

時空間地球統計学を用いた水文データ補間分布推定手法

浜口 俊雄・佐藤 嘉展・小尻 利治

要 旨

本研究では、時間距離換算パラメータを提案・導入し、地下水文データに対する kriging の時空間適用への拡張を試みた。地下水位分布を解析例として時空間統計モデルの最適化した分布を推定したところ、概ね良好な結果を得た。これより本提案手法は時空間の相関性を考えた分布として大変有効であると言え、特に観測データの相関性を時空間で一元的に考えて分布モデル評価をする場合に有用性・応用性に富むとがわかった。その際、モデル構造に注意しながら使うことが必要である。

キーワード：地下水，水文データ，時空間分布，時空間統計モデル，時間距離換算パラメータ

1. 序論

気温・降水量など所与の気象観測データあるいは河川流量・水位や地下水位などの水文観測データは水文現象の再現計算時に入力データまたは再現検証として用いられる。ゆえにそれらの空間分布と時間変化は重要である。従来の入力データ作成時は、点在する観測データを空間分布として扱うべく、Voronoi 分割で定めた多角形有効域毎に観測値/代表値が一様であると見なして与える Thiessen 法、または、観測値/代表値の影響度が距離に反比例すると仮定し、距離の逆数から定めた重みを乗じた観測値の和で空間補間的に各格子セル毎の値を算出する距離逆数法を採用していることが多い。しかし前者は概して粗く不連続な階段状の分布となる点、後者は観測値の影響が距離に反比例することの理論的根拠が必ずしも保障されないという点に問題がある。さらに後者は、比較的細かい分布を扱う際に、降水量のようなゼロ領域を部分的に有する空間変量に対しても連続かつ滑らかな空間分布を全域に算出し、望まない不都合な結果を生じてしまうこともある。故に、空間全域の再現性よりも部分領域的な(数箇所の観測点での)再現性を重要視する場合に向く。

筆者らは水文モデルパラメータ分布について、空間分布として同定(浜口ら, 2006)あるいはアップスケール時の等価化(浜口ら, 2007)に関する手法を各々提案してきた。ただし、これらはパラメータがモデル定数のため、時間に依存しない分布を扱っていた。本研究では、時間変動

を伴う空間分布を推定補間対象に考えるため、従来は空間補間を目的としている地球統計学の kriging を時間方向にも拡張し、時空間分布を推定するための新たな kriging を提案するとともに検証する。

2. 時空間統計モデル

2.1 位相トレンド構造

kriging 推定時、式(1)の様に対象変量 $\phi(z)$ をトレンド成分 $m(z)$ とランダム成分 $w(z)$ に分解し、ランダム成分の定常性を確保する。

$$\phi(z) = m(z) + w(z) \quad (1)$$

ここに、 z : 空間座標とする。本稿では式(2)の様に両成分を時空間関数形に変更し、時空間対応型に拡張させることを考える。

$$\phi(z, t) = m(z, t) + w(z, t) \quad (2)$$

ただし、 t : 時間とする。そのトレンド関数 $m(z, t)$ は次数が一次以上のものも想定してもよいが、ここでは単純な次数で検討しやすくするために、空間分布に xy 平面空間を想定すれば式(3)の様な一次多項式で与えられる。

$$m(z, t) = b_1 + b_2x + b_3y + b_4t \quad (3)$$

ここに、 b_i : 係数とする。トレンド関数の従来モデルと式(3)との違いは第4項の時間トレンド成分項にある。対象変量は空間と時間という異なる性質の変数を同時に有しているものの、多項式形式で表されているため、多項式

各項はその係数によって容易に対象変量と同じ次元で扱え得る。後述の「時間距離換算パラメータ」を提案・導入する場合でも、多項式形式は矛盾点を出すことなく扱えるかたちになっている。よって時空間に対応した上記構造のトレンド関数を「位相トレンド関数」と定義する。

