

# ハイブリッド遺伝アルゴリズムを用いた土壤汚染 概況調査における試料採取地点最適配置探索

木内智明<sup>1</sup>・米田稔<sup>2</sup>・森澤眞輔<sup>3</sup>・大塚順基<sup>4</sup>

<sup>1</sup>学生会員 京都大学大学院 工学研究科環境地球工学専攻 (〒606-8501 京都市左京区吉田本町)

<sup>2</sup>正会員 工博 京都大学助教授 工学研究科環境地球工学専攻 (〒606-8501 京都市左京区吉田本町)

<sup>3</sup>フェロー会員 工博 京都大学教授 工学研究科環境地球工学専攻 (〒606-8501 京都市左京区吉田本町)

<sup>4</sup>非会員 工修 (株)三機工業 (〒242-0001 大和市中鶴間1742の7)

本研究では一般に無情報条件下で行われる土壤汚染概況調査において試料採取地点の最適配置を決定するために、空間的相関の強さを表すパラメータの無情報事前分布を仮定し、領域全体での空間平均値を有限個の観測値の平均値で推定する場合の推定分散に基づいた評価関数を設定した。そして遺伝アルゴリズムと最良地点逐次移動法を組み合わせたハイブリッドアルゴリズムを構築し各手法の有効性の比較を行った。そして種々の形状の領域に対して最適配置探索を行い、任意形状領域への任意の地点数による採取地点配置の探索も試みた。その結果、試料採取地点は領域内に概ね均等に分布することを確認した。

**Key Words :** *soil pollution, sampling points, optimal arrangement, noninformative condition  
hybrid genetic algorithm, spatial correlation*

## 1. 序論

現在、土壤汚染の概況調査において、「土壤のサンプリング密度は概ね1000m<sup>2</sup>につき1ヶ所で、試料採取は1ヶ所につき、基本的には中心1地点及び周辺4方位の5mから10mまでの間からそれぞれ1地点の合計5地点で行う」という5地点混合方式が、環境庁監修の「土壤・地下水汚染対策ハンドブック」により提示されている<sup>1)</sup>。これは、約1000m<sup>2</sup>の面積での平均土壤汚染濃度を、その領域内の5地点での測定値の平均値で代表することを意味するが、その理論的な根拠については何ら述べられていない。

米田ら<sup>2)</sup>は、いくつかの統計構造を持つ場について、遺伝アルゴリズム(GA)を用いて5地点混合方式における試料採取地点の最適配置を求めているが、最適配置の探索に失敗するケースも見られた。それを踏まえて木内ら<sup>3)</sup>は、GAに加えてニューラルネットワークにおいて見られる手法を利用したアルゴリズムを構築するとともに、それを用いてより多様な条件下において最適配置を探索した。他にも最適な試料採取地点やモニタリング地点の探索を行った例としてはCarreraら<sup>4)</sup>、Rouhani<sup>5)</sup>、Meyer and Brill<sup>6)</sup>、Tucciarelli and Pinder<sup>7)</sup>、McKinney and Loucks<sup>8)</sup>、Andricevic<sup>9)</sup>、James and Freeze<sup>10)</sup>、

Mayerら<sup>11)</sup>、James and Gorelick<sup>12)</sup>、Andricevic<sup>13)</sup>、Cieniawskiら<sup>14)</sup>、Wagner<sup>15)</sup>の研究などがあるが、それらは全て事前に汚染分布の空間的相関構造が既知であるとしている。しかし現実の場合にはこのような相関構造に関する事前情報は与えられていない場合がほとんどである。

本研究ではまず、汚染分布の空間的相関構造が未知であることを前提として、採取地点配置を評価する関数を導出する。そして、より性能の高い最適配置探索アルゴリズムを構築し、それを用いて任意の地形・任意の地点数における採取地点配置の探索を行うことを試みる。

なお本研究では水平2次元方向の濃度分布のみについて考え、垂直方向の濃度分布については考えないものとする。

## 2. 評価関数の設定

2次元座標を表すベクトルを $X$ 、水平2次元空間で分布する確率変数を $Z(X)$ とする。 $Z(X)$ に統計的弱定常性と等方性を仮定すると、 $Z(X)$ の期待値 $m$ は場所によって変化せず、2地点 $x_1$ と $x_2$ での $Z(X)$ の値 $Z(x_1)$ 、 $Z(x_2)$ の共分散

$\text{Cov}(Z(X_1), Z(X_2))$ は、次式のように $x_1, x_2$ 間の距離 $h$ のみの関数となる。

$$\text{Cov}(Z(X_1), Z(X_2)) = C(h) \quad (1)$$

$\sigma$ を $Z(x)$ の標準偏差とすると、相関係数 $r(h)$ は次式で定義される。

$$r(h) = \frac{C(h)}{\sigma^2} \quad (2)$$

$r(h)$ の関数形としては、正定値性の条件<sup>10</sup>を満たすものならば様々なものが考えられるが、本研究ではそれらのうち、代表的なものとして次式で示す指数型相関関数形を適用した。

$$r(h) = \exp\left(-\frac{h}{L}\right) \quad (3)$$

ここで $L$ は、相関の強さを決定するパラメータであり、この $L$ の値を分布させることによって、現実中存在する十分様々な場の相関構造を表現できると考えた。

本研究では、それぞれの採取地点配置を評価する関数として次式に示す推定分散を用いる。推定分散式の導出過程を付録に示す。

$$\begin{aligned} \sigma_E^2 &= \frac{1}{S^2} \int_S dX \int_S C(|X - X'|) dX' \\ &+ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n C(|X_{N_i} - X_{N_j}|) \\ &- \frac{2}{nS} \sum_{i=1}^n \int_S C(|X - X_{N_i}|) dX \end{aligned} \quad (4)$$

ここで $n$ は採取地点数、 $S$ は対象領域の全面積、 $X_{N_i}$  ( $i=1, 2, \dots, n$ )は $n$ 個の採取地点の内、 $i$ 番目の採取地点の座標、 $|X_{N_i} - X_{N_j}|$ は2つの採取地点 $X_{N_i}$ と $X_{N_j}$ 間の距離、 $|X - X_{N_i}|$ は領域内の任意の座標 $X$ と採取地点 $X_{N_i}$ 間の距離を表す。

本研究では、式(4)を $Z(X)$ の分散 $\sigma^2$ で割った次式を評価関数とする。

$$\begin{aligned} \frac{\sigma_E^2}{\sigma^2} &= \frac{1}{S^2} \int_S dX \int_S r(|X - X'|) dX' \\ &+ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n r(|X_{N_i} - X_{N_j}|) \\ &- \frac{2}{nS} \sum_{i=1}^n \int_S r(|X - X_{N_i}|) dX \end{aligned} \quad (5)$$

