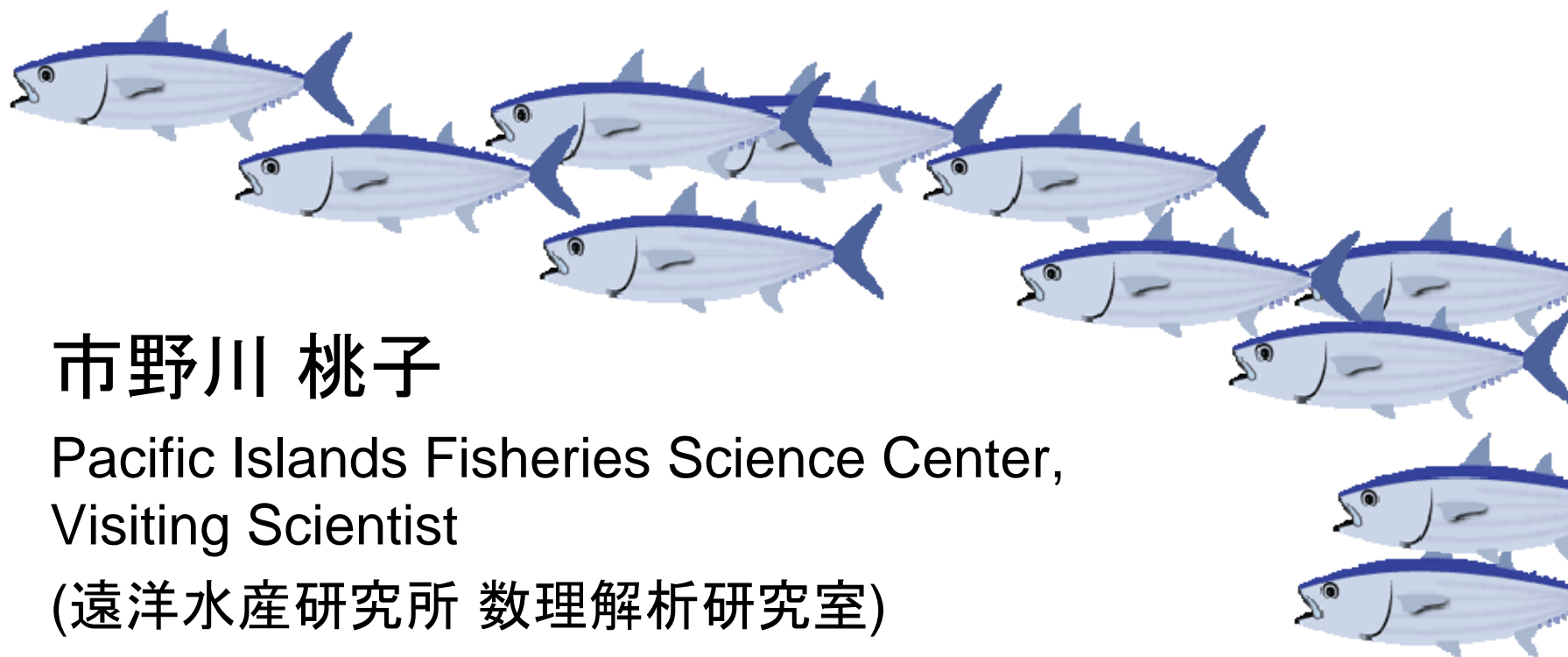




# Rと水産資源学

## ～Rを使って漁業データ解析～



市野川 桃子

Pacific Islands Fisheries Science Center,  
Visiting Scientist

(遠洋水産研究所 数理解析研究室)

# 自己紹介

R0.63

1999年～2004年3月 大学  
院修士・博士課程  
プランクトン食物網の研究

R1.90

2004年4月～  
遠洋水産研究所(任期つき  
の派遣社員)

2006年2月 博士号取得

2007年9月  
NOAA/NMFS(アメリカ合衆  
国大気海洋局), Pacific  
Islands Fisheries Science  
Centerからの招聘研究員と  
して, 改めて, 遠洋水産研  
究所で働く

R2.4.1

現在

- Rで北太平洋クロマグロの将来予測
- 延縄CPUE標準化
  - ◻ 北太平洋ビンナガ, クロマグロ,  
マカジキ, メカジキ
- ビンナガ通常タグ実験結果の解析
- 資源評価
  - ◻ 北太平洋クロマグロ, VPA/統合  
モデル(SS2)
  - ◻ 中西部太平洋メバチ・キハダ, 統  
合モデル(MFCL)
- オペレーティングモデルを用いた資  
源評価モデル評価の試み(途中)



