

# プローブ旅行時間データの データクレンジング手法の適用に関する研究

立川 太一<sup>1</sup>・橋本 浩良<sup>1</sup>・瀬戸下 伸介<sup>1</sup>・松島 敏和<sup>2</sup>

<sup>1</sup>正会員 国土交通省 国土技術政策総合研究所 道路研究室 (〒305-0804 茨城県つくば市旭1番地)

E-mail: tachikawa-t924a@nilim.go.jp, hashimoto-h22ab@nilim.go.jp, setoshita-s2n9@nilim.go.jp

<sup>2</sup>正会員 中央復建コンサルタンツ株式会社 交通計画グループ (〒533-0033 大阪市東淀川区東中島4-11-10)

E-mail: matsushima\_t@cfk.co.jp

近年、自動車の移動履歴情報から得られるプローブ旅行時間データが渋滞状況の分析などに活用されている。プローブ旅行時間データには路側での駐停車等の特異値が含まれており、分析に適さない特異値のクレンジングが必要である。本研究では松島らが提案したデータクレンジング手法について、特異値が多い区間の特性を把握し、個車の走行データの挙動からその有効性を確認した。その結果、特異値は速度変化が比較的大きい時間帯に生じやすいことを確認した。一方で、同一時間帯に交通状況が大きく変化する区間、混雑時でも車両ごとの旅行速度差が大きい区間においては特異値を除去できない可能性があることが分かった。

**Key Words :** Probe data, Travel time, Data cleansing, ETC2.0probe data

## 1. はじめに

近年のICT技術の進展により、カーナビ等から取得される自動車の移動履歴情報（以下「プローブデータ」という。）が大量に収集されるようになってきている。また、プローブデータをデジタル道路地図（以下「DRM」という。）などのリンク別の旅行時間に加工したデータ（以下「プローブ旅行時間データ」という。）が渋滞状況の分析など道路交通状況の把握に広く活用されている。

プローブ旅行時間データを構成する個車の走行データには、個人の運転特性や路側での駐停車などによる特異な走行データが存在する。道路の一般的な交通状況を把握することを考えると、分析に適さない特異なデータの除去（以下「データクレンジング」という。）を行うことが必要である。

プローブ旅行時間データのデータクレンジングについては、現状、分析者に委ねられている。統一かつ汎用的なデータクレンジング手法を確立することが、分析の効率化や成果の質の向上に資する。松島ら<sup>1</sup>は、汎用的なデータクレンジング手法として、同一時間帯の個別車両間の所要時間差に着目したデータクレンジング手法（以下「本手法」という。）を提案している。

松島らは、表-1に示すように、路側での駐停車などの特異な走行と想定されるデータの除去結果について、DRMリンク延長別・旅行速度別に分類し、除去対象件数の割合を確認している。しかし、実際に特異値として除去対象となるデータ及び除去対象となるデータの同一時間帯の他のデータの挙動についての検証はなされていない。

本研究では、本手法の有効性の検証をねらいとして、除去対象となるデータの挙動、特異値データの同一時間帯の他のデータの挙動を確認し、路側での駐停車と想定されるデータが除去されているか、また駐停車をせずに走行しているデータが除去されていないか等を検証するものである。

本稿の構成は以下の通りである。第2章ではデータクレンジングに関する既往研究について整理し、本研究の位置づけを整理する。第3章では東京都全域を対象として本手法による特異値の発生状況を区間別に確認する。第4章では個別の走行データにおける挙動について検証する。第5章で成果と今後の課題を整理し、本稿をまとめる。

表-1 松島らによるデータクレンジングの除去対象件数の割合

速度ランク\DRMリンク延長	500m未満	500m以上 1000m未満	1000m以上 2000m未満	2000m以上 3000m未満	3000m以上	合計
1km/h未満	100.00%	100.00%	100.00%	-	-	100.00%
1km/h以上10km/h未満	0.14%	5.36%	15.15%	43.88%	55.15%	0.40%
10km/h以上20km/h未満	0.00%	0.00%	0.00%	3.77%	23.77%	0.01%
20km/h以上40km/h未満	0.00%	0.00%	0.00%	0.00%	0.39%	0.00%
40km/h以上60km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
60km/h以上80km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
80km/h以上100km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
100km/h以上120km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
120km/h以上150km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
150km/h以上	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
合計	0.34%	0.19%	0.16%	0.18%	0.17%	0.30%

凡例：割合  大 小  
 ※除去対象件数の割合(%) = 除去対象データ対象件数 / データクレンジング前のデータ件数

## 2. データクレンジングに関する既往研究と本研究の位置づけ

### (1) プローブ旅行時間データの特異値

プローブデータには、時刻と位置（緯度・経度）、前後・左右加速度といった挙動履歴にかかわるデータが存在する。本研究で着目しているプローブ旅行時間データとは、1つのリンクを走行する際の所要時間である。

プローブ旅行時間データを構成する個車の走行データは、通常通り走行しているデータに加え、走行中に駐停車等を含むデータも存在する。走行中の駐停車を含むデータは駐停車時間も所要時間として計上されるため、特異な所要時間となる可能性がある。道路の一般的な交通状況を把握することを考えると、駐停車を含むような特異値データを除去することが必要となる。

### (2) データクレンジング手法に関する既往研究

既往研究においては、ナンバープレート調査による地点間の旅行時間から平均旅行速度を算出する際に、同一時間帯のZスコアを用いる方法<sup>2)</sup>や標準偏差を用いる手法<sup>3)</sup>による特異値除去の検証がなされている。

これらのデータクレンジングは、特異値データを旅行時間の外れ値と捉え、統計的な閾値について感度分析を実施したものである。

ナンバープレート調査のように、同一区間、同一時間帯の母集団の分布が明確である場合は、Zスコアを用いる方法や標準偏差を用いる手法によるデータクレンジングが可能である。

### (3) 本手の特徴

ETC20プローブデータ等のデータはサンプルデータであり、現状サンプル数は多くない。そのためZスコアを用いる方法や標準偏差を用いる手法によるデータクレン

ジング手法は適用しづらい。

そこで、松島らは本手法を提案している。

本手法は、同一時間帯に取得されたプローブ旅行時間の相互比較により、特異な所要時間データを除去するというものである。具体的には、①閾値による高速度・低速度データの除去、②同一時刻帯の道路交通状況を考慮した特異な所要時間データの除去の2段階のデータクレンジングを行うものである（表-2）。

②が本手法の特徴である。図-1の例のような駐停車のデータは道路交通状況分析には適さないため、除去対象とすべきものであるものの、個車の所要時間のみでは混雑時の走行のデータと区別ができない。そこで、駐停車車両はほかの車両との所要時間差が大きく、混雑による低速度車はほかの車両との所要時間差が小さいことに着目し、当該リンクの同一時間帯の最少所要時間（＝最高速度）のサンプルから一定時間以上遅れる車両を特異な走行とみなし、該当するデータを除去している。松島らは、②のパラメータである同一時間帯最少所要時間からの差の閾値（以下「時間閾値」という。）は、分析目的に応じて可変であるものの、汎用性を考慮したデフォルト値としては600秒程度が妥当であるとしている。

リンク延長が長くなると最少所要時間からの時間差が大きくなるため、最少所要時間が非常に小さい場合、通常の走行データであってもデータクレンジング対象となる可能性がある。この影響を抑制するため、高速道路では80km/h、一般道では30km/hの走行時の所要時間を最少所要時間の下限とする。

### (5) 本手法の検証結果

松島らは、自身が提案するデータクレンジング手法について、2015年4月～6月、全国、全道路種別、昼間12時間（の速度ランク別DRMリンク延長別旅行速度データの距離帯別・速度帯別のデータ件数を集計することで除

去対象件数の割合の検証を行っている。表-1に示す通り、1km/h未満の全てのデータと、距離に応じて1km/h以上のデータであっても除去できているというのがその内容である。

(6) 本手法の課題及び本研究の位置づけ

松島らの研究においては、実際に特異値として除去対象となるデータの位置情報や、除去対象となるデータの同一時間帯の他のデータの挙動といった、詳細な検証はできていない。また、本手法によるデータクレンジングが機能しづらい区間等が存在する可能性もあると考える。

