

# 不完全データを用いた GGM による交通量推計手法に関する研究

鈴木 裕司<sup>1</sup>・柳原 正実<sup>2</sup>・小根山 裕之<sup>3</sup>

<sup>1</sup> 非会員 首都大学東京大学院 都市環境科学研究科 (〒192-0397 東京都八王子市南大沢 1-1)

<sup>2</sup> 正会員 首都大学東京大学院 都市環境科学研究科 助教 (〒192-0397 東京都八王子市南大沢 1-1)  
E-mail: yanagihara@tmu.ac.jp

<sup>3</sup> 正会員 首都大学東京大学院 都市環境科学研究科 教授 (〒192-0397 東京都八王子市南大沢 1-1)  
E-mail: oneyama@tmu.ac.jp

道路ネットワークの管理・運営や道路交通の円滑化に資する施策の実施に当たり、道路ネットワーク全体にわたって交通量を的確に把握することが必要である。近年利活用が進められているプローブカーシステムは低コストで広域な情報収集が可能であるが、その混入率が低い現状では情報が収集できない道路区間や時間帯が存在する。そこで本研究では、ネットワーク全体の交通量を的確に把握することを目的とし、これまでに提案されている Gaussian Graphical Model 型の推計モデルの運用に関して、欠損を許容する交通量データセットを効率的に利用できる交通量推計手法を構築した。特にパラメータの推定段階では固有値分解を用いた閉形式の効率的な計算手法を提案する。また、シミュレーションを用いた適用実験を行いモデルのパフォーマンスを検証した。

**Key Words:** traffic volume, Gaussian Graphical model, Bayes' theorem, EM algorithm

## 1. はじめに

道路ネットワークの管理・運営や道路交通の円滑化に資する施策の実施に当たり、道路ネットワーク全体にわたって交通量を的確に把握することが必要である。交通量の計測には車両感知器が用いられるが、設置・管理にコストがかかるため、車両感知器が設置される箇所は自ずと限られ、設置されない区間の交通量は計測できない。一方、近年利活用が進められているプローブカーシステムは低コストで広域な情報の収集が可能である。しかし、車両自体がセンサーであるプローブカーは時空間的に散在するため、データ数に時空間的なばらつきが生じるという特性を持つ。

既存の研究<sup>1)</sup>では、画像補修フィルタ技術であるガウシアングラフィカルモデル<sup>2)</sup> (Gaussian Graphical Model : GGM) を応用して交通量を補間推定する手法が提案されている。まず GGM とは、多変量正規分布で確率的相関構造を表現したものである。GGM を応用して設置箇所が限られる車両感知器による計測交通量から、車両感知器が設置されない区間の交通量を推計する手法が提案されているが、この手法の GGM のパラメータ推定過程では、車両感知器が設置されている区間と設置されてい

ない区間の関係性を求めるために、全ての道路区間のプローブデータを用いている。一方で、実際の時空間的に散在するプローブデータを効率的に活用するには、欠損を許容したデータセットからパラメータ推定が可能な手法の構築が必要である。

欠損を許容したデータセットを利用する方法として、原ら<sup>3)</sup> は GGM の構造に各道路区間の速度分布を対応させたモデルを提案している。しかし、交通量保存則などの法則をモデルで表現するためには、交通量を確率変数として扱うモデルを構築する必要がある。また、瀬尾ら<sup>4)</sup>によってプローブカーデータより交通量を算出する手法も提案されており、広域の交通量の観測値を得ることも将来的には可能であると考えられる。

本研究では、道路ネットワーク全体の交通量を的確に把握・補完することを目的とし、GGM 型の推計モデルの運用に関して、欠損を許容する交通量データセットを効率的に利用できる交通量推計手法を構築する。

本モデルの運用においては、大きく分けて学習ステップと推計ステップの 2 ステップが存在する。まず学習ステップでは、欠損を含む交通量データセットを利用した GGM のパラメータ推計を行う。推計ステップでは、学習ステップで推定したパラメータと観測道路の交通量を

