Development and evaluation of analysis methods for science data based on statistical learning theory

学位論文要旨

統計的学習理論に基づく
自然科学データ解析手法の開発および評価

Development and evaluation of analysis methods
for science data based on statistical learning theory

自然科学研究科 電子情報科学専攻

太田 守

Abstract

Natural disasters make us recognize their awfulness and they are also important to predict for our daily lives. Understanding of natural phenomena is therefore imperative in observations of natural disasters. We developed new analysis methods for scientific data based on statistical learning theory. We addressed the following two issues.

1. The remote sensing technology which used weather radar for the analysis of global precipitation phenomena is exploited widely. In space observation, the precipitation particles are usually assumed as raindrops. This assumption is not satisfied for snowflakes or graupel because the scattering characteristics of electromagnetic wave can change variously. It is therefore convenient for understanding of the scattering characteristics to classify the particle type. We proposed an approximation method for Bayesian inference of the Gamma distribution. Applying the proposed method, we estimated the parameters of diameter-velocity distribution of precipitation particles.

2. Investigating characteristics of plasma waves observed by scientific satellites in the Earth's plasmasphere/magnetosphere is the effective key to understand not only generation mechanisms of the waves but also a plasma environment which influences its generation and propagation conditions. The wave distribution function (WDF) method can estimate a WDF as directional distribution of wave energy density. However, in order to determine the solution uniquely, we must assume an adequate model for WDF. We developed new estimation methods for the WDF method, and evaluated the performance by using computer-generated test data.

1　序論

1.1　研究の背景

自然現象は我々の生活と不可分に関わり合い, その神秘性は現象に対する興味関心を惹きつけている. 古来より人々を魅了し続けるオーロラは, 幻想的で荘厳な美しさから興味の対象であるのと同時に畏怖の対象でもあった. また, 2011 年 3 月 11 日に発生した東北地方太平洋沖地震とそれに伴う津波, 2016 年 10 月 8 日に発生した阿蘇山の爆発的噴火, 近年増加している集中豪雨による水害や土砂災害を含む大規模災害からは自然の脅威を感じずにはいられない. その一方で, 人間を含むあらゆる生物の生命存続にとって必要不可欠な水資源は, 自然がもたらす恩恵の一つでもある. 自然現象が様々な恩恵や脅威をもたらす存在であることを深く認識し, 同時に自然とうまく関わっていくために, 現象の定量的な解明および理解を目指す自然科学の手法が発展してきた. 一例として近年注目されている電磁波によるリモートセンシングは, 遠方の環境を広範囲かつ瞬時に観測可能であり, 自然防災利用のために特に有効な観測技術として注目されている. 加えて 2014 年 2 月 28 日には, 初の地球全域の降水の詳細観測を行う全球降水観測計画主衛星が打ち上げられ, また 2016 年 12 月 20 日に放射線帯でのプラズマ総合観測を目的とした ERG 衛星が打ち上げられるなど, これまで実現困難であった高分解能・高精度な観測を行うミッションが数多く計画されている. 自然現象の解明において, 今後得られる観測データはより価値の高い情報を含み, 同時にそのデータ量は膨大になることが予想される. 自然現象の理解および予測の観点において, 計測技術の発達により取得可能となった大規模かつ高精度なデータには非常に多くの有益な情報が含まれていると考えられるが, それにともないデータ解釈の困難性は増しているという問題に直面している. 近年において, 膨大なデータを解析するために統計的学習を応用した特徴抽出手法の研究が盛んに行われているが, その多くは画像・音声といったメディアへの適用に留まり, 自然科学データに対して適用した事例は少ない. 自然現象の解明・予測の重要性を鑑みて, 自然科学データへの統計的学習手法の適用は急務であるといえる.

1.2　研究の目的

本研究の目的は, 自然科学データ解析における統計的学習手法の開発及びその評価を行うことである. 具体的には, 自然現象の理解・災害予測の観点において特に重要と考えられる以下の二つの課題に取り組む.

1. 降水粒子の粒径-落下速度分布のパラメータ推定技術の開発　厳密解を得ることが困難なガンマ分布のベイズ推定の近似手法を確立し, その近似手法により効率的な学習が可能となる降水粒子の粒径-落下速度分布のパラメータ推定を実データ解析に応用する.

2. プラズマ波動の到来方向推定手法の開発　伝搬特性の重要な要素として波動の到来方向に着目し, 衛星による電磁場観測データから到来方向推定を行う手法の開発を目指す. VLF 波動の伝搬方向推定手法の入力データとしてはスペクトルマトリクスが一般的に用いられている. 本研究では, スペクトルマトリクスを入力データとする推定手法の開発と推定の信頼性評価を行う.

2　気象レーダーによる降水強度予測の基礎研究

降水現象は, 水資源として我々の生活に多くの恵みを与えてくれるが, 同時に集中豪雨や, 局所的な豪雪などの自然災害をもたらす. 近年の雪害の増加に伴って, 局所的な降水現象の詳細な解析が求められている. 全球的な降水現象の解析のために, 気象レーダーを用いたリモートセンシング技術が広く使用されているが, これまでの解析では対象となる空間中の降水粒子を雨滴と仮定している. 一般に, 雪片やあられなどの降水粒子の種類によって電波の散乱特性は大きく変化するため, この仮定は成立せず, 降雪現象の解析には従来の手法は適切でない場合がある. このため, 降水粒子を分類し, 各降水粒子ごとの電波散乱特性を把握することは有益である. 特に, 気象レーダーの観測量であるドップラースペクトルと推定対象の降水強度との関係は, 降水粒子の粒径-落下速度分布のモーメントとして与えられるため, 粒径-落下速度分布のパラメータ推定が高精度な降水強度予測の基礎技術として重要となる.

