

# 動的サイト占有モデル

## —状態の不確実性を考慮したサイト占有動態の統計的推測

深谷 肇<sup>†</sup>

(受付 2015 年 7 月 23 日; 改訂 2016 年 1 月 4 日; 採択 3 月 1 日)

### 要 旨

生態学研究で行われる種々のサイト占有状態の調査では分類誤差が発生することがあり、サイト占有状態の観測は不確実性を伴う。このような占有状態の不確実性を無視することは、一般的に、関心のあるサイト占有状態そのものとサイト占有動態を駆動する生態学的過程の推測にバイアスをもたらすことになるため、分類誤差はデータの収集とその解析の過程において適切に考慮されることが重要である。本稿では、分類誤差を考慮してサイト占有動態の統計的推測を行う動的サイト占有モデルと、それを一般化した枠組みである多状態動的サイト占有モデルについて、その動機や適合する調査デザイン、およびモデルの定式化についてその概要を示す。また様々な生態学的問題に適用できる多状態動的サイト占有モデルの 1 つの例として、固着性生物の観察で生じる特定の観測誤差を考慮した群集動態モデルを紹介し、固着性生物群集動態の推測において動的サイト占有モデルを用いることの利点を述べる。

キーワード：階層モデル、隠れマルコフモデル、生態学、不完全な発見、Pollock のロバストデザイン。

### 1. はじめに

生態学では、ある生息地における種の分布や繁殖状況を明らかにするために、複数の調査地(サイト)を設けてその占有状態(個体の在不在や繁殖の有無など)の調査が行われることがある。このような調査の対象となる生物は、哺乳類や鳥類、昆虫などの無脊椎動物から植物まで様々であり、占有状態の調査方法もまた、調査員による目視やカメラトラップなどの自動記録装置、糞や足あとなどの形跡の確認まで多様である。同じように、サイトの定義も研究ごとに異なり、数キロメートル四方の地理的区画から局所的な微小生息地まで、様々なものがある。

野外における観測はしばしば不完全であり、このようなサイト占有状態の観測データには観測誤差が含まれることが多い。例えば、占有されているサイトにおいてその種の個体が発見されない、繁殖が行われているサイトにおいて繁殖を示す証拠が見つからない、といった観測結果が得られることがある。こうした観測誤差はサイト占有状態を示す証拠の検出に関する誤り、あるいは占有状態の分類に関する誤りであるので、検出誤差や分類誤差と呼ばれる。このような観測誤差はサイトの生態学的状態に不確実性をもたらす、背後にある生態学的な規則や過程に関する推測の精度や偏りに影響する。そのためサイト占有調査における観測誤差は、もとより野外調査においてその頻度が最小化されることが望ましいものの、推測の過程においてもま

<sup>†</sup> 統計数理研究所：〒 190-8562 東京都立川市緑町 10-3

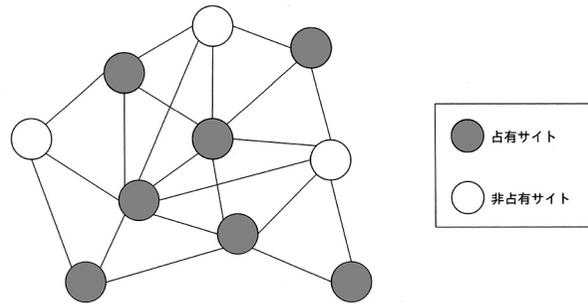


図1. メタ個体群の概念図。メタ個体群は、個体の分散(移出入)によって連結した局所個体群(同種個体の局所集団)の集合である。局所個体群が成立する空間をサイトやパッチと呼ぶ。サイトの占有状態は局所的な「絶滅」と他のサイトからの「住み着き(個体の移入)」によって変化する。これらの過程により駆動されるメタ個体群全体のダイナミクスがメタ個体群動態である(Hanski, 1998)。この図では明示していないが、実際のメタ個体群では、個々のサイトの面積や空間配置などの要因がその動態に本質的な影響を及ぼす。

た適切に考慮される必要がある(このような観測誤差の生じる理由が、観測者の技量不足にあるとは限らない。根本にあるのは、調査時の気象やサイトの立地条件、調査努力量によっては完全な野外観測を達成することが難しいという事実である)。

生態学においては、このような分類誤差を考慮してサイト占有状態の推測を行うための統計モデルがサイト占有モデルと呼ばれており、2000年代以降様々なモデルの開発と体系化が進んできている(MacKenzie et al., 2006; Royle and Dorazio, 2008; Kéry and Schaub, 2012; Bailey et al., 2014; Kéry and Royle, 2016)。本稿では、サイト占有モデルの中でも特に、占有状態の観測時系列に適合した統計モデルである動的サイト占有モデルとその拡張について概説する。これらのモデルは、いずれも隠れマルコフモデルと呼ばれる時系列モデルの特殊な例と見ることができ、様々な生態学的問題に適用可能な、有用な枠組みである。これ以降の本稿の構成は以下のとおりである。2節では、個体の発見・非発見データの時系列からサイトの占有状態(占有・非占有)とその動態を推測する動的サイト占有モデルについて説明する。3節では動的サイト占有モデルの拡張である多状態動的サイト占有モデルを扱い、サイト占有状態と観測のカテゴリの数が2より多い場合のモデリングについて説明する。4節では、多状態動的サイト占有モデルの1つの応用として観測誤差を考慮した固着性生物群集動態の推測に関して説明し、最後に5節でまとめを行う。

## 2. 動的サイト占有モデル

### 2.1 動機・調査デザイン・データ

動的サイト占有モデルは、野外におけるメタ個体群動態(Hanski, 1998; 図1)あるいは複数サイトにおける占有動態を推測するための統計モデルである。特定の種のメタ個体群動態を理解する上で重要な変数として、サイト占有率や局所住み着き確率、局所絶滅確率などが挙げられる。これを明らかにするためには、複数のサイト(生物学的に自然に定義される場合と、操作上の単位として任意に設定される場合がある; MacKenzie et al. 2006を参照されたい)において、対象種によるサイトの占有を継続的に調査する必要がある。ところで、すでに述べたように、野外では生息個体を全て確実に観察できるとは限らないのが野生生物の生態調査の難しいところである。サイト占有調査の文脈で言えば、実際にはサイトが占有されていても、特定の調査で

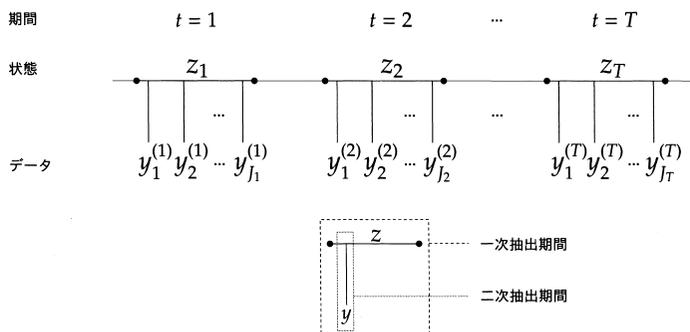


図 2. Pollock のロバストデザイン. 時間について入れ子状の調査デザインとなっており, サイトの占有状態 ( $z_t$ ) が変化しない期間 (一時抽出期間, または季節と呼ばれる) の間に独立な調査が反復される. 住み着きや絶滅のためにサイトの占有状態は時間変化するが, そのような変化は異なる季節の間で生じると仮定される. 簡単のため, サイトの添字 ( $i$ ) を省略した.

はたまたま対象種の個体が発見されず, 種のサイト占有が確認されない場合が起こりうる. このような場合, 観測される不在データには, サイトが占有されていなかったために生じたものと, 占有されていたが個体が発見されなかったために生じたものの 2 種類が混ざっていることになる. 後者の不在データは「偽物」であるため, このような検出誤差は偽陰性誤差 (false negative error) と呼ばれる. 偽陰性誤差が存在する場合には, 当然見かけ上のサイト占有率は実際のそれよりも低くなり, また, そのことを考慮しないメタ個体群動態のパラメータ推定はときに大きなバイアスを生じることが知られている (Moilanen, 2002). つまり野外調査の観測過程が不完全であるために, 関心のある生態学的過程 (メタ個体群動態) の推測が容易ではないのである.

このような偽陰性誤差の問題を考慮して, メタ個体群動態に関するパラメータの適切な推測を目指した統計モデルが MacKenzie et al. (2003) による動的サイト占有モデルである. このモデルは, 占有状態が一定とみなせる特定の短期間におけるメタ個体群のサイト占有率を, 偽陰性誤差を考慮して推定するサイト占有モデル (MacKenzie et al., 2002; Tyre et al., 2003 も参照) の一般化として提案された. サイト占有モデルとは異なり, 動的サイト占有モデルでは局所個体群の占有状態の時間変化がモデル化され, したがってパラメータとして局所住み着き確率と局所絶滅確率が推定される.