用いて未観測道路の交通量を推計する。特にパラメータの推定段階では固有値分解を用いた閉式形の効率的な計算手法を提案する。

本稿の構成は以下の通りである。第 2 章では、本研究で用いるモデルの定式化過程を示す。続く第 3 章では、シミュレーションを用いた本モデルの適用実験手法と結果を示す。最後に第 4 章では、結論と今後の課題について述べる。

## 2. モデルの定式化

### (1) ネットワークの設定

まず、一般的な道路ネットワークは、道路の結節点にノードをおき、ノードとノードで結ばれる区間をリンクとする。一方、本研究で用いるネットワークの構造は、結節点に接続される道路の流入地点、流出地点にそれぞれノードをおき、ネットワークの中で車両の方向別交通量を明示的に表現できるようにした。一般的な道路ネットワークの構造を図-1 に示し、方向別交通量を表現する道路ネットワークの構造を図-2 に示す。

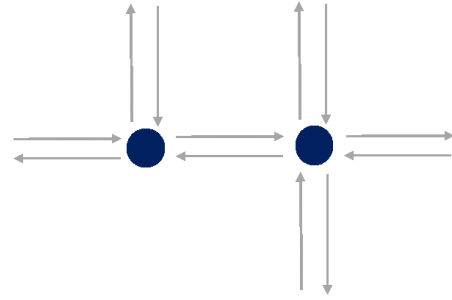


図-1 一般的な道路ネットワーク

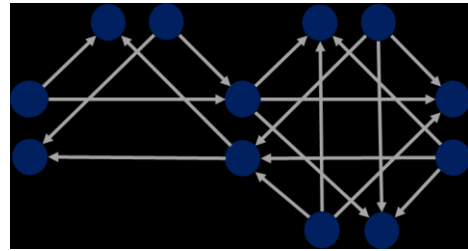


図-2 GGMのネットワーク構造

### (2) 交通量保存則を考慮した GGM 型確率密度関数

本研究では、交通量保存則を考慮した GGM 型の確率密度関数を利用する。当該関数は式(1)のように定義される。ただし、 $d$  時点における各道路の交通量を  $z_{d,i}$  とし、道路  $i$  の集合を  $V$  とする。  $J_{in}$  はある交差点に流入する道路の集合、  $J_{out}$  は流出する道路の集合とする。  $\beta_i$  は各道路の交通量の平均に関するパラメータであり、  $\xi_i$  は各道路の交通量の分散に関するパラメータ、  $C$  は規格化定数である。

$$p(z_d | \beta, \xi) = C \exp \left\{ \sum_{i \in V} \beta_i z_{d,i} - \frac{1}{2} \sum_{i \in V} \xi_i z_{d,i}^2 - \frac{1}{2} \sum_{(i,j) \in J} \left( \sum_{i \in J_{in}} z_{d,i} - \sum_{j \in J_{out}} z_{d,j} \right)^2 \right\} \quad (1)$$

### (3) パラメータ推計

学習ステップでは、GGM 型の確率密度関数を用いた EM アルゴリズム<sup>6)</sup>(Expectation - Maximization algorithm)によるモデルパラメータの最尤推定を行う。EM アルゴリズムとは、不完全データに基づく統計モデル一般に適用される最尤推定導出のためのアルゴリズムであり、対数尤度関数の期待値を計算するステップを E ステップ、これを最大化するステップを M ステップとする。EM アルゴリズムでは、不完全データを広義にとらえるこ