# 今日の話の流れ

1. はじめに: Rとは？
2. 漁業データ解析とR
  - Rを使ってできることを目的別に概観
3. Rの使用例
  - ① CPUE標準化の簡単なシミュレーション(プログラミングの実例も含めて)
  - ② 複数の資源評価モデルを評価するためのオペレーティングモデルの紹介
4. まとめ

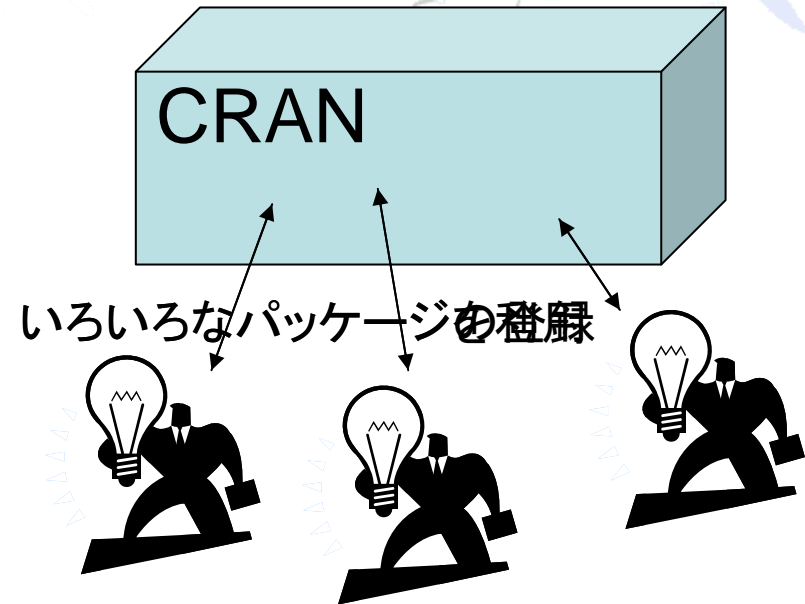


## はじめに: Rとは？

- 統計言語であるSの思想に基づいて開発されたフリーのソフトウェア
- WWW上のRのためのサーバー-CRAN (Comprehensive R Archive Network)から誰でも入手可能
- 多様なプラットフォームに対応
  - Unix系OS、Mac OS X, Windows
- 豊富なパッケージ群
  - パッケージ: 共通の目的を達成するための関数群

# Rの「パッケージ」システム

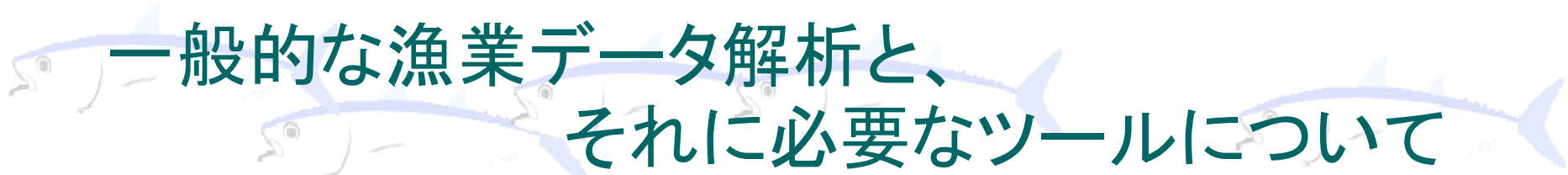
- 一定のフォーマットを満たすような自作のパッケージをCRANに登録し、皆に使ってもらうことが可能  
→ 世界中から集まる便利なパッケージをいつでも利用できる



- 漁業資源解析に特化したパッケージもある: FLR project (a framework for fisheries modelling and management strategy simulation in R)
- パッケージのダウンロード&インストール

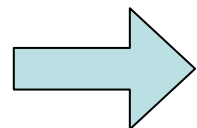
R> install.packages("パッケージ名") または、メニューバー→「パッケージ」→「パッケージのインストール」

R> library("パッケージ名") または、メニューバー→「パッケージ」→「パッケージの呼び出し」



## 一般的な漁業データ解析と、 それに必要なツールについて

- 漁獲データの可視化, グラフ作成
  - Excelなどのスプレッドシートソフト
- 地図データの解析
  - ArcView, GMTなど
- CPUE標準化等、統計的解析
  - 統計ソフト。SAS, S-Plus..
- パラメータ推定
  - Excelのソルバー、AD model builder