本研究の意義は、本手法が600秒以上の駐停車またはそれに準ずる走行データを特異値として除去するといった目的を達成できているかを検証するとともに、改善に向けた手法を検討することにある。

表-2 データクレンジング手法の概要

項目	内容
① 閾値によるデータの除去	・ 低速度側 1km/h 未満, 高速度側 150km/h 以上となるデータを除去
② 同一時間帯の道路交通状況を考慮したデータの除去	・ 同一時間帯における最少所要時間 (最高速度) のサンプルから一定時間以上遅れるサンプルのデータを除去 ・ 但し, 高速道路では 80km/h, 一般道では 30km/h の走行時の所要時間を最少所要時間の下限とする

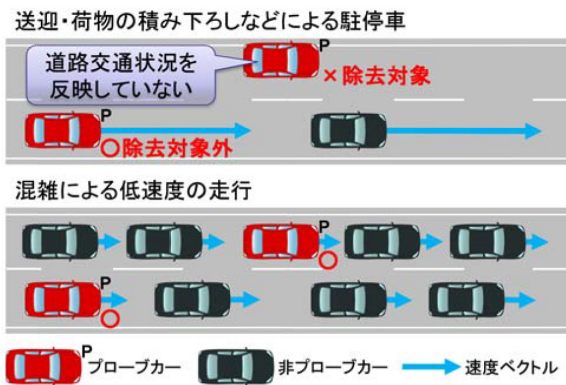


図-1 除去対象とする走行データの例

3. 特異値の発生状況の特性把握

(1) 検証概要

a) 検証のねらい

本章では、本手法により特異値として除去対象となるデータの発生状況の特性を検証する。具体的には除去対象となるデータがどのような日時・時間帯・道路種別・

用途・DRM区間において多くみられるかといった傾向を把握することをねらいとする。

b) 検証対象

本手法を適用する際は、同一区間の同一時間帯にサンプルが複数存在することが前提条件となるため、検証対象としてはサンプル数が多い地域が望ましい。そこで、2016年6月時点で約22万台<sup>4)</sup>と、ETC2.0車載器セットアップ台数が最も多い東京都を対象とする。対象期間は2016年4月18日(月)～24日(日)の1週間とする。

c) 検証方法

本章では、ETC2.0プローブデータのDRMリンク単位別の旅行時間を用いる。同一区間・同一時間帯の最少所要時間(但し、高速道路では80km/h、一般道では30km/hでの旅行速度による所要時間を下限値とする)に時間閾値600秒を加算した時間よりも遅れるものを特異値データとして抽出した。その結果をもとに、特異値の発生状況の特性を検証する。

(2) 検証結果

a) 日別集計

表-3に示すように特異値の件数は4月24日(日)が最も多い。比率でみると4月22日(金)が最も多い。

b) 時間帯別集計

表-4に示すように7～9時台が件数、比率ともに最も多い。一方で0～5時台は件数10未満、比率0.005%未満と他の時間帯と比較して少ない。

この結果をみると、朝ピーク時のような速度変化が比較的大きい時間帯において除去対象データが生じやすいことがうかがえる。

c) 道路種別別集計

表-5に示すように件数は都市内高速が最も多く、そのほかの種別では概ね近い件数となっている。一方で比率でみると都市間高速が目立って高くなっている。

d) 用途別集計

表-6に示すように除去対象となるデータのほとんどは用途が乗用であるものの、比率でみると用途が乗合であるものが最も多い。これは、タクシー等の影響があると考えられる。また、図-2に示すように、用途別・道路種別別にみると、一般国道や一般都道府県道では用途が乗合の比率が高くなっている。

e) DRMリンク別集計

表-7に示すように除去対象となるデータ件数が最も多い区間は136件で、比率でみても約2.5%と他の区間における除去件数よりも目立って多くなっている。

同区間は東名高速道路における東名川崎ICから東京出口方向への東向き区間であり、平成27年速報の渋滞ワーストランキング<sup>5)</sup>では年間合計で2位、午前ピーク1時間でも朝7時台が2位、速度40km/hの渋滞損失時間は1位

となっており、全国的にみても渋滞が顕著な区間である。

また、図-3に示すように、一般道路では除去対象件数が数件しかない区間がほとんどであるのに対し、高速道路では除去対象件数が多い区間が存在することが分かる。

f) DRMリンク距離別旅行速度帯別集計

表-8に示すように除去件数が最も多いのは200~400m区間における2km/h未満の走行データとなっている。また、DRM区間延長が2km以上になると、10km/h以上の走行サンプルも除去対象となっている。

表-3 日別除去対象件数及び比率

	特異値 (件)	特異値以外 (件)	全サンプル (件)	特異値比率 (%)
4月18日(月)	193	1,557,411	1,557,604	0.012%
4月19日(火)	162	1,582,262	1,582,424	0.010%
4月20日(水)	151	1,349,902	1,350,053	0.011%
4月21日(木)	87	954,685	954,772	0.009%
4月22日(金)	265	1,658,767	1,659,032	0.016%
4月23日(土)	212	1,795,454	1,795,666	0.012%
4月24日(日)	271	2,114,465	2,114,736	0.013%

表-4 時間帯別除去対象件数及び比率

時刻 (hh)	特異値 (件)	特異値以外 (件)	全サンプル (件)	特異値比率 (%)
00	4	151,192	151,196	0.003%
01	4	108,550	108,554	0.004%
02	1	91,236	91,237	0.001%
03	1	90,470	90,471	0.001%
04	4	125,413	125,417	0.003%
05	8	306,859	306,867	0.003%
06	89	549,231	549,320	0.016%
07	159	679,501	679,660	0.023%
08	193	706,274	706,467	0.027%
09	121	612,253	612,374	0.020%
10	92	616,041	616,133	0.015%
11	59	584,666	584,725	0.010%
12	84	584,275	584,359	0.014%
13	56	582,238	582,294	0.010%
14	56	589,662	589,718	0.009%
15	57	626,972	627,029	0.009%
16	55	628,609	628,664	0.009%
17	87	673,193	673,280	0.013%
18	62	651,266	651,328	0.010%
19	43	546,490	546,533	0.008%
20	31	502,484	502,515	0.006%
21	41	437,719	437,760	0.009%
22	20	331,463	331,483	0.006%
23	14	236,889	236,903	0.006%
合計	1,341	11,012,946	11,014,287	0.012%

表-5 道路種別別除去対象件数及び比率

	特異値 (件)	特異値以外 (件)	全サンプル (件)	特異値比率 (%)
都市間高速	251	512,003	512,254	0.049%
都市内高速	372	3,151,296	3,151,668	0.012%
一般国道	222	2,778,753	2,778,975	0.008%
主要地方道	271	3,286,851	3,287,122	0.008%
一般都道府県道	225	1,283,798	1,284,023	0.018%
他の市町村道 (含東京23区)	0	245	245	0.000%
道路種別計	1,341	11,012,946	11,014,287	0.012%

表-6 用途別除去対象件数及び比率

	特異値 (件)	特異値以外 (件)	全サンプル (件)	特異値比率 (%)
乗用	1,122	9,718,476	9,719,598	0.012%
貨物	124	871,144	871,268	0.014%
特殊	13	85,761	85,774	0.015%
乗合	82	337,558	337,640	0.024%
未定義	0	7	7	0.000%
用途計	1,341	11,012,946	11,014,287	0.012%