とで、欠損データの問題に限らず、切断や打ち切りデータ、有限混合分布モデル、ロバスト分布モデルや潜在変数モデル、ベイジモデルと、その適用範囲が広いことが特徴である。まず、利用する交通量データセットとして、一部の道路に設置された車両感知器及び時空間的に散在するプローブカーから推計された交通量を想定するため、 $d$  時点における各道路の交通量の完全データ  $z_d$  は車両感知器及びプローブカーの有無に対応して、道路状況が観測されていない道路の交通量データ  $x_d$  と観測されている道路の交通量データ  $y_d$  にかけて定義される。不完全データセット  $y_d$  が与えられた下での未観測道路の交通量の事後確率は式(2)のように定義する。

$$p(x_d | y_d, \beta, \xi) = C \exp \left\{ \left( \sum_{i \in U^d} \beta_i x_{d,i} + \sum_{i \in O^d} \beta_i y_{d,i} \right) - \frac{1}{2} \left( \sum_{i \in U^d} \xi_i x_{d,i}^2 + \sum_{i \in O^d} \xi_i y_{d,i}^2 \right) - \frac{1}{2} \sum_{(i,j) \in J} \left( \sum_{i \in J_{in}} z_{d,i} - \sum_{j \in J_{out}} z_{d,j} \right)^2 \right\} \quad (2)$$

式(1)及び式(2)の確率を用いて、EM アルゴリズムにおける E ステップで利用する条件付き対数尤度関数の期待値は式(3)のように表現できる。

$$\begin{aligned}
 & Q(\boldsymbol{\beta}, \boldsymbol{\xi} | \boldsymbol{\beta}^{old}, \boldsymbol{\xi}^{old}) \\
 &= \sum_d \int_{x_d} p(x_d | y_d, \boldsymbol{\beta}^{old}, \boldsymbol{\xi}^{old}) \ln p(z_d | \boldsymbol{\beta}, \boldsymbol{\xi}) dx_d \quad (3) \\
 &= \sum_d I_d
 \end{aligned}$$

式(3)の積分  $I_d$  は高次元の積分であり、数値積分の計算コストが非常に大きい。したがって本研究では式(2)の計算効率を上げるため、固有値分解を利用し導いた積分形のない閉形式(closed-form)を利用する。まず、式(3)の中の積分  $I_d$  は既存研究に従って式(4)の形に変形することができる。ただし、 $a_d^{(*)}$ 、ベクトル  $\mathbf{b}_d^{(*)}$ 、行列  $C_d^{(*)}$  は積分変数  $\mathbf{x}$  に対して定数であり、特に行列  $C_d^{(*)}$  は実対象行列であるものとする。

$$I_d = \int_{\mathbf{x}} \left\{ \left( a_d^{(1)} + \mathbf{x}^T \mathbf{b}_d^{(1)} - \mathbf{x}^T C_d^{(1)} \mathbf{x} \right) \times \exp \left( a_d^{(2)} + \mathbf{x}^T \mathbf{b}_d^{(2)} - \mathbf{x}^T C_d^{(2)} \mathbf{x} \right) \right\} d\mathbf{x} \quad (4)$$

式(4)の積分変数を式(5)のように置換し、式(6)を得る。

$$\mathbf{x}' - \frac{1}{2} C_d^{(2)-1} \mathbf{b}_d^{(2)} = \mathbf{x} \quad (5)$$

$$I_d = \int_{\mathbf{x}'} \left\{ \left( a_d^{(3)} + \mathbf{x}'^T \mathbf{b}_d^{(3)} - \mathbf{x}'^T C_d^{(3)} \mathbf{x}' \right) \times \exp \left( a_d^{(4)} - \mathbf{x}'^T C_d^{(2)} \mathbf{x}' \right) \right\} d\mathbf{x}' \quad (6)$$

ここで、実対象行列  $C_d^{(2)}$  を式(7a)~式(7c)のように固有値分解し、積分変数を  $\mathbf{x}'' = P_d \mathbf{x}'$  と置くことで、ガウス積分系の積と和で表現された式(8)を得る。ただし、 $\lambda_{d,i}$  は固有値、 $P_d$  は固有ベクトルの集合である行列を表す。

$$C_d^{(2)} = P_d^T A_d P_d \quad (7a)$$

$$P_d^{-1} = P_d^T, \quad |P_d| = 1 \quad (7b)$$

$$\{A_d\}_{ij} = \begin{cases} \lambda_{d,i} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (7c)$$

$$I_d = \int_{\mathbf{x}''} \left\{ \left( a_d^{(3)} + \mathbf{x}''^T \mathbf{b}_d^{(3)} - \mathbf{x}''^T C_d^{(3)} \mathbf{x}'' \right) \times \exp \left( a_d^{(4)} - \mathbf{x}''^T A_d \mathbf{x}'' \right) \right\} d\mathbf{x}'' \quad (8)$$

