

遺伝的プログラミングによる 潮位偏差の予測に関する研究

STUDY ON PREDICTION OF SEA LEVEL ANOMALY USING GENETIC PROGRAMMING

西田 渉¹・松田 良寛²・野口 正人³

Wataru NISHIDA, Yoshitomo MATSUDA and Masato NOGUCHI

¹正会員 博士(工) 長崎大学助教授 工学部社会開発工学科 (〒852-8521 長崎県長崎市文教町1-14)

²学生会員 長崎大学大学院 生産科学研究科博士前期課程 (同上)

³フェロー 工博 長崎大学教授 工学部社会開発工学科 (同上)

Recently, as the progress of information- communication technologies, lots of valuable data related to water environment are automatically observed and some of them are opened to public in real time through media such as internet etc. It is expected that these data will contribute to the solving problems occurred at real water areas.

In this study, regarding the facts that sufficient data are recently available to realize the data- driven modelling, genetic programming was newly developed and its applicability to the prediction of sea level departure was discussed with comparing to the results obtained from artificial neural network model. Although the training of genetic programming was executed for limited data, construction of tree was successfully established by regarded genetic procedures. Obtained tree considerably simulates the temporal changes of sea level anomaly observed at Kuchinotsu port. While, for obtaining the tree which outputs results with smaller errors, computational conditions, such as observed data fed to model and parameters on genetic procedures, have to be chosen with care.

Key Words : *data-driven model, genetic programming, artificial neural network, sea level anomaly*

1. はじめに

河口や沿岸域の流動状況を予測する手法としては、所定の支配方程式を基礎とする数値モデルを数値シミュレーションする手法が最も一般的であり、現地への適用をとおして、これまでに多くの成果を上げてきている。ただ、予測精度の高い結果を得るためには、サブグリッドスケールの物理現象のパラメタリゼーションによって導入された各種のモデル係数を適切に評価しておく必要がある。これらの係数値については、遺伝的アルゴリズム(GA)やAdaptive Cluster Covering Method などの大域的最適化手法による推定が試みられてきているが⁽¹⁻⁴⁾、値の算出に相応の計算時間を要することもあり、実際には限られた一定期間の実測情報に基づいて現地への適用性が議論されるに止まっている。

ところで、近年、ICT (Information- Communication Technology)の進展に伴って、実水域における各種の物理現象が実時間的にモニタリングされると共に、実測値

が電子データとして蓄積されてきている。これらの情報の活用について、モデル解析の範囲で言えば、物理モデルの検定、検証に用いられている。その一方で、実現象の予測にあたっては、これら貴重なデータに基づいて、計算機に実現象の傾向を自己学習させ、実用的なモデルを構築させていくData-Driven Model (DDM)による手法も、蓄積データ数の増加によって実現可能になってきていると言えそうである。DDMは、比較的長期間の実測値を使って、実証的にモデルが生成されるため、その長期予測への適用が期待される。もちろん、こうした手法の予測結果は実測値の精度に依存するため、計算に際しては実測誤差が可能な限り除去されておく必要があることは言うまでもない。

DDMの代表例としてニューラル・ネットワークモデルが上げられるが、本研究では、進化論的手法に基づくDDM手法の一つとして、水工学分野においても、その利用が始められている遺伝的プログラミング手法を用いることとし、これを潮位偏差の予測に適用を試みることにした^(5),6)。具体的には、遺伝的プログラミングの評価結

果をニューラル・ネットワークによる結果と比較しながら計算精度の検討を行なった．また，長期変化の予測に本手法を適用する際の問題点についても検討した．

2．DDMの概要

ここで，本研究で用いた二つのモデルの概要について記しておく

(1) 遺伝的プログラミング

遺伝的プログラミングは，遺伝的アルゴリズムと同様に，進化論的思考に基づく計算手法の一つであり，生物の遺伝子の複製，選択・淘汰といった自然界で観察されるメカニズムを基本として構成されている．ただし，遺伝的プログラミングでは，Machine Learning によってプログラムの自動合成を目指している点が遺伝的アルゴリズムにはない特徴の一つとなっている⁶⁻⁸⁾．

