

# Spatio-Temporal Statistics with R

## Chapter 3

# Spatio-Temporal Statistical Models

## 時空間統計モデル

Spatio-Temporal Statistics with R 輪読

2019年5月10日

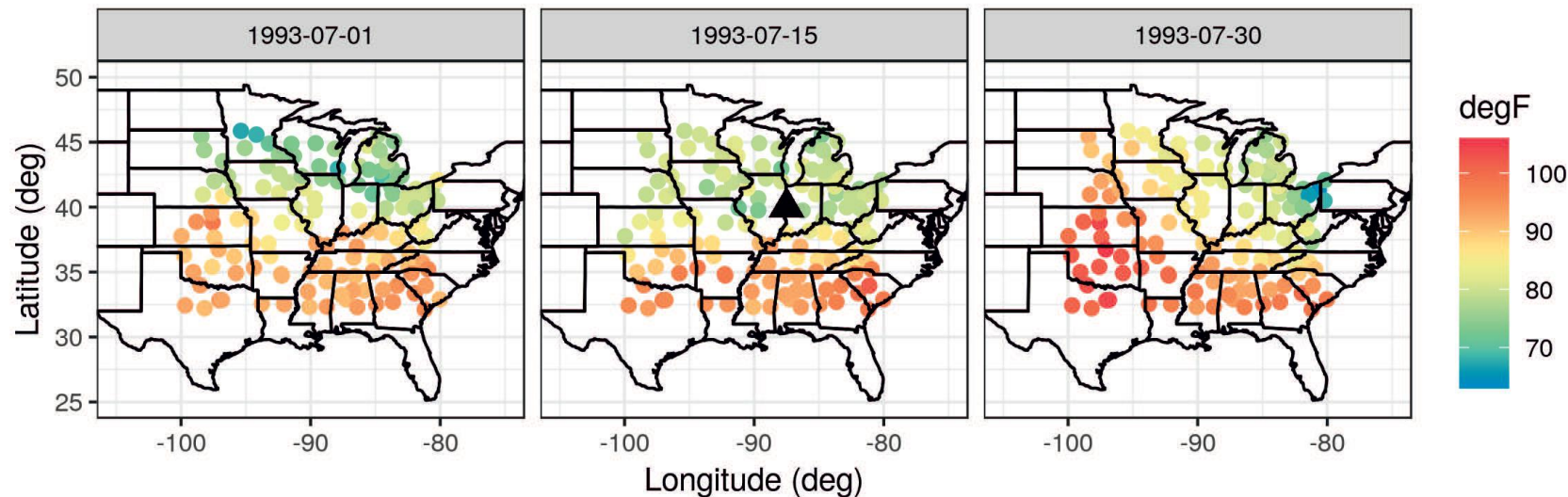
担当：澤田紘太

# 3. 時空間統計モデル

- 時空間統計モデリングの基本目標
  - 観測期間内のある地点における応答変数の値を予測し、その予測の不確実性も示す
  - 時空間的な非独立性のもとで、応答変数に対する共変量の重要性を科学的に推定する
  - ある場所における応答変数の未来の値を予報し、その予報の不確実性も示す

## 3.1 時空間的な予測

- 目標：▲地点の1993年7月15日の最高気温を予測したい
  - 同じ日付の周辺の温度データだけでなく、前後の日付のデータも使う
    - 前も後もデータがあるので、smoothing
      - 7月15日までしかデータがないならfiltering
      - 7月15日より前までしかデータがないならforecasting



# 決定論的予測

- 逆距離加重法 (inverse distance weighting, IDW)

- 全データ点の加重平均 (予測したい点に近いほど加重)

- 各時点 $t_j$ にそれぞれ $m_j$ 地点のデータがある

$$\{Z(\mathbf{s}_{11}; t_1), Z(\mathbf{s}_{21}; t_1), \dots, Z(\mathbf{s}_{m_1 1}; t_1), \dots, Z(\mathbf{s}_{1T}; t_T), Z(\mathbf{s}_{2T}; t_T), \dots, Z(\mathbf{s}_{m_T T}; t_T)\}$$

- IDW予測子 
$$\hat{Z}(\mathbf{s}_0; t_0) = \sum_{j=1}^T \sum_{i=1}^{m_j} w_{ij}(\mathbf{s}_0; t_0) Z(\mathbf{s}_{ij}; t_j)$$