式(5)に式(3)を代入すると、

$$\begin{aligned} \frac{\sigma_E^2}{\sigma^2} &= \frac{1}{S^2} \int_S dX \int_S \exp\left(-\frac{|X - X'|}{L}\right) dX' \\ &+ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \exp\left(-\frac{|X_{N_i} - X_{N_j}|}{L}\right) \\ &- \frac{2}{nS} \sum_{i=1}^n \int_S \exp\left(-\frac{|X - X_{N_i}|}{L}\right) dX \end{aligned} \quad (6)$$

式(6)を $L$ の関数 $g(L)$ として表す。

木内ら<sup>11</sup>は $L$ の値を3,15,60mとした場合についてそれぞれ $g(L)$ を評価関数として配置の優劣を評価したが、実際の概況調査では $L$ の値は事前には不明である。いま、 $L$ は正の値しかとらず、他には何ら $L$ の確率分布に関する情報が与えられていないとすると、このような場合、ベイズ統計理論では、その対数値が一様分布するという無情報事前分布がよく採用される。本研究でも $L$ の確率分布として、その対数値が一様分布すると仮定した。このとき、 $L$ の確率密度関数 $p(L)$ は次式で表すことができる。

$$p(L) = \frac{c}{L} \quad (7)$$

ここで $c$ は $p(L)$ の全ての $L$ についての積分値が1となるように決められる定数である。

よって本研究では、次式で表される $g(L)$ の期待値によって配置を評価することとした。

$$\begin{aligned} &\int_{L_1}^{L_2} p(L)g(L)dL \\ &= \frac{c^2}{S^2} \int_S dX \int_S \left\{ E_i\left(-\frac{|X - X'|}{L_1}\right) - E_i\left(-\frac{|X - X'|}{L_2}\right) \right\} dX' \\ &+ \frac{c}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left\{ E_i\left(-\frac{|X_{N_i} - X_{N_j}|}{L_1}\right) - E_i\left(-\frac{|X_{N_i} - X_{N_j}|}{L_2}\right) \right\} \\ &- \frac{2c}{nS} \sum_{i=1}^n \int_S \left\{ E_i\left(-\frac{|X - X_{N_i}|}{L_1}\right) - E_i\left(-\frac{|X - X_{N_i}|}{L_2}\right) \right\} dX \end{aligned} \quad (8)$$

ここで $E_i(\cdot)$ は積分指数関数を表す。

$$E_i(x) = - \int_{-x}^0 \frac{\exp(-t)}{t} dt \quad (9)$$

$L_1, L_2$ はそれぞれLのとりうる最小値と最大値である。Lについての積分区間の下限 $L_1$ は0をとり、上限 $L_2$ としては、式(3)の値が対象領域内の任意の2地点間距離 $h$ でほぼ1であり、その結果、式(8)の値の差が実質上領域内の $n$ 個の観測地点の位置に依存しなくなる値にとった。ここでは $h$ が対象領域内における任意の2地点の最大長さ $h_1$ をとるとき、式(3)の値が約0.99となるLの値である $L_2=100h_1$ とした。

本研究で採用した評価関数の導出ではいくつかの仮定を置いている。まず、「場の統計的弱定常性等方性」については、ある場所が特に高濃度である可能性が高いという事前情報がある場合（統計的弱定常性の仮定が不適当）や、特にある方向への汚染濃度の相関が強いという事前情報がある場合（等方性の仮定が不適当）を除いて、場の分布の統計的性質として無理のない仮定であると考えている。また実際、筆者らにはある実際の汚染サイトにおいて統計的弱定常性等方性の仮定の下で濃度分布を統計的に解析し、仮定が妥当であったという結果を得た経験がある。場の構造としての相関関数形の選択については、現実には真の相関関数形を求めるなどということは不可能であり、観測データから求めたいいくつかの距離ごとの相関係数の値にどのような関数をフィッティングさせるか、という問題になる。この場合、本研究で仮定した「指数型相関関数形」は、相関の強さを表すパラメータであるLの値を変えることによって、いくらかの誤差を許容すれば、現実の観測値から求められる距離ごとの相関係数の値の分布を近似するためのかなりな自由度を有していると考えられる。また、このことから実際、指数型相関関数形は2次元確率場の解析のためによく使用されている。また、無情報事前分布としてLの「対数値が一様分布する」という仮定については、ベイズ統計学において様々な種類のデータに基づく経験から得られた仮定と考えられ、本研究のようにLの分布についての情報が、正の値をとるといふ以外存在しない段階においては妥当な仮定と考えている。しかし「指数型相関関数形」の仮定や「対数値が一様分布する」という仮定については、今後、土壌汚染データの統計的性質の解析を数多く経験することによって、特に土壌汚染データの解析において妥当なものが求められる可能性があり、その場合には式(8)の形は異なってくると考えられる。

### 3. 最適配置探索の手法

式(8)で示される評価関数の値を最小にする試料採取地点の取り方を求める手法として、本研究では、遺伝アルゴリズム (GA) を利用した最適化アルゴリズムを用

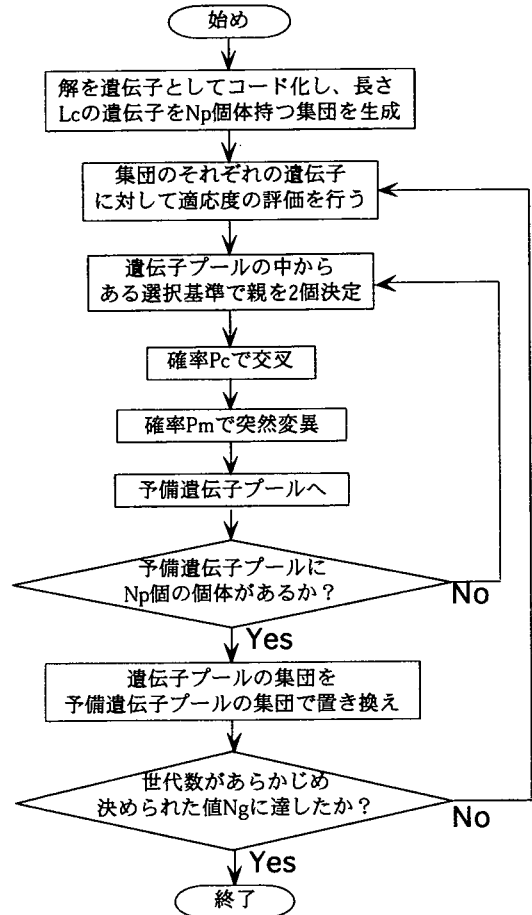


図-1 GAのフローチャート

いて解析した。

#### (1) 遺伝アルゴリズム (GA)<sup>17)</sup>

GAは、生態系における遺伝と進化のメカニズムを適当に簡略化し、それをコンピュータ上で模倣することによって最適解に到ろうとする手法である。GAはまだまだ発展途上の研究領域ではあるが、その特徴として極めて広範囲の複雑な最適化問題に適用可能であるという利点を持っており、問題解決に対する有効性が示されている。近年、サンプリング地点の選択手法としてGAを用いた研究もいくつか出てきている。

