

1 表題 階層ベイズによる複雑な生態学的プロセスの推定：ブナの花粉散布空間パターン推
2 定を例に

3 簡略表題 階層ベイズによる生態学的プロセスの推定

4 表題の英訳 Application of hierarchical Bayes model on estimating complex ecological
5 processes: an example from spatial pattern of pollen dispersal of Siebold's beech

6 著者 1 富田 基史 (Motoshi Tomita)*

7 東北大学大学院農学研究科 (Graduate School of Agricultural Science, Tohoku
8 University)

9 著者 2 花岡 創 (So Hanaoka)

10 岐阜大学大学院連合農学研究科 (United Graduate School of Agricultural Science,
11 Gifu University)

12 著者 1 の連絡先

13 住所 989-6711 宮城県大崎市鳴子温泉字蓬田 232-3

14 電話番号 0229-84-7359

15 Email motoshi@bios.tohoku.ac.jp

16 著者 2 の連絡先

17 住所 501-1193 岐阜市柳戸 1-1

18 電話番号 058-293-2856

19 Email so1223@nifty.com

20 要旨

21 野外の生物を対象に観測したデータは，多くの生態的なプロセスが複雑に作用し合った
22 結果として得られたものであるため，そこから興味のあるプロセスを推定することがしば
23 しば困難なことがある．本稿ではブナの花粉散布空間パターンを例に，階層ベイズを用い
24 て複雑な生態学的プロセスを推定するための基本的な考え方と実際の手順を解説する．ま
25 ず比較的簡単なモデルを用いて，尤度を元にした統計モデリングの基本的な考え方を紹介
26 し，最尤推定によってパラメータを推定する．そして，簡単なモデルの問題点を明らか
27 にしたうえで，階層ベイズモデルによってモデルの改良を試みる．最後に Markov chain
28 Monte Carlo (MCMC) 法によってモデルのパラメータを推定し，最尤推定の結果と比較
29 することで階層ベイズの利点を示す．これらの過程を通じて，統計モデリングと階層ベ
30 ズの考え方を紹介したい．

31 はじめに

32 野外の生物を対象に観測したデータは、多くの生態的なプロセスが複雑に作用し合った
33 結果として得られたものである。生態学研究者は、興味がある生態的プロセスを観測した
34 データから推定して議論するわけだが、すべての要因を観測して、それらの相互作用を包
35 括的に推定して議論することは不可能である。したがって多くの要因を一定にコントロー
36 ルした野外実験系を用いるか、それらの影響を無視して少数の要因のみを対象に議論する
37 ことが普通である。このようにある仮説のもとで現象を単純化したモデルを構築して、よ
38 りデータに適合するようにパラメータを推定することが、統計モデルによるデータ解析の
39 基本であると言える。

40 生態学研究者が統計モデルについて意識するきっかけとなったのは、一般化線形モデル
41 (GLM) の普及によって様々な形式のデータに対して使える選択肢が増えたことだろう。
42 誤差構造やパラメータセットが異なるいくつかのモデルを試すことが一般的となったこと
43 で研究者は、各要因の効果が「統計的に有意」であるかだけでなく、モデル(仮説)がど
44 れだけよく現象を説明できるかという点を意識するようになった。しかしこれは、GLM
45 などの単純な統計モデルの限界を研究者に気づかせるきっかけにもなった。

46 単純化した統計モデルの限界としてまず挙げられるのが、観測していない要因の影響で
47 ある。野外で観測したデータはしばしば、着目した要因だけでは説明しきれないほどのば
48 らつきを持つことがあり、モデルのあてはまりを著しく低下させるばかりか、着目した
49 要因の効果を誤って評価させることすらある(詳しくは久保・粕谷(2006)を参照されたい)。
50 また観測した要因同士の関係が非線形であったり複雑な因果関係を持っているなど
51 の場合についても、これまでの統計モデルでは対処できないことが多かった。前者につい
52 ては、観測していない要因の効果を「個体」や「場所」ごとに「ランダム効果」としてモ
53 デルに組み込むことで、簡単な例についてはこの問題を解決できる(一般化線形混合モデ
54 ル: GLMM など)。後者については、着目する要因をしぼるなどモデルを単純化すること
55 で、部分的な推論はできる。

56 しかしながら大規模な野外観測プロジェクトなどにより従来とは比べ物にならないほど
57 のデータが蓄積し、生態学研究者の興味・関心が個々の単純なプロセスからより複雑なプ
58 ロセス全体を理解することへと広がっている今、これまで一般的であった統計モデルの手
59 法だけでその要望に答えることには限界がある。より複雑な仮説を柔軟に検証できる手法
60 が必要である。

61 柔軟なモデリングが可能であることから、階層ベイズモデルはこれまでの統計学的モデ

62 ルでは不可能であった複雑な生態的プロセスを推定を可能にした．本稿ではまず，モデ
63 リングの基本的な考え方を解説したうえで，具体例として筆者らが取り組んだ花粉散布
64 パターンの推定を例に，階層ベイズを用いて複雑な生態的プロセスを推定する例を紹介
65 する．

66 回帰モデルや GLM・GLMM などの従来よく使われてきたの統計的モデルでは，パラ

67 実例：ブナの花粉散布パターンの推定

68 階層ベイズを用いて複雑な生態学的プロセスを推定する例として，本稿では筆者らが
69 取り組んだブナ (*Fagus crenata* Blume) の花粉散布の空間パターン推定をとりあげる
70 (Hanaoka et al. 投稿中)．固着生の樹木にとって花粉散布は自らの遺伝子を移動させる
71 もっとも重要なプロセスであるため，集団の遺伝的な多様性や構造の研究において古くか
72 ら議論されてきた．近年では，植物の繁殖戦略や花粉媒介者の効率，自殖・近親交配の程
73 度の評価，森林の分断化と遺伝子流動の関係など，生態学分野の多くの研究におて花粉散
74 布パターンの調査が行われている (田中ほか 2006)．また応用科学分野においては，例え
75 ば遺伝子組み換え作物や植林活動に伴う自然集団への遺伝子流出リスクを評価するため
76 に，「花粉はどこまで・どれくらい飛んでいくのか？」といった散布パターンの定量的な
77 評価が求められている．

