Vol.2015-MPS-103 No.35
Vol.2015-BIO-42 No.35
2015/6/24

MCMCサンプリングに基づく海洋環境因子ネットワーク解析

岡西 孝真[1,a)]　加藤 有己[1]　藤村 弘行[2]　中野 義勝[3]　須田 彰一郎[2]　曽根 秀子[4]　藤渕 航[1]

概要：サンゴ礁域は生物多様性の高い領域であるが，近年の環境変化によりその生育環境も大きく影響を受けていることが知られている．そのため，生態系の正確な把握とともに，その理解に基づくリスク変動予測を行うことが重要となる．本稿では，サンゴ礁域における経時的環境情報測定データを用いて，マルコフ連鎖モンテカルロ（MCMC）サンプリングに基づき，ベイジアンネットワークを仮定した環境因子ネットワークの推定を行う．これにより，環境指標となる重要な情報の選別が可能となり，環境リスク予測モデルの発展に寄与することが期待される．

1. はじめに

サンゴには褐虫藻と呼ばれる植物性プランクトンが共生している．この褐中藻の光合成によって生産される有機物は，サンゴの成長を助け，また他の微生物や甲殻類，サンゴを隠れ家とする魚類などが集まり，多様な生態系を形成していることが知られている．しかし近年の環境変化に伴い，その生態系は大きな影響を受けている．生態系に有害な影響を及ぼす恐れを環境リスクという．この生態系を健全に維持する為には，生態系の正確な把握とともに，他の様々な因子がどのように影響を及ぼしているのか理解する必要があり，それに基づくリスク変動予測を行うことが重要とされる．

本稿では，沖縄のサンゴ礁域である，瀬底南の観測データ（水温，pH，光量など）を用いて，マルコフ連鎖モンテカルロ（以後 MCMC という）サンプリング [2] に基づき，ベイジアンネットワーク [3] を仮定した環境因子ネットワーク解析を行う．ベイジアンネットワークとは，ベイズ推論とグラフ構造が合わさったものであり，得られたデータからある事象が起こりうる確率分布を推定するために用いることが多い．その際，パラメータが増えると推定計算が困難になるため，多数パラメータを持つモデル推定には MCMC のようなサンプリングに基づく手法が有効である．本稿では，まずベイジアンネットワークに基づいた MCMC サンプリングついて紹介し，それにより得られた環境因子ネットワークの推定結果を用いて考察する．

[1] 京都大学 iPS 細胞研究所
Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606–8507, Japan
[2] 琉球大学 理学部
University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903–0213, Japan
[3] 琉球大学 熱帯生物圏研究センター
Tropical Biosphere Research Center, University of the Ryukyus, 3422 Sesoko, Motobu, Okinawa 905–0227, Japan
[4] 国立環境研究所 環境リスク研究センター
National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305–8506, Japan
[a)] t.okanishi@cira.kyoto-u.ac.jp

2. ベイジアンネットワークに基づいたMCMCサンプリング

2.1 環境因子ネットワーク

まず，環境因子ネットワークとしてベイジアンネットワークを仮定し，以下 [4], [5] に基づいた記述を行う．P 個の因子 $G_1, G_2, \ldots, G_P$ の相互作用について考える．ある因子 G_j が他の因子から影響を受ける相互作用を，β_{ij} をパラメータとする対数線形関数を用いて次のように表す:

$$E[\log(G_j)] = \sum_{i=1;i\neq j}^{P} I_{ij}\beta_{ij}\log(g_i).$$

ここで，I_{ij} は相互作用の有無を示す 2 値のパラメータで，$I_{ij}=1$ であれば因子 i は j に影響を与えており，$I_{ij}=0$ であれば影響を与えていないことを示す．I_{ij} を要素とする遷移行列を T，β_{ij} を要素とする行列を B で表すことにする．すなわち，$T=(I_{ij})$，$B=(\beta_{ij})$ である．

2.2 事前分布

対数線形モデルにおいて，パラメータは各因子 G_j の分布の分散 ${\sigma_j}^2$，相互作用を示すパラメータ β_{ij}，およびネットワークの形状を示す I_{ij} によって表される．なお，事前

