

氏 名	金 秀 梃
授与した学位	博 士
専攻分野の名称	学 術
学位授与番号	博甲第5160号
学位授与の日付	平成27年 3月25日
学位授与の要件	環境生命科学研究科 環境科学専攻 (学位規則第5条第1項該当)
学位論文の題目	The Estimation of the Variogram in Geostatistical Data with Outliers (地球統計データにおけるバリオグラムの推定と外れ値に関する研究)
論文審査委員	教授 栗原 考次 教授 坂本 亘 准教授 苗田 薫

学位論文内容の要旨

空間統計学 (spatial statistics) はいろいろな分野で得られた空間データを分析する学問であり、一般的な統計学と多少異なる発達をしてきた応用統計学の一つである。空間統計学は地理学の鉱産業をはじめとし、経済学、生態学や生物学など多くの学問分野で適用されている。多くの分野で得られた空間データは時空間的な形態を持つ。

本論文では空間統計学の中で、空間予測の問題に焦点を当てた。空間統計学の中で、空間予測する方法はいくつかあるが、クリギング (kriging) という方法を用いて空間予測を行った。クリギング法は、地球統計学 (geostatistical statistics) の中心的な手法であり、特に未観測位置 $\mathbf{s}$ における確率場の値 $Z(\mathbf{s})$ を予測するものである。十分多くの位置で予測を繰り返せば、確率場曲面全体の予測を行うことが可能になる。したがってクリギング法は、空間的予測・補間とも呼ばれる。

地球統計データの分析プロセスとしては、(1) 標本バリオグラム (sample variogram) の推定、(2) 推定された標本バリオグラムへの理論バリオグラム (theoretical variogram) モデルの当てはめ、(3) 当てはめたバリオグラムを用いてクリギングによる空間予測 (クリギング) を行う、の手順で分析が行われる。

地球統計データの分析プロセスにおいて二つに焦点を当てて分析を行った。

一つ目は、バリオグラムの推定における当てはめについてである。地球統計データの分析プロセスの「(2) 推定された標本バリオグラムへの理論バリオグラムモデルの当てはめ」の段階においてよく使われるのは、目で見て決定する方法、最小二乗法、最尤推定法などがある。当てはめの手段として最小二乗法の中で一般化最小二乗法 (ordinary least squares; OLS) と最尤推定法 (maximum likelihood; ML) を使って、両方の方法を比較した。特に、一般化最小二乗法を用いる際にはラグに対するグループを分割することにより間接的バリオグラムパラメータを推定する。このラグの数 (number of lags) は、バリオグラムの推定量に多くの影響を与え、その推定値を用いたクリギング (空間予測) の結果にも影響を与える可能性があるためデータ数による適切なラグの数の選択は重要な問題になる。データ数による最適なラグの数を求めるために leave-one-out 交差検証 (leave-one-out cross-validation; LOOCV) と赤池情報量規準 (Akaike information criteria; AIC) を用いてラグの数の設定するシミュレーション研究を行った。

二つ目は、影響力が大きい観測値の検知についてである。地球統計データにおいて影響力が大きい観測値がある場合には、前で述べた地球統計データの分析プロセスに大きく影響が及び、最終的な目的である未知の位置での値の空間予測 (クリギング) にバイアス (bias) を発生させる。地球統計データを分析する際に最尤法と赤池情報量規準を用いて、標本影響関数 (sample influence functions; SIF) を導いて影響力がある単数あるいは複数の観測値を見つける方法を提案した。そして、シミュレーション研究では、大きい影響力を持つデータを作成し、提案手順にしたがって外れ値を検知できた。シミュレーション研究により、提案方法の有効性を確認できた。適用例として、2010年1月から2012年12月までの日本中国地域の119個の観測所で観測された3年間の最大日降水量データを用いた。提案方法を用いて、影響力が大きい観測値 (外れ値) が検知された。そして、検知された外れ値を取り除く前と外れ値を取り除いた後のバリオグラムモデルのパラメータとクリギングを比較した。外れ値を取り除いた後、バリオグラムモデルのパラメータとクリギングの観点から予測性能が改善されたことが確認された。

論文審査結果の要旨

本論文では、空間統計学の中で地球統計学に関する研究を取り扱っている。地球統計学は、本来鉱床の空間分布の記述と予測を行う鉱山学の分析手法であったが、現在は環境科学、気象学、森林学等の分野でも広く利用されている。地球統計学の中心的手法であるクリギング (kriging) は、空間領域で定義される確率場において、未観測位置の確率場の値を予測・補間する手法である。クリギングは、(1) バリオグラム雲による標本バリオグラムの推定、(2) 理論バリオグラムによる標本バリオグラムのモデル化、(3) 理論バリオグラムを用いたクリギングによる空間予測、の手順で分析が行われる。本論文では、クリギングの精度向上のため2つのアプローチを行っている。

最初のアプローチは、理論バリオグラムのパラメータ推定である。一般に、空間データはある距離まで空間的自己相関が徐々に減少し、ある距離を超えると自己相関が0になる。理論バリオグラムのモデル化では、2点間の非類似度を距離に対してプロットしたバリオグラム雲に対して、プロットを距離によりラグ(区間)に分割し、各ラグでの非類似の平均値に基づき最小二乗法によりパラメータの推定を行う。ラグの数は、理論バリオグラムの推定量及びその推定値を用いたクリギングの結果にも影響を与える可能性がある。本研究では、leave-one-out 交差検証 (LOOCV) と赤池情報量規準 (Akaike information criterion; AIC) を用いて、種々の条件下でシミュレーション研究により最適なラグの数を求める方法について提案している。

次のアプローチは、影響力が大きい観測値の検知についてである。影響力の大きい観測値がある場合には、解析結果に大きく影響が及び、最終的な目的である未知の位置での値の空間予測にバイアスを発生させる。本研究では、地球統計データを分析する際に最大尤度(maximum likelihood)と赤池情報量規準(AIC)を用いて、標本影響関数 (sample influence functions; SIF) を導いて影響力がある単数あるいは複数の観測値を見つける方法を提案している。まず、シミュレーション研究により、大きい影響力を持つデータに対して、提案手法により外れ値を検知出来ることを示し、その有効性を検証している。また、適用例として、2010年1月から2012年12月までの日本中国地域の119カ所の観測所で観測された3年間の最大日降水量データ提案方法を用いて、影響力が大きい観測値(外れ値)を検出している。さらに、検知された外れ値を削除する前と外れ値を削除した後のバリオグラムモデルのパラメータとクリギングを比較し、外れ値を削除した後のバリオグラムモデルのパラメータとクリギングによる予測性能が改善されたこと示している。

これらの研究成果は、3編の査読付き参考論文、2編の関連論文、4件の口頭発表などにおいて公表しており、The 8th Asian Regional Section of the International Association of Statistical Computing において Wakimoto Memorial Fund を受賞している。これらの研究は、空間統計学に関する理論と応用への貢献は大きい。以上により、本論文は博士(学術)に値すると判断した。