

道路環境アセスメントとしての大気汚染の予測

奥谷 巖*

Prediction of Air Pollution for Road Environmental Assesment

Iwao Okutani *

沿道地域の環境アセスメントの一環として、本研究ではNO_xによる大気汚染濃度を統計的手法を用いて予測することを試みた。データは関東、近畿、中国、九州の4都市における10数日間の計測値を用いた。手法的には重回帰分析法、数量化理論I類および一般化された重回帰分析法を採用したが、結論的には説明変数の非線形的影響を考慮しうる数量化理論I類による予測結果が最もよいことが判明した。また、説明変数としては交通量の偏相関係数が最も大きく影響力の大なることをうかがわせるが、気温、雲量をはじめとする他の変数はあまり大きな説明力をもっていないことが明らかとなった。

まえがき

経済の沈滞化傾向の中にあつて、ひとりモータリゼーションは着実な歩みのものに進行しつつあり、幹線道路は慢性的供給不足に悩まされている。これに伴ない沿道地域の車による環境破壊も深刻の度合いを高めており、騒音や大気汚染の問題は社会問題と化しているといえよう。

こうしたことに鑑みて、本研究では沿道地域の環境アセスメントという観点から、車による大気汚染という問題を探り挙げ、交通量をはじめとする道路環境要因から、いかにして大気汚染濃度を的確に予測評価しうるかについて、関東、近畿、中国および九州における4都市の実測データをもとに、実証的な検討を行う。

大気汚染物質として代表的なものは、一酸化炭素、硫黄酸化物および窒素酸化物であるが、前二者がガソリン精製技術の向上あるいは自動車エンジンの改良により、かなりの程度除去されつつあるのに対し、後者の窒素酸化物(NO_x)は依然としてその除去技術が進んでいない。したがって、本研究ではこのNO_x濃度を具体的な対象として分析を進める。

道路環境要因

NO_xによる大気汚染濃度は、道路をとりまくさまざまな環境により影響を受け変化するが、ここでは以下に示すような8つの要因を、具体的なNO_x濃度説明因子として考える。

- | | |
|----------------|------------------------------|
| 1. 交通量(台/時) | 2. X方向の風速(m/秒) |
| 3. Y方向の風速(m/秒) | 4. 気温(°C) |
| 5. 湿度(%) | 6. 日射量(cal/cm ²) |

*信州大学工学部土木工学科 Fac. Eng., Shinshu Univ.

7. 雲量(割)

8. 大気安定度

ここに、X方向というのは道路に直角な方向であり、Y方向というのは、道路に沿う交通の流れの方向である。道路の環境アセスメントという観点からのNO_x濃度の予測は、NO_x濃度そのものの計測データを説明変数

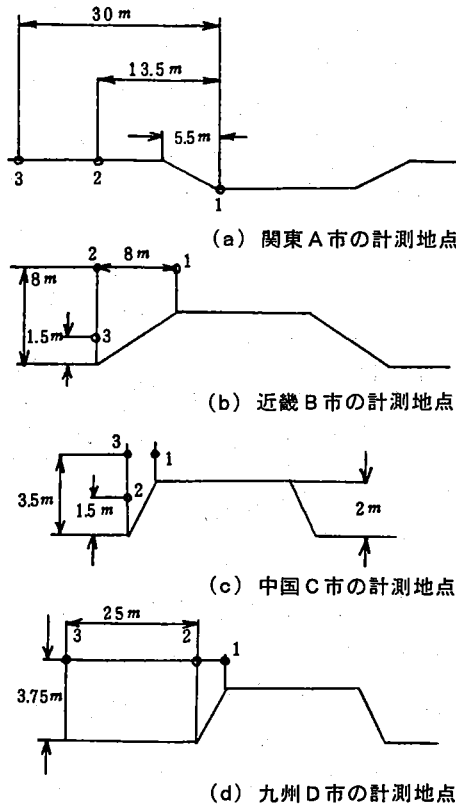


図1. 各地域のNO_x計測地点

とすることなく、それをいかにうまく他の道路環境要因から予測するかということであるが、一般に大気汚染濃度は非常に不安定であり、時々刻々の道路環境によって大きく変動するし、またいかなる環境要因がそうした変動に寄与しているかも明らかではない。

したがって、本研究では各地域で共通して計測可能であり、大気汚染濃度に関与している可能性があると考えられる道路環境要因をできるだけ多く採り挙げるといふ視点から、上記8要因に着目して、統計的分析をおとしてその説明因子としての有意性を検討しようとしたものである。

被説明変数としてのNO_x濃度は、各地域ごとに図1(a)~(d) (道路断面図) に示した1, 2, 3なる3つの計測地点の1時間の平均値(ppb)で与えた。データは、このNO_x濃度を含め、1日24時間分、合計10数日間のデータを準備した。

