

# Notes on data sets

theodor.westny@liu.se

June 2021

## 1 Introduction

For model training and evaluation, publicly available data sets provided by the Federal Highway Administration (FHWA) are used. The data sets contain recorded highway traffic data captured at 10 Hz from US-101 [1] and Interstate 80 (I80) [2]. Both data sets provide 45 min worth of recordings divided into three 15-minute subsets simulating mildly, moderately, and heavily congested traffic. The NGSIM highway data sets are consistently used in many studies, but have been reported to suffer from considerable tracking noise [3] and even physically impossible trajectories (collisions) [4].

## 2 Method

For the purpose of our paper, we have performed an extensive investigation of the data in search for vehicles that would break the feature extraction script. Since we rely on the given lane id's for extraction, tracking noise that results in vehicles jumping from lane to lane in fraction of seconds would yield erroneous trajectories for training. The procedure we employed was therefore to detect these instances and asses their state. If deemed repairable, we pass the trajectories through simple smoothing filters and re-label the current lane. Furthermore, some instances where the same lane is visited twice (in a reasonable manner, e.g. overtaking), are split into two for easier use. However, some instances ( $\leq 5$ ) deemed non-repairable are removed from the data sets. A list of all the vehicle id's concerned for the individual data sets can be found in the appendix.

The final data sets are contained within the github

repository<sup>1</sup>, and labeled FIXED. They are made freely available under creative commons license<sup>2</sup>, as per original license of the NGSIM data sets. The use of the FIXED data sets are optional but encouraged to replicate results. The use of the original data sets are supported, but the extraction script will automatically detect and remove instances that fall under the category described above — resulting in a loss of total maneuvers.

## References

- [1] J. Colyar and J. Halkias, “US Highway 101 Dataset. Federal Highway Administration Research and Technology fact sheet. Publication number: FHWA-HRT-07-030,” *Tech. Rep.*, 2007, <http://doi.org/10.21949/1504477>.
- [2] J. Halkias and J. Colyar, “Interstate 80 Freeway Dataset. Federal Highway Administration Research and Technology fact sheet. Publication number: FHWA-HRT-06-137,” *Tech. Rep.*, 2006, <http://doi.org/10.21949/1504477>.
- [3] M. Montanino and V. Punzo, “Making NGSIM data usable for studies on traffic flow theory: Multistep method for vehicle trajectory reconstruction,” *Transportation Research Record*, vol. 2390, no. 1, pp. 99–111, 2013.
- [4] B. Coifman and L. Li, “A critical evaluation of the Next Generation Simulation (NGSIM) vehicle trajectory dataset,” *Transportation Research Part B: Methodological*, vol. 105, pp. 362–377, 2017.

<sup>1</sup><https://github.com/westny/imb-behavior-prediction>

<sup>2</sup><https://creativecommons.org/licenses/by-sa/4.0/>

# Appendices

## A US-101 0750am-0805am

Table 1: US-101 0750am-0805am

Vehicle ID	Process	New ID
6	Smoothened	N/A
37	Split	3000
49	Split	3001
104	Smoothened	N/A
114	Smoothened	N/A
115	Split	3002
170	Smoothened	N/A
190	Smoothened	N/A
233	Smoothened	N/A
237	Smoothened	N/A
250	Smoothened	N/A
305	Smoothened	N/A
309	Smoothened	N/A
350	Smoothened	N/A
375	Smoothened	N/A
384	Smoothened	N/A
387	Smoothened	N/A
435	Smoothened	N/A
457	Smoothened	N/A
472	Smoothened	N/A
513	Smoothened	N/A
522	Smoothened	N/A
525	Smoothened	N/A
582	Smoothened	N/A
586	Smoothened	N/A
590	Smoothened	N/A
597	Smoothened	N/A
632	Smoothened	N/A
668	Smoothened	N/A
698	Smoothened	N/A
704	Removed	N/A
719	Smoothened	N/A
726	Smoothened	N/A
735	Removed	N/A