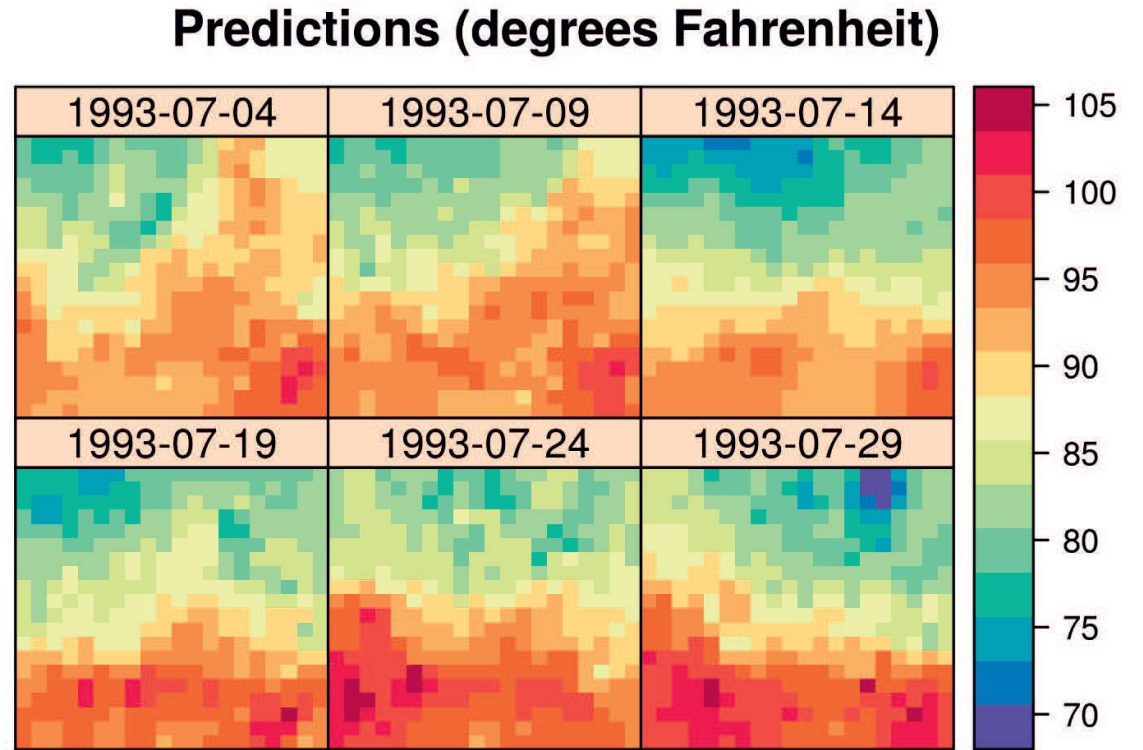
- 重みづけ 
$$w_{ij}(\mathbf{s}_0; t_0) \equiv \frac{\tilde{w}_{ij}(\mathbf{s}_0; t_0)}{\sum_{k=1}^T \sum_{\ell=1}^{m_k} \tilde{w}_{\ell k}(\mathbf{s}_0; t_0)},$$

$$\tilde{w}_{ij}(\mathbf{s}_0; t_0) \equiv \frac{1}{d((\mathbf{s}_{ij}; t_j), (\mathbf{s}_0; t_0))^\alpha},$$

- $\alpha$ は正の実数、 $d()$ は何らかの距離 (e.g. ユークリッド)

# IDWによる予測

- データのない日（7月14日）の予測はより滑らか
  - 他の方法でもそうなりやすい
- $\alpha = 5$ で計算
  - 小さくするとより滑らか
  - CVで最適値を選ぶ（後で）



# カーネル予測子

- IDWはカーネル予測子の一種

$$\tilde{w}_{ij}(\mathbf{s}_0; t_0) = k((\mathbf{s}_{ij}; t_j), (\mathbf{s}_0; t_0); \theta)$$

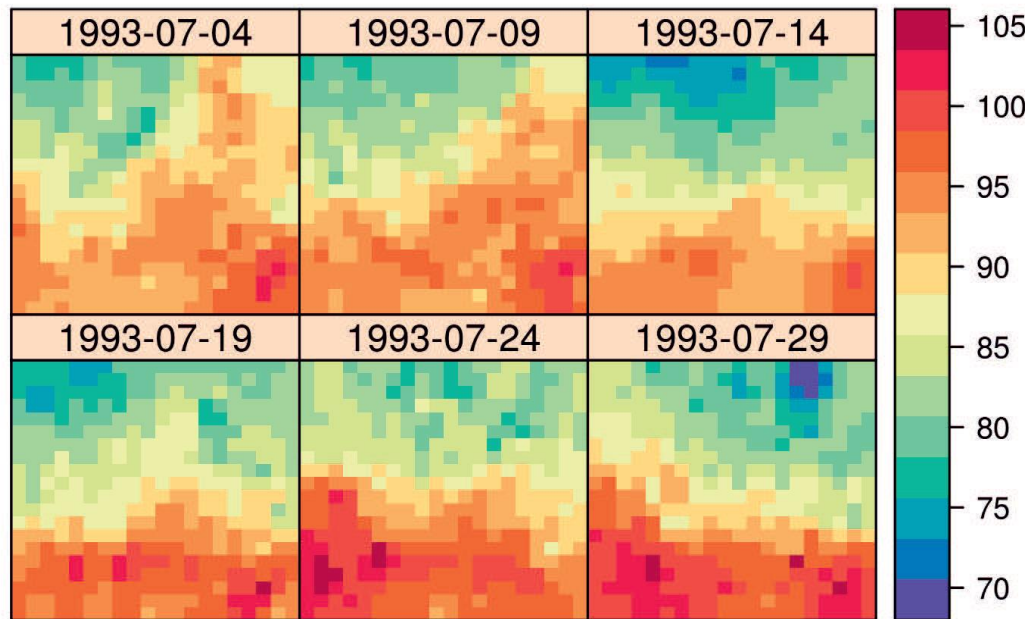
- $k()$ はカーネル関数（2点間の類似性を示す関数）
  - $\theta$ はバンド幅パラメータ、 $\alpha$ に対応
  - 台はコンパクトでない（コンパクト=閾値以上の距離で重みゼロ）
- 他のカーネル予測子の例
    - Gaussian radial basis kernel

$$k((\mathbf{s}_{ij}; t_j), (\mathbf{s}_0; t_0); \theta) \equiv \exp \left( -\frac{1}{\theta} d((\mathbf{s}_{ij}; t_j), (\mathbf{s}_0; t_0))^2 \right)$$