ほとんどのソフトは有料！

## Rはタダでできます①

### データの可視化, グラフ作成

- 単純なグラフの作成はもちろん、様々なパッケージを使うことで縦横無尽なグラフ作成
- どのようなグラフを作れるのかは、demo関数で確かめられる

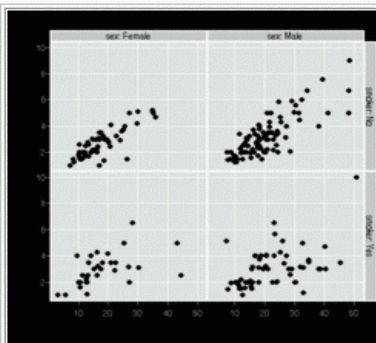
```
R> demo("graphics")
```

- R graphical manuals というサイト (<http://bg9.imslab.co.jp/Rhelp>) のimage browserでは、12,994枚のRで作成されたグラフの画像が一覧できる

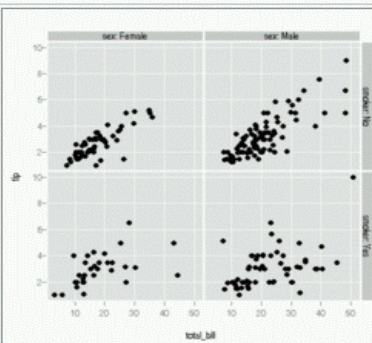
# <http://bg9.imslab.co.jp/Rhelp>より

Hit: 12994 Page: 813 1 2 3 4 5 6 7 8 9 10 11 ... Jump to

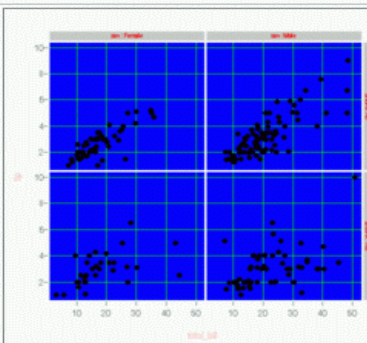
next >



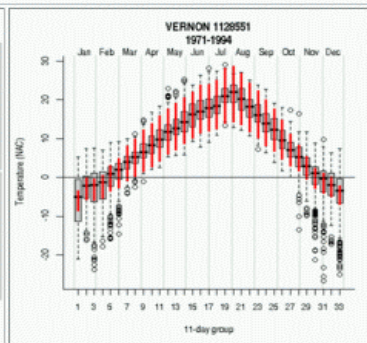
.build\_options



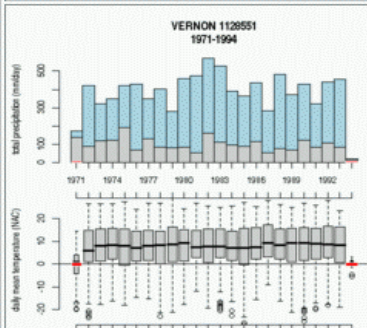
.build\_options



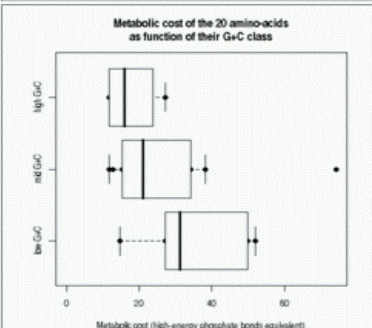
.build\_options



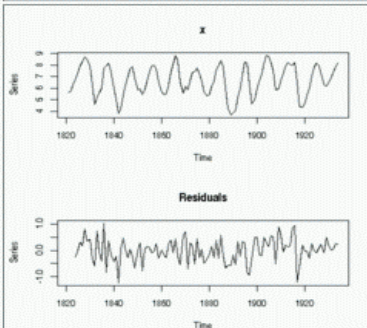
A1128551.DLY



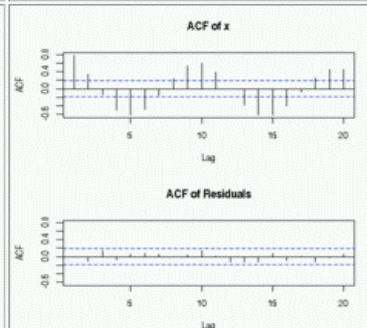
A1128551.DLY



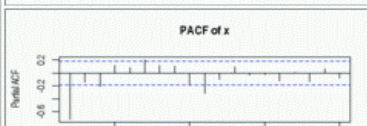
aacost



AAR



AAR

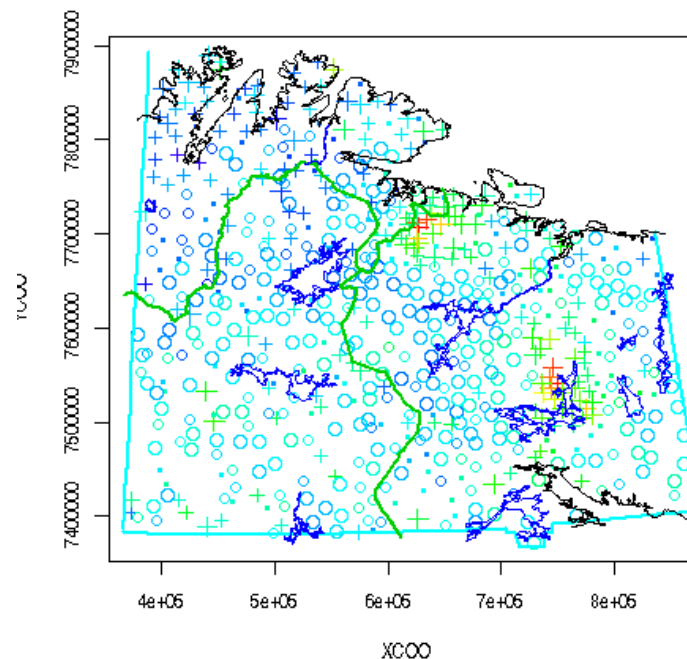
