因子分析による都市規模自動車交通シミュレーション結果の解釈

内 種 岳 詞*・伊 藤 伸 泰*,**

Applying Factor Analysis to Describe Urban Scale Vehicle Traffic Simulation Results

Takeshi UCHITANE* and Nobuyasu ITO*,**

In this paper, we apply a factor analysis which is a kind of multivariate statistics to the results of vehicle traffic simulation of Kobe-city, Japan. City scale traffic simulations usually output multivariate results. However, both in the context of data assimilation and in the context of statistical hypothesis testing, such results are often treated as independent variables. Lack of concerning such result correlations results in worse accuracy of the simulation result. Therefore, better statistical comparisons and better statistical analyses between simulation results are strongly required. From the results of factor analysis, it showed that the large number of simulation results were described by 33 factors. Obtained factors which had higher contribution rate were robust against the time and space of origin and destination and against the number of vehicles. Moreover, the higher contribution rate factors can be distinguished from other factors which can be obtained even if simulation settings were completely random. Though existences of such factors are out of range of this study, the knowledge of factor analysis results may help us not only explain the city traffic simply but also improve accuracy of large scale vehicle traffic simulations.

Key Words: vehicle traffic simulation, large scale road network, multivariate statistics, factor analysis

1. はじめに

自動車交通における渋滞は、経済的ロスや環境汚染の原因となる社会問題である。交通シミュレーションは、将来の交通量予測やバイパス建造など道路工事の影響予測を行なう道具として広く利用が始まっている¹⁾。そして、利用できる実データの増加と、計算機の性能向上により、都市規模を対象とした交通シミュレーションの利用が検討され始めている。都市規模の自動車現象を扱った先行研究では、シミュレーションでの交通現象再現性や交通ビッグデータからの交通予測精度が議論された。Asano らは、神戸市都心部の自動車交通現象を自動車交通シミュレーションで再現する試みを報告した²⁾。シミュレーションにおける調整パラメータとして道の最高速度を変更したが、シミュレーション結果は交通センサスデータに十分に合わないことが報告された。Asif らは、シンガポールの自動車交通の実データから、将来の通行速度を予測するモデルを構築した³⁾。多くの道で自動車交通の時系列データ

が取得可能であるならば、予測モデルをシミュレーションなしで構築できたことが報告された。ただし、この成果は、現状の交通現象を予測可能にするものであり、道路工事などの影響調査や環境汚染への影響の見積もりは、シミュレーションなしに実現できない。これらの先行研究より、都市規模の自動車交通現象を再現することにさへ膨大な実データが必要であり、部分的なデータを利用したシミュレーションによる都市交通の再現は未だ不十分であることが示唆されている。

再現が困難な理由として、都市規模のシミュレーションで利用する実データが多変量であることが挙げられる。自動車交通の再現や評価に用いる交通現象の実データは、

- 道路交通情勢調査 (道路交通センサス)
- 総合都市交通体系調査 (パーソントリップ調査)
- 物資流動調査
- 全国貨物純流動調査 (物流センサス)
- 大都市交通センサス

が考えられてきた^{4),5)}。シミュレーションによる交通現象再現の文脈においては、このような複数種類データ、かつ、それぞれの時系列データに対し、シミュレーション結果を同時に合わせる必要がある。そのとき、複数データの観測場所や観測時間の相関は無視できないバイアスである。たとえば、道ごとの交通量は、道の地理的つながりから何かしらの関係性が存在することは容易に想定される。すなわち、自動車は道に沿って移動するため、ある交通量の多い道の移動先の道も交通量が多くなるような傾向が期待される。このような関係

* 理化学研究所計算科学研究機構
神戸市中央区港島南町 7-1-26

** 東京大学大学院工学系研究科 東京都文京区本郷 7 丁目 3-1

* RIKEN Advanced Institute for Computational Science,
7-1-26 Minatojimaminami-machi, Chuo-ku, Kobe

** School of Engineering, The University of Tokyo, 7-3-1
Hongo, Bunkyo-ku, Tokyo

(Received June 12, 2015)

(Revised July 15, 2016)

性が存在する実データとシミュレーション結果の比較において、関係性はバイアスとして考慮されなければならない。また、現象に含まれる関係性を考慮できれば、実データの観測場所や観測頻度を削減できる利点もある。しかし、実データから関係性を発見するデータマイニングを実施するための、実データ収集コストは非常に高いことやそのような関係性を発見する方法が確立されていないことから、このような関係性は一般的に考慮されない。よって、実データは独立したデータとして取り扱われることが多く、その結果、シミュレーション結果と実データの誤差はより大きく見積もられる。

多変量データの取り扱いによる困難さは、シミュレーション結果の解釈時にも存在する。エージェントベースの自動車交通シミュレーションでは、交通現象を車1台ごとに再現するため、都市規模のシミュレーションからは、車の数や都市に含まれる道の数と同じ次元数の結果が得られる。そして、現実と同様の関係性はシミュレーション結果にも含まれることが見込まれる。また、シミュレーションが比較的低コストで実施できるとはいえ、計算資源の都合上、結果の次元数より少ない回数のシミュレーションしか実施できないことは起こりえる。すると、統計的に有意だが、実在し得ない見せかけの関係性が得られる可能性がある。シミュレーション結果から導かれた見せかけの関係性は、結果の解釈を誤らせる原因になる。