2.1 ガンマ分布に対するベイズ推定の近似手法の提案

降水粒子の粒径分布のモデルとしては, ガンマ分布が用いられることが多い [1, 2]. ガンマ分布は指数分布族であるため事後分布のパラメータの更新は容易に行えるが, 事後分布の期待値の厳密計算は困難である. このため, ガンマ分布に対する効率的なベイズ推定のための近似手法を提案した.

2.1.1 ガンマ分布の諸性質

確率変数 x が以下の確率密度関数

$$\mathrm{Gamma}(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x) \quad (1)$$

に従うとき, x は母数 α, β のガンマ分布に従うという. ここで, $\Gamma(\cdot)$ はガンマ関数である. ガンマ分布の共役事前分布の分布族は, ハイパーパラメータを γ, ε, ν として,

$$p(\alpha,\beta;\gamma,\varepsilon,\nu) \propto \left[\frac{\beta^\alpha\gamma^{\alpha-1}}{\Gamma(\alpha)}\exp(-\beta\varepsilon)\right]^\nu \quad (2)$$

と表現できる. ガンマ分布の共役事前分布のハイパーパラメータが $\{\gamma_0, \varepsilon_0, \nu_0\}$ であるとき, 事後分布のパラメータ $\{\gamma^*, \varepsilon^*, \nu^*\}$ は次式から計算できる.

$$\gamma^* = (\gamma_0^{\nu_0}\mathbb{G}[x]^N)^{\frac{1}{\nu_0+N}} \quad (3)$$

$$\varepsilon^* = \frac{\nu_0\varepsilon_0 + N\bar{x}}{\nu_0+N} \quad (4)$$

$$\nu^* = \nu_0 + N \quad (5)$$

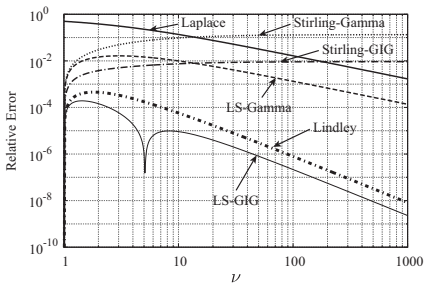
ここで, $\bar{x}$ と $\mathbb{G}[x]$ はそれぞれサンプル数 N からなるデータ群の標本平均と幾何平均を表す.

2.1.2 提案する近似手法

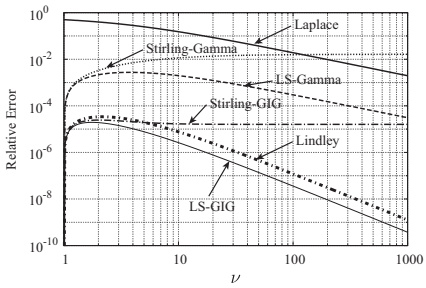
$p(\alpha,\beta;\gamma,\varepsilon,\nu)$ を β で周辺化して得られる分布 $p(\alpha)$ は, Stirling の近似により一般化逆ガウス分布で近似できる. この近似を Stirling-GIG 近似と呼び, Stirling の近似において α^{-1} の項を無視することで得られるガンマ分布による近似を Stirling-Gamma 近似と呼ぶことにする. また, $p(\alpha)$ の最頻値 $\hat{\alpha}$ における対数尤度関数の高階微分係数を用いた近似で, 一般化逆ガウス分布としたものを LS-GIG 近似, ガンマ分布としたものを LS-Gamma 近似と呼ぶことにする. 以上の 4 つの提案手法と, Laplace 近似 [3], Lindley の近似 [4] について誤差評価を行った.

2.1.3 提案近似手法の相対誤差の評価

図 1 に各近似手法により計算した期待値 $\mathbb{E}[\alpha]$ と数値積分による結果との相対誤差を示す. どの $\hat{\alpha}$ の場合



(a) $\hat{\alpha} = 1.0$



(b) $\hat{\alpha} = 10.0$

図 1: $\mathbb{E}[\alpha]$ の相対誤差

にも共通する性質があり, Stirling-Gamma, Stirling-GIG は標本数が多くなるとともに相対誤差が一定値に収束している. これは最頻値が厳密解からずれていることが原因である. その他の Laplace 近似, LS-Gamma, LS-GIG, Lindley の近似は最頻値を厳密に求めているため, 標本数が多くなるにつれて誤差が小さくなる傾向がみられる. 図 1 より, LS-GIG 近似が最も $\mathbb{E}[\alpha]$ の相対誤差が小さいことが分かる. また, var$[\alpha]$ や Hellinger 距離による誤差評価を行うことで, 本提案の LS-GIG 近似によるものが比較した近似手法の中で最適であることを示した. ガンマ分布のパラメータの期待値計算に LS-GIG 近似を用いることで, 数値積分を行うことなく良好な近似値を効率的に計算可能となった.

2.2 降水粒子の粒径-落下速度分布の解析

2.2.1 降雪粒子観測装置と粒径-落下速度分布

富山高等専門学校本郷キャンパスの屋上に図 2 に示す降雪粒子観測装置が設置してある. この観測装置は, 観測領域 (16.4 cm ×12.3 cm ×20.0 cm) の降雪粒子をカメラで撮影する. 得られた画像から降雪粒

