5 Dynamic Spatio-Temporal Models

5.3 Process and Parameter Dimension Reduction

5.3.1 Parameter Dimension Reduction

5.3.2 Dimension Reduction in the Process Model

5.4 Nonlinear DSTMs

5.3 Process and Parameter Dimension Reduction

Now, denoting the process vector $\mathbf{Y}_t \equiv (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))'$, (5.10) can be written in vector–matrix form as a linear first-order vector autoregression DSTM,

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \tag{5.11}$$

where the $n \times n$ transition matrix is given by $\mathbf{M}$ with elements $\{m_{ij}\}$, and the additive spatial error process $\boldsymbol{\eta}_t \equiv (\eta_t(\mathbf{s}_1), \ldots, \eta_t(\mathbf{s}_n))'$ is independent of $\mathbf{Y}_{t-1}$ and is specified to be mean-zero and Gaussian with spatial covariance matrix $\mathbf{C}_\eta$. The stability (non-explosive)

- spatial locations (n) >> time replications (T)の場合, 推定するパラメータ数はn^2オーダー
- パラメータの数を減らす and/or spatio-temporal dynamic processの次元を落としたい
  - lagged-nearest-neighbor representations

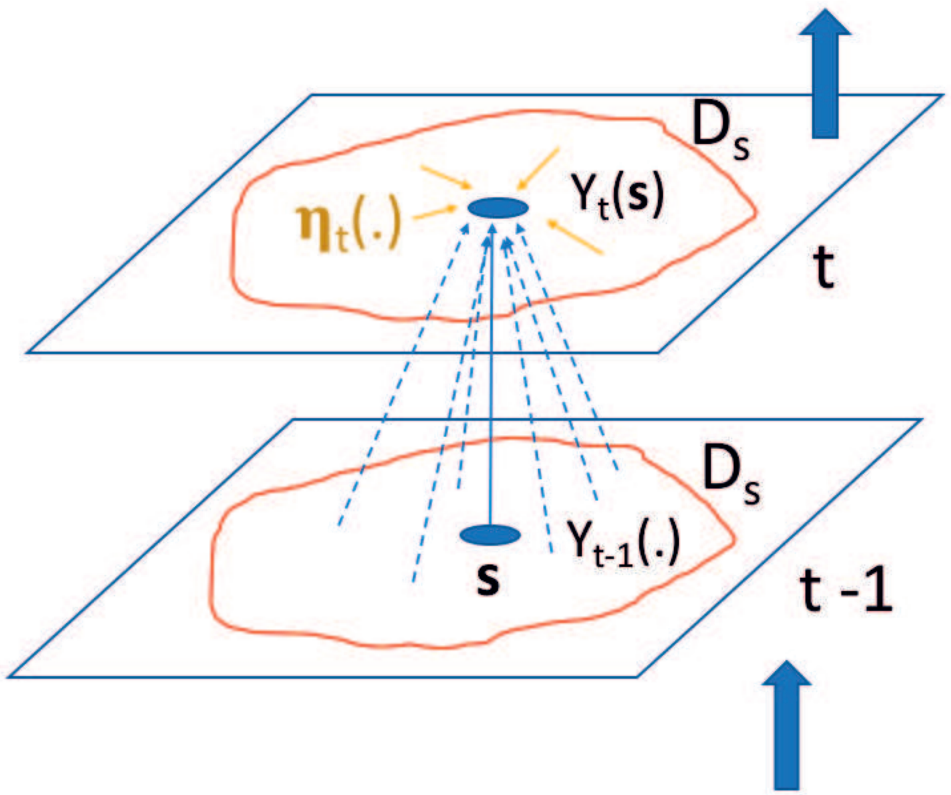


Figure 5.1: Cartoon illustration of a linear DSTM. The process at spatial location s and time t, $Y_t(s)$, is constructed from a linear combination of the process values at the previous time, $Y_{t-1}(\cdot)$, plus an "instantaneous" random spatial error process, $\eta_t(\cdot)$. The thick arrows indicate the passage from past to present to future.

