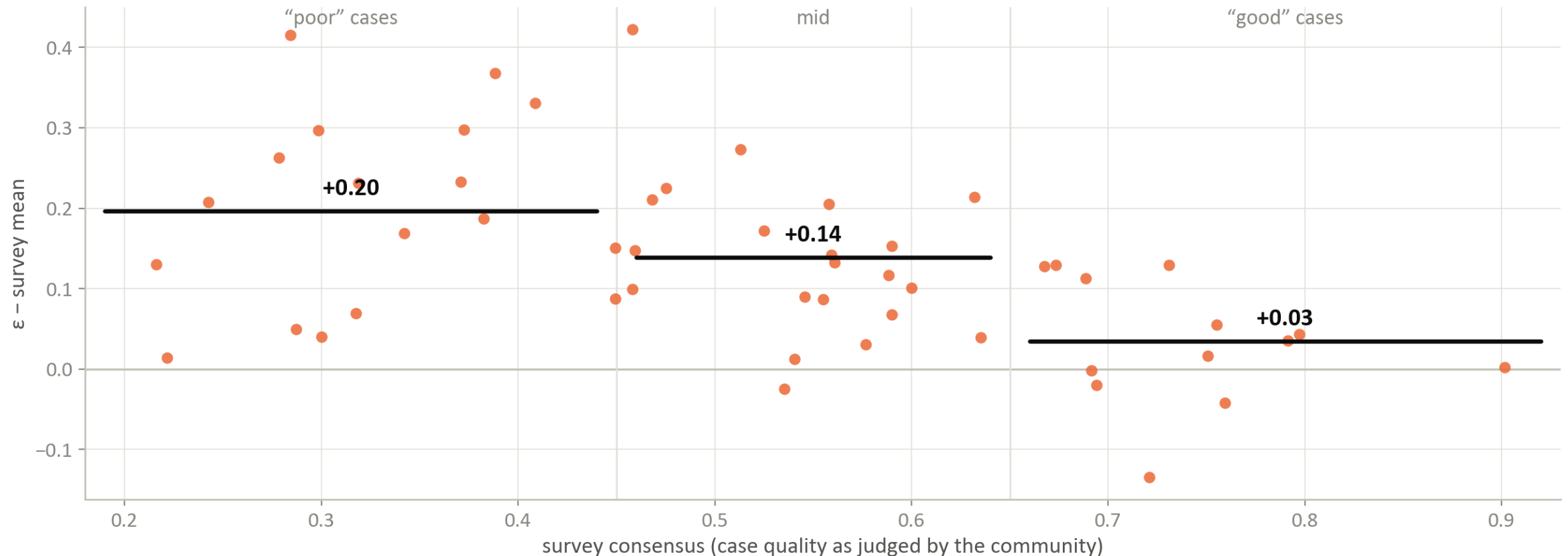
AIAA 2017 Run 1 (double horn): opinions split