重回帰分析法による予測

まず最初に、この種の解析には常套的手法である重回帰分析法を用いた予測について調べてみる。

いま、 $y(t)$ を時刻 t におけるNO_x濃度、 $x_i(t)$ を先に示した道路環境要因のうち第 i 番目の変数の時刻 t における値としたとき

$$y(t) = a_0 x(t) + a_1 x(t-1) + \dots + a_r x(t-r) + b + \epsilon(t) \quad (1)$$

なる構造式を考える。ここに、 $x(t)$ は $x_i(t)$ を要素とするベクトル、 a_l は $x(t-l)$ にかかるパラメータベクトル、 b は定数項(スカラー)、 $\epsilon(t)$ は残差である。

本研究で目指しているものは、時々刻々の大気汚染濃度をダイナミックに予測することではなくて、むしろ道路環境アセスメントとしての大気汚染水準の予測評価である。この意味において、式(1)の構造中、 $x(t-1)$ 、 $x(t-2)$ 、 \dots 、 $x(t-r)$ の項は必ずしも妥当ではない。すなわち、評価対象道路におけるNO_x濃度も、その道路がおかれている交通量等の環境条件によつて的確に予測しうるか否かが重要な鍵なのであって、環境アセスメントの立場からは道路環境条件の時系列データを利用することが前提にできない場面が多いからである。

したがって、ここではまず $x(t-l)$ ($x > 0$) を予測構造式に含めたとして、その効果はどの程度期待しうるかをみるために、 r を0~9まで変化させたモデルについて、NO_x濃度の予測精度を対象4地域について調べてみた。 $y(t)$ および $x_1(t) \sim x_8(t)$ のデータは、生データをそのまま使う場合、一次指数平滑化を施した場合、二次指数平滑化を施した場合、同一時刻について3日間の平均値を採った場合の4ケースを考えた。平滑化データ、平均化データを考えたのは、NO_x濃度および8つ

の道路環境要因の時間変動が激しい場合が多く、生データのままであると相互関係がうまく把握できない恐れがあると判断したためである。

なお、予測精度を評価するために、本研究では下記のような6つの指標を考えた。

①単純平均誤差率

これは、予測誤差の絶対値の真値に対する比率の平均値で、いま求める単純誤差率を ϵ 、予測値を $\hat{y}(t)$ 、真値を $y(t)$ としたとき

$$\epsilon = \frac{\sum_t |\hat{y}(t) - y(t)|}{\sum_t y(t)}$$

のように表わされる。

②荷重平均誤差率

真値が大きい場合、上で述べた誤差率がより大きく作用するようにしたもので、いま荷重平均誤差率を δ とすると

$$\delta = \sqrt{\frac{1}{\sum_t y(t)} \sum_t \left(\frac{\hat{y}(t) - y(t)}{y(t)} \right)^2 y(t)}$$

となる。

③最大誤差率

いま最大誤差率を ϵ_{\max} とすると

$$\epsilon_{\max} = \max_t \frac{|\hat{y}(t) - y(t)|}{y(t)}$$

として与えられる。

④RMS誤差

Root Mean Square 誤差の略称で通常最もよく用いられる誤差指標である。この指標は、いまサンプル数を N としたとき

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_t (\hat{y}(t) - y(t))^2} \quad N: \text{サンプル数}$$

のように定義される。

⑤一致係数

H. Theilが与えた不一致係数を1から引いたもので、いま求める一致係数を η とすると

$$\eta = 1 - \frac{(1/N) \sum_t (\hat{y}(t) - y(t))^2}{(1/N) \sum_t \hat{y}(t)^2 + (1/N) \sum_t y(t)^2}$$

のように表わされる。 η は $0 \leq \eta \leq 1$ の範囲にあり、1に近ければ予測精度が高いことがいえる。①~④の指標が、それぞれの数値の大きさに適当に予測精度を判定しなければならなかったのに対し、この一致係数 η は指数化されており、判定が容易であるという利点を有する。

⑥相関係数

いま相関係数を ρ とすると

$$\rho = \frac{(1/N) \sum_t (y(t) - \bar{y}(t)) (\hat{y}(t) - \bar{\hat{y}}(t))}{\sqrt{(1/N) \sum_t (y(t) - \bar{y}(t))^2 - (1/N) \sum_t (\hat{y}(t) - \bar{\hat{y}}(t))^2}}$$

となる。ここに、バーは平均値を表わすものとする。一般に、相関係数は変動の傾向の類似性を表わす指標であって、一致の度合いを計る指標としては、必ずしもよいものではないことに注意する必要がある。

さて、まず生データを用いた場合についての予測精度を地域別に示すと、表1(a)~表1(d)のようになる。表中、比較データというのは真値として用いるNOxデータのことであって、生データと一次平滑化データの2つを考えている。