さらに、式(8)に対して式(9a)、式(9b)のガウス積分の公式を適用することで式(10)のような閉形式の式を得ることができる。

$$\int_{-\infty}^{+\infty} x^{2n} e^{-ax^2} dx = \left( \frac{1}{2a} \right)^n \sqrt{\frac{\pi}{a}} \quad (9a)$$

$$\int_{-\infty}^{+\infty} x^{2n+1} e^{-ax^2} dx = 0 \quad (9b)$$

$$I_d = \exp(a_d^{(4)}) \left( \prod_i \sqrt{\frac{\pi}{\lambda_{d,i}}} \right) \left( a_d^{(3)} + \sum_i \frac{\{C_d^{(3)}\}_{ii}}{2\lambda_{d,i}} \right) \quad (10)$$

実際の数値計算では、 $a_d^{(*)}$ 、ベクトル  $\mathbf{b}_d^{(*)}$ 、行列  $C_d^{(*)}$  と固有値  $\lambda_{d,i}$  を求めたのち式(10)を計算する。これらの変数は観測値と推定パラメータに依存する変数である。

次に、M ステップでは、対数尤度関数の期待値を最大化するパラメータを推定する。その推定値は以下の通りに定義される。

$$\boldsymbol{\beta}^{est}, \boldsymbol{\xi}^{est} = \arg \max_{\boldsymbol{\beta}, \boldsymbol{\xi}} Q(\boldsymbol{\beta}, \boldsymbol{\xi} | \boldsymbol{\beta}^{old}, \boldsymbol{\xi}^{old}) \quad (11)$$

このときに対数尤度の期待値が収束しているかを調べ、収束していなければ、

$$\begin{aligned}
 \boldsymbol{\beta}^{est} &\rightarrow \boldsymbol{\beta}^{old} \\
 \boldsymbol{\xi}^{est} &\rightarrow \boldsymbol{\xi}^{old}
 \end{aligned} \quad (12)$$

として E ステップに戻る。これを繰り返し行いパラメータを推定する。これを学習ステップとする。なお、本研究では、パラメータ群  $\boldsymbol{\beta}$  及び  $\boldsymbol{\xi}$  の最小二乗和が十分に小さくなった場合を EM アルゴリズムの収束条件とした。

#### (4) 未観測道路の交通量推計

推計ステップでは、既存研究等<sup>7)8)9)10)</sup>に従い、交通量保存則を考慮した GGM の確率密度関数を事前分布とする。観測交通量として一部の道路に交通量を与えたときの尤度を観測交通量の誤差が限りなく小さいと仮定し、ディラックのデルタ関数を利用して、式(13)のように設定する。

$$p(\mathbf{y} | \mathbf{x}) = \prod_{i \in O_d} \delta(y_i - x_i) \quad (13)$$

ここで、道路状況が観測されている道路の交通量を  $y_{d,i}$  とし、それらの道路の集合を  $O_d$  とする。また、道路状況が観測されていない道路の交通量を  $x_{d,i}$  とし、それらの道路の集合を  $U_d$  とする。ベイズ推論により

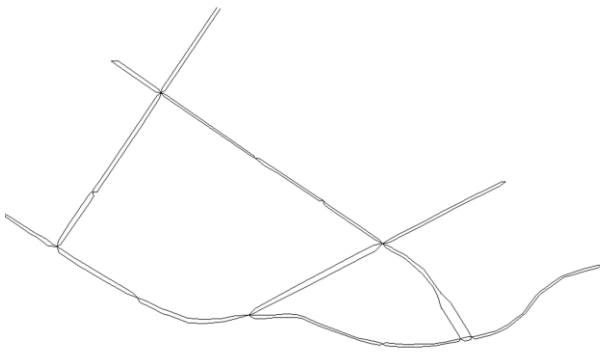


図-3 仮想ネットワーク

表-4 推定パラメータセットと利用データ

推定パラメータセット	推定に利用したデータセットの欠損率
I	0%
II	5%
III	10%

未観測道路の交通量の事後確率分布は式(14)のように定義される。

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\xi})}{p(\mathbf{y})} \propto p(\mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\xi}) \prod_{i \in O_d} \delta(y_i - x_i) \quad (14)$$

