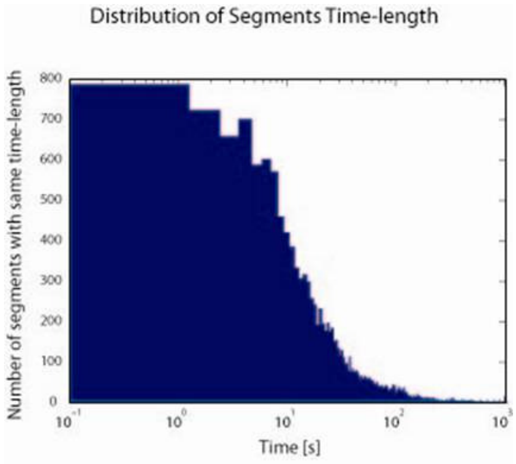
Chunking: a method to increase robustness of field-operational-test data analysis

Marco Dozza & Jonas Bärgman - Chalmers University of Technology.



When using field operational test (FOT) data to evaluate advanced driver assistance systems (ADASs), selecting relevant data from the FOT database is a crucial step which significantly influences any following step in data analysis. In the methodology project SeMiFOT [https://www.chalmers.se/safer/EN/projects/traffic-safety-analysis/semifot], a collaboration between SAFER-partners [http://www.chalmers.se/safer/SV/] and University of Michigan Transportation Research Institute, data extraction queries were defined to select data relevant for safety analysis of several ADASs such as lane departure warning. Independently of how these queries are designed, they select segments of continuous-time in the FOT database based on some condition (e.g. baseline and treatment for velocities over 70km/h). As the top-right Figure shows, these segments may have a very different length and short segments are likely to be more represented than long segments. Once segments are extracted, one or more parameters such as mean speed or standard deviation of lane position (SDLP) are typically calculated for each individual segment. The parameters for all segments are combined for each condition into an overall value such as mean speed or median of SDLP. These combined parameters are then used for hypothesis testing. The Table shows that different ways to compute the above mentioned parameters result in significantly different analysis outcomes.

Results from parameter calculation on a simple query that extracts segments of data were vehicle velocity is greater than 70km/h, are reported in the Table. These results suggest that a method to split data segments into equally long sub-segments – which we call *chunking* – may provide a more robust estimation of the overall parameters than a simple systematic calculation of the same parameters over each segment. This *chunking* method may be applied when parameters from data segments of different time- (or distance-) length need to be combined. In specific cases, such as when arithmetic mean is used for calculation of the segments parameter values, a time- or distance weighted mean may be used instead of *chunking* for performing the segment combination. In addition, a time filter could be implemented without using *chunking* to avoid calculating parameters on too short segments. However, neither of these approaches is equivalent to *chunking* and will in most cases have significantly lower statistical power (number of data segments in Table). In the SDLP example, this parameter is not robust for segments shorter than 60-s making its values in the forth column much more reliable than the ones in the third column. Further, the statistical power is more than 4 times greater using chunking than filtering at 60s. One drawback of *chunking* is that when segment duration does not add up to an integer multiple of the *chunk* duration, the residuals are discarded. When considering applying *chunking* and choosing *chunk*-length, at least parameter robustness, statistical power, vehicle dynamics, computation time and data loss should be taken into account. *Chunking* may help new users of FOT data in calculating robust parameters and to increase statistical power.

	NO Chunking	5-s Chunk	20-s Chunk	60-s Chunk	180-s Chunk	600-s Chunk	No Chunking & weighted mean	No Chunking & filter above 60s
Average segment lenght [s]	91.3	5	20	60	180	600	91.3	289.2
Number of data segments	15729	279518	65074	18846	4681	745	15729	4328
Used Data [h]	399	388	362	314	234	124	399	348
Mean Speed [km/h]	77.9	93.8	95.0	96.8	99.6	102.9	93.2	88.6
Median SDLP [m]	0.271	0.102	0.211	0.299	0.35	0.353	0.332	0.387

Authors Contacts:

Marco Dozza – Chalmers University of Technology, dept. of Applied Mechanics.

Tel: +46 31 772 3621, e-mail: marco.dozza@chalmers.se

SAFER - Box 8077 - S-402 78 - Göteborg – Sweden

Jonas Bärgman – Chalmers University of Technology, dept. of Applied Mechanics.

Tel: +46 31 772 5846, e-mail: jonas.bargman@chalmers.se

SAFER - Box 8077 - S-402 78 - Göteborg - Sweden