これらの表から総じて言えることは、rが大きくなると重相関係数は大きくなるものの、予測精度は必ずしも向上しないか、かえって減少するということである。

平滑化データ、平均化データについてもほぼ同様のことがいえ、実際RMS誤差についてその変化の様子をグラフに示すと、図2のようになる。一部を除き、rの増

表1. rによる予測精度の変化

(a) 生データを用いた場合の手算精度 (大気汚染濃度平均値 = 329.127 (関東A市))

| 回帰データ | 比較データ | 単純平均誤差率 | 両重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 真相関係数 |
|-------|-------|---------|---------|-------|---------|-------|-------|--------|-------|
| 0 | 生データ | 0.374 | 0.397 | 1.659 | 109.190 | 0.794 | 0.719 | 66.344 | 0.817 |
| 時点前 | 平滑値 | 0.289 | 0.359 | 1.117 | 106.447 | 0.812 | 0.376 | | |
| 0~1 | 生データ | 0.361 | 0.380 | 1.699 | 106.196 | 0.800 | 0.733 | 36.526 | 0.834 |
| 時点前 | 平滑値 | 0.287 | 0.342 | 1.027 | 102.489 | 0.819 | 0.380 | | |
| 0~2 | 生データ | 0.355 | 0.374 | 2.009 | 105.724 | 0.800 | 0.743 | | |
| 時点前 | 平滑値 | 0.280 | 0.334 | 0.751 | 101.157 | 0.821 | 0.392 | 25.839 | 0.846 |
| 0~4 | 生データ | 0.359 | 0.381 | 1.825 | 104.789 | 0.803 | 0.749 | 17.163 | 0.866 |
| 時点前 | 平滑値 | 0.277 | 0.330 | 0.983 | 99.167 | 0.826 | 0.415 | | |
| 0~6 | 生データ | 0.336 | 0.367 | 1.101 | 101.478 | 0.811 | 0.749 | 12.776 | 0.879 |
| 時点前 | 平滑値 | 0.278 | 0.327 | 0.914 | 87.910 | 0.830 | 0.394 | | |
| 0~9 | 生データ | 0.369 | 0.380 | 2.415 | 102.374 | 0.813 | 0.898 | 10.018 | 0.902 |
| 時点前 | 平滑値 | 0.296 | 0.349 | 1.338 | 101.734 | 0.827 | 0.340 | | |

(b)

生データを用いた場合の手算精度 (大気汚染濃度平均値 = 137.292 (近畿B市))

| 回帰データ | 比較データ | 単純平均誤差率 | 両重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 真相関係数 |
|-------|-------|---------|---------|-------|--------|-------|-------|-------|-------|
| 0 | 生データ | 0.507 | 0.515 | 5.683 | 27.612 | 0.737 | 0.344 | | |
| 時点前 | 平滑値 | 0.388 | 0.431 | 3.113 | 69.844 | 0.763 | 0.193 | 7.436 | 0.438 |
| 0~1 | 生データ | 0.515 | 0.521 | 5.872 | 27.372 | 0.738 | 0.348 | | |
| 時点前 | 平滑値 | 0.372 | 0.416 | 3.222 | 67.861 | 0.771 | 0.225 | 4.922 | 0.485 |
| 0~2 | 生データ | 0.521 | 0.513 | 6.046 | 27.817 | 0.738 | 0.313 | | |
| 時点前 | 平滑値 | 0.374 | 0.421 | 3.316 | 67.058 | 0.774 | 0.211 | 3.667 | 0.513 |
| 0~4 | 生データ | 0.520 | 0.538 | 6.063 | 27.785 | 0.738 | 0.326 | | |
| 時点前 | 平滑値 | 0.386 | 0.429 | 3.347 | 68.678 | 0.770 | 0.199 | 2.241 | 0.531 |
| 0~6 | 生データ | 0.500 | 0.513 | 4.913 | 29.391 | 0.729 | 0.345 | | |
| 時点前 | 平滑値 | 0.391 | 0.431 | 2.639 | 70.878 | 0.762 | 0.236 | 1.888 | 0.578 |
| 0~9 | 生データ | 0.523 | 0.511 | 4.104 | 31.302 | 0.683 | 0.213 | | |
| 時点前 | 平滑値 | 0.406 | 0.470 | 2.196 | 81.655 | 0.721 | 0.068 | 2.187 | 0.639 |

(c)