本研究では米田らの研究<sup>7)</sup>で採用されている図-1に示すフローチャートに従って、GAによる試料採取地点の最適配置探索を行った。以下に図-1中の操作について説明を加える。

##### a) 解のコード化

まず、対象とする調査領域を多くの小さなメッシュに分割し、各節点に番号をつける。そして、これらの節

点の中から  $L_c$  個の試料採取地点を選択することとし、この試料採取地点番号の並びを遺伝子とする。

**b) 初期個体集団の生成**

長さ  $L_c$  の遺伝子 ( $L_c$  個の試料採取地点番号の並び) を持つ個体をランダムに  $N_p$  個生成し遺伝子プールの集団とする。

**c) 評価関数の計算**

集団のそれぞれの遺伝子に対して、適応度を評価する (評価関数を計算する)。

**d) 選択・交叉・突然変異**

以下のステップ (i) (ii) (iii) (iv) を、 $N_p$  個の子孫が生成されるまで繰り返す。

**(i) 親の選択**

遺伝子プールの中から、ある選択基準で親を 2 個決める (適応度の大きい個体ほど選択される確率が大きくなるように設定する)。

**(ii) 交叉**

選択された 2 つの親の遺伝子の中で、確率  $P_c$  により、ランダムに選択された位置で遺伝子型を入れ替えて、2 つの子孫を生成する。

**(iii) 突然変異**

それぞれの子孫について 1 遺伝子当たり確率  $P_m$  により、ランダムな位置の番号をランダムな値に変化させる。

**(iv) 得られた遺伝子を予備遺伝子プールの中に入れる。**

**e) 世代の進行**

遺伝子プールの集団を、予備遺伝子プールの集団で置き換え、その後 c) に戻る。この操作により、世代が 1 つ進んだことになる。

d) (i) の操作における親の選択方法としては米田らと同じく、ランク方式とエリート主義戦略を採用した。ランク方式とは、各個体を適応度によってランク (順位) づけし、それらが選択される期待値を実際の適応度ではなくあらかじめ各ランクに対して決められた値によって決定するものである。本研究では適応度の順位の逆数に比例した期待値をもつようにした。エリート主義戦略とは、最良個体が選択・交叉・突然変異により失われてしまうことを防止するために、各世代における最良個体のいくつかを強制的に次の世代に残すものである。本研究では、最良個体 1 つについてエリート主義戦略の対象とした。さらに初期個体集団の中に、現行 5 地点混合方式に倣い図-2 に示す配置の遺伝子を 1 つ入れておくことにより、エリート主義戦略によって悪くとも図-2 に示す配置より良い試料採取地点配置が求まることを保証した。

**(2) 最良地点逐次移動法**

本研究ではまた、ニューラルネットワーク (N.N.) ア

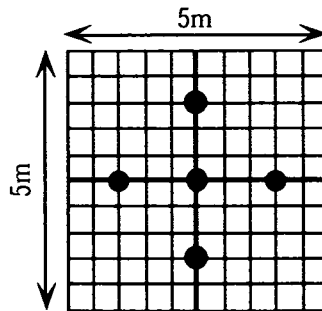


図-2 5地点混合方式に倣った採取地点配置

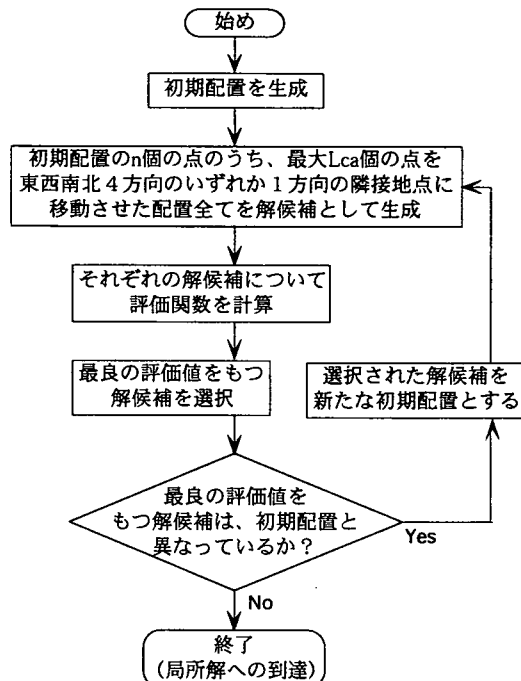


図-3 最良地点逐次移動法のフローチャート

ルゴリズム<sup>10)</sup>などにおける最適手法として用いられる最急降下法にヒントを得た配置探索アルゴリズム (図-3 参照) を構築した。本研究ではこの配置探索法を「最良地点逐次移動法」と呼ぶこととする。

以下に図-3 中の操作について説明を加える。

**a) 初期配置の生成**

まず、対象とする調査領域内に試料採取地点の候補となる節点を設定し、各節点に番号をつける。そして、これらの節点の中から  $L_c$  個の試料採取地点をランダムに選択する。

**b) 解候補の生成**

初期配置の  $L_c$  個の点のうち、1 回の動作において移動可能な採取地点の最大数を  $L_{ca}$  個とし、1 個から  $L_{ca}$  個の採取地点を、隣接する節点のいずれか 1 つに移動

させた全ての組み合わせからなる集団を生成する。この集団が、最適配置の候補の集合となる。

c) 評価関数の計算

b) で生成したそれぞれの候補について、評価関数を計算する。

d) 候補の選択

候補の集合の中から、評価値の最も優れている候補を1つ選択する。

e) 初期配置の更新

d) で選択された候補の評価関数の値と初期配置の評価関数の値を比較する。選択された候補の評価関数の値のほうが良い評価関数の値を持っているなら、その候補を新たな初期配置としてb)に戻る。初期配置のほうが良い評価関数の値を持っている場合は、この探索がそこで局所解に到達したことを示す。そこで探索を終了する。

