

IV-8 予測誤差による交通需要予測プロセスの検討

北海道大学 学生員 ○ 桐越 信
北海道大学 正 員 山形 耕一
北海道大学 正 員 五十嵐日出夫

1 はじめに

都市交通計画では、交通需要予測の最も一般的プロセスとして、発生交通量と集中交通量の予測—分布交通量の予測—交通機関別交通量の予測—配分交通量の予測という総合的交通需要予測プロセスを採用している。総合的交通需要予測プロセスでは数多くの数学的モデルが用いられるが、モデルによる予測値には、その操作の諸段階において様々な誤差が介入する。これらの誤差とその発生段階により分類してみると、現象解析の誤り、説明変数の省略、代理変数の採用、線形化等の構造の簡略化等のモデル設定段階での誤差、パラメータの推定法や調査データの誤差等のパラメータ推定段階での誤差、さらにモデルに取り入れられた説明変数の予測誤差等の予測段階での誤差の3つに大別される。総合的交通需要予測プロセスでは、これらの誤差が発生、集中から配分に至る各段階で生じると共に各段階の予測値は次の段階の入力となり誤差が以降の予測段階に伝播する。このことが交通需要予測の予測誤差の正確な把握を困難なものにしている。そこで本研究では、交通需要予測の予測誤差に影響を与える要因をいくつか取り上げ、これらが注目する予測量にどのように影響するかを分析し、予測誤差による交通需要予測プロセスの検討を試みた。

2 予測誤差の伝播の理論的検討

本研究では、予測誤差への影響要因として、①モデルの精度、②モデルに取り入れられた説明変数の予測誤差を取り上げた。予測誤差とそれへの影響要因①、②との関係を図1に示す。ここで m_1 は第1段階の真の予測モデルを示し、 m_2 は第2段階の真の予測モデルを示す。 $\hat{m}_1$ 、 $\hat{m}_2$ はそれぞれの段階での推定された予測モデルである。第1段階において、外的基準 y_0 のモデルによる期待値と $m_1(x_0)$ 、 y_0 と $m_1(x_0)$ との差 u_0 をモデルに取り込まれなかった要因の偶然的影響と考え、

$$y_0 = m_1(x_0) + u_0 \quad (1)$$

となる。(1)式の関係より $m_1(x_0)$ の推定モデル $\hat{m}_1(x)$ が得られる。予測対象を y_0 とし説明変数の予測誤差を Δx_0 とすると推定モデル $\hat{m}_1(x)$ による予測誤差 ω_1^* は

$$\omega_1^* = m_1(x_0) + u_0 - \hat{m}_1(x_0 + \Delta x_0) \quad (2)$$

となるので、その分散は

$$E(\omega_1^{*2}) = E\{[m_1(x_0) + u_0 - \hat{m}_1(x_0 + \Delta x_0)]^2\} \quad (3)$$

となる。第2段階において y_0 をモデルに取り込まれなかった要因による偶然的影響と考え、

$$z_0 = m_2(y_0) + v_0 \quad (4)$$

によって $m_2(y)$ の推定モデル $\hat{m}_2(y)$ が得られ、予測対象を z_0 とすると推定モデル $\hat{m}_2(y)$ による予測誤差 ω_2^* は、第1段階の出力が第2段階の入力となるので

$$\omega_2^* = m_2(y_0) + v_0 - \hat{m}_2(y_0^*) = m_2(m_1(x_0) + u_0) + v_0 - \hat{m}_2(\hat{m}_1(x_0 + \Delta x_0)) \quad (5)$$

となる。その分散は、

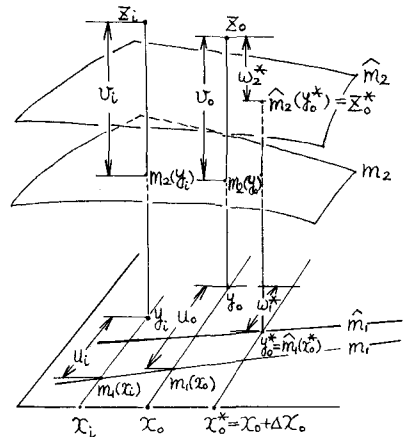


図1 予測対象と予測誤差

$$E[\omega_2^{**}] = E[\{m_2(m_1(x_0) + u_0) + v_0 - \hat{m}_2(\hat{m}_1(x_0 + \Delta x_0))\}^2] \quad (6)$$