Vehicle ID	Process	New ID
752	Smoothened	N/A
774	Removed	N/A
784	Smoothened	N/A
808	Smoothened	N/A
816	Smoothened	N/A
817	Smoothened	N/A
890	Smoothened	N/A
925	Split	3003
932	Smoothened	N/A
949	Smoothened	N/A
983	Smoothened	N/A
986	Smoothened	N/A
996	Smoothened	N/A
1000	Smoothened	N/A
1014	Smoothened	N/A
1043	Smoothened	N/A
1045	Smoothened	N/A
1054	Smoothened	N/A
1068	Smoothened	N/A
1075	Smoothened	N/A
1079	Smoothened	N/A
1098	Smoothened	N/A
1106	Smoothened	N/A
1110	Smoothened	N/A
1132	Smoothened	N/A
1154	Smoothened	N/A
1165	Smoothened	N/A
1198	Smoothened	N/A
1235	Smoothened	N/A
1244	Smoothened	N/A
1255	Removed	N/A
1269	Smoothened	N/A
1302	Smoothened	N/A
1321	Split	3004
1334	Smoothened	N/A
1407	Smoothened	N/A

Vehicle ID	Process	New ID
1432	Smoothened	N/A
1446	Smoothened	N/A
1461	Smoothened	N/A
1493	Smoothened	N/A
1556	Smoothened	N/A
1564	Split	3005
1565	Smoothened	N/A
1571	Smoothened	N/A
1618	Split	3006
1647	Smoothened	N/A
1713	Smoothened	N/A
1769	Split	3007
1779	Smoothened	N/A
1791	Smoothened	N/A
1944	Smoothened	N/A
1976	Split	3008
2015	Smoothened	N/A
2068	Smoothened	N/A
2098	Smoothened	N/A
2119	Smoothened	N/A
2146	Smoothened	N/A
2150	Smoothened	N/A
2204	Split	3009
2276	Smoothened	N/A
2349	Smoothened	N/A
2402	Smoothened	N/A
2427	Smoothened	N/A
2473	Smoothened	N/A
2586	Smoothened	N/A
2613	Smoothened	N/A
2629	Smoothened	N/A
2631	Smoothened	N/A
2680	Smoothened	N/A
2724	Smoothened	N/A
2748	Smoothened	N/A
2752	Smoothened	N/A
2753	Smoothened	N/A

## B US-101 0805am-0820am

Table 2: US-101 0805am-0820am		
Vehicle ID	Process	New ID
11	Smoothened	N/A
20	Smoothened	N/A
31	Smoothened	N/A
36	Smoothened	N/A
63	Smoothened	N/A
80	Smoothened	N/A
99	Smoothened	N/A
100	Smoothened	N/A
146	Smoothened	N/A
181	Smoothened	N/A
204	Smoothened	N/A
244	Smoothened	N/A
275	Smoothened	N/A
344	Split	3001
420	Split	3000
540	Smoothened	N/A
579	Removed	N/A
644	Smoothened	N/A
753	Smoothened	N/A
758	Split	3002
889	Split	3005
1176	Smoothened	N/A
1213	Smoothened	N/A
1230	Smoothened	N/A
1360	Smoothened	N/A
1453	Split	3003
1486	Smoothened	N/A
1562	Smoothened	N/A
1686	Smoothened	N/A
1818	Smoothened	N/A
1835	Split	3004
1906	Smoothened	N/A

## C US-101 0820am-0835am

Table 3: US-101 0820am-0835am

Vehicle ID	Process	New ID
60	Smoothened	N/A
146	Smoothened	N/A
366	Smoothened	N/A
381	Smoothened	N/A
445	Smoothened	N/A
585	Split	3000
721	Smoothened	N/A
823	Removed beginning then split	3005
848	Split	3003, 3004
849	Split	3001
1107	Smoothened	N/A
1154	Smoothened	N/A
1162	Split	3002
1313	Smoothened	N/A