(3) ハイブリッド遺伝アルゴリズム

種の適応を扱うGAでは、解空間上に複数の探索点を設定して大域的な探索を行うため、最適解の周辺には早く近づくが、局所探索能力が弱いという問題が一般に指摘されている。一方、最良地点逐次移動法は一個体ごとの局所探索を行うため局所解への到達は非常に早いですが、大域的最適解への到達に失敗して局所的最適解に陥ってしまう可能性が大きいという欠点がある。

本研究では、この2つの手法が相互補完的な特徴を有している点を利用し、2つの手法を組み合わせることによって、局所的最適解に陥ることなく、より早く大域的最適解に到達することを目指したハイブリッドアルゴリズム(図-4参照)を構築した。

ハイブリッド(1)は、生成された初期配置それぞれに対して最良地点逐次移動法による探索を行い、得られた配置の集合に対してGAを1世代分ずつ、この動作を繰り返す、というものである。

ハイブリッド(2)は、まずGAを何世代か進ませて最適解の近傍への大まかな収束を行い、GA終了後の遺伝子プールの集団の中で評価関数の値の優れた上位いくつかの個体を初期配置として最良地点逐次移動法による局所探索を行う、というものである。本研究ではGA終了後の上位10個体について最良地点逐次移動法による探索を行った。

4. 各手法の有効性の比較

本研究ではまず、5m四方の正方形領域を0.5m四方のメッシュに区切って11×11=121個の節点を生成し、その中から $L_c=5$ 個の試料採取地点を選択する問題により、

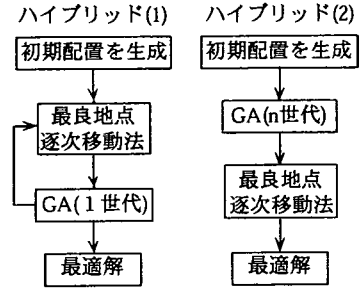


図-4 本研究で用いたハイブリッド遺伝アルゴリズムのアイデア

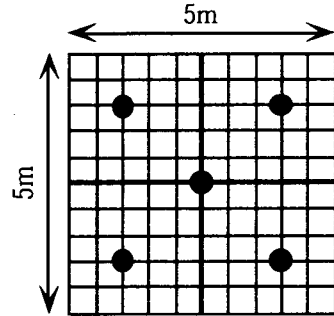


図-5 5m四方領域における最適配置

アルゴリズムの性能を評価することにした。これは、試料採取地点の配置の対称性を無視すれば ${}_{121}C_5 = 198,792,594$ 通りの組み合わせの中から1つの最適解を選ぶ問題となるが、すべての組み合わせについて評価関数を計算した結果、最適配置は図-5のようになった。

GAと最良地点逐次移動法において、それぞれ異なる乱数系列を用いた10回の探索を行い、GAについては各世代における評価関数の値と計算コスト(本研究では、計算時間の大部分を占めると考えられる評価関数の計算回数を計算コストとする)それぞれ10個の平均をそれぞれの期待値とした。最良地点逐次移動法においては、10通りの計算コストの増加に伴う評価関数の減少を内挿する曲線を求め、この10本の曲線の平均を期待値の変化とした。また別に最適配置の場合の評価関数の値も計算し、それと各手法で求めた評価関数の期待値との差を縦軸にとり、横軸に計算コストの期待値をとった図を作成することにより各手法の有効性について検討した。

(1) GAによる最適配置探索

個体数 $N_p=100$ として、いくつかの交叉確率 $P_c$ と突然変異確率 $P_m$ の組み合わせによるGAを用いた最適配置探索の結果を図-6に示す。いずれのパラメータ設定においても図-2型配置の評価関数の値(最適配置のものとの差)0.0727より評価関数の値を著しく減少させてはいるものの、最適配置の発見には至って

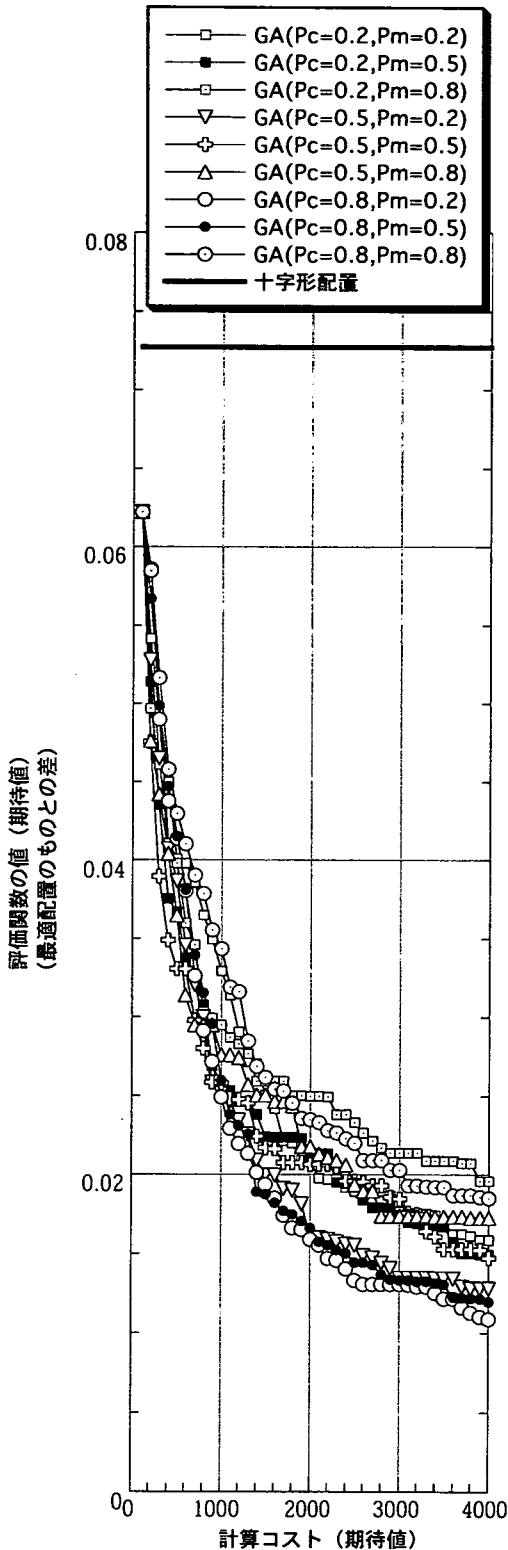


図-6 GAによる最適配置探索結果 (5m四方正方形領域)