ϵ performance

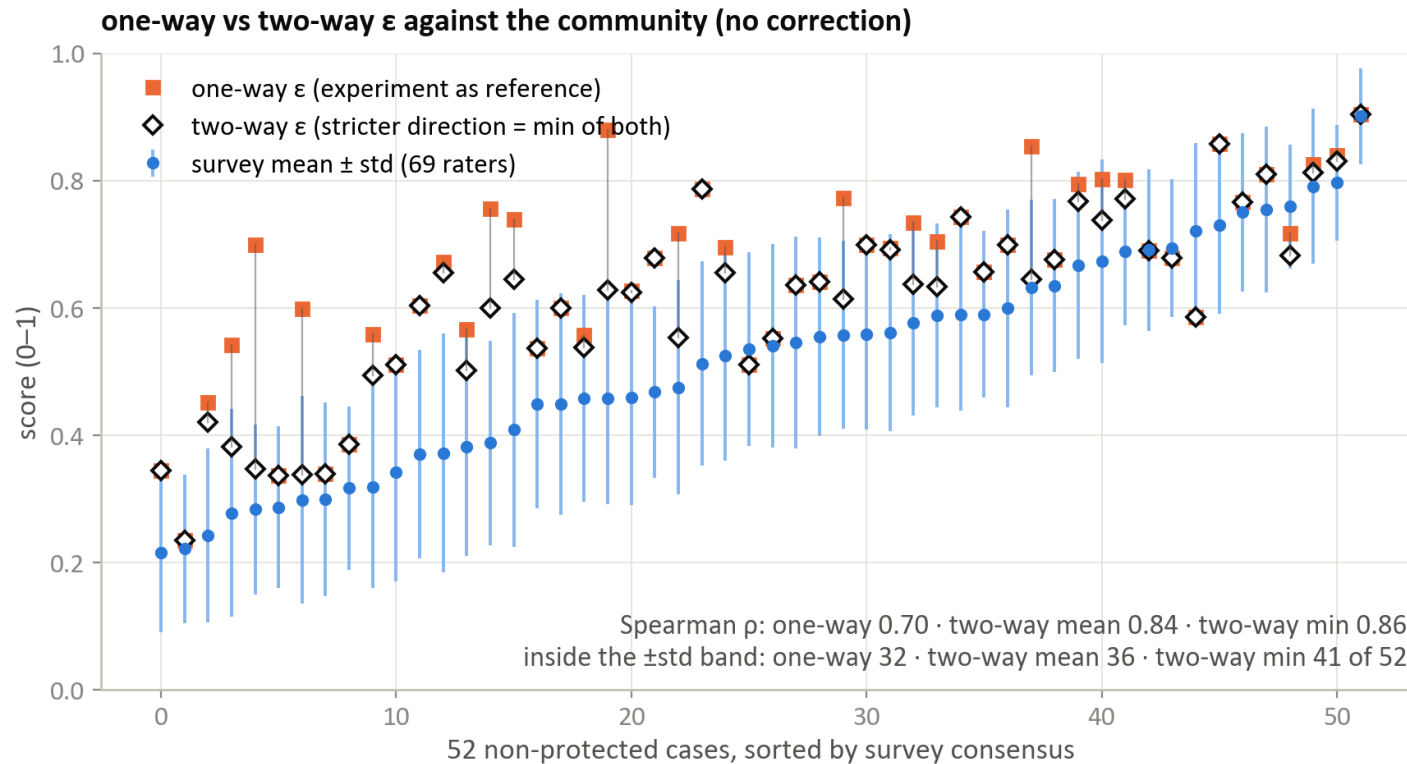
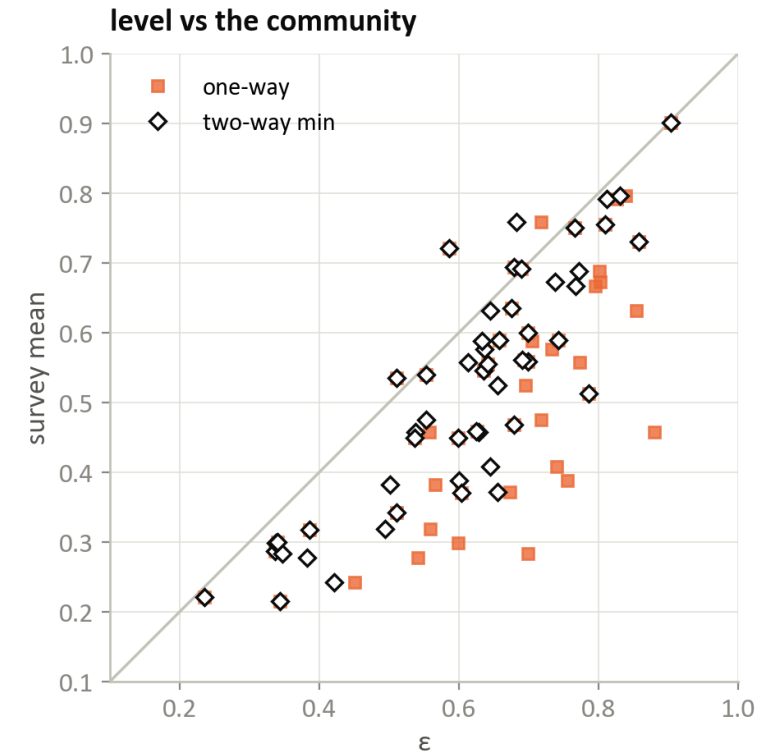
... ϵ is too forgiving for poor shapes

Where one-way ϵ deviates from the community — 52 non-protected cases (band averages in black)