## D I80 0400pm-0415pm

Table 4: I80 0400pm-0415pm

Vehicle ID	Process	New ID
5	Smoothened	N/A
108	Smoothened	N/A
186	Smoothened	N/A
210	Smoothened	N/A
268	Smoothened	N/A
365	Smoothened	N/A
376	Smoothened	N/A
399	Smoothened	N/A
409	Smoothened	N/A
447	Smoothened	N/A
458	Smoothened	N/A
518	Smoothened	N/A
545	Split	3000
743	Smoothened	N/A
843	Split	3001
844	Smoothened	N/A
860	Split	3002
1097	Smoothened	N/A
1178	Split	3003
1215	Smoothened	N/A
1229	Smoothened	N/A
1235	Smoothened	N/A
1349	Smoothened	N/A
1383	Smoothened	N/A
1408	Smoothened	N/A
1432	Smoothened	N/A
1455	Smoothened	N/A
1628	Smoothened	N/A
1713	Smoothened	N/A
1724	Smoothened	N/A
1729	Smoothened	N/A
1762	Smoothened	N/A
1778	Smoothened	N/A
1791	Smoothened	N/A
1828	Smoothened	N/A
1854	Smoothened	N/A

Vehicle ID	Process	New ID
1909	Smoothened	N/A
1912	Smoothened	N/A
1979	Smoothened	N/A
1989	Smoothened	N/A
2087	Smoothened	N/A
2147	Smoothened	N/A
2182	Split	3003
2196	Smoothened	N/A
2234	Smoothened	N/A
2313	Smoothened	N/A
2326	Smoothened	N/A
2379	Smoothened	N/A
2404	Smoothened	N/A
2437	Smoothened	N/A
2761	Smoothened	N/A
2767	Smoothened	N/A
2783	Smoothened	N/A
2785	Smoothened	N/A
2812	Smoothened	N/A

## E I80 0500pm-0515pm

Table 5: I80 0500pm-0515pm

Vehicle ID	Process	New ID
342	Smoothened	N/A
491	Smoothened	N/A
515	Smoothened	N/A
522	Smoothened	N/A
654	Smoothened	N/A
658	Smoothened	N/A
662	Smoothened	N/A
706	Smoothened	N/A
750	Smoothened	N/A
819	Smoothened	N/A
876	Smoothened	N/A
909	Smoothened	N/A
990	Smoothened	N/A
1004	Split	3000
1009	Smoothened	N/A
1018	Smoothened	N/A
1087	Smoothened	N/A
1118	Smoothened	N/A
1164	Smoothened	N/A
1211	Smoothened	N/A
1215	Smoothened	N/A
1225	Smoothened	N/A
1226	Smoothened	N/A
1281	Smoothened	N/A
1288	Smoothened	N/A
1319	Smoothened	N/A
1376	Smoothened	N/A
1382	Smoothened	N/A
1392	Smoothened	N/A
1407	Smoothened	N/A

Vehicle ID	Process	New ID
1424	Smoothened	N/A
1441	Smoothened	N/A
1462	Smoothened	N/A
1464	Smoothened	N/A
1468	Smoothened	N/A
1482	Smoothened	N/A
1506	Smoothened	N/A
1530	Smoothened	N/A
1535	Smoothened	N/A
1546	Smoothened	N/A
1574	Smoothened	N/A
1594	Smoothened	N/A
1619	Smoothened	N/A
1647	Smoothened	N/A
1680	Split	3001
1688	Smoothened	N/A
1754	Smoothened	N/A
1801	Smoothened	N/A
1832	Smoothened	N/A
1912	Smoothened	N/A

## F I80 0515pm-0530pm

Table 6: I80 0515pm-0530pm

Vehicle ID	Process	New ID
26	Smoothened	N/A
38	Smoothened	N/A
148	Smoothened	N/A
195	Smoothened	N/A
302	Smoothened	N/A
383	Smoothened	N/A
401	Smoothened	N/A
410	Smoothened	N/A
497	Smoothened	N/A
503	Smoothened	N/A
512	Smoothened	N/A
550	Smoothened	N/A
611	Smoothened	N/A
629	Split	3000
686	Smoothened	N/A
726	Smoothened	N/A
845	Smoothened	N/A
886	Smoothened	N/A
950	Smoothened	N/A
955	Smoothened	N/A
975	Smoothened	N/A
1154	Smoothened	N/A
1163	Smoothened	N/A
1254	Smoothened	N/A
1302	Smoothened	N/A
1466	Split	3001
1512	Smoothened	N/A
1557	Split	3002
1584	Smoothened	N/A
1620	Smoothened	N/A
1714	Smoothened	N/A