# カーネル予測子による予測

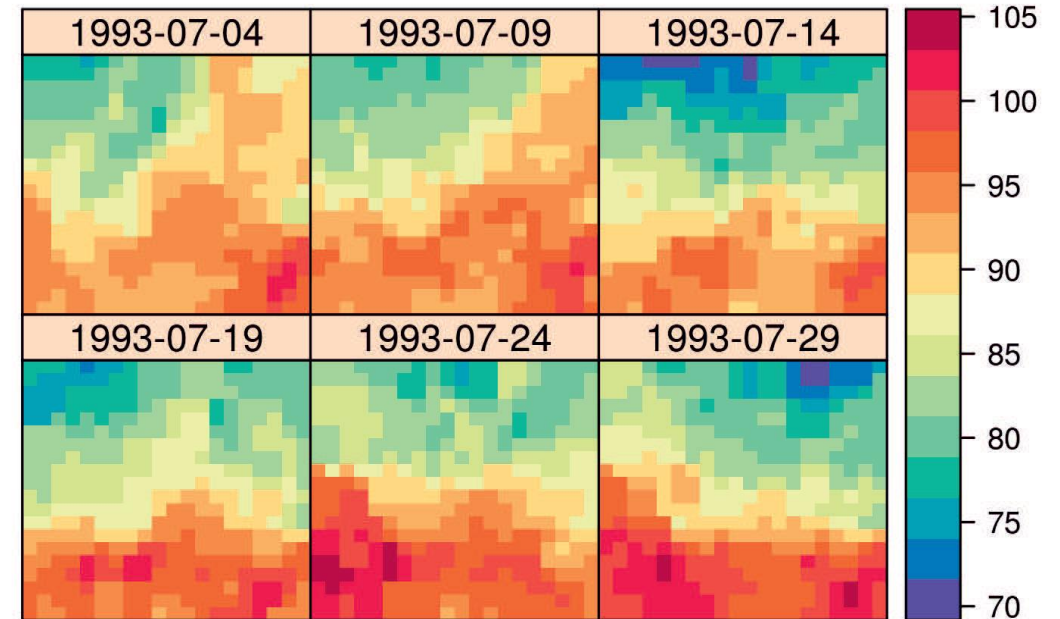
IDW

Predictions (degrees Fahrenheit)



Gaussian radial basis kernel

Predictions (degrees Fahrenheit)



- どっちがいい？ → 交差検証(cross-validation, CV)

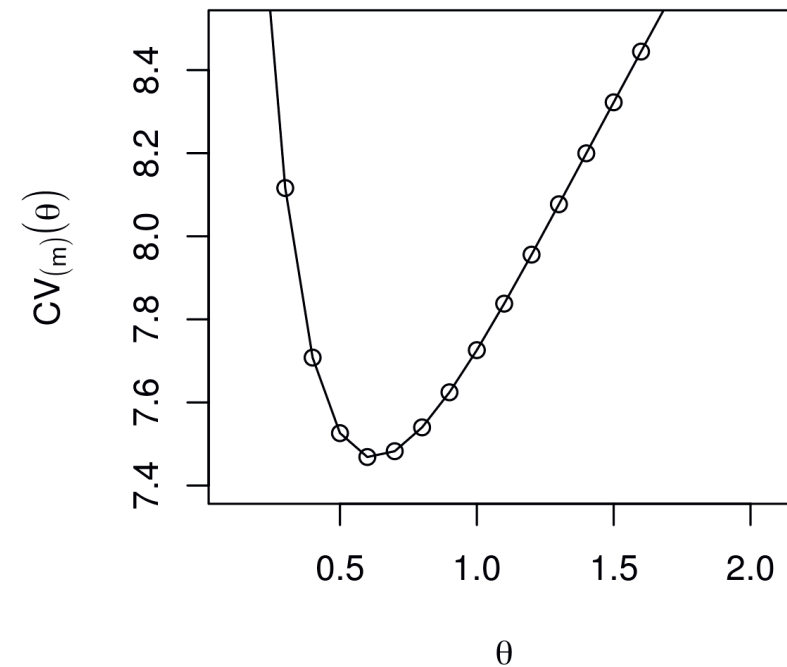
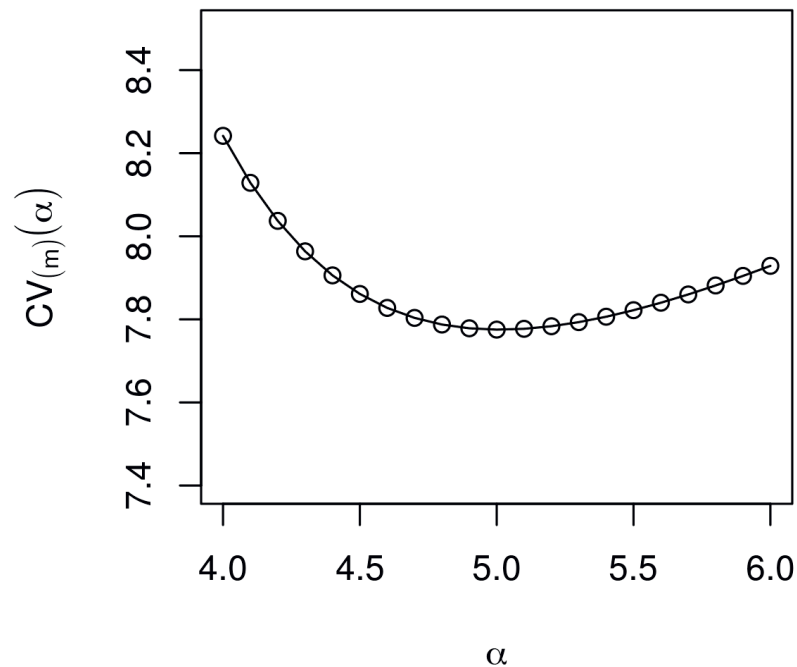
# Technical Note 3.1: Cross-Validation