ここで，本研究で用いられた遺伝的プログラミングの処理手続きを示すと，図-1のとおりである．今回の計算では，各個体は，入力データに対する関数となるように表現するものとし，その木構造をLisp言語のS式で表わすようにしている．処理手続きについては，まず，個体群として M 個の個体をランダムに生成し，各個体の表現木から得られた計算値と実測値から適合度を評価する．続いて，個体の適合度の良否に基づいて，個体群に遺伝オペレータを作用させ，次世代の個体群を生成させるものとした．なお，この処理の中で適用されるオペレータは，GAでも標準的に適用される選択操作と遺伝子操作（突然変異，交叉）としている．その内，突然変異では，表現木の終端・非終端ノードラベルの変更と，部分木の新規生成を行っている．交叉では，個体群から選択された親となる1対の表現木に対して，ルートノードを除く部分木の取り替えを実行している．この取替え位置の深さに関しては，部分木が効果的に積み上げられるよう，浅い層のノードが選択される確率が高くなる深さ依存型交叉とした．本計算では，表現木を固定長として取り扱うものとしていることから，交叉による部分木の取り替えは，同じ階層にあるノード間でなされた．部分木の削除は，終端ノードが全て零で構成される部分木に入れ替える突然変異によって実現させるものとした．

表現木に導入した終端記号と非終端記号は，表-1のとおりである．すなわち，計算対象となる潮位偏差は，S式で構成される表現木で算出される．また，終端記号としては，潮位偏差に影響をもたらすと考えられる気象学的項目として，現時刻と過去6時間の大気圧，時間降雨量，緯度・経度方向の風速を取り上げることとし，これらに加えて，表現木に含まれる係数値として，ランダムに発生させた定数が用いられた．なお，気象学的項目については，各項目の最大値で基準化されており，係数値は-1から1までの実数とされている．非終端記号として

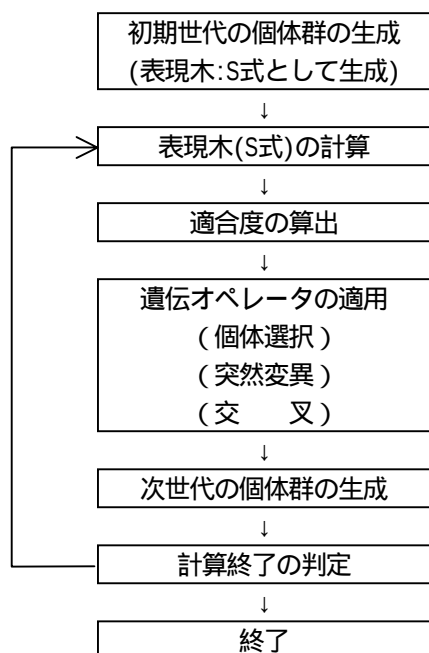


図-1 遺伝的プログラミングの処理手続き

表-1 終端記号と非終端記号

終端記号	気象学的時系列データ 大気圧，時間降雨量，風速 [当該時刻と過去6時間の1時間 毎の値を最大値で基準化]
	定数値 [-1～0,0～1の実数を初期に200 個づつランダムに生成]
非終端記号	+, * LispのS式で(+ A B), (* C D) と記述される．A,B,C,D は終端記号または部分木

表-2 遺伝的プログラミングの設定条件

設定項目	値
表現木の最大層数	7
1世代の個体数(M)	1000
エリート選出確率(P_{er})	0.15
交叉確率(P_{cr})	0.7
突然変異確率(P_{mt})	0.7

は，当初，四則演算子並びに累乗・指数関数としていたが，予備計算の結果から，解の収束が進み難かったこと，また係数値に負の値を含めたことから，{+ *}を用いた．これらは，いずれも二つの引数を持つ関数記号である．

各表現木の適合度については，算出値と実測値を用いて，以下のように評価することにした．

$$ER = \frac{1}{N} \sum_{n=1}^N w_{obs,n} |OBS_n - CAL_n|^2 \quad (1)$$

ここに， N ：実測データの総数， $w_{obs,n}$ ：実測データの番号 n に対する重み， OBS_n ， CAL_n ：それぞれ実測データ

と算出値，である． $w_{obs,n}$ の値については，基本的に 1 としたが，台風の影響と見られる潮位偏差の増加があったものは，幾つかの試行計算の結果を参考に 1.2 とした．また，モデルに入力する時系列データを入手できなかった期間については上記の重みを零とすることで適合度の評価対象から外している．