2.2 位相ランダム構造

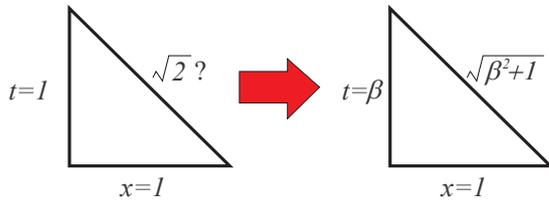


Fig.1 Schematic of parameter β

ランダム成分は、その相関式において観測データ間の相関距離が重要となる。ただし、時間と空間は異なる性質の座標変数であるため、同一空間内で計量し得る場合は位相空間で考える必要がある。本研究では、時間距離1(単位時間)に匹敵する空間距離を換算すると $\beta(>0)$ になると考え、その β を時間距離換算パラメータと定義する。Fig.1はその概念を示している。空間方向に1、時間方向に1だけ離れた場合の時空間での距離は $\sqrt{2}$ ではない。時間を空間距離に換算して β だけ離れていると考えれば、その時空間距離は $\sqrt{\beta^2+1}$ となる。

これを踏まえて、時空間距離すなわち位相距離を算出する方法を以下の様に提案する。位相一次元距離を d_r 、空間二次元距離を d_t とすると、

$$d_r = \sqrt{d_x^2 + d_y^2 + (\beta d_t)^2} \quad (4)$$

$$d_t = \sqrt{d_x^2 + d_y^2} \quad (5)$$

である。位相一次元距離で相関関係を考える際は式(6)~(8)の様な共分散関数を用意すればよい。ここに以降のパラメータ a_i は位相相関距離と呼ぶと定義する。

・位相一次元指数型 :

$$C(\mathbf{d}) = \sigma^2 \exp \left\{ -\frac{d_r}{a_r} \right\} \quad (6)$$

・位相一次元ガウス型 :

$$C(\mathbf{d}) = \sigma^2 \exp \left\{ -\left(\frac{d_r}{a_r} \right)^2 \right\} \quad (7)$$

・位相一次元球状型 :

$$C(\mathbf{d}) = \begin{cases} \sigma^2 \left\{ 1 - 1.5 \left(\frac{d_r}{a_r} \right) + 0.5 \left(\frac{d_r}{a_r} \right)^3 \right\} & (0 \leq d_r \leq a_r) \\ 0 & (d_r > a_r) \end{cases} \quad (8)$$

位相二次元では空間・時間別に一次元相関距離を考える

ため、式(9),(10)の様な共分散関数でよい。同様に位相三次元では式(11),(12)の様な関数になる。

・位相二次元指数型 :

$$C(\mathbf{d}) = \sigma^2 \exp \left\{ -\sqrt{\left(\frac{d_t}{a_t} \right)^2 + \left(\frac{d_t}{a_t} \right)^2} \right\} \quad (9)$$

・位相二次元ガウス型 :

$$C(\mathbf{d}) = \sigma^2 \exp \left[-\left\{ \left(\frac{d_t}{a_t} \right)^2 + \left(\frac{d_t}{a_t} \right)^2 \right\} \right] \quad (10)$$

・位相三次元指数型 :

$$C(\mathbf{d}) = \sigma^2 \exp \left\{ -\sqrt{\left(\frac{d_x}{a_x} \right)^2 + \left(\frac{d_y}{a_y} \right)^2 + \left(\frac{d_t}{a_t} \right)^2} \right\} \quad (11)$$

・位相三次元ガウス型 :

$$C(\mathbf{d}) = \sigma^2 \exp \left[-\left\{ \left(\frac{d_x}{a_x} \right)^2 + \left(\frac{d_y}{a_y} \right)^2 + \left(\frac{d_t}{a_t} \right)^2 \right\} \right] \quad (12)$$