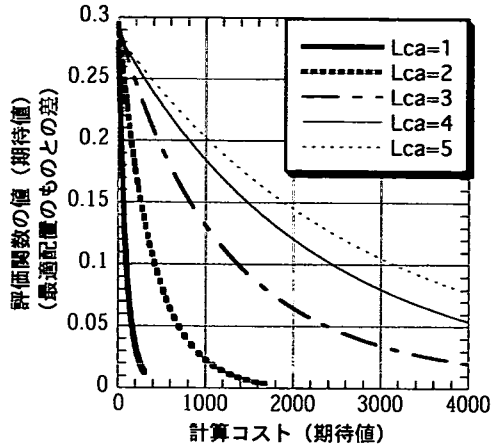


図-7 最良地点逐次移動法による最適配置探索結果 (5m四方正方形領域)

いない。また、 $(Pc, Pm) = (0.8, 0.2)$  のものが計算コストが1000以上で最もよく評価関数の値を下げているため、以後のハイブリッドアルゴリズムにおけるGA部分のパラメータとして $(Pc, Pm) = (0.8, 0.2)$ をとった。

### (2) 最良地点逐次移動法による最適配置探索

最良地点逐次移動法においては、それぞれの $Lca$ の値ごとに最適配置の探索を行った。図-7より、 $Lca$ の値が増加するに従い計算コストが増大することがわかる。本研究では少ない計算コストで局所解に到達する点を重視して $Lca=1$ を以後のハイブリッドアルゴリズムにおけるパラメータとして採用した。

### (3) ハイブリッド(1)による最適配置探索

個体数 $Np=5, 10$ としてGAの世代数を変化させた場合のハイブリッド(1)による最適配置探索を行った結果を図-8に示す。図-8には比較のために、 $Np=100$ でのGAのみの場合と、 $Lca=1$ での最良地点逐次移動法のみの場合の結果も示している。GAあるいは最良地点逐次移動法を単独で行う場合に比べて、ハイブリッド(1)では評価関数の値をさらに減少させており、特に個体数 $Np=10$ の場合は最適配置の探索に成功している。

### (4) ハイブリッド(2)による最適配置探索

個体数 $Np=10, 50, 100$ としてGAの世代数を変化させた場合のハイブリッド(2)による最適配置探索を行った結果を図-9に示す。図-8同様にGAのみの場合と最良地点逐次移動法のみの場合の結果も示している。図-9よりハイブリッド(2)においては個体数の増加が探索結果の向上に必ず結びついているとは言い難く、また同じ個体数でもGAの世代数の増加による

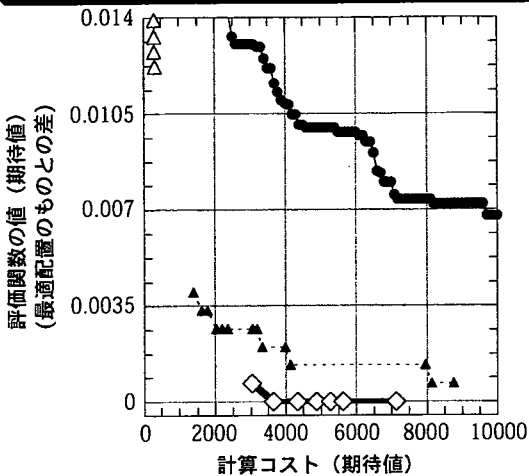
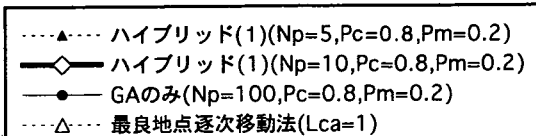


図-8 ハイブリッド(1)による最適配置探索結果 (5m四方正方形領域)

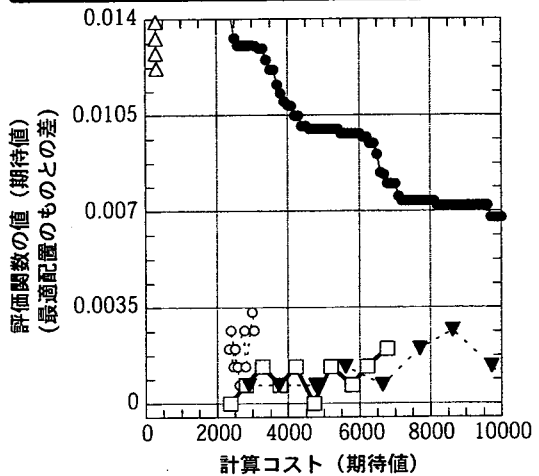
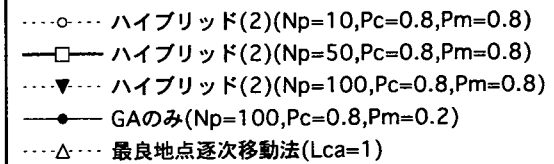


図-10 ハイブリッド(2)( $P_m=0.8$ )による最適配置探索結果 (5m四方正方形領域)

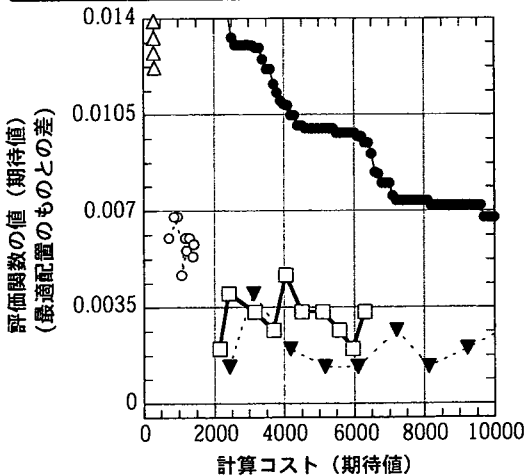
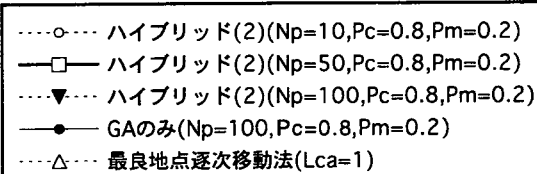


図-9 ハイブリッド(2)( $P_m=0.2$ )による最適配置探索結果 (5m四方正方形領域)

計算コストの増加が必ずしも探索結果の向上につながっていないことがわかる。ハイブリッド(1)と同様こちらでもGAあるいは最良地点逐次移動法を単独で行う場合に比べて評価関数の値を大幅に減少させているものの、いずれのパラメータ設定においても最適配置の探索に失敗するケースが見られた。