ここで，以降の計算で用いた遺伝的プログラミングの設定条件を表-2にまとめておく．一個体の表現木の最大層数は 7 層とした．その結果，最下層には 128 個の終端ノードが配されている．個体群は，1000 個の個体で形成されており，個体の選択は，適合度の計算結果に基づくエリート戦略とルーレットルールを併用したものである．選出された個体の交叉確率は 0.7 とされた．突然変異確率については，条件の異なる幾つかの予備計算の結果から，この確率値を低く設定すると，世代交代が進むにつれて最大適合度は順調に高くなるものの，ある適合度以上には適合度の改善が見られなかった．これは，表現木の形式が早期に収斂し，局所解に陥ったためと推察される．そのため，今回の計算では，突然変異確率を 0.7 とすることで個体群の多様性を高め，局所解からの脱出を図ることとした．個体の適合度は，部分木の非終端記号の構成とともに，終端記号の定数値の与え方によっても値が向上することも予想されることから，ある回数の世代交代がなされた時に定数ノードの設定値の改善を図るために，各定数にランダムサーチを施すことにした．計算の環境については，OS を MS-Windows として Visual Basic で計算の全体的な流れを制御し，表現木を記述している Lisp スクリプトの計算は，UNIX エミュレータの下で guile-1.6.5 を使って行なっている．

(2) ニューラル・ネットワークモデル

ニューラル・ネットワークモデルは，遺伝的プログラミングの計算結果の比較に用いられた．

このモデルの構成は，図-2のとおりであり，構成が比較的簡単なフィードフォワード型の多層モデルとされた．

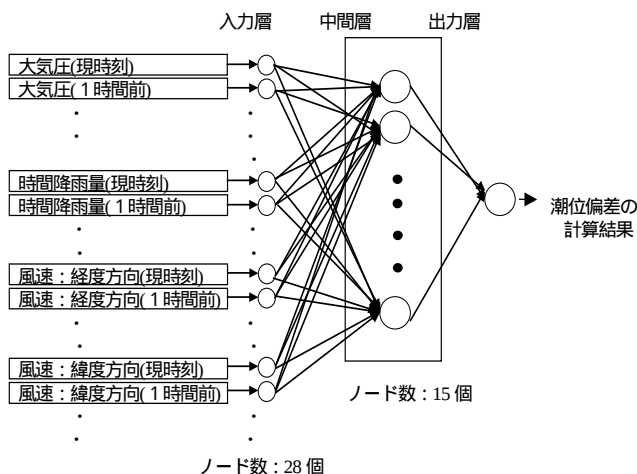


図-2 ニューラル・ネットワークの構成

すなわち，中間層を 1 層とした全 3 層型とし，中間層に 15 個，出力層に 1 個のノードを配置した．モデルの学習は，計算値と実測値との差に基づいて，バックプロパゲーションによって行なうことにした．1 セットの学習用データに対する計算回数は 5 万回としている．入力層に与える実測値については，遺伝的プログラミングで与えたものと同一種の 28 個のデータとしているが，実測値の最大値と最小値を用いて 0 から 1 の範囲に変換されている．モデル中に配されたデータ処理ノードでは，一般に用いられるシグモイド関数： $f(x)=1/\{1+\exp(-ax)\}$ を用いて出力値を算出することにした．なお，この式にある係数 a については，実際に観察された学習の進行状況から 0.05 としている．

3．計算結果の比較

(1) 計算対象地点と入力データ

前章で説明したモデルによる潮位偏差の計算は，図-3 に示される島原湾湾口部に位置する口之津港を対象に行なわれた．この港湾が取り上げられたのは，気象庁から 1 時間毎の潮位偏差のデータが配信されており，十分な



図-3 口之津港の位置図

(図中の [+] は，入力データとした MSM のグリッドポイントを指す)

実測結果を入手可能であったこと，また，これまで著者らは島原湾の潮流の数値シミュレーションを実施してきており⁴⁾，海上の気象学的条件の変化を考慮したシミュレーションに本計算結果を反映させる目的もあったことによっている．