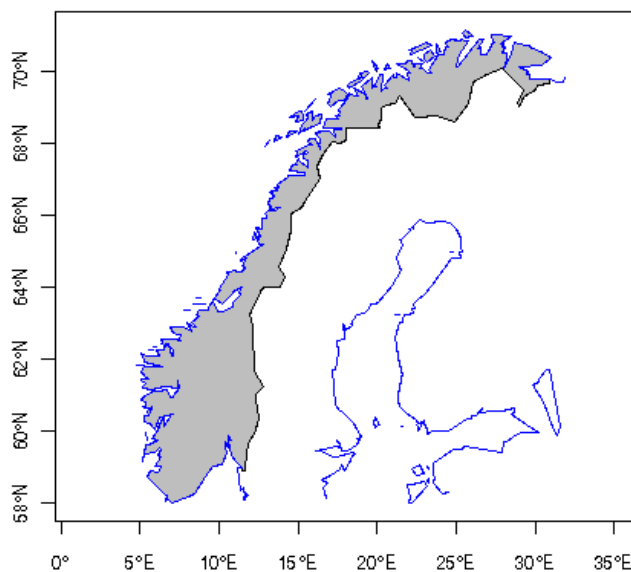




# Rはタダでできます②

## 地図データの解析 & 可視化

- maps, maptools, mapdataのパッケージを使って世界地図を描く → そのうえに漁場の分布などの絵をかく
- 例: <http://bg9.imslab.co.jp/Rhelp>より



# Rはタダでできます③

## CPUE標準化をはじめとした統計解析

- よく使われる関数たち
  - lm 線形モデル用関数
  - glm 一般化線形モデル
  - gam 一般化加法モデル(mgcvパッケージ内)
  - glmML, glmSQL, lme4 一般化線形混合モデル(追加パッケージ)
- 欠点もある: あんまり大きなデータを扱えない & (SASなどに比べて)遅い
  - 9万件の漁業データで、50年×3カ月の交互作用を考慮したglm→約41秒。生成される結果のオブジェクト164M。
  - あんまり遅いようならSASなどへの移行を
  - biglm というパッケージもある → このパッケージなら、上述の計算が3秒、結果オブジェクトが0.1M。
  - しかし、北太平洋全域の延縄縄操業データ(100万件以上)を一括して解析することはできない、。。

## Rはタダでできます④ パラメータ推定

- よく使われる関数たち

optim 最小化関数

optimize 一変数関数専用

constrOptim 線形不等式制約付きの最適化関数

nlm 汎用非線形最小化関数

- 欠点もある: あんまり大きなデータを扱えない & (AD model builder などに比べて)遅い

- 50くらいのパラメータを推定するタグデータ解析のときに optimを使ったが、3時間くらいかかるので、あきらめて AD model builderに移行したところ、3分以下で計算が終わった

- 10個以下くらいのパラメータ推定だったら問題ない?