Vol.2015-MPS-103 No.35
Vol.2015-BIO-42 No.35
2015/6/24

情報がある場合と無い場合で与え方が異なる．分散 ${\sigma_j}^2$ に関する事前分布を $h({\sigma_j}^2)$ とし，β_{ij} に関する事前分布は，G_i から G_j へ相互作用がある場合と無い場合に分けて，$h(\beta_{ij} \mid I_{ij}=1)$ および $h(\beta_{ij} \mid I_{ij}=0)$ で表す．また，I_{ij} に関する事前分布を $h(I_{ij})$ で表す．$h({\sigma_j}^2)$ は逆ガンマ分布により与え，2 つのパラメータ a_1、a_2 を用いて $\mathrm{invGamma}(\frac{a_1}{2}, \frac{a_2}{2})$ とする．$h(\beta_{ij} \mid I_{ij})$ は，平均 0，分散 $\sigma(\beta)^2(I_{ij}+(1-I_{ij})C)^2$ の正規分布に従うとする．C は定数であり，$I_{ij}=0$ のとき分散が十分小さくなるように定める．B の j 列目のベクトルから j 番目の要素自身を抜いたベクトルを $\boldsymbol{\beta_{.j}}$ により表すと，このベクトルは因子 G_j が他の因子から受ける相互作用のパラメータを並べたものである．C は対角要素 (i,i) を $\sigma(\beta)(I_{ij}+(1-I_{ij})C)$ とする対角行列であり，$\boldsymbol{I_{.j}}$ は T の j 列目のベクトルから j 番目の要素自身を抜いたベクトルである．I_{ij} の事前分布は

$$P(I_{ij}=1) = P_{ij},$$
$$P(I_{ij}=0) = 1-P_{ij}$$

となるベルヌーイ分布に従うとする．事前情報がない場合 $P_{ij}=0.5$ とし，事前に相互作用が予測される場合にはより高い数値を設定する．

2.3 事後分布

事後分布はベイズ推定より，尤度と事前分布の積で表されることが知られている．よって，前節で求めた事前分布を用いて事後分布の推定を行う．

$\boldsymbol{g}=(\boldsymbol{g_1},\ldots,\boldsymbol{g_P}) \in \mathbb{R}^{n\times P}$ を正規化し対数を取った因子データとする．また，因子 G_i に関するものを除く観測データを $\boldsymbol{g(i)}=(\boldsymbol{g_1},\boldsymbol{g_{i-1}},\boldsymbol{g_{i+1}},\ldots,\boldsymbol{g_P})$ と書く．G_j の対数は以下の正規分布に従うものとする:

$$\log(G_j) \sim \mathcal{N}\left(\sum_{i=1;i\neq j}^{P} \beta_{ij}\log(\boldsymbol{g_i}), {\sigma_i}^2\right).$$

ここで $\boldsymbol{S}=(\sigma_1,\sigma_2,...,\sigma_P)$ とすると，パラメータ T,B,S を持つネットワークにおいて $\log(\boldsymbol{g_j})$ の確率密度は，

$$f(\boldsymbol{g_j} \mid T,B,S)$$
$$= \frac{1}{\sqrt{2\pi{\sigma_j}^2}}\exp\left(-\frac{1}{2{\sigma_j}^2}\left[\log(\boldsymbol{g_j}) - \sum_{i=1;i\neq j}^{P}\beta_{ij}\log(\boldsymbol{g_i})\right]\right)$$

で表される．全ての j に対して上式を掛け合わせたものが尤度関数となる．よって事後分布は，

$$f(T,B,S \mid \boldsymbol{g})$$
$$= \frac{P(\boldsymbol{g} \mid T,B,S)h(T)h(B \mid T)h(S)}{\sum_{T,B,S} P(\boldsymbol{g} \mid T,B,S)h(T)h(B \mid T)h(S)}$$
$$\propto f(\boldsymbol{g} \mid T,B,S)h(T)h(B \mid T)h(S)$$
$$= \prod_{j=1}^{P} f(\boldsymbol{g_j} \mid \boldsymbol{I_{.j}},\boldsymbol{\beta_{.j}},\sigma_j)h(\sigma_j)h(\boldsymbol{\beta_{.j}} \mid \boldsymbol{I_{.j}})h(\boldsymbol{I_{.j}})$$