生データを用いた場合の手算精度 (大気汚染濃度平均値 = 72.482 (中国C市))

| 回帰データ | 比較データ | 単純平均誤差率 | 両重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 真相関係数 |
|-------|-------|---------|---------|-------|--------|-------|-------|--------|-------|
| 0 | 生データ | 0.587 | 0.566 | 3.677 | 42.327 | 0.735 | 0.342 | | |
| 時点前 | 平滑値 | 0.404 | 0.456 | 3.176 | 39.094 | 0.771 | 0.965 | 25.272 | 0.659 |
| 0~1 | 生データ | 0.546 | 0.547 | 2.743 | 42.283 | 0.737 | 0.575 | | |
| 時点前 | 平滑値 | 0.402 | 0.451 | 2.630 | 39.081 | 0.775 | 0.414 | 14.538 | 0.690 |
| 0~2 | 生データ | 0.574 | 0.563 | 2.888 | 42.248 | 0.738 | 0.610 | | |
| 時点前 | 平滑値 | 0.394 | 0.439 | 2.463 | 38.395 | 0.781 | 0.459 | 10.430 | 0.710 |
| 0~4 | 生データ | 0.668 | 0.620 | 3.727 | 42.522 | 0.743 | 0.639 | | |
| 時点前 | 平滑値 | 0.423 | 0.453 | 1.466 | 36.655 | 0.796 | 0.546 | 7.292 | 0.749 |
| 0~6 | 生データ | 0.767 | 0.715 | 3.263 | 53.258 | 0.681 | 0.551 | | |
| 時点前 | 平滑値 | 0.454 | 0.518 | 1.365 | 47.665 | 0.720 | 0.443 | 5.424 | 0.769 |
| 0~9 | 生データ | 1.023 | 0.930 | 4.064 | 68.579 | 0.557 | 0.425 | | |
| 時点前 | 平滑値 | 0.634 | 0.693 | 2.542 | 63.009 | 0.620 | 0.310 | 3.758 | 0.789 |

(d)

生データを用いた場合の手算精度 (大気汚染濃度平均値 = 19.887 (九州D市))

| 回帰データ | 比較データ | 単純平均誤差率 | 両重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 真相関係数 |
|-------|-------|---------|---------|-------|--------|-------|-------|--------|-------|
| 0 | 生データ | 1.887 | 1.550 | 7.869 | 11.485 | 0.657 | 0.715 | | |
| 時点前 | 平滑値 | 1.479 | 1.309 | 4.742 | 10.893 | 0.682 | 0.723 | 16.830 | 0.560 |
| 0~1 | 生データ | 1.922 | 1.553 | 6.177 | 12.058 | 0.659 | 0.710 | | |
| 時点前 | 平滑値 | 1.446 | 1.348 | 4.251 | 12.119 | 0.691 | 0.661 | 12.442 | 0.561 |
| 0~2 | 生データ | 1.789 | 1.546 | 6.153 | 12.069 | 0.653 | 0.712 | | |
| 時点前 | 平滑値 | 1.382 | 1.314 | 4.609 | 11.084 | 0.663 | 0.689 | 8.576 | 0.674 |
| 0~4 | 生データ | 1.927 | 1.637 | 6.983 | 12.761 | 0.639 | 0.678 | | |
| 時点前 | 平滑値 | 1.482 | 1.393 | 4.950 | 12.540 | 0.651 | 0.644 | 5.696 | 0.707 |
| 0~6 | 生データ | 1.944 | 1.656 | 7.880 | 12.741 | 0.640 | 0.708 | | |
| 時点前 | 平滑値 | 1.397 | 1.426 | 4.266 | 12.807 | 0.646 | 0.635 | 4.170 | 0.725 |
| 0~9 | 生データ | 1.978 | 1.693 | 7.650 | 13.197 | 0.621 | 0.685 | | |
| 時点前 | 平滑値 | 1.543 | 1.435 | 4.223 | 12.935 | 0.643 | 0.601 | 2.786 | 0.741 |

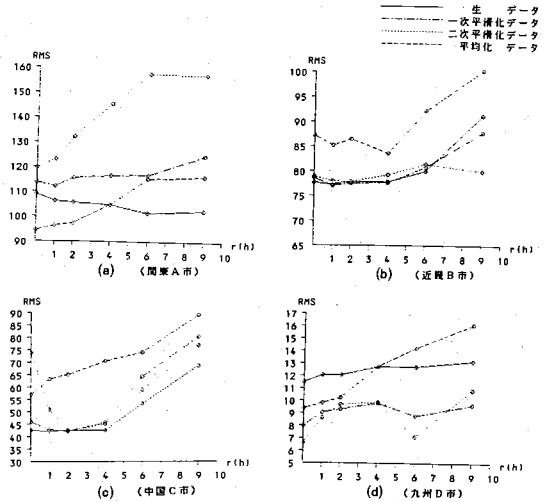


図2. rを変化させた場合のRMS誤差水準

加に伴ない誤差も増える傾向があることがわかる。このことは先に述べた意味において都合がよく、道路の環境アセスメントという観点からは、その道路がおかれている時点の交通量を始めとする要因から予測すればよいということがほぼ結論づけられる。