計算の対象期間は 2004 年 9 月であり，先述のように，口之津港における同期間の潮位偏差の実測値は，気象庁の報告値とした⁹⁾．大気圧や時間降雨量等の気象学的時系列データについては，海上での気象諸量の変化が潮位偏差を生じさせる主たる要因と考えられることから，以下の計算では，現在，気象庁で運用されている気象擾乱の短期予測モデルの一つである Mesoscale モデルによって数値予報された値を利用することにした．この予測モ

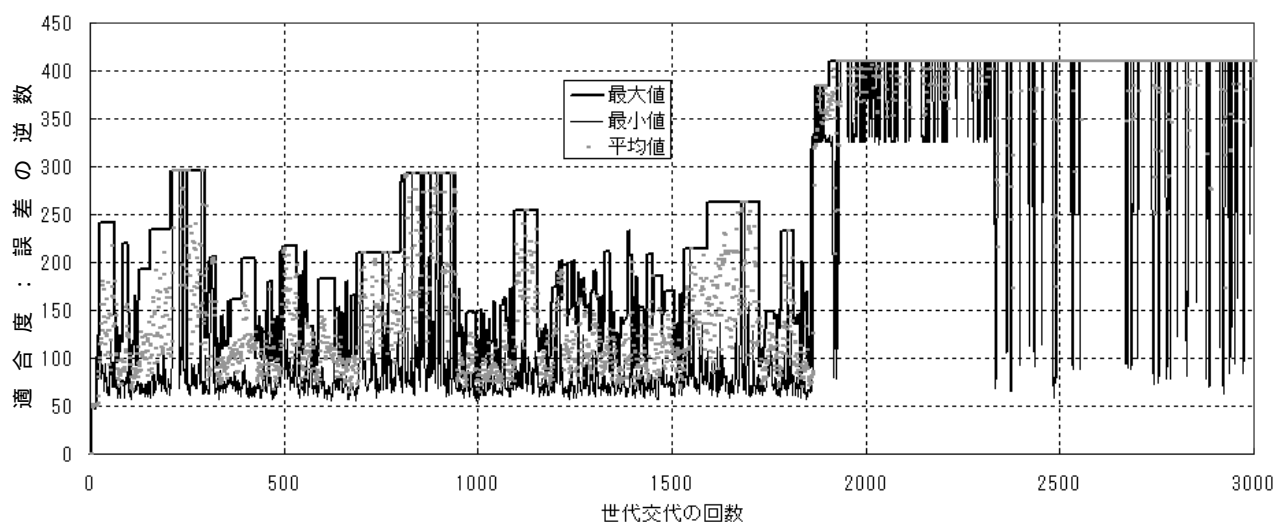


図-4 遺伝的プログラミングの表現木の適合度の変化

デルの解像度は、緯度、経度方向に $0.1^{\circ} \times 0.125^{\circ}$ であり、距離にして約10kmとなっている¹⁰⁾。なお、二つのDDMの学習に用いたデータの総数は、数値予報値を取得できなかった期間を除くと610個となっている。

(2) モデルの計算結果

ここで、まず、遺伝的プログラミングによって算出された適合度について、世代交代に伴う変化の様子を図-4に示す。同図には、一つの個体群の中で、適合度の評価結果の良い上位10位までの算出結果の平均値と、最大値並びに最小値が逆数として表示されている。また、図に描かれた世代交代の回数は3000回までの結果である。

この図を見ると、計算開始後、直ちに最大値と最小値は増加するが、1863回目の世代交代がなされるまでの間において、とくに最大値は、凹凸状の分布を示しながら大きく変化していることが分かる。その一方で、最小値は最大値と同程度の値に近づく様子も見られるが、概ね60から100の範囲にある。こうした結果が得られたのは、最大値の分布にあっては、個体群からの個体の選出処理の中に取り入れられたエリート戦略によって、世代交代の際に適合度の高い個体が優先的に選出され、次世代に引き継がれるものの、全く同一の表現木を持つ個体で上位が占められておらず、これらの表現木の構成がルール選択や部分木の交叉によって変化しやすかったためと考えられる。このことは、平均値が最大値に近い部分に偏っていないことから推察される。併せて、突然変異確率を比較的高い値に設定していることによって、世代交代の履歴に余り影響されていない適応度の高い個体が新規に生成されていることも最大値がこのような分布になった原因の一つと考えられる。