上記の背景より、統計的により妥当なことはもちろん、実データのよりロバストな比較方法やシミュレーション結果のよりロバストな解釈方法が求められる。実データを収集するコストは高いが、交通シミュレーションは比較的低コストでさまざまなシナリオの擬似交通現象を発生できる。そこで、本論文では、シミュレーション結果を多変量データとして取り扱う解釈方法に着目した。交通シミュレータを利用して得られた擬似データに対して、統計的により妥当、かつ、ロバストなシミュレーション結果の解釈方法を考えた。

既存の研究で、内種らは、都市規模の自動車交通シミュレーション結果に関係性が含まれることを報告した⁶⁾。神戸市の都心を対象とした自動車交通シミュレーションの結果には、高速道路と並走する国道の間の自動車通過台数に負の相関関係が存在した。この関係性は、シミュレーション内の高速道路が混雑すれば、並走する国道は空くと解釈できる。しかし、この関係性を発見するための分析は、35952本の道のうち、交通量が日に2700台以上通過した53本の道に限られたものであった。また、関係性の数や強さやロバスト性などは不明であった。よって、より多くの道に対して分析を実施する必要がある。

本論文では、都市規模の自動車交通シミュレーションの結果を扱うため、実在の都市である神戸市規模の自動車交通シミュレーションを実施し、シミュレーション結果にデータマイニングの1手法である因子分析を適用した。データマイニングの文脈では、多変量データにおけるデータ間の関係性は、データの次元数より低い自由度のモデルで表現される。たとえば、1本の国道はほかの複数の道路と交差し自動車が入り・流出するため交差点間の交通量は変動するが、交差点間のよ

うに地理的に連続した道の交通量は独立したデータとはいえない。また、国道と地理的には連続しない並走する高速道路にもそのような関係性が存在する可能性があり、国道と高速道路の交通量も独立したデータとはいえない。よって、因子分析の結果、交通量が独立していると統計的にはいえない複数の道はまとめて1つの道路グループとして表現される。また、因子分析により得られた道路グループが、シミュレーション設定に与えた摂動に対してロバストかを調べた。

実データのデータマイニングから得られた関係性は、実データの解釈を助けるが、シミュレーション結果から得られた結果間の関係性は、シミュレーション結果の解釈を助ける。そのため、得られた因子分析の結果は、現実の現象や実データに含まれる関係性を直接解釈していない。本論文で実施したシミュレーション実験は現実の交通現象の再現を目的としないため、得られた道路グループが現実の交通にも存在するという妥当性の評価は行なわないが、現実の現象を再現したシミュレーション結果が存在すると仮定したとき、そのシミュレーション結果に対して因子分析を適用した結果は、実データの解釈と一致する可能性がある。シミュレーション結果が妥当だと仮定した場合に、得られた関係性が都市規模自動車交通シミュレーション技術にどのように寄与するかを考察した。

2. 都市規模の交通シミュレーション

本研究の目的は、都市規模の自動車交通シミュレーション結果から各道の通過交通量の関係性を因子分析を利用してモデル化することである。都市規模の交通量をモデル化するため、神戸市規模の道路地図を利用した自動車交通シミュレーションを実施した。

ここでは、「道」は交差点間の道のりであり、複数の交差点を経由し進行できる道のりである「道路」と区別して表現する。都市に含まれるすべての道の通過交通量をシミュレーション結果として得るには、すべての道を車が通過できなくてはならない。そのため、自動車交通シミュレーションでよく利用されている Open Street Map⁸⁾では、道間の接続性の面で不十分である。そこで、浅野らが(株)ゼンリンのデジタル道路地図から構築した Simulation of Urban Mobility (SUMO)⁹⁾用道路ネットワークデータ²⁾を利用した。Fig.1に示した道路ネットワークは、神戸市の都心部とその周辺地域の道をすべて含んでおり、ある道からほかの道へ到達する経路の存在が保証されている。また、道路ネットワーク上の道は、高速道路、国道、その他(県道と市道)の区別が可能で、高速道路は法定速度 100 km/h、国道は法定速度 60 km/h、県道と市道は法定速度 30 km/h と定めた。各道の配置と法定速度の関係を Fig.1 に示す。

本研究の目的を達成するために、実データに合ったシミュレーション設定の利用は好ましいが、神戸市の自動車交通シミュレーションにおいて、道路交通センサスデータと一致する交通流を発生させることが困難であることが報告されている²⁾。しかし、完全にランダムな自動車走行台数や移動経路

Fig. 1 Road map of Kobe-city, Japan is drawn. 35952 roads and 14438 intersections are included. Red circles represent traffic right. High-speed ways with 100 km/h legal speed, national routes with 60 km/h legal speed and Other ways with 30 km/h legal speed are drawn with red, green and black lines respectively

Fig. 2 Road map is divided into 6 areas. Area 1, 2, 3 and 4 are divided so that degree of longitude is divided by the same width as each other. The center of city lies both on area 2 and area 3. Origin-Destination(OD)s from area 1 to 4 and from area 4 to 1 are had by vehicles which pass the center of city. Area 5 and area 6 which are related with background traffic are divided by 555.70 m from top in considering the cluster of road network. The coordinate data of each area is shown in Table 1

の設定からは、意図せぬ渋滞が発生し、過大な旅行時間を要する恐れがある。そのため、ここでは簡易的に、大域的な自動車交通の特長についてのみ実データに合わせることを考えた。神戸市都心部の自動車交通として報告された特徴¹⁰⁾は、

- 都心に流入する自動車交通の7割を通過交通が占めている(1日に通過する10万台の内7万台は通過交通である)。
- 小規模な駐車場が散在しており、うろつき自動車交通が発生している。

の2点であり、これらの条件を満たす人工的な自動車交通を発生させた。

Fig. 3 Plot of the number of running vehicles on the map in each second is shown. Since each vehicle is assigned a more appropriate route as the number of iterations increases, all vehicle can reach their destinations without stopping for a long time by a traffic jam