子の粒径 D と落下速度 v が計算されている．本研究で使用したカメラは Manta G-033 であり，解像度を 656×492 pixel，フレームレートを 60 fps とし，シャッタースピードは 16000 μs と設定したものを用いた．図 3 に実際に観測装置で観測された粒径-落下速度分布 $N(D, v)$ の例を示す．粒径-落下速度分布 $N(D, v)$ を落下速度で周辺化した $N(D)$ を粒径分布という．降水粒子の粒径分布 $N(D)$ は，ガンマ分布でモデル化されることが多く，

$$N(D; \alpha, \beta) = N_0 \cdot \mathrm{Gamma}(D|\alpha, \beta) \tag{6}$$

と表現される [1, 2]．ここで，N_0 は降水粒子の空間数密度を表す．また，粒径と落下速度の関係式として，

$$v(D; \boldsymbol{w}) = w_0 D^{w_1} \tag{7}$$

が実データに対して当てはまりがよいと報告されている [5, 6]．ここで，$\boldsymbol{w} = [w_0, w_1]^{\mathrm{T}}$ は粒径と落下速度の関係を決定する降水粒子の種類に依存するパラメータである．降水粒子観測装置により取得された粒径，落下速度データから，各パラメータ α, β, $\boldsymbol{w}$ を推定した．

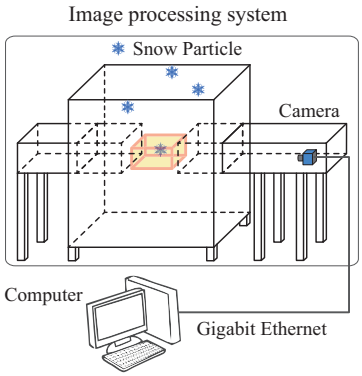


図 2: 降雪粒子観測装置

2.2.2 粒径-落下速度分布の確率モデル

粒径-落下速度分布のパラメータ推定に用いた確率モデルを図 4 に示す．この図は，モデルで扱われる確率変数の依存関係を矢印で表している．$\boldsymbol{D}$ と $\boldsymbol{v}$ はそれぞれ粒径と落下速度の観測データを表す．K はモデルに便宜的に設けられる最大クラスタ数であり，N は全粒子数を表す．また，クラスタ番号 k の粒子種の粒径 D と落下速度 v の関係式として $v = w_{0k} D^{w_{1k}}$ を採用する．加えて，粒径分布をガンマ分布でモデル化することとし，k 番目のクラスタのパラメータを α_k, β_k

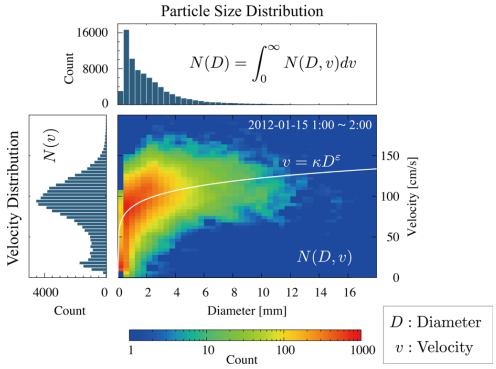


図 3: 粒径-落下速度分布の例

とする．以上をまとめて，$\boldsymbol{\alpha} = \{\, \alpha_k \,\}$, $\boldsymbol{\beta} = \{\, \beta_k \,\}$, $\boldsymbol{w} = \left\{\, [w_{0k}, w_{1k}]^{\mathrm{T}} \,\right\}$ と図中に表記している．また，各粒子 (D_n, v_n) の所属クラスタは多項分布に従う確率変数 z_n によって，クラスタの混合比は確率変数 $\boldsymbol{\pi}$ によって与えられる．降水粒子を特徴づけるモデルパラメータ $\left\{\, \alpha_k, \beta_k, [w_{0k}, w_{1k}]^{\mathrm{T}} \,\right\}$，実効的なクラスタ数を与える混合比 $\boldsymbol{\pi}$，また n 番目の降水粒子データ (D_n, v_n) の所属クラスタ z_n を推定する．なお，本モデルの同時確率は次式で定めた．

$$\begin{aligned} &p(\boldsymbol{D}, \boldsymbol{v}, Z, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}; \boldsymbol{\varphi}_0, \boldsymbol{s}_0, \boldsymbol{c}_0, \boldsymbol{d}_0, \boldsymbol{e}_0, \boldsymbol{f}_0) \\ &= p(\boldsymbol{v}|\boldsymbol{D}, Z, \boldsymbol{w}, \boldsymbol{b}) p(\boldsymbol{D}|Z, \boldsymbol{\alpha}, \boldsymbol{\beta}) p(Z|\boldsymbol{\pi}) p(\boldsymbol{w}|\boldsymbol{a}) \\ &\quad \cdot p(\boldsymbol{a}; \boldsymbol{c}_0, \boldsymbol{d}_0) p(\boldsymbol{b}; \boldsymbol{e}_0, \boldsymbol{f}_0) p(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\varphi}_0) p(\boldsymbol{\pi}; \boldsymbol{s}_0) \end{aligned} \tag{8}$$

ここで，$Z = \{\, z_n \,\}$ とした．$\boldsymbol{\varphi}_0 = \{\, \gamma_0, \varepsilon_0, \nu_0 \,\}$, $\boldsymbol{s}_0$, $\boldsymbol{c}_0$, $\boldsymbol{d}_0$, $\boldsymbol{e}_0$, $\boldsymbol{f}_0$ は各確率変数のハイパーパラメータを表し，条件付き確率はそれぞれ

$$p(\boldsymbol{v}|\boldsymbol{D}, Z, \boldsymbol{w}, \boldsymbol{b}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{LN}(v_n | \boldsymbol{w}_k^T \boldsymbol{\phi}(D_n), b_k^{-1})^{z_{nk}} \tag{9}$$

$$p(\boldsymbol{D}|Z, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mathrm{Gamma}(D_n | \alpha_k, \beta_k)^{z_{nk}} \tag{10}$$

$$p(Z|\boldsymbol{\pi}) = \prod_{n=1}^{N} \prod_{k=1}^{K} {\pi_k}^{z_{nk}} \tag{11}$$

$$p(\boldsymbol{w}|\boldsymbol{a}) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{w}_k | \boldsymbol{0}, a_k^{-1} \boldsymbol{I}) \tag{12}$$

$$p(\boldsymbol{a}; \boldsymbol{c}_0, \boldsymbol{d}_0) = \prod_{k=1}^{K} \mathrm{Gamma}(a_k | c_0, d_0) \tag{13}$$

$$p(\boldsymbol{b}; \boldsymbol{e}_0, \boldsymbol{f}_0) = \prod_{k=1}^{K} \mathrm{Gamma}(b_k | e_0, f_0) \tag{14}$$