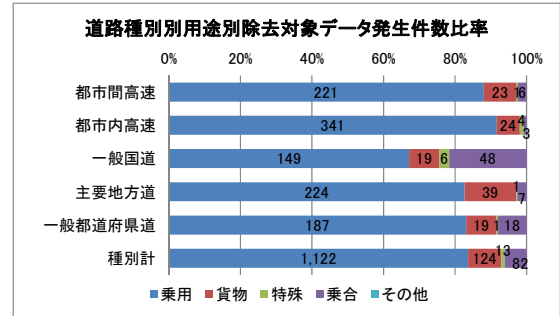


図-2 道路種別別除去対象データ発生件数分布

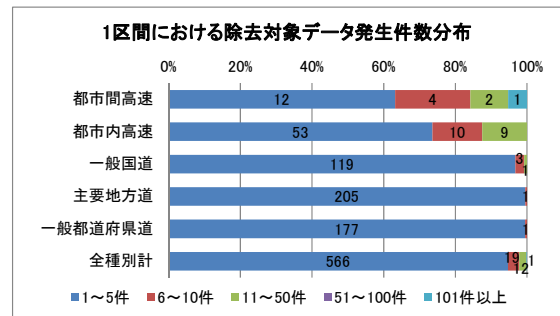


図-3 道路種別別除去対象データ発生件数分布

表-7 DRMリンク別除去対象件数及び比率

2次メッシュ番号	流入ノード	流出ノード	道路種別	特異値 (件)	特異値以外 (件)	全サンプル (件)	特異値比率 (%)
533934	001315	000489	都市間高速	136	5,382	5,518	2.46%
533926	000279	000274	都市内高速	44	4,268	4,312	1.02%
533931	000094	000089	都市間高速	43	2,558	2,601	1.65%
533935	004296	003146	都市内高速	43	5,613	5,656	0.76%
533955	003347	000677	都市内高速	35	2,243	2,278	1.54%
533945	004818	004266	一般国道	28	1,485	1,513	1.85%
533955	000383	000670	都市内高速	21	2,878	2,899	0.72%
533934	000924	002172	都市間高速	20	5,234	5,254	0.38%
533955	004497	003347	都市内高速	17	3,394	3,411	0.50%
533945	000204	000115	都市内高速	16	2,236	2,252	0.71%
533946	00A531	002321	都市内高速	14	3,719	3,733	0.38%
533946	003042	00A685	都市内高速	12	1,741	1,753	0.68%
533956	000808	000143	都市内高速	12	1,833	1,845	0.65%
533944	001810	001808	一般国道	10	1,240	1,250	0.80%
533932	001174	001422	都市間高速	9	2,272	2,281	0.39%
533936	001849	004206	都市内高速	9	6,077	6,086	0.15%
533945	003239	004815	一般国道	9	1,119	1,128	0.80%
533945	004972	002822	都市内高速	9	1,731	1,740	0.52%
533946	00A997	00B042	一般都道府県道	8	387	395	2.03%
533926	000293	000822	都市内高速	7	1,944	1,951	0.36%
533932	001422	001449	都市間高速	7	2,147	2,154	0.32%
533944	102342	102188	主要地方道	7	1,451	1,458	0.48%
533946	003101	003865	都市内高速	7	3,190	3,197	0.22%
533955	004490	000387	都市内高速	7	3,209	3,216	0.22%
533956	003696	001242	都市内高速	7	3,873	3,880	0.18%

除去対象件数が1件以上の区間: 589区間



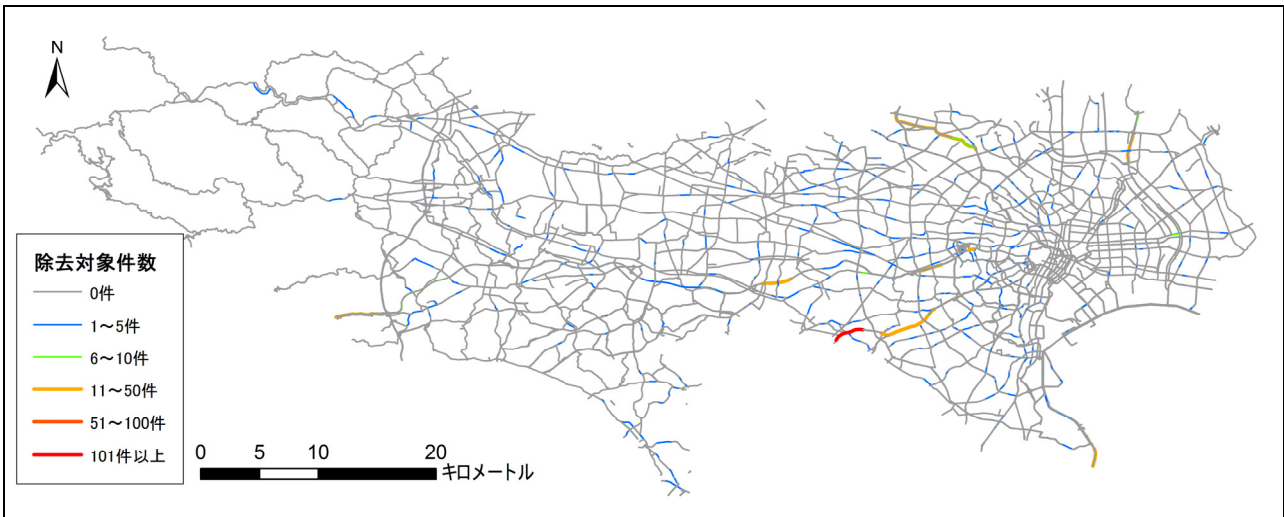


図4 除去対象件数が多いDRMリンクの位置図

表8 DRMリンク距離帯別旅行速度帯別除去対象件数

		旅行速度											合計	
		~2km/h	2km/h~ 4km/h	4km/h~ 6km/h	6km/h~ 8km/h	8km/h~ 10km/h	10km/h~ 12km/h	12km/h~ 14km/h	14km/h~ 16km/h	16km/h~ 18km/h	18km/h~ 20km/h	20km/h~		
DRM 区間 延長	~200m	10	0	0	0	0	0	0	0	0	0	0	0	10
	200m~400m	395	13	0	0	0	0	0	0	0	0	0	0	408
	400m~600m	127	100	0	0	0	0	0	0	0	0	0	0	227
	600m~800m	15	82	6	0	0	0	0	0	0	0	0	0	103
	800m~1000m	4	42	18	0	0	0	0	0	0	0	0	0	64
	1000m~1200m	1	10	24	1	0	0	0	0	0	0	0	0	36
	1200m~1400m	0	6	20	11	0	0	0	0	0	0	0	0	37
	1400m~1600m	0	6	29	42	1	0	0	0	0	0	0	0	78
	1600m~1800m	0	0	4	12	4	0	0	0	0	0	0	0	20
	1800m~2000m	0	1	2	0	1	1	0	0	0	0	0	0	5
	2000m~3000m	0	0	39	68	97	45	5	1	0	0	0	0	255
	3000m~4000m	0	0	1	13	8	6	10	3	3	0	0	0	44
4000m~	0	0	0	0	4	12	14	12	9	3	0	0	54	
合計		552	260	143	147	115	64	29	16	12	3	0	1,341	

(3) 考察

2016年4月18日（月）～24日（日）の東京都の1週間の除去対象となるデータを確認したところ、表4に示すように朝の7～9時台等旅行速度の変化が比較的大きい時間帯に除去対象となるデータが多く生じる傾向がみられた。

また、除去対象データが目立って多く存在する区間があり、その傾向は高速道路において顕著であった。用途の多くは乗用であった。一方、一般道路では多くの区間で除去対象となるデータが数件ずつ発生している傾向にあり、用途においては乗合が占める割合が比較的高い傾向がみられた。

以上を踏まえ、以降では除去対象件数が多く存在する区間に着目し、路側での駐停車と想定されるデータが除去されているか、また駐停車をせずに走行しているデータが除去されていないか等を検証する。

4. 個別の走行データの挙動の検証

(1) 検証概要

a) 検証のねらい

本章では、個別の走行データの挙動について検証する。具体的には個別の走行データの位置情報を確認し、本手法のねらいである600秒の駐停車と想定されるデータが除去できているか、通常の走行と想定されるデータが除去されていないか等を検証する。

b) 検証対象

除去対象となるデータ及び除去対象となるデータの同一時間帯の他のデータの挙動を確認することを踏まえると、除去件数が多い区間を対象とすることが望ましい。また、高速道路と一般道では最少所要時間の下限が異なり、第3章で示したように特異値の現れ方にも違いが生じているため、高速道路と一般道路の双方において検証を行う。

以上より、高速道路で除去対象件数が136件と最も多い区間と、一般道路で除去件数が28件と最も多い区間について、個別の走行データの挙動を検証する。検証区間を表9に示す。