Table 1 Coordinate data of areas

area	x range (m)	y range (m)
all	[0.00, 11469.73]	[0.00, 9308.76]
area1	[9385.50, 11469.80]	[1659.30, 6722.00]
area2	[7332.80, 9385.50]	[2785.35, 9308.80]
area3	[5264.30, 7332.80]	[4055.20, 9308.80]
area4	[1387.10, 5264.30]	[6188.00, 9308.80]
area5	[0.00, 4900.50]	[0.00, 5557.70]
area6	[0.00, 1387.10]	[5557.70, 9308.80]

第1に、自動車のOrigin-Destination (OD) 分布が条件を満たすように、神戸都心とその周辺地域を **Fig. 2** のように6つのareaに分割した。また、ある道の始点があるareaに含まれるとき、その道はそのareaに属すると定めた。Fig. 2の道路地図は、上方向が北向きで、北西の角を原点とし東向きをx軸方向、南向きをy軸方向として、x座標の範囲は[0, 11469.73] (m)であり、y座標の範囲は[0, 9308.76] (m)である。6つのareaはそれぞれ、**Table 1**に示される範囲の矩形で定義した。areaの大きさや位置関係が揃っていないことは、以下の都心の定義のためである。area1, area2, area3, area4を合わせて都心と定め、area1に属する道から発生しarea4に属する道に到着する自動車とその逆にarea4から発生しarea1へ到着する自動車を合わせて通過交通と定めた。また、ほかのarea間の移動はうろつき自動車交通を含む背景流と考えた。シミュレーション内時間6時間でそれぞれのareaから発生する自動車の台数と到着台数を**Table 2**に定めた。area1からarea4へ移動する8750台およびその逆の8750台が都心を通過する交通であり、合計17500台が6時間で都心を通過する。すなわち24時間では、70000台の自動車が都心を通過することに相当する。ここで、都心の交通に含まれない交通は、area5とarea6とから流出し、area5とarea6とに流入する4通りの交通である。**Table 2**において、都心の交通に含まれない自動車の発生台数は、下線で示される。よって、**Table 2**に示された全組み合わせ36通りの内、

Table 2 OD distributions per 6 hours are described. The total number of OD generations is correspond to 100 thousands per a day. The number of OD generations which are had by vehicles running on the center of city (area 2 or area 3), is correspond to 70 thousands per a day

		Departure					
		area1	area2	area3	area4	area5	area6
Arrive	area1	250	250	250	8750	250	250
	area2	250	250	250	250	250	250
	area3	250	250	250	250	250	250
	area4	8750	250	250	250	250	250
	area5	250	250	250	250	250	250
	area6	250	250	250	250	250	250

都心の通過交通の2通りと都心に含まれない交通の4通りを除いた30通りが、都心の背景流として発生する。都心の背景流は、6時間あたり250台の自動車30通り、合計で7500台発生する。すなわち、24時間では、30000台の自動車が都心の背景流として発生する。なお、都心の背景流に含まれないarea5, area6における交通が存在しないことは、area5, area6と都心間で流入する自動車交通に与える影響からも不自然であると考え、都心の背景流と同様の背景流を設定した。

第2に、area内における自動車発生時間と自動車発生場所は乱数を利用して定めた。発生時刻(秒)は、エリアごとに一様分布 $U[0, 21600]$ からサンプリングした。また、発生場所は、車線数の多い道ほど発生頻度が高いと考え、道の車線数の自乗に比例した確率で道を選択し、その道の始点を発生地点とした。到着場所となる道も、同様に、車線数に比例した確率で決定し、その道の終点を到着地点とした。

第3に、自動車の経路選択戦略を定めた。既存の自動車の経路選択戦略の中には、OD間の最短経路を選択する方法や法定速度を考慮した旅行時間に基づく経路を選択する方法がある。しかし、これらの戦略では、同じ経路に多くの車が集中する可能性があり渋滞が容易に発生する。シミュレーション結果に統計的に妥当な解釈を与える目的を達成するため、渋滞の発生・成長・解消という非定常的な現象が試行ごとに発生していることは望ましくない。よって、試行ごとに渋滞による平均旅行時間の変化が少ないように自動車の経路を定める必要がある。そのような経路を生成する方法として、SUMOのツールとして公開されている、dua.iterate¹¹⁾を採用した。dua.iterateは、自動車交通シミュレーションを繰り返し行ない、より旅行時間の少ない経路を再計算する経路選択戦略である。シミュレーションの各繰り返しにおいて、自動車が道を通るのに要した時間を道の重みとして記憶し、つぎに実行されるシミュレーションでは、再計算された道の重みに従って旅行時間を最小とする経路を求める。シミュレーションを繰り返すことにより、ほかのすべての車の旅行時間を延ばすことなく、旅行時間を最小化する経路が得られ、平均旅行時間のばらつきは小さく抑えられる。

第4に、SUMOでシミュレーションを実行するとき、これまで示したほかにも、信号の現示パターンや車の追従モデル

Fig. 4 Box plots of number of vehicles on each road are drawn. On the top figure, roads on which the average values are larger than 500 are included. On the bottom left figure, roads on which the top 20 of averaged values are included. On the bottom right figure, roads on which the worst 20 of averaged values are included