図に示された計算結果の後半部については、最大値に微小な適合度の向上があるものの、1/ERの値は、410程度で分布している。これは、2600回目付近では最小値も最大値に等しくなるなど、最小値と最大値の差が前半部

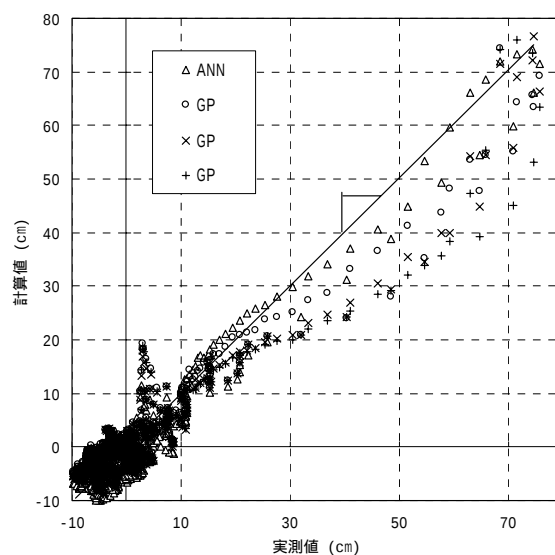


図-5 二種類のモデルの計算結果と実測結果との比較 (GP:遺伝的プログラミング, ANN:ニューラル・ネットワークモデル)

表-3 ERとTCの評価結果

	計算条件 種別	ANN	GP		
ER	学習用 2004年9月	16.2	18.0	20.1	23.7
TC		0.75	0.60	0.62	0.50
ER	検証用 2004年10 ~12月	153	113	114	115
TC		0.26	0.25	0.18	0.23

に比べて小さくなっており、上位の表現木の構成にばらつきが少なくなった結果、計算の手続きに含められた遺伝オペレータ処理に対して、堅牢な個体群へと遷移したことによるものと思われる。なお、ここでは、生成された表現木の変化状況を具体的に示していないが、図-4の最大値が急増した時の構成の変化を調べてみると、1913回目に1/ERが383から409に変化した時や、適応度に微小