表-9 検証対象区間

	2次メッシュ番号	流入ノード	流出ノード	区間概要
区間①	533934	001315	000489	東名高速道路 東京出口への 東向き方向
区間②	533934	004818	004266	新宿跨線橋 橋西向き方向

c) 検証方法

本章では、同一時間帯の個車の走行データの所要時間及び旅行速度を比較して、除去対象データが特異値といえるかどうかを検証する。用いるデータはETC2.0プローブデータのDRMリンク単位別の旅行時間である。また、600秒以上の駐停車を行ったかを確認するため、ETC2.0プローブデータの個車の位置情報の取得時刻及び緯度経度情報について検証する。

(2) 区間① 東名高速道路 東京出口への東向き方向

a) 区間の概要

本区間は東名高速道路における東名川崎ICから東京出口方向への東向きの区間で東京都内に存在する部分で、DRMリンクの終点は東京出口への分岐点となっている。区間長は2,208mである。

本区間は前述のとおり平成27年速報の渋滞ワーストランキングでは年間合計で2位、午前ピーク1時間でも朝7時台が2位、速度40km/hの渋滞損失時間は1位となっており、全国的にみても渋滞が顕著な区間である。

b) 日別時間帯別集計

本区間における日別時間帯別の除去対象となる個別の走行データは、表-10に示すように平日、休日に依らず、6~9時台までに集中している。また、1件と少ないが、22時台にも除去対象データが存在している。そこで、本区間については6~9時台に除去対象となる個別の走行データが多数存在する4月22日と、22時台に除去対象となる個別の走行データが存在する4月20日の分析を行う。

c) 本区間における除去対象車両の車種及び用途

表-11に示すように、本区間において除去対象となる

車種及び用途は、ほとんどが車種：普通の使用：乗用となっている。

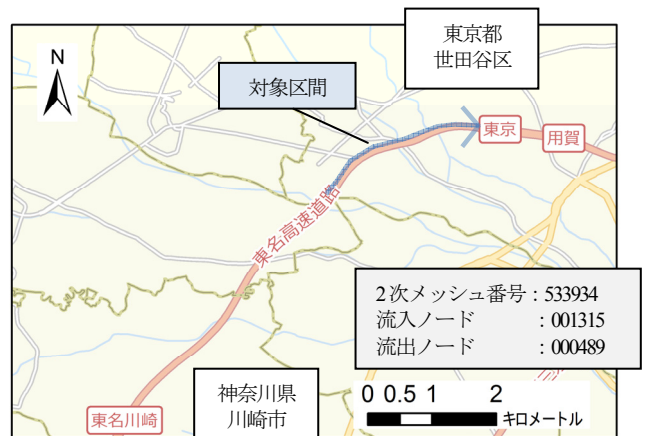


図-5 区間①の対象区間

表-11 区間①の除去対象車両の車種及び用途

	乗用	貨物	特殊	乗合	用途計
軽二輪	0	0	0	0	0
大型	0	0	0	0	0
普通	104	2	0	5	111
小型	7	13	0	0	20
軽自動車	3	2	0	0	5
車種計	114	17	0	5	136

d) 4月22日の個別の走行データの走行速度

図-6に示すように、本区間は6時台から速度低下が始まり、9時台の終わり頃に速度低下が解消している。

6時台は速度低下が発生し始める時間帯である。6時台の最速車両は6:04に本区間に流入しており、旅行速度は85.8km/hと80km/hを上回っている。そのため所要時間の下限値（80km/hで走行した際の所要時間）に600秒を加算した699.4秒が本区間の閾値になる。旅行速度で示すと11.4km/hとなる（表-12）。

7時台、8時台は終始速度低下が生じている。7時台、8時台の最速の個別の走行データはそれぞれ28.1km/h、53.8km/hと下限値の80km/hを下回っている。そのため、それぞれ最速となる個別の走行データの所要時間に600秒を加算した883.2秒、747.7秒が本区間の閾値になる。旅行速度で示すと9.0km/h、10.6km/hとなる（表-12）。

表-10 区間①における日別時間帯別除去対象サンプル数

	(時)																							合計	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
4月18日(月)	0	0	0	0	0	0	0	13	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23
4月19日(火)	0	0	0	0	0	0	0	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
4月20日(水)	0	0	0	0	0	0	0	9	17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	27
4月21日(木)	0	0	0	0	0	0	9	1	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
4月22日(金)	0	0	0	0	0	0	1	6	26	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43
4月23日(土)	0	0	0	0	0	0	4	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
4月24日(日)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
合計	0	0	0	0	0	0	14	42	69	10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	136

9時台は速度低下が解消していく時間帯である。9時台の最速の個別の走行データは速度低下が解消したといえる9:59発に本区間に流入しており、旅行速度は85.1km/hと80km/hを上回っている。そのため所要時間の下限値(80km/hで走行した際の所要時間)に600秒を加算した699.4秒が本区間の閾値となる。旅行速度で示すと11.4km/hとなる(表-12)。

4月22日における速度低下の状況をみると、図-7に示すように、6時台や9時台のように同一時間帯の中で交通状況が変動し速度が高い時間帯と低い時間帯が混在する場合、閾値は高い速度に依存してしまい、個別の走行データの走行時における最高速度の低下状況をうまく閾値に反映できない区間があることが確認できた。

なお、7、8時台の個別の走行データをみると、閾値の前後の所要時間となっているデータが多くみられており、駐車による特異値データが除去対象になっていないか、単に混雑により速度が低下していたデータが除去されてしまった可能性もある。この点については、それぞれの個別の走行データの位置情報を見ることで確認する。

表-12 4月22日の6~9時台の閾値

	最高速度 (km/h)	最高速度時の 旅行時間(s)	80km/hでの 旅行時間(s)	閾値となる 所要時間(s)	閾値の 所要時間の 旅行速度 (km/h)
	①	②	③	MAX(②、③) +600秒	
6時台	85.8	92.6	99.4	699.4	11.4
7時台	28.1	283.2	99.4	883.2	9.0
8時台	53.8	147.7	99.4	747.7	10.6
9時台	85.1	93.4	99.4	699.4	11.4

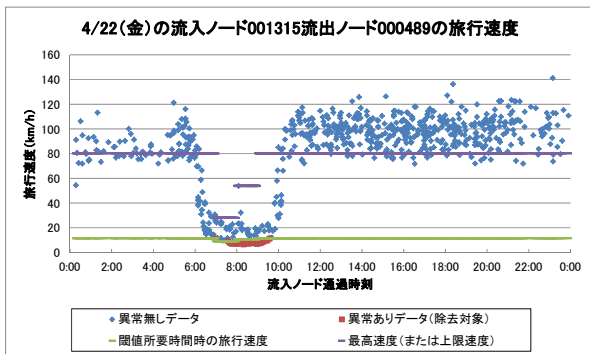


図-6 4月22日の個別の走行データの走行速度

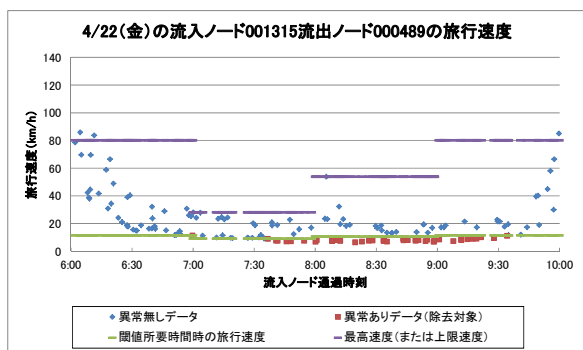


図-7 4月22日の個別の走行データの走行速度 (6~9時台)

e) 4月22日の個別の走行データの位置情報の確認

7時台の最遅車両に着目し、600秒以上の駐停車をしているか確認した。その結果、表-13に示すように同車両のおおよそ100m間隔での移動時間は概ね20~80秒程度で、5~20km/h前後で走行していることが分かった。2km/hで走行するような時間帯もあるものの、所要時間はいずれも3分以内であり、600秒以上の駐停車は伴っていない。以上より、同車両は渋滞時の走行とみなすことができ、本手法の目的に照らせ合わせると除去したくないデータであったといえる。