- シンプルなDSTMの場合
  - ランダムウォーク　($\mathbf{M} = \mathbf{I}$)
  - 空間的に均一なARプロセス　($\mathbf{M} = \theta_p\mathbf{I}$)
  - 空間的に不均一なARプロセス　($\mathbf{M} = \mathrm{diag}(\boldsymbol{\theta}_p)$)

5.3.1 Parameter Dimension Reduction

As an example of the last parameterization described above, consider the process model where $\mathbf{C}_\eta = \sigma_\eta^2 \mathbf{I}$, and $\mathbf{M} = \text{diag}(\boldsymbol{\theta}_p)$. We can decompose the first-order conditional distributions in this case as

$$[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_p, \sigma_\eta^2] = \prod_{i=1}^{n} [Y_t(\mathbf{s}_i) | Y_{t-1}(\mathbf{s}_i), \theta_p(i), \sigma_\eta^2], \quad t = 1, 2, \ldots.$$

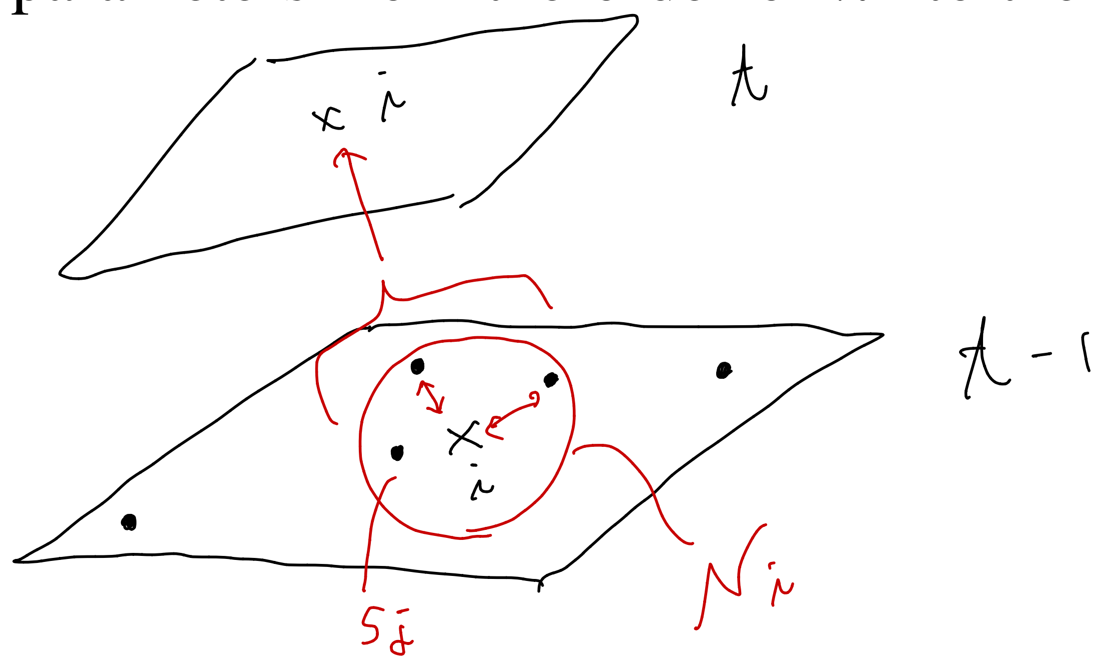
Thus, conditional on the parameters $\boldsymbol{\theta}_p = (\theta_p(1), \ldots, \theta_p(n))'$, we have spatially independent univariate AR(1) processes at each spatial location (i.e., only the Y-value at the previous time at the same spatial location influences the transition). However, if $\boldsymbol{\theta}_p$ is *random* and has spatial dependence, then if we integrate it out, the marginal conditional distribution, $[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \sigma_\eta^2]$, can imply that all of the elements of $\mathbf{Y}_{t-1}$ influence the transition to time t at all spatial locations (i.e., this is a non-separable spatio-temporal process). Recall

どーゆーこと?

パラメータの数を減らす　lagged-nearest-neighbor representations

$$Y_t(\mathbf{s}_i) = \sum_{\mathbf{s}_j \in \mathcal{N}_i} m_{ij} Y_{t-1}(\mathbf{s}_j) + \eta_t(\mathbf{s}_i), \tag{5.13}$$

where $\mathcal{N}_i$ corresponds to a pre-specified neighborhood of the location $\mathbf{s}_i$ (including $\mathbf{s}_i$), for $i = 1, \ldots, n$, and where we specify $m_{ij} = 0$, for all $\mathbf{s}_j \notin \mathcal{N}_i$. Such a parameterization reduces the number of free parameters from the order of n^2 to the order of n.