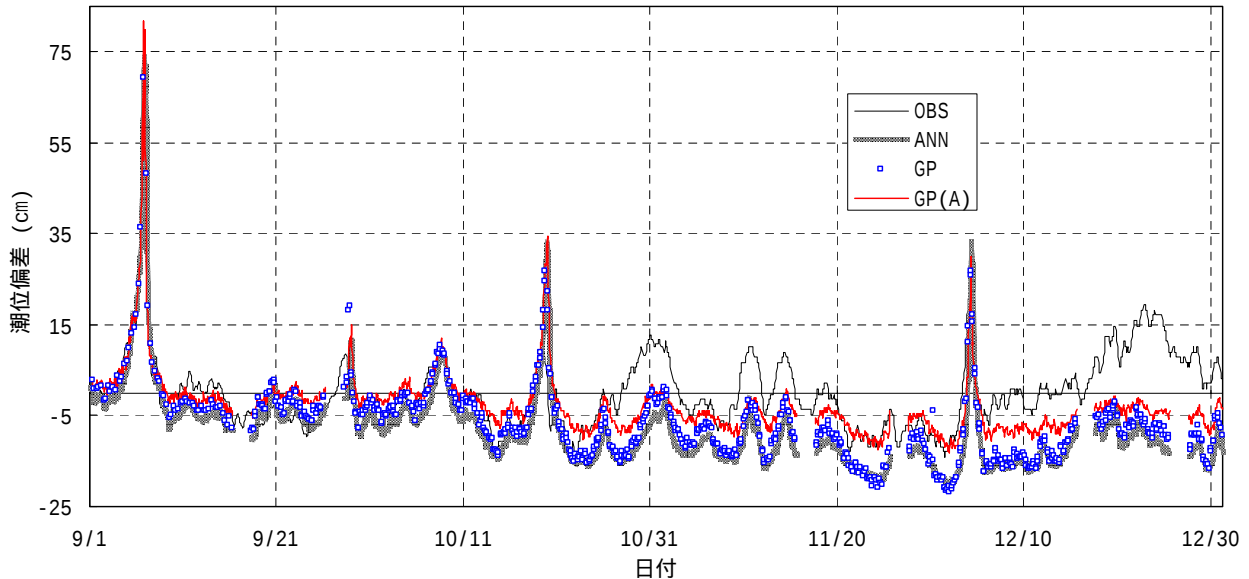


図-6 潮位偏差の実測値と計算結果

な改善があった場合などに、一つ前の個体群で上位の個体群を母体とする部分木の変化が影響を与えているようであったが、それ以外では、部分木の構成の変化状況から、下位の個体同士の交叉や突然変異によって適合度が改善されているようである。

つぎに、遺伝的プログラミングとニューラル・ネットワークモデルの計算結果の比較として、実測結果への対応状況を示すと、図-5のとおりである。この図には、遺伝的プログラミングによって得られた表現木の内、図-4に記入された、の各世代の最良個体による算出結果が描かれている。

図-5から、いずれのモデルも実測結果を表現する傾向にあり、実測値が20cm程度以下では、ばらつきがあるものの、同程度の算出精度となっている。ただ、表-3に表記されているように、9月の学習データに対するER値は、ニューラル・ネットワークモデルにおいて最も小さく、実測値の全体的な再現性が他に比べて高い。遺伝的プログラミングについては、式(1)の重みを変化させたこともあり、潮位偏差の最大値付近において実測値との差が縮まっているが、偏差が30～70cmとなる範囲で値が低めに算出されている。この程度はERの値が高くなるにつれて大きくなる傾向にある。

ここで、計算結果の再現性を評価する別の指標として、以下の式(2)として定義される値(TC)を求めると、表-3の下欄のとおりである。

$$TC = \frac{\sum_{n=1}^N \delta OBS_n \cdot \delta CAL_n}{\sqrt{\sum_{n=1}^N \delta OBS_n^2} \sqrt{\sum_{n=1}^N \delta CAL_n^2}} \quad (2)$$

ここに、 N ：実測データの総数、 δOBS_n 、 δCAL_n ：それぞれ実測データと算出値の時間変化量、である。

式(2)による算出結果を見ると、ERの結果が良好であったこともあり、ニューラル・ネットワークモデルにおいてTCの値が他者に比べて高い。一方で、遺伝的プログラミングでは、ERの結果が表現木の順に良かったが、TCの良否については、それらの順序が逆転しており、表現木の計算結果が実測値の時間変化の傾向をより良く再現していることを示している。今回の計算処理では、個体の適合性の評価にTCの値を反映させていないが、この評価結果を併用する必要性を示唆しているものと思われる。ただし、これら二つの指標は、表現している誤差の内容が異なることもあり、取り上げられる問題によって議論されるべきものと考えられる。ERのような全体的な誤差の結果にTCの結果を具体的にどの程度寄与させるかは、今後検討していきたい。

以上に記した学習後のモデルの検証として、2004年10月から12月までの潮位偏差の評価に適用した結果を表-3に記している。なお、この期間の計算に用いたデータは学習計算のそれと同種であり、その数は1977個である。また、一連の計算から得られた2004年9月から12月までの結果は、図-6に示されるとおりである。

表に記載された結果から、各モデルのERとTCのいずれも精度が低下しており、ニューラル・ネットワークモデルのTCは遺伝的プログラミングの表現木と同程度の値になっている。予測結果の様子を図-6に見ると、10月中旬までの変化や12月初旬に台風が接近したことによる潮位偏差の上昇などは表現されているが、10月末からの平時の変化は低く評価されている。こうした結果になった理由として、計算条件に関しては、過学習などモデルの学習の程度や、季節風の変化といった未学習データの混入等が挙げられる。