動的サイト占有モデルでは, メタ個体群デザインと Pollock のロバストデザインと呼ばれる, 2 つの調査デザインの組み合わせが前提となっている. メタ個体群デザインとは, メタ個体群 (図 1) をサンプリングすることの比喩であり, 研究対象の生息域に独立な調査サイトを複数設置して個体群の空間的な反復を確保し, 各サイトで複数回の調査を行うことで時間的な反復を得る調査デザインである (Kéry and Schaub, 2012). ただしここでの「メタ個体群」は操作的な概念であり, 実際にはサイトの集合が生態学的な実体としてのメタ個体群とはならない場合も起こりうることに注意されたい. 一方で Pollock のロバストデザイン (Pollock, 1982) は, 時間軸に沿って入れ子状に調査が反復される調査デザインである (図 2). ロバストデザインによる階層的なサンプリングは, データから個体の発見確率を推定するために非常に重要である (e.g., Dail and Madsen, 2013; Knappe and Korner-Nievergelt, 2015).

データは以下のように取得される (図 2). 調査サイトを合計  $I$  箇所設定し, 合計  $T$  期間に渡る各サイトの占有状態の変化に関心があるとする. ここで, 各期間  $t$  ( $t = 1, \dots, T$ ) の内でサイ

トの占有状態は一定であり、一方で隣り合う期間の間隔は占有状態の変化が生じるのに十分な時間スケールであるとする。各期間  $t$  の間に、それぞれのサイト  $i$  ( $i = 1, \dots, I$ ) で独立な占有状態調査が合計  $J_{it}$  回行われる(つまり、調査は占有状態の変化が生じない短い期間内に反復して実施される)。こうして、サイト  $i$  ( $i = 1, \dots, I$ )、期間  $t$  ( $t = 1, \dots, T$ )、調査  $j$  ( $j = 1, \dots, J_{it}$ ) のそれぞれにつき、個体の発見の有無を表す 2 値データ  $y_j^{(it)}$  が観測されることになる。個体の発見を  $y = 1$ 、非発見を  $y = 0$  と表す。動的サイト占有モデルの目標は、この発見データの配列から不完全な発見(偽陰性誤差)を考慮してサイト占有率や局所住み着き確率、局所絶滅確率などに関する推測を行うことである。

モデルの説明に進む前に、サイト占有モデルに関連する用語について少し整理しておく。上記の期間  $t$  と調査  $j$  はそれぞれ、一次抽出期間(primary sampling period)と二次抽出期間(secondary sampling period)と呼ばれることがある(MacKenzie et al., 2003; 図 2)。また、MacKenzie et al. (2006) は一次抽出期間を「季節(season)」と定義し、複数の季節を扱う動的サイト占有モデルを多季節サイト占有モデル、1 季節のみを扱うサイト占有モデルを 1 季節サイト占有モデルとそれぞれ呼んでいる。占有モデルでは、「季節」の長さが研究の対象や状況によって異なることに注意されたい。例えば、数週間ほどの短い期間に研究が行われる場合には、「季節」が繁殖期などの生態学的な意味を持つことがあるが、1 年おきの調査が行われている場合などはその限りではない(MacKenzie et al., 2006)。

## 2.2 モデルと推測

ここでは Royle and Kéry (2007) に倣って、基本的な動的サイト占有モデル(MacKenzie et al., 2003)の状態空間表現(階層表現とも呼ばれる)を与える。状態空間表現は、状態変数と呼ばれる直接的には観測されない潜在確率変数を用いて、

- 状態変数の初期分布に関するモデル
- 状態変数の時間発展を表すシステムモデル
- 状態変数に依存した観測データの確率分布を表す観測モデル

の組み合わせによって観測時系列の生成過程を表すものである(Royle and Dorazio, 2008; Kéry and Schaub, 2012; 深谷, 2016)。以下で説明されるように、動的サイト占有モデルの状態変数は各期間における各サイトの占有状態である。状態空間表現を用いた Royle and Kéry (2007) とは対照的に、MacKenzie et al. (2003) では状態変数を明示せずに動的サイト占有モデルが説明されている。Royle and Kéry (2007) と MacKenzie et al. (2003) はどちらも、基本的に同じ動的サイト占有モデルを説明したものであるが、生態学的過程と観測過程をはっきり区別して表す状態空間表現を用いた方が、モデルで仮定されるデータ生成過程を直感的に理解しやすいように思われる。システムのモデルと観測のモデルを区別して表すことは、それぞれの要素の柔軟な拡張を考える上でも有用である(Royle and Dorazio, 2008; Kéry and Schaub, 2012; Kéry and Royle, 2016)。

直接的には観測されない状態変数として、サイト  $i$  ( $i = 1, \dots, I$ )、期間  $t$  ( $t = 1, \dots, T$ ) におけるサイトの占有状態を導入し、これを  $z_{it}$  と表す。 $z = 1$  は種によるサイトの占有を、 $z = 0$  は非占有(種の不在)を表す。占有状態の時間変化は、種の局所住み着きと局所絶滅によって生じる。非占有サイト  $i$  において期間  $t$  から  $t+1$  の間に新たに住み着きが生じる確率  $\Pr(z_{i,t+1} = 1 | z_{it} = 0)$  を局所住み着き確率と定義し、 $\gamma_t$  と表す。また、占有サイトで局所絶滅が生じる確率  $\Pr(z_{i,t+1} = 0 | z_{it} = 1)$  を局所絶滅確率と定義し、これを  $\epsilon_t$  と表す。MacKenzie et al. (2003) は局所絶滅確率をモデルパラメータとしているが、以下では Royle and Kéry (2007) に倣い、局所絶滅確率の関数としてサイトの「生存確率」 $\Pr(z_{i,t+1} = 1 | z_{it} = 1) = 1 - \epsilon_t$  を定義

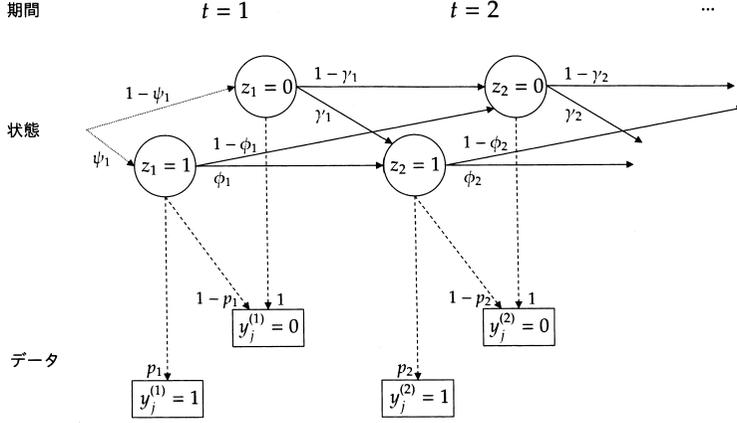


図 3. 2 状態の動的サイト占有モデル. 点線の矢印は初期分布モデル(式(2.1)), 実線の矢印はシステムモデル(式(2.2)), 破線の矢印は観測モデル(式(2.3))をそれぞれ表す. 矢印のそばの値は各過程の条件付き確率である. 簡単のため, サイトの添字 ( $i$ ) を省略した.

し, これを  $\phi_t$  とおいて説明する. 期間  $t$  にサイト  $i$  が占有されていた場合に, 1 回の調査あたり少なくとも 1 個体が発見される確率  $\Pr(y_j^{(it)} = 1 | z_{it} = 1)$  を発見確率と定義し, これを  $p_t$  と表す. また,  $t = 1$  においてサイトが占有されている確率  $\Pr(z_{i1} = 1)$  を初期占有確率として  $\psi_1$  とする. これらの要素を用いて, 動的サイト占有モデルは以下のように表される (図 3).

初期分布. サイト  $i = 1, \dots, I$  の初期占有状態  $z_{i1}$  はそれぞれ独立に, 初期占有確率  $\psi_1$  をパラメータとしたベルヌーイ分布に従う.

$$(2.1) \quad z_{i1} \sim \text{Bernoulli}(\psi_1)$$

システムモデル. 期間  $t = 2, \dots, T$  における各サイトの占有状態  $z_{it}$  は, 1 つ前の期間の占有状態  $z_{i,t-1}$  に条件付けられたベルヌーイ分布に従う.

$$(2.2) \quad z_{it} \sim \text{Bernoulli}(z_{i,t-1}\phi_{t-1} + [1 - z_{i,t-1}]\gamma_{t-1})$$

つまり期間  $t > 1$  におけるサイト  $i$  の占有確率は, 前の期間にサイトが占有 ( $z_{i,t-1} = 1$ ) されていればサイトの生存確率  $\phi_{t-1}$  に等しく, またサイトが占有されていなければ ( $z_{i,t-1} = 0$ ) 局所住み着き確率  $\gamma_{t-1}$  に等しい. 各サイト, 各期間の占有状態は独立にこの条件付き分布に従うと仮定される.

