Warping Points in Arc-Length-Based Averages Have a Substantial Effect on Computed Biomechanical Response Corridors and Objective Ratings Methods

Devon C. Hartlen, Duane S. Cronin

I. INTRODUCTION

The adoption of numerical models to assess vehicle and occupant safety, especially by regulatory agencies [1], has led to a clear and present need to assess model performance against experimental data for model qualification. Euro NCAP CP550 has proposed HBM qualification procedures that assess HBM performance against a series of experimental test cases. Euro NCAP CP550 uses ARCGen [2] to calculate an average curve and statistical response corridors from the experimental test data used for model assessment, ranging from time-series kinematic data to hysteretic force-displacement responses. The application of ARCGen requires the number of warping points as a user-definable signal registration parameter to compute accurate and meaningful averages and corridors. Warping points are used to align shared response features across experimental responses with respect to arc-length, with more warping points providing greater alignment flexibility, especially for noisy or oscillatory experimental data. However, too many warping points introduces the risk of overfitting. In this paper, we quantify the effect of the number of warping points on the resulting average and corridors using exemplar oscillatory head acceleration data to demonstrate the impact of the number of warping points on three objective rating methods, and provide guidance for determining the number of warping points for a specific case.

II. METHODS

As an exemplar of oscillatory data, ARCGen was applied to the NBDL volunteer sled test dataset [3], specifically the 10 g frontal sled pulse and x-axis head acceleration-time responses (n=18) (Fig. 1(A)). Averages and response corridors were calculated with a varying number of warping points from 0 (no signal registration) to 8. All averages were computed using 500 resampling points. The sum of squared errors (SSE) was computed incrementally between averages as the number of warping points increased to identify the number of warping points at which the average curve converged. The computed average response with the lowest incremental SSE was considered the converged average curve. To quantify the effect of the number of warping points on metric outcomes, each of the nine computed averages and corridors, corresponding to the range of 0 to 8 warping points, was rated against the converged curve (lowest SSE) using three objective rating methods: CORA [4]; ISO 18571 [5]; and Euro NCAP CP550 [1].

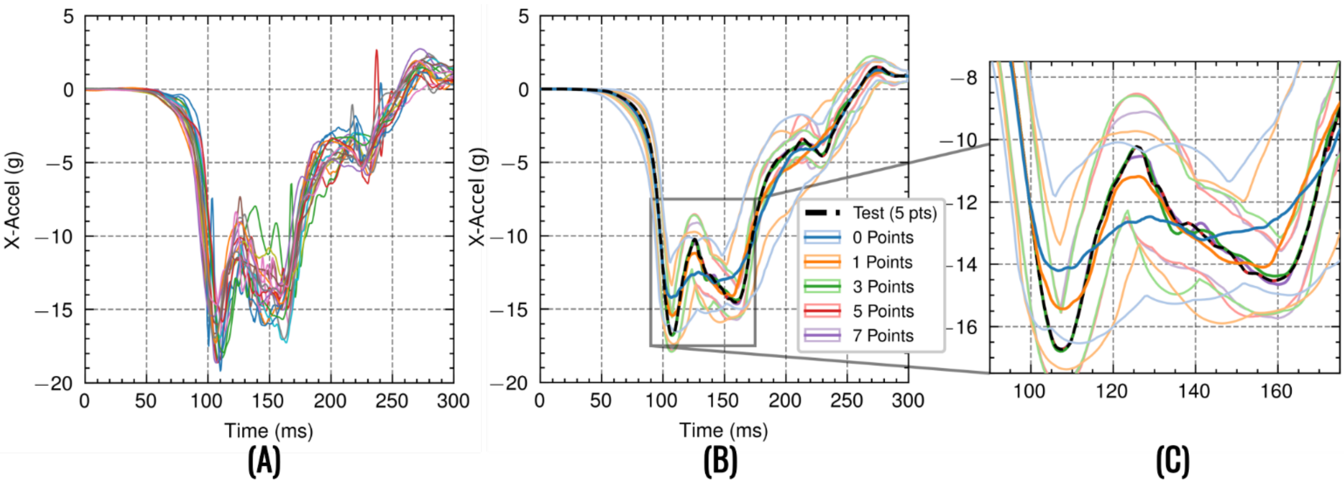


Fig. 1. (A) 18 x-axis head acceleration-time responses from the 10 g NBDL dataset. (B) A subset of averages and corridors for 0, 1, 3, 5, and 7 warping points. The inset plot (C) zooms in on the three local extrema between 100 ms and 175 ms. Black dashed line is the converged average calculated with 5 warping points.