- transition coefficientsをパラメータ化すると，メカニスティックな理解（decay rateとasymmetry）が可能に
- Appendix D.1を使うと，

$$\mathbf{Y}_t = \underline{\mathbf{M}(\boldsymbol{\theta}_p)}\mathbf{Y}_{t-1} + \mathbf{M}_b(\boldsymbol{\theta}_p)\mathbf{Y}_{b,t} + \boldsymbol{\eta}_t,$$

5つの対角要素を持つ推移行列

- spread rateに関する関数×2
- advectionに関する関数×2
- 離散化パラメータ×1

5.3.2 Dimension Reduction in the Process Model

the first basis-expansion　　residual basis-expansion

$$Y_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta} + \sum_{i=1}^{n_\alpha} \phi_i(\mathbf{s})\alpha_{i,t} + \sum_{j=1}^{n_\xi} \psi_j(\mathbf{s})\xi_{j,t} + \nu_t(\mathbf{s}), \tag{5.15}$$

固定効果　　基底関数（空間）　　ランダム効果（時間）　　小スケールのランダム効果（時間）

the vector form of (5.15):

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\Psi}\boldsymbol{\xi}_t + \boldsymbol{\nu}_t, \tag{5.16}$$

The evolution of the latent process $\{\boldsymbol{\alpha}_t\}$ can proceed according to the linear equations involving a transition matrix, discussed earlier. For example, one could specify a first-order vector autoregressive model (VAR(1)),

$$\boldsymbol{\alpha}_t = \mathbf{M}_\alpha \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \tag{5.17}$$

where $\mathbf{M}_\alpha$ is the $n_\alpha \times n_\alpha$ transition matrix, and $\boldsymbol{\eta}_t \sim Gau(\mathbf{0}, \mathbf{C}_\eta)$

- ランダムウォーク　($\mathbf{M} = \mathbf{I}$)
- 空間的に均一なARプロセス　($\mathbf{M} = \theta_p\mathbf{I}$)
- nearest-neighbor model　($\mathbf{M} = \mathrm{diag}(\boldsymbol{\theta}_p)$)　　を使ってさらにパラメータを減らす

5.3.2 Dimension Reduction in the Process Model

the first basis-expansion　residual basis-expansion

$$Y_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta} + \sum_{i=1}^{n_\alpha} \phi_i(\mathbf{s})\alpha_{i,t} + \sum_{j=1}^{n_\xi} \psi_j(\mathbf{s})\xi_{j,t} + \nu_t(\mathbf{s}), \tag{5.15}$$

固定効果

基底関数
(空間)

ランダム効果
(時間)

小スケールのランダム効果
(時間)

the vector form of (5.15):

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\Psi}\boldsymbol{\xi}_t + \boldsymbol{\nu}_t, \tag{5.16}$$

DSTMでは，空間的な交互作用を扱える基底関数が良い -> **Lab. 5.3 (基底関数に直行関数を使う)**

Empirical orthogonal functions (EOFs) are ideal basis functions to use in this case, since they capture most of the variability in the observed signal, by design. In this Lab we look at the SST data set, take the EOFs that we generated in Lab 2.3, and estimate all unknown parameters, first within a classical time-series framework based on a vector autoregression and using the method of moments (see Appendix C.1), and then in a state-space framework using the EM algorithm (see Appendix C.2).