位相二・三次元共分散関数では、時間からの換算の位相相関距離 a_t が β を含んだかたちのため、 β を直接同定せずに考えられる。同様に、先述の位相トレンド関数においても b_4 には β が含まれた係数と見なしでもよいが、こちらは従来モデルのかたちと変わりはないため、計算は時間距離換算パラメータ値を意識することなく実行できる。ところでkriging推定値 $\phi^*(z)$ は元来、重み付き観測値の総和に帰着する。いま、 λ :観測値の重み係数ベクトル、 ϕ :観測値ベクトル、 $f(z)$:位相トレンド多項式で用いる基底関数ベクトル、 F :観測点毎の $f(z)$ のベクトル量を転置して観測位置順に縦に並べて出来た行列、 C :観測点間でのランダム成分の共分散行列、 $c(z)$:1つの推定位置のランダム成分と各観測点でのランダム成分との共分散値を縦に並べて出来たベクトル、とおくと、

$$\phi^*(z) = \lambda^T \phi \quad (13)$$

$$\lambda = C^{-1} c(z)$$

$$-C^{-1} F \left(F^T K^{-1} F \right)^{-1} \left\{ F^T C^{-1} c(z) - f(z) \right\} \quad (14)$$

と書ける。従来はその重みを時間毎に求めていたため、実質的に重みが時間依存となっていた。すなわち、上記式(13),(14)の $f(z)$ 、 F 、 $c(z)$ 、 C のパラメータが時間毎に求められ、その時間方向の結合関係は考慮されずにいたことになる。本研究では、これらの行列・ベクトルが全て時間方向への結合も考慮されており、重み λ も当然のように時空間距離(位相空間距離)で定まることが式(14)からも判る。また本手法では各パラメータが時間に依存しない定数で得られるため、一度定まった重みはどの時刻にも使え、降雨量のGCM出力に対するダウンスケ-

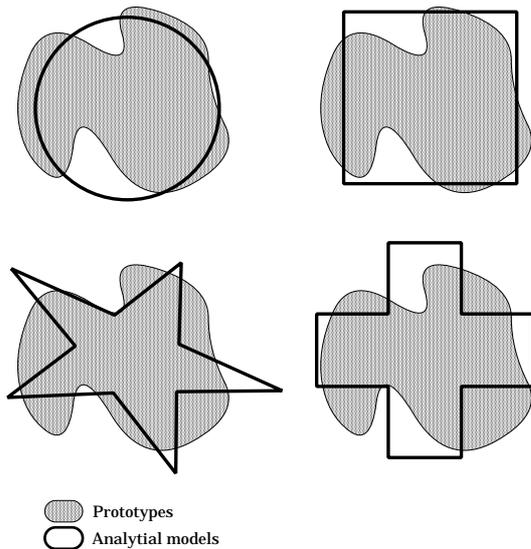


Fig.2 Given analytical models in good/bad agreement with the unknown prototype

ル時の分布再現にも利用可能と期待される。

2.3 情報量規準 AIC

Fig.2は、“実物”ないしは“実現象”であるプロトタイプに対して、解析モデルが様々当てはめられた状況を抽象的に示したものである。本来、プロトタイプ(網掛け部)の構造は未知なものであることが多い。これを解析する場合、連続性、保存性、法則性といった理論的考察から派生したような何らかの型にはまった解析モデル(太実線内部)で表現しなければならない。するとFig.2のように、モデル構造上のズレが多かれ少なかれ必ず生じることになる。一般的には、このズレを直接計ることで個々のモデルの適合度を調べればよいが、上述にもあるようにプロトタイプは未知なため、両者のズレを直接的かつ絶対的に評価することは不可能と言える。しかし、プロトタイプが唯一無二であって不変であることに留意すると、ズレを表すモデル誤差評価式において、プロトタイプの構造を示す項は定数と見なせることに気が付く。よって解析モデル間の適合性の優劣だけでよいのであれば、モデル誤差評価式の解析モデル構造を示す項だけを抽出し、比較することでズレ(モデル適合度)の相対的評価が実現可能となる。その指標が情報量規準(Information Criterion)と呼ばれており、本稿でもモデルの適合度を判断する指標として採用している。

本研究での統計モデルは、1通りのトレンドモデルと複数のランダムモデルの組み合わせが考えられる。一連の統計モデル構造同定の際に同時に複数のモデルから最適な統計モデルを選定する。まず各統計モデルに対してパラメータを最尤推定すると同時に、各モデル毎に情報量