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\varphi}_0) \propto \prod_{k=1}^{K} \left[\frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} {\gamma_0}^{\alpha_k - 1} \exp(-\beta_k \varepsilon_0) \right]^{\nu_0} \tag{15}$$

$$p(\boldsymbol{\pi}; \boldsymbol{s}_0) \propto \prod_{k=1}^{K} \pi_k^{s_0 - 1} \tag{16}$$

と定義した．ここで，$\boldsymbol{\phi}(D_n) = [1 \ \ln(D_n)]^{\mathrm{T}}$ である．

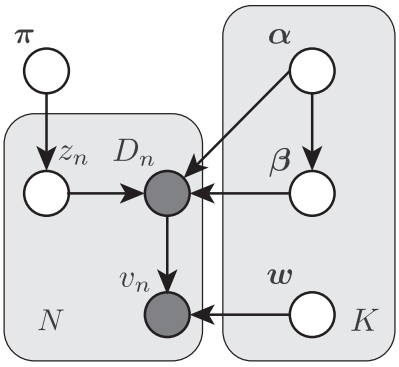


図 4: 粒径-落下速度分布のグラフィカルモデル

また，各確率変数の事前分布には共役事前分布を与えた．$\mathcal{N}(x|\mu, \sigma^2)$ と $\mathcal{LN}(x|\mu, \sigma^2)$ はそれぞれ正規分布と対数正規分布の確率密度関数を表す．本研究では，以上で述べたモデルパラメータを次式で定義される変分事後分布を用いて変分ベイズ法により推定した．

$$\begin{aligned} &q(Z, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) \\ &\qquad = q(Z) q(\boldsymbol{w}) q(\boldsymbol{a}) q(\boldsymbol{b}) q(\boldsymbol{\alpha}, \boldsymbol{\beta}) q(\boldsymbol{\pi}) \end{aligned} \tag{17}$$

このモデルは K を十分大きくとることで，関連度自動決定 [7] により最適なクラスタ数が自動的に決定される．また，変分事後分布の落下速度に関わるパラメータについて無情報事前分布を仮定し更新則を適用しないことで粒径分布に対する混合ガンマ分布のクラスタリング手法として本モデルを用いることができる．また，変分ベイズ法の更新則に現れるガンマ分布のパラメータの期待値計算に本提案の LS-GIG 近似を用いて推定の効率化を図ることで大規模データの解析に適用が可能となった．

擬似データによる検討の結果，良好に関連度自動決定が機能した $K = 10$ として，周辺尤度の下限の最大化によるハイパーパラメータの最適化を行い，粒径-落下速度分布のパラメータ推定を行った．

2.2.3 粒径データを用いたパラメータ推定結果

図 5 に雪片とあられが同時に降っている期間の粒径分布と，前項で述べた推定手法によるクラスタリング結果を示す．用意された 10 個のクラスタの内，関連度自動決定により 8 個のクラスタの混合比は零へと縮退し，自動的に 2 つのクラスタが存在していることを学習しており，目視で確認した降雪種と合致する推定結果となった．

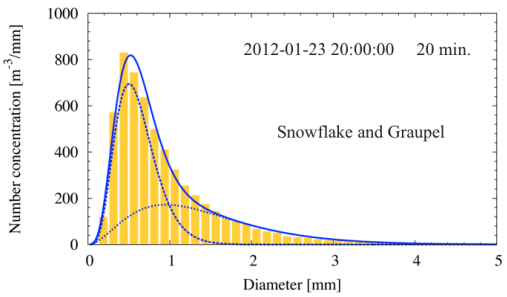


図 5: 粒径分布のパラメータ推定結果

2.2.4 粒径，落下速度データを用いた推定結果

図 6 に粒径-落下速度分布のパラメータ推定により得られた各クラスタの平均粒径と平均落下速度の時系列変化を示す．平均粒径の大きいクラスタについて，時間とともに次第に平均粒径が大きくなっていることから，この時間帯において小さい雪片から次第に大きな雪片へと遷移していく過程が確認できる．平均粒径の時系列変化からは，平均粒径の大きな雪片と平均粒径が小さい降雪粒子の 2 種しか判別できないが，同時刻の平均粒径の小さいクラスタの平均落下速度をみると，あられの時間帯では速度が大きく，それ以降の時間帯での速度が小さくなっていることが分かる．つまり，粒径-落下速度速度分布のクラスタリングにより小さい降水粒子，雪片，あられの 3 種類に分類できる可能性が示された．

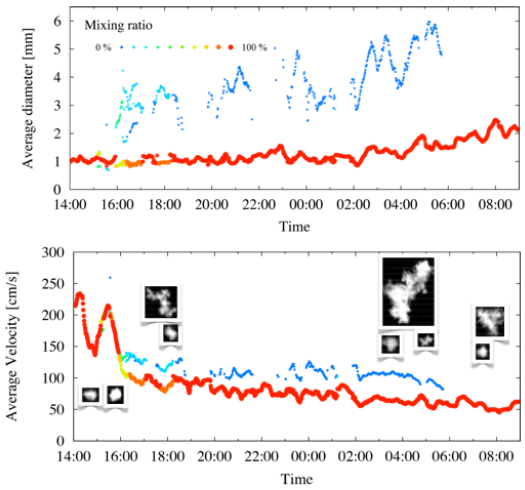


図 6: 平均粒径，落下速度の時系列変化

2.2.5 ドップラースペクトルの事後予測分布

ドップラーレーダーの観測量であるドップラースペクトル $Z_{\mathrm{doppler}}(v)$ は, 粒径-落下速度分布のパラメータにより

$$Z_{\mathrm{doppler}}(v) = \sigma(D, v) \cdot N_0 \cdot \mathrm{Gamma}(D|\alpha, \beta) \cdot \frac{\partial D}{\partial v} \quad (18)$$