観測モデル. サイト  $i = 1, \dots, I$ , 期間  $t = 1, \dots, T$ , 調査  $j = 1, \dots, J_{it}$  の発見データ  $y_j^{(it)}$  は, 状態変数である占有状態  $z_{it}$  の条件付きベルヌーイ分布に従う.

$$(2.3) \quad y_j^{(it)} \sim \text{Bernoulli}(z_{it}p_t)$$

つまり, サイトが占有されていれば ( $z_{it} = 1$ ) 占有が確認される確率は発見確率  $p_t$  に等しく, サイトが占有されていなければ ( $z_{it} = 0$ ) 占有が確認される確率は 0 である. このような定式化によって偽陰性誤差がモデル化されていることが理解されよう (図 3). 一方で, 偽陽性誤差 (false positive error; 非占有サイトで個体が検出されること) はないと仮定されている. 偽陽性誤差は, 次節で説明される多状態動的サイト占有モデルの枠組みで扱われている.  $z_{it}$  を所与として, 発見データ  $y_j^{(it)}$  は互いに独立と仮定される. なお, ここで仮定するモデルは発見確率が  $j$  によら

ないため、観測モデルを二項分布  $\sum_j y_j^{(it)} \sim \text{Binomial}(J_{it}, z_{it} p_t)$  としてもよい。

このように表される動的サイト占有モデルには、初期占有確率  $\psi_1$ 、各期間のサイト「生存確率」 $\{\phi_t\}_{t=1}^{T-1}$ 、局所住み着き確率  $\{\gamma_t\}_{t=1}^{T-1}$  および発見確率  $\{p_t\}_{t=1}^T$  がパラメータとして含まれている。データからパラメータの値を推定するためのアプローチは2つある。1つはモデルの尤度を最大化するパラメータを特定する最尤推定である (MacKenzie et al., 2003)。もう1つのアプローチは、パラメータに対して事前分布を指定し、パラメータと状態変数の事後分布を求めるベイズ推定である。上記の動的サイト占有モデルでは各パラメータに対してベータ分布を指定すれば未知変数の全条件付き分布が明示的に求まるため、ギブスサンプリングによる事後分布からの効率的なサンプリングが実現できる (Royle and Kéry, 2007)。

上記の動的サイト占有モデルでは、生存確率、局所住み着き確率、発見確率の季節変化が仮定されているが、季節変化の他にも、これらのパラメータの変動は様々な形でモデル化できる (MacKenzie et al., 2003, 2006; Royle and Kéry, 2007)。パラメータの変動性は共変量を用いてモデル化されることが多い。つまり、パラメータを測定された共変量の関数として表現し、両者の関連を推定する。例えばサイト  $i$ 、期間  $t$  に測定された共変量  $x_{it}$  によってサイトと期間による生存確率の変動を説明する場合、ロジットリンク関数を用いて以下のようにモデル化することが多い。

$$(2.4) \quad \text{logit}(\phi_{it}) = \log(\phi_{it}/[1 - \phi_{it}]) = a + bx_{it}$$

$a$  は生存確率の(ロジット軸上での)切片、 $b$  は傾きのパラメータである。共変量としてはサイトと期間の組み合わせごとに測定された量 ( $x_{it}$ ) の他にも、サイトごと ( $x_i$ )、期間ごと ( $x_t$ ) に測定された量や調査ごとに測定された量 ( $x_j^{(it)}$ ) がそれぞれ利用されることがある (調査ごとに測定された共変量は  $j$  に依存するため、発見確率の変動を説明するために用いられる)。共変量を用いることなくパラメータの確率的なサイト間変動をモデル化する方法として、サイトランダム効果の導入が考えられる (Royle and Kéry, 2007)。

$$(2.5) \quad \text{logit}(\phi_{it}) = a_t + u_i$$

$$(2.6) \quad u_i \sim \text{Normal}(0, \sigma_\phi^2)$$

$a_t$  は期間ごとの生存確率の水準を定めるパラメータ、 $u_i$  は平均 0、分散  $\sigma_\phi^2$  の正規分布に従う生存確率のサイトランダム効果である。このようなランダム効果を含む動的サイト占有モデルの推定には、一般的にベイズ推定のアプローチが取られる (MacKenzie et al., 2006; Royle and Kéry, 2007)。メタ個体群では、局所個体群の占有動態が近傍サイトからの移入 (レスキュー効果; Brown and Kodric-Brown, 1977) の影響を受けると考えられ、パラメータ ( $\phi, \gamma$ ) には近傍サイトの占有状態に依存した空間変動性の存在が予想される。このようなパラメータの空間相関のモデル化の例は Risk et al. (2011), Bled et al. (2011a, 2011b), Yackulic et al. (2012), Sutherland et al. (2014) などを参照されたい。

MacKenzie et al. (2003) では、偽陰性誤差を考慮した動的サイト占有モデルの推定値と偽陰性誤差を考慮しない素朴な推定値が比較されている。ニシアメリカフクロウの亜種 (*Strix occidentalis caurina*) を対象としたデータの例では、推定された発見確率が 0.38–0.59 (モデル平均推定値) 程度の状況において、局所住み着き確率と局所絶滅確率に関する上記 2 種類の推定値の間に、最も大きなもので確率軸上で 0.14 の違いが報告されている (AIC 最良モデルとの比較; MacKenzie et al. (2003) の Table 2 を参照されたい)。素朴な推定値と比べると、動的サイト占有モデルによる動態パラメータの推定値は概して値が小さく、偽陰性誤差を考慮しない推測ではサイト (この例ではフクロウの縄張り) 利用の回転率が過大評価される傾向が見られている (MacKenzie et

al., 2003).

動的サイト占有モデルは統計モデルとして、捕獲再捕獲法のモデル(特に個体の一時的な移出入を考慮した捕獲再捕獲法モデル; Kendall et al., 1997)と類似性がある。また、より一般的には、動的サイト占有モデルは隠れマルコフモデル(離散的な状態変数が仮定された状態空間モデル; Zucchini and MacDonald, 2009)の1つである。状態変数が2値であり偽陰性誤差がモデル化されていること、メタ個体群デザインのために複数のサイトの時系列が同時にモデル化されていること、Pollockのロバストデザインのために期間内に複数のデータがありうることなどが動的サイト占有モデルの特徴である。

### 3. 複数の状態カテゴリ・観測カテゴリへの拡張

#### 3.1 動機・調査デザイン・データ

前節では、偽陰性の観測誤差を考慮した動的サイト占有モデルについて説明した。野外生態調査で偽陰性の検出誤差を完全になくすことは難しいため、サイト占有動態のバイアスの少ない推定値を得るという観点からは、動的サイト占有モデルの適用と、それを可能とする調査デザインを採用することが望ましい。一方で、種の確実な判別が難しい場合などには偽陽性の検出誤差が発生することがある。例えば対象種がサイトを占有していないにも関わらず、別の類似種を対象種と誤判別してしまうなどの理由で、観測された在データに「偽物」が混ざるような場合である。偽陽性誤差は一般的に、偽陰性誤差よりも頻度は低いと考えられるが、これもまたサイト占有率やサイト占有動態の推定に深刻なバイアスをもたらす要因となることが知られている(Royle and Link, 2006; McClintock et al., 2010a; Miller et al., 2011)。

また、研究の目的によっては、サイトの占有状態に関してより詳細な分類に関心がある場合もあるだろう。例えばサイトの占有・非占有に加えて占有個体(群)の繁殖の有無に関心があり、サイトが占有されていて繁殖が行われている、サイトは占有されているが繁殖は行われていない、サイトは占有されていない、の3状態が調べられるような場合である(Nichols et al., 2007; Martin et al., 2009)。あるいは、個体群の大きさなどに関連した離散的な指標に関心がある場合なども考えられよう(Royle and Link, 2005; Fiske et al., 2014)。このような場合には、観測されるデータが発見・非発見の2値よりも多くのカテゴリに分類されることが一般的であり、また背後にある状態も占有・非占有という2値ではなく、3つ以上のカテゴリのどれかに属すると考えることが自然である。

以下では動的サイト占有モデルの拡張(MacKenzie et al., 2009; Miller et al., 2013)を説明する。動的サイト占有モデルでは仮定される状態のカテゴリと観測のカテゴリの数がどちらも2つ(0か1)であったが、以下で説明されるモデルではこれらの片方、または両方が3つ以上のカテゴリから構成される。このようなモデルクラスをここでは多状態動的サイト占有モデル(multistate dynamic site occupancy model)と呼ぶことにする。多状態動的サイト占有モデルでは占有状態の不確実性を考慮してサイト占有動態が推測され、サイト占有調査データに関する広範な問題を統一的に取扱うことができる。実際に多状態動的サイト占有モデルの適用範囲は多岐にわたり、偽陰性誤差だけでなく偽陽性誤差も考慮したメタ個体群動態パラメータの推定バイアスの補正(Miller et al., 2013)、繁殖状況や個体数指標などによって表される個体群動態の推測(MacKenzie et al., 2009; Martin et al., 2009; Fiske et al., 2014)のほか、生息地の状況と個体群動態を同時に考慮した推測(MacKenzie et al., 2011; Miller et al., 2012; Martin et al., 2010)や感染症の動態の推測(McClintock et al., 2010b; Conn and Cooch, 2009も参照)、複数種からなる群集動態の推測(Miller et al., 2012; Fukaya and Royle, 2013; Yackulic et al., 2014)など様々である。