78 樹木の花粉散布パターンを調べる方法はいくつかあるが，遺伝マーカを用いた親子解析
79 技術を用いるのが近年では一般的である．親子解析によって分かることは，特定の母親に
80 それぞれの花粉親がどれだけの割合で受粉させることができたかという情報だけである．
81 それに個体同士の位置関係を組み合わせることで，分子生態学の研究者は樹木の花粉散布
82 パターンを議論してきた．ところが，遺伝マーカで推定できる親子関係には，花粉散布パ
83 ターン以外にも個体ごと花粉生産量や開花時期の同調性などといった多くの要因が影響す
84 ることが予想できる．さらに，ブナのような風媒花をつける樹木の花粉散布は，開花期の
85 気象条件や地形・植生の影響を受ける．例えば，開花期の風向・風速は花粉が移動する方
86 位・距離に影響を与える．また，尾根や密な林分は花粉が移動する障壁となることが知ら
87 れている (e.g. Levin, Kerster 1974)．したがって，花粉散布パターンそのものもこれら
88 の要因によって複雑なパターンを示すと考えられる．そのため，複雑な生態的プロセスの
89 推定が可能な階層ベイズモデルは，花粉散布パターンの定量的な解析にきわめて有効であ
90 ると思われる．

91 この節ではこれまで一般的だった親子解析にもとづく花粉散布研究の流れとその問題点
92 を明らかにしたうえで，まず単純な仮定のもとで花粉散布パターンの統計学的なモデリン

93 グに取り組む．そして階層ベイズの考え方を利用して単純なモデルを改良し，より複雑な
94 モデルを構築する実際の手順を紹介する．最後に，単純なモデルと階層ベイズモデルの推
95 定結果を比較することで，階層ベイズの有効性を議論したい．

96 調査の概要

97 調査地は，岐阜大学応用生物科学部附属岐阜フィールド科学教育研究センター位山演習
98 林内の約 9ha の天然林であり，主な構成樹種はブナとミズナラ (*Quercus mongolica* var.
99 *grosseserrata* Rehd. et Wils.) である．このうち中心部に約 4ha のプロットを設置して，
100 すべての成木個体の位置とサイズを測定し葉を採取した．ブナが開花・結実した 2005 年
101 の 9 月に調査地の中心部から 4 個体の母親を選んで，種子を採取した．

102 成木 143 個体・種子 673 個から DNA を抽出して，マイクロサテライトマーカ 7 遺伝子
103 座の遺伝子型を解析した．ソフトウェア FaMoz (Gerber et al. 2003) を用いて親子解析
104 を行い，493 個の種子の花粉親を特定できた (図 1)．本稿ではこれらのデータを用いる．

105 親子解析で樹木の花粉散布を調べる

106 もっとも簡単な例である図 2a は，親子解析によって推定された母親-花粉親ペア間の距
107 離の頻度分布を全ペア間の距離の頻度で割ったものである．ここから母親に近い花粉親か
108 ら多くの花粉が飛んできていることが容易に分かる．同様の指標を母樹から見た方位ごと
109 にプロットしてみると，すべての方向から均一に花粉が飛んできているというよりは，特
110 定の方角に偏っているように見える (図 2b)．

111 ところが，これらの図はいずれも母親から見て花粉親となった個体がどの距離・方位に
112 どれだけいるかということを表しているだけで，花粉散布パターンそのものを表している
113 わけではない．例えば，花粉親の中に他の個体よりも多く花粉を散布するものがいた場
114 合，それらの個体が含まれる距離・方位から飛んでくる花粉の量を過大評価することにな
115 る．また，ある方位には遠距離の個体が多いなど，個体の分布に偏りがあった場合にも同
116 様に結果に偏りが生じる．このように，これら図が表しているのは個体の位置やサイズな
117 どの要因と花粉散布・花粉生産などの生態的プロセスの結果として観察された“見かけ上
118 の花粉散布パターン”にすぎず，「花粉は近い個体から多く来ているようだ」「特定の方向
119 から多く来ているようだ」といった定性的なことしか分からない．花粉散布パターンを定
120 量的に評価するためには，“見かけ上の花粉散布パターン”を眺めるのではなく，生態的プ
121 ロセスそのものである“真の花粉散布パターン”を何らかの方法で推定する必要がある．

122 花粉散布パターンのモデル化

123 階層ベイズによって複雑なプロセスを推定する前にまず、比較的単純なモデルを用いて
124 統計モデリングの基本的な手順を解説したい。親子解析で得られたデータは、花粉散布な
125 どのいくつかの生態学的プロセスが作用した結果である。生態学における統計モデリング
126 の目的は、生態学的プロセス同士やそれらとデータとの関係を様々な仮定のもとで“単純
127 化して”、観測したデータからプロセスを推定することである。花粉散布研究においては、
128 これまで Neighborhood モデル (Burczyk et al. 1996; Burczyk and Prat 1997; Burczyk
129 et al. 2002) と呼ばれるモデルが多く使われてきた。本節ではこの Neighborhood モデル
130 を親子解析データにあわせて単純化したモデルを紹介する。

131 風媒性の植物の場合、花粉親からの距離が離れるにしたがって概ね花粉の密度が減少す
132 る (Levin, Kerster 1974)。したがって、距離にともなう花粉密度の減少関数として花粉
133 散布パターンをモデル化することが、もっとも単純であろう。この関数を花粉散布カーネ
134 ル (pollen dispersal kernel) と呼ぶ。関数の形は任意であるが、よく使われるのは以下の
135 指数べき乗分布 (Austerlitz et al. 2004) である。その他の花粉散布カーネルについては
136 Austerlitz et al. (2004) に詳しく解説されているので、そちらも参照されたい。

$$f(x|a, b) = \frac{b}{2\pi a^2 \Gamma(2/b)} \exp \left[- \left(\frac{x}{a} \right)^b \right] \quad (1)$$

137 ここで、 x は花粉親からの距離、 a, b は散布カーネルの形を決めるパラメータである。

138 個体間距離だけでなく、それぞれの花粉親がどれだけ花粉を生産したかも、観測した親
139 子解析データに影響すると考えられる。ここでは個体の花粉生産量 $Q(\cdot|\cdot)$ が以下の式の
140 ように個体サイズ (胸高直径) d で決まると仮定する。

$$\log Q(d|\alpha) = \alpha \log d \quad (2)$$

141 ある花粉親から x だけ離れた地点における花粉の密度 λ は、式 (1) と式 (2) の積で表す
142 ことができる (式 3)。式 (3) は花粉散布という生態的なプロセスを単純化して模したもの
143 であるので、ここではこれをプロセスモデルと呼ぶことにする。

$$\lambda = f(x|a, b) \cdot Q(d|\alpha) \quad (3)$$

144 プロセスモデルのパラメータを推定することが、統計モデリングの最初の目的である。
145 しかし、データからパラメータを推定するためには、プロセスモデルとデータを結びつけ