規準値を算出する。本稿では以下に記す AIC : 赤池情報量規準 (Akaike, 1974) , を採用して算出する。

$$AIC = -2MLL + 2m \quad (15)$$

ただし、MLL : 最大対数尤度、 m : パラメータ数を表す。

次に数通りの β 値毎に AIC を大小比較する。小さな値ほど適合度は高いと判断してモデル間の相対的優劣を考える。この作業から行い、各モデルに対する最小の β となる値がだまかに見えてくる。これは β がハイパーパラメータとして扱われているためによる。

3. 数値実験

本研究では、河川水位は河道域のみの分布となり、降水量は晴天時に時間的不連続となることなどから、時空間連続性・変動性が常時あって検討するには比較的扱いやすい流域内の地下水位分布をまず検討対象とすることにした。

本提案手法の効用について検討すべく、宮古島砂川地下水盆地流域の1993年10月20日と11月20日における地下水位分布一斉水位観測データを用いた。単位時間は1 dayで設定、単位長さは1mとし、それに対する色々な値の β ($0 \leq \beta \leq 2$) を使って比較検討した。

3.1 位相一次元指数型の場合

トレンド関数・共分散関数のパラメータを最尤法で同定した結果は

【 $\beta = 1/30$ のとき】

$$m(z, t) = 30.3666 - 0.009170x + 0.006854y - 0.04127t \quad (16)$$

$$C(d) = 33.5213^2 \exp\left(-\frac{d_r}{170.4173}\right) \quad (17)$$

【 $\beta = 1/3$ のとき】

$$m(z, t) = 38.3943 - 0.01035x + 0.006163y - 0.006140t \quad (18)$$

$$C(d) = 50.4371^2 \exp\left(-\frac{d_r}{2074.04}\right) \quad (19)$$

【 $\beta = 2/3$ のとき】

$$m(z, t) = 39.3464 - 0.01044x + 0.006001y - 0.003645t \quad (20)$$

$$C(d) = 50.0789^2 \exp\left(-\frac{d_r}{2948.94}\right) \quad (21)$$

【 $\beta = 1$ のとき】

$$m(z, t) = 39.6524 - 0.01046x + 0.005944y - 0.002643t \quad (22)$$

$$C(d) = 48.9596^2 \exp\left(-\frac{d_r}{3315.27}\right) \quad (23)$$

【 $\beta = 2$ のとき】

$$m(z, t) = 39.9872 - 0.01048x + 0.005861y - 0.001439t \quad (24)$$

$$C(d) = 47.4622^2 \exp\left(-\frac{d_r}{3635.75}\right) \quad (25)$$

となる。このうち、どの場合も類似した分布の特徴が見受けられたため。代表例として、ここに $\beta = 2/3$ のときの分布推定結果のみを Fig.3,4 に示す。

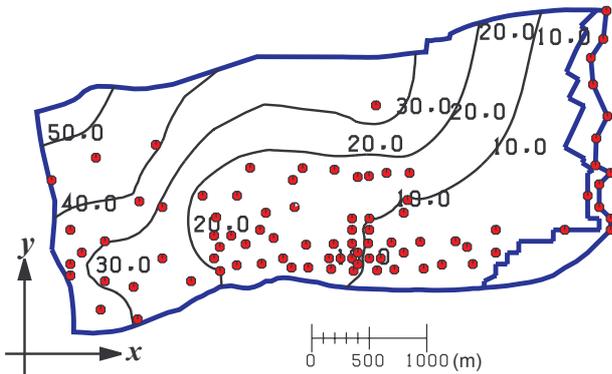


Fig.3 Estimated groundwater level distribution on Oct. 20, 1993 based on exponential type in topological one dimension

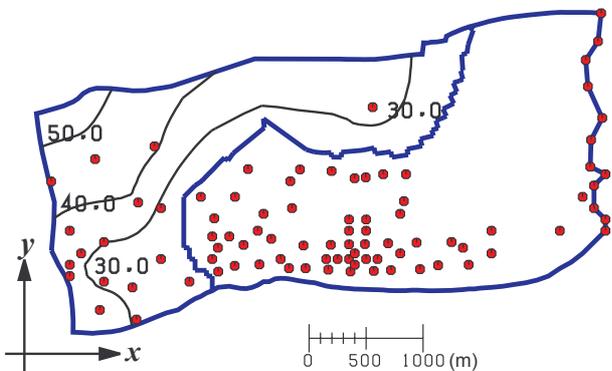


Fig.4 Estimated groundwater level distribution on Nov. 20, 1993 based on exponential type in topological one dimension