となる．それぞれの事前分布および尤度関数を代入することにより，${\sigma_j}^2, \boldsymbol{\beta_{.j}}, \boldsymbol{I_{ij}}$ に関する事後確率を求めることができる．

${\sigma_j}^2$ に関する条件つき事後確率は，

$${\sigma_j}^2 \cong f({\sigma_j}^2 \mid \boldsymbol{g(j)},\boldsymbol{\beta_{.j}},\boldsymbol{I_{.j}})$$
$$= \mathrm{invGamma}\left(\frac{n+a_1}{2}, \frac{|\boldsymbol{g_j}-\boldsymbol{g(j)}\boldsymbol{\beta_{.j}}|+a_2}{2}\right)$$

と表される．事前情報がない場合には a_1, a_2 は 0 と選択される．

$\boldsymbol{\beta_{.j}}$ に関する条件つき事後確率は，

$$\boldsymbol{\beta_{.j}} \cong f(\boldsymbol{\beta_{.j}} \mid \boldsymbol{g},\boldsymbol{\beta_{.j}},\boldsymbol{I_{.j}})$$
$$= \mathcal{N}_{\mathcal{P}}(A{\sigma_j}^{-2}\boldsymbol{g(j)}^t\boldsymbol{g_j}, A),$$
$$A^{-1} = {\sigma_j}^{-2}\boldsymbol{g(j)}^t\boldsymbol{g_j} + (CRC)^{-1}$$

と表される．行列 R は，$g(j)^t g_j$，あるいは $P-1$ 次元単位行列 $I(P-1)$ と選ぶことができる．

I_{ij} に関する条件つき事後確率は，

$$I_{ij} \cong f(I_{ij} \mid \boldsymbol{g},\boldsymbol{\beta_{.j}},\sigma_j,T/I_{ij})$$
$$= f(I_{ij} \mid \boldsymbol{\beta_{.j}},\sigma_j,T/I_{ij})$$

と表される．この分布より順にサンプリングすることにより連結行列 T が得られる．T/I_{ij} は I_{ij} 除く T の要素である．このように自身を除くのは，パス $I_{ij}=1$ が追加されることにより閉路が生成されるのを防ぐためである．サンプリング手順に以下の処理を加える．I_{ij} は次のベルヌーイ分布によりサンプリングされる．

$$P(I_{ij}=1 \mid \beta_{.j},\sigma_j,T/I_{ij}) = \frac{a}{a+b}.$$

$I_{ij}=1$ が閉路となるとき $a=0,\ b=1$ とし，新たにパスが追加される確率を 0 とする．それ以外の場合には，以下の式によりパスの追加を判定する．

$$a = f(\boldsymbol{\beta_{.j}} \mid T/I_{ij},I_{ij}=1)f(\sigma_j \mid T/I_{ij},I_{ij}=1)f(I_{ij}=1).$$
$$b = f(\boldsymbol{\beta_{.j}} \mid T/I_{ij},I_{ij}=0)f(\sigma_j \mid T/I_{ij},I_{ij}=0)f(I_{ij}=0).$$

これらの手順により，全ての j に対して事後確率から 1 回のサンプルを取ることができる．

2.4 サンプリングアルゴリズム

上記で求めた事後確率からのサンプリング方法を用いて，ギブスサンプリング [1] を行う．T,B,S の事後確率は以下の式で表される:

$$f(T,B,S \mid \boldsymbol{g})$$
$$\propto f(\boldsymbol{g} \mid T,B,S)h(T)h(B \mid T)h(S)$$
$$= \prod_{j=1}^{P} f(\boldsymbol{g_j} \mid \boldsymbol{I_{.j}},\boldsymbol{\beta_{.j}},\sigma_j)h(\sigma_j)h(\boldsymbol{\beta_{.j}} \mid \boldsymbol{I_{.j}})h(I_{.j}).$$