GAにおいてエリート主義戦略をとると、その時点での最良解であるエリート個体の遺伝子が集団中に急速に広がる可能性が高いため探索の幅が狭くなり、局所解に陥る危険性が大きくなる。そこでこの弊害を防ぐために、突然変異確率 $P_m$ の値を0.2から0.8へと増加させて、ハイブリッド(2)による同様の探索を行った。その結果を図-10に示す。図-10には図-8,9同様にGAのみの場合と最良地点逐次移動法のみの場合の結果も示している。 $P_m=0.2$ の場合と比べて全体的に探索結果が向上し、さらに、ハイブリッド(1)よりも少ない計算コストで最適配置の探索に成功しているものもあった。

#### (5) アルゴリズムの性能の比較・検討

以上の結果より各アルゴリズムの性能を比較すると、GAと最良地点逐次移動法をハイブリッド化させることにより探索能力は向上しており、さらにハイブリッド(2)に比べてハイブリッド(1)のアルゴリズムによる探索のほうが比較的少ない計算コストで安定して

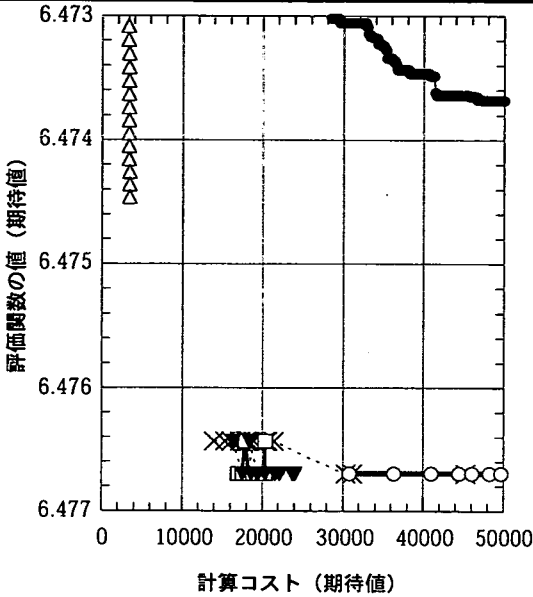
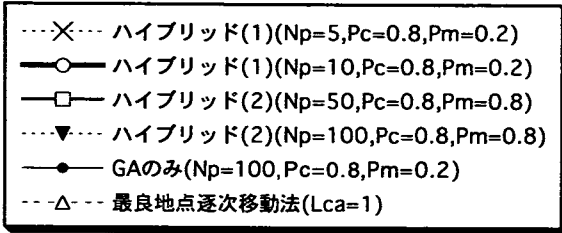


図-11 ハイブリッドアルゴリズムによる最適配置探索結果 (50m四方正方形領域)

最適解を得ている。しかしこの原因として、対象領域が5m四方と小さく、アルゴリズムの初期において局所探索を行うハイブリッド(1)のほうがアルゴリズムの初期において大域的探索を行うハイブリッド(2)に比べて有利な条件であった可能性が考えられる。そこで、大きな領域として50m四方の正方形領域を0.5m角のメッシュに分割し、同様に5地点の最適配置探索を行った結果を図-11に示す。ただし、この領域における最適配置は計算時間の制約から求められなかったため、式(8)で示される評価関数のうち、場の相関構造のみに依存して採取地点の配置に依存しない右辺第1項を除いた式に、 $-1$ を乗じたものを縦軸にとった。図-11で、ハイブリッド(1)はある程度以上の計算コストをかければ安定して良好な解が得られている。一方、ハイブリッド(2)においてはより少ない計算コストで良好な解を得ることが多いものの、図-9、図-10と同様にGAの世代数の増加による計算コストの増加が必ずしも探索結果の向上につながらない不安定さが見られた。以上の最適配置探

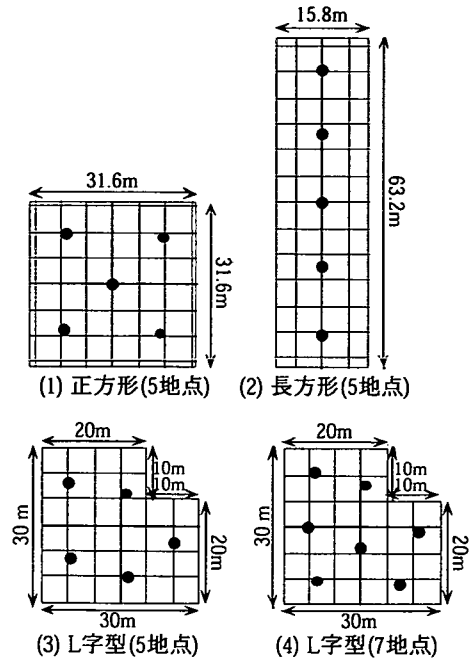


図-12 種々の形状の領域における最適配置探索の結果

索結果から2種類のハイブリッドアルゴリズムの優劣を一概に決定するのは難しいが、比較的大きな領域においてもハイブリッド(1)の有効性が確認されたため、本研究では良好な解への到達の安定性を重視してハイブリッド(1)( $N_p=10, P_c=0.8, P_m=0.2$ )のアルゴリズムを次章での最適配置探索に用いることとした。

## 5. 任意形状の領域における最適試料採取地点の探索

まず、比較的単純な数種類の形状の領域についてハイブリッド(1)のアルゴリズム(個体数 $N_p=10$ ,世代数 $N_g=5$ )を用いた最適配置の探索を行った。その結果を図-12に示す。図-12の(1)は現行の5地点混合方式と同じく面積が $1000m^2$ の正方形領域における5地点の配置であるが、領域の中心に1点、残りの4点是对角線上にそれぞれ中心から約13m離れた配置となった。(2)は(1)の領域の縦方向の長さを2倍、横方向の長さを0.5倍にした領域における配置である。この形状の領域での最適配置は、(1)のような対角線上ではなく、領域の縦方向の中央線上に、領域の中心に1点、残りはそれぞれ12.5m~13m間隔で並んだ。(3)は図のようなL字型の領域における5地点の配置であり、(4)は(3)と同じ形状の領域において、7地点での配



置の探索を試みた結果である。

さらに、対象領域がより複雑な地形をしている場合にも適用可能な最適配置探索システムを構築した。その手法の概要を図-13に示す。

#### a) 対象領域の平面図の取り込み

対象領域の平面図をスキャナ等でパソコンに読み込ませる。

#### b) 試料採取可能な領域と不可能な領域との区別、縦横の節点数の設定

読み込んだ平面図を画像編集ソフト（本研究では「Adobe Photoshop 5.5」を使用）により開き、試料採取可能な領域と不可能な領域とを塗り分け、それをモノクロ2階調化する。この操作により例えば、試料採取可能な領域を黒色で、不可能な領域を白色で表す。また、画像編集ソフトにより画像の縦横のピクセル数を設定することができるが、この縦横のピクセル数がそのまま領域の縦横の節点数となるため、希望の節点数・接点間隔が得られるよう縦横のピクセル数を設定する。