Fig.3に関しては最下流部の水位が過小推定になったために最下流部には誤差が認められるが、全体は概ね誤差も小さく、分布推定が良好であるとわかる。Fig.4は明らかに中流域から下流域にかけての水位が過小評価となっており、下流域の水深がゼロになってしまっている。これは10月20日の水位分布の推定に強い影響を受けた状態でパラメータが同定されたためと思われる。

3.2 位相三次元指数型の場合

トレンド関数・共分散関数のパラメータを最尤法で同定した結果は

【 $\beta = 1/30$ のとき】

$$m(z, t) = 35.0986 - 0.009946x + 0.006663y - 0.07457t \quad (26)$$

$$C(d) = 15.5726^2 \exp\left\{-\sqrt{\left(\frac{d_x}{811.2706}\right)^2 + \left(\frac{d_y}{690.6625}\right)^2 + \left(\frac{d_t}{29.9608}\right)^2}\right\} \quad (27)$$

となる。位相三次元指数型において、 $\beta = 1/30$ のときの結果のみ表示している理由は次節にて詳述する。このときの分布推定結果を Fig.5,6 に示す。

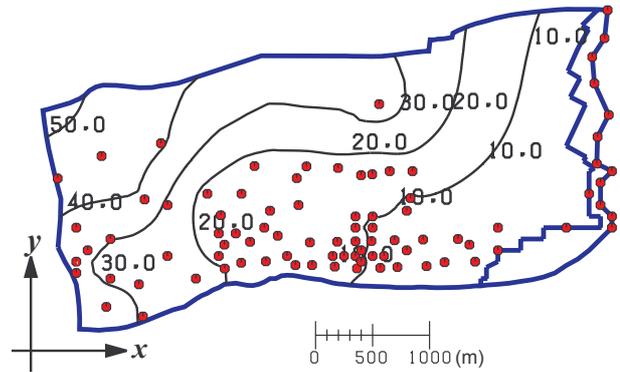


Fig.5 Estimated groundwater level distribution on Oct. 20, 1993 based on exponential type in topological three dimensions

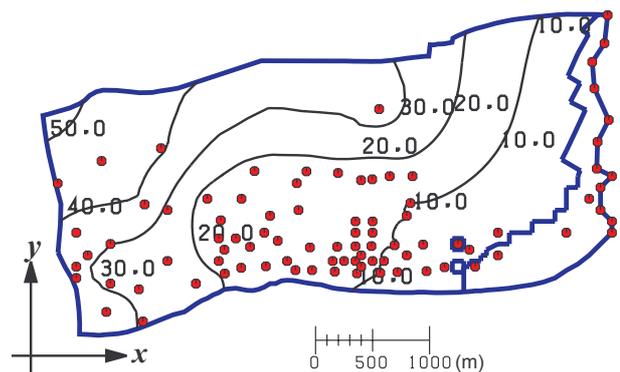


Fig.6 Estimated groundwater level distribution on Nov. 20, 1993 based on exponential type in topological three dimensions

両結果とも Fig.3と同様に最下流部の水位が過小推定になっているが、全体は概ね誤差も小さくて分布推定が良好であるとわかる。

Table 1 Relationship between β and AIC in parameter identification

Covariance func. model		$\beta = 1/30$	$\beta = 1/3$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$
Dim.	Type					
1 Topo-logical	Exponential	720.0568	611.9140	608.1317	616.2344	645.1787
	Gaussian	744.4187	711.0623	724.1948	770.9871	939.0329
	Spherical	743.7970	600.1422	591.9690	598.1137	625.1465
3 Topo-logical	Exponential	630.0193	640.4219	641.0675	659.3442	680.6591
	Gaussian	737.6235	764.3799	778.7095	749.4559	824.9632
	Spherical	618.4129	618.4135	618.4137	618.4136	618.4127