D. C. Hartlen (e-mail: Devon.Hartlen@UWaterloo.ca) is a Post-Doctoral Research Associate and D. S. Cronin is a Professor and the Canada Research Chair in Trauma Biomechanics and Injury Prevention at the University of Waterloo, Canada.

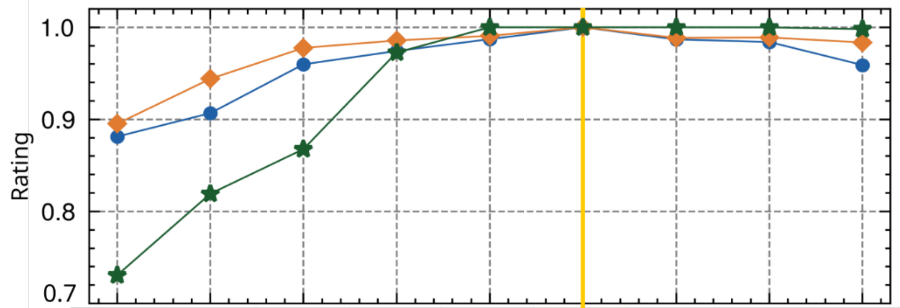
III. INITIAL FINDINGS

Increasing the number of warping points substantially impacted the average and response corridors calculated for the data (Fig. 1(B), (C)), especially between 0 and 2 warping points (Fig. 1(B)). These averages did not capture the three local extrema between 100 ms and 175 ms (valley-peak-valley, Fig. 1(C)) and had large, unrepresentative corridors that did not reflect experimental spread (Fig. 1(A)). Increased warping points led to improved representation of experimental features later in time, such as the reversal at 225 ms and the final local peak at 275 ms. For this dataset, SSE was minimised at 5 warping points (Fig. 2(A)), with the computed average (converged curve) used as the test signal in all rating calculations.

The result of comparing the curves for each number of warping points to the converged value increased up to the expected rating of 1.0 at 5 warping points (Fig. 2(B)). The largest changes occurred between 0 and 2 warping points, with a slight decrease in rating beyond 5 warping points.

Warping Increment	SSE
0→1	136.51
1→2	128.74
2→3	14.41
3→4	8.95
4→5	4.28
5→6	5.53
6→7	5.22
7→8	14.57

(A)



	0	1	2	3	4	5	6	7	8
CORA	0.8813	0.9067	0.9598	0.9742	0.9871	1.000	0.9868	0.9840	0.9590
ISO	0.8952	0.9440	0.9776	0.9857	0.9907	1.000	0.9885	0.9889	0.9835
EuroNCAP	0.7307	0.8192	0.9726	0.9726	1.000	1.000	1.000	0.9999	0.9980

Warping Points

(B)

Fig. 2. (A) Incremental sum of squared errors (SSE) for average responses computed with 0–8 warping points.(B) Objective rating methods for averages and corridors computed with 0–8 warping points compared to the converged value at 5 warping points.

IV. DISCUSSION

This study demonstrates that the number of warping points can substantially affect the fidelity of computed averages and corridors to experimental data and needs to be considered in any ARCGen calculation. With a small number of warping points (2 or fewer), signal registration could not align local extrema with respect to arc-length, producing poorly trending averages and artificially wide corridors. Increasing the number of warping points improved the alignment of later-occurring features and extrema that characterise this dataset (Fig. 2(B)). Five warping points were found to minimise incremental SSE for the example case considered, which correlated well with the rule of thumb proposed in [2]: that the number of warping points should match the number of key shared features (such as shared local extrema, inflections, plateaus, or reversals) in the experimental responses. Beyond 5 warping points, incremental SSE decreased because signal registration tended to over-register the average, which manifests as points clustering near key features rather than the visual fit of the average degrading, thereby impacting CORA and ISO scores more than Euro NCAP.

Appropriate identification of the number of warping points was important, as the choice has downstream effects on model qualification, with the greatest impact when few or no warping points were used. However, for this dataset, there was a range of warping points between 4 and 6 that produced similar averages, corridors and ratings, suggesting some flexibility in the selection of warping points. Nevertheless, this study demonstrates that without an appropriate, dataset-specific selection of warping points, users can negatively impact the evaluation of biofidelity when comparing computational HBMs to experimental data. Future work will present an automated approach to guide the identification of an appropriate number of warping points for the ARCGen method.

V. REFERENCES

[1] Ressi, F., *et al.*, *IRCOBI*, 2025. [2] Hartlen, D. C. & Cronin, D. S., *Front Bioeng Biotechnol*, 2022. [3] Ewing, C. & Thomas, D., USAARL 73-1, 1972. [4] Gehre, C., *et al.*, 21st ESV Conf., 2009. [5] ISO/TS 18571:2024.