Lab 5.3

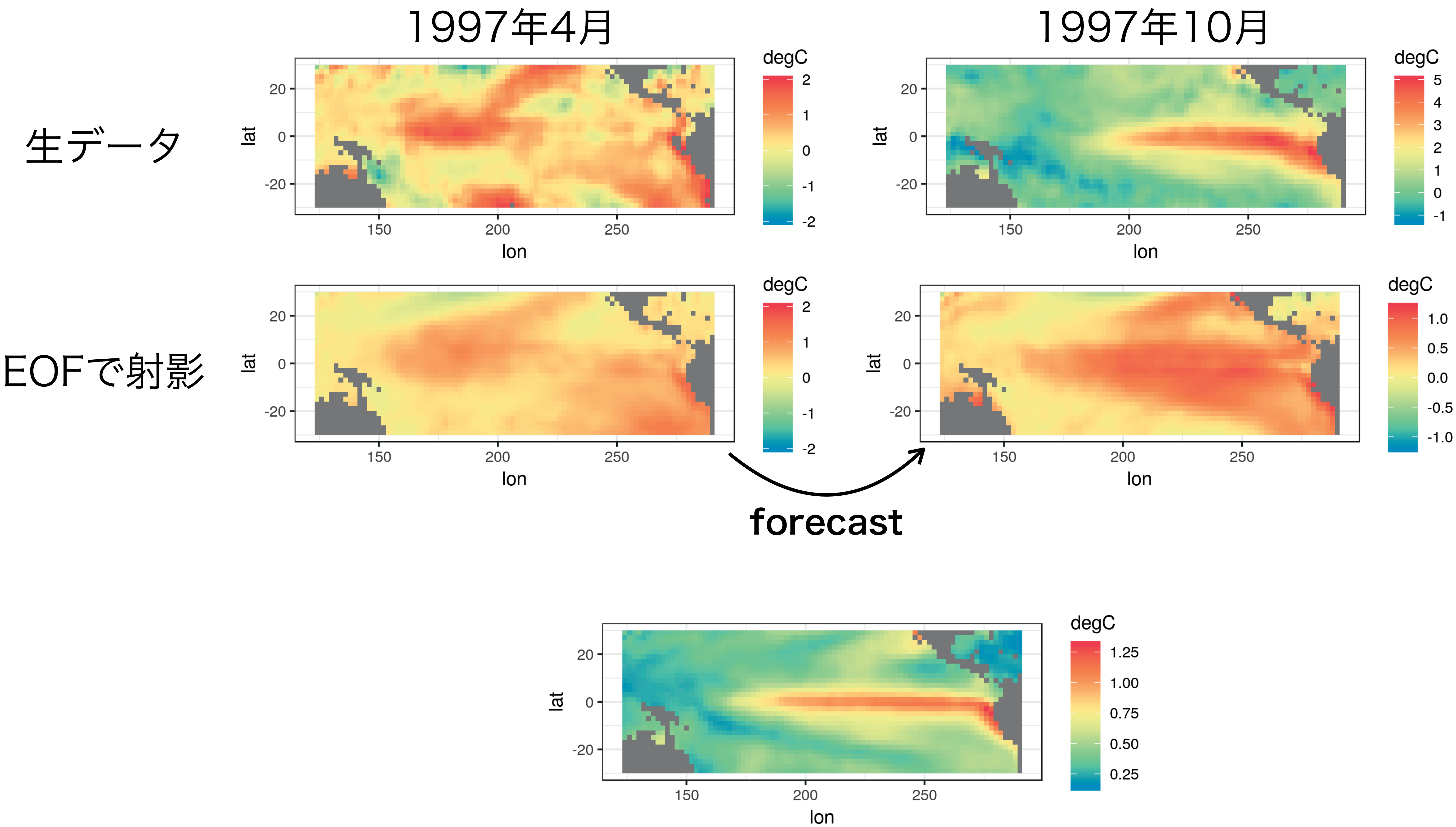
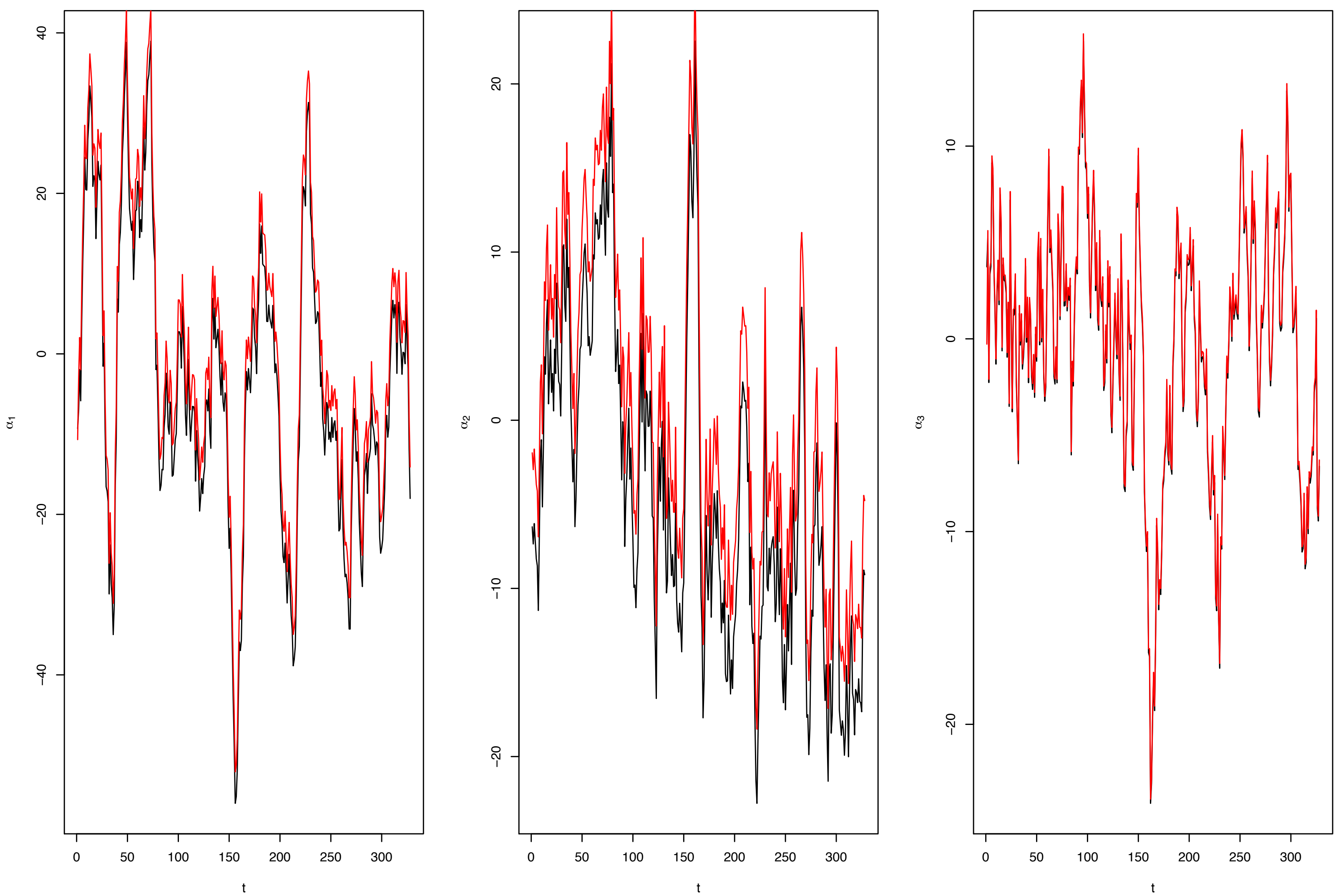