- データをtraining sampleとvalidation sampleに分け、前者に当てはめたモデルを後者で評価する
- K分割交差検証法K-fold cross-validation
  - $m$ 点のデータをK個のほぼ同じサイズのfoldにランダムに分ける
  - 各foldについて、残りのK-1個のfoldでモデルを当てはめる
  - k-th foldに含まれる $m_k$ 点について予測 $\hat{Z}_i^{(-k)}$  ( $i = 1, \dots, m_k$ )
  - k-th foldについての平均二乗予測誤差  $MSPE_k = \frac{1}{m_k} \sum_{i=1}^{m_k} (Z_i - \hat{Z}_i^{(-k)})^2$
  - K-fold cross-validation score  $CV_{(K)} = \frac{1}{K} \sum_{k=1}^K MSPE_k$
- Leave-one-out cross validation (LOOCV)
  - K分割交差検証法で $K = m$ の場合（各データ点が1つのfoldになる）
  - LOOCVスコア  $CV_{(m)} = \frac{1}{m} \sum_{i=1}^m MSPE_i$
- LOOCVは期待二乗誤差の推定値としてはバイアスが小さいが、分散が大きい
  - $K=5, 10$ くらいが妥当な落としどころ



# LOOCVによる比較

- IDWの $\alpha$ は5、Gaussの $\theta$ は0.6がベスト
- Gaussのほうが最適なバンド幅でのCVスコアが低く、うまく予測できている



## 3.2 回帰（傾向面）推定

- ふつうの（重）回帰モデルを当てはめる
  - 時空間的な非独立性は”傾向 trend”=共変量で説明できると仮定
  - モデル内で誤差・予測誤差分散を得ることができる

$$Z(\mathbf{s}_i; t_j) = \beta_0 + \beta_1 X_1(\mathbf{s}_i; t_j) + \dots + \beta_p X_p(\mathbf{s}_i; t_j) + e(\mathbf{s}_i; t_j)$$

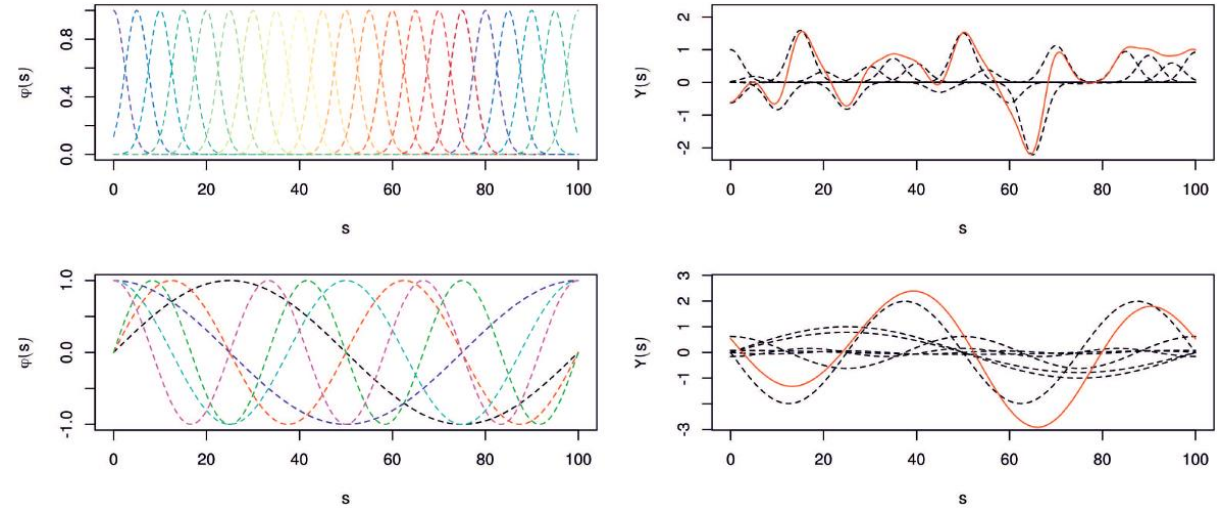
- 一旦、誤差はi.i.d.と仮定  $e(\mathbf{s}_i; t_j) \sim \text{indep. } N(0, \sigma_e^2)$
- 共変量
  - 空間的に変わるもの e.g. 標高
  - 時間的に変わるもの e.g. 季節
  - 時空間的に変わるもの e.g. 湿度
  - 基底関数basis functions