のパラメータを変更することが可能である。しかし、現実のデータの取得が困難なことから、パラメータの調整が困難なことから、ここでは、SUMOの規定値を利用した。

2.1 交通シミュレーション結果

Fig. 3に、Table 2に従って発生させたOD分布における走行中の自動車台数の時間推移を示した。図中の線の色は、それぞれdua.iterateの繰り返し回数に対応している。シミュレーションが開始された直後の数百秒間は、地図上に自動車が点在しはじめ、目的地に到着する自動車も存在しないため、走行中の自動車数は単調に増加した。dua.iterateの繰り返し数0回目に相当する黒の線は、道の長さや法定速度から求めた旅行時間を最短にする経路を自動車が行き通る場合である。dua.iterateの繰り返し数0回目では、走行中の自動車台数が単調に増加していることから、渋滞が発生し自動車が目的地に到達するのに過大な時間を要していることがわかる。dua.iterateの繰り返し数が増えるにつれ、走行中の自動車台数は、約1800台でより長い時間定常化した。また、走行中の自動車台数は、シミュレーション開始後、約2000秒から定常化した。走行中の自動車台数が定常化していることは、自動車が地図上に発生する台数と目的地に到着し地図から消える台数が時間によらず同数程度であり、平均旅行時間のばらつきが小さく抑えられたことを意味している。dua.iterateの繰り返し数を増加させると計算コストも増加するので、ここでは、すべての試行で繰り返し数6回と定めた。得られた人工的な交通は、朝の通勤時間の渋滞の成長のような現実の交通シナリオとは異なるが、データマイニングのデータとしては渋滞のバイアスを考慮する必要がなく扱いやすい。

エリア間の移動台数は等しいが試行ごとにODの場所および時間の分布が異なるシミュレーションを160試行実施した。**Fig. 4**は160試行から得られた各道における通過台数の箱ひげ図を平均交通量の多い順に並べた図である。なお、通過交

通が極端に少ない道が多数存在することから、平均通過交通量が500台以上の道2097本のみを表示している。Fig. 4より、平均通過台数の多い道ほど分散が大きい傾向があった。また、分散の大きさは、同程度の平均通過台数の道でも大きく異なる。よって、観測点ごとに等しい分散でないことからだけでも、道の通過台数の比較から都市規模の自動車交通シミュレーションの再現性を評価するのは困難である。その上、観測点ごとの位置関係が通過交通量に与えるバイアスを考慮することは、より困難である。

3. 因子分析

2章で行なった自動車交通シミュレーション160試行における自動車通過台数のうち平均値が500台以上の道の自動車通過台数に因子分析を適用し、本論文の目的であるシミュレーション結果の解釈を行なった。

因子分析⁷⁾は、 D 次元の確率変数を $X \in \mathbb{R}^{D \times 1}$ 、データの母平均を $\bar{X} \in \mathbb{R}^{D \times 1}$ 、 n 個の因子中 i 番目の共通因子を λ_i 、 i 番目の因子負荷量を $V_i \in \mathbb{R}^d$ 、 i 番目の独自因子を ϵ_i として、(1)式のように定義される。

$$X - \bar{X} = \begin{bmatrix} V_1 V_2 & \dots & V_n \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \\ = V\Lambda + \epsilon \quad (1)$$

ここで、 X の相関行列 R を考えると、(2)式に示すように、共通因子 Λ と独自因子 ϵ は分散 $SD(X)^2$ によって標準化され、相関行列は因子負荷量と標準化された独自因子の和となる。

$$R = \frac{E[(X - \bar{X})(X - \bar{X})^T]}{[SD(X)][SD(X)]^T} \\ = V \left[\frac{\Lambda}{SD(X)} \right] \left[\frac{\Lambda}{SD(X)} \right]^T V^T \\ + \left[\frac{\epsilon}{SD(X)} \right] \left[\frac{\epsilon}{SD(X)} \right]^T \\ = VV^T + \epsilon'\epsilon'^T \quad (2)$$

すなわち、因子分析では相関行列 R を、よりランクの低い因子負荷量 V で近似する。この近似により、確率変数への寄与が因子負荷量の対応するベクトル $V_1, V_2, \dots, V_n$ として各因子ごとに数値化される。

因子分析で因子負荷量を求める手順として、共通因子の算出法決定、因子数の特定、因子負荷量の回転が順に行なわれる。共通因子の決定法として最小二乗法を、因子数の特定法として平行分析を、因子負荷量の回転方法としてバリマックス回転をそれぞれ採用した。なお、各手順において、統計ソフト“R version 3.1.3”¹²⁾とライブラリ“psych version 1.4.8”¹³⁾による解析方法に準拠する。

Fig. 5 Scree plot is drawn. Eigenvalues of simulation results are described with circle markers and eigenvalues of randomly sampled data are described with triangle markers

最小二乗法により、相関行列と近似値との残差を最小とする共通性を得た。共通性の初期値としては、重相関係数の二乗である SMC を用いることが一般的であるが、データ数に変数の次元数より少なく、また、強い相関をもつ次元が存在することから相関行列が計算上ランク落ちするため、相関行列の逆行列から求められる SMC の代わりに初期値 0.5 を用いた。

平行分析は、元のデータと同じ数の変数、および、元のデータと同じ数のサンプル数の正規乱数より得た分散共分散行列の固有値を求め、乱数列から得られた固有値より因子分析で得られた因子の固有値が大きい因子までを因子数と定める方法である¹⁴⁾。平行分析により、乱数列からさえ得られるような因子を取り除くことが期待される。平行分析の結果として、Fig. 5 に、因子分析で得られた固有値と正規乱数の相関行列から求めた固有値の比較を示す。Fig. 5 より、33 個の因子を統計的に有意な因子と見なした。