3.3 モデル最適選択

次に β の最適値について検討した。本研究では β が時空間統計モデルに独立なパラメータとして見なせば、一種のハイパーパラメータとして扱うことになるため、パラメータ同定時のAIC(赤池情報量規準)値が最小となる β をもって、 β の最適値と判断する。各 β に対するAIC算出結果をTable 1に示す。

これから、共分散関数が一次元の場合、AICは指数型・球状型であれば $1/3 \leq \beta \leq 1$ で最小が存在し、ガウス型であれば $1/30 \leq \beta \leq 1/3$ で最小が存在することがわかる。本稿では、 β を5通りしか準備していないため、前者は $\beta=1/2$ 、後者は $\beta=1/3$ で最適値になると言える。一方、共分散関数が三次元の場合は、AICが単調増加または極値が複数となるなど β には一意に最適値があるとは断言できない。先の3.2節で同定結果を $\beta=1/30$ のみにしていた理由はその最適値の存在の問題があるためである。これは一次元モデルが時空間でひとまとめにした1次元距離の相関性を取っていて、どちらからも相関性が取れる形で同定されるため、 β に最適値が現れるが、時空間的に相関性を方向毎に同定(調整)される統計モデルであることから、空間距離との整合性をとる β の機能が薄れて、観測データの時空間各軸の変動特性からパラメータが独立的に同定されてしまうため、 β が十分小さくていいと算出されるか、複数の最適候補が小さな差異の範囲で現れ、最適値の存在については一概に言えなくなってしまうと推察される。さらには β にほぼ反比例して、トレンド関数の時間項係数 b_4 は小さくなっている。三次元モデルは時間トレンド成分を小さくして、時間成分のランダム性を強めて考えるモデルの方が優れていくという結果を有する。これはやはり、時間変動が空間変動に比べて小さくて、トレンド性はなくていいというモデルになってしまい、時間方向の影響だけでパラメータ同定す

るために β が小さい方が優秀という結果を招いていることも窺える。よって、より複雑な時空間分布を有する降水量分布などへの適用するならば、その前に統計モデルの改良、例えば、今回定数とした共分散関数の分散値を時間依存にするモデルの採用なども検討する価値があると思われる。また、 β の最適に向けての扱い方に関して更なる検討が必要と思われる。

4. 結論

本稿では、時間距離換算パラメータ β を提案・導入し、krigingの時空間適用への拡張を試みて場合毎にモデル最適化した推定結果を得た。これより本手法はデータの相関性を時空間で一元的に考える場合に特に有用性・応用性に富むとわかった。ただし、時間と空間の相関性を一元的に考える場合は β が効果的に評価に関わるため、その最適値が存在するとわかるが、各軸で相関性をとるようにすれば時間距離換算の効果が薄れ、最適値の存在は保障の限りでないと言える。

参考文献

- 浜口俊雄・小尻利治・中北英一：地球統計学的な疑似生成観測値の利用によるパラメータの空間分布同定，京都大学防災研究所年報，第49号B，pp.633-639，2006。
- 浜口俊雄・小尻利治・Mohamed Saber：均質化理論に基づくアップスケーリングの水文学的適用法，京都大学防災研究所年報，第50号B，pp.759-764，2007。
- Akaike, H.: A new look at the statistical model identification, IEEE Trans. Automat. Control, AC-19, pp.716-723, 1974.

Distribution Estimation Method for Interpolation of Hydrological Data via Spatiotemporal Geostatistics

Toshio HAMAGUCHI, Yoshinobu SATO and Toshiharu KOJIRI

Synopsis

This study presents a newly developed method for estimation of hydrological distributions in space and time based on observed data. To extend spatial estimates into spatiotemporal ones, a conversion parameter of temporal-distance is newly defined and introduced into the kriging equations using the trend and covariance functions. This parameter equivalently converts a temporal gap into a spatial distance in the time-related terms of those functions. In topological space in one dimension, the above approach is most effective and helpful to krig a hydrological distribution. It can be shown that the developed method of kriging estimation in space and time with a conversion parameter is useful and applicable to a spatiotemporal interpolation through the numerical tests.

Keywords : Groundwater, Hydrological data, Spatiotemporal distribution, Spatiotemporally statistical-based model, Conversion parameter of temporal-distance