また、予測に用いるデータとしては、汚染濃度が高い関東A市の結果から判断すれば、平均化データが最も精度の高い予測値を与えており、ついで生データとなっている。

全体的にみた場合、予測精度は必ずしも高くなく、単純平均誤差率で30%以上、RMS誤差の平均値に対する割合も50%以上となる場合が多い。しかし、一致係数ではほぼ0.7以上の水準は確保できると思われる。

数量化理論I類による予測

重回帰分析法による予測は、暗黙のうちに各説明変数の被説明変数に対する影響が線形であることを前提としているが、実際にはその影響が非線形である可能性は十分考えられる。また、ここで採用している8つの説明変数はすべて定量的変量とはいえ、その値はその時々状況に応じてランダムに変動している性質のものである。こうしたことを考慮して、本節では数量化理論I類を用いてNOx濃度を予測することを試みる。本理論では、すべての説明変数をいくつかのカテゴリーに分け、各カテゴリーごとに、変数とそのカテゴリーに反応したか否かで1, 0の値をとるダミー変数を考えて、それに掛かる係数を最小二乗法で決めることによって、予測構造式を確定する。

いま、変数jのカテゴリーkを表わすダミー変数を $\delta_{jk}(t)$ 、それに対する係数を a_{jk} としたとき

$$y(t) = \sum_{j=1}^p \sum_{k=1}^C a_{jk} \delta_{jk}(t) + \varepsilon(t)$$

のように予測式が与えられる。なお、ここでは前節の検討から、式(1)の γ に対応する量は0としていることに注意する。

生データを用いた場合について、各変数のカテゴリー区分と重み係数 a_{jk} 、レンジ、偏相関係数を示すと表2のようになる。重相関係数はいずれの地域も重回帰分析の時より大きくなる傾向がみられ、予測能が向上するという改善が得られたが、このことはやはり説明変数に潜んでいる非線形的影響を物語るものと思われる。

各変数の影響の程度は、九州のD市を除いて、レンジ、偏相関係数ともに交通量が高い値を示しており、他の変数は地域によってまちまちの影響を示している。雲量、気温はいずれにしても大きな影響を及ぼさないと予想さ

表3. 数量化理論I類を用いた場合の予測精度

| 大気汚染濃度平均値 = 229.427 (関東A市) | | | | | | | | |
|----------------------------|-------|---------|---------|-------|--------|-------|-------|-------|
| 使用データ | 比較データ | 単純平均誤差率 | 荷重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 偏相関係数 |
| 生データ | 生データ | 0.285 | 0.315 | 1.100 | 91.240 | 0.831 | 0.769 | 0.836 |
| データ | 平滑値 | 0.237 | 0.287 | 1.055 | 87.508 | 0.848 | 0.430 | 0.836 |
| 平均化データ | 生データ | 0.355 | 0.387 | 2.255 | 99.789 | 0.816 | 0.732 | 0.802 |
| データ | 平滑値 | 0.256 | 0.319 | 1.087 | 95.644 | 0.836 | 0.308 | 0.802 |

| 大気汚染濃度平均値 = 137.292 (近畿B市) | | | | | | | | |
|----------------------------|-------|---------|---------|-------|--------|-------|-------|-------|
| 使用データ | 比較データ | 単純平均誤差率 | 荷重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 偏相関係数 |
| 生データ | 生データ | 0.435 | 0.455 | 4.601 | 68.367 | 0.781 | 0.442 | 0.575 |
| データ | 平滑値 | 0.331 | 0.373 | 2.447 | 59.467 | 0.809 | 0.358 | 0.575 |
| 平均化データ | 生データ | 0.484 | 0.511 | 3.367 | 85.448 | 0.703 | 0.189 | 0.515 |
| データ | 平滑値 | 0.391 | 0.464 | 1.688 | 81.103 | 0.719 | 0.132 | 0.515 |

| 大気汚染濃度平均値 = 72.482 (中国C市) | | | | | | | | |
|---------------------------|-------|---------|---------|-------|--------|-------|-------|-------|
| 使用データ | 比較データ | 単純平均誤差率 | 荷重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 偏相関係数 |
| 生データ | 生データ | 0.624 | 0.584 | 3.364 | 36.774 | 0.786 | 0.624 | 0.705 |
| データ | 平滑値 | 0.486 | 0.561 | 1.174 | 37.468 | 0.797 | 0.435 | 0.705 |
| 平均化データ | 生データ | 0.572 | 0.584 | 4.112 | 40.641 | 0.752 | 0.621 | 0.784 |
| データ | 平滑値 | 0.410 | 0.462 | 3.352 | 37.983 | 0.785 | 0.543 | 0.784 |

