

# VASTの基礎1

金森由妃（中央水研）

vast\_workshop2020

Ver 1.0

# 準備

1. テストコードをRStudioのソースエディタにコピペ
2. ワークショップ用フォルダ“vastws”を作成
  - 解析するデータを入れる
  - パスの確認
3. 作業ディレクトリの変更: 2行目をvastwsのパスに変更
4. 1~11行目を実行

# テストコードの構成

1. 各種設定
  2. データセットの作成
  3. パラメータの推定
  4. 作図（ほぼ説明しない）
- ✓ “1.4 Model settings” が最も重要（その他の部分はほぼ触らない）
  - ✓ **解析したいモデルに合わせて変更**する必要がある

# 目標

- ・ 今回解析するモデル

- ① 遭遇確率

$$\text{遭遇確率} = \underbrace{\text{年}}_{\text{固定効果}} + \underbrace{\text{場所} + \text{年} \times \text{場所}}_{\text{ランダム効果}}$$

- ② 生物が採集された場合の現存量

$$\text{現存量} = \underbrace{\text{年}} + \underbrace{\text{場所} + \text{年} \times \text{場所}}$$

- ・ 単一種のみを扱い， 年の効果は独立とする

# 目標

- ・ 今回解析するモデル

- ① 遭遇確率

$$p_1(i) = \beta_1(c_i, t_i) + \underbrace{\sum_{f=1}^{n_{\omega 1}} L_{\omega 1}(c_i, f) \omega_1(s_i, f)}_{\text{ランダム効果}} + \underbrace{\sum_{f=1}^{n_{\varepsilon 1}} L_{\varepsilon 1}(c_i, f) \varepsilon_1(s_i, f, t_i)}_{\text{ランダム効果}}$$

固定効果

ランダム効果

- ② 生物が採集された場合の量

$$p_2(i) = \beta_2(c_i, t_i) + \underbrace{\sum_{f=1}^{n_{\omega 2}} L_{\omega 2}(c_i, f) \omega_2(s_i, f)}_{\text{ランダム効果}} + \underbrace{\sum_{f=1}^{n_{\varepsilon 2}} L_{\varepsilon 2}(c_i, f) \varepsilon_2(s_i, f, t_i)}_{\text{ランダム効果}}$$

# 目標

- ・ 今回解析するモデル

- ① 遭遇確率

$$p_1(i) = \beta_1(c_i, t_i) + \underbrace{\sum_{f=1}^{n_{\omega 1}} L_{\omega 1}(c_i, f) \omega_1(s_i, f)}_{\text{ランダム効果}} + \underbrace{\sum_{f=1}^{n_{\varepsilon 1}} L_{\varepsilon 1}(c_i, f) \varepsilon_1(s_i, f, t_i)}_{\text{ランダム効果}}$$

固定効果

ランダム効果

- ② 生物が採集された場合の量

$$p_2(i) = \beta_2(c_i, t_i) + \underbrace{\sum_{f=1}^{n_{\omega 2}} L_{\omega 2}(c_i, f) \omega_2(s_i, f)}_{\text{ランダム効果}} + \underbrace{\sum_{f=1}^{n_{\varepsilon 2}} L_{\varepsilon 2}(c_i, f) \varepsilon_2(s_i, f, t_i)}_{\text{ランダム効果}}$$

- ・ 式で出てくるギリシャ文字はVASTでもパラメータ名として使われているので、数式が苦手でも我慢
- ・ 単一種（つまり  $c = 1, f = 1, L_{\omega} = 1, L_{\varepsilon} = 1$ ）
- ・ 年の効果は独立（ $\beta$  と  $\varepsilon$  に関係）

# 目標

- ・ 今回解析するモデル

- ① 遭遇確率

$$p_i = \underbrace{\beta_p(t_i)}_{\text{固定効果}} + \underbrace{\omega_p(s_i) + \varepsilon_p(s_i, t_i)}_{\text{ランダム効果}}$$

固定効果

ランダム効果

- ② 生物が採集された場合の量

$$d_i = \underbrace{\beta_d(t_i)}_{\text{固定効果}} + \underbrace{\omega_d(s_i) + \varepsilon_d(s_i, t_i)}_{\text{ランダム効果}}$$

- ・ 式で出てくるギリシャ文字はVASTでもパラメータ名として使われているので、数式が苦手でも我慢
- ・ 単一種（つまり  $c = 1, f = 1, L_\omega = 1, L_\varepsilon = 1$ ）
- ・ 年の効果は独立（ $\beta$  と  $\varepsilon$  に関係）