# Rの使用例(1) 資源量指数の標準化に関する シミュレーションモデルをRで

知りたい資源の資源量のトレンドを知るための指数＝  
資源量指数

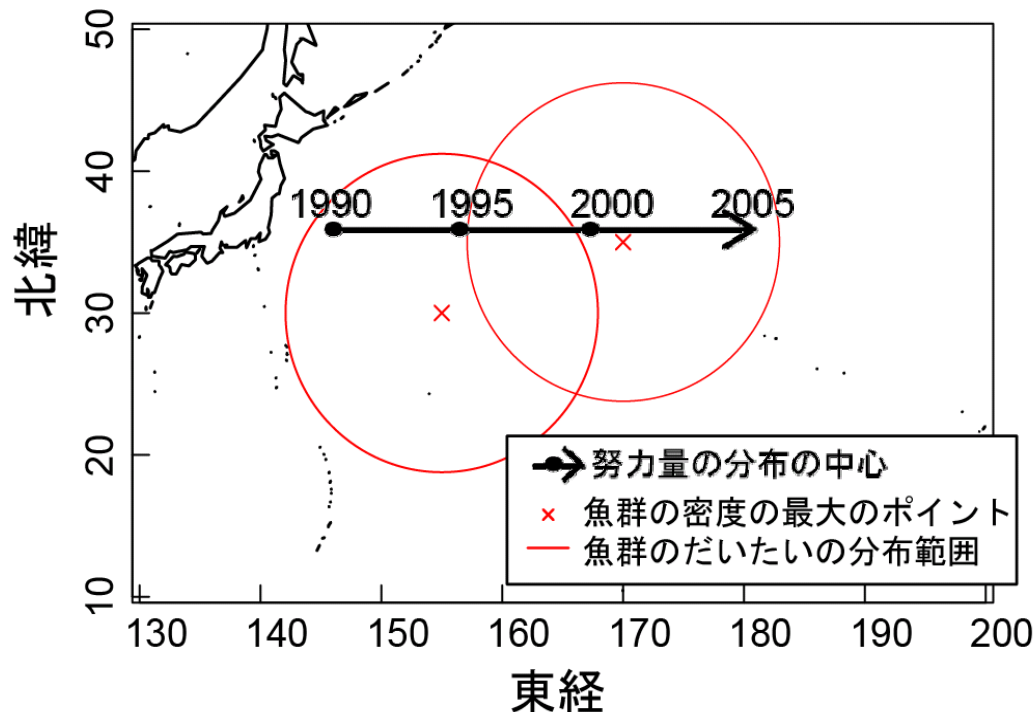
- 単位努力量あたりの漁獲尾数
- 漁業データはしばしば偏っている→CPUEに影響を与える要素を考慮して「標準化」する必要がある

[問題設定] 魚はある海域に集中しているが、それを漁獲する漁業は、それとは関係なく努力量分布をシフトさせる場合、正しい資源指数を推定することができるだろうか？

- 混獲の魚種などではよくある話
- ターゲットがシフトする場合にも同じような問題がしばしば見られる(北太平洋のマカジキ→メバチへ)

# シミュレーションモデルのシナリオ

- 北太平洋のある地点を中心に魚が分布(正規分布)→資源量は、1990年から2005年まで単調減少しているとする
- 漁業は1990年からスタートし、日本沿岸からだんだん沖に漁場がシフトする状況を考える
- 操業の中心場所から正規分布に従って漁場をランダムに選択し、近くにいる魚をすべて漁獲するとする



## 漁業データ

- 操業位置(緯度・経度)
- 操業年
- 操業あたりの漁獲尾数
- 一年で 1000 操業
- testdata.csv という形で保存



# シミュレーションから得られた漁獲量データを使って、正しい資源量トレンドを推定する

0. 努力量が時空間的に偏ったようなシミュレーションデータ (testdata.csv)