| 大気汚染濃度平均値 = 19.887 (九州D市) | | | | | | | | |
|---------------------------|-------|---------|---------|-------|--------|-------|-------|-------|
| 使用データ | 比較データ | 単純平均誤差率 | 荷重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 偏相関係数 |
| 生データ | 生データ | 1.280 | 1.229 | 5.213 | 9.890 | 0.690 | 0.646 | 0.632 |
| データ | 平滑値 | 1.138 | 1.131 | 4.518 | 10.274 | 0.690 | 0.627 | 0.643 |
| 平均化データ | 生データ | 1.388 | 1.355 | 6.271 | 19.169 | 0.678 | 0.552 | 0.643 |
| データ | 平滑値 | 1.125 | 1.184 | 4.719 | 10.334 | 0.682 | 0.525 | 0.643 |

れ、説明変数から省いても差し支えないと考えられる。

データとして、平滑化データを用いた場合も、平均化データを用いた場合も、上記のことと同様の結論が得られた。

生データを用いて予測した場合と平均化データを用いて予測した場合について、各地域の予測精度を示すと表3のようになる。

表1の $\gamma=0$ の場合の結果と比較すると、生データを用いかつ生データを真値としたときの結果は、次のようになる。

まず、関東A市の場合、単純平均誤差率および荷重平均誤差率は、数量化理論I類を用いた方が、約8%~9%減少する。最大誤差率は約50%減少する。RMS誤差は109 ppbから91 ppbへと約20%減少し、一致係数、相関係数はそれぞれ0.04、0.07増加する。

その他の地域についてもRMS誤差で比較してみると、

表2. 数量化理論I類による分析結果 (生データを用いた場合)

| 変数 | カテゴリ区分 | 重み係数 | 偏相関係数 | |
|--------------|-----------|----------|---------|-------|
| | | | レンジ | 偏相関係数 |
| 交通量 (台/時) | ~600 | -147.697 | 300.599 | 0.761 |
| | 600~1000 | -86.681 | | |
| | 1000~1500 | 21.457 | | |
| | 1500~2000 | 37.184 | | |
| | 2000~2500 | 73.769 | | |
| X方向の風速 (m/秒) | ~1.0 | -91.112 | 100.109 | 0.356 |
| Y方向の風速 (m/秒) | ~1.0 | -96.467 | | |
| 気温 (度) | ~10.0 | - | 116.597 | 0.304 |
| | 10.0~20.0 | - | | |
| | 20.0~30.0 | - | | |
| 湿度 (%) | ~40.0 | -120.546 | 157.707 | 0.525 |
| | 40.0~60.0 | -67.888 | | |
| | 60.0~80.0 | 3.956 | | |
| | 80.0~ | 37.184 | | |
| 日照量 (ca1) | ~0 | -6.581 | 21.084 | 0.123 |
| | 0~50 | 14.423 | | |
| | 50~100 | - | | |
| 雲量 (%) | ~3.5 | 14.794 | 25.981 | 0.124 |
| | 3.5~5.5 | -11.688 | | |
| | 5.5~7.5 | -4.507 | | |

| 偏相関係数 = 0.836 (関東A市) | | | | |
|----------------------|-----------|---------|---------|-------|
| 変数 | カテゴリ区分 | 重み係数 | レンジ | 偏相関係数 |
| 交通量 (台/時) | ~600 | -51.666 | 135.184 | 0.428 |
| | 600~1000 | 12.688 | | |
| | 1000~1500 | 25.807 | | |
| | 1500~2000 | 24.821 | | |
| | 2000~2500 | 25.221 | | |
| X方向の風速 (m/秒) | ~1.0 | 12.005 | 58.156 | 0.225 |
| Y方向の風速 (m/秒) | ~1.0 | -18.842 | | |
| 気温 (度) | ~10.0 | -8.642 | 16.244 | 0.112 |
| | 10.0~20.0 | 5.050 | | |
| | 20.0~30.0 | -11.192 | | |
| 湿度 (%) | ~40.0 | -19.745 | 31.685 | 0.165 |
| | 40.0~60.0 | -8.791 | | |
| | 60.0~80.0 | -11.846 | | |
| | 80.0~ | 7.412 | | |
| 日照量 (ca1) | ~0 | 23.241 | 67.032 | 0.322 |
| | 0~50 | -23.505 | | |
| | 50~100 | -43.792 | | |
| 雲量 (%) | ~3.5 | 10.875 | 22.958 | 0.179 |
| | 3.5~5.5 | 5.223 | | |
| | 5.5~7.5 | -4.507 | | |