#### c) デジタル化（2値表現）

モノクロ2階調化した平面図をファイルコンバートソフト（本研究では「GraphicConverter」(フリーソフト)を使用)によりASCII形式に変換することで、試料採取可能な領域（黒色の部分）の中の節点を「1」、不可能な領域（白色の部分）の中の節点を「0」で表す。

#### d) 最適配置探索プログラムへの読み込み

以上の処理を施した対象領域の平面図を、本研究における最適配置探索プログラムに読み込ませる。

#### e) 最適配置探索アルゴリズムの実行

上記(c)の手順において「1」で表された節点を探索候補地点として最適配置探索アルゴリズムを実行する。

この手法の適用例として、対象とする領域が道路などにより分断されているケースでの最適配置探索の結果を図-14に示す。

図-12、図-14から、いずれの領域形状、いずれの地点数においても採取地点の最適配置は概ね領域内に均等に分布することがわかる。通常、領域内の汚染分布に関する情報が全く与えられていないケースにおいて主観的に採取地点配置を決定する場合、採取地点を領域内に均等に分布させるのは妥当であると思われるが、その配置は人によって異なりうる。本研究の手法による最適配置では、同様に採取地点が領域内に均等に分布する傾向を持っているが、本研究の手法はそのような地点配置を客観的基準に基づいて一意的に決定することができるという大きな利点を持っている。

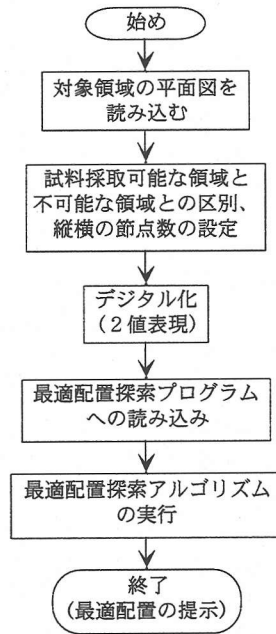
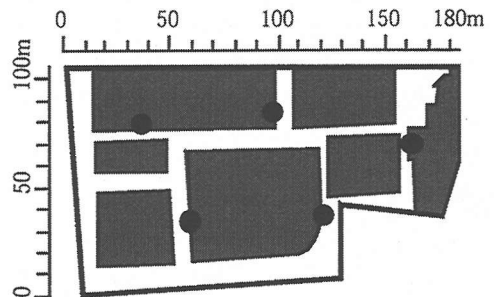
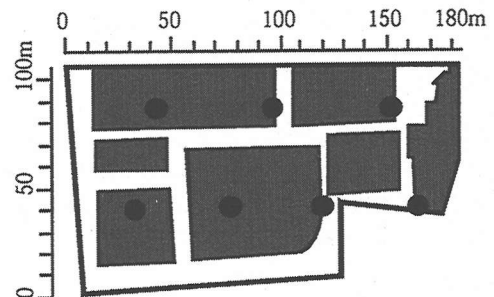


図-13 最適配置探索アルゴリズムの任意地形への適用手法



(1) 5地点の配置



(2) 7地点の配置

図-14 任意形状の領域における最適配置の例（灰色の部分が生料採取可能な領域）

## 6. 結論

本研究では無情報条件下での土壌汚染概況調査における試料採取地点の最適配置を決定するために、領域全体での空間平均値を有限個の観測値の平均値で推定する場合の推定分散に基づいた評価関数を設定した。そして、GAと最良地点逐次移動法を組み合わせたハイブリッドアルゴリズムを構築し、各手法の有効性の比較を行った結果、最良地点逐次移動法の適用結果の集合に対してGAを1世代進め、それを繰り返すハイブリッド(1)のアルゴリズムにより最も安定して良好な解を得た。このアルゴリズムを用いて種々の形状の領域における最適配置探索を行い、さらに任意の領域形状、任意の採取地点数によるサンプリング地点配置の探索が可能なシステムを構築し、複雑な形状の領域への適用を試みた。これらの結果より、試料採取地点は領域内に概ね均等に分布することを確認した。

本研究で用いた評価関数は、汚染物質濃度の平均値を有限個の観測値から最も有効に推定するためのものである。今後はさらに最適配置探索アルゴリズムの性能を向上させるとともに、別の評価関数を用いて汚染の空間分布を最も有効に推定するための配置を求める方法や、汚染物質分布に関する事前情報の取り扱いなども考慮したより応用性の高いサンプリング戦略を決定する手法の確立を目指していく。

## 付録 推定分散式の導出

水平2次元空間で分布する確率変数 $Z(X)$ について、 $Z_k$ を $Z(x_{k,i})$  ( $i=1,2,\dots,k$ )の $k$ 個の地点での算術平均、 $Z_N$ を $Z(x_{N,j})$  ( $j=1,2,\dots,n$ )の $n$ 個の地点での算術平均とし、 $Z_k$ を $Z_N$ で推定するものとする。このとき、有限個の測定地点での平均値 $Z_N$ の期待値は $Z_k$ の期待値に一致する。 $Z(X)$ の期待値が場所によらず一定である場合、以下の式が成り立つ。

$$E[Z_N] = E[Z_k] = m \quad (a-1)$$

ここで $E[\ ]$ は期待値操作を意味し、 $m$ は定数である。

推定分散を $Z_k$ と $Z_N$ との差を2乗したものの期待値として定義すると、推定分散式は以下のように導かれる。

$$\begin{aligned} \sigma_E^2 &= E\left[(Z_k - Z_N)^2\right] \\ &= E[Z_k^2] + E[Z_N^2] - 2E[Z_k Z_N] \end{aligned} \quad (a-2)$$

ここで、共分散関数について以下の式を導く。

$$\begin{aligned} C(h) &= C\left(|X_{k,i} - X_{k,j}|\right) \\ &= \text{Cov}\left(Z(X_{k,i}), Z(X_{k,j})\right) \\ &= E\left[\left(Z(X_{k,i}) - E[Z(X_{k,i})]\right)\left(Z(X_{k,j}) - E[Z(X_{k,j})]\right)\right] \\ &= E\left[Z(X_{k,i}) \cdot Z(X_{k,j})\right] - E[Z(X_{k,i})] \cdot E[Z(X_{k,j})] \end{aligned} \quad (a-3)$$