また、同じく7時台において除去対象外となる最遅車両に着目し、その位置情報を確認した。その結果、表-14に示すように同車両のおおよそ200m間隔での移動時間は概ね40~120秒程度で、5~20km/h前後で走行していることが分かった。同車両は表-13で検証した、7時台の除去対象となる最遅車両の11分前に同区間に流入しており、走行状況は類似していたと想定される。なお、位置情報の取得間隔が異なるのはETC2.0のカーナビのタイプの違いによるものである。

表-13、表-14の位置情報を見ると、両車両とも渋滞時の走行とみなすことができ、表-13で検証した車両は本手法の目的に照らせ合わせると除去したくないデータであったといえる。

表-13 4月22日7時台の除去対象となる最遅車両の位置情報

測位時刻	直前の測位点からの所要時間(s)	直前の測位点からの移動距離(m)	直前の測位点からの移動距離/所要時間(km/h)
7:46:13			
7:46:50	37	94	9.1
7:47:09	19	91	17.2
7:47:26	17	94	19.9
7:47:48	22	95	15.5
7:49:00	72	101	5.1
7:49:52	52	103	7.1
7:51:02	70	103	5.3
7:51:34	32	102	11.5
7:52:28	54	104	6.9
7:54:13	105	106	3.6
7:54:44	31	105	12.2
7:56:09	85	102	4.3
7:59:35	146	107	2.6
8:00:26	111	101	3.3
8:01:05	39	104	9.6
8:01:27	22	103	16.9
8:01:46	19	105	19.9
8:02:00	14	103	26.5
8:02:48	48	108	8.1
8:04:08	80	105	4.7
8:04:37	29	105	13.0

表-14 4月22日7時台の除去対象外となる最遅車両の位置情報

測位時刻	直前の測位点からの所要時間(s)	直前の測位点からの移動距離(m)	直前の測位点からの移動距離/所要時間(km/h)
7:35:59			
7:36:46	47	190	14.6
7:38:26	100	197	7.1
7:39:16	50	203	14.6
7:41:09	113	210	6.7
7:42:35	86	207	8.7
7:43:35	60	203	12.2
7:44:24	49	206	15.1
7:46:32	128	208	5.9
7:47:17	45	213	17.0

f) 4月20日の個別走行データの走行速度

図-8に示すように、本区間も4月22日と同様に、6時台から速度低下が始まり、9時台の終わり頃に速度低下が解消している。なお、6時台には除去対象データは生じていない。

6時台は速度低下が発生し始める時間帯である。6時台の最速車両は6:04に本区間に流入しており、旅行速度は85.8km/hと80km/hを上回っている。そのため所要時間の下限値（80km/hで走行した際の所要時間）に600秒を加算した699.4秒が本区間の閾値になる。旅行時間で示すと11.4km/hとなる（表-15）。

7時台は終始速度低下が生じている。7時台の最速車両は45.7km/hと下限値の80km/hを下回っているため、最速車両の所要時間に600秒を加算した774.1秒が本区間の閾値になる。旅行速度で示すと10.3km/hとなる（表-15）。

8時台から速度低下が解消し始め、8時台の最速車両は8:49発に本区間に流入しており、旅行速度は104.4km/hと80km/hを上回っている。そのため所要時間の下限値（80km/hで走行した際の所要時間）に600秒を加算した699.4秒が本区間の閾値となる。旅行速度で示すと11.4km/hとなる（表-15）。

また、22時台をみると1車両のみ約10.6km/hで走行しているものの、同時間帯の最速車両は130km/h程度と80km/hを上回っている。そのため所要時間の下限値（80km/hで走行した際の所要時間）に600秒を加算した699.4秒が本区間の閾値となる。旅行速度で示すと11.4km/hとなる。旅行速度が10.6km/hの車両の所要時間は約745秒と閾値を上回るため、除去対象データとなっている。

4月20日における速度低下の状況を見ると、同一時間帯の最高速度の低下に伴い閾値が低下していることが確認できた。また、22時台に1サンプルのみ混入している10km/h程度の個別の走行データも除去できることが分かった。

しかし、4月22日と同様に、閾値の前後の値となっているものが多くみられているため、それぞれの個別の走行データの位置情報を見ることで確認する。

表-15 4月20日の6～9時台の閾値

	最高速度(km/h)	最高速度時の旅行時間	80km/hでの旅行時間(s)	閾値となる所要時間(s)	閾値の所要時間の旅行速度(km/h)
	①	②	③	MAX(②、③)+600秒	
6時台	93.0	85.5	99.4	699.4	11.4
7時台	45.7	174.1	99.4	774.1	10.3
8時台	104.4	76.1	99.4	699.4	11.4
9時台	106.7	74.5	99.4	699.4	11.4

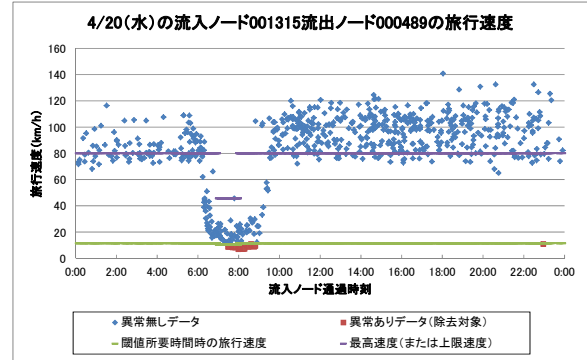


図-8 4月20日の個別の走行データの走行速度

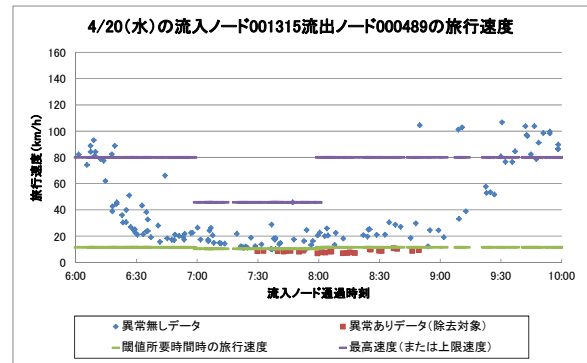


図-9 4月20日の個別の走行データの走行速度（6～9時台）

g) 4月20日の個別走行データの位置情報の確認（6～9時台）

4月20日の7, 8時台の最遅車両に着目し、600秒以上の駐停車をしているかを確認した。その結果、表-16に示すように、同車両のおおよそ100m間隔での移動時間は概ね20～70秒程度で、5～20km/h前後で走行していることが分かる。2km/hで走行するような時間帯もあるものの、所要時間はいずれも3分以内となっており、600秒以上の駐停車は行っていない。

また、4月20日の7, 8時台において除去対象外となる最遅車両なものに着目し、その位置情報を確認した。その結果、表-17に示すように、同車両のおおよそ200m間隔での移動時間は概ね30～90秒程度で、10～20km/h前後で走行していることが分かった。同車両は表-13で検証した、除去対象となる最遅車両の6分前に同区間に流入しているため、走行状況は類似していたと想定される。

表-16, 表-17の位置情報を見ると、両車両とも渋滞時



の走行とみなすことができ、表-16で検証した車両は本手法の目的に照らせ合わせると除去したくないデータであったといえる。

表-16 4月20日の7, 8時台において除去対象となる最遅車両の位置情報

測位時刻	直前の測位点からの所要時間(s)	直前の測位点からの移動距離(m)	直前の測位点からの移動距離/所要時間(km/h)
7:59:55			
8:00:26	31	92	10.7
8:00:45	19	93	17.6
8:01:35	50	95	6.8
8:02:12	37	96	9.3
8:02:47	35	100	10.3
8:03:21	34	102	10.8
8:03:51	30	104	12.5
8:04:12	21	104	17.8
8:05:11	59	103	6.3
8:07:34	143	104	2.6
8:08:28	54	102	6.8
8:08:50	22	103	16.9
8:09:14	24	101	15.2
8:10:23	69	102	5.3
8:11:33	70	103	5.3
8:12:05	32	103	11.6
8:16:03	178	106	2.1
8:16:24	21	107	18.3
8:16:30	6	51	30.6
8:16:31	1	2	7.2
8:16:54	23	52	8.1
8:18:29	95	105	4.0