バリマックス回転は、因子負荷量の分散を最大化する回転を行なう。すなわち、バリマックス回転により、ある因子に寄与する変数がより明確に定まる。よって、自動車交通の因子分析においては、変数に対応する道がどの因子によって解釈されるかをより明確に示せると考える。なお、自動車交通において因子間の関係性も無視できない要素かもしれないが、ここでは因子間の無相関を仮定し、直交回転のバリマックス回転を利用した。

3.1 分析結果

因子分析により得られた結果は、自動車交通シミュレーション結果を説明するモデルであり、モデルは因子数に相当する自由度をもつ。得られた因子数は 33 であることから、2097 本の道を表わす観測可能変数が 33 変数で統計的に有意に表現できることが示された。各因子の寄与率を、それぞれ Table 3 の 10 万台/日の列に示した。因子の寄与率は、それぞれ因子の固有値の大きさに対応している。Table 3 より、累計因子寄与率は 69.0%であった。すなわち、シミュレーション結果の平均からの差を考慮したときに、その差の 69.0%までが 33 個の一次独立な変数の線形結合によって説明された。

3.2 因子分析結果のロバスト性

得られた 33 の関係性は、尤もらしい設定のシミュレーショ

Table 3 Contribution rates of factor analyses are shown. The rate of found robust factors for variance of total number of vehicles are described as bold font. Such values are larger values than the other. The number of found factors (33) are similar to each other even if factor analysis is applied to data from randomly generating OD distributions. In random OD row, the average contribution rates are higher than the others

	100,000 vehicles/day 6 areas	80,000 vehicles/day 6 areas	120,000 vehicles/day 6 areas	100,000 vehicles/day Random OD
#	Contribution Rate			
1	10.3%	7.8%	4.9%	3.9%
2	4.8%	5.7%	4.4%	3.8%
3	2.8%	3.1%	3.7%	3.4%
4	2.7%	2.9%	3.6%	3.4%
5	2.7%	2.6%	3.3%	3.3%
6	2.6%	2.6%	2.9%	3.2%
7	2.5%	2.5%	2.8%	3.1%
8	2.3%	2.3%	2.6%	3.1%
9	2.3%	2.3%	2.2%	2.6%
10	2.3%	2.2%	2.1%	2.6%
11	2.2%	2.2%	2.0%	2.6%
12	2.0%	2.1%	2.0%	2.4%
13	1.9%	2.0%	1.9%	2.4%
14	1.9%	2.0%	1.8%	2.4%
15	1.8%	2.0%	1.6%	2.3%
16	1.7%	1.9%	1.6%	2.3%
17	1.7%	1.8%	1.6%	2.3%
18	1.6%	1.8%	1.6%	2.2%
19	1.5%	1.7%	1.4%	2.2%
20	1.5%	1.7%	1.4%	2.1%
21	1.4%	1.6%	1.4%	2.0%
22	1.3%	1.6%	1.4%	1.9%
23	1.3%	1.5%	1.4%	1.8%
24	1.3%	1.5%	1.4%	1.7%
25	1.2%	1.5%	1.3%	1.6%
26	1.2%	1.4%	1.3%	1.6%
27	1.2%	1.3%	1.3%	1.5%
28	1.2%	1.3%	1.3%	1.4%
29	1.2%	1.3%	1.2%	1.4%
30	1.2%	1.3%	1.2%	1.3%
31	1.1%	1.3%	1.2%	1.3%
32	1.1%	1.2%	1.1%	1.2%
33	1.0%	1.1%	1.1%	1.1%
34	—	—	1.1%	1.1%
	Total			
	69.0%	71.2	67.3%	76.4

ン結果から得られたものである。しかし、得られた関係性がロバストであるかは不明である。得られた関係性のロバスト性を検証するため、異なるシミュレーション設定の関係性と比較した。

地図と SUMO シミュレータの規定値を除き、関係性に影響を与える可能性があるシミュレーション設定は、自動車の発生台数 (25000 台/6 時間)、発生エリア分割方法 (6 area)、分析に利用する道の選択方法 (平均 500 台以上)、分析に利用する試行数 (160 試行) である。自動車の発生台数は、現実では日によって変動する設定であり、発生台数の変化は道の平均

通過台数に影響する。そのため、因子分析結果も影響を受ける可能性がある。検証のため、規定値の 10 万台/日より 2 割増減した発生台数で因子の比較を行なった。さらに、発生エリアの分割方法の変化は、車の移動経路に影響する。そのため、道の平均移動台数も変化し、因子分析結果も影響を受ける可能性がある。検証のため、エリア分割をせず全域の道を対象としてランダムに OD を設定し因子の比較を行なった。分析に利用する道の数や分析に利用する試行数は、統計的に有意な因子の区別に影響を与える。そのため、多くの試行データを利用した因子分析では、異なる因子が発見される可能性がある。しかし、多くの試行データを利用した因子分析は、多大な計算時間や必要メモリー容量の増加が見込まれる。したがって、ここではより多くの道を選択した分析や試行数を増やした分析は行わず、考察を示すのみとした。よって、追加実験は、発生台数を増減した 2 通りと、地図全域に対してランダム OD を発生させた 1 通りの合計 3 通りである。

Fig. 6 に平均 500 台以上通過した道と因子分析により通過交通量がモデル化された道をそれぞれ示した。ここで、自動車交通のシミュレーションでは、因子負荷量の要素の値の絶対値が十分に大きいことは、対応する因子によって道の自動車通過台数の平均からの差が良く説明されていることを意味する。因子負荷量の要素値の範囲は $[-1.0, 1.0]$ であり、絶対値が 0.3 より大きな要素である道を因子がモデル化した道だと考えた。Fig. 6 より、平均 500 台以上通過した道の多くが因子分析により通過交通量をモデル化されていた。発生する台数が多くなるに連れ、平均 500 台以上通過交通がある道は増えるが、増えた道もなにかしらの因子によって説明された。また、ランダム OD の設定より得られた平均 500 台以上通過した道の図は、OD の発生エリアを指定し発生台数を変えた設定から得られた図と明らかに異なった。

