



# 時空間現象の統計モデリング

## 西井 龍映

学位: 理学博士(広島大学)

専門分野: 統計学、パターン認識、画像解析

データ取得・転送・蓄積技術の発達により、超高次元データ(例:コンビニでのPOSデータ、ゲノムデータ、人工衛星による地球観測データ等)が観測されるようになりました。このような膨大なデジタルデータ・画像に対し、既存の統計的モデル・解析手法を直接応用しても、有用な情報は得られないどころか、モデルそのものが破綻することがあります。私の研究は膨大なデータから有用な情報を取り出すための統計的手法やモデリングについて考察することです。また時空間データからカテゴリを推測するパターン認識についても研究しています。研究内容として A)、B)2つに大別できます。

### A)時空間データのモデリングと解析手法の開発

企業との共同研究で解析対象となる時系列データや画像について、データの背景および解析目的を理解した上で、次のループで解析手法を考察します。

(1)種々の数理モデル・統計モデルを提案し、(2)モデル推定の方法論の開発およびモデル評価方法を導出し、(3)解析プログラムを開発し、(4)実データで提案手法を評価し、(1)に戻ります。

共同研究が実り多い成果を得るには、企業研究者との密接なコミュニケーションが必要不可欠です。このようなオーダーメイドの共同研究は、多くの場合既存の手法の微修正で解決できますが、新し

い統計的問題に繋がることもあります。それを克服することが共同研究の醍醐味です。なお、某メーカーとの共同研究の成果を共同特許として出願するべく、現在申請書類を作成中です。

### B)多変量データに対する解析方法論の開発

手書き文字認識や生体個人認証のように、パターン認識は多くの重要な応用例があります。時空間データに対しては画素それぞれのカテゴリを推定する問題が議論されています。その場合、空間依存性を数理モデルに取り入れて判別性能を向上させることが重要です。図は広島県の各画素における森林面積比を人口と起伏量から予測したものです。図(c)は(b)に比べてはるかに真のカテゴリ(a)に近いことがわかります。このことから判別においてもデータの空間依存性を適切にモデリングすることが重要であると確認できます。一般に環境データのモデリングと応用を研究課題としています。

また統計手法のうち、最も多く用いられている回帰モデルの推定法について方法論を開発することにも興味があります。標本数より説明変数が多いときの説明変数の選択は、モデル選択基準よりもL1ノルムに基づく縮小推定量が良い予測を与えます。さらに非線形回帰や目的変数が複数の場合のモデリングも重要な研究課題と考えています。

図:広島県の森林比率の予測