表-17 4月20日の7, 8時台の除去対象外となる最遅車両の位置情報

測位時刻	直前の測位点からの所要時間(s)	直前の測位点からの移動距離(m)	直前の測位点からの移動距離/所要時間(km/h)
7:53:42			
7:54:25	43	182	15.2
7:55:00	35	186	19.1
7:55:53	53	204	13.9
7:58:00	67	212	11.4
7:58:50	50	206	14.8
8:00:17	87	207	8.6
8:01:34	77	202	9.4
8:02:40	66	201	11.0
8:06:10	90	210	8.4
8:07:00	50	210	15.1

h) 4月20日の個別走行データの位置情報の確認 (22時, 23時台)

さらに、22時台の除去対象の個別の走行データに着目すると、表-18に示すように、600秒以上の駐停車は行っていない。しかし同一時間帯に他の全車両が高速で走行していることを考えると、駐停車は行っていないものの特異な走行データであると考えられる。本手法ではこのようなデータも除去できる仕組みになっていることが確認できた。

なお、同一時間帯の位置情報を詳細にみると、表-19に示すように、同一区間に3分遅れて流入した車両は区

間全体を80km/h前後で走行しており、この結果からも22時台の最遅サンプルは特異な走行をしていたといえる。

表-18 4月20日の22時台の最遅車両の位置情報 (除去対象)

測位時刻	直前の測位点からの所要時間(s)	直前の測位点からの移動距離(m)	直前の測位点からの移動距離/所要時間(km/h)
22:57:56			
22:58:31	35	172	17.7
22:58:34	3	12	14.4
22:59:03	29	163	20.2
22:59:11	8	32	14.4
22:59:41	30	108	13.0
23:00:08	13	95	26.3
23:01:35	87	186	7.7
23:01:46	11	22	7.2
23:02:52	66	209	11.4
23:03:33	41	154	13.5
23:03:48	15	53	12.7
23:03:56	8	49	22.1
23:04:27	31	94	10.9
23:04:55	28	59	7.6
23:05:13	18	69	13.8
23:06:18	65	139	7.7
23:06:20	2	4	7.2
23:07:02	42	45	3.9
23:07:32	30	64	7.7
23:08:31	59	70	4.3
23:08:42	11	29	9.5
23:08:43	1	2	7.2
23:09:51	68	208	11.0

表-19 4月20日の22時台の最遅車両の3分後に同一区間に流入した車両の位置情報

測位時刻	直前の測位点からの所要時間(s)	直前の測位点からの移動距離(m)	直前の測位点からの移動距離/所要時間(km/h)
23:00:07			
23:00:16	9	179	71.6
23:00:26	10	214	77.0
23:00:34	8	172	77.4
23:00:43	9	206	82.4
23:00:52	9	204	81.6
23:01:01	13	203	56.2
23:01:10	9	201	80.4
23:01:19	9	205	82.0
23:01:29	10	228	82.1
23:01:38	9	203	81.2

i) 本区間での検証まとめ

本区間では、混雑時など同一時間帯の最速車両の速度低下が生じる区間において、閾値となる所要時間の増大(=旅行速度の低下)が確認できた。これは、本手法の特徴である他の車両との旅行速度差を用いて除去対象サンプルを抽出する機能が働いていることを意味する。

一方で、同一時間帯に交通状況が大きく変動する場合、所要時間が短い(旅行速度が速い)走行データによる所要時間をもとに閾値が算出される可能性があり、混雑により所要時間が長い(旅行速度が遅い)走行データが除

去対象となる可能性がある。

さらに、混雑時間帯であっても、図-8のように7時台に40km/h以上の速度で走行しているデータが存在すると、同データの所要時間をもとに閾値が算出されるため、混雑により所要時間が長い（旅行速度が遅い）走行データが除去対象となる可能性がある。

まとめると、本区間のような①同一時間帯に混雑が発生または解消した際の旅行速度差が大きい、②混雑時でも車両ごとの旅行速度差が大きいといった傾向がみられると、本来は除去すべきでないサンプルが除去対象となってしまう可能性があることが確認できた。

①については最速サンプルを抽出する時間幅を短縮することが考えられる。図-10は、本区間において実験的に閾値となる最短所要時間を抽出する幅を10分に短縮した際の閾値及び除去対象データである。6時台でも6:30前後から7時にかけては閾値の低下がみられる。

また、②については最速サンプルの所要時間に加算する時間600秒をDRMリンク長によって変動させることが考えられる、図-11は、本区間において実験的に現在の時間閾値にさらに600秒を追加（時間閾値を1,200秒に設定）した際の閾値及び除去対象データである。本区間の全ての走行データが除去対象外になることが分かる。

なお、図-10、図-11は、本区間において本手法の目的に照らし合わせて望ましいデータ処理を行うためにあくまでに試行的に条件を変動させた結果である。実際の適用にはより詳細な分析が必要になる。

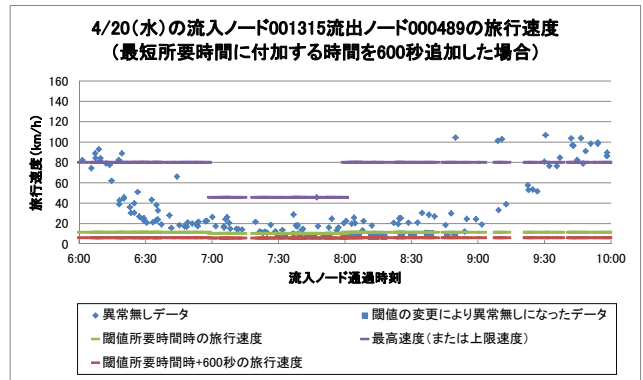


図-11 4月20日の個別走行データにおいて最短所要時間に付加する時間を仮に600秒追加した場合の特異値判定（6～9時台）

(2) 区間② 新宿跨線橋仮橋 西向き方向

a) 区間の概要

本区間は新宿跨線橋仮橋の西向き区間で、国道20号において、都道芝新宿王子線と交差する新宿4丁目から都道四谷角筈線と交差する西新宿1丁目までの区間の一部である。区間長は313mである。

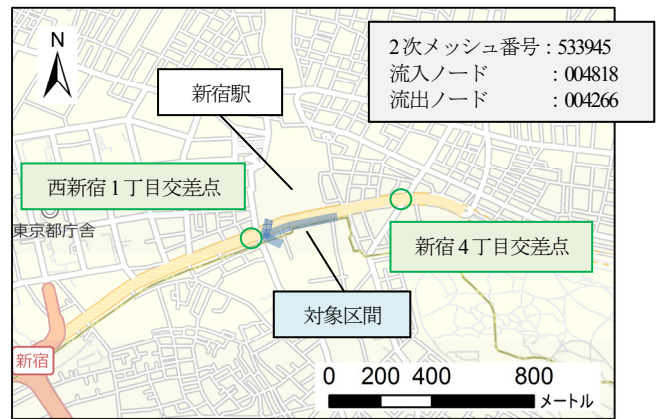


図-12 区間②の対象区間

b) 日別時間帯別集計

表-20に示すように、本区間における日別時間帯別の除去対象となる個別の走行データは、発生時間帯にあまり偏りはみられない。また、平日、休日に依らず、いずれの日にも数サンプルが除去対象となっている。1日あたりの除去対象個別走行データ数でみると、4月20日（水）が最も多くなっている。そこで本区間については、4月20日の個別走行データを検証する。

