

## 欠測データ集合を扱う神経回路網法 CQSAR を用いた 環境指標データの補完と分布図

神部順子<sup>a</sup>、長嶋雲兵<sup>b</sup>、青山 智夫<sup>c</sup>

<sup>a</sup>江戸川大学メディアコミュニケーション学部，〒270-0198 流山市駒木 474、

<sup>b</sup>産業技術総合研究所計算科学研究部門，〒305-8568 つくば市梅園 1-1-1

<sup>c</sup>宮崎大学工学部電気電子工学科，〒889-2192 宮崎市学園木花台西 1-1

*E-mail: kambe@edogawa-u.ac.jp*

### 1. はじめに

今日、環境問題は個別の地域を研究する時代から、大域的な地域を関連付けて考えなければならない時代になっている。日本の西に位置し、日本の大気に大きな影響を及ぼす中国・東南アジアは、経済発展の著しい地域であり、そこでの大気汚染はかなり憂慮される状況である。

Figure 1 に 2005 年 4 月の東京都、千葉県、埼玉県、川崎市の SPM の月間平均測定点を示す [1]。この例では測定点が東京湾岸に集中していることが判るが、SPM の濃度分布を直感的に理解することはむずかしい。ところが、大気汚染指標である窒素酸化物 (NO<sub>x</sub>) や浮遊粒子状物質 (Suspended Particulate Matter: SPM) などのデータの測定点は離散的であり、かつ疎であり、また一部に偏っている。さらにそれらは欠測が多く、従来の統計的解析では欠測を含むデータのほとんどが捨て去られるため、多くの情報が失われている。

この種の分布を直感的な表現とするためには、等高線による分布図が用いられる。ところが、大気環境データの測定点は離散的であり、かつ疎、また一部に偏っているという性質があり通常の方法では、分布図の作成は難しい。そのため本研究では、欠測データ集合を扱う神経回路網法 CQSAR: Compensation Quantitative Structure Activity Relationships [2,3] を用いて格子点上のデータ補完を行い、そのデータを用いて分布図を書くことを試みた。

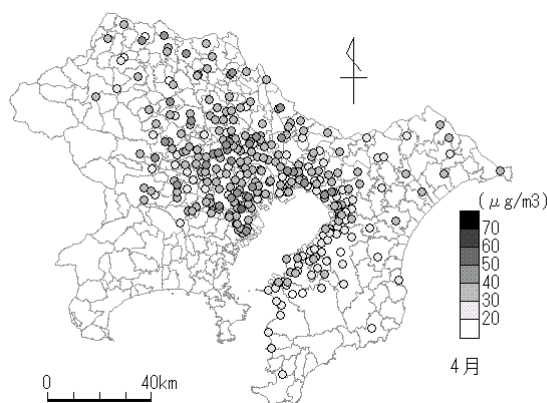


Fig.1. Observation points and monthly averaged density of SPM in and around Tokyo on April, 2005.

### 2. データと手順

CQSAR [2, 3] は階層型神経回路網を用いた新しいデータ補完法として使うことができる。階層型神経回路網は因果関係が不明な現象の多変量解析に用いられ、主な用途は環境問題解析、薬理活性解析などである。

CQSAR 処理はおおまかに 2 つのステップで構成されている。その第 1 ステップでは、欠測部を各説明変数、対象現象の観測値について独立に補完する。CQSAR 処理の第 2 ステップは第 1 ステ

ップで完全化したデータの神経回路網による QSAR をおこなう。第 2 ステップでの QSAR 用の神経回路網学習では simulated annealing 操作を行った。学習は一般的な back propagation 学習よりもニューロン間結合強度消去操作によりネットワーク構造を最小化する再構築学習法 reconstruction learning[4]を用いた。

### 3 . 結果

学習用サンプルデータは、仮想的に作成した鞍点を持つ谷のデータ 961 点 (=31×31) (Figure 2) から 10 点をランダムにサンプルした。Figure 3 に描画に用いた 131 点(CQSAR による補間点 121 + 教師データ 10 点)とそれらから作られる分布図を示す。Figure 4 上の図で格子点の上にはない 10 点が教師データである。

Figure 2 と比較してみると、教師データ周辺は良く再現されており、全体的にたかだか 10 点のサンプリングと、CQSAR の補間によって鞍点を持つ谷の形状が良く再現されていることがわかる。サンプリングが全く無い左下および右上側に見られるように、外挿点の精度はわるいが、大まかな形は良くとらえられている。

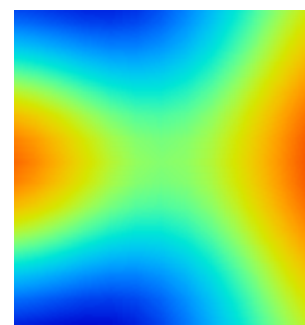


Fig.2 Schematic distribution

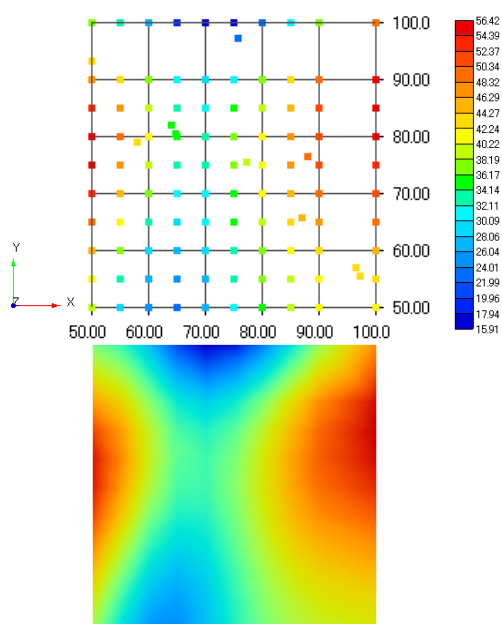


Fig. 3 Distribution of estimated data and map

### 謝辞

本研究の一部は、地球環境研究総合推進費 H19 地球環境研究革新型研究課題浮遊粒子状物質 (SPM) および大気汚染物質の脳型多変量解析技法の開発(FY2007 - FY2008)代表：神部順子の補助を受けている。

[1] 神部順子、中山栄子、青山智夫、長嶋雲兵、2007 日本コンピュータ化学会秋季年会 1P05、2007 年、兵庫。

<http://media.laic.u-hyogo.ac.jp/%7Ehayashi/SCCJ2007/abstract-pdf/SCCJ2007au-1P05.pdf>

[2] Bitou, K., Yuan, Y., Aoyama, T., Nagashima, U., *Proceedings of International Conference on Control Automation and Systems 2003*, CD-ROM (TA06-04),10.22-25, 2003.

[3] 青山智夫, 神部順子, 長嶋雲兵, *J. Comput. Chem. Jpn.*, Vol. 6, No. 5, pp. 263–274 (2007).

[4] Aoyama, T., Ichikawa, H., *Chem. Pharm. Bull.*, 39,1222-1228(1991).