動的サイト占有モデルと同様に、多状態動的サイト占有モデルでは基本的にメタ個体群デザ

インと Pollock のロバストデザインを組み合わせた調査デザインを前提とする．すなわち，サイト  $i$  ( $i = 1, \dots, I$ )，期間  $t$  ( $t = 1, \dots, T$ )，調査  $j$  ( $j = 1, \dots, J_{it}$ ) のそれぞれに，サイト占有状態に関する合計で  $M \geq 2$  カテゴリーのデータ  $y_j^{(it)} \in \{1, \dots, M\}$  が得られているとする．前節とは異なり，観測されるデータは 2 値に限られないことに注意されたい．ある期間  $t$  の間に占有状態は変化しないが，期間  $t-1$  や  $t+1$  では占有状態が異なる場合があり，調査は状態の変化がないと考えられる期間に反復して行われることは動的サイト占有モデルの仮定と同じである (図 2)．多状態動的サイト占有モデルの目標は，このサイト占有状態データの配列から，占有状態の観測の不確実性を考慮してサイトの占有状態およびその推移確率に関する推測を行うことである．

### 3.2 モデルと推測

ここでは，多状態動的サイト占有モデルの一般的な状態空間表現を与える．前節と同様，直接的には観測されない状態変数としてサイト  $i$  ( $i = 1, \dots, I$ )，期間  $t$  ( $t = 1, \dots, T$ ) におけるサイトの占有状態を表す変数を導入し，これを  $z_{it}$  と表す．サイト占有状態には合計で  $N \geq 2$  個のカテゴリの存在を仮定する ( $z_{it} \in \{1, \dots, N\}$ )．観測データと同じく，状態も 2 値に限られないことに注意されたい．占有状態の時間変化は，一般的な推移確率によって表現される．期間  $t$  に状態  $k$  であったサイトが期間  $t+1$  で状態  $l$  となる確率を  $p_{lk}^{(t)} = \Pr(z_{i,t+1} = l | z_{it} = k)$  とする．各期間の推移確率を  $N$  行  $N$  列の行列としてまとめて  $\mathbf{P}_t$  ( $t = 1, \dots, T-1$ ) とすると，これは期間  $t$  の推移確率行列を構成する．

$$(3.1) \quad \mathbf{P}_t = \begin{pmatrix} p_{11}^{(t)} & p_{12}^{(t)} & \cdots & p_{1N}^{(t)} \\ p_{21}^{(t)} & p_{22}^{(t)} & \cdots & p_{2N}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}^{(t)} & p_{N2}^{(t)} & \cdots & p_{NN}^{(t)} \end{pmatrix}$$

一方で観測データの変動は，サイト占有状態に依存した条件付き確率分布によって表現される．サイト  $i$ ，期間  $t$  の占有状態が  $z_{it} = n$  であった場合に，観測データ  $y_j^{(it)} = m$  が得られる確率を  $q_{mn}^{(t)} = \Pr(y_j^{(it)} = m | z_{it} = n)$  と表す．各期間の観測確率を  $M$  行  $N$  列の行列としてまとめたものを  $\mathbf{Q}_t$  ( $t = 1, \dots, T$ ) とすると，これは期間  $t$  の観測確率行列となる．

$$(3.2) \quad \mathbf{Q}_t = \begin{pmatrix} q_{11}^{(t)} & q_{12}^{(t)} & \cdots & q_{1N}^{(t)} \\ q_{21}^{(t)} & q_{22}^{(t)} & \cdots & q_{2N}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{M1}^{(t)} & q_{M2}^{(t)} & \cdots & q_{MN}^{(t)} \end{pmatrix}$$

$\mathbf{P}_t$  と  $\mathbf{Q}_t$  は，各列の要素の和が 1 となる確率行列であることに注意されたい．また， $t=1$  においてサイトの占有状態が  $n$  である確率を  $\psi_n = \Pr(z_{i1} = n)$  とし，これをまとめた初期占有確率ベクトルを  $\boldsymbol{\Psi}_1 = (\psi_1, \dots, \psi_N)$  とする．これらの要素を用いて，多状態動的サイト占有モデルは以下のように表現される．

初期分布．以下では，試行回数が 1 の多項分布をカテゴリカル分布 (categorical distribution) とよぶ．サイト  $i = 1, \dots, I$  の初期占有状態  $z_{i1}$  はそれぞれ独立に，初期占有確率  $\boldsymbol{\Psi}_1$  をパラメータとしたカテゴリカル分布に従う．

$$(3.3) \quad z_{i1} \sim \text{Categorical}(\boldsymbol{\Psi}_1)$$

システムモデル． $t = 2, \dots, T$  における各サイトの占有状態  $z_{it}$  は，1 つ前の期間の占有状態

$z_{i,t-1}$  に条件付けられたカテゴリカル分布に従う確率変数である。

$$(3.4) \quad z_{it} \sim \text{Categorical}(\mathbf{P}_{t-1}(z_{i,t-1}))$$

ここで  $\mathbf{P}_t(n)$  は推移確率行列  $\mathbf{P}_t$  の  $n$  列目のベクトル  $(p_{1n}^{(t)}, \dots, p_{Nn}^{(t)})$  である。各サイト、各期間の占有状態は独立にこの条件付き分布に従うと仮定される。

観測モデル。サイト  $i = 1, \dots, I$ , 期間  $t = 1, \dots, T$ , 調査  $j = 1, \dots, J_{it}$  の観測データ  $y_j^{(it)}$  は、状態変数である占有状態  $z_{it}$  の条件付きカテゴリカル分布に従う。

$$(3.5) \quad y_j^{(it)} \sim \text{Categorical}(\mathbf{Q}_t(z_{it}))$$

ここで  $\mathbf{Q}_t(n)$  は観測確率行列  $\mathbf{Q}_t$  の  $n$  列目のベクトル  $(q_{1n}^{(t)}, \dots, q_{Mn}^{(t)})$  である。 $z_{it}$  を所与として、発見データ  $y_j^{(it)}$  は互いに独立であることが仮定される。

このように、2つ以上のカテゴリに分類される占有状態および観測データを扱うために、多状態動的サイト占有モデルではベルヌーイ分布の代わりにカテゴリカル分布が用いられる。多状態動的サイト占有モデルのパラメータ推定および共変量の導入などの拡張に関しては、動的サイト占有モデルと同様に考えることができる (MacKenzie et al., 2009)。ただし、上記の一般的な多状態動的サイト占有モデルはカテゴリ数の増加に伴って推移確率行列や観測確率行列が大きくなるため、モデルに含まれるパラメータの数が多くなる。扱われる具体的な問題に応じて、推移確率行列あるいは観測確率行列、またはその両方に制約が置かれることも少なくない。観測過程が「階層的」であると考えられる場合、すなわち上位のカテゴリ(サイト占有かつ繁殖成功など)の観測結果は占有状態が下位のカテゴリ(サイト非占有など)の場合には生じないような状況が仮定される場合には、観測行列が三角行列に制約されることがある。例えば(1:非占有, 2:占有, 3:占有かつ繁殖あり)という3つの占有状態に対して、観測カテゴリが(1:個体を発見せず, 2:個体を発見, 3:個体を発見して繁殖の証拠も発見)に分けられるとき、以下のような観測行列が考えられる (MacKenzie et al., 2009)。

$$(3.6) \quad \mathbf{Q}_t = \begin{pmatrix} 1 & q_{12}^{(t)} & q_{13}^{(t)} \\ 0 & q_{22}^{(t)} & q_{23}^{(t)} \\ 0 & 0 & q_{33}^{(t)} \end{pmatrix}$$

これは個体の発見と繁殖の確認に関して、偽陰性の検出誤差を仮定していることに相当する。一般的に3行3列の観測確率行列に含まれる自由パラメータの数は6つであるが、この形に制約された場合は3つとなる。

また、推移確率行列や観測確率行列のいくつかの要素が条件付き確率の積の形で与えられることがある。例えば式(3.6)の観測確率行列は以下のようにも表すことができる (Nichols et al., 2007; MacKenzie et al., 2009)。

$$(3.7) \quad \mathbf{Q}_t = \begin{pmatrix} 1 & 1 - q_2^{(t)} & 1 - q_3^{(t)} \\ 0 & q_2^{(t)} & q_3^{(t)}(1 - r^{(t)}) \\ 0 & 0 & q_3^{(t)}r^{(t)} \end{pmatrix}$$

ここで  $q_n^{(t)}$ ,  $n = 2, 3$  は状態が  $n$  のときに個体が発見される確率、 $r^{(t)}$  は状態3(繁殖が行われている)サイトで個体が発見された際に、同時に繁殖も確認される確率である。つまりこの形では個体の発見確率と繁殖の確認確率が区別されており、階層的な観測過程に関するより明示的な情報が得られる表現となっている。

Miller et al. (2013)は偽陽性誤差を考慮した多状態動的サイト占有モデルを提案している。このモデルはここで説明されたものより一般的なものであり、複数の観測確率行列を導入して、

異なる種類の観測データを同時に扱う。

以上、占有状態の不確実性を考慮し、サイト占有動態に関する広範な問題に適用可能な多状態動的サイト占有モデルの一般的な状態空間表現を見てきた。捕獲再捕獲法で multievent model と呼ばれている統計モデル (Pradel, 2005) は、多状態動的サイト占有モデルと類似性が高い。動的サイト占有モデルと同様に、多状態動的サイト占有モデルもまた隠れマルコフモデルの特殊な例である (Fiske et al., 2014)。多状態動的サイト占有モデルの1つの具体例として、次節では観測誤差を考慮した固着性生物群集動態の推測について紹介する。