Vol.2015-MPS-103 No.35
Vol.2015-BIO-42 No.35
2015/6/24

このように分解することにより，σ_j および $\boldsymbol{\beta_{\cdot j}}$ は前述の通り，

$$\begin{aligned}\sigma_j^2 &\cong f(\sigma_j^2 \mid \boldsymbol{g(j)}, \boldsymbol{\beta_{\cdot j}}, \boldsymbol{I_{\cdot j}}) \\ &= \mathrm{invGamma}\left(\frac{n+a_1}{2}, \frac{|\boldsymbol{g_j} - \boldsymbol{g(j)}\boldsymbol{\beta_{\cdot j}}| + a_2}{2}\right), \\ \boldsymbol{\beta_{\cdot j}} &\cong f(\boldsymbol{\beta_{\cdot j}} \mid \boldsymbol{g}, \boldsymbol{\beta_{\cdot j}}, \boldsymbol{I_{\cdot j}}) \\ &= \mathcal{N}_{\mathcal{P}}(A\sigma_j^{-2}\boldsymbol{g(j)}^t\boldsymbol{g_j}, A), \\ A^{-1} &= \sigma_j^{-2}\boldsymbol{g(j)}^t\boldsymbol{g_j} + (CRC)^{-1}\end{aligned}$$

によりサンプリングすることができる．$\boldsymbol{I_{\cdot j}}$ をサンプリングする場合は，

$$P(I_{ij} = 1 \mid \boldsymbol{\beta_{\cdot j}}, \sigma_j, T/I_{ij}) = \frac{a}{a+b}$$

を用いるが，閉路の出現を避けるために因子 G_j に関する接続以外も調べる必要がある．

3. 実験

3.1 経時的環境情報測定データ

ここでは，瀬底南にあるサンゴ礁域について，2014 年 10 月から 2015 年 2 月の約 4 か月にわたって経時的環境情報の測定を行った．実際にネットワーク解析に用いた環境因子は，水温 (Temp: °C)，塩分 (Sal: psu)，pH (pH: pH unit)，溶存酸素 (ODO: % sat)，濁度 (Turbidity: NTU)，クロロフィル量 (Chlorophy11: μg/L)，シアノバクテリア色素 (BGA-PC: μg/L)，蛍光性溶存有機物 (fDOM: QSU)，水深 (Depth: m)，水中光量 (μmol/m^2/s)，気圧 (Pressure: mbar)，湿度 (RH: %)，風速の東西方向成分 (Wind x: m/s)，風速の南北方向成分 (Wind y: m/s)，雨量 (Rain: mm) の 15 個である．このうち，気圧，湿度，風速，雨量の 4 因子は琉球大学の熱帯生物圏研究センター瀬底研究施設内に設置してある気象ステーションのデータを用いた．海域データは多項目水質計 (YSI, exo2) および光量子計 (JFE Advantech, ALW-CMP) を用いて，2014 年 10 月から 2015 年 2 月まで，各月ごとに 15 分間隔で約 2 週間連続的に取得した．

3.2 計算機実験

2.3 節で定義した因子ベクトル $\boldsymbol{g}$ として上記経時的環境情報測定データを用い，8 並列の計算機システムを用いて MCMC サンプリングを行った．なお，ここでは収束のしやすさの観点から Replica Exchange 法を用い，200,000 回のサンプリングを行った．図 1 に瀬底南に対する環境ネットワークの推定結果を示す．