Table 3 に摂動を与えたシミュレーション結果から得られた因子数と寄与率をそれぞれ示した。自動車の発生台数の変化やランダム OD によって、因子数は大きく変化しないことがわかった。しかし、発生台数を変化させた因子寄与率では、寄与率の特に大きな因子が存在するのに対し、ランダム OD の因子寄与率では、寄与率の特に大きな因子が存在しなかった。そこで、特に寄与率の大きな因子に着目し、10 万台/日の発生台数から得られた因子 No. 1 と因子 No. 2 がモデル化した道の地図上での分布を比較した。

Fig. 7 に、因子 No. 1 と、発生台数を 2 割減したときに得られた因子 No. 1 と発生台数を 2 割増したときに得られた因子 No. 1 とが説明する道を示した地図を示した。さらに、**Fig. 8** に、因子 No. 2 と、発生台数を 2 割減したときに得られた因子 No. 1 と発生台数を 2 割増したときに得られた因子 No. 4 とが説明する道を示した地図をそれぞれ示した。これらの地図の右側には、説明する道間の相関行列をカラーマップで示した図もそれぞれ示した。相関行列の行および列に含まれる道は、ほかの道と正の相関をもつ頻度が多いものから順に並んでいる。また、青色の道は、相関行列の第 1 列、第

Fig. 6 A Comparison of found roads by applying factor analysis. 4 left plots are drawn with roads on which 500 or more vehicles are passed, and 4 right plots are drawn with found roads by applying factor analysis. Where, from top to bottom, visualization of results of 100 thousands per a day, 80 thousands per a day, 120 thousands per a day and 10 thousands per a day with random OD are drawn respectively. Though roads on which vehicles passed increase as the number of generated vehicles get larger, almost of such road are explained by results of factor analyses. On the other hand, in random OD, distributions of roads on which 500 or more vehicles passed is even different from the others

1 行に対応する道と通過交通量において正の相関を示す道である。一方、赤色の道は、相関行列の第 1 列、第 1 行に対応する道と通過交通量において負の相関を示す道である。相関行列より、因子が説明する道のグループは、その相関関係によって、さらに 2 種類のグループに分割できる。分割されたグループの道路は、一方が混めば他方が空くような関係になっている。

Fig. 7 と Fig. 8 より、自動車台数の増加により説明する道が増えたことが示されたにもかかわらず、因子寄与率の特に大きな因子がモデル化した道の分布の変化は小さいことが示された。異なる設定で、共通の因子が発見されたことは、そ

の設定の変更に対して因子がロバストであることを意味する。このことより、車の発生台数に対してロバストな寄与率の高い因子が存在したといえる。しかし、寄与率が高くても、車の発生台数を変化させることにより、説明する道が異なる因子も存在した。たとえば、発生台数を 2 割増したときに得られた第 2 因子や第 3 因子は、寄与率が高いが、因子が説明する道の分布は、ほかの因子分析結果には存在しなかった。このことより、寄与率が高い因子がロバストであるかは不明である。一方、Fig. 7 と Fig. 8 で示された以外の因子の組み合わせでは、同様の分布を発見できなかった。このことより、因子寄与率の比較的低い因子は、シミュレーション結果から統計的に有意であるとして発見されたが、ランダム OD の設定でも見つかるような、実在しない関係性であると考えられる。例として、**Fig. 9** に、第 22 因子と第 26 因子がモデル化した道の地図上での分布を示す。

分析に利用する道の選択方法 (平均 500 台以上)、分析に利用する試行数 (160 試行) についての追加実験は実施しなかったが、それぞれの変化に対して発見される因子への影響を考察する。まず、分析に利用する道の数を変更した場合、同様の因子分析結果が得られることが予測される。その根拠として、自動車発生台数を増減したときに、平均 500 台以上通過する道の分布はランダム OD の結果ほど変わらなかった。そのため、より少ない台数以上通過した道を含めたデータも似たような道の分布をすることが予測され、因子分析でもロバストな因子が発見されると期待される。つぎに、試行数を増やすことにより、より多くの因子で擬似データがモデル化されることが予測される。すると、発見されたロバストな因子は複数の因子に分割される可能性があり、ロバストな因子の発見のために複数の因子を組み合わせる必要があると予測される。これまでの検証では、地図上の道の分布を比較してロバスト性を検証したが、より多数の因子の比較は、なにかしらの指標を導入し比較しなければ実現できないと予測される。よって、計算コストの面のみならず、ロバスト性の検証においても、現状では試行数を増やすことはできない。

これらの検証結果より、得られた知見はつぎの点である。

- 都市規模の自動車交通シミュレーション結果には、統計的に有意な複数の関係性が存在した。
- シミュレーション設定の摂動に対してロバストかつ因子寄与率が特に大きな因子が存在した。
- 因子寄与率が比較的低い因子は、エリアを考慮しないランダム OD で得られた見せかけの関係性と区別できなかった。都市の自動車交通の実データが得られたとして、実データから考慮すべき関係性を得るためには、統計的に有意な関係性を発見するだけでなく、ランダムな設定で発見されるような見せかけの関係性を無視する必要があると考えられる。データに含まれる見せかけの関係性を特定するために、社会実験でランダム設定のデータを得ることは困難であり、シミュレーションの活用が期待される。