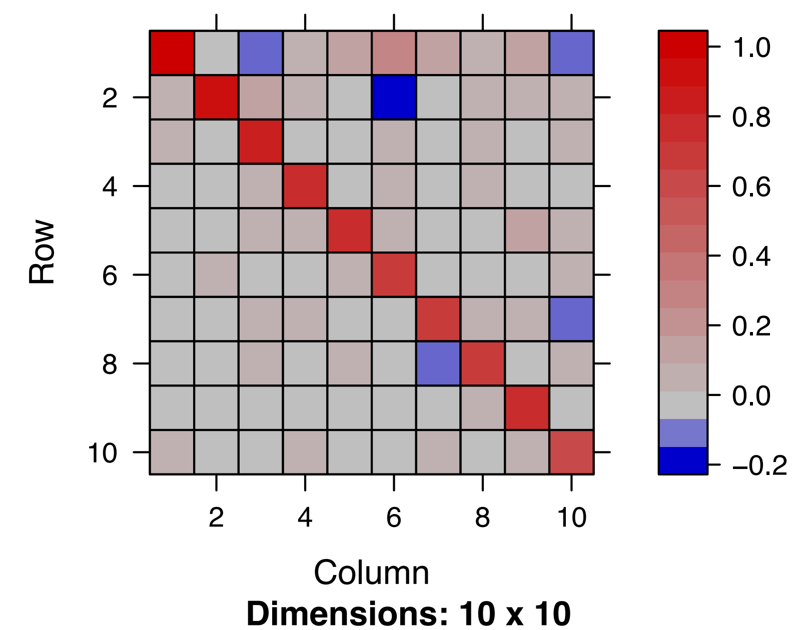
Figure 5.7: Top: SST data for April 1997 (left) and October 1997 (right). Middle: the EOF projection for April 1997 (left), and the forecast for October 1997 (right). Note the different color scales for the predictions (up to 1°C) and for the observations (up to 5°C). Bottom: Prediction standard errors for the forecast.