## 1. 漁業データの概観

- 地理情報のプロットなど

## 2. データの層化

- 概観したデータをもとに、データを作り変える

## 3. 標準化の実行

- glm関数を使って

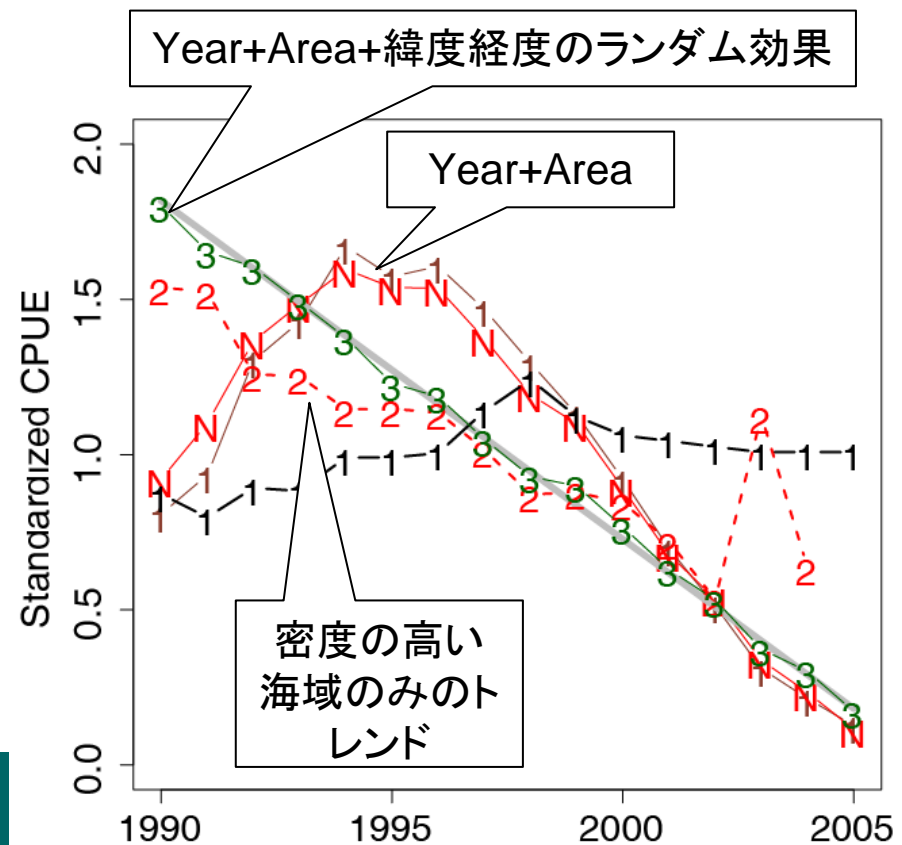
普通の標準化のプロセス

## 4. 結果の評価 (真の資源量との比較)

→ Rで行う例を紹介

# RによるCPUE標準化シミュレーション (まとめ)

- ちょっとでも油断すると, とんでもなく間違った結論を導くことがある
- GLMで示したのは極端な例なので, 実際は, 海区割りの方法 & 海域・年の相互作用を工夫することで, GLMでも正しい資源量は推定できるだろう



→正しい海区割り&相互作用の入れ方が重要!

# (参考)CPUEシミュレーションデータ 作成のためのRコード

```
yearrange <- 1990:2005
abundance <- seq(from=10000,to=1000,length=length(yearrange)) # 100匹を一群れとする群が7000~
  10000群ある
effort <- rep(1000,length(yearrange)) # 毎年1000操業
catch.position <- data.frame(lat=seq(from=35,to=35,length=length(yearrange)),
  lon=seq(from=145,to=180,length=length(yearrange))) # 操業の中心地の設定
catch <- data.frame(lon=0,lat=0,year=0,N=0) # 漁獲データはcatchというオブジェクトに格納
for(y in 1:length(yearrange)){
  fishdist <-
    data.frame(lon=c(rnorm(abundance[y]/2,sd=8,mean=155),rnorm(abundance[y]/2,sd=10,mean=17
    0)),
      lat=c(rnorm(abundance[y]/2,sd=5,mean=30),rnorm(abundance[y]/2,sd=5,mean=35)))
  # 魚の群の分布
  catch.tmp <- data.frame(lon=rnorm(effort[y],sd=4,mean=catch.position$lon[y]),
    lat=rnorm(effort[y],sd=4,mean=catch.position$lat[y]),
    year=yearrange[y],N=0) # 操業位置をランダムに決める
  for(i in 1:effort[y]){
    # 半径2度のなかにいくつ群があるか
    a <- sum(((fishdist$lon-catch.tmp$lon[i])^2+(fishdist$lat-catch.tmp$lat[i])^2)<0.5)
    if(a!=0) catch.tmp$N[i] <- floor(sum(exp(rnorm(a,mean=2))))
    # 半径2度の円の中に群がいたら、各群からランダム匹数だけ漁獲する
  }
  catch <- rbind(catch,catch.tmp)
}
catch <- catch[-1,]
write.csv(catch,file="testdata.csv")
```