Fig. 7 Road distributions of similar results, factor No. 1 in 100 thousands per a day (Top), factor No. 1 in 80 thousands per a day (Middle) and factor No. 1 in 120 thousands per a day (Bottom) are drawn

3.3 得られた因子の利用について

本論文では、多変量であるシミュレーション結果の解釈のために因子分析を適用し、得られた因子の性質を明らかにするまでを目的とした。そのため、シミュレーション結果が現実の交通現象の再現として妥当かどうかの検証は行っていない。しかし、得られた因子の有用性について議論することは、重要であると考え。そこで、以後の考察では、シミュレーション結果が現実の交通を再現したと仮定して、因子の利用方法に関する考察を行なった。

データ同化における推定精度向上や工事の影響推定における予測精度向上の文脈において、得られたロバストな因子のバイアスを考慮できることは、推定精度や予測精度の向上につながる。なぜなら、ある変数のばらつきが因子の影響の重ね合わせとより小さなばらつきのノイズとしてモデル化されるため、データとの誤差や予測値のばらつきを小さく見積もることができる。また、道路工事による地図の改変など、因子が説明する道の分布が大きく変わるような摂動が与えられたとき、因子が説明する道の分布の変化から工事が影響する道や範囲を特定することができる。さらに、実データを得るための観測地点や観測頻度を、因子からデザインすることが可能となる。

一方、シミュレーション結果から得られた知見を説明するとき、因子を利用した説明は概要ではあるが自由度が低く受

Fig. 8 Road distributions of similar results, factor No. 2 in 100 thousands per a day (Top), factor No. 2 in 80 thousands per a day (Middle) and factor No. 4 in 120 thousands per a day (Bottom) are drawn

Fig. 9 Road distributions of unsure results, factor No. 22 in 100 thousands per a day (Top) and factor No. 26 in 100 thousands per a day (Bottom) are drawn

け入れやすい。たとえば、得られた因子を現実の因子に対応づけるため、因子の区別を行なった。区別には、道の種別として、高速道路、国道、県道、市道の区別を、道の進行方向として、東西方向の区別を、道の場所として、areaの区別をそれぞれ利用した。Table 4に因子の区別結果を示した。Table 4

より、特に寄与率の高い因子でありロバストな第 1 因子と第 2 因子で説明される道は進行方向が区別できる因子であった。この結果は、神戸市は、北を山に、南を海に囲まれており、都心を通過するような長距離の自動車移動は東西方向に制限されるためだと考えられる。また、ロバストな因子は、比較的エリア全域に道が分布しており、かつ、高速道路や国道、県道を多く含んでいた。そのため、この因子が説明している道は互いに影響し合う神戸市の主要な自動車交通だと考えることができる。よって、道路工事などによる影響は、工事の実施地域だけでなく、広く伝搬することが因子を利用して説明できる。一方、因子寄与率の比較的低い因子には、道の進行方向では区別できないなど、神戸市のどのような交通因子か判断が困難なものも存在した。たとえば、第 22 因子は、因子で説明される道が地図上で離れて点在しておりどのような交通因子かを説明することが困難であった。また、第 24, 26, 31, 32 因子は、因子がモデル化した道の数少なく、どのような交通因子かを説明することが困難であった。第 22 因子と第 26 因子における道の地図上での位置とその関係性は Fig. 9 にそれぞれ示した。これらの因子は、ランダム OD から得られる因子と区別できない因子であったため、実在する可能性はあるかもしれないが実証は困難である。

4. ま と め

本論文では、都市規模の交通シミュレーション結果に対して統計的な解釈を与えるため、多変量解析である因子分析を適用した。そして、神戸市の道路地図を利用した自動車交通シミュレーション結果より、2097 の観測可能な変数が 33 の因子で統計的に説明された。しかし、得られた因子の中には、自動車発生台数の増減を摂動として加えることにより発見できなくなる因子が存在した。よって、統計的に有意だけでなく、ロバストな因子の発見が重要であることを示した。統計的に有意だけでなく、ロバストな因子は、シミュレーション結果を解釈するモデルとして利用できる。

本論文では、一般的な因子分析手法を適用したが、多変量データのモデル化手法は、多様なものが提案されている。どのようなモデル化手法が交通現象の関係性を説明するのによりふさわしいかを知ることが重要であり、さらなるデータ分析が期待される。

発見されたロバストな因子は、実際の交通現象との比較によりその妥当性が検証されるが、妥当性の検証は、交通シミュレーションの解釈方法を提案する本論文の目的を超えている。なお、妥当性の検証のためには、道路構造の変化や信号の位置やタイミングの変化など、より多様な条件での検討はもちろん、ほかの都市における結果についても検討することが必要である。そのため、因子分析の結果に関しては、検証を行わず考察のみとした。因子の利用方法の考察より、因子寄与率が高くロバストな因子は、実データの収集コスト削減や比較精度の向上にも貢献すると期待される。また、発見された因子は、神戸市の主要な交通である可能性を示した。自動

Table 4 Attributes of road groups are distinguished. Attributes include a kind of roads, a direction of roads and areas in which the found roads belong. Kinds of roads include in high-speed way (HS), national routes (NR), prefectural road (PR) or city road (CR). Directions of roads are to east (E) or to west (W). Areas are shown in Fig. 2