146 必要がある．これをプロセスモデルに対してデータモデルと呼ぶことにする．

147 本稿では花粉散布パターンを推定するためのデータとして，親子解析の結果を用いてい
148 る．親子解析データが持っている情報は，特定の母親に対してそれぞれの花粉親が受粉さ
149 せることができた割合である．母樹 i から解析した N_i 個の種子のうちそれぞれの花粉親
150 $1, 2, \dots, j, \dots, J$ が受粉させた個数を $n_{i1}, n_{i2}, \dots, n_{ij}, \dots, n_{iJ}$ として，それぞれの割合の期待
151 値を $\phi_{i1}, \phi_{i2}, \dots, \phi_{ij}, \dots, \phi_{iJ}$ とすると，このデータモデルは多項分布を用いて以下の式 (4)
152 のように表せる．

$$P(n_{i1}, n_{i2}, \dots, n_{ij}, \dots, n_{iJ}, N_i | \phi_{i1}, \phi_{i2}, \dots, \phi_{ij}, \dots, \phi_{iJ}) = \frac{N_i!}{\prod_{j=1}^J n_{ij}!} \prod_{j=1}^J (\phi_{ij})^{n_{ij}} \quad (4)$$

153 ここで ϕ_{ij} は相対的な受粉貢献度であるので，

$$\begin{aligned} \phi_{ij} &= \frac{\lambda_{ij}}{\sum_{j=1}^J \lambda_{ij}} \\ &= \frac{f(x_{ij}|a, b) \cdot Q(d_j|\alpha)}{\sum_{j=1}^J f(x_{ij}|a, b) \cdot Q(d_j|\alpha)} \end{aligned} \quad (5)$$

154 となる．

155 すべての母親について式 (4) の積をとったものがモデルの尤度である (式 (6))．最尤推
156 定ではほとんどの場合，尤度の代わりに対数尤度 (式 (7)) を最大化するようなパラメータ
157 の組み合わせ a, b, α を探索する．

$$L(a, b, \alpha) = \prod_{i=1}^I P(n_{i1}, \dots, n_{iJ}, N_i | \phi_{i1}, \dots, \phi_{iJ}) \quad (6)$$

$$\log L(a, b, \alpha) = \sum_{i=1}^I \log P(n_{i1}, \dots, n_{iJ}, N_i | \phi_{i1}, \dots, \phi_{iJ}) \quad (7)$$

158 ここではまず，Nelder-Mead 法を用いて推定した結果を見てみよう．

159 最尤推定モデルの検証

160 最尤推定の結果，それぞれのパラメータは $\hat{a} = 5.894 \pm 2.368$, $\hat{b} = 0.5636 \pm 0.0664$, $\hat{\alpha} =$
161 2.716 ± 0.140 (推定値 \pm SE) と推定された．この結果をもとに描いた花粉散布カーネル

162 と個体サイズと花粉生産量の関係が図 3 である．このモデルの仮定があっているならば，
163 ここで推定した花粉散布カーネルは“真の花粉散布パターン”を表しているはずであるが，
164 果たしてこの推定結果は正しいのだろうか？そこでこれらの推定結果がどれだけデータに
165 合っているのか，ブートストラップを用いた簡単なテストを試してみることにしよう．

166 まず，式 (5) に推定したパラメータ $\hat{a}, \hat{b}, \hat{\alpha}$ を代入して，母親 i に対する各花粉親の相対
167 的な受粉貢献度 $\hat{\phi}$ を算出する．つぎに，母親 i について確率 $\hat{\phi}_i$ にもとづいてランダムに
168 N_i 個の花粉親をサンプリングする．これをすべての母親について 10,000 回繰り返し，図
169 2 のように距離・方位ごとの頻度分布を算出する．この 95% 信頼区間と実際のデータを
170 比較することで，モデルがどれだけデータに合っているか確かめられるわけである．

171 図 4 をみると，距離・方位ともに有意に 95% 信頼区間から外れている箇所は少ないも
172 のの，信頼区間の形と実際のデータの分布が合っておらず，次節で行った階層ベイズによ
173 る推定結果 (図 7) に比べて信頼区間の幅が広すぎるように見える．これらのテストに明
174 確な規準はないが，少なくとも上記のモデルは実際の花粉散布パターンをよく表している
175 とは言えないだろう．

176 この理由として考えられるのが，モデルから予測した相対的な受粉貢献度 $\hat{\phi}$ のばらつき
177 が過小評価され，実際よりもランダム交配に近い状態でブートストラップを行ったのでは
178 ないかという点である．このモデルではそれぞれの花粉親が生産する花粉の量は式 (2) の
179 ように個体サイズのみで決まると仮定した．しかし，一般的に個体の花粉生産量や繁殖成
180 功度は個体サイズと有意な相関を示すものの，そのばらつきは極めて大きいことが指摘さ
181 れている (Klein et al. 2008)．この点を考慮しなかったことが，モデルの当てはまりを悪
182 くした第一の原因と考えられる．

183 また，このモデルでは花粉散布パターンは式 (1) のように距離だけで決まり，すべての
184 方角に同じ密度で花粉が飛ぶと仮定しているが，実際の花粉散布パターンは開花期の風
185 向・風速に影響されると予想される．花粉散布パターンと風向・風速の関係を検討した研
186 究はいくつかあるが，それらはいずれも「特定の方角への散布量が多く，そこから離れる
187 につれて飛散量が減少する (Burczyk et al. 1996; Burczyk and Prat 1997)」，「花粉は個
188 体を中心に楕円状に散布される (Austerlitz et al. 2007)」といった単純なパターンしか
189 仮定してない．しかし，実際の風媒性の樹木の花粉散布は開花期間中の風向・風速に大き
190 く影響され，複雑な空間パターンをとる (Nikkanen et al. 2002)．したがって，花粉散布パ
191 ターンをモデル化するには，距離や方角だけでは説明できない複雑な空間パターンを同
192 時に考慮したほうがより実際の花粉散布パターンをうまく表現できるだろう．