ここで $|x_{k,i} - x_{k,j}|$ は $X_{k,i}, X_{k,j}$ 間の距離を表す。

式(a-3)に式(a-1)を代入すると、

$$C\left(|X_{k,i} - X_{k,j}|\right) = E\left[Z(X_{k,i}) \cdot Z(X_{k,j})\right] - m^2 \quad (a-4)$$

式(a-4)を利用すると、式(a-2)の右辺の各項は以下のように表される。

$$\begin{aligned} E[Z_k^2] &= E\left[\frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k Z(X_{k,i}) Z(X_{k,j})\right] \\ &= \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k E\left[Z(X_{k,i}) Z(X_{k,j})\right] \\ &= \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k \left\{ C\left(|X_{k,i} - X_{k,j}|\right) + m^2 \right\} \\ &= \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k C\left(|X_{k,i} - X_{k,j}|\right) + m^2 \end{aligned} \quad (a-5)$$

同様に、

$$E[Z_N^2] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n C\left(|X_{N,i} - X_{N,j}|\right) + m^2 \quad (a-6)$$

$$E[Z_k Z_N] = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n C\left(|X_{k,i} - X_{N,j}|\right) + m^2 \quad (a-7)$$

式(a-5),(a-6),(a-7)を式(a-2)に代入すると、

$$\begin{aligned} \sigma_E^2 &= \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k C\left(|X_{k,i} - X_{k,j}|\right) \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n C\left(|X_{N,i} - X_{N,j}|\right) \\ &\quad - \frac{2}{kn} \sum_{i=1}^k \sum_{j=1}^n C\left(|X_{k,i} - X_{N,j}|\right) \end{aligned} \quad (a-8)$$

いま、 $X_{k,i}$  ( $i=1,2,\dots,k$ ) は領域全体に分布しているとし、 $k \rightarrow \infty$ の極限を考えると、 $Z_k$ は領域全体 $S$ での空間平均と考えることができる。つまり、

$$Z_k = \frac{1}{S} \int_S Z(X) dX \quad (a-9)$$

この積分は水平2次元の領域全体 $S$ での面積分を表

す。このとき式(a-8)は次式のようになる。

$$\begin{aligned} \sigma_E^2 = & \frac{1}{S^2} \int_S dX \int_S C(|X - X'|) dX' \\ & + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n C(|X_{N_i} - X_{N_j}|) \\ & - \frac{2}{nS} \sum_{i=1}^n \int_S C(|X - X_{N_i}|) dX \end{aligned} \quad (a-10)$$

式(a-10)で表される $\sigma_E^2$ が、領域全体での空間平均値を有限個の観測値の平均値で推定した場合の推定分散となる。

#### 参考文献

- 1) 環境庁水質保全局水質管理課土壌農業課監修：土壌・地下水汚染対策ハンドブック，p.114，公害研究対策センター，1996.
- 2) 米田稔，森澤眞輔，西村留美：土壌汚染概況調査の5地点混合方式における試料採取地点最適配置，土木学会論文集，No.622/VII-2，pp.51-58，1999.
- 3) 木内智明，米田稔，大塚順基，森澤眞輔：土壌汚染概況調査における試料採取地点の最適配置，土木学会第54回年次学術講演会講演概要集第7部，pp.546-547，1999.
- 4) Carrera, J., Usunoff, E. and Szidarovszky, F.: A method of optimal observation network design for groundwater management, *Journal of Hydrology*, 73, 147-163, 1984.
- 5) Rouhani, S.: Variance-reduction analysis, *Water Resource Research*, 21, 834-846, 1978.
- 6) Meyer, P. D. and Brill, E. D.: A method for locating wells in a groundwater monitoring network under uncertainty, *Water Resource Research*, 24, 1277-1282, 1978.
- 7) Tucciarelli, T. and Pinder, G.: Optimal data acquisition strategy for

the development of a transport model for groundwater remediation, *Water Resource Research*, 27, 577-588, 1991.

- 8) McKinry, D. C. and Loucks, D. P.: Network design for predicting groundwater contamination, *Water Resource Research*, 28, 133-147, 1992.
- 9) Andricevic, R.: Coupled withdrawal and sampling design for groundwater supply models, *Water Resource Research*, 29, 5-16, 1993.
- 10) James, B. R. and Freeze, R. A.: The worth of data in predicting aquitard continuity in hydrogeological design, *Water Resource Research*, 29, 2049-2065, 1993.
- 11) Meyer, P. D., Valocchi, A. J. and Eheart, J. W.: Monitoring network design to provide initial detection of groundwater contamination, *Water Resource Research*, 30, 2647-2659, 1994.
- 12) James, B. R. and Gorelick, S. M.: When enough is enough: The worth of monitoring data in aquifer remediation design, *Water Resource Research*, 30, 3499-3513, 1994.
- 13) Andricevic, R.: Evaluation of sampling in the subsurface, *Water Resource Research*, 32, 863-874, 1996.
- 14) Cieniawski, S. E., Eheart, J. W. and Ranjithan, S.: Using genetic algorithms to solve a multiobjective groundwater monitoring problem, *Water Resource Research*, 31, 399-409, 1995.
- 15) Wagner, B. J.: Sampling design methods for groundwater modeling under uncertainty, *Water Resource Research*, 31, 2581-2891, 1995.
- 16) Journel, A.G. and Huijbregts, CH.J.: *Mining Geostatistics*, p.35, Academic Press, 1978.
- 17) (例えば) L. デービス：遺伝アルゴリズムハンドブック，森北出版，1997.
- 18) (例えば) 馬場則夫，小島史男，小澤誠一：ニューラルネットの基礎と応用，共立出版，1994.

(2001.3. 16 受付)

## SEARCH FOR THE OPTIMAL ARRANGEMENT OF SAMPLING POINTS IN THE INITIAL SURVEY OF SOIL POLLUTION USING HYBRID GENETIC ALGORITHM

Tomoaki KIUCHI, Minoru YONEDA, Shinsuke MORISAWA and Junki OTSUKA

Supposing noninformative prior distribution of a parameter which expressed the strength of spatial correlation, the evaluation function was deduced in order to search for the optimal arrangement of sampling points in the initial survey of the general condition of soil pollution without any prior information. Then, we developed the optimization algorithms which used the genetic algorithm (GA), the method of stepwise movement to the best point (SMBP) and two types of their hybrid algorithms. The results showed that hybrid algorithm (1), which carry out searches by SMBP for all initial arrangements and then carry out GA using local solutions obtained by the SMBP, got the most stably good performance. Using the algorithm, some optimum solutions were searched, and the results showed that sampling points were approximately equally distributed in the site.