4. 考察

瀬底南の環境データから，水中光量と溶存酸素に高い相関が見られた．これは，光量の増加に伴ってサンゴ礁環境に生息する様々な生物のうち，植物の光合成が活発となり，海水中の酸素が増えた事を示している．造礁サンゴは体内に褐虫藻と呼ばれる藻類を共生させている．観測地点のサンゴの被度は高く，主に光強度に応じた褐虫藻の光合成を示しているものだと考えられる．また，一般に光合成では酸素が放出されるのと同時に海水中の二酸化炭素が消費され，pH が上昇する．実際に環境ネットワークからも，溶存酸素と pH の高い相関が見られ整合的である．ただし，瀬底南の 1 月と 2 月はこのような溶存酸素と pH の相関は見られたものの，その前提となる光量と溶存酸素の相関が見られなかった．溶存酸素は大気と海水がよく攪拌され平衡状態にあれば，酸素飽和度は 100% となる．瀬底南はサンゴ礁の緑の部分に砕波帯があり，ここで大気への逃散と海水への溶け込みがあるため，とくに北風の強い冬場は光合成による変動が小さくなる傾向にある．このため，光量との相関が相対的に低下し，溶存酸素と pH の相関だけが現れたものだと考えられる．蛍光性溶存有機物と溶存酸素の間には負の相関があり，溶存酸素が減少する夜間に溶存有機物が水中に増加することを示している．これは，日中に光合成で生成した有機物をサンゴ礁の生物群集が夜間に呼吸によって分解し，その一部が溶存有機物として放出されている事を示しているのではないかと考えられる．物理的な環境因子同士のネットワークでは，水温が上昇すると，湿度が高くなる相関関係が見られた．さらに，これらは風速とも関係しており，湿度が低下している時は南北成分の風が卓越する関係が示された．解析したデータは主に冬のデータであり，乾いた北よりの風を表しているものと考えられる．浅海域の海水の流れは風向と風速に大きく依存するため，風に起因する海流が地点の水質に影響することを考えていたが，環境ネットワークからそのような関係は見られなかった．

このように説明が可能な因子だけでなく，例えば気圧が溶存酸素や pH と相関を示すことが一部のネットワークで見受けられるなど，説明がつかない相関関係も多数あり，当該月の海況や天気など今後多方面からの見当が必要である．また，浅海域では多量の降雨で一時的に塩分の低下が観測される事があるが，今回のデータにおいて，雨量はどの因子とも相関関係が見られなかった．これは多量の降雨が期待される梅雨の時期が今回の解析に含まれていないためであり，今後の継続した観測の結果が待たれる．

5. おわりに

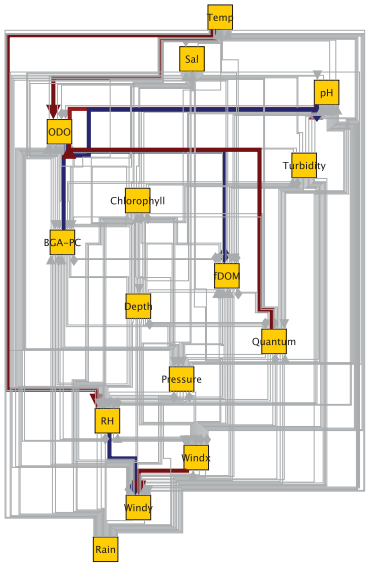
本研究ではベイジアンネットワークによるモデル推定として MCMC サンプリングを行い，その結果から環境因子同士の依存関係をみることができた．これは生態系を正確に知るための一部に過ぎないが，環境が生態系に与える影響を情報解析により把握する道を開く一助となろう．今後より柔軟なモデルを考慮することで，環境リスク予想モデルへの発展に寄与することが期待される．