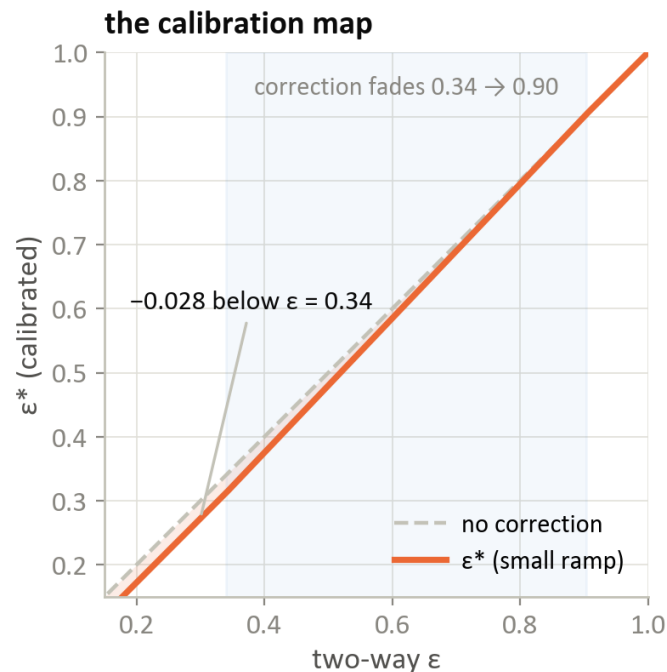
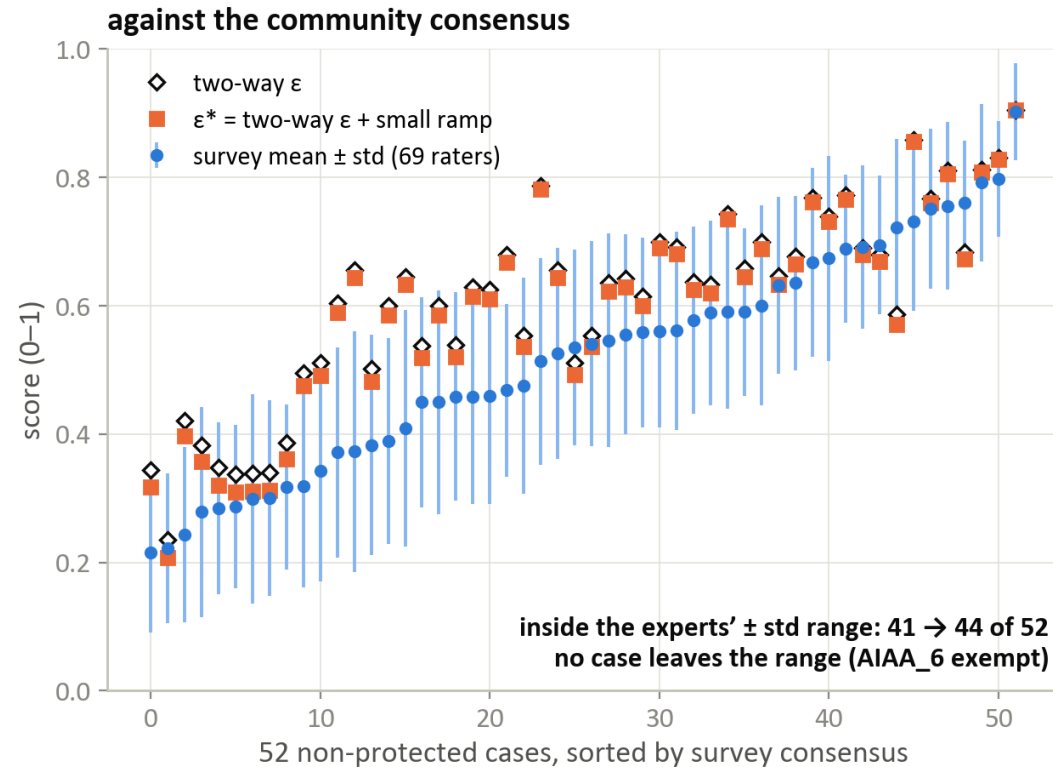
Two-way min ϵ alternative (ϵ^*)

Score each pair from both sides and keep the stricter: two-way min ϵ (ϵ^*)