未観測道路の交通量の事後確率を式(15)、式(16)で示すガウス・ザイデル法により最大化し、未観測道路の交通量を推計する。

$$x_n^{(t+1)} \leftarrow \frac{1}{C_{ij}} \left( \beta_i - \sum_{m=1}^{n-1} C_{nm} z_m^{(t+1)} - \sum_{m=n+1}^{|V|} C_{nm} z_m^{(t)} \right) \quad (15)$$

$$z_m^{(t)} := \begin{cases} y_m & m \in O \\ x_m^{(t)} & m \in U \end{cases} \quad (16)$$

### 3. シミュレーション実験手法と結果

#### (1) 検証用データの生成方法

本モデルの検証実験には交通流シミュレーションによって生成されたデータセットを用いる。利用した仮想ネットワークは図-3に示す道路リンク数が40、交差点数が5で構成される。まず11の交通需要パターンを作成し、その11パターンをランダムに組み合わせ100日分の仮想

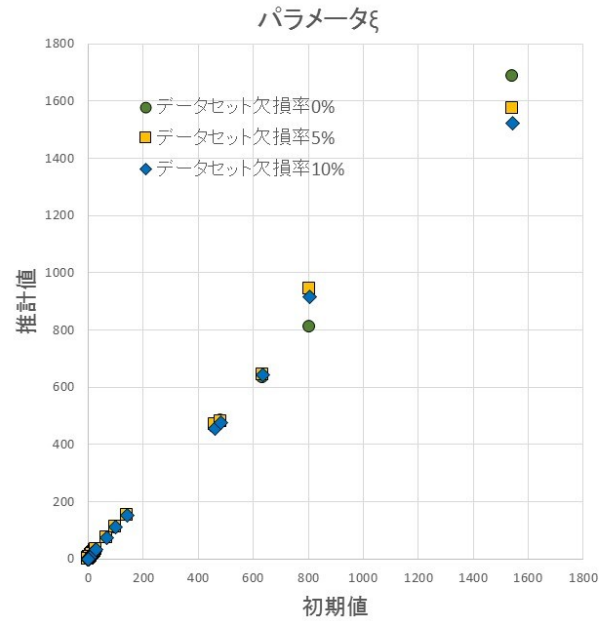


図-5 パラメータ  $\xi$  の初期値と各推定パラメータセットの比較

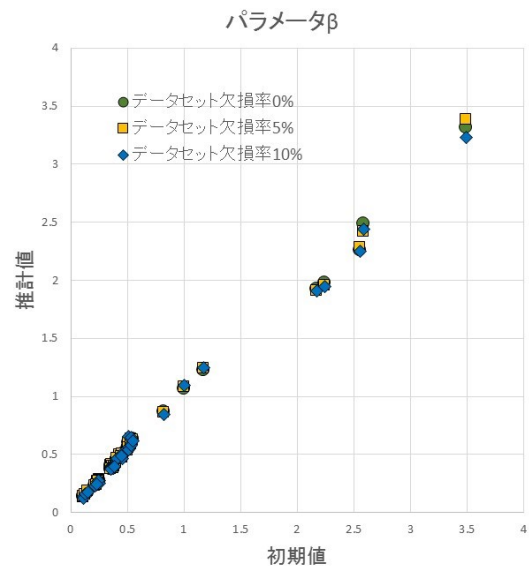


図-6 パラメータ  $\beta$  の初期値と各推定パラメータセットの比較

交通状況についてのシミュレーション結果を生成した。

#### (2) パラメータ推計

学習ステップでは、表-4に示すデータセットの欠損率が0%、5%、10%の3ケースについてパラメータ推計を行った。各時点での欠損する道路は、空間的に散在するプローブカーの特性を考慮してランダムに設定した。