193 このような複雑なパターンを推定するためには，多くのパラメータを同時に推定する必
194 要がある．階層ベイズはパラメータが多く，しかもそれらが独立でないといった複雑なモ

195 デルの推定において威力を発揮する．以下ではどのようにして階層ベイズモデルが複雑な
196 プロセスを扱うのか，具体的に見てみよう．

197 階層ベイズによる花粉モデルの拡張

198 階層ベイズモデルに関する基本的な解説は深澤・角谷 (この特集号) などに譲るが，最
199 尤法などの非ベイズ統計学との違いは，パラメータを一つの値ではなく確率分布と考える
200 点である．すると，深澤・角谷 (この特集号) の式 (1) のようにデータとパラメータの間
201 には，

$$P(\theta|\text{Data}) \propto P(\text{Data}|\theta)P(\theta) \quad (8)$$

202 という関係が成り立つ．

203 ここで $P(\text{Data}|\theta)$ はモデルの尤度である．ベイズ統計学では $P(\theta)$ をパラメータの事
204 前分布， $P(\theta|\text{Data})$ をパラメータの事後分布と呼び，それぞれパラメータに関する“事前
205 知識”とデータを得た後のパラメータがとりうる確率分布を意味している．最尤法では尤
206 度を最大化するパラメータの値を推定したが，ベイズ統計学では事後分布の形が推定の対
207 象となる．

208 事前分布を設定する利点としてまず挙げられるのが，先行研究によって特定のパラメー
209 タがとりうる値のおおよその範囲が分かっている場合に，これまでの知見を推定に反映さ
210 せることができる点である．逆にパラメータのとりうる範囲に関する情報がない場合は，
211 一様分布や分散の大きい正規分布のような“無情報”な分布を設定することで，ほぼデー
212 タのみを使ってパラメータを推定することもできる．

213 事前分布に反映させることができる情報はパラメータの値がとりうる範囲だけではな
214 い．例えば，この節で扱う例題のように地点ごとに推定すべきパラメータがある場合，各
215 地点のパラメータの値は独立に決まるのではなく，近い地点ほど似通った値をとることが
216 予想される．このような場合は，近い地点ほど似通った値をとるように事前分布を設定す
217 ることで，それぞれの地点のパラメータがバラバラな値をとらないように制約をかけるこ
218 とができる (深澤ら この特集号) ．

219 このように事前分布を様々な形に設定することによってパラメータの挙動さえもモデル
220 化することができる点が，階層ベイズの特徴である (深澤・角谷 この特集号) ．以下では，
221 パラメータモデルを導入することによって前述の最尤推定にもとづく単純なモデルを拡張
222 し，より複雑な花粉散布プロセスをモデル化することを試みる．

223 距離だけでは説明できない複雑な花粉散布パターンをモデル化するためにここでは，式

224 (1) の花粉散布カーネルを“変形”させることを試みた．具体的には，まず平面上の花粉
 225 散布カーネルを花粉親を中心に $5 \times 5\text{m}$ のメッシュに分割し，それぞれのマスごとに式 (1)
 226 で決まる花粉密度からずれる大きさのパラメータ ρ を設定した．そして，花粉親を原点と
 227 したときにある地点 \mathbf{z} における花粉密度 $f'(\mathbf{z}|\dots)$ は，式 (1) と ρ の積であると定義した．

$$f'(\mathbf{z}|a, b, \rho) = f(\|\mathbf{z}\||a, b) \cdot \underline{\rho_{m_{\mathbf{z}}}} \quad (9)$$

228 ただし， $m_{\mathbf{z}}$ は地点 \mathbf{z} が位置するマスの ID である．

229 ρ は空間構造を持ったパラメータであるので，それぞれの地点におけるパラメータの値
 230 は独立ではなく，近いもの同士が似通った値をとるような空間自己相関を持っていると考
 231 えられる．ここでは Intrinsic CAR モデル (深澤ほか この特集号) によって， ρ をモデル
 232 化した． ρ_m 以外のすべてのマス ρ_{-m} が与えられたとき， $\log \rho_m$ が以下のように周囲の
 233 マス δ_m の対数をとったものの平均を期待値に持つ正規分布にしたがうと仮定した．

$$\log \rho_m | \rho_{-m} \sim \text{Normal} \left(\frac{\sum_{l \in \delta_m} w_{ml} \log \rho_l}{w_{m+}}, \frac{\tau_\rho}{w_{m+}} \right) \quad (10)$$

234 ここで， δ_m はマス m の周辺のマスの ID， w_{ml} はマス m 近傍のマス l に対する重みづけ
 235 (ここではすべて 1 とした)， w_{m+} はマス m の近傍のマスの重みの合計である．

236 散布カーネルのパラメータのうち， a については $1/a \sim \text{Gamma}(0.1, 0.01)$ という無情
 237 報に近い事前分布を設定した*1．一方， b についても a と同様に無情報に近い事前分布を
 238 設定して推定を試みたが，うまくパラメータを推定できなかった．この理由としては， b
 239 が距離 d のべき指数であるために (式 (1))，わずかな値の変化でも花粉密度に関係する他
 240 のパラメータへの影響が大きいためであると考えられる．無情報事前分布ではパラメー
 241 タがとりうる範囲が広すぎるのである．

242 このような場合は，パラメータに関する事前知識を活用して恣意的に事前分布を設定す
 243 ることで，推定がうまくいく場合がある．指数べき乗分布の形状パラメータ b はほとんどの
 244 風媒性の樹木では 1 よりも小さい値をとることがこれまでの研究から知られている．そ
 245 こでここでは， $b \sim \text{Gamma}(30, 60)$ (平均 $0.5 \cdot$ 分散 8.3×10^{-3}) という事前分布を設定
 246 した．

247 また，個体の花粉生産量についてもサイズだけでは説明できないばらつきの効果を考慮
 248 する必要があるだろう．ここでは以下のように，式 (2) にサイズだけでは決まらない効果
 249 の大きさを表すパラメータ β を加えることで，花粉生産量 $Q'(\cdot|\cdot)$ をモデル化した．