#### 4. 観測誤差を考慮した固着性生物群集動態の推測

##### 4.1 動機・調査デザイン・データ

推移確率行列は、土壌や岩礁などの基盤に定着して生活する固着性生物群集 (例えば草本やサンゴなど) のサイト占有動態と、その結果生じる各種の相対優占度 (群集全体の生物量に対して各種が占める割合) の動態を要約するパラメータとして、群集生態学研究において推測の対象となってきた (例えば Tanner et al., 1994; Wootton, 2001; Hill et al., 2004; Tsujino et al., 2010)。群集動態の文脈における推移確率とは、種や分類群といった、群集を構成する要素の間で起こるサイト占有状態の推移しやすさを表したものであり、これを明らかにすることで群集構成種の相対優占度の時間変化の過程やその帰結を考察できる。以下で詳しく説明されるように、ここで想定されるサイトとは、空間的な広がり非常に小さい、生息地の一部分である。野外では利用可能な空間の全てが固着性生物に占有されるとは限らないため、生態学的状態の集合には生物だけでなく「空き地」という状態が含まれることが普通である。また、特定の種の出現頻度が低いためにその種に関する状態の推移がほとんど観察されない場合や、多種多様な種が出現する中で種の同定自体がそもそも難しいような場合には、複数の種を (例えば機能や形態が類似したもの同士で) グループにまとめることが多い。したがって実証研究においては、いくつかの種、種のグループ、および空き地などから構成された「生物群集」の動態が推移確率行列を用いてモデル化される。

野外群集の推移確率を推定するにはどうすればよいだろうか。固着性生物は場所を変えずに生息することから、推移確率が一定と考えられる空間領域 (対象種や環境に依存するが、例えば数十センチメートル～数メートル四方) に占有状態を調査する微小な領域 (例えば半径数ミリメートル) である固定調査点を複数設置し、各点の占有状態の経時的な観測によって推移の頻度を調べればよいはずである (図 4(A))。いま  $n_{lk}$  を、ある時点で占有状態が  $k$  であり、次の時点には状態  $l$  に推移した調査点の数とすると、推移確率の素朴な推定量として

$$(4.1) \quad \hat{p}_{lk} = \frac{n_{lk}}{\sum_s n_{sk}}$$

を考えることができる。これは状態  $k$  から始まる推移に関する単純な多項分布モデルの最尤推定量に相当する (Spencer and Susko, 2005)。

この推定量は固定された調査点が毎度正確に調査されている場合には妥当であるが、そうではない場合には推移確率の偏った推定量となる (Conway-Cranos and Doak, 2011; Fukaya and Royle, 2013)。つまり、調査を行うべき位置とは異なる場所を誤って観測してしまうと、調査点上で生じた実際の状態の推移を観測していないことになるため、推移確率は正しく推定されないのである。言うまでもなく、調査点上の占有状態の観測には細心の注意が払われるべきであるが、それでも野外調査においては固定された調査点の、その正しい位置を正確に捉えて再調査することが難しいことも少なくない (図 4(B))。その場合、観測される生態学的状態の中には、実際に調査点を占有している生態学的状態とは異なったものが含まれてしまう可能性がある。

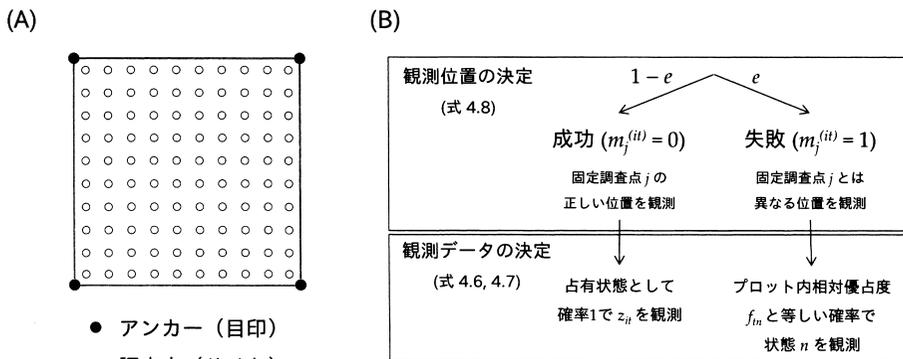


図 4. (A) 推移確率を推定するための固着性生物群集動態の調査デザイン。岩礁などの基盤上に、占有状態を調査する微小な領域を固定調査点(サイト)として多数設置する。調査点は固定されたアンカーなどを目印とした調査領域の内側に設置される。調査点を内包するような調査対象の空間領域をプロットと呼ぶ。各調査点は、アンカーを目印として調査点の位置を示す観測器具などを用いて、その位置を特定してから観測される。(B) Fukaya and Royle (2013) のモデルで仮定される階層的な観測過程。

このような観測誤差は、固定調査点が小さく、また固定調査点の観測を行うための器具の精度があまり高くないことにその一因がある。特に対象となる生物が小さい場合などには、観測位置のわずかなずれが観測結果に大きな影響を及ぼしかねない(Conway-Cranos and Doak, 2011)。

こうした状況では、観測誤差の発生機構が考慮された統計モデルを用いて推移確率を推定する必要があるだろう。固着性生物群集動態研究の文脈でこのアイデアを最初に提案し、実用的な枠組みを提示したのは Conway-Cranos and Doak (2011) である。この研究では観測誤差率を新しいパラメータとして含む多項分布モデルの拡張を考え、観測誤差を考慮した推移確率の最尤推定法が提案された。

ところで、この新しい推定量と素朴な推定量(式(4.1))は、どちらも集約された推移頻度データ  $n_{ij}$  に基づいたものであり、データがどの調査点の観測から得られたものであるかは考慮されない。これとは異なるアプローチとして、個々の調査点で生じる占有動態と観測過程をモデル化することによって、データの得られた調査点の情報を捨てることなく観測誤差を考慮した推移確率の推定を行うこともできるはずである。このようなアイデアに基づき、多状態動的サイト占有モデルの枠組みを用いてバイアスの少ない推移確率の推定を目指したモデルを提案したのが Fukaya and Royle (2013) である。以下ではこのモデルについて説明する。

これまでと同様、想定される調査デザインはメタ個体群デザインと Pollock のロバストデザインの組み合わせである。推移確率を推定するために固定調査点(サイト)を多数設置することから(図 4(A))、データはメタ個体群デザインで得られた占有動態データと形式的に同じ構造を持つ。つまり固着性生物群集調査の文脈では、個々の固定調査点がメタ個体群研究におけるサイト(局所個体群)に対応し、調査点の集合がメタ個体群に対応する。ただし、ここでの関心は群集の局所的な動態にあることを考えると、「メタ個体群デザイン」という呼び方に生態学的に適切な意味はない。固着性生物群集の調査において、調査点は格子状に規則正しく配置されることが多いようであるが、無作為に配置されていても構わない。全ての固定調査点を包含する調査対象となる空間領域を、以下ではプロットと呼ぶ(図 4(A))。

サイト  $i$  ( $i = 1, \dots, I$ )、期間  $t$  ( $t = 1, \dots, T$ )、調査  $j$  ( $j = 1, \dots, J_{it}$ ) のそれぞれについて、サ

イト占有状態に関する合計で  $N \geq 2$  カテゴリーのデータ  $y_j^{(it)} \in \{1, 2, \dots, N\}$  が得られているとする. 先に述べたように,  $y_j^{(it)}$  はサイトを占有する生態学的状態の1つであり, それは特定の種, 種のグループ, または「空き地」のどれかである(状態の集合の1例を挙げると, [種 A, 種 B, 種 C-E のグループ, 空き地] のようになる). またサイトの占有状態  $z_{it}$  もこれらの生態学的状態のいずれかであり, したがって占有状態のカテゴリの数は観測と同じ  $N$  である.

#### 4.2 モデルと推測

このモデルでは, 占有状態の観測について「成功」と「失敗」の2種類が考慮される(図4(B)). 観測の成功とは観測者が固定調査点の正しい位置を観測できた場合に相当し, その場合はサイトを占有している生態学的状態が確率1で観測されると仮定される. 一方で観測の失敗とは, 観測者が誤った位置で占有状態の観測を行ってしまった場合に相当し, この場合はプロットにおける各状態の相対優占度に等しい確率でどれか1つの状態が観測されると仮定される. 1回の観測あたりに観測が失敗する確率(観測誤差率)を  $e$  とし, これが時間的・空間的に一定であると仮定すれば, 式(3.2)に対応する各期間の観測確率行列は以下のように表される.

$$(4.2) \quad Q_t = \begin{pmatrix} (1-e) + ef_{t1} & ef_{t1} & \dots & ef_{t1} \\ ef_{t2} & (1-e) + ef_{t2} & \dots & ef_{t2} \\ \vdots & \vdots & \ddots & \vdots \\ ef_{tN} & ef_{tN} & \dots & (1-e) + ef_{tN} \end{pmatrix}$$

ここで  $f_{tn}$  は期間  $t$  における状態  $n$  のプロット内相対優占度 ( $0 \leq f_{tn} \leq 1, \sum_m f_{tm} = 1$ ) である. つまり  $z_{it} = n$  である場合に  $n$  以外の状態  $m$  が観測される確率は, 観測に失敗して, かつ  $m$  が観測される確率  $ef_{tm}$  に等しく, 一方  $n$  が観測される確率は観測が成功する確率  $(1-e)$  と観測に失敗して  $n$  を観測する確率  $ef_{tn}$  の和に等しい.