となる。ここでモデル m_1 として重回帰モデル $G = \beta_0 + \sum_{k=1}^p \beta_k x_k$ を、モデル m_2 としてグラビティモデル $T = \alpha_0 G^{\alpha_1} A^{\alpha_2} D^{\alpha_3}$ を採用し、 u_i , v_i の分散とそれぞれ σ_u^2 , σ_v^2 とすると、

$$E[\omega_1^{**}] = \sigma_u^2 + \sum_{k=1}^p V[\Delta x_k] \beta_k^2 \quad (7)$$

$$E[\omega_2^{**}] = \sigma_v^2 + V[\Delta G] \left(\frac{\partial T}{\partial G}\right)^2 + V[\Delta A] \left(\frac{\partial T}{\partial A}\right)^2 \quad (8)$$

となるが、 $V[\Delta G]$, $V[\Delta A]$ はともに (7) 式で表わせるので、

$$V[\Delta G] = \sigma_{u_1}^2 + \sum_{k=1}^p V[\Delta x_{k1}] \beta_{k1}^2, \quad V[\Delta A] = \sigma_{u_2}^2 + \sum_{k=2}^p V[\Delta x_{k2}] \beta_{k2}^2$$

として、(8) 式に代入すると

$$E[\omega_2^{**}] = \sigma_v^2 + (\sigma_{u_1}^2 + \sum_{k=1}^p V[\Delta x_{k1}] \beta_{k1}^2) \left(\frac{\partial T}{\partial G}\right)^2 + (\sigma_{u_2}^2 + \sum_{k=2}^p V[\Delta x_{k2}] \beta_{k2}^2) \left(\frac{\partial T}{\partial A}\right)^2 \quad (9)$$

となる。(9) 式は、第1段階の予測誤差を含めた形での第2段階までの予測誤差を表わしており、 σ_v^2 , $\sigma_{u_1}^2$, $\sigma_{u_2}^2$, $\sum_{k=1}^p V[\Delta x_{k1}] \beta_{k1}^2$, $\sum_{k=2}^p V[\Delta x_{k2}] \beta_{k2}^2$ が予測誤差の影響要因に対応している。

3 シミュレーションによる予測誤差分析

本研究では、交通需要予測プロセスのうち、「発生交通量の予測と集中交通量の予測—分布交通量の予測」の部分を検討の対象とし、前節で示された①モデルの精度、②モデルに取り込まれた説明変数の予測誤差についてシミュレーションによって分析した。

3.1 モデルの精度

モデルの精度は、発生交通モデル、集中交通モデルにおいては、モデルに取り込む説明変数の差異による適合性の違いにより表現し、分布交通モデルにおいては、パラメータの推定方法の違いによる適合性の違いで表現した。発生交通モデル、集中交通モデルには重回帰モデルを用いモデルに取り込む説明変数を1~4変数として、4種類のモデルを検討の対象とした。分布交通モデルにはグラビティモデルを用い、パラメータの推定を対数変換により線形化して推定する方法と非線形回帰法で推定する方法との2通りの方法で行い予測誤差への影響を検討した。モデルの精度を示すのが、フローチャートの①の発生交通モデル、集中交通モデルの適合性、②の分布交通モデルの適合性である。フローチャートの③は、発生交通、集中交通モデルの出力が分布交通モデルの入力として用いられ、誤差の伝播を含む複合誤差を検討するものである。

3.2 モデルに取り込まれた説明変数の予測誤差

説明変数の予測誤差の影響をみるために、説明変数 X に一定水準の誤差を与えた説明変数の予測値 W を作成し、 W を入力したときの予測値 $\hat{Y}$ と原データ Y_i との間の一致性の検討を行った。フローチャートの④の発生交通、集中交通モデルの予測性は、発生交通量および集中交通量の予測段階での精度であり、⑤の分布交通量の予測性は、発生交通量、集中交通量の予測段階での誤差の伝播を含む分布交通量の予測段階における複合誤差を示す。一定水準の誤差を与えた説明変数の予測値 W は次のようにして作成した。