また、学習ステップにおける対数尤度関数の期待値の最大化では、与えた初期値への依存性が強いいため、本検証実験ではデータセットの平均値と不変分散にGGMのそれらが合致するようなパラメータの値を用いた。さら



に交通状況がデータセットの平均値と大きく異なるケースにおいて、欠損率が大きいデータに対する最尤推定計算の収束が不安定になったため、本研究では欠損率が大きすぎるデータを学習データから除いている。この学習ステップにおける最尤推定には6時間以上の計算時間を要した。計算の効率化を行わない場合はこの場合よりも遥かに長い計算時間が必要となるため計算の効率化は必要不可欠であるといえる。

図-5はパラメータ  $\xi$  に関して初期値と各推定パラメータセットを比較したものである。図-6は、パラメータ  $\beta$  に関して図-5と同様に示したものである。これらは、EMアルゴリズムを用いた学習ステップでのパラメータ推計過程で、パラメータ群が初期値からどれだけ推移したものを示す。図-7は、パラメータ  $\xi$  に関して推定パラメータセット I と推定パラメータセット II 及び III を比較したものである。図-8は、パラメータ  $\beta$  に関して図-7と同様に示したものである。これらの図より、学習ステップに欠損を含む不完全なデータセットを使用した場合の推定値が、欠損を含まない完全なデータセットを使用した場合の推定値と比較してどれくらい差が生じるかを確認する。

図-5より交通量が多い傾向にある道路ほどパラメータの初期値と推定値に大きな差が生じ、初期値よりも大きな値が推計されたことがわかる。図-6からは、交通量が多い傾向にある道路では推定値が初期値より小さく推計され、交通量が少ない傾向にある傾向にある道路では推定値が初期値より幾分大きな値として推計されたことがわかる。図-7からは、交通量が多い傾向にある道路では、データセットの欠損率を変化させることで推計値のばらつきが大きくなり、交通量が多い傾向にある道路では、データセットの欠損率を変化させることでの推計値のばらつきは小さくなりことが見てとれる。図-8では、交通量の多い道路の一部で欠損率が大きくなることで推計値が大きくなるものがあるが、その他多くのパラメータは初期値と同等の値であった。以上より、データセットに欠損が生じた場合においても、学習ステップにおけるパラメータ推定では、欠損が存在しないデータセットによる推定値と大きな差異が生じないことがあることが確認された。

(3) 未観測道路の交通量推計

次に推計ステップで推計される交通量の検証として、5ケースの交通状況において、学習ステップのデータセット欠損率0%, 5%, 10%に場合の検証を行った。各推定パラメータを用いて、推計ステップでの欠損率20%(40本中8本)を未観測道路とし、未観測道路の交通量推を推計した。図-9は、ある1ケースの交通状況において、各