として与えられる．ここで, $\sigma(D, v)$ は粒径 D, 落下速度 v の降水粒子の後方散乱断面積である．3 種の降水粒子に分類されたという前項の結果を用いて計算した, 各降水粒子ごとのドップラースペクトルの事後予測分布を図 7 に示す．各降雪種でドップラースペクトルが大きく異なる形状をとることから, 実際にドップラーレーダーの観測量から観測領域の降雪粒子種の比率を推定可能であることが示唆された．各降水種ごとのドップラースペクトルと降水強度の関係を計算しておくことで, 降水強度の高精度推定が可能になると考えられる．

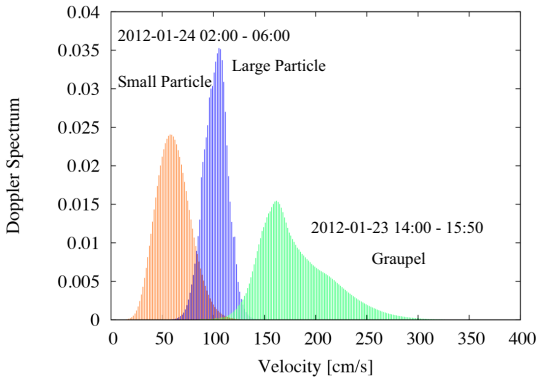


図 7: ドップラースペクトルの事後予測分布

3 スペクトルマトリクスを用いた到来方向推定手法の開発

宇宙空間の本格的な利活用に伴い, 地球磁気圏内の物理現象を解明することの重要性が増している．地球磁気圏を飛翔する科学衛星の電磁場観測データは, 地球周辺環境の把握のために必要な情報を豊富に含んでいると考えられる．情報の活用のためには, データから波動の伝搬特性を推定するための手法の開発が欠かせない．

3.1 到来方向推定手法の概要

3.1.1 スペクトルマトリクスの定義

科学衛星に搭載されるプラズマ波動観測器は, 衛星の位置における電場と磁場の波形観測を行う．到来方向推定を行う際, 観測信号はバンドパスフィルタにより解析対象の周波数帯域の解析信号 $\boldsymbol{z}(t)$ へと変換される．観測時刻 $\{ t_1, t_2, \ldots, t_N \}$ における観測データ $\{ \boldsymbol{z}(t_n) \}$ の相互相関行列

$$\boldsymbol{S} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{z}(t_n) \boldsymbol{z}(t_n)^{\dagger} \quad (19)$$

はスペクトルマトリクスと呼ばれ, 伝搬ベクトル推定において広く用いられている．以降では, N をスペクトルマトリクスの平均回数と呼ぶことにする．また, 定義よりスペクトルマトリクス $\boldsymbol{S}$ は観測信号の成分数を D として D 次の正定値エルミート行列である．

3.1.2 波動分布関数法

Storey ら [8] によって提案された波動分布関数法では, 観測点に到来する波を異なる周波数, 伝搬方向をもつ無限個の平面波と考え, エルゴード的な定常不規則過程とみなせると仮定する．また, 観測点に到来する複数の波動のエネルギー密度分布を, 波動分布関数 $F(\theta, \phi)$ として表現する．ここで, 伝搬方向を表すベクトル $\hat{\boldsymbol{k}}$ に対して, $\hat{\boldsymbol{k}}$ と地球磁場とのなす角を天頂角 θ, $\hat{\boldsymbol{k}}$ を x-y 面に投影したベクトルが x 軸となす角を方位角 ϕ と定義した．スペクトルマトリクスは波動分布関数と以下の積分方程式で関連づけられる．

$$\mathbb{E}[\boldsymbol{S}] = \frac{\pi}{2} \int_0^{2\pi} \int_0^{\pi} \boldsymbol{a}(\theta, \phi) F(\theta, \phi) \sin\theta \mathrm{d}\theta \mathrm{d}\phi \quad (20)$$

スペクトルマトリクス $\boldsymbol{S}$ の平均回数が十分に多ければ確率的な揺らぎは小さくなり, 上式の右辺の期待値を観測値に置き換えた近似式が適用できる．式中に表れる $\boldsymbol{a}(\theta, \phi)$ は (θ, ϕ) 方向の伝搬ベクトルを持つ波動の単位エネルギー当たりのスペクトルマトリクスであり, 積分核と呼ぶ [9]．積分核は波動の周波数と伝搬モード, サイクロトロン周波数とプラズマ周波数をもとに理論的に計算可能である．式 (20) に基づき観測によって得られるスペクトルマトリクスと理論から計算される積分核を既知として, $F(\theta, \phi)$ を推定する手法を波動分布関数法という [8]．この方程式の解は一般的には一意に定まらないため, 波動分布関数に何らかのモデルを与える必要がある．