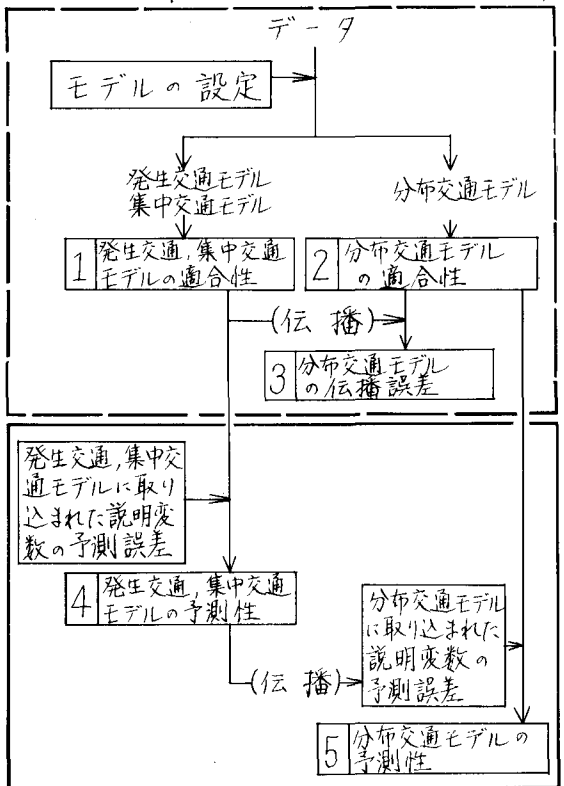


図2 予測誤差分析のフローチャート

$$W = X \times (1.0 + RSD(X) \times R(0, 1)) \text{ ————— (10)}$$

ここで、 $RSD(X)$ は X に与えた相対誤差であり、 $R(0, 1)$ は $N(0, 1)$ の正規乱数である。予測誤差分析における予測値 $\hat{y}_i$ と原データ y_i との一様性は、

$$C = 1.0 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \text{ ————— (11)}$$

なる一様性指標によって表わした。使用したデータは昭和47年道央圏パーソントリップ調査のデータである。

4 予測誤差分析の分析結果

4.1 発生交通量の予測誤差分析

重回帰分析によって作成された発生交通モデルを表1である。目的は全目的である。当然のことであるが、集中交通モデルについてもパラメータ、寄与率と同一様の結果を得ている。次に発生交通モデルに取り込まれた説明変数の予測誤差の影響を表2に示す。ここで、④の発生交通量の予測性については、いずれのCASEにおいても説明変数の予測誤差 $RSD(X)$ が大きくなるに従い予測精度は低下するが、CASE3、CASE4においては、この傾向が甚しく

表1 発生交通モデル

CASE 1	${}_1G_1 = 21633.63 + 3.04 X_{11}$	$R^2 = 0.889$
CASE 2	${}_2G_1 = 952.35 + 2.86 X_{11} + 1.20 X_{21}$	$R^2 = 0.977$
CASE 3	${}_3G_1 = -179.21 + 3.08 X_{11} + 5.00 X_{21}$ $- 8.21 X_{31}$	$R^2 = 0.993$
CASE 4	${}_4G_1 = 711.39 + 3.40 X_{11} + 4.67 X_{21}$ $- 7.28 X_{31} - 2.49 X_{41}$	$R^2 = 0.995$

X_1 : 従業人口総数, X_2 : 夜間人口, X_3 : 就業人口, X_4 : 第2次従業人口

表2 説明変数が予測誤差をもつ場合
発生交通量の予測誤差分析

$RSD(X)$ が15%を越え、CASE1、CASE2よりも予測精度は低下する。このことは、複雑なモデルほど、モデルに取り込まれた説明変数の予測誤差に敏感であり、誤差の水準が低いよりは高い予測精度を維持できるが、予測誤差の水準が高くなると予測精度は著しく悪化する傾向にある。F検定をベースとした重回帰分析の変数選択法のひとつである変数増減法による変数選択では、変数4個を取り込んだCASE4が最良の発生交通モデルとなるが、この例に示されたごとく、多く変数を取り込んだ複雑なモデルは、操作性や論理性において予測モデルとして不適確となり易いばかりでなく、予測性においても不適確となる傾向にある。以上のことから、発生交通モデルとしては、CASE2のモデルが操作性、論理性、予測性のうえから適当である。

$RSD(X)$	CASE 1	CASE 2	CASE 3	CASE 4
0.025	0.888	0.975	0.988	0.990
0.050	0.886	0.974	0.977	0.982
0.075	0.880	0.968	0.951	0.942
0.100	0.875	0.966	0.914	0.918
0.125	0.871	0.965	0.876	0.896
0.150	0.871	0.955	0.827	0.850
0.175	0.866	0.948	0.769	0.783
0.200	0.851	0.944	0.666	0.711
0.225	0.841	0.933	0.545	0.605
0.250	0.835	0.920	0.555	0.599

表3 分布交通モデル

4.2 分布交通量の予測誤差分析

表3は分布交通モデルとして採用したグラビティモデルであるが、(1)は対数変換によって線形化してパラメータを推定した場合であり、(2)は非線形回帰法によってパラメータを推定した場合である。当然のことであるが、非線形回帰法によるほうが寄与率は0.80とかなり高い値となっている。

$$(1) \quad T = 0.03 \times \frac{G^{0.64} A^{0.63}}{D^{1.39}} \quad R^2 = 0.610$$

$$(2) \quad T = 0.00017 \times \frac{G^{0.86} A^{0.86}}{D^{1.17}} \quad R^2 = 0.802$$