| 偏相関係数 = 0.705 (中国C市) | | | | |
|----------------------|-----------|---------|--------|-------|
| 変数 | カテゴリ区分 | 重み係数 | レンジ | 偏相関係数 |
| 交通量 (台/時) | ~600 | -36.253 | 65.470 | 0.433 |
| | 600~1000 | -32.891 | | |
| | 1000~1500 | -23.277 | | |
| | 1500~2000 | 24.805 | | |
| | 2000~2500 | 21.271 | | |
| X方向の風速 (m/秒) | ~1.0 | 12.397 | 21.793 | 0.154 |
| Y方向の風速 (m/秒) | ~1.0 | -2.042 | | |
| 気温 (度) | ~10.0 | 16.459 | 18.898 | 0.123 |
| | 10.0~20.0 | -2.440 | | |
| | 20.0~30.0 | 5.105 | | |
| 湿度 (%) | ~40.0 | 24.250 | 29.079 | 0.232 |
| | 40.0~60.0 | -4.829 | | |
| | 60.0~80.0 | 24.123 | | |
| | 80.0~ | 26.094 | | |
| 日照量 (ca1) | ~0 | 17.025 | 46.526 | 0.369 |
| | 0~50 | -6.520 | | |
| | 50~100 | -31.262 | | |
| | 100~150 | 23.557 | | |
| 雲量 (%) | ~3.5 | -3.791 | 51.524 | 0.271 |
| | 3.5~5.5 | -13.861 | | |
| | 5.5~7.5 | 37.663 | | |

| 偏相関係数 = 0.632 (九州D市) | | | | |
|----------------------|-----------|---------|--------|-------|
| 変数 | カテゴリ区分 | 重み係数 | レンジ | 偏相関係数 |
| 交通量 (台/時) | ~600 | -3.637 | 5.998 | 0.148 |
| | 600~1000 | -1.159 | | |
| | 1000~1500 | 2.381 | | |
| | 1500~2000 | 1.835 | | |
| | 2000~2500 | 1.835 | | |
| X方向の風速 (m/秒) | ~1.0 | 4.224 | 14.581 | 0.338 |
| Y方向の風速 (m/秒) | ~1.0 | -10.357 | | |
| 気温 (度) | ~10.0 | 6.048 | 10.369 | 0.273 |
| | 10.0~20.0 | -0.292 | | |
| | 20.0~30.0 | -4.321 | | |
| 湿度 (%) | ~40.0 | -5.675 | 12.444 | 0.301 |
| | 40.0~60.0 | -6.235 | | |
| | 60.0~80.0 | 0.024 | | |
| | 80.0~ | 6.209 | | |
| 日照量 (ca1) | ~0 | 6.183 | 12.123 | 0.353 |
| | 0~50 | 5.940 | | |
| | 50~100 | - | | |
| 雲量 (%) | ~3.5 | -1.295 | 4.887 | 0.175 |
| | 3.5~5.5 | -0.139 | | |
| | 5.5~7.5 | 3.592 | | |

近畿B市の場合で数理化I類の方が約12%少なくなっており、同様に中国C市、九州D市でも、それぞれ8%、14%減少している。

一般化された重回帰分析法による予測

重回帰分析法では説明変数をすべて定量的なものとして扱いかい、逆に数理化理論I類では説明変数はすべてカテゴリー化し定性的なものとして扱おうが、実際問題としては連続的な定量的変数とカテゴリー化された変数とを混在させた形で分析したい場合が少なくないし、そうすることによって予測精度が上昇する可能性もある。したがって、本節ではこの方法によりNO_x濃度を予測し、その適用性について検討してみる。

本方法は、一部の説明変数が定量的であり、一部の説明変数が定性的であることを許すという意味において、ここでは一般化された重回帰分析法とよぶが、見方を変えれば一般化された数理化理論I類とよんでもよいであろう。

この方法の本問題への適用にあたっては、8つの説明変数のうちいずれの変数をカテゴリー化し、いずれの変数をそのまま定量的変数として残しておくかについてまず考えなければならないが、ここでは最も影響が大きく変動幅も大きい交通量を定量的変数とし、残りの変数についてはその影響の仕方はせいぜい段階的であろうという判断の下に、すべて数理化理論I類と同じようにカテゴリー化して与える。