*1 これは a に相当する正規分布の分散の逆数の共役事前分布がガンマ分布であることから，それと同様にモ
 デル化したためである．計算上特に便利であるというわけではない．

$$\log Q'(d_j|\alpha, \beta_j) = \alpha \log d_j + \underline{\beta_j} \quad (11)$$

250 β はそれぞれが独立したパラメータではなく，全体があるばらつきを持った正規分布に
251 したがうランダム効果として，モデル化した．

$$\beta \sim \text{Normal}(0, \tau_\beta) \quad (12)$$

252 なお，式 (10),(12) の超パラメータである τ_ρ, τ_β はどちらも無情報に近い事前分布
253 $1/\tau \sim \text{Gamma}(0.1, 0.01)$ を設定した．

254 これらのパラメータの事後分布は WinBUGS 1.4.2 を用いて，Malkov chain Monte
255 Carlo (MCMC) 法によって推定した．パラメータの収束には時間を要し，150,000
256 MCMC steps \times 3 chains の計算におよそ3日かかった．

257 階層ベイズモデルの推定結果と結果の検証

258 各パラメータの事後分布の平均はそれぞれ， $\hat{a} = 2.335 \pm 2.346$ (0.120, 8.926), $\hat{b} =$
259 0.433 ± 0.075 (0.307, 0.595), $\hat{\alpha} = 2.907 \pm 0.314$ (2.291, 3.511), $\hat{\tau}_\rho = 4.736 \pm$
260 1.064 (3.007, 7.288), $\hat{\tau}_\beta = 1.104 \pm 0.216$ (0.651, 1.503) であった (事後平均 \pm SD
261 (95% 確信区間))．さきほどのモデルと共通のパラメータ a, b, α のうち a はやや小さい値
262 となったが，その他の2つのパラメータはどちらのモデルでもほぼ同じ値となった．

263 パラメータの事後平均から花粉散布カーネルと個体の花粉生産量とサイズの関係に加えて，
264 それらにマス・個体ごとのランダム効果を加えた値をプロットしたものが，図5であ
265 る．最尤推定の結果をもとに描いた図3と比べると，より近い個体からの散布の割合が階
266 層ベイズで多くなったように見える．また，花粉散布カーネル・花粉生産量ともに距離・
267 サイズだけで説明できない部分の影響が大きく，特に固定効果に対してランダム効果を考
268 慮したときの花粉生産量のばらつきが顕著であった．ここから個体の花粉生産量はサイズ
269 だけでは決まらず，ランダム効果としてここでは推定されたその他の要因が大きく影響し
270 ていると思われる．

271 つぎに花粉散布カーネルを空間的に見てみよう．図6a,bは式(9)で定義した花粉散布
272 カーネルのうち，それぞれ距離で決まる固定効果とそれ以外のランダム効果の大きさを空
273 間上にプロットしたものであり，両者をあわせたものが図6cである．他の図と比較でき
274 るようにするためにこれらの図は，本来の花粉散布パターンの上下・左右を反転させて，
275 母樹を中心に見たときにどこから・どれだけの花粉が飛んでくるかを表すようにしてあ
276 る．図6cを見ると，図6aのようにすべての方角から同じように花粉が飛んでくるわけ

277 はなく，北東と南西からより多くの花粉が飛んできているように見える．開花期間中に風
278 向・風速の観測を行った結果，北東と南西からの吹く風の頻度が圧倒的に多かったことか
279 ら（花岡ら，未発表データ），ブナの花粉散布パターンが開花期の風向・風速によって影響
280 を受けることが分かった．

281 これらの結果がどれだけデータと合っているかを検証するために，最尤推定と同様に
282 パラメータの事後平均から距離・方位ごとの花粉散布パターンを 10,000 回のシミュレー
283 ションによって推定し，実際のデータと比較した（図 7）．図 7 を見ると，距離・方位とも
284 に 95% 信頼区間から外れる階級が少しあるものの最尤推定の結果（図 4）よりも信頼区間
285 の形がデータに近いので，それなりにうまく推定できていると言ってよいだろう．

286 ただし，これらの推定結果を解釈する際にはデータやモデルの構造に起因するバイアス
287 が存在するかどうかを同時に検討する必要がある．この例では，推定された花粉散布の空
288 間パターン（図 6c）の母樹から見て南西（SW）方向 50m 付近に，周囲に比べて密度が高い
289 点がある．これは，その地点に受粉貢献度が高い個体が 1 個体存在するために，その地点
290 から飛んでくる花粉の量が多いと推定されたためと考えられる．これは全体の推定結果に
291 大きな偏りをもたらすものではないが，バイアスを含んだうえでの結果であることに留意
292 して考察を行うべきであろう．

293 さらにこの例では，推定された花粉散布の空間パターン（図 6c）と個体の空間分布（図
294 1）の形が似ており，推定結果に何らかの影響を及ぼす可能性も考えられる．これについ
295 て筆者らは，空間的に偏りがないと考えられる母樹に近い個体のデータのみを用いて同様
296 の解析を行い，推定結果に大きな偏りが無いことを確認した（Hanaoka et al. 投稿中）．
297 この例のように，空間パターンやランダムな個体差など，独立でない多くの変数を扱うよ
298 うなモデルにおいては，個々のパラメータにバイアスが生じることがよくある．したがっ
299 て，複雑なモデルを扱う際には推定結果の妥当性を注意深く検討する必要がある．

300 まとめ

301 ここまで，ブナの花粉散布パターン推定を例に，統計モデリングの基本的な考え方と階
302 層ベイズを用いて複雑なプロセスを推定するまでの流れを解説してきた．その結果，花粉
303 親からの距離によって花粉密度が減少するといった単純なモデルでは表現できなかった複
304 雑な花粉散布の空間パターンを推定することができた．これはパラメータモデルによる柔
305 軟なモデリングが可能な階層ベイズの大きな利点である．

306 柔軟なモデリングが可能であるということは一方で，研究者自身が生態学的プロセス
307 に関するモデルをかなり“主観的”に構築しなければならないことを意味している．一般

