

## 科学技術計算システム活用事例紹介

科学技術計算システムは、スーパーコンピュータ（いわゆるスパコン）をメインで提供しています。スパコンというとFortranやCなどのプログラムで大規模な数値シミュレーションを動かすといったイメージが強いかもしれませんが、オープンソースのRやPythonを使うこともでき、以前に比べて利用の幅が広がっています。今回ご紹介するのは、ご自身の研究を進めるためにスパコンをうまく活用いただいた事例です。モデル高度化やデータ量の増大によってPCの計算時間が増えて困っている方はぜひスパコンの利用をご検討ください。

### 私がスーパーコンピューターを利用するまで

山浦悠一 森林総合研究所・四国支所

令和元年になってから、科学技術計算システムのスーパーコンピューター（以下、SCSスパコン）を利用して、論文を二本出版することができました(1, 2)。特に前者の論文は、野外調査をコンピュータ上で模擬的に3,840回実施するというもので、スパコンがなければ到底遂行できませんでした(図1)。私は未だにコンピュータに疎い人間です。今回機会をいただきましたので、私がどのようにしてスパコンを使ってこのような研究を行なうことになったのか、振り返ってみたいと思います。

ここ10年、私は野外調査で収集したデータを解析するために、階層モデルという統計的手法の適用に取り組んでいます。野外調査では、あらかじめ設定した回数調査を実施できないことがまあり、これは特に長期・広域的な調査で顕著になります。また、一般市民が収集した「市民データ」と呼ばれる野外調査データを解析する機会も増えました。市民データの調査地点は市街地の近くに多く、その分布が空間的に不均一なのが普通です。こうした「不規則な(irregular)」データを従来の方法で解析すると、例えば「原始的な自然を愛好する生物は大都市に多い」といった誤った結論を導いてしまいます。そこで、階層モデルを用いて、野外調査を実施した上でデータが得られるという過程を考慮したい、と考えました。

SCSスパコンの利用を考え始めたのは5年ほど前です。その頃私は、上記の問題意識のもと、階層モデルを用いた生物群集の解析手法（階層群集モデル）を考案し、その性能をコンピュータ上で実験的に調査していました(3)。階層モデルの短所は構造が複雑で、計算（解析）を実行するのに時間がかかってしまうことです。階層群集モデルのような複雑なモデルではベイズ推定が一般に用いられます。上記の論文ではJust Another Gibbs Sampler (JAGS) というベイズ推定用のプログラムを統計プラットフォームのRから呼び出しましたが、一つの解析を実行するのに3時間 (!) ほどかかっていました。また、この研究では1,080回の模擬的な調査を12回の解析コードに分割して職場のデスクトップPCで行っていました。しかし、予算をはたいて購入した

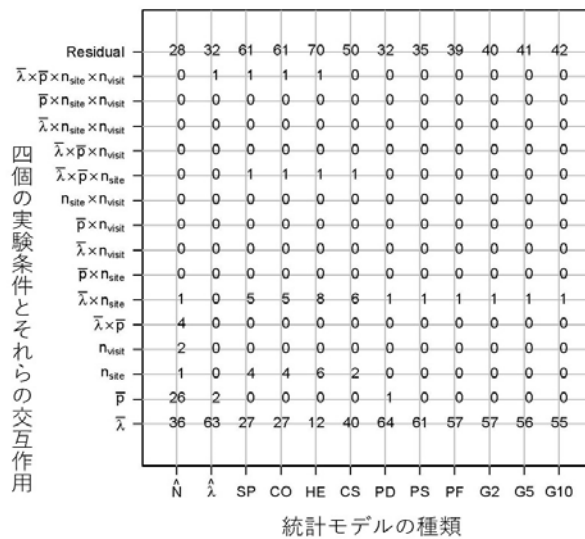


図1. 統計モデルの推定精度に実験条件が及ぼす影響

コンピュータ上に仮定の生物群集を生成して野外調査を行ないました(1)。野外調査は、群集構成種の平均個体数  $\bar{\lambda}$ 、平均発見率  $\bar{p}$ 、調査地点数  $n_{\text{site}}$ 、訪問回数  $n_{\text{visit}}$  が異なる条件下で行ない、各種の各調査地点における個体数を12個の統計モデルで推定し、その精度を求めました。統計モデル別（横軸）に、4個の実験条件とその交互作用（縦軸）を要因とした分散分析を行ない、各要因で説明された精度のばらつき（割合）を示したのがこの図です。モデルにより、精度を左右する実験条件が異なることが分かります。

デスクトップPCでも解析コードを6個回すとCPUの使用率が100%に達し、文書作成すら困難な有様でした。フォルダの中には、解析が終わるのを待つコードが常に待機していました。自宅のノートPCを走らせることもできましたが、冷却ファンがフル作動し、妻が「大丈夫？」と心配します。こうして、職場のデスクトップPCだけで粛々と作業を続けることになりました。解析の追加ややり直しなどもあり、実験開始から論文受理まで、結局3年弱かかりました。この類の研究を職場のデスクトップPC一台のみで行なうことの限界は明白でした。