T : 分布交通量, G : 発生交通量
 A : 集中交通量, D : 時間距離

表4は発生交通モデル、集中交通モデルの精度と分布交通モデルの精度とを複合した場合の分布交通量の予測における予測精度を表わしている。表4の(1)はグラビティモデルのパラメータを対数変換により線形化して推定した場合に対応し、(2)は非線形回帰法によって推定した場合に対応する。CASE1～CASE4は採用した発生交通モデル、集中交通モデルに対応している。発生交通モデル、集中交通モデルに取り込まれた説明変数の予測誤差がないと考えられる場合には、この一様性指標によって全プロセスの予測誤差が評価される。表4

では、(1)、(2)とともに発生交通モデル、集中交通モデルとして多くの説明変数を取り込んだほうが予測精度は良くなっている。これはモデルに取り込んだ説明変数の予測誤差を考慮していないことによる。表5、表6は発生交通モデル、集中交通モデルの精度、発生交通モデル、集中交通モデルに取り込まれた説明変数の予測誤差、さらには分布交通モデルの精度と複合した場合の分布交通量の予測精度を示している。表5はグラビティモデルのパラメータを対数変換によって線形化して推定した場合であり、表6は非線形回帰法によって推定した場合である。表5、表6における予測精度は本研究で考えたプロセス全体を通しての予測の一致性を示し、最終的評価対象となる。ここでは、説明変数の予測誤差に基き、表2の発生交通量、集中交通量段階での予測誤差が伝播され、説明変数の予測誤差の水準RSD(X)の増加とともに、分布交通量の予測の一致性は低下している。とくに、CASE3、CASE4ではこの傾向が著しく、RSD(X)が15%を越えるとCASE1、CASE2の方が良い一致性を示し、表4の結果とは一致しない。

5 結 論

本研究では、交通需要予測の一連のプロセスのうち、「発生交通量、集中交通量の予測—分布交通量の予測」を取り上げ、予測誤差への影響要因として、①モデルの精度、②モデルに取り込まれた説明変数の予測誤差を考え、これらが注目する予測量にどのように影響するかと理論的に検討し、これをシミュレーションによって分析し、交通需要予測プロセスにおけるモデルの作成、予測作業の精度面での関連性について把握した。その結果について要約すると、

- 1) 予測誤差への影響要因を明らかにし、その相互関係を(1)式で表現した。 $\sigma_u^2 + \sum_{k=1}^p V(\alpha_k) \beta_k^2$ はモデルの複雑さを介して関係があり、一般に互いにトレードオフの関係にある。
- 2) モデルに取り込まれた説明変数の予測誤差は、相対誤差で10~15%となると影響が現われてくる。とくに多くの説明変数を含む場合にその影響が甚しい。このことより予測モデルの簡略化の方向が示唆され、変数の選択も含めて、採用すべきモデルの決定・需要予測の一連のプロセスを通じての予測誤差を考慮してなされるべきであることがわかる。

参考文献

- 1) 山形耕一、相越 信：交通需要予測プロセスの予測精度について、第1回土木計画学研究発表会講演集、1979年
- 2) 太田勝敏：交通需要予測の誤差分析、土木学会第28回年次学術講演概要集第4部、1973年、P133~P134
- 3) 杉恵順子、松岡康弘：交通量推定モデルにおける累積誤差、第13回日本都市計画学会学術研究発表会論文集、1978年、P181~P186

表4 分布交通量の予測における複合誤差

	(1)	(2)
CASE1	0.519	0.646
CASE2	0.565	0.719
CASE3	0.604	0.792
CASE4	0.616	0.806

表5 説明変数が予測誤差をもつ場合の分布交通量の予測における複合誤差 (対数変換)

RSD(X)	CASE1	CASE2	CASE3	CASE4
0.025	0.517	0.562	0.599	0.611
0.050	0.514	0.561	0.595	0.607
0.075	0.515	0.560	0.573	0.589
0.100	0.520	0.559	0.563	0.583
0.125	0.512	0.558	0.527	0.552
0.150	0.513	0.553	0.509	0.537
0.175	0.497	0.536	0.461	0.493
0.200	0.501	0.542	0.438	0.467
0.225	0.497	0.536	0.401	0.440
0.250	0.494	0.518	0.319	0.368

表6 説明変数が予測誤差をもつ場合の分布交通量の予測における複合誤差 (非線形回帰法)

RSD(X)	CASE1	CASE2	CASE3	CASE4
0.025	0.645	0.716	0.785	0.799
0.050	0.643	0.709	0.772	0.788
0.075	0.630	0.703	0.703	0.724
0.100	0.626	0.694	0.661	0.692
0.125	0.615	0.690	0.634	0.670
0.150	0.604	0.679	0.542	0.585
0.175	0.599	0.634	0.466	0.521
0.200	0.575	0.628	0.365	0.424
0.225	0.567	0.618	0.240	0.399
0.250	0.546	0.591	0.231	0.349