308 化線形モデル (GLM) などの比較的単純なモデルでは，そのモデルが表現しようとして
309 いる生態的プロセスや因果関係を意識することはあまり無く，AIC によるモデル選択や
310 Akaike weight による model averaging によって説明変数を絞ったうえで，個別の要因の
311 効果を検討することが多いと思う．しかし，階層ベイズには確立されたモデル選択手法が
312 なく，また仮にモデル選択ができたとしても，様々な要因の組み合わせの中から個々のプ
313 ロセスの効果を一つ一つ検討していく作業には莫大な時間と労力を要する．したがって，
314 階層ベイズによってモデルを構築する際には，個々のプロセスを表現するモデルやプロセ
315 ス同士の因果関係が生態学的に妥当であるか，注意深く検討する必要があるだろう．

316 また独立でないパラメータを扱う複雑なモデルでは，パラメータ推定が困難になる場合
317 がある．実際の計算においては，異なる MCMC 間のパラメータの挙動を比較したり，パ
318 ラメータの発散によるエラーに直面しながら少しずつ事前分布やモデルの構造を改良して
319 いくことがよくある．本稿のように計算に時間を要するモデルの場合，試行錯誤によるモ
320 デルの改良に数週間から数ヶ月を要することがある．したがって，大きなデータセットを用
321 いて複雑なモデルのパラメータを推定しようとする場合は，モデルの生態学的な妥当性を
322 検討するだけでなく，パラメータの推定プロセスにも留意してモデルを構築する必要があ
323 るだろう．

324 本稿の例では MCMC 計算に WinBUGS を用いた．WinBUGS は研究者がベイズ統計
325 学や MCMC の原理を理解していなくても直感的にモデルを記述できるという点で便利な
326 ソフトウェアであるが，前述のような試行錯誤によってモデルを改良するためには，ベ
327 ィズ統計学や MCMC を研究者自身がある程度理解していることが望ましいと筆者は考えて
328 いる．本稿ではこれらの基本的な事項について解説しなかったが，和書では豊田 (2008);
329 古谷 (2008)，洋書では Clark and Gelfand (2006); McCarthy (2007) あたりが参考にな
330 るだろう．

331 本稿で紹介したブナの花粉散布空間パターンを推定した例は階層ベイズの一応用例に過
332 ぎず，生態学における階層ベイズの応用例やその威力を十分に解説できたとは言い難い．
333 しかし，WinBUGS などのソフトウェアや階層ベイズ・MCMC に関する解説書などが充
334 実している現在，生態学者が階層ベイズに取り組む敷居はかなり低くなっていると感じ
335 る．本稿によって読者が階層ベイズに興味を持ったり，最初の取り掛かりの助けとなれば
336 幸いである．

337 謝辞

338 本論文で紹介した研究に対しては，向井譲氏，後藤晋氏，島谷健一郎氏をはじめ，多く
339 の方から有益な助言をいただいた．また原稿の執筆に際しては，久保拓弥氏，深澤圭太氏，
340 および査読者の方々より貴重な意見をいただいた．以上の方々に厚く御礼申し上げます．

341 参考文献

342 Austerlitz F, Dick CW, Dutech C, Klein EK, Oddou-Muratorio S, Smouse PE, Sork,
343 VL (2004) Using genetic markers to estimate the pollen dispersal curve. *Mol Ecol*
344 13:937-954

345 Austerlitz F, Dutech C, Smouse PE, Davis F, and Sork VL (2007) Estimating
346 anisotropic pollen dispersal: a case study in *Quercus lobata*. *Heredity* 99:193-204

347 Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating
348 revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the
349 neighbourhood model. *Molecular Ecology* 11:2379-2391

350 Burczyk J, Adams WT, Shimizu JY (1996) Mating patterns and pollen dispersal in
351 a natural knobcone pine (*Pinus attenuata* Lemmon.) stand. *Heredity* 77:251-260

352 Burczyk J, Prat D (1997) Male reproductive success in *Pseudotsuga menziesii* (Mirb.)
353 Franco: the effects of spacial structure and flowering characteristics. *Heredity*
354 79:638-647

355 Clark JS, Gelfand AE (2006) Hierarchical Modelling for the Environmental Sciences:
356 Statistical Methods and Applications. New York, Oxford University Press

357 古谷知之 (2008) ベイズ統計データ分析. 朝倉書店, 東京

358 Gerber S, Chabrier P, Kremer A (2003) FaMoz: a software for parentage analysis
359 using dominant, codominant and uniparentally inherited markers. *Mol Ecol Notes*
360 3:479

361 Klein EK, Desassis N, Oddou-Muratorio S (2008) Pollen flow in the wildservice tree,
362 *Sorbus torminalis* (L.) Crantz. IV. Whole interindividual variance of male fecundity
363 estimated jointly with the dispersal kernel. *Mol Ecol* 17:3323-3336

364 久保拓弥・粕谷英一 (2006) 「個体差」の統計モデリング. *日本生態学会誌* 56, pp181-190

365 Levin DA, Kerster HW (1974) Gene flow in seed plants. *Evol Biol* 7:139-220

- 366 McCarthy MA (2007) Bayesian Methods for Ecology. New York, Cambridge Univer-
367 sity Press
- 368 Nikkanen T, Pakkanen A, Heinonen J (2002) Temporal and spatial variation in air-
369 borne pollen and quality of the seed crop in a Norway spruce seed orchard. *Forest*
370 *Genet* 9:243-255
- 371 田中健太・北本尚子・島谷健一郎・後藤晋 (2006) 植物の花粉散布研究における遺伝マー
372 カの役割. *日本生態学会誌* 56, pp134-144
- 373 豊田秀樹 (編) (2008) マルコフ連鎖モンテカルロ法. 朝倉書店, 東京

図の説明

図 1 調査地の成木個体位置および親子解析によって推定された母親-花粉親ペアの分布。花粉親のシンボルの大きさが受粉貢献度を表す。

図 2 (a) 距離および (b) 方位別にみた花粉親の頻度分布。親子解析によって推定された母親-花粉親ペアの数を、個体間距離および母親から見た花粉親の方位ごとに集計し、それぞれのクラスに含まれるすべてのペア数で割った値をプロットした。