一方で, 推移確率行列については時間的に一定であるという仮定が置かれる. したがって, 式(3.1)に対応する各期間の推移確率行列は以下ようになる.

$$(4.3) \quad P_t = P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix}$$

前節で与えられた多状態動的サイト占有モデルの一般的な状態空間表現((3.3)–(3.5)式)において, 以上の推移確率行列と観測確率行列を用いれば Fukaya and Royle (2013)で提案された群集動態モデルが得られる. 実際には, この論文で与えられた状態空間表現は若干異なり, 各観測における誤差の有無を表す補助的な指示変数( $m$ )が導入されている. この補助変数を導入することで, 階層的な観測過程(観測位置の決定と, それに続く観測データの決定; 図4(B))がより明示的に表現されるとともに, このモデルを階層ベイズモデルとして定式化した際に全ての未知変数についてのギブスサンプリングを構成できる. 以下ではこの補助変数を含む形でのモデルの状態空間表現を与え, さらに Fukaya and Royle (2013)に倣い, パラメータに事前分布を設定して階層ベイズモデルとして表現する. フリーのMCMCソフトウェアであるJAGSを用いてこのモデルを当てはめるためのモデルスクリプトを図5に示す.

初期分布. サイト  $i = 1, \dots, I$  の初期占有状態  $z_{i1}$  はそれぞれ独立に, 初期占有確率  $\Psi_1$  をパラメータとしたカテゴリカル分布に従う.

$$(4.4) \quad z_{i1} \sim \text{Categorical}(\Psi_1)$$

```

model {
  ## 初期分布およびシステムモデル
  for (i in 1:I) {
    z[i, 1] ~ dcat(Psi_1[]) # (4.4)式

    for (t in 2:T) {
      z[i, t] ~ dcat(P[, z[i, t-1]]) # (4.5)式
    }
  }

  ## 観測モデル
  for (t in 1:T) {
    for (n in 1:N) {
      f[t, n] <- sum(z[, t] == n) / I # (4.9)式
    }
    for (i in 1:I) {
      for (j in 1:J[i, t]) {
        m[i, t, j] ~ dbern(e) # (4.8)式
        for (n in 1:N) {
          # q[i, t, j, ]: 観測確率ベクトル
          q[i, t, j, n] <-
            ((1 - m[i, t, j]) * equals(z[i, t], n)
            + m[i, t, j] * f[t, n])
        }
        y[i, t, j] ~ dcat(q[i, t, j, ]) # (4.6, 4.7)式
      }
    }
  }

  ## 事前分布の指定
  e ~ dbeta(1, 1) # (4.10)式
  for (n in 1:N) {
    P[1:N, n] ~ ddirch(v1N) # (4.11)式
  }
  Psi_1[1:N] ~ ddirch(v1N) # (4.12)式
}

```

図 5. 観測誤差を考慮した群集動態モデルの JAGS コード.

システムモデル.  $t = 2, \dots, T$  における各サイトの占有状態  $z_{it}$  は, 1 つ前の期間の占有状態  $z_{i,t-1}$  に条件付けられたカテゴリカル分布に従う確率変数である.

$$(4.5) \quad z_{it} \sim \text{Categorical}(\mathbf{P}(z_{i,t-1}))$$

$\mathbf{P}(n)$  は推移確率行列  $\mathbf{P}$  の  $n$  列目のベクトルである. 各サイト, 各期間の占有状態は独立にこの条件付き分布に従うと仮定される.

観測モデル.  $\mathbf{f}_t = (f_{t1}, \dots, f_{tN})$  を期間  $t$  における状態のプロット内相対優占度ベクトル,  $m_j^{(it)}$  を,  $m = 0$  が観測の成功を,  $m = 1$  が観測の失敗を表す補助変数とする. サイト  $i = 1, \dots, I$ , 期間  $t = 1, \dots, T$ , 調査  $j = 1, \dots, J_{it}$  の観測データ  $y_j^{(it)}$  は, サイト占有状態  $z_{it}$  および潜在変数  $m_j^{(it)}$ , 相対優占度  $\mathbf{f}_t$  の条件付きカテゴリカル分布に従う.

$$(4.6) \quad y_j^{(it)} \sim \text{Categorical}(\mathbf{1}'_N(z_{it})) \quad \text{when } m_j^{(it)} = 0$$

$$(4.7) \quad y_j^{(it)} \sim \text{Categorical}(\mathbf{f}_t) \quad \text{when } m_j^{(it)} = 1$$

$\mathbf{1}'_N(n)$  は  $n$  番目の要素が 1, それ以外は 0 からなる長さ  $N$  のベクトルである. 観測誤差の有無  $m_j^{(it)}$  は, 調査ごとに独立に誤差率  $e$  のベルヌーイ分布に従うと仮定される.

$$(4.8) \quad m_j^{(it)} \sim \text{Bernoulli}(e)$$

また、各期間  $t$  の相対優占度  $f_{tn}, n = 1, \dots, N$  は導出パラメータとして、 $z_{it}$  の実現値から以下のように計算される。

$$(4.9) \quad f_{tn} = \frac{\sum_{i=1}^I I(z_{it} = n)}{I}$$

ここで  $I(z = n)$  は  $z = n$  の場合に 1、それ以外で 0 となる指示関数である。

事前分布の指定. このモデルのパラメータ  $(e, \mathbf{P}, \Psi_1)$  に対する事前分布は任意に指定できるが、以下の(無情報)事前分布を用いることでギブスサンプリングに必要な全条件付き分布を明示的に求めることができる。

$$(4.10) \quad e \sim \text{Beta}(1, 1)$$

$$(4.11) \quad \mathbf{P}(n) \sim \text{Dirichlet}(\mathbf{1}_N), n = 1, \dots, N$$

$$(4.12) \quad \Psi_1 \sim \text{Dirichlet}(\mathbf{1}_N)$$

$\mathbf{1}_N$  は要素が全て 1 からなる長さ  $N$  のベクトルである。

Fukaya and Royle (2013) では  $I = 200, T = 10, N = 5$  という条件のもと、いくつかの  $e$  の値における推移確率の推定バイアスがシミュレーションによって評価されている(推定バイアスは、 $\mathbf{P}$  の事後分布の spatial median を点推定値として評価された)。この条件のもとでは、観測誤差を考慮しない素朴な推定量(4.1)は  $e > 0$  の場合にバイアスを生じるが、Conway-Cranos and Doak (2011) と Fukaya and Royle (2013) の方法ではバイアスが大きく改善され、例えば  $e \leq 0.45$  といった状況ではバイアスがほとんどないことが示されている。また同じシミュレーションの中で、少なくとも推移確率行列と観測誤差率が一定という条件のもとでは、Pollock のロバストデザインが成立しない状況(つまり全ての  $i, t$  について反復となる観測がなされておらず、 $J_{it} = 1$  である)においても観測誤差率  $e$  が推定可能であり、反復調査のデータがなくとも推移確率のバイアスの少ない推定値が得られることが確かめられている(Fukaya and Royle, 2013)。

固着性生物の群集動態を多状態動的サイト占有モデルを用いてモデル化することの利点として、各調査点における占有状態が明示的に考慮されることによって、群集構造に依存した過程を柔軟にモデル化できることが挙げられる。観測過程のモデル化において式(4.9)を用いることはその1つの例といえる。ところで相対優占度  $f_{tn}$  に関する式(4.9)の仮定は、プロット内における各生態学的状態の空間的な分布がランダムなモザイクであることが暗に仮定されていることになるが、そのような仮定はあまり現実的ではない。なぜなら固着性生物の空間分布はしばしば集中的なパターンを示すからである(群集の空間パターンがランダムであるという仮定は、Conway-Cranos and Doak (2011) の方法でも観測誤差の影響を考慮するために置かれている)。Fukaya and Royle (2013) の枠組みでこのような空間パターンの影響を考慮するための方法として、各調査点の空間的位置に関する情報を導入し、局所的な相対優占度をモデル化することが考えられる(この場合、相対優占度ベクトルはサイト  $i$  に依存する)。具体的な方法として例えば、相対優占度の推定においてプロット内の全ての調査点を等しく用いるのではなく、近傍の調査点のみを用いたモデル化(Bled et al., 2011a, 2011b; Yackulic et al., 2012)が考えられる。 $A_i$  と  $a_i$  をそれぞれサイト  $i$  の近傍と定義されたサイトの添字集合とその数として、局所相対優占度は以下のように推定される。

$$(4.13) \quad f_{itn} = \frac{\sum_{j \in A_i} I(z_{jt} = n)}{a_i}$$

距離に応じた適当な重み付けを考えることもできるだろう。 $\mathbf{x}_i$  をサイト  $i$  の空間座標、 $K(\bullet)$  を

距離に依存した重みを与える関数として、局所相対優占度は以下のように推定される。

$$(4.14) \quad f_{itn} = \frac{\sum_{j=1}^I K(\|\mathbf{x}_j - \mathbf{x}_i\|) I(z_{jt} = n)}{\sum_{k=1}^N \sum_{j=1}^I K(\|\mathbf{x}_j - \mathbf{x}_i\|) I(z_{jt} = k)}$$