3.2 マルコフ確率場モデルに基づく波動分布関数法

3.2.1 提案手法の概要

マルコフ確率場 (MRF) モデルに基づく波動分布関数法 (以下, 本手法と表記する) では, 到来方向を離散化して推定問題を取り扱う. 本手法では, 離散化数を K として波動分布関数を表すベクトル $\boldsymbol{f} \in \mathbb{R}^K$ と, 観測されたスペクトルマトリクスの独立成分 L 個からなる実ベクトル $\hat{\boldsymbol{s}} \in \mathbb{R}^L$ との関係を, 積分核行列 $\boldsymbol{A} \in \mathbb{R}^{L\times K}$ を用いて

$$\hat{\boldsymbol{s}} = \boldsymbol{A}\boldsymbol{f} + \varepsilon \boldsymbol{a}_\varepsilon + \boldsymbol{e} \tag{21}$$

と表現する. この式は, 式 (20) を離散化して得られるものである. ここで, $\boldsymbol{a}_\varepsilon$ は白色ノイズに対応する積分核であり, その強度は確率変数 ε で与えられる. また, $\boldsymbol{e} \in \mathbb{R}^L$ は誤差ベクトルを表す. 図 8 に提案する確率モデルのパラメータ依存関係を示す. ここで, 図中の α, β はそれぞれ $\hat{\boldsymbol{s}}$, $\boldsymbol{f}$ の事前分布のパラメータであり本手法では確率変数として取り扱われる. また, 図 8 に示した依存関係から, MRF モデルの同時確率は

$$p(\hat{\boldsymbol{s}}, \boldsymbol{f}, \varepsilon, \alpha, \beta) = p(\hat{\boldsymbol{s}}|\boldsymbol{f}, \varepsilon, \alpha)p(\boldsymbol{f}|\beta)p(\varepsilon)p(\alpha)p(\beta) \tag{22}$$

と表現できる. 以下に, 各モデルパラメータに与えた確率分布を示す.

$$p(\hat{\boldsymbol{s}}|\boldsymbol{f}, \varepsilon, \alpha) = \mathcal{N}_L(\hat{\boldsymbol{s}}|\boldsymbol{A}\boldsymbol{f} + \varepsilon\boldsymbol{a}_\varepsilon, \alpha^{-1}\boldsymbol{I}) \tag{23}$$

$$p(\boldsymbol{f}|\beta) = \frac{\beta^{\frac{K}{2}}|\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}|^{\frac{1}{2}}}{(2\pi)^{\frac{K}{2}}} \exp\left(-\frac{\beta}{2}\|\boldsymbol{C}\boldsymbol{f}\|_2^2\right) \tag{24}$$

ここで $\mathcal{N}_L(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ は確率変数 $\boldsymbol{x}$ が平均 $\boldsymbol{\mu}$, 分散共分散行列 $\boldsymbol{\Sigma}$ の L 変量正規分布に従うことを表す. また, β^{-1} は事前分布の分散に関するスケール変数であり, $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} \in \mathbb{R}^{K\times K}$ は MRF の相関行列を表す. 本手法では, ϵ, α, β の事前分布には共役事前分布を与えた. また, 変分事後分布を

$$q(\boldsymbol{f}, \varepsilon, \alpha, \beta) = q(\boldsymbol{f})q(\varepsilon)q(\alpha)q(\beta) \tag{25}$$

として, 変分ベイズ法を用いて近似的にモデルパラメータの事後分布を推定する. このとき, 変数の定義により $\boldsymbol{f}$ と ε が非負の値をとることから $\mathbb{E}[\boldsymbol{f}] \geq \boldsymbol{0}$, $\mathbb{E}[\varepsilon] \geq 0$ の制約条件のもと推定を行う. また, 本手法の推定結果を学習された近似事後分布の最頻値として定義する.

3.2.2 疑似データを用いた提案手法の評価

本提案手法 (MRF) の波動分布関数の推定像を, 波動の到来方向が既知の疑似データを用いて評価した. 比較対象として, 最大エントロピー法 (MEM), Phillips-Tikhonov 正則化法 (PTM) を用いた. プラズマ周波数を 300 kHz, サイクロトロン周波数を 50 kHz, 波動の周波数を 3 kHz としてホイスラーモードを仮定した擬似データを作成した.

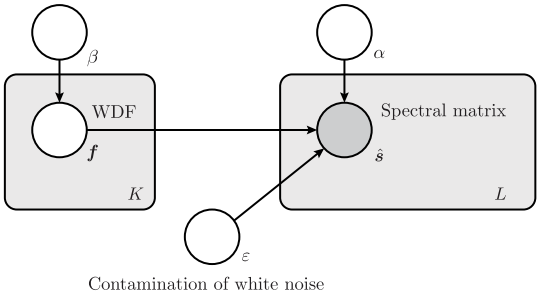
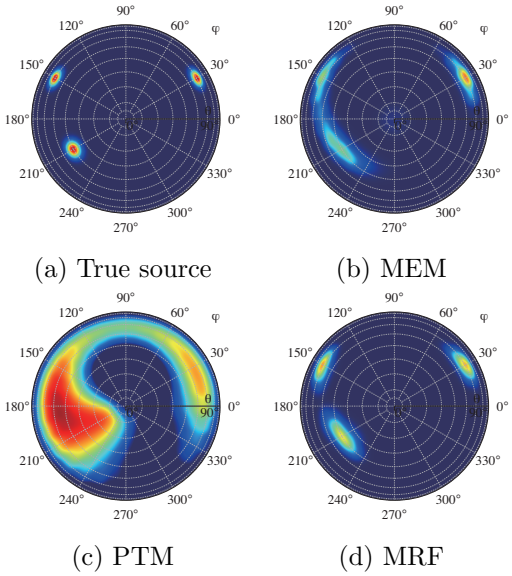


図 8: 提案する確率モデルのパラメータ依存関係



(a) True source (b) MEM

(c) PTM (d) MRF

図 9: 三つの点波源の場合の各モデルの推定結果

1. ノイズを含まない擬似データを用いた評価 強度, 広がりがともに等しい三つの到来波を仮定し, 各波の到来方向が $(\theta_1, \phi_1) = (60°, 30°)$, $(\theta_2, \phi_2) = (60°, 150°)$, $(\theta_3, \phi_3) = (40°, 210°)$ である疑似データに対する各手法の推定結果を図 9 に示す. PTM では, 全ての波源を覆うような広がった推定像が得られ, MEM では二つ目の波源と三つ目の波源が一つの広がった波源として推定されていることが分かる. そ

れに対して, MRF は波源の広がりは元データに比べて, やや大きくなるものの三つの波源の到来方向が良好に推定されていることが分かる.