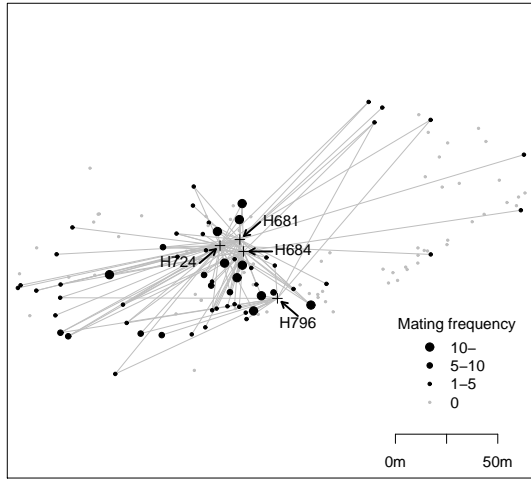
図 3 最尤推定によって推定された (a) 花粉散布カーネルと (b) 個体の花粉生産量とサイズの関係。

図 4 最尤推定値をもとに予測した花粉散布パターンを、図 2 に重ね書きしたもの。実線はそれぞれのクラスの 95% 区間を表している。

図 5 階層ベイズモデルによって推定された (a) 花粉散布カーネルと (b) 個体の花粉生産量とサイズの関係。実線は固定パラメータの事後平均のみで推定された値 (図 3 に相当) で、白丸がランダム効果を含んだマス・個体ごとの値を表す。

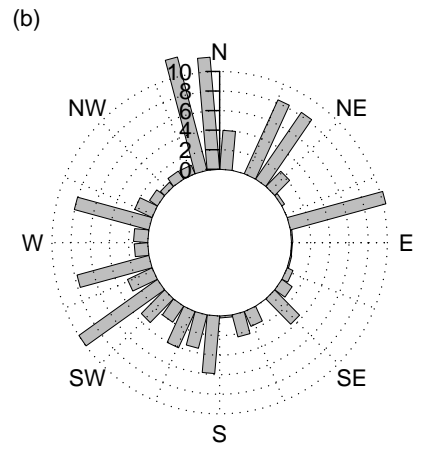
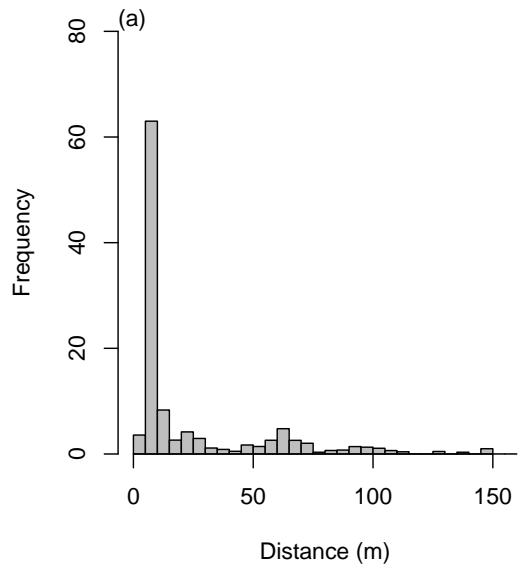
図 6 階層ベイズモデルによって推定した花粉散布の空間パターン。(a) 固定効果 a, b のみ、(b) 空間ランダム効果 ρ のみ、(c)(a),(b) を組み合わせたもの。その他の図と対応させるためにこれらの図は、本来の花粉散布パターンの上下・左右を反転させ、母親を中心にしたときにどこから・どれだけの花粉が飛んでくるかを表すようにした。

図 7 階層ベイズモデルの事後平均をもとに予測した花粉散布パターンを、図 2 に重ね書きしたもの。実線はそれぞれのクラスの 95% 区間を表している。



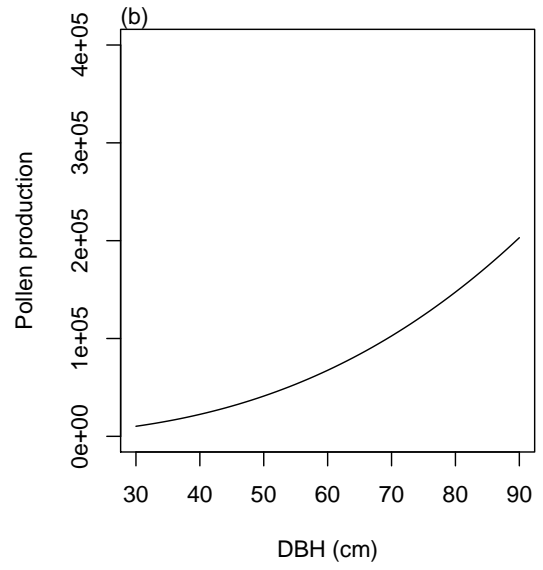
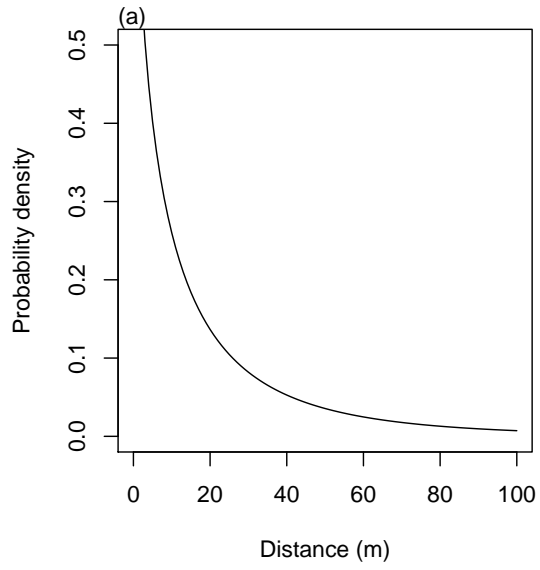
375

376 富田・花岡 (図 1)



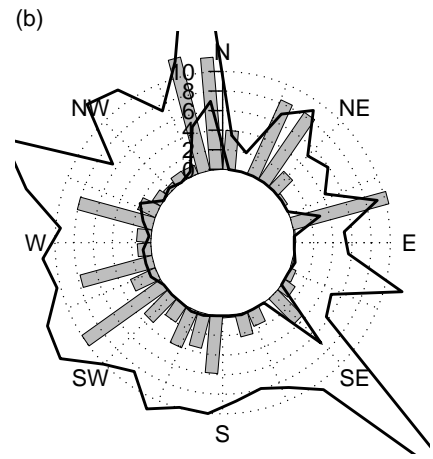
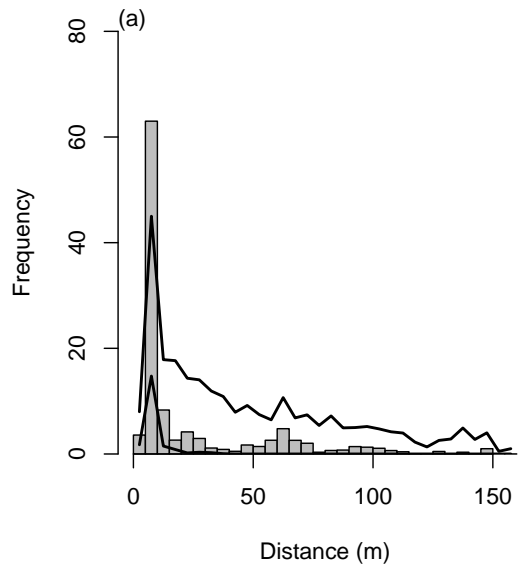
377