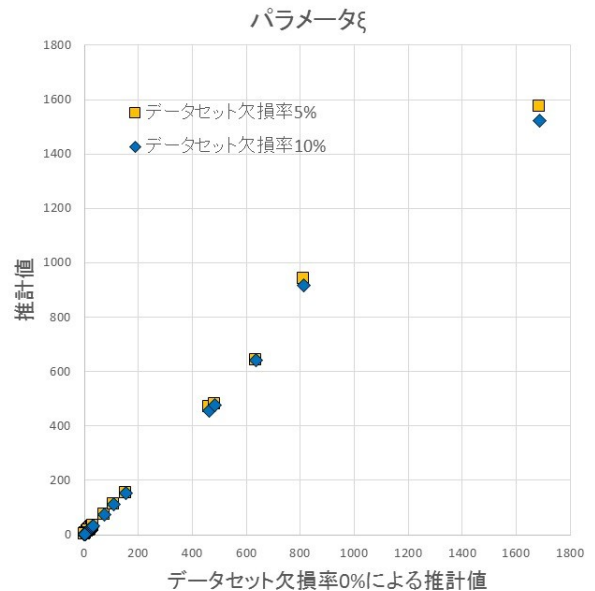


図-7 推定パラメータセット I と推定パラメータセット II, III の  $\xi$  比較

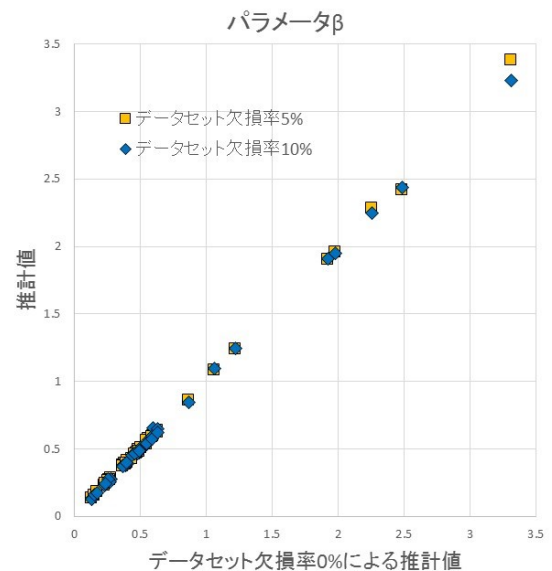


図-8 推定パラメータセット I と推定パラメータセット II, III の  $\beta$  比較

パラメータによる未観測道路の推計値と真値を比較したものである。また図-10は、学習ステップのデータセット欠損率10%の場合で、推計されたパラメータを用いて、推計ステップで交通状況の異なる5ケースを設定した際の未観測道路の推計交通量と真値を比較したものである。ここで、5ケースの内、ケースAはネットワーク全体での交通量が最も大きく、ケースB,C,D,Eの準にネットワーク全体での交通量が小さいものとする。

図-9での比較結果及び、各推定パラメータセット I ~ III による推計値のRMSEは、それぞれ19.85, 19.95, 20.70

[veh/h]であることから、学習ステップにおけるデータセットの欠損の有無によって、未観測道路の交通量推計結果に大きな差が生じないことが確認された。

また、ケースAでは未観測道路の推計交通量が過小評価となり、ケースD及びケースEでは推計結果が過大評価となった。これは、ケースAの交通量の平均値が5.4%推計データセットに比べ大きく、逆にケースD、Eの交通量の平均値は、それぞれ5.6%、15.7%小さいことが要因として考えられる。ケースA及びケースEでは、推計交通量の交通量保存則が十分に満たされていなかったことから、学習ステップにおける交通量保存則の制約を与えるパラメータの推定精度が低かったと考えられる。

#### 4. まとめ

本研究ではデータの欠損を許容した上で、精度の高い結果を得られる交通量保存則を考慮した交通量推計モデルを構築した。特に現実的な計算時間で計算が終了するよう効率的な計算式を提案した。しかし、学習ステップにおける最尤推定には6時間以上の計算時間を要したため、今後はより広域のネットワークへの適用に向けて、更なる計算の効率化と現実的に利用できるネットワークの大きさを確認してゆく必要がある。

また、本研究では収束計算を安定させるため学習ステップにおいて交通量保存則の制約の強さを与えるパラメータを固定して計算しており、このことがモデルの精度を下げる要因となっていることが示唆された。したがって今後は当該パラメータの適切な推定方法を構築することが望まれる。