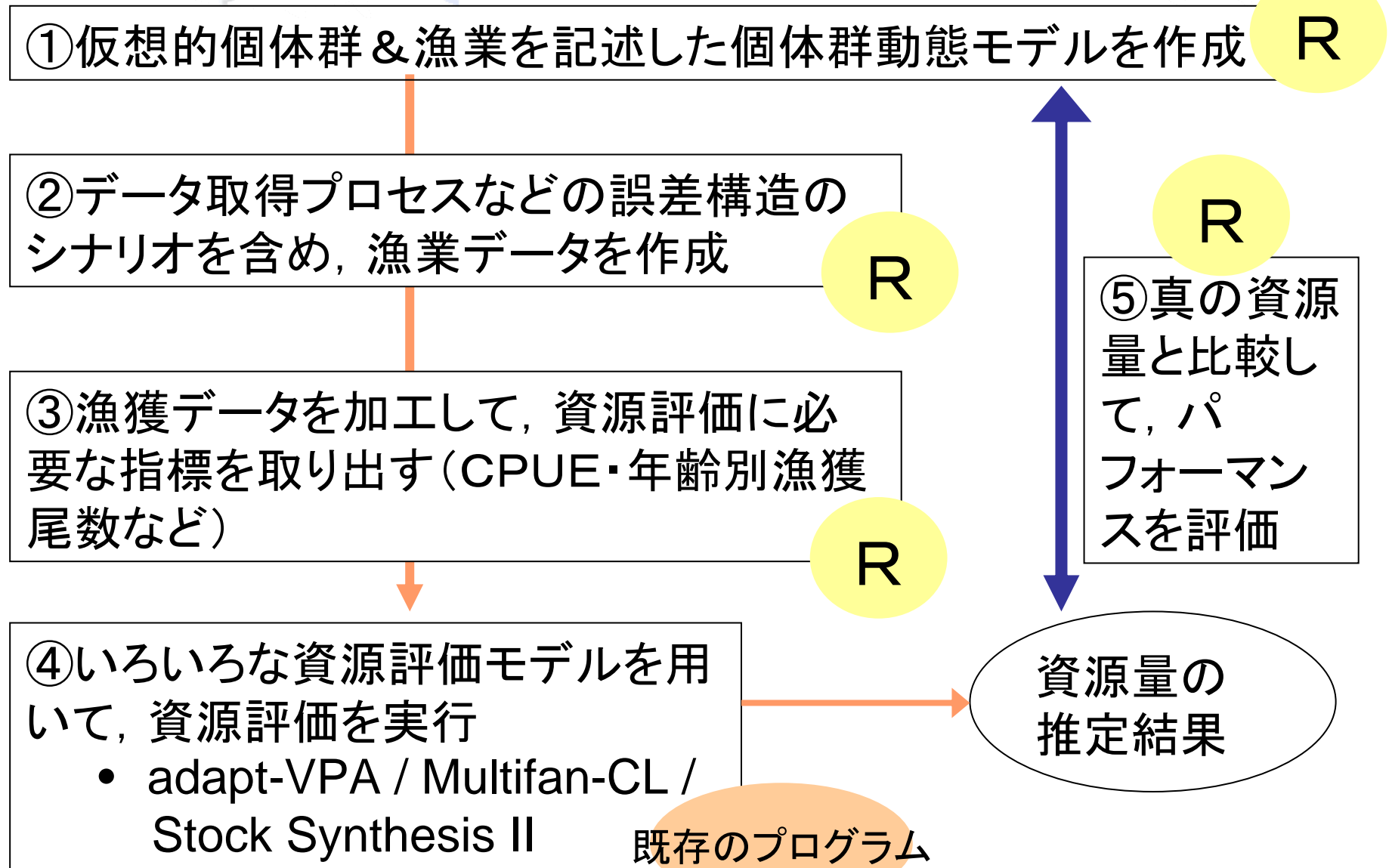
# Rの使用例(3) 複数の資源量評価モデルの評価のためのオペレーティングモデル

(遠洋水産研究所 竹内幸夫氏との共同研究)

- ・ 今の国際浮魚資源の資源評価に関する問題点: プロダクションモデルやVPAのような単純なモデルから, 複雑な資源評価モデル(統合モデル)への移行
- ・ 統合モデルとは?
  - Multifan-CL (Founier 2001)やStock-Synthesis II (Methot 2005)のような, 体長ベースの年齢構造化モデル
- ・ 統合モデルの利点
  - 体長データやタグデータをインプットデータとして, モデル内で成長曲線・年齢別漁獲尾数から資源量まで全てを推定(パラメータ数は数十から数百)
- ・ 統合モデルの欠点
  - 推定パラメータが多すぎるため, モデルの挙動が不安定になる場合も.