適用結果を予測精度についてまとめて示すと表4のよ
表4. 一般化された重回帰分析(数理化理論I類)を用いた場合の予測精度

| 大気汚染濃度平均値 = 229.427 (関東A市) | | | | | | | | | |
|----------------------------|-------|---------|---------|-------|---------|-------|-------|--------|-------|
| 使用データ | 比較データ | 単純平均誤差率 | 多重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 重回帰係数 |
| 生 | 生データ | 0.294 | 0.347 | 1.172 | 100.196 | 0.812 | 0.764 | | |
| データ | 平均値 | 0.274 | 0.326 | 0.773 | 99.595 | 0.825 | 0.574 | 45.079 | 0.830 |
| 平均化 | 生データ | 0.996 | 0.408 | 1.824 | 105.031 | 0.799 | 0.702 | | |
| データ | 平均値 | 0.281 | 0.353 | 1.043 | 105.683 | 0.818 | 0.259 | 25.667 | 0.782 |
| 大気汚染濃度平均値 = 137.382 (近畿B市) | | | | | | | | | |
| 使用データ | 比較データ | 単純平均誤差率 | 多重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 重回帰係数 |
| 生 | 生データ | 0.458 | 0.471 | 4.984 | 69.062 | 0.779 | 0.403 | | |
| データ | 平均値 | 0.320 | 0.368 | 3.683 | 60.071 | 0.805 | 0.363 | 6.107 | 0.519 |
| 平均化 | 生データ | 0.495 | 0.523 | 3.723 | 90.165 | 0.680 | 0.171 | | |
| データ | 平均値 | 0.415 | 0.478 | 1.507 | 84.570 | 0.695 | 0.113 | 4.336 | 0.476 |
| 大気汚染濃度平均値 = 72.482 (中国C市) | | | | | | | | | |
| 使用データ | 比較データ | 単純平均誤差率 | 多重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 重回帰係数 |
| 生 | 生データ | 0.760 | 0.892 | 3.857 | 44.443 | 0.732 | 0.491 | | |
| データ | 平均値 | 0.553 | 0.582 | 4.522 | 42.688 | 0.760 | 0.327 | 15.525 | 0.726 |
| 平均化 | 生データ | 0.728 | 0.731 | 4.723 | 52.535 | 0.669 | 0.325 | | |
| データ | 平均値 | 0.563 | 0.591 | 2.085 | 50.824 | 0.696 | 0.329 | 28.527 | 0.822 |
| 大気汚染濃度平均値 = 19.887 (九州D市) | | | | | | | | | |
| 使用データ | 比較データ | 単純平均誤差率 | 多重平均誤差率 | 最大誤差率 | RMS | 一致係数 | 相関係数 | 分散比 | 重回帰係数 |
| 生 | 生データ | 1.755 | 1.832 | 7.919 | 18.091 | 0.620 | 0.574 | | |
| データ | 平均値 | 1.500 | 1.483 | 5.723 | 13.428 | 0.630 | 0.548 | 12.923 | 0.637 |
| 平均化 | 生データ | 1.779 | 1.623 | 6.770 | 12.076 | 0.644 | 0.588 | | |
| データ | 平均値 | 1.340 | 1.347 | 4.438 | 11.843 | 0.655 | 0.544 | 7.835 | 0.569 |

うになる。RMS誤差で評価した場合、関東A市、近畿B市で重回帰分析法の場合より約10%予測精度が向上しているが、数理化理論I類の場合の精度より劣っている。また、中国C市、九州D市の場合には、重回帰分析法よりかえって悪い予測を行う結果となっていることがわかる。

む す び

以上、本研究では関東、近畿、中国および九州の4都市の実測データを用い、NO_x汚染濃度を道路の環境アセスメントという観点から、3つの相異なる統計的手法によって予測することを試みた。すなわち、重回帰分析法、数理化理論I類、一般化された重回帰分析法の3方法による予測である。これらの方法は、最初の方法が定量的変数のみを説明変数とし、その影響の仕方を線形と仮定するのに対し、第2番目の方法はすべての変数を、最後の方法は一部の説明変数のみを定性的変数として、あるいはカテゴリー化された変数として取り扱おうことができ、したがって線形構造を有しつつ定量的変数の非線形性を考慮しようという点に特色を有している。

適用の結果は、交通量、風速など本研究でとり挙げた8つの説明要因すべてについてその非線形性を仮定する数理化理論I類による予測が最も好ましい予測値を与えるという結論に達した。また、説明力としては、交通量は際立って大きいものを持っており、気温、雲量といった変数のそれは逆にあまり大きくないことがわかった。

しかしながら、最良の予測精度をあげている結果でも、RMS誤差の平均値に対する割合が40%となっており、未だ満足のものではない。これは予測理論の不備というよりも、予測対象のNO_x濃度のサンプル値を1時間の平均値にとっており、そのランダムな変動を統計的な予測手法で精度高く追うことには一つの限界があるという基本的な事由に根ざしていると考えられる。

したがって、今後の目指すべき方向としては、限られたデータの中でサンプル数の減少という犠牲を払うことは覚悟して、たとえば連続した8時間の平均汚染濃度を、やはり同じ時間で平均化した説明変数で予測するといったようなことが考えられる。

参 考 文 献

奥野忠一他：多変量解析法 日科技連