

降雨流出モデルへの MCMC 法の適用に関する基礎的検討

岐阜大学大学院自然科学技術研究科 学生会員 ○范顔楊
 岐阜大学工学部 花田浩平・林浩平・Vivien Wanda-Rahma
 (株)日水コン中央研究所 川口智哉
 岐阜大学工学部 正会員 篠田成郎

1. 研究の背景と目的

これまでに各種の降雨流出モデルが提案されるとともに、そのモデルパラメータの適切な同定方法(例えば、SCE-UA 法¹⁾、遺伝的アルゴリズム法²⁾、進化戦略法³⁾、群知能最適化法⁴⁾など)が論じられている。これらのいずれの方法でも、対象とする流域特性に応じてパラメータの下限値、上限値および初期値を適切に与えることが必要になる。一方、気象学や海洋工学など大規模なデータとモデルを必要とする分野において、データ同化手法の適用が進められ、近年では、生物学や社会科学にも応用が広がってきている⁵⁾。データ同化手法は、ベイズの定理に基づくパラメータの確率密度関数の推定から尤もらしいパラメータ値として同定するものであり、実測データを利用してモデルの精度を向上させることを目的にしている⁶⁾。また、物理モデルを適用することで欠測データを補完することも可能となる。しかし、降雨流出モデルにこうしたデータ同化手法を適用した例はない。

本研究では、データ同化手法を降雨流出モデルに適用した場合、どのような課題が生じるのか、あるいはこうした適用のメリットは何なのかについて検討することを目的とする。データ同化手法としては、MCMC(モンテカルロ・マルコフ連鎖)法、アンサンブル・カルマンフィルター法、アジョイント法などがあるが、ここでは、最も簡単な手法とされる MCMC 法の適用性について検討する。

2. MCMC 法の概要

観測値 y が与えられたときにモデルパラメータが θ となる確率密度 $p(\theta|y)$ は、モデルパラメータ θ の確率密度 $p(\theta)$ 、観測値 y の確率 $p(y)$ およびモデルパラメータ θ を与えたときに観測値 y が得られる確率密度 $L(y|\theta)$ を用いてベイズの定理より次のように表される。

$$p(\theta|y) = \frac{L(y|\theta)p(\theta)}{p(y)} \dots\dots(1)$$

MCMC 法では、パラメータ θ の候補値をマルコフ連鎖として乱数によって生成して得られるモデル推定値と実測値との差の確率密度関数を尤度関数として、事後確率 $p(\theta|y)$ を最大化するモデルパラメータ θ が求められる。この際、乱数発生を繰り返すことにより得られるパラメータ θ の系列は定常なマルコフ過程となる必要がある。MCMC 法での定常性とは、時間や位置によってパラメータ θ の確率分布が変化しないことであり、その平均や分散も時間や位置によって変化しない。一般に、MCMC 法での定常性は Geweke の方法⁷⁾で確認できる。マルコフ性は、将来状態の条件付き確率分布が現在状態のみに依存し、過去のいかなる状態にも依存しない特性であり、自己相関で確認です。

MCMC 法の計算アルゴリズムとしては、メトロポリス法(M 法)⁸⁾とメトロポリス・ヘイスティンクス法(M-H 法)⁹⁾が有名である。いずれのアルゴリズムでも、次式で定義される事前・事後確率比 α によって更新パラメータの採否が決められる。現在 t のパラメータを $\theta^{(t)}$ とし

て、候補値 $\theta' (= \theta^{(t)} + \varepsilon)$ を正規乱数 ε により発生する。式(2)の確率比 α によって、現在のパラメータ $\theta^{(t)}$ と候補値 θ' の事後確率を比較し、パラメータを更新するかどうか判断する。

$$\alpha = \frac{P(\theta'|y)q(\theta^{(t)}|\theta')}{P(\theta^{(t)}|y)q(\theta'|\theta^{(t)})} \dots\dots(2)$$

ここに、 $q(\theta^{(t)}|\theta')$ はパラメータ θ' の生起の後にパラメータ $\theta^{(t)}$ が生起する条件付き確率であり、推移確率と呼ばれる。推移確率が往復で等しい場合には式(2)の q は相殺され、これを M 法と呼ぶ。一方、式(2)の q が残る場合が M-H 法となる。ここでは、M-H 法を適用することにする。