The function `DSTM_EM`, provided with the package **STRbook**, runs the EM algorithm that carries out maximum likelihood estimation in a state-space model. The function takes the data `Z`, the initial covariance $\mathbf{C}_0$ in `Cov0`, the initial state $\boldsymbol{\mu}_0$ in `muinit`, the evolution operator $\mathbf{M}$ in `M`, the covariance matrix $\mathbf{C}_\eta$ in `Ceta`, the measurement-error variance σ_ϵ^2 in `sigma2_eps`, the matrix $\mathbf{H}$ in `H`, the maximum number of EM iterations in `itermax`, and the tolerance in `tol` (the tolerance is the smallest change in the log-likelihood, multiplied by 2, required across two consecutive iterations of the EM algorithm, before terminating). All parameters supplied to the function need to be initial guesses (usually those from the method of moments suffice); these will be updated using the EM algorithm.

```
DSTM_Results <- DSTM_EM(Z = SSTdata,
                        Cov0 = Cov0,
                        muinit = matrix(0, n, 1),
                        M = Mest,
                        Ceta = Ceta,
                        sigma2_eps = 0.1,
                        H = H <- E$v[, 1:n],
                        itermax = 10,
                        tol = 1)
```

```r
par(mfrow = c(1,3))
for(i in 1:3) {
  plot(DSTM_Results$alpha_smooth[i, ], type = 'l',
       xlab = "t", ylab = bquote(alpha[.(i)]))
  lines(TS[, i], lty = 'dashed', col = 'red')
}
```





推定した推移行列

赤: momentsで推定したα，黒:EMで推定で推定したα

で，似てるのか似てないのか，どう解釈すれば良いのかは不明

model. These two models and their respective inferences can be expected to differ when there is more nonlinearity in the process and/or the data are less complete in space and time.

In our concluding remarks, we remind the reader that in this Lab we considered a *linear* DSTM for modeling SSTs. Recent research has suggested that *nonlinear* DSTMs may provide superior prediction performance; see the case study in Appendix F.

→ 5.4に続く.