このように推定される局所的な群集構造は、観測過程のみならず生態的過程（つまりサイト占有動態）とも関連付けることができる (Bled et al., 2011a, 2011b; Risk et al., 2011)。

## 5. 終わりに

本稿では、メタ個体群デザインおよび Pollock のロバストデザインを用いて得られたサイト占有状態の時系列に適合する動的サイト占有モデルと、その一般化である多状態動的サイト占有モデルについて概説した。これらの枠組みは、関心のある生態学的動態に加えて観測過程を同時にモデル化することにより、観測の不完全性に由来するサイト占有状態の不確実性を考慮した推測を実現するものである。偽陰性検出誤差のような観測の不完全性がある場合、これを無視したパラメータの推定は大きくバイアスしてしまうことがあり、占有状態とその動態に関する偏りの少ない推測を行うためにはこれらの統計的枠組みを用いることが望ましい。動的サイト占有モデル（特に多状態動的サイト占有モデル）は潜在的に様々な問題に適用可能な枠組みであり、広い範囲の生態学研究でその有効な応用を考えることができると考えられる (Bailey et al., 2014)。

長期的な観測に大きなコストがかかりがちな生態学研究では珍しくないことであるが、動的サイト占有モデルが当てはめられるデータは、時系列データとしては非常に短いことが多い。期間の数が 10 よりも少ない適用例も多く (例えば MacKenzie et al., 2003)、長期研究の場合でも、期間の数が例えば 30 を超えるものはほとんどないと思われる。つまり多くの適用例で、動的サイト占有モデルの動態パラメータの推定はメタ個体群デザインによる空間的な反復によって達成されていると言える。そのため動的サイト占有モデルは、統計モデルとしては時系列モデルである隠れマルコフモデル (Zucchini and MacDonald, 2009) と等価であるが、状況としては、医学統計分野などに見られる複数の被験者を対象に反復測定されたデータ (経時データ) の解析 (藤越, 2009; 船渡川・船渡川, 2015) により近いものがある。経時データでは、被験者数がある程度多くても時点数は限られていることが少なくない (船渡川・船渡川, 2015)。また、多状態モデルにおいては、観測の次元や仮定される状態の次元も基本的には高くなく、これまでの適用例では多くの場合で 10 次元を超すことはないようである。そのため、多状態動的サイト占有モデルのほとんどは一般的な隠れマルコフモデルとしては単純なものであり、状態の次元の増加に伴うパラメータ数や尤度の計算負荷の劇的な増加といった本質的な推定の難しさ (Zucchini and MacDonald, 2009) はほとんどない。

しかしながら、生態学研究の感覚からすると、動的サイト占有モデルで要求されるデータの量は必ずしも少なくない。モデルが単純でも、情報が限られた状況では状態変数を仮定するサイト占有モデルのパラメータ推定は不安定になりがちである (Moreno and Lele, 2010; Welsh et al., 2013)。また、サイト占有モデルでは観測誤差の生じる頻度に関する情報を収集できるような調査計画が不可欠であり、ロバストデザインに基づく反復調査はこれを保証するために非常に重要である。反復を取らない場合と比べると調査にかかるコストは増大してしまうものの、観測誤差を考慮した推測を行うためにはその実施が推奨される。反復調査は必ずしも全ての期間とサイトで等しく行う必要はなく、部分的な実施によっても観測誤差率の推定に役立つが、関心のあるパラメータの変動を識別するためには実施する時間と場所に関して注意深い検討が必要だろう。

反復調査に限らず、動的サイト占有モデルの適用を考える場合にはモデルの仮定に適合するデータを野外で取得するために考慮すべき点を研究の計画段階で認識しておくことが重要である。本稿ではこのことについて具体的には取り上げなかったが、サイト占有状態の調査デザインに関する主要な問題として調査地の設定 (MacKenzie and Royle, 2005; MacKenzie et al., 2006; Efford and Dawson, 2012), 反復調査に関する注意 (MacKenzie and Royle, 2005; MacKenzie et al., 2006; Kendall and White, 2009), 必要な調査努力や最適な努力量の配分 (MacKenzie and Royle, 2005; MacKenzie et al., 2006; Bailey et al., 2007; Guillera-Arroita et al., 2010; Guillera-Arroita and Lahoz-Monfort, 2012; McKann et al., 2013) などが議論されている。

データ収集の要求は大きいものの、観測誤差を考慮して生態的過程の信頼性の高い推測を実現できることは動的サイト占有モデルの大きな利点である。最尤法によって動的サイト占有モデルを当てはめるソフトウェアとして、PRESENCE (<http://www.mbr-pwrc.usgs.gov/software/presence.html>), MARK (White and Burnham, 1999), E-SURGE (Choquet et al., 2009), R のパッケージとして unmarked (Fiske and Chandler, 2011) などが公開されている (Gimenez et al., 2014)。JAGS (Plummer, 2003) や WinBUGS (Lunn et al., 2000) などのソフトウェアは、モデルの状態空間表現をほぼそのままの形で記述したスクリプト (JAGS コードや BUGS コードなどと呼ばれる) を用意することで、動的サイト占有モデルのベイズ推定 (MCMC による事後分布推定) を行うことができる。ベイズ的なアプローチによる動的サイト占有モデルの当てはめについては Royle and Kéry (2007), Royle and Dorazio (2008), Kéry and Schaub (2012) などを参照されたい。

動的サイト占有モデルは複数のサイトの占有状態とその動態に関する推測を行うための統計モデルであるが、同様のモデルは複数の標識個体の離散時系列データを生じる捕獲再捕獲法でも用いられる。実際、サイト占有状態の推測と標識個体の状態の推測の間には密接な関連があり、動的サイト占有モデルにおいて「サイト」を「個体」に置きかえれば、類似したシステムモデルと観測モデルによって種々の捕獲再捕獲法データの変動が表現される。Royle and Dorazio (2008) や Kéry and Schaub (2012) は状態空間表現を用いて両者の類似性を理解するのに適した文献である。

## 謝 辞

小泉逸郎氏, 船渡川伊久子氏, 比嘉基紀氏, 持橋大地氏および匿名の査読者より原稿に対する有益な助言をいただいた。本研究は JSPS 科研費 15K18617, 15H04418 の助成を受けて行われた。

## 参 考 文 献

- Bailey, L. L., Hines, J. E., Nichols, J. D. and MacKenzie, D. I. (2007). Sampling design trade-offs in occupancy studies with imperfect detection: Examples and software, *Ecological Applications*, **17**, 281–290.
- Bailey, L. L., MacKenzie, D. I. and Nichols, J. D. (2014). Advances and applications of occupancy models, *Methods in Ecology and Evolution*, **5**, 1269–1279.
- Bled, F., Royle, J. A. and Cam, E. (2011a). Assessing hypotheses about nesting site occupancy dynamics, *Ecology*, **92**, 938–951.
- Bled, F., Royle, J. A. and Cam, E. (2011b). Hierarchical modeling of an invasive spread: The Eurasian Collared-Dove *Streptopelia decaocto* in the United States, *Ecological Applications*, **21**, 290–302.