M-H 法によるパラメータ同定のフローチャートを図-1に示す。まず、 $\alpha > 1$ (θ' の事後確率は $\theta^{(t)}$ より大きい)の場合にはパラメータを更新し、 $\alpha \leq 1$ の場合、ある確率(一般には一様乱数によって、更新・非更新を判定する)で更新する。こうしたアルゴリズムにより、正規乱数 ε の繰り返し生成に伴い、更新パラメータは事後確率 $p(\theta|y)$ を最大とする θ の値の周りで大きく変動する挙動を示すようになる。上記の定常性とマルコフ性が満足されるようになる繰り返し回数は Burn-in 回数と呼ばれ、Burn-in 回数後の更新パラメータの最頻値が同定値として採択される。

なお、複数のモデルパラメータを同定する場合には、全てのパラメータの組み合わせに対して M-H 法を適用するのではなく、パラメータの種類を順次変更しつつ、その値を更新する方法が採られる。これをギブス・サンプリング¹⁰⁾と呼ぶ。

3. 集中型コンパートメントモデルへの MCMC 法の適用

降雨流出モデルへの MCMC 法の適用性を検証するため、最も単純な集中型コンパートメントモデルを例として考察する。このモデルでは、鉛直方向 3 つの土層(上から A 層・B 層・C 層)において 4 種類のパラメータ(初期水位、土層間隙率、水平流出係数、鉛直流出係数)が設けられているため、合計 12 個のパラメータ同定が必要になる。なお、用いる観測データは岐阜県中津川市二ツ森山東斜面に位置する森ヶ洞集水域における 1998 年の降雨流出を対象とした¹¹⁾。

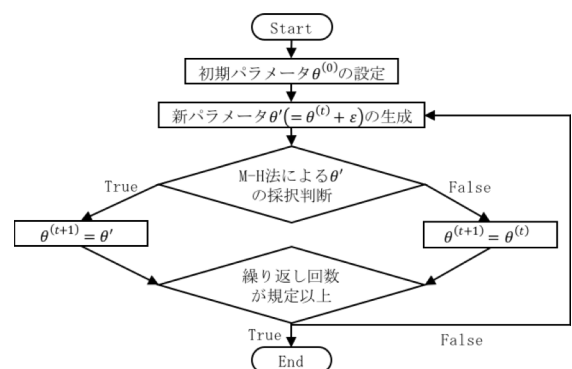


図-1. MCMC 法(M-H 法)によるパラメータ同定の手順

A層での初期水位 h_{A0} と水平流出係数 α_{AH} を例として、繰り返し回数に伴うパラメータ同定値の変動を図-2に示す。2000回弱および9000回弱の時点で h_{A0} と α_{AH} の値が同時に低下している。これら以外のパラメータにおいても同様な挙動を示しており、パラメータが相互に連動していることが原因と考えられる。何回かのパラメータ組み合わせの試行錯誤により、連動性の高いパラメータを特定したところ、各層の水平流出係数のみを同定対象のパラメータとすることでMCMC法による同定が可能になることが分かった。その結果を図-3の中段に示す。この例では、繰り返し回数が数百回を超えると、一定の値の周りでパラメータが変動している。このことから、MCMC法では、パラメータ相互の連動性を排除した同定対象パラメータの選択が重要になると言える。こうした結果より、集中型コンパートメントモデルにおける12個のパラメータ同定の手順として、まず、3つの初期水位、土壌間隙率および鉛直流出係数の仮の値を定め、残りの3つの水平流出係数をMCMC法で同定し、その結果を用いて、順次、パラメータの種類ごとにMCMC法で同定を繰り返す方法が適切になることが判明した。

また、パラメータ生成時の正規乱数の標準偏差 σ_ε を変えた場合について、図-3に併せて示す。これらの結果より、パラメータ更新の幅を小さくすると適切な値を見つかることが出来ず、一方、幅を大きくしても変動に長周期のトレンドが生じるようになることが分かる。前述のように、既往の流出モデルにおけるパラメータ同定ではその下限値・上限値を定める必要があったが、MCMC法ではこうした標準偏差を適切に与えなければならぬ。ただし、Burn-in後の変動パラメータ値に関する平均、分散、およびGeweke統計量および確率比 α によるパラメータ棄却率(50%程度が良いとされている)により、同定結果の定常性を確認できるため、最初には大きな σ_ε を与えておき、その後、 σ_ε を小さくしていくことで適切なパラメータ同定が可能になる。表-1はパラメータ α_{AH} についてこれらの統計量を比較した結果であり、 $\sigma_\varepsilon = 0.025$ のケースが最も高い定常性を示し、適切な同定結果と判断される。

図-4はパラメータ α_{AH} に関する自己相関を示したものであり、 σ_ε が大きくなるに従い小さな遅延で自己相関がゼロになっている。 σ_ε が大きい場合には、更新されるパラメータ値はその前の値と大きく異なる可能性が高くなるため、早く自己相関が小さくなる。しかし、その反面、パラメータ更新時の棄却率が大きくなりすぎる。これらより、MCMC法では、出来るだけ大域的にパラメータを探索しつつ、定常性と自己相関性を確認してパラメータ同定結果を判定することが大切になる。