そこで、学習用の入力データに11月の値を追加して遺伝的プログラミングの再学習を試みた。学習データの数

は1236個である．この条件で得られた結果は，図-6にGP(A)として併記されているとおりであり，ERとTCの値は学習期間で68.7，0.69，また検証期間で103，0.28であった．図-6から，11月の再現性は期待した以上には改善されていないものの，10月の計算精度を概ね保持しつつ，全体的な算出結果が改善されていることが分かる．モデルの構造から容易に予想されることではあるが，本研究で主に取り上げた遺伝的プログラミングの計算精度は，ニューラル・ネットワークモデルと同様に，やはり学習用データの数や質などに強く依存するものと言える．とくに，10月や12月末期の計算結果の向上にあたっては，モデルの計算条件の検討と共に入力データの適切な選定が必要になるものと考えられる．

4．おわりに

本研究では，近年，多くの機関の努力によって水工学的に有用な実測値が蓄積されつつあること，また，これら豊富な情報を活用したモデル解析が可能となっていることを考えて，DDMの一つである遺伝的プログラミングを用いた潮位偏差の予測を試みた．

その結果，今回の計算条件では適合度の評価値が複雑に変化しており，世代交代の停止条件の設定が容易ではないものの，計算条件が適切になされれば，ニューラル・ネットワークモデルと同程度の予測精度を備えた表現木が得られることが示されたものと考えている．その一方で，計算精度の向上には，学習用のデータや計算条件の設定など，DDMに特有の問題を解決せねばならないことも併せて示された．なお，ここで利用した二つのモデルは基本的な構成としており，それぞれに改良の余地があることは言うまでもない．とくに遺伝的プログラミングについては，予測計算の手続き(プログラム)を自動生成できるという点がニューラル・ネットワークにはない特徴である．今回の計算では，関数式の導出を目的として演算子のみが非終端記号に適用されたが，これに加えて，条件分岐などを表す特別式を導入することも可能である．こうしたプログラムの生成条件に自由度を増した場合の適用性に関しても検討したい．

今後，これまでに記した課題に取り組むと共に，ケーススタディを増やすことで，遺伝的プログラミングの実問題への適用性について引き続き検討していきたいと考えている．

謝辞：本研究を遂行するにあたり，UNESCO-IHEのDimitri P. Solomatine 助教授からDDMの活用に関して有益なアドバイスを戴いた．ここに記して深甚なる謝意を表します．また，入力データの前処理並びに計算の実施にあたっては，岡崎久典氏を初めとする本学河川工学研究室内の諸氏に協力を戴いた．ここに謝意を表します．

参考文献

- 1) Solomatine, D.P. and Dibike, Y. B. : Automatic calibration of groundwater models using global optimization techniques, *J. Hydrological Sciences*, 44 (6), pp.879-894, 1999.
- 2) Maskey, S., Jonoski, A., and Solomatine, D. P. : Groundwater remediation strategy using global optimization algorithms, *J. of Water Resources Planning and Management*, Vol. 128, No.6, pp. 431-440, 2002.
- 3) Noguchi, M., Solomatine, D. P. and Nishida, W.: Calibration of water quality model by global optimization techniques, *Proc. 5th Int'l conf. on Hydroinformatics* Vol. 1, pp. 464-469, 2002.
- 4) 西田 渉，野口 正人，D.P.Solomatine：気象予報情報の潮流予測への適用に関する研究，*水工学論文集*，第49巻，pp.1279-1284，2005．
- 5) Giustolisi, O. and Laucelli, D.: Modeling rainfall-runoff by genetic programming: results from two experimental urban basins, *Proc. 5th Int'l conf. on Hydroinformatics* Vol. 2, pp. 1484-1491, 2002.
- 6) Babovic, V., Harrington, J. R. and Keijzer, M.: Generation of settling velocity equations for sand grains using genetic programming, *Proc. 6th Int'l conf. on Hydroinformatics* Vol. 2, pp. 1631-1638, 2004.
- 7) 伊庭 斉志：進化論的計算の方法，東京大学出版会，1999．
- 8) 伊庭 斉志：遺伝的プログラミング，東京電気大学出版会，2001．
- 9) 気象庁：ホームページ 海洋の情報資料，URL <http://www.data.kishou.go.jp>，2004．
- 10) 気象庁予報部：メソ数値予報の実用化に向けて，数値予報課報告，別冊第44号，p.79，1998．

(2005.9.30受付)