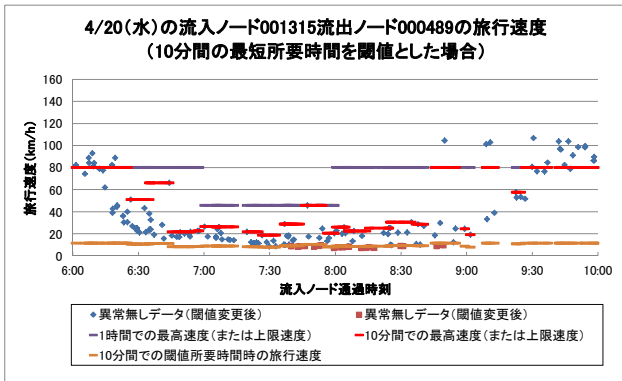


図-10 4月20日の個別走行データにおいて最短所要時間幅を仮に10分とした場合の特異値判定（6～9時台）

表-20 区間②における日別時間帯別除去対象サンプル数

	(時)																							合計		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23		
4月18日(月)	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	1	0	4	
4月19日(火)	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	1	0	1	0	6	
4月20日(水)	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	0	0	0	0	7	
4月21日(木)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	3	
4月22日(金)	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
4月23日(土)	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	3	
4月24日(日)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	
合計	0	0	0	0	0	0	1	3	0	4	2	1	1	0	1	2	1	2	2	2	2	2	1	2	1	28

c) 本区間における除去対象車両の車種及び用途

表-21に示すように、本区間において除去対象となる車種及び用途は、全て車種：普通用途：乗合となっている。これはタクシー等の旅客運送用の車両に多い組み合わせである。本区間の600秒以上の駐停車はタクシーの移動特性が関係している可能性がうかがえる。

表-21 区間②の除去対象車両の車種及び用途

	乗用	貨物	特殊	乗合	用途計
軽二輪	0	0	0	0	0
大型	0	0	0	0	0
普通	0	0	0	28	28
小型	0	0	0	0	0
軽自動車	0	0	0	0	0
車種計	0	0	0	28	28

d) 4月20日の個別の走行データの走行速度

図-13に示すように、本区間は1日を通して低速度から高速度の車両が混在している。また表-22に示すように、15時台、17時台では最速サンプルの速度が30km/h（一般道の最少所要時間の下限値における走行速度）未滿となっているものの、そのほかの時間帯では30km/h以上となっている。

図-14に示すように、4月20日における速度分布をみると、8時～11時台では除去対象となる個別の走行データと近い速度の除去対象外の個別の走行データがいくつか見られる。一方で、図-15に示すように、14時台～17時台においては除去対象と除去対象外の個別の走行データには一定の速度差がみられる。

以降では、除去対象の走行データ及び除去対象外の低速な走行データの位置情報を確認する。なお、本区間では、マップマッチングに失敗した位置情報が多かったため、マップマッチング前の位置情報を用いて検証を行う。

表-22 4月20日の14～17時台の閾値

	最高速度	最高速度時の	30km/hでの	閾値となる	閾値の
	(km/h)	旅行時間(s)	旅行時間(s)	所要時間(s)	所要時間の
	①	②	③	MAX(②、③) +600秒	旅行速度 (km/h)
14時台	37.6	30.0	37.6	637.6	1.77
15時台	27.4	41.2	37.6	641.2	1.76
16時台	34.9	32.2	37.6	637.6	1.77
17時台	27.7	40.7	37.6	640.7	1.76

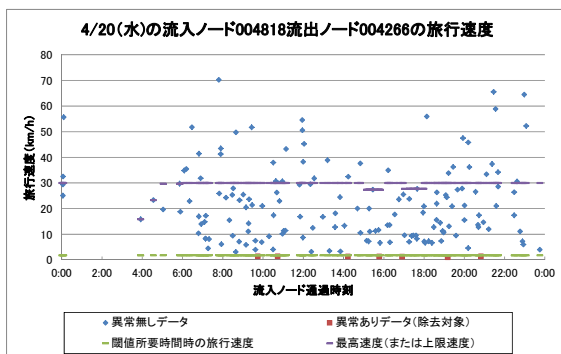


図-13 4月20日の個別の走行データの走行速度

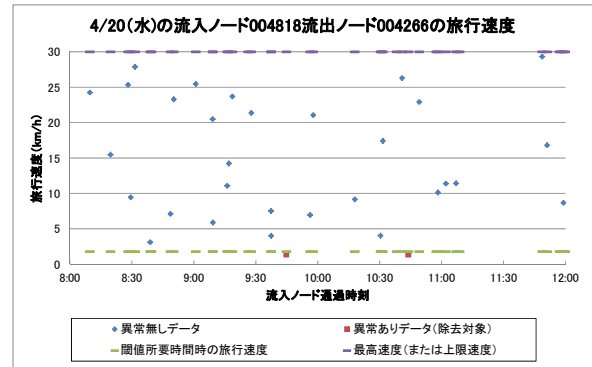


図-14 4月20日の8～12時台の個別の走行データの走行速度

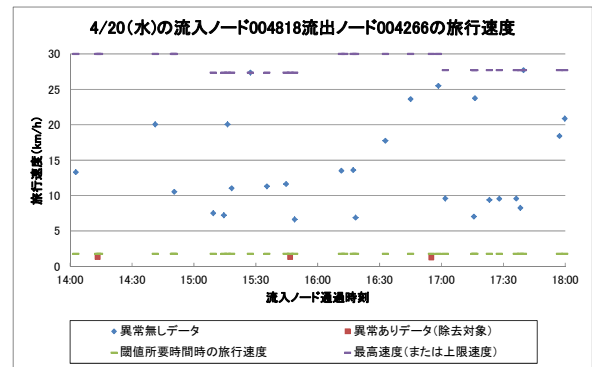


図-15 4月20日の14～18時台の個別の走行データの走行速度

e) 4月20日の個別の走行データのマップマッチング前の位置情報の確認（除去対象の最遅車両）

4月20日の最遅車両に着目し、600秒以上の駐停車をしているか確認した。

本車両の運行ID（同一車両であれば1日間常に同一になる番号）は18460である。本車両は9:11に本区間に流入し、19:29に本区間を流出している。マップマッチング前の位置情報を見ると、表-23に示すように通し番号（一連の移動に与えられる連番）22と23の間に流入し、36と37の間に流出している。

図-16に示すように、本車両は該当区間の南側に位置情報が集中している箇所が存在する。このエリアはバスタ新宿があり、3階はタクシー乗降所となっている。位置情報がバスタ新宿内外を転々としていることから、位置情報の取得誤差が生じているか、バスタ新宿に流入している可能性がうかがえる。また、通し番号24と25の間に14分程度間隔が空いていることから、600秒以上の駐停車している可能性が高いと想定される。

以上より、本個別走行データは、本区間をずっと走行しているか、本区間とその周辺を転々としているものの、いずれにしても600秒以上駐停車等をしてきた可能性の高い走行データであると想定される。



表-23 4月20日の除去対象の最遅車両のマップマッチング前の位置情報

運行ID	通し番号	測位時刻	直前の測位点からの所要時間(s)
18460	22	19:10:44	19
18460	23	19:11:21	37
18460	24	19:12:10	49
18460	25	19:25:57	827
18460	26	19:26:01	4
18460	27	19:26:16	15
18460	28	19:26:22	6
18460	29	19:26:28	6
18460	30	19:28:32	124
18460	31	19:28:36	4
18460	32	19:28:55	19
18460	33	19:29:06	11
18460	34	19:29:19	13
18460	35	19:29:30	11
18460	36	19:29:43	13
18460	37	19:30:08	25

以上より、本個別走行データは、本区間かその周辺で600秒以上駐停車等をしてきた走行データであると想定される。

表-24 4月20日の除去対象の最速車両のマップマッチング前の位置情報

運行ID	通し番号	測位時刻	直前の測位点からの所要時間(s)
37535	33	10:43:52	51
37535	34	10:44:05	13
37535	35	10:44:13	8
37535	36	10:55:27	674
37535	37	10:55:29	2
37535	38	10:55:37	8
37535	39	10:55:51	14
37535	40	10:55:55	4
37535	41	10:56:03	8
37535	42	10:56:59	56
37535	43	10:57:07	8
37535	44	10:57:27	20
37535	45	10:57:40	13
37535	46	10:58:16	36



図-16 4月20日の除去対象の最遅車両のマップマッチング前の位置情報 (同区間周辺)



図-17 4月20日の除去対象の最速車両のマップマッチング前の位置情報 (同区間周辺)

f) 4月20日の個別の走行データのマップマッチング前の位置情報の確認 (除去対象の最遅車両)