# 基底関数

- 初等的な基底関数の線形結合で複雑な曲線・曲面を表現する

$$Y(s) = \alpha_1 \phi_1(s) + \alpha_2 \phi_2(s) + \dots + \alpha_r \phi_r(s)$$

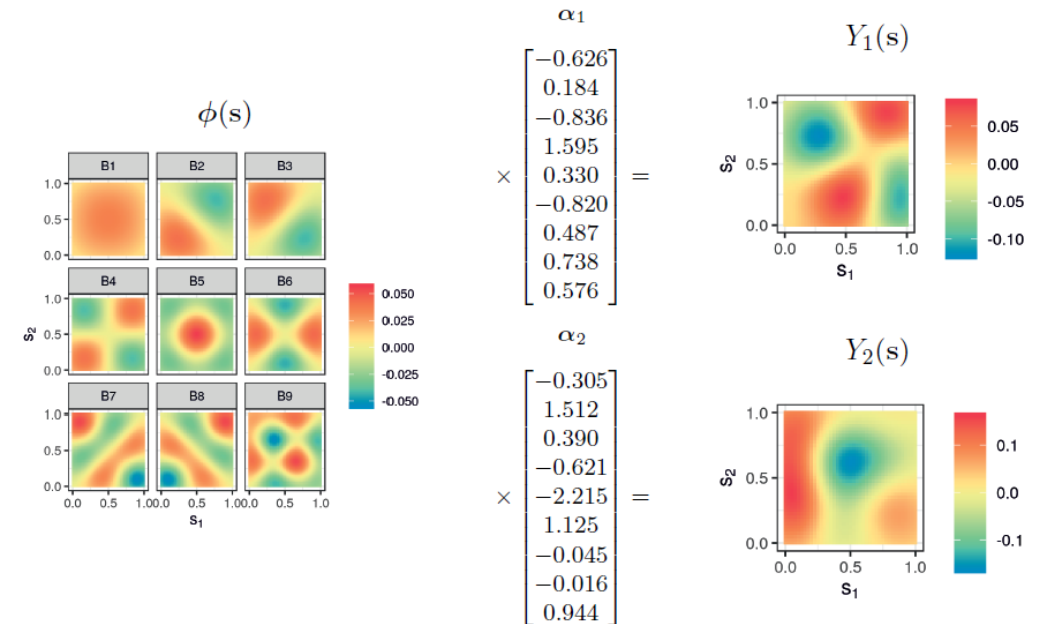
- 基底関数に係数（重み）をかけて足し合わせる



Spatial Basis Functions

Coefficients

Process

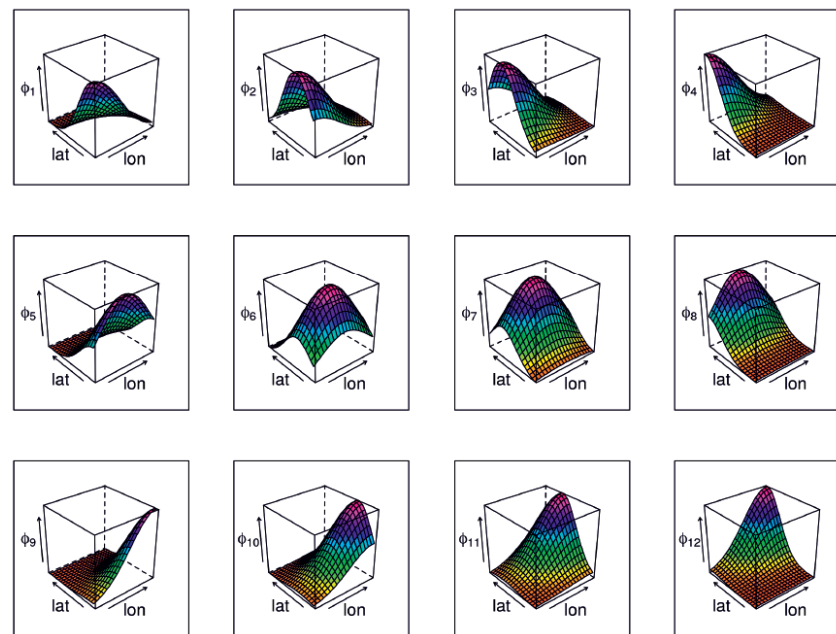


# 回帰モデル

- 下記の基底関数（共変量）の線形結合モデル

- 全体平均（切片）
- 線形の緯度
- 線形の経度
- 線形の時間（日）トレンド
- 緯度と経度の交互作用
- 緯度と時間の交互作用
- 経度と時間の交互作用
- 空間のみの基底関数12個

- 切片と18個の回帰係数を推定



# 最小二乗法 Ordinary least squares, OLS

- 残差平方和RSSを最小化するパラメータの組を求める  
 $\{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p\}$