# **1. 各種設定**

2. データセットの作成

3. パラメータの推定

4. 作図（ほぼ説明しない）



# 1.2 Version for cpp code

- TMBのコード
- VAST自体のバージョンとは異なる
- 最新版では色々なオプションがある（が、使ったことない）
- テストコードでコンパイルエラーが出た人は、テストコードが動いたバージョンに変更してください

# 1.3 Spatial settings

K平均法によるknot決めに関する部分

- **Method = c("Grid", "Mesh", "Spherical\_mesh")[#]**
  - Grid:
  - Mesh:
  - Spherical\_mesh
- **Kmeans\_Config = list(...)**
  - 変更の必要なし
- **grid\_size\_km = 25**
  - MethodがGridの時に必要な情報
  - 基本はMeshなので関係ないが、NULLとすると2.3でエラーが出る
- **n\_x = 100**
  - knotの数の指定
  - 100以上が好ましい (Thorson et al. 2019)
  - knot数が多いほど、AICは下がる

# 1.4 Model settings: 超重要

- **FieldConfig = c(Omega1 = \_\_\_, Epsilon1 = \_\_\_, Omega2 = \_\_\_, Epsilon2 = \_\_\_)**
  - ファクターの数（カテゴリの数が上限）
- **RhoConfig = c(Beta1 = \_\_\_, Beta2 = \_\_\_, Epsilon1 = \_\_\_, Epsilon2 = \_\_\_)**
  - 年の効果をどう考えるか？
  - 0 = 固定効果, 1 = 独立, 2 = RW, 3 = 定数, 4 = AR
- **OverdispersionConfig = c("Eta1" = 0, "Eta2" = 0)**
  - VASTの基礎2でのお楽しみ
- **ObsModel = c(PosDist = \_\_\_, Link = \_\_\_)**
  - 事前分布とリンク関数の設定
  - ?make\_dataを参照（めちゃくちゃたくさんある）

# 1.5~1.7

## 1.5 Stratification for results

- **strata.limits = data.frame('STRATA'="All\_areas")**
  - Grid:
  - 変更の必要なし

## 1.6 Derived objects: 解析するデータの場所の指定

- **Region = "British\_Columbia"**
  - 自分のデータを解析する場合は, “other”に変更

## 1.7 Save settings: ここまでの設定を保存する

- **DateFile = paste0(getwd(), '/VAST\_output/')**
  - 今回の解析用に作成するフォルダの名前
  - 例えば私は, paste0(getwd(), "/vast", Sys.Date(), "\_lnorm\_log", n\_x, sakana, "fixed")として, フォルダ名でモデル・knot数・魚種が分かるようにしている

1. 各種設定
- 2. データセットの作成**
3. パラメータの推定
4. 作図（ほぼ説明しない）

# 2.1 Data-frame for catch-rate data

解析用のデータフレームを作成する

- いわゆるtidyなデータを作成し、**最終的なオブジェクト名を Data\_Geostat にする！**
- **列名はオリジナルで作らない！キャメルケースで書く！**
  - Year, Month, Lon, Lat, Catch\_KG, Vessel など
  - キャメルでない場合は、コードの修正が必要
  - Catch\_KG は変更すると以降でエラーが出る場合もある
- 52-61行目部分を変更し、自分のデータでData\_Geostatを作成
  - 今回必要なのは、Year, Lon, Lat, Catch\_KG

例を入れるか、R上で見せる

## 2.2 Extrapolation grid

- データの位置情報を取得
- 緯度経度をUTM(Universal Transverse Mercator)座標へ変換
- 日本は51~56の範囲
- 検出したデータの位置情報を教えてくれるので、確認をする

Using strata 1

convUL: For the UTM conversion, automatically detected zone 9.  
convUL: Converting coordinates within the northern hemisphere.