また、4月20日の除去対象データのうち最速となる車両に着目し、600秒以上の駐停車をしているか確認した。

本車両の運行IDは37535である。本車両は10:43に本区間に流入し、10:57に本区間を流出している。マップマッチング前の位置情報を見ると、表-24に示すように通し番号33と34の間に流入し、45と46の間に流出している。

図-17に示すように、本車両も該当区間の南側に位置情報が集中している箇所がある。東側からタクシー乗り場に流入し、そこで1周し、タクシー乗り場から流出する様子が見える。

表-21に示すように、本区間でのマップマッチング前の位置情報のうち通し番号35と36の間に11分程度間隔が空いている。図-17でも、通し番号35と36の間にタクシー乗り場等で駐停車を行っていたことが想定される。

g) 4月20日の個別の走行データのマップマッチング前の位置情報の確認 (除去対象外の最遅車両)

さらに、4月20日において除去対象外のデータのうち最速となる車両に着目し、600秒以上の駐停車をしているか確認した。

本車両の運行IDは3713である。本車両は8:38に本区間に流入し、8:44に本区間を流出している。マップマッチング前の位置情報を見ると、表-25に示すように通し番号11と12の間に流入し、25と26の間に流出している。

図-18に示すように、本車両も該当区間の南側に位置情報が集中している箇所がある。東側からタクシー乗り場に流入し、そこで1周し、タクシー乗り場から流出する様子が見える。



表-25に示すように、本区間でのマップマッチング前の位置情報のうち通し番号11と12の間に7分程度、通し番号21と22の間に3分程度間隔が空いている。なお本車両は通し番号11と12の間で本区間に流入していることから、通し番号11と12の間の7分間終始駐停車をしていたということはない。また、図-18に示すように通し番号21と22間に本車両は位置を変えている。距離としては1kmもない移動であるため、多少の駐停車を伴っている可能性はあるものの、600秒以上の駐停車はしていないと想定される。

表-25 4月20日の除去されていない最も低速な個別走行データのマップマッチング前の位置情報

運行ID	通し番号	測位時刻	直前の測位点からの所要時間(s)
3713	11	8:33:12	812
3713	12	8:40:15	423
3713	13	8:40:28	13
3713	14	8:40:33	5
3713	15	8:40:38	5
3713	16	8:40:46	8
3713	17	8:41:00	14
3713	18	8:41:03	3
3713	19	8:41:06	3
3713	20	8:41:09	3
3713	21	8:41:12	3
3713	22	8:43:59	167
3713	23	8:44:03	4
3713	24	8:44:26	23
3713	25	8:44:42	16
3713	26	8:45:02	20



図-18 4月20日の除去対象外の最遅車両のマップマッチング前の位置情報 (同区間周辺)

f) 本区間での検証まとめ

本区間においても高速道路と同様に、混雑時などの同一時間帯の最速の個別走行データの速度低下が生じる区間において、閾値となる所要時間の増大 (= 旅行速度の低下) が確認できた。これは、本手法の特徴である他の

車両との旅行速度差を用いて除去対象サンプルを抽出する機能が働いていることを意味する。

また、マップマッチング前の位置情報から600秒以上の駐停車が発生したと想定されるサンプルを除去対象として抽出できていることも確認できた。さらに、600秒に満たない駐停車と想定される車両は、除去していないことも確認できた。

本区間において特異値の除去が目的通りにできた要因としては、①同一時間帯に混雑が発生または解消した際の旅行速度差が小さい、②混雑時でも車両ごとの旅行速度差が小さく、同一時間帯において30km/hを大きく下回るサンプルが存在しなかったことが考えられる。

なお、本研究の意義である「本手法が600秒以上の駐停車またはそれに準ずる走行データを特異値として除去するといった目的を達成できているかの検証」とは視点が異なるものの、600秒未満であっても駐停車をしていると想定される走行データは本来平均旅行時間からは除くことが望ましい。本手法では600秒未満の駐停車時間を所要時間に計上している。対応としては時間閾値を短縮することが考えられるが、一方で駐停車していない走行データを除去してしまう可能性がある。松島らは時間閾値を600秒とすることで、特異値でないデータを残すこと及び特異値データを除去することの均衡が取れることを感度分析により確認しているが、同時に分析目的に合致する時間閾値の整理を課題として挙げている。本研究からタクシー乗り場にアクセスする区間を分析したい場合は時間閾値を短くすることで駐停車による所要時間が除去できるようになることが分かることから、引き続き分析目的に合致する時間閾値の整理についても分析を進めることが望ましい。

5. おわりに

本研究では、同一区間の同一時間帯に走行データが複数存在する区間において、本手法によって特異値と判別されるデータが多い区間の特性について検証した。その結果、本データクレンジング手法により特異値と判別されるデータは旅行速度の変化が比較的大きい時間帯に多く見られること、特定区間に偏ってみられることを把握した。

また、除去対象データが多く存在する区間において、特異値と判別されたデータとそうでないデータを比較し、本手法の適用性について検証した。その結果、本手法の着眼点であった同一時間帯の最高速度の低下に伴い閾値が変動することによって、600秒程度駐停車をしている車両及びそれに準ずる異常と想定される走行をしている車両を除外できることを確認した。

一方で、以下の課題を確認した。

- ①同一時間帯に交通状況が大きく変動する場合、所要時間が短い（旅行速度が速い）走行データによる所要時間をもとに閾値が算出される可能性があり、混雑により所要時間が長い（旅行速度が遅い）走行データが除去対象となる可能性がある
- ②混雑時でも車両ごとの旅行速度差が大きい区間においては、最速データでの所要時間をもとに閾値が算出されるため、混雑により所要時間が長い（旅行速度が遅い）走行データが除去対象となる可能性がある

①については、閾値を設定する時間帯幅を600秒、400秒、200秒等と狭めていく対応が考えられる。但し、時間帯の閾値を短くすると、時間帯毎に取得できるサンプル数が減少することに留意が必要である。今回検証した東名高速自動車道は一定数のサンプルが確保できているため、図-10のように時間帯の閾値を短くすることで駐停車ではないと想定される幾つかの走行データを除去対象外にすることができたが、今後東京都以外の様々な地域で本手法での特異値除去を行っていくことを想定すると、閾値を設定する時間帯幅を狭める際は取得サンプル数に留意する必要がある。取得サンプル数との関係を含め、時間帯の閾値についても今後より厳密に検証を行うことが望ましい。

②については、閾値算出の際に用いる「600秒」を距離帯によって調整する対応が考えられる。今回検証した東名高速自動車道では最少所要時間に合計で1,200秒加

算した所要時間を閾値とすることで駐停車ではないと想定される走行データを除去対象外にすることができたが、この「1,200秒」はリンク長、混雑時の最短所要時間データとその他データの所要時間差等の影響を受けると考えられる。リンク長や同一時間帯の他の走行データの速度との関係を含め、閾値算出の際に用いる「600秒」についても今後より厳密に検証を行うことが望ましい。

引き続き、本手法の吟味を進め、データクレンジング手法としての精緻化を目指す。

#### 参考文献

- 1) 松島敏和, 橋本浩良, 高宮進: プローブ旅行速度データのデータクレンジング手法の開発, 土木計画学研究発表会・講演集, Vol.53, 2016
- 2) 橋本浩良: 旅行時間データのサンプル数と旅行時間算定値の代表性の関係, 土木技術資料 56-10,
- 3) 野間真俊, 奥谷正, 橋本浩良: 道路ネットワークの評価における時間信頼性指標の適用に関する研究, 土木計画学研究・講演集, Vol.37, 2008
- 4) ETC2.0総合情報ポータルサイト: セットアップ件数の推移-ETC2.0 (DSRC)
- 5) 国道交通省HP: 渋滞ワーストラランキング平成27年速報値

(2016.?? 受付)

## A STUDY OF AN APPLICATION OF A DATA CLEANSING METHOD FOR PROBE TRAVEL TIME DATA

Taichi TACHIKAWA, Hiroyoshi HASHIMOTO, Shinsuke SETOSHITA and Toshikazu MATSUSHIMA