参考文献

- 1) Duan, Q., et al.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of Hydrology*, 158, 265-284, 1994.
- 2) 田中丸治哉: タンクモデル定数の大域的探索, 農業土木学会論文集, 178, 103-112, 1995.
- 3) 藤原洋一ら: 進化戦略による流出モデル定数の最適化, 農業土木学会論文集, 227, 119-129, 2003.
- 4) 小槻峻司ら: 群知能最適化手法を用いた分布型流出モデルのパラメータ同定, 土木学会 論文集 B1, 68, 523-528, 2012.
- 5) 長尾大道ら: スパースモデリングとの融合によるデータ同化研究の新展開, システム/制御/情報, 61(4), 132-137, 2017.
- 6) 加藤知道: データ同化と陸域生物圏モデル, 生物と気象, 13, 1-17,

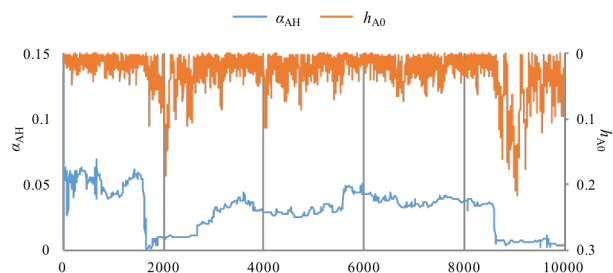


図-2. 繰り返し回数に伴うモデルパラメータ変動の例 (パラメータ数12個の場合)

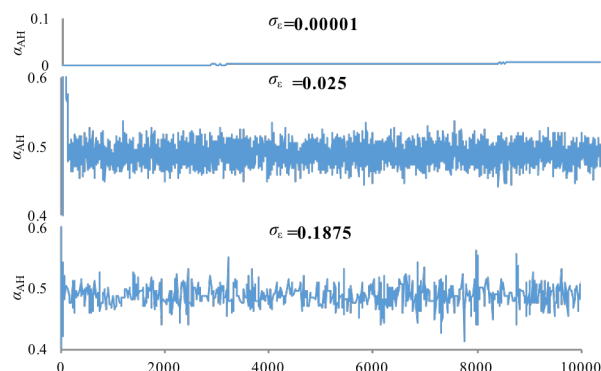


図-3. 新パラメータ生成時乱数の分散によるモデルパラメータ変動の比較(パラメータ数3個の場合)

表-1 パラメータ α_{AH} の同定における Burn-in 後の前半10%区間と後半50%区間での統計量の比較

σ_ε	データ数	平均	分散	Geweke統計量	棄却率
1.0×10^{-5}	700	2.76×10^{-3}	1.12×10^{-7}	-180.76 (N. G.)	4%
	3500	5.86×10^{-3}	4.66×10^{-7}		
2.5×10^{-2}	700	4.89×10^{-1}	2.03×10^{-4}	-0.27 (O. K.)	51%
	3500	4.89×10^{-1}	1.79×10^{-4}		
1.875×10^{-1}	700	4.88×10^{-1}	2.98×10^{-4}	-1.78 (O. K.)	92%
	3500	4.90×10^{-1}	3.01×10^{-4}		

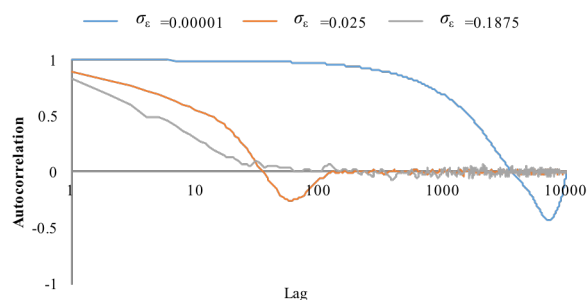


図-4. 新パラメータ生成時乱数の分散によるモデルパラメータ変動の自己相関の比較

- 2013.
- 7) Geweke, J.: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *Bayesian Statistics*, 4, 169-193, 1992.
- 8) Metropolis, N. et al.: Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, 21, 1087-1092, 1953.
- 9) Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57:97-109, 1970.
- 10) Geman, S. et al.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transaction Pattern Anal Mach Intell*, 6:721-741, 1984.
- 11) 花田浩平: データ同化手法を用いた降雨流出解析による流出特性の評価, 令和元年度中部支部土木学会研究発表会概要, 2020.