## 2.3 derived objects for spatio-temporal estimation

調査・漁業地点をknotに変換（1. 3の設定を使っている）

- **Save\_Results = TRUE**

- Kmeans-100.RDataがフォルダに作成される
- 二度目以降はKmeans-100.RDataを使ってくれる

convUL: Converting coordinates within the northern hemisphere.

convUL: For the UTM conversion, used zone 9 as specified

convUL: Converting coordinates within the northern hemisphere.

convUL: For the UTM conversion, used zone 9 as specified

Num=1 Current\_Best=Inf New=172166.9

Num=2 Current\_Best=172166.9 New=159856.2

Num=3 Current\_Best=159856.2 New=160643.3

⋮

Num=99 Current\_Best=153309.7 New=161469.7

Num=100 Current\_Best=153309.7 New=159542.5

Calculated and saved to /Users/Yuki/Dropbox/vastws/VAST\_output//Kmeans-100.RData

convUL: Converting coordinates within the northern hemisphere.

convUL: Converting coordinates within the northern hemisphere.



1. 各種設定
2. データセットの作成
- 3. パラメータの推定**
4. 作図（ほぼ説明しない）

# 3 Build and run model: make\_data()

TMB用のデータセットを作成

- **c\_iz**
  - カテゴリ数 (e.g., 魚種, 年齢, 銘柄 . . . ) . 0始まりで
- **b\_i**
  - 生物の量 (応答変数)
- **a\_i**
  - 調査・漁業の面積
- **s\_i**
  - knot数. 0始まりで
- **t\_i**
  - 年
- **Aniso = TRUE**
  - 空間相関の歪みを考慮するかどうか
- その他については ?make\_data

# 3 Build and run model: make\_data()

FieldConfig\_input is:

	Component_1	Component_2
Omega	1	1
Epsilon	1	1
Beta	-2	-2

OverdispersionConfig\_input is:

Eta1	Eta2
-1	-1

Calculating range shift for stratum #1:

# 3 Build and run model: make\_model()

cppファイルをコンパイルし，推定するパラメータが列挙される

List of estimated fixed and random effects:

	Coefficient_name	Number_of_coefficients	Type
1	beta1_ct	7	Fixed
2	beta2_ct	7	Fixed
3	L_epsilon1_z	1	Fixed
4	L_epsilon2_z	1	Fixed
5	L_omega1_z	1	Fixed
6	L_omega2_z	1	Fixed
7	ln_H_input	2	Fixed
8	logkappa1	1	Fixed
9	logkappa2	1	Fixed
10	logSigmaM	1	Fixed
11	Epsiloninput1_sft	1276	Random
12	Epsiloninput2_sft	1276	Random
13	Omegainput1_sf	116	Random
14	Omegainput2_sf	116	Random

# 3 Build and run model: fit\_tmb()

## パラメータの推定

```
Constructing atomic D_lgamma
Optimizing tape... Done
iter: 1  value: 13012.14 mgc: 36.81998 ustep: 1
iter: 2  value: 12951.89 mgc: 9.56431 ustep: 1
iter: 3  value: 12949.05 mgc: 2.199174 ustep: 1
Matching hessian patterns... Done
outer mgc: 3081.279
      ⋮
iter: 1  value: 11789.58 mgc: 0.005095535 ustep: 1
iter: 2  value: 11789.58 mgc: 1.857035e-06 ustep: 1
iter: 3  mgc: 1.665644e-13
outer mgc: 3.115627
iter: 1  mgc: 2.867521e-11
outer mgc: 0.004092186
Optimizing tape... Done
iter: 1  mgc: 2.867521e-11
Matching hessian patterns... Done
outer mgc: 31832.82
#####
The model is likely not converged
#####
```

1. 各種設定
2. データセットの作成
3. パラメータの推定
- 4. 作図**

# 注意点

- Save.RDataを使って、あとから4. Figuresを動かすことはできない
- 作図の体裁は基本的に変更できないので、綺麗な図を描きたければ自分で頑張るしかない

# 1.5~1.7

## 4.7 Density surface for each year

- **Dens\_xt = plot\_maps(plot\_set = c(3), · · ·)**
  - 設定を変えると色々な図が作成される
  - ?plot\_maps参照

## 4.8 Index of abundance

- **Index = plot\_biomass\_index(···, use\_biasecorr = TRUE)**
  - バイアスコレクションは必須 (Thorson & Kristensen 2016)

Using bias-corrected estimates for abundance index (natural-scale)...

Using bias-corrected estimates for abundance index (log-scale)...

## 4.9 Center of gravity and range expansion/contraction

- **plot\_range\_index()**

Plotting center-of-gravity...

Using bias-corrected estimates for center of gravity...

Plotting effective area occupied...

Using bias-corrected estimates for effective area occupied (natural scale)...

Using bias-corrected estimates for effective area occupied (log scale)...