ϵ^* + a small calibration

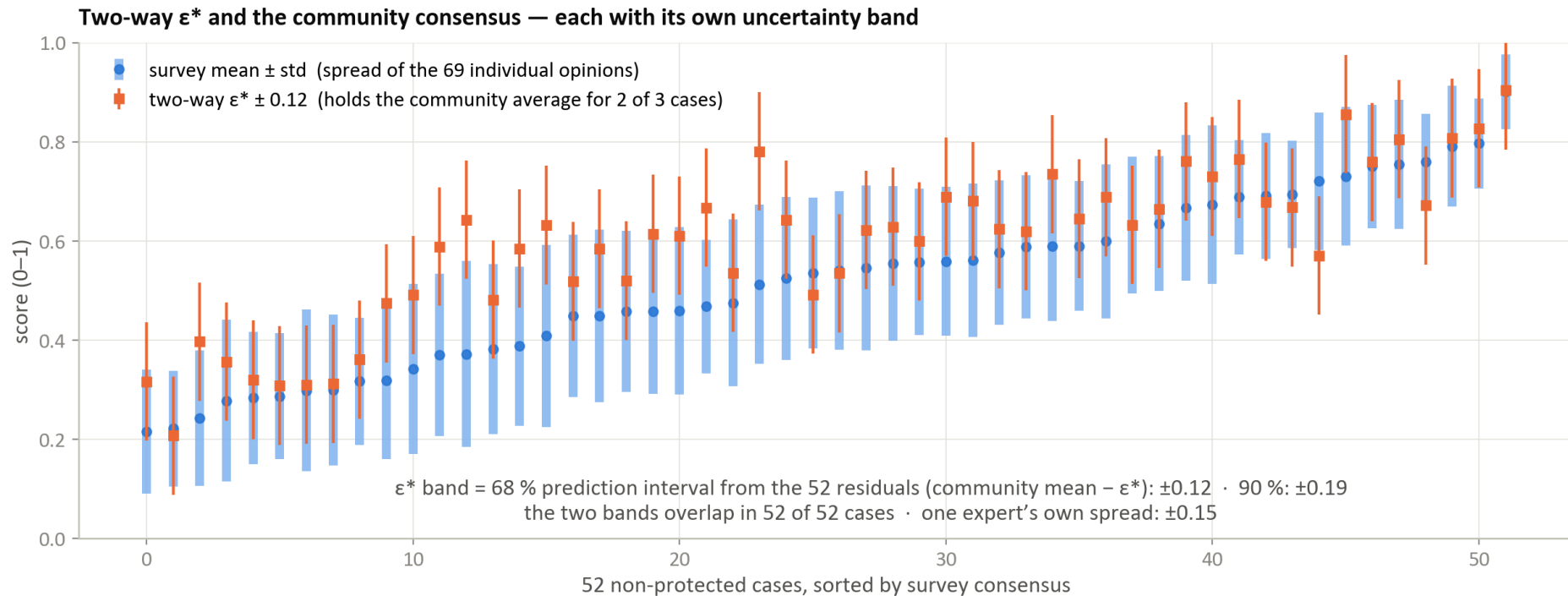
We can further improve with a small correction...



inside the experts' $\pm$ std range: one-way 33 $\rightarrow$ two-way 41 $\rightarrow$ + calibration 44 of 52 · poor-shape overshoot +0.20 $\rightarrow$ +0.13 $\rightarrow$ +0.11
concordance with the consensus (Lin's CCC — correlation that also requires the values to match): 0.54 $\rightarrow$ 0.75 $\rightarrow$ 0.78

ϵ^* : reliability

Two-way min ϵ is about as reliable as one expert: $\epsilon^* \pm 0.12$



ϵ^* band = 68 % prediction interval from the 52 residuals (community average - ϵ^*)
 ± 0.12 holds the community's average for 2 of 3 cases (± 0.19 for 9 of 10) — comparable to one expert's own spread (± 0.15)

Overview

No single number is perfect; but we can work with a range

- Proposal: score both directions and keep the stricter (two-way min ϵ), for single comparisons of non-protected ice shapes; report $\epsilon^* \pm 0.12$
- The small calibration adds only 3 cases: two-way min ϵ does most of the job to survey agreement
- The band is about one expert's own spread (± 0.15): expert judgment is not deterministic, so no single number can be
- Practical rule: don't over-interpret ϵ^* differences below ≈ 0.1
- Currently out of scope: protected cases; separate investigation needed
- The overall method is under review and open for discussion; feedback welcome

Takeaways

- The icing community shares one judgment of ice-shape agreement (Kendall's $W = 0.69$), and the 69-rater average is precise to ± 0.02
- One-way ϵ ranks shapes only roughly like the experts (Spearman $\rho = 0.71$) and is too forgiving where agreement is poor ($+0.20$): ice outside the reference footprint is not penalized
- Scoring both directions and keeping the stricter (two-way min ϵ) fixes most of it without fitting ($\rho = 0.86$); a small calibration adds 3 more cases to the survey band
- Individual expert judgement keeps a ± 0.15 spread, no single number is fully deterministic. A band $\epsilon^* \pm 0.12$ can be used to account for that

Thank you for your attention, and especially to the 69 survey respondents!