謝辞 本研究は一部，CREST（海洋生物多様性および生態系の保全・再生に資する基盤技術の創出）からの助成金を受けている．

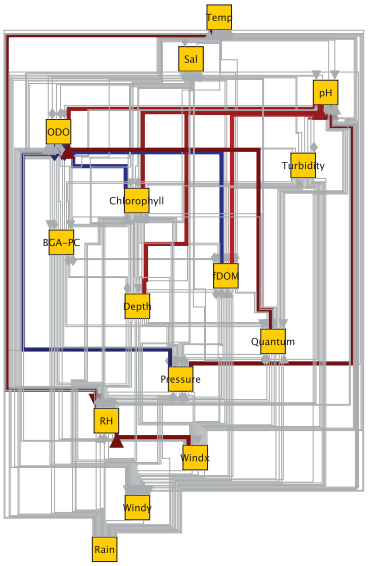
参考文献

[1] George E.I. and McCulloch R.E.: Variable selection via Gibbs sampling, *J. Am. Stat. Associat.*, 88(423):881–889 (1993).
[2] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21:1087–1092 (1953).
[3] Pearl J.: Bayesian networks: A model of self-activated memory for evidential reasoning, *Proc. 7th Conference of the Cognitive Science Society*, 329–334 (1985).
[4] Toyoshiba H., Yamanaka T., Sone H., Parham F.M., Walker N.J., Martinez J. and Portier C.J.: Gene interaction network suggests dioxin induces a significant linkage between aryl hydrocarbon receptor and retinoic acid receptor beta, *Environ. Health Perspect.*, 112(12):1217–1224 (2004).
[5] Toyoshiba H., Portier C.J., Sone H., Parham F.M., Irwin R.D. and Boorman G.A.: Inference for Bayesian network via Gibbs sampling, *preprint*.

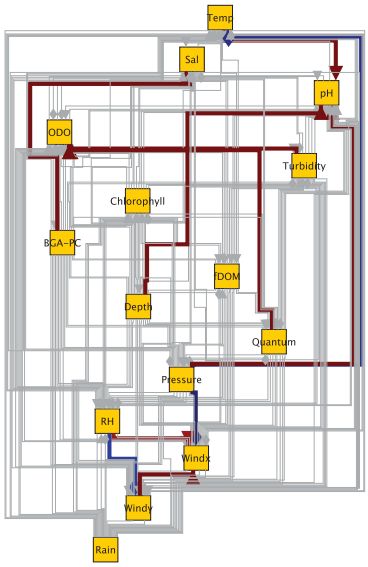
Vol.2015-MPS-103 No.35
Vol.2015-BIO-42 No.35
2015/6/24



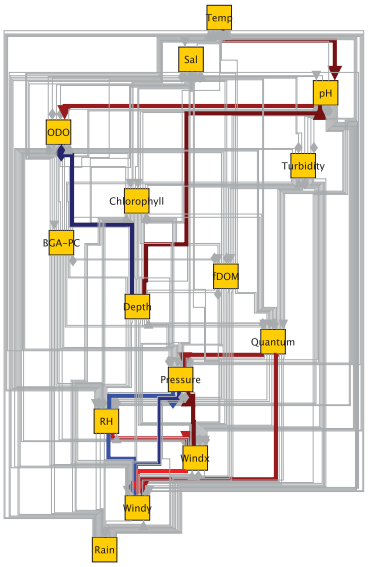
(a) 2014 年 10 月 16 日–2014 年 11 月 6 日



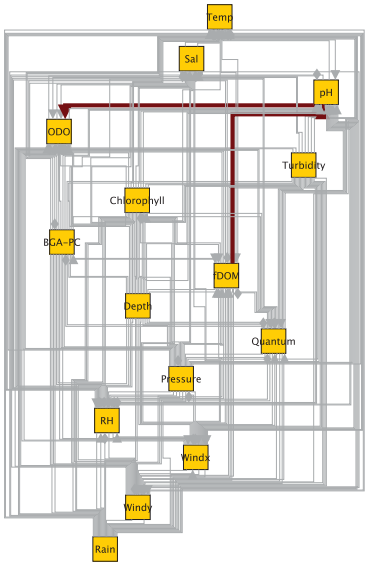
(b) 2014 年 11 月 17 日–2014 年 12 月 3 日



(c) 2014 年 12 月 24 日–2014 年 1 月 6 日



(d) 2015 年 1 月 20 日–2015 年 2 月 3 日



(e) 2015 年 2 月 10 日–2015 年 2 月 23 日

図 1 瀬底南に対する環境因子ネットワーク解析図．なお，図中の赤線は正の相関を示し，青線は負の相関を示している．