2. ノイズを含む疑似データを用いた評価　到来方向が $(\theta_1, \phi_1) = (50°, 120°)$ である到来波を仮定し, S/N 比を変えて擬似データを作成し, 疑似データの波動分布関数と各手法の推定像の平均二乗誤差 (MSE) を計算した結果を図 10 に示す. 各 S/N 比ごとに異なるノイズを 100 パターン作成し, 図には 100 回の試行での MSE の平均値を点で, エラーバーでその試行での MSE の最小値, 最大値を示した. S/N 比が 45, 50[dB] の場合にはほぼノイズの影響はないと考えられるため, 点波源を良好に推定できていた MEM, MRF の MSE は小さい値をとり, 推定像が広がってしまう PTM では MSE が大きくでていることが分かる. また, S/N 比が低くなるにつれ提案手法 (MRF) は従来手法に比べ MSE が小さくなり, ノイズに頑健であることが明らかとなった.

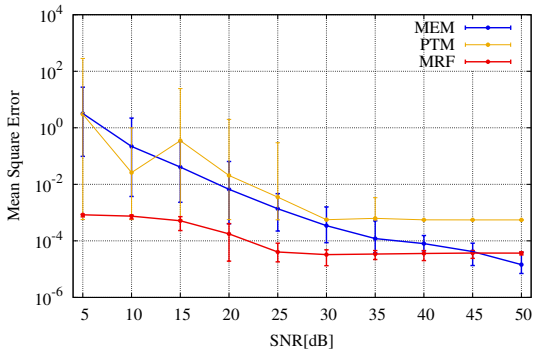


図 10: S/N 比に対する各推定手法の推定像の MSE

3.3 恣意的なモデルを排除した波動分布関数法

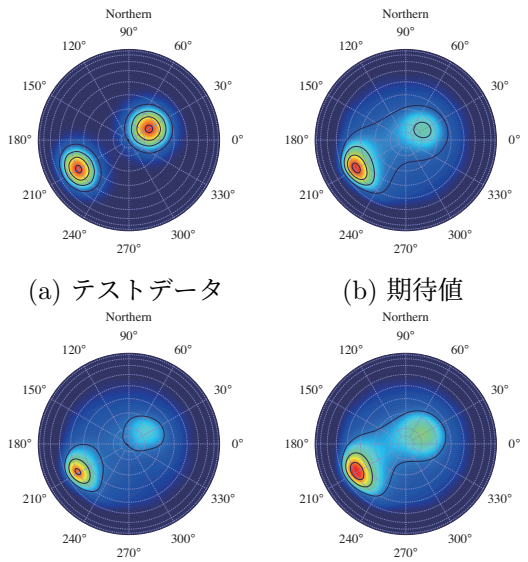
p 次元の観測信号が, 平均 $\mathbf{0}$ で積分核から理論的に計算されるスペクトルマトリクス $\boldsymbol{S}$ を分散共分散行列としてもつ定常ガウス過程に従うことを仮定する. このとき, $\boldsymbol{S}$ は積分核と波動分布関数と以下の関係をもつ.

$$\boldsymbol{S} = \frac{\pi}{2}\int_0^{2\pi}\int_0^{\pi} \boldsymbol{a}(\theta,\phi)F(\theta,\phi)\sin\theta \mathrm{d}\theta\mathrm{d}\phi \tag{26}$$

このガウス過程の母数としてのスペクトルマトリクスと観測値を区別するために, 科学衛星の観測データを用いて平均回数 N で作成されるスペクトルマトリクスの観測値を $\boldsymbol{W}$ と表現する. このとき, 平均回数 N のスペクトルマトリクス $\boldsymbol{W}$ の生起確率を表す尤度分布は,

$$p(\boldsymbol{W}|\boldsymbol{S}) = \frac{N^{Np}\cdot|\boldsymbol{W}|^{N-p}e^{-N\cdot\mathrm{tr}(\boldsymbol{S}^{-1}\boldsymbol{W})}}{\pi^{\frac{p(p-1)}{2}}|\boldsymbol{S}|^N\prod_{k=1}^{p}\Gamma(N-k+1)} \tag{27}$$

で与えられる. この尤度分布と無情報事前分布から得られる波動分布関数の事後分布から, その実現値をギブスサンプリング [7] を用いて複数個計算することで, その期待値と信用区間を求めることができる. 図 11 に二つの波源からなる擬似データを, 平均回数を 200 回として推定した結果を示す. 期待値と信用区間の両方において二つの峰をもつ分布構造が現れており, 二つの波源を特定できていることが分かる. このように信用区間と期待値を合わせて確認することで, 平均回数に対して推定像にどの程度の揺らぎが存在するのかを恣意的なモデルを用いずに評価できる. この方法は, 波動分布関数法の推定像の性質の解明や, より高速で信頼性の高い推定手法を開発していくための理論基盤として有効に利用できると考えられる.



(a) テストデータ　(b) 期待値

(c) 95%下側信用区間　(d) 95%上側信用区間

図 11: 二つの広がった波源の場合の推定結果 (平均回数 $N = 200$)

4 結論

統計的学習理論に基づく自然科学データの解析技術の開発およびその評価として二つの課題に取り組んだ.