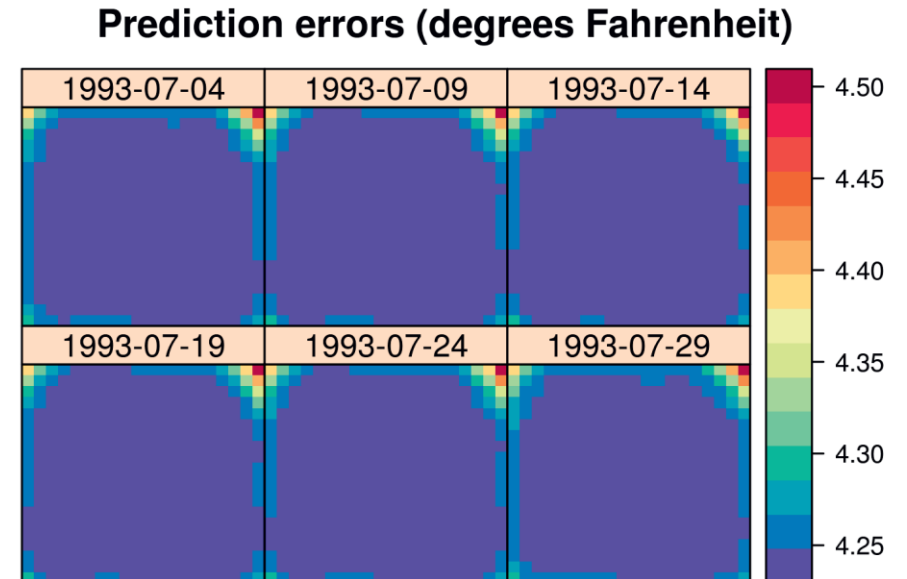
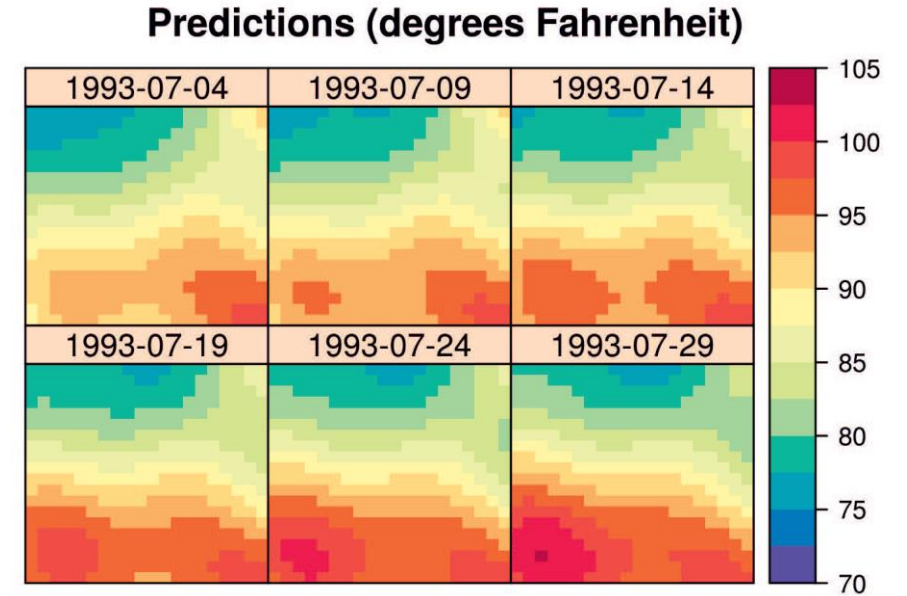
$$RSS = \sum_{j=1}^T \sum_{i=1}^m (Z(\mathbf{s}_i; t_j) - \hat{Z}(\mathbf{s}_i; t_j))^2$$

$$\hat{Z}(\mathbf{s}; t) = \hat{\beta}_0 + \hat{\beta}_1 X_1(\mathbf{s}; t) + \dots + \hat{\beta}_p X_p(\mathbf{s}; t)$$

- 分散パラメータも同時に得られる  $\hat{\sigma}_e^2 = \frac{RSS}{mT-p-1}$

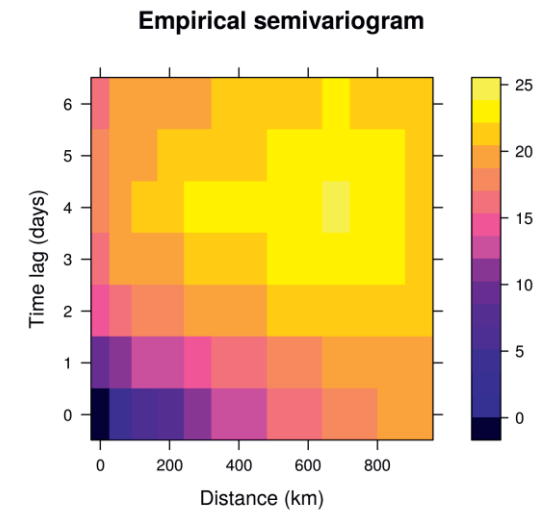
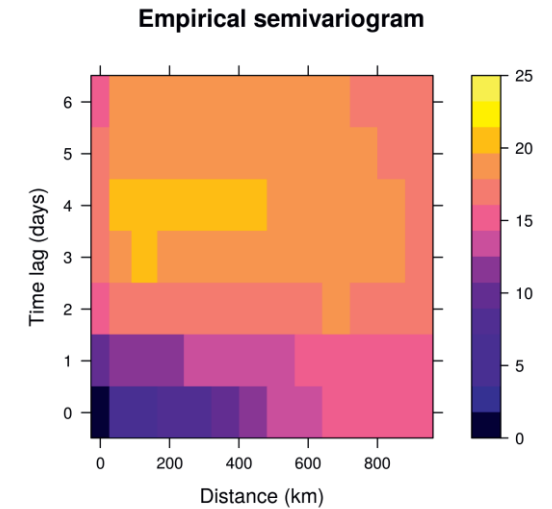
# OLSによる予測