- Brown, J. H. and Kodric-Brown, A. (1977). Turnover rates in insular biogeography: Effect of immigration on extinction, *Ecology*, **58**, 445–449.
- Choquet, R., Rouan, L. and Pradel, R. (2009). Program E-SURGE: A software application for fitting multievent models, *Modeling Demographic Processes in Marked Populations* (eds. D. L. Thomson, E. G. Cooch and M. J. Conroy), 845–865, Springer, New York.
- Conn, P. B. and Cooch, E. G. (2009). Multistate capture-recapture analysis under imperfect state observation: An application to disease models, *Journal of Applied Ecology*, **46**, 486–492.
- Conway-Cranos, L. L. and Doak, D. F. (2011). Sampling errors create bias in Markov models for community dynamics: The problem and a method for its solution, *Oecologia*, **167**, 199–207.
- Dail, D. and Madsen, L. (2013). Estimating open population site occupancy from presence-absence data lacking the robust design, *Biometrics*, **69**, 146–156.
- Efford, M. G. and Dawson, D. K. (2012). Occupancy in continuous habitat, *Ecosphere*, **3**, p.32.
- Fiske, I. J. and Chandler, R. B. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance, *Journal of Statistical Software*, **43**, 1–23.
- Fiske, I. J., Royle, J. A. and Gross, K. (2014). Inference for finite-sample trajectories in dynamic multi-state site-occupancy models using hidden Markov model smoothing, *Environmental and Ecological Statistics*, **21**, 313–328.
- 藤越康祝 (2009). 『経時データ解析の数理』, 朝倉書店, 東京.
- 深谷肇一 (2016). 状態空間モデルによる時系列解析とその生態学への応用, 日本生態学会誌, **66**(2) (印刷中).
- Fukaya, K. and Royle, J. A. (2013). Markov models for community dynamics allowing for observation error, *Ecology*, **94**, 2670–2677.
- 船渡川 伊久子, 船渡川 隆 (2015). 『経時データ解析』, 朝倉書店, 東京.
- Gimenez, O., Blanc, L., Besnard, A., Pradel, R., Doherty, P. F., Marboutin, E. and Choquet, R. (2014). Fitting occupancy models with E-SURGE: Hidden Markov modelling of presence-absence data, *Methods in Ecology and Evolution*, **5**, 592–597.
- Guillera-Arroita, G. and Lahoz-Monfort, J. J. (2012). Designing studies to detect differences in species occupancy: Power analysis under imperfect detection, *Methods in Ecology and Evolution*, **3**, 860–869.
- Guillera-Arroita, G., Ridout, M. S. and Morgan, B. J. T. (2010). Design of occupancy studies with imperfect detection, *Methods in Ecology and Evolution*, **1**, 131–139.
- Hanski, I. (1998). Metapopulation dynamics, *Nature*, **396**, 41–49.
- Hill, M. F., Witman, J. D. and Caswell, H. (2004). Markov chain analysis of succession in a rocky subtidal community, *American Naturalist*, **164**, E46–61.
- Kendall, W. L. and White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy, *Journal of Applied Ecology*, **46**, 1182–1188.
- Kendall, W. L., Nichols, J. D. and Hines, J. E. (1997). Estimating temporary emigration using capture-recapture data with Pollock’s robust design, *Ecology*, **78**, 563–578.
- Kéry, M. and Royle, J. A. (2016). *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS, Volume 1: Prelude and Static Models*, Academic Press, Massachusetts.
- Kéry, M. and Schaub, M. (2012). *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective*, Academic Press, Massachusetts.
- Knape, J. and Korner-Nievergelt, F. (2015). Estimates from non-replicated population surveys rely on critical assumptions, *Methods in Ecology and Evolution*, **6**, 298–306.
- Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS — A Bayesian modelling framework: Concepts, structure, and extensibility, *Statistics and Computing*, **10**, 325–337.

- MacKenzie, D. I. and Royle, J. A. (2005). Designing occupancy studies: General advice and allocating survey effort, *Journal of Applied Ecology*, **42**, 1105–1114.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one, *Ecology*, **83**, 2248–2255.
- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G. and Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly, *Ecology*, **84**, 2200–2207.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L. and Hines, J. E. (2006). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*, Academic Press, Massachusetts.
- MacKenzie, D. I., Nichols, J. D., Seamans, M. E. and Gutiérrez, R. J. (2009). Modeling species occurrence dynamics with multiple states and imperfect detection, *Ecology*, **90**, 823–835.
- MacKenzie, D. I., Bailey, L. L., Hines, J. E. and Nichols, J. D. (2011). An integrated model of habitat and species occurrence dynamics, *Methods in Ecology and Evolution*, **2**, 612–622.
- Martin, J., McIntyre, C. L., Hines, J. E., Nichols, J. D., Schmutz, J. A. and MacCluskie, M. C. (2009). Dynamic multistate site occupancy models to evaluate hypotheses relevant to conservation of Golden Eagles in Denali National Park, Alaska, *Biological Conservation*, **142**, 2726–2731.
- Martin, J., Chamaillé-Jammes, S., Nichols, J. D., Fritz, H., Hines, J. E., Fonnesebeck, C. J., MacKenzie, D. I. and Bailey, L. L. (2010). Simultaneous modeling of habitat suitability, occupancy, and relative abundance: African elephants in Zimbabwe, *Ecological Applications*, **20**, 1173–1182.
- McClintock, B. T., Bailey, L. L., Pollock, K. H. and Simons, T. R. (2010a). Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections, *Ecology*, **91**, 2446–2454.
- McClintock, B. T., Nichols, J. D., Bailey, L. L., MacKenzie, D. I., Kendall, W. L. and Franklin, A. B. (2010b). Seeking a second opinion: Uncertainty in disease ecology, *Ecology Letters*, **13**, 659–674.
- McKann, P. C., Gray, B. R. and Thogmartin, W. E. (2013). Small sample bias in dynamic occupancy models, *Journal of Wildlife Management*, **77**, 172–180.
- Miller, D. A., Nichols, J. D., McClintock, B. T., Grant, E. H. C. and Bailey, L. L. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification, *Ecology*, **92**, 1422–1428.
- Miller, D. A. W., Brehme, C. S., Hines, J. E., Nichols, J. D. and Fisher, R. N. (2012). Joint estimation of habitat dynamics and species interactions: Disturbance reduces co-occurrence of non-native predators with an endangered toad, *Journal of Animal Ecology*, **81**, 1288–1297.
- Miller, D. A. W., Nichols, J. D., Gude, J. A., Rich, L. N., Podruzny, K. M., Hines, J. E. and Mitchell, M. S. (2013). Determining occurrence dynamics when false positives occur: Estimating the range dynamics of wolves from public survey data, *PLoS ONE*, **8**, e65808.
- Moilanen, A. (2002). Implications of empirical data quality to metapopulation model parameter estimation and application, *Oikos*, **96**, 516–530.
- Moreno, M. and Lele, S. R. (2010). Improved estimation of site occupancy using penalized likelihood, *Ecology*, **91**, 341–346.
- Nichols, J. D., Hines, J. E., MacKenzie, D. I., Seamans, M. and Gutiérrez, R. J. (2007). Occupancy estimation and modeling with multiple states and state uncertainty, *Ecology*, **88**, 1395–1400.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)* (eds. K. Hornik, F. Leisch and A. Zeileis), Technische Universität Wien, Austria.
- Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture, *Journal of*

- Wildlife Management*, **46**, 752–757.
- Pradel, R. (2005). Multievent: An extension of multistate capture-recapture models to uncertain states, *Biometrics*, **61**, 442–447.
- Risk, B. B., de Valpine, P. and Beissinger, S. R. (2011). A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails, *Ecology*, **92**, 462–474.
- Royle, J. A. and Dorazio, R. M. (2008). *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*, Academic Press, New York.
- Royle, J. A. and Kéry, M. (2007). A Bayesian state-space formulation of dynamic occupancy models, *Ecology*, **88**, 1813–1823.
- Royle, J. A. and Link, W. A. (2005). A general class of multinomial mixture models for anuran calling survey data, *Ecology*, **86**, 2505–2512.
- Royle, J. A. and Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors, *Ecology*, **87**, 835–841.
- Spencer, M. and Susko, E. (2005). Continuous-time Markov models for species interactions, *Ecology*, **86**, 3272–3278.
- Sutherland, C. S., Elston, D. A. and Lambin, X. (2014). A demographic, spatially explicit patch occupancy model of metapopulation dynamics and persistence, *Ecology*, **95**, 3149–3160.
- Tanner, J. E., Hughes, T. P. and Connell, J. H. (1994). Species coexistence, keystone species, and succession: A sensitivity analysis, *Ecology*, **75**, 2204–2219.
- Tsujino, M., Hori, M., Okuda, T., Nakaoka, M., Yamamoto, T. and Noda, T. (2010). Distance decay of community dynamics in rocky intertidal sessile assemblages evaluated by transition matrix models, *Population Ecology*, **52**, 171–180.
- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K. and Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates, *Ecological Applications*, **13**, 1790–1801.
- Welsh, A. H., Lindenmayer, D. B. and Donnelly, C. F. (2013). Fitting and interpreting occupancy models, *PLoS ONE*, **8**, e52015.
- White, G. C. and Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals, *Bird Study*, **46**, S120–S139.
- Wootton, J. T. (2001). Prediction in complex communities: Analysis of empirically derived Markov models, *Ecology*, **82**, 580–598.
- Yackulic, C. B., Reid, J., Davis, R., Hines, J. E., Nichols, J. D. and Forsman, E. (2012). Neighborhood and habitat effects on vital rates: Expansion of the Barred Owl in the Oregon Coast Ranges, *Ecology*, **93**, 1953–1966.
- Yackulic, C. B., Reid, J., Nichols, J. D., Hines, J. E., Davis, R. and Forsman, E. (2014). The roles of competition and habitat in the dynamics of populations and species distributions, *Ecology*, **95**, 265–279.
- Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction Using R*, CRC Press, New York.

## Dynamic Site Occupancy Models: Statistical Inference of Site Occupancy Dynamics Accounting for State Uncertainty

Keiichi Fukaya

The Institute of Statistical Mathematics

In various ecological surveys of site occupancy, observation of occupying state is typically uncertain because of classification error. Since ignoring such a state uncertainty in general induces bias in the inference of the occupying state and the ecological processes of occupancy dynamics, classification errors need to be taken into account properly in the processes of data collection and analyses. This paper outlines the dynamic site occupancy model as a statistical model for the inference of site occupancy dynamics that account for classification error and the multistate dynamic site occupancy model as a generalization of the dynamic site occupancy model, in terms of motivations, required census design and model formulation. As an example of the multistate dynamic site occupancy model, which can potentially be applied to various ecological problems, I also present a statistical model for community dynamics that accounts for an observation error that can occur in the observation of sessile organisms and note some merits of using a dynamic site occupancy modeling framework in the inference of sessile community dynamics.