#### 参考文献

- 1) 花岡洋平, 安田宗樹, 桑原雅夫, 原祐輔: ガウシアングラフィカルモデルを用いたネットワーク交通状況の補間推定に関する研究, 土木計画学研究・講演集, 2013.
- 2) 田中和之: 確率モデルによる画像処理技術入門, 森北出版, 2006.
- 3) 原祐輔, 花岡洋平, 桑原雅夫: 道路ネットワーク内の関係性に着目した長期観測プローブデータによるプローブ未観測リンクの交通状態補間, 交通工学論文集, Vol. 2, No. 1, pp. 1-10, 2016.
- 4) 瀬尾亨, 日下部貴彦, 朝倉康夫: 車間距離を計測するプローブカーを前提とした交通状態の推定手法, 土木学会論文集, 2013.
- 5) 瀬尾亨, 日下部貴彦, 朝倉康夫: 車両軌跡に基づく流率密度関係の推定法—基本的な枠組みと数値実験, 第35回交通工学研究会発表会論文集, pp. 271-278, 2015.
- 6) Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm, Journal of

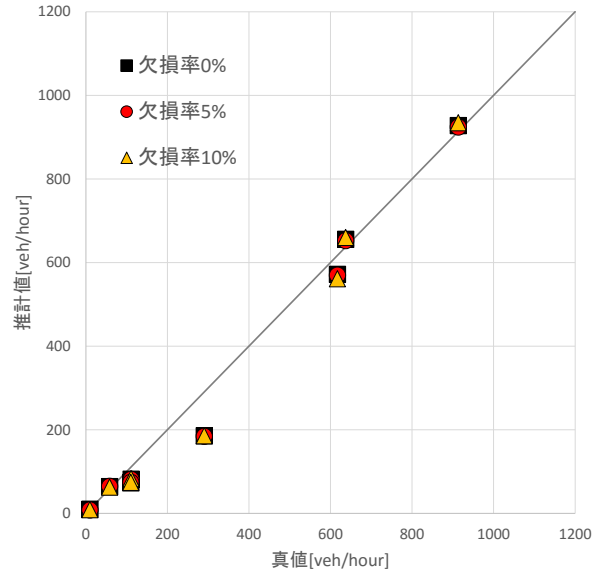


図-9 各推定パラメータセットによる未観測道路の真値と推計値の比較

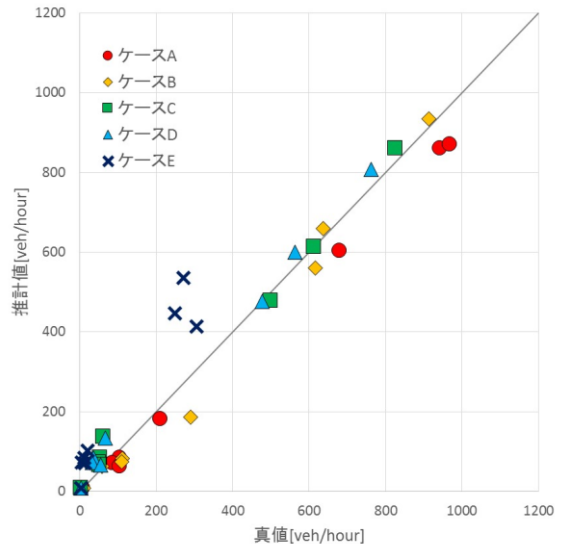


図-10 推定パラメータセットIIIによる各ケースの未観測道路の真値と推計値の比較

the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1, pp.1-38, 1977.

- 7) 三輪富生, 山本俊行, 竹下知範, 森川高行: プローブカーの速度情報を用いた動的 OD 交通量の推定可能性に関する研究, 土木学会論文集, 2007.
- 8) 片岡駿, 安田宗樹, 田中和之: マルコフネットを用いた交通量の統計的予測, 日本物理学会講演概要集, 2012.
- 9) 豊田秀樹: マルコフ連鎖モンテカルロ法, 朝倉書店, 2008.
- 10) 久保拓弥: データ解析のための統計モデリング入門—一般化線形モデル・階層ベイズモデル・MCMC (確率と情報の科学), 岩波書店, 2012.

(2016.4.22 受付)

TRAFFIC STATE INTERPOLATION OF UNOBSERVED ROAD LINKS BY  
GAUSSIAN GRAPHICAL MODEL

Yuji SUZUKI, Masami YANAGIHARA and Hiroyuki ONEYAMA