- カーネル法より滑らか
  - 滑らかな基底関数を使っているため
- 予測誤差に時空間的パターンはあまりない
  - 共変量で時空間的ばらつきの大部分を説明できているため
- 予測誤差は端のほうで大きい
  - 外挿に近くなる
- 観測誤差とモデルの当てはまりの悪さが区別できない



## 3.2.1 モデル診断: 独立でない誤差

- 残差のセミバリオグラム
  - 元データのセミバリオグラムと比較
- 共変量でデータの時空間変動をある程度説明できているが、すべてではない



# 非独立性の検定 (Tech Note 3.2含む)

- ダービン・ワトソン検定 Durbin-Watson test

- 時系列の連続する点での非独立性
- 自己相関が強いと小さくなる

$$d = \frac{\sum_{t=2}^T (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^T \hat{e}_t^2}$$

- モランのI統計量 Moran's I

- 空間的自己相関の統計量
- 距離で重みづけしたピアソンの相関係数
- 時間も含めるとSpace-time Index

$$I = \frac{m \sum_{i=1}^m \sum_{j=1}^m w_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{(\sum_{i=1}^m \sum_{j=1}^m w_{ij})(\sum_{i=1}^m (Z_i - \bar{Z})^2)}$$

- ダービン・ワトソン検定の時空間版

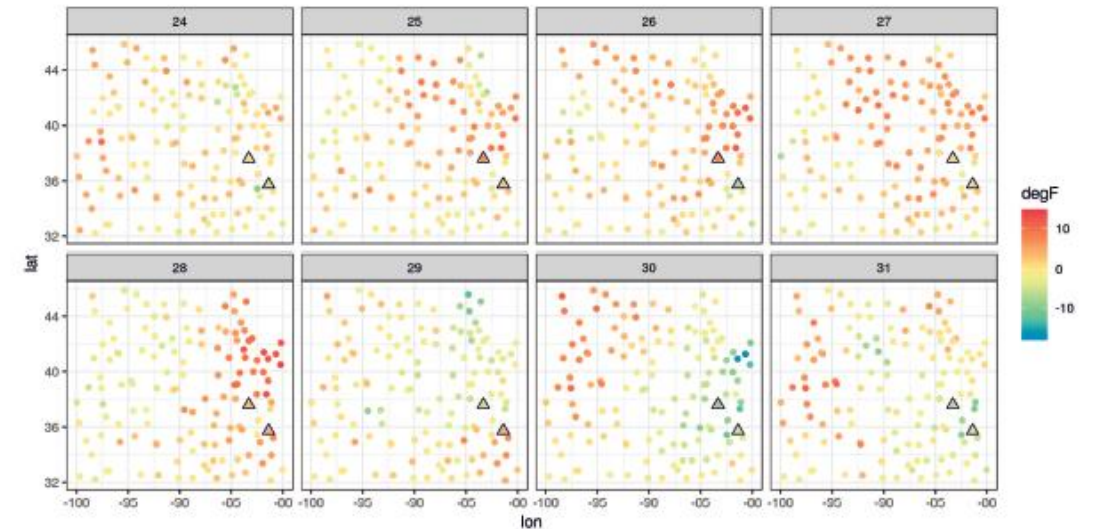
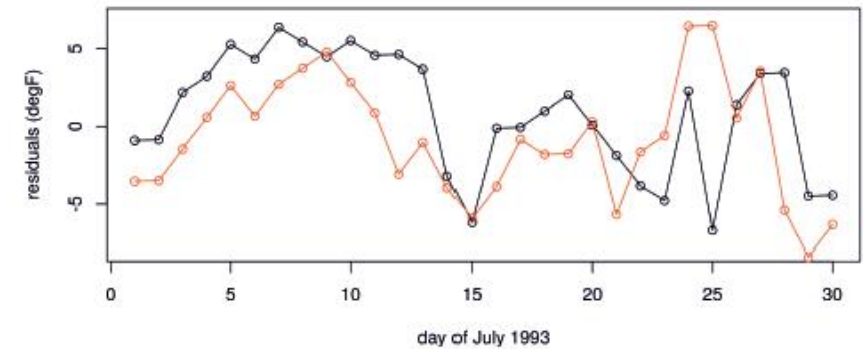
- $\hat{\gamma}_e(\|\mathbf{h}_1\|; \tau_1)$ : 最小の時間・空間ラグでのセミバリオグラム

$$F \equiv \left| \frac{\hat{\gamma}_e(\|\mathbf{h}_1\|; \tau_1)}{\hat{\sigma}_e^2} - 1 \right|$$



# OLS残差の自己相関

- 時間的・空間的な自己相関
  - 検定はLabで
- モデルが時空間的変動を説明しきれていない
  - 環境変動の解析ではよくあること



# 非独立性への対処

- 一般化最小二乗法 Generalized least squares, GLS
  - 誤差の独立性という仮定を緩め、誤差間の共分散を許す
  - $e \sim N(0, C_e)$   $C_e$ : 時空間共分散行列
  - 共分散行列が既知である必要  $\Rightarrow$  4章
    - データのない点を含む、予測したい範囲内の全点について
- 目的によっては気にしなくてもよい
  - OLSのパラメータ推定値・予測値は非独立性を無視しても不偏
  - 標準誤差・予測標準誤差はバイアス
    - 正の自己相関なら過少推定  $\rightarrow$  推定・予測の精度を過大評価