378 富田・花岡 (図 2)



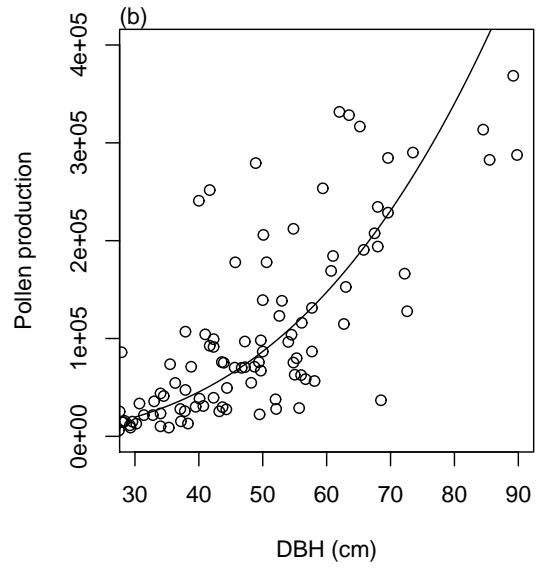
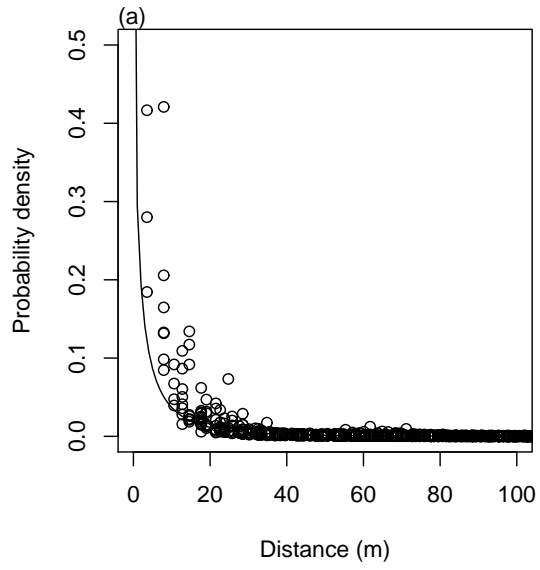
379

380 富田・花岡 (図 3)



381

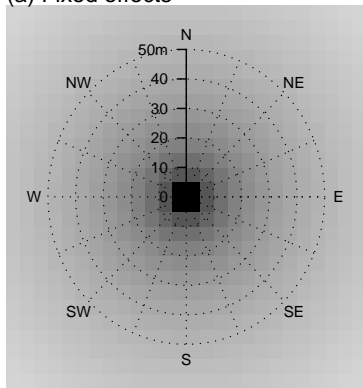
382 富田・花岡 (図 4)



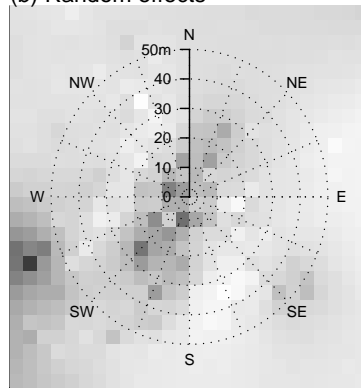
383

384 富田・花岡 (図 5)

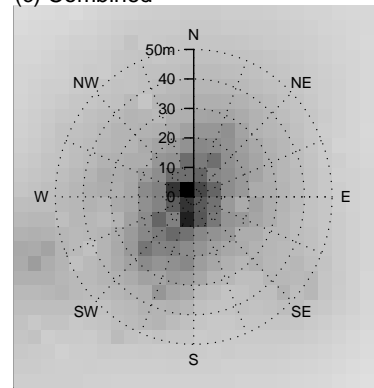
(a) Fixed effects



(b) Random effects

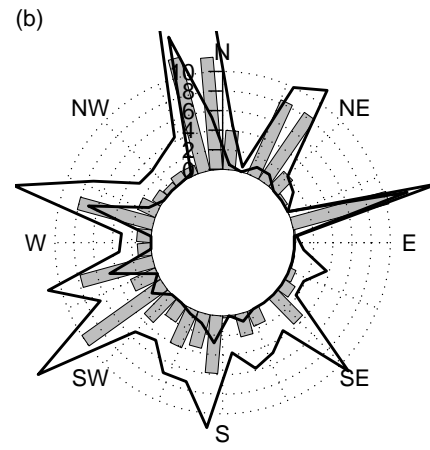
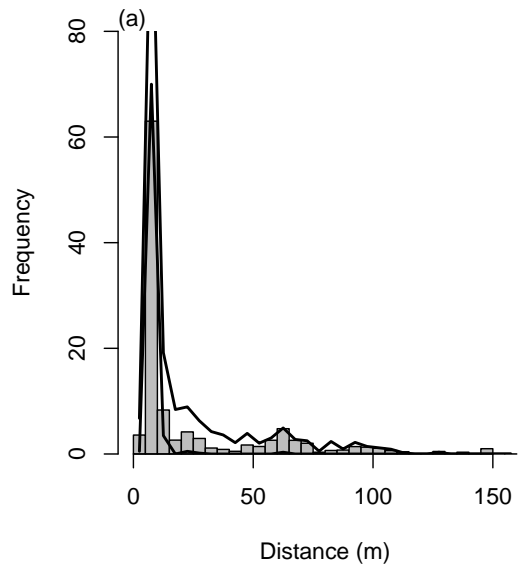


(c) Combined



385

386 富田・花岡 (図 6)



387

388 富田・花岡 (図 7)