#	kind	direction	area
1	HS, NR, PR	E	area1,2,3,4
2	HS, PR	W	area1,2,3,4 north
3	HS, NR	W	area2,3 middle
4	NR, PR	E	area4 north
5	HS, NR	E	area3,4 north
6	CR	E	area4 north & middle
7	NR, PR	W	area1,2 north
8	PR	W	area3,4 north
9	CR	W	area3,4 north
10	CR	W	area4 middle
11	HS, PR	E	area1,2 middle & south
12	CR	E	area4 north
13	PR, CR	E	area2 north
14	HS, NR	W	area2 south
15	HS	E	area2 south
16	PR	E	area2,3 middle
17	CR	E	area4 north
18	CR	W(E)	area1(4) north
19	NR	E	area2 south
20	CR	E	area3 middle
21	CR	W	area2 middle
22	CR	—	area4 middle
23	NR, CR	E	area1 south
24	CR	—	area3 north & south
25	CR	W	area1 south
26	CR	—	area3,4 north
27	NR	—	from area5 to south
28	CR	W	area4 middle
29	NR, CR	E	area1 middle & south
30	NR	—	from south to area5
31	CR	—	area3 north & south
32	CR	—	area3 north
			area4 north & middle
33	NR	E	area3,4 south

車交通シミュレーションの影響を説明するときに、このような因子による簡約化した説明は、多くの説明を受けるより受け入れやすい。

最後に、本研究におけるシミュレーションの 1 試行には、Intel Xeon E5-1620 3.60 GHz で約 3 時間必要とした。また、R による因子分析は、同環境で 160 試行に対し約 2 時間、かつ、約 8 GB のメモリーを必要とした。より大規模な地図を対象とした交通シミュレーションやより多くのエージェントを含むシミュレーションの実施、また、より精度の高い結果を得るために試行数を増やすことで、ロバストな関係性を得るための計算コストは爆発的に増える。このことより、都心だけでなく県や全国の規模を考慮したシミュレーション結果の分析を実現するためには、より多くの並列実行が可能でメモリー搭載量の多いハイパフォーマンスコンピュータやスーパーコンピュータの利用が不可欠となるだろう。

謝辞 本研究は JST CREST の支援を受けたものである。本研究では、都市規模の交通シミュレーションをさまざまなシナリオで正しく実行し、それぞれの実験結果に因子分析を適用する必要があった。シミュレーションの分散実行と結果分析の補助ソフトに、シミュレーション実行・管理ソフト *OACIS*¹⁵⁾ を利用した。ここに *OACIS* の開発者らに謝意を表する。

参 考 文 献

- 1) Japan Society of Traffic Engineers : 交通シミュレーション活用のススメ, 丸善出版 (2012)
- 2) Y. Asano, N. Ito, H. Inaoka, T. Imai and T. Uchitane: Traffic Simulation of Kobe-city, Proceedings of the International Conference on Social Modeling and Simulation, plus Econophysics Colloquium 2014, 255/264 (2015)
- 3) M.T. Asif, J. Dauwels, C.Y. Goh, A. Oran, E. Fathi, M. Xu, M.M. Dhanya, N. Mitrovic and P. Jaillet: Spatiotemporal Patterns in Large-Scale Traffic Speed Prediction, IEEE Transactions on Intelligent Transportation Systems, **15**-2, 794/804 (2014)
- 4) 玉本, 堀口, 桑原: 交通シミュレーション SOUND の大規模ネットワークへの適用, 第 30 回土木計画学研究発表会 (秋大会) 講演論文集, 290 (2004)
- 5) 佐藤, 堀口, 桑原: 大規模ネットワークにおける動的シミュレーション適用の現在とこれから, 第 29 回土木計画学研究発表会講演集, 62 (2004)
- 6) 内種, 伊藤: 交通シミュレーションによる交通解析, 第 21 回社会における AI 研究会, 講演原稿, 3 (2014)
- 7) 浅野長一郎: 因子分析法通論, 共立出版 (1972)
- 8) OpenStreetMap: <http://www.openstreetmap.org>
- 9) D. Krajzewicz, J. Erdmann, M. Behrisch and L. Bieker: Recent Development and Applications of SUMO — Simulation of Urban MObility, International Journal on Advances in Systems and Measurements, **5**-3&4, 128/138 (2012)
- 10) 神戸市都市計画総局: 「都市の概況と人の移動実態」および「想定される課題」, 神戸市総合交通計画検討委員会第 1 回委員会 (平成 24 年 8 月 7 日開催) 資料 4, http://www.city.kobe.lg.jp/information/project/urban/sogokotsu/commission_sogokotsu.html (2012)
- 11) D. Krajzewicz, M. Behrisch, J. Erdmann and Y.-P. Floetteroed: Demand/Dynamic User Assignment, http://sumo.dlr.de/wiki/Demand/Dynamic.User_Assignment (2008)
- 12) R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, <http://www.R-project.org/> (2015)
- 13) W. Revelle: psych: Procedures for Personality and Psychological Research, <http://CRAN.R-project.org/package=psych> (2015)
- 14) H. John: A rationale and test for the number of factors in factor analysis, Psychometrika, **30**-2, 179/185 (1965)
- 15) Y. Murase, T. Uchitane and N. Ito: A tool for parameter-space explorations, Physics Procedia, **57**, 73/76 (2014)

[著 者 紹 介]

内 種 岳 詞 (正会員)

大阪大学で、学士、修士、博士をそれぞれ 2008, 2010, 2012 年に取得。現職は、理化学研究所計算科学研究機構研究部門特別研究員。「京」コンピュータによる離散事象シミュレーション研究に従事。IEEE, 日本知能情報ファジィ学会, 進化計算学会の会員。

伊 藤 伸 泰

東京大学で、理学博士を 1991 年に取得。現職は、東京大学准教授、理化学研究所計算科学研究機構研究部門チームリーダー。日本物理学会の会員。