ニュースはWebで発信しています。

AFFRIT Portal » 農林水産研究情報総合センターニュース  
<https://itcweb.cc.affrc.go.jp/affrit/inside/publications/affrit-news/start>

### トピックス:

AFFRIT Portal  
よくある質問  
(FAQ)

AFFRIT PortalのFAQでは、過去のお問い合わせへの回答をまとめて掲載しています。サービスの利用や手続き等でお困りの際にご参照ください。

<https://itcweb.cc.affrc.go.jp/affrit/faq>

この研究の後の大きな出来事の一つは、JAGSの並列計算ができるようになったことです(jagsUIというRのパッケージでは、コマンドでparallel = TRUEと打てばいいだけ！)。CPUの複数のコアにベイズ推定の計算を個別に割り当てることで、トータルの計算時間を短縮することができます。jagsUIの例題で試したところ、並列計算すると計算時間がおおよそ半減していました(ただしCPUの使用率は倍増していました)。

もう一つの大きな出来事は、SCSスパコンを使えるようになったことです。科学技術計算システムの講習会に参加してシステムの基本を勉強し、スパコンにJAGSをインストールしました(苦労しました)。時間を見つけながらの作業で、AFFRIT担当者のお力添えいただきながら、講習からここまで6か月かかりました。これにより、スパコン内でJAGSをRから並列計算し、さらにそのための解析コードを複数個回すことができます。現在の職場のデスクトップPCでは最大4個の解析コードしか回せないの、スパコンを使用できるメリットは大きいです。16個の解析

コードを実行しても、スパコンの使用状況に応じて解析コードは順繰りに実施され、私は待っているだけでいいのです。さらに大事なことは、スパコンでの解析は、自宅や出張中でもできるということです。子供の世話や野外調査で職場に不在がちな私にとって、このメリットはかけがえのないものです。関連ソフトウェア(JAGSなど)のインストール、そして解析コマンド(Rのスク립トなど)をLinux言語で操作するハードルを越えれば——SCSスパコンは悩める計算機集約的なユーザーの大きな助けになってくれると思います。



今回利用されたスパコン(分散並列型クラスタシステム)

1. Y. Yamaura, F. G. Blanchet, M. Higa, Analyzing community structure subject to incomplete sampling: hierarchical community model vs canonical ordinations. *Ecology* 100, e02759 (2019).
2. Y. Yamaura et al., A spatially explicit empirical model of structural development processes in natural forests based on climate and topography. *Conserv. Biol.*, (2019).
3. Y. Yamaura, M. Kery, J. A. Royle, Study of biological communities subject to imperfect detection: bias and precision of community *N*-mixture abundance models in small-sample situations. *Ecol. Res.* 31, 289-305 (2016).

## 科学技術計算システム(SCS)サービスのご案内

### ■SCSの提供サービス

- スーパーコンピュータ・分散並列型クラスタシステム(TypeA)  
・大規模共有メモリ型システム(TypeB)
- 科学技術計算アプリケーション  
(SAS, R, Maple, MOE, ANSYS 他)
- 農林水産基礎数値データベース(NDB)

◎利用案内サイト(まずはこちらをご確認ください)  
AFFRIT Portal > 科学技術計算システム  
(運用情報、利用の手引き、FAQ、研究支援サービスなど)  
<https://itcweb.cc.affrc.go.jp/affrit/inside/scs2016/>

◎メールお問い合わせ先  
情報総合センター総合受付アドレス  
request@affrc.go.jp



### <お知らせ>

○夏季ご利用をお勧めします！  
9月末までの夏季期間は、計算機リソースに余裕があります。大規模な計算、長時間の計算は、この期間を利用されることをお勧めします。  
なお、下記のメンテナンス時には計算が停止しますので、ご注意ください。

○SCS定期メンテナンス(予告)  
9月27日(金) 15時 ~ 10月1日(火) 13時に、法定電気設備点検(停電)と定期メンテナンスにともなうSCSの運用停止を予定しています。  
計算機、アプリケーション、NDBすべてのサービスが停止となります。(実行中のジョブは停止、キューは削除します)  
停電に伴うセンター全体のサービス停止の詳細は近日中にお知らせいたします。

## 文献情報DBセミナー開催のご案内 / 2019年9月12日(木)

第1部(10:00-)「BIOSIS」「CAB」「MEDLINE」  
第2部(13:00-)「Web of Science」「Current Contents Connect」、文献ツール「EndNote Basic」  
場所:筑波産学連携支援センター 情報通信共同利用館(電農館)3階 セミナー室  
オンライン受講有り！  
お申し込みはWebで！ [https://ml-wiki.sys.affrc.go.jp/kyoudouriyou\\_sisetu/seminar](https://ml-wiki.sys.affrc.go.jp/kyoudouriyou_sisetu/seminar)

受講者の約8割が  
“セミナーは有意義だった”と回答

受講者の声:  
年々、検索のやり方も進化しているように感じられます。  
チャットで参加できる点はとても良かったです(オンライン)。



ニュースはWebで  
発信しています。

AFFRIT Portal >> 農林水産研究情報総合センターニュース  
<https://itcweb.cc.affrc.go.jp/affrit/inside/publications/affrit-news/start>