## 3.2.2 時空間モデルのパラメータ推定

- どの変数が重要なのかを知りたい
- GLSだとSE大きい傾向は同じ
- 緯度・経度・緯度:経度は重要でなさそう（信頼区間に0を含む）
  - 緯度が有意でないのは意外だが、緯度と他の変数の交互作用などがあるためか

	Dependent variable:	
	Max. Temperature (°F)	
	$\hat{\beta}_{ols} (SE(\hat{\beta}_{ols}))$	$\hat{\beta}_{gls} (SE(\hat{\beta}_{gls}))$
Intercept	192.240** (97.854)	195.320** (98.845)
Longitude	1.757 (1.088)	1.780 (1.097)
Latitude	-1.317 (2.556)	-0.974 (2.597)
Day	-1.216*** (0.134)	-1.237*** (0.136)
Longitude × Latitude	-0.026 (0.028)	-0.022 (0.029)
Longitude × Day	-0.023*** (0.001)	-0.023*** (0.001)
Latitude × Day	-0.019*** (0.002)	-0.019*** (0.002)
$\alpha_1$	16.647*** (4.832)	19.174*** (4.849)
$\alpha_2$	18.528*** (3.056)	16.224*** (3.125)
$\alpha_3$	-6.607** (3.172)	-4.204 (3.199)
$\alpha_4$	30.545*** (4.370)	27.500*** (4.493)
$\alpha_5$	14.739*** (2.747)	13.957*** (2.759)
$\alpha_6$	-17.541*** (3.423)	-15.779*** (3.461)
$\alpha_7$	28.472*** (3.552)	25.985*** (3.613)
$\alpha_8$	-27.348*** (3.164)	-25.230*** (3.202)
$\alpha_9$	-10.235** (4.457)	-7.401 (4.556)
$\alpha_{10}$	10.558*** (3.327)	8.561** (3.396)
$\alpha_{11}$	-22.758*** (3.533)	-19.834*** (3.569)
$\alpha_{12}$	21.864*** (4.813)	17.771*** (5.041)
Observations	3,989	3,989

Note:

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

# 他の問題

- 多重共線性
  - 一部変数の線形結合が他の変数と近似
  - 標準誤差が増大する
    - e.g. 基底関数5は有意だが、基底関数5にノイズを加えた関数をモデルに入れると有意でなくなる (Lab 3.2)
- 交絡
  - 重要な変数が無視されている/無関係な変数が含まれていることで解釈や有意性が変わる
    - 残差が非独立ということは、モデルにない共変量が重要
    - 予測が目的なら、非独立性をモデルに組み込んでしまえばよい

## 3.2.3 変数選択

- 重要な共変量だけを選び出す
- 総当たり
  - 全共変量の有無を総当たりで試し、予測力が最大のものを選ぶ
  - 共変量 $p$ 個で $2^p$ 個のモデルを計算する必要がある
- 効率的なアルゴリズムを用いたbest subset法
- 自動的な変数選択アルゴリズム
  - 変数増加法forward selection :  $p > n$ でも使える
  - 変数減少法backward selection
  - Mixed selection
  - LASSO, Ridge

# 変数増加法forward selection

- 切片のみモデルから始め、事前に決めた停止条件を満たすまで、評価基準を最も改善する変数を一つずつ追加していく
- AIC基準で4変数までのモデル
  - フルモデルで有意でない緯度が最初
  - 緯度:日が入ると係数は減少
- RSS基準にすると少し結果が変わる

Dependent variable:					
Max. Temperature (°F)					
	$\hat{\beta}_{ols}$				
	(1)	(2)	(3)	(4)	(5)
Intercept	88.673***	148.940***	147.840***	136.810***	138.420***
Latitude		-1.559***	-1.559***	-1.274***	-1.273***
Day			0.069***	0.755***	0.755***
Latitude × Day				-0.018***	-0.018***
Longitude					0.019
Observations	3,989	3,989	3,989	3,989	3,989
Residual Std. Error	7.726	4.710	4.669	4.626	4.625

Dependent variable:					
Max. Temperature (°F)					
	$\hat{\beta}_{ols}$				
	(1)	(2)	(3)	(4)	(5)
Intercept	88.673***	148.940***	147.780***	140.420***	122.020***
Latitude		-1.559***	-1.560***	-1.366***	-0.838***
Longitude × Day			-0.001***	-0.006***	-0.011***
Latitude × Day				-0.012***	-0.023***
$\alpha_{10}$					-6.927***
Observations	3,989	3,989	3,989	3,989	3,989
Residual Std. Error	7.726	4.710	4.661	4.607	4.470

# Technical Note 3.4: 正則化

- RSSに罰則項を足してパラメータ推定値をshrinkさせる

$$\sum_{j=1}^T \sum_{i=1}^m [Z(\mathbf{s}_i; t_j) - (\beta_0 + \beta_1 X_1(\mathbf{s}_i; t_j) + \dots + \beta_p X_p(\mathbf{s}_i; t_j))]^2 + \lambda \sum_{\ell=1}^p |\beta_{\ell}|^q$$

- Ridge回帰： $q = 2$  係数が0に近く推定される
- Lasso回帰： $q = 1$  一部の係数が0になる→変数選択
- ベイズ法：係数に事前分布を与える より一般的
  - Ridge・Lassoに相当するものもベイズで定式化できる
- 罰則付きの推定値はスケール不変でないので、Xをスケールリング・センタリングする