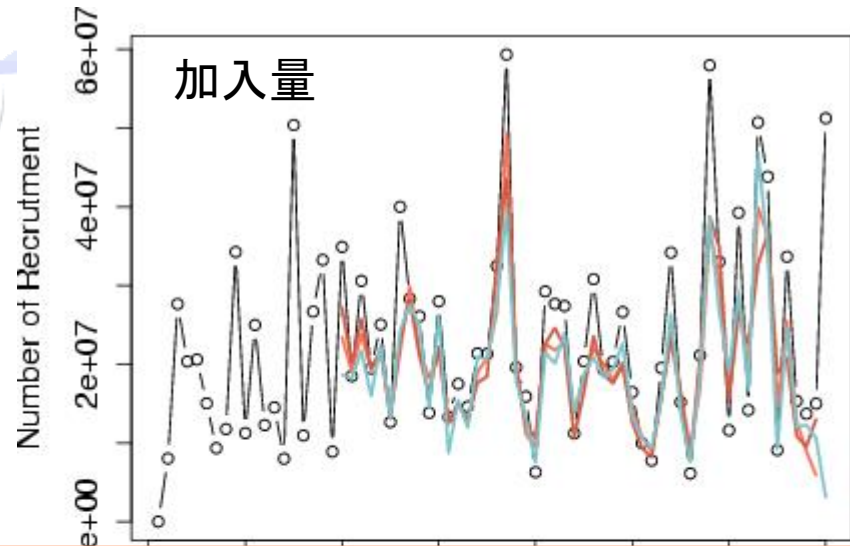
➔ 単純なモデルと, 複雑なモデル, どちらが優れているのか?  
また, 特殊な状況に対するロバスト性は?

# 資源評価モデルを評価するオペレーティングモデル → Rを使って半自動化



## OMの結果の例

推定した加入量と資源量の真の値との比較



- 現実でありそうな漁業データのサンプリングプロセスの偏りが、どのくらい資源評価結果に影響を及ぼすか & 各モデルの頑健性を調べることができる
  - 過去ほどデータの精度（漁獲量、サイズサンプルデータの減少）が悪くなる
  - データの精度の悪さはどのくらいまで許せるか？過去の資源評価の結果はどのくらい信頼できるか？



## まとめ

- データの可視化 & 解析をインタラクティブに行うことで漁業データのより深い理解が可能に
- 資源解析の欠点 = 真の解がわからない
  - 簡単なシミュレーションを行い, 方法の頑健性を試すことによって, より建設的な議論を行うことができる
    - CPUEシミュレーション: Goodyear (2006) Col. Vol. Sci. Pap. ICCAT 59(1) 211-233
    - 統合モデルのオペレーティングモデルによる評価: Labelle (2005) Fish. Res. 71, 311-344

# このプレゼンを作る際にお世話になった情報源 & その他の情報(敬称略)

- 最もよく行くサイトRのサイト
  - Rjpwiki: <http://www.okada.jp.org/RWiki/>
- 参考書
  - 初級編 「The R Tipsーデータ解析環境Rの基本技・グラフィックス活用集 船尾暢男 編」
  - 中級編 「Rの基礎とプログラミング技法 U.リゲス著 石田基広訳」
- 混合モデルについて
  - 生態学データ解析 GLMM (久保拓弥)  
<http://hosho.eec.hokudai.ac.jp/~kubo/ce/LinksGlmm.html>
- お魚の画像 魚と釣りの素材「海の素材集」サイト  
(<http://www.otomiya.com/sozai/>)

## おわり

このシンポジウムの開催と講演の機会を  
提供して下さった岡村寛様,  
また, ご清聴くださった皆様,  
ありがとうございました

このプレゼンで使ったプログラムは, 近日中に市野川の  
ホームページにアップします

(<http://dolphin.c.u-tokyo.ac.jp/~momoko7/>)