まず, レーダー気象学にとって重要な基礎研究の一つである降水粒子の粒径-落下速度分布のパラメータ推定問題を扱った. 粒径分布のモデルとしてよく用いられているガンマ分布のベイズ推定は, 事後分布の期待値が解析的に計算できない. この問題点を解決するためにガンマ分布のベイズ推定のための近似期待値計算手法の提案を行った. 提案手法を用いることで大規模な降水粒子観測データに対して推定手法を適用できるようになった. この粒径-落下速度分布のパラメータ推定手法を, 降水粒子観測装置の観測データに適用し, 良好に降水粒子のクラスタリングを行えていることを確認した. また, 粒径-落下速度分布のクラスタリングにより得られたモデルパラメータから, 各降雪種のドップラースペクトルの事後予測分布を計算した. ドップラーレーダによって取得されるドップラースペクトルとの比較によりレーダの観測値のみから降水粒子種の比率を推定できる可能性と, その結果を用いた高精度な降水強度推定の可能性が示唆された.

次に, 科学衛星によって得られる観測波形データから作成されるスペクトルマトリクスを用いた伝搬ベクトル推定手法を開発した. また, マルコフ確率場モデルに基づく波動分布関数法の提案し, 擬似データを用いて従来手法の推定像との比較, 評価を行った. その結果, ノイズを含む擬似データに対しても本提案手法が平均二乗誤差の観点で従来手法に比べて安定に推定可能であることが分かった.

最後に, スペクトルマトリクスの平均回数を考慮した尤度分布を用いることで, 解を一意に定める必要を無くし恣意的なモデルを排除した波動分布関数法による推定を可能とした. この手法を用いることで, 平均回数による推定像の揺らぎを把握でき客観的に解の良さが評価可能になる. また, 効率的な到来方向推定アルゴリズムの開発に当たっての推定像評価の判断基準として本手法が有効に応用できると考えられる.

本研究で用いた手法は, 統計的学習理論に基づくデータの性質によらない汎用的なものであり, その応用範囲は広いといえる. 本研究で扱った, 工学的な応用として重要なクラスタリングと逆問題解法の開発・評価により, 自然現象の機構解明の一助となるだけではなく, 実観測データを用いた学習手法の定量評価を通じて機械学習分野の発展に貢献できれば幸いである.

参考文献

[1] 深尾昌一郎, 浜津享助. 気象と大気のレーダーリモートセンシング. 京都大学学術出版会, 2009.

[2] Guifu Zhang, Juanzhen Sun, and Edward A. Brandes. Improving parameterization of rain microphysics with disdrometer and radar observation. *Journal of the Atmospheric Sciences*, Vol. 63, No. 4, pp. 1273–1290, 2006.

[3] 元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇（編）. パターン認識と機械学習（上）: ベイズ理論による統計的予測. 丸善出版, 東京, 2007.

[4] D. V. Lindley. Approximate bayesian methods. *Trabajos de Estadistica Y de Investigacion Operativa*, Vol. 31, No. 1, pp. 223–245, 1980.

[5] 椎名徹, 村本健一郎. 降雪粒子の落下速度と落下中の粒径および融解直径の関係. 電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理, Vol. 81, No. 6, pp. 1108–1114, jun 1998.

[6] 村本健一郎, 椎名徹. 画像処理による降雪粒子の大きさと落下速度の自動測定. 電子情報通信学会論文誌 D-2 情報・システム, Vol. 72, No. 9, pp. p1382–1387, sep 1989.

[7] 元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇（編）. パターン認識と機械学習（下）: ベイズ理論による統計的予測. 丸善出版, 東京, 2008.

[8] L. R. O. Storey and F. Lefeuvre. The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma I. The direct problem. *Geophysical Journal International*, Vol. 56, No. 2, pp. 255–269, 1979.

[9] L. R. O. Storey and F. Lefeuvre. The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma — II. The integration kernels. *Geophysical Journal International*, Vol. 62, No. 1, pp. 173–194, 1980.

平成　29年　2月　1日

学位論文審査報告書（甲）

1．学位論文題目（外国語の場合は和訳を付けること。）

統計的学習理論に基づく自然科学データ解析手法の開発および評価

2．論文提出者　(1) 所　　属　電子情報科学　　専攻

(2) 氏　　名　太田　守（おおた　まもる）

3．審査結果の要旨（600～650字）

平成29年2月1日に第1回学位論文審査委員会を開催した後、口頭発表を行った。その直後に、第2回審査委員会を開いて慎重審議を行った結果、以下の通り判定した。なお、口頭発表における質疑を最終試験に代えるものとした。

本論文は、統計的学習理論に基づく自然科学データの解析技術およびその評価法の開発を目的として、二つの課題に取り組んでいる。まず、レーダー気象学にとって重要な降水粒子の粒径-落下速度分布のパラメータ推定問題に対し、近似解法を用いてパラメータを効率的に求める方法を提案した。さらに同手法を降水粒子種毎の粒径-落下速度分布のパラメータ推定に応用し、レーダー観測データから降水粒子の混合率が推定できることを示した。次に、科学衛星で計測した電磁波の伝搬方向推定法の改良に取り組んだ。まず、マルコフ確率場モデルに基づく波動分布関数の求解法を新たに提案し、データに雑音が含まれる場合でも頑健かつ信頼性の高い推定解が得られることを示した。さらに、測定データの平均回数に基づく尤度分布を利用することで、恣意的な仮定なしに推定結果を得る方法を開発した。このように本研究は、多種多様なパラメータに依存する自然科学観測データから、客観的かつ信頼度の高い解析結果を得る手法を提案したもので、大変重要な成果といえる。以上より、本論文は博士（工学）に値すると判定した。

4．審査結果　(1) 判　定（いずれかに○印）　合　格　・　不合格

(2) 授与学位　博　士（　工　学　）