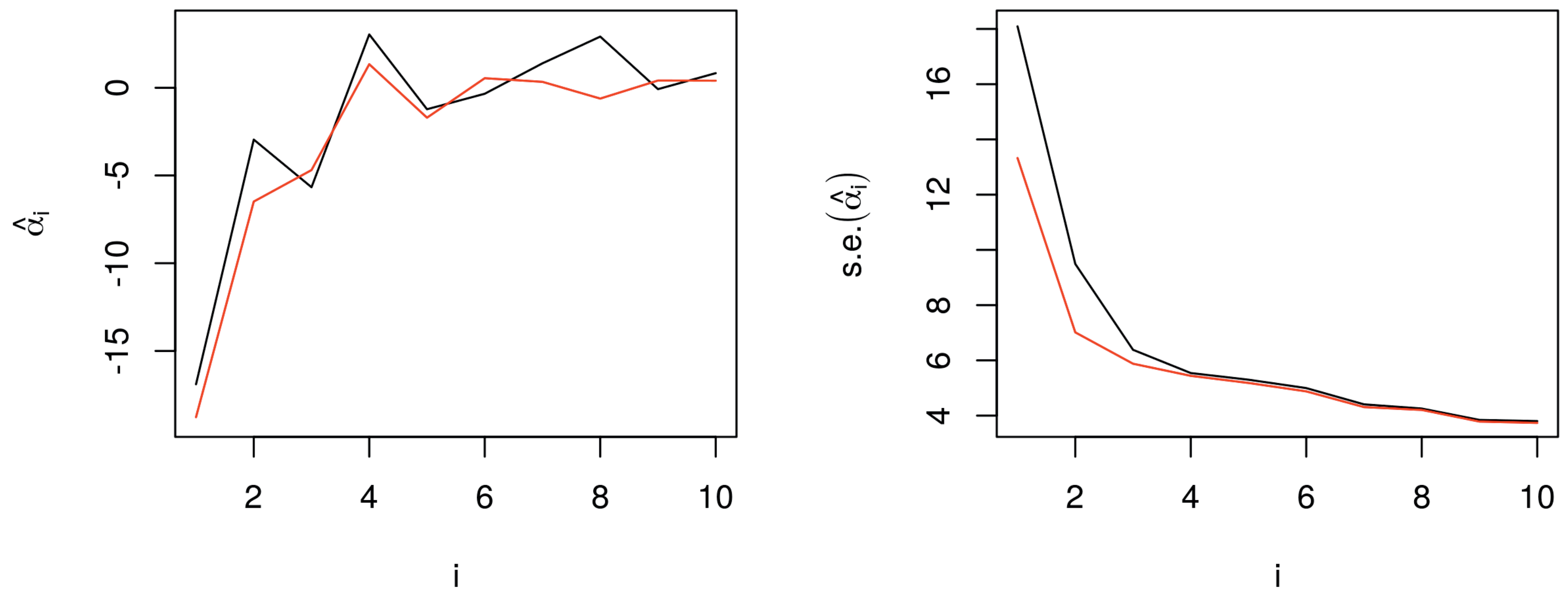


Figure 5.10: Forecasts (left) and prediction standard errors (right) for the EOF coefficients in October 1997 using a lag-6 time-series model estimated using the method of moments (black) and a lag-1 state-space model estimated using the EM algorithm (red).

5.4 Nonlinear DSTMs

$$Y_t(\cdot) = \mathcal{M}(Y_{t-1}(\cdot), \eta_t(\cdot); \boldsymbol{\theta}_p), \quad t = 1, 2, \ldots, \tag{5.18}$$

where $\mathcal{M}$ is a nonlinear function that models the process transition from time $t-1$ to t, $\eta_t(\cdot)$ is an error process, and $\boldsymbol{\theta}_p$ are parameters.

この節で紹介されるモデル

- **state-dependent model**
  - threshold vector autoregressive model
- **general quadratic nonlinear DSTM (連続・離散)**
- **その他: deep learning**
  - convolutional and recurrent neural networks (RNNs): 計算が大変
  - echo state network (ESN): RNNsより計算が楽
    c.f. McDermott & Wikle 2017, Appendix F
  - agent-based model (individual-based model)
    c.f. Cressie & Wikle 2011 (Section7.3.4), Wikle & Hooten (2016)

state-dependent modelの一般式

$$\mathbf{Y}_t = \mathbf{M}(\mathbf{Y}_{t-1};\boldsymbol{\theta}_p)\,\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \tag{5.19}$$

threshold vector autoregressive model

$$\mathbf{Y}_t = \begin{cases} \mathbf{M}_1\mathbf{Y}_{t-1} + \boldsymbol{\eta}_{1,t}\,, & \text{if } f(\omega_t) \in d_1, \\ \quad\vdots & \quad\vdots \\ \mathbf{M}_K\mathbf{Y}_{t-1} + \boldsymbol{\eta}_{K,t}\,, & \text{if } f(\omega_t) \in d_K, \end{cases} \tag{5.20}$$

where $f(\omega_t)$ is a function of a time-varying parameter ω_t that can itself be a function of the process, $\mathbf{Y}_{t-1}$, in which case it is a state-dependent model. We implicitly assume that conditions on the right-hand side of (5.20) are mutually exclusive; that is, $d_1, \ldots, d_K$ are disjoint. A simpler threshold model results if the parameters $\{\omega_t\}$ do not depend on the process.

general quadratic nonlinearity

reaction-diffusion PDE

$$\frac{\partial Y}{\partial t} = \frac{\partial}{\partial x}\left(\delta \frac{\partial Y}{\partial x}\right) + Y \exp\left(\gamma_0 \left(1 - \frac{Y}{\gamma_1}\right)\right), \tag{5.21}$$

where the first term corresponds to a diffusion (spread) term that depends on a parameter δ, and the second term corresponds to a density-dependent (Ricker) growth term with growth parameter γ_0 and carrying capacity parameter γ_1. More generally, each of these parameters could vary with space and/or time. Notice that the diffusion term is linear in Y but the density-dependent growth term is nonlinear in that it is a function of Y multiplied by a non-linear transformation of Y. This can be considered a general case of a quadratic interaction.

離散時間・空間

$$Y_t(\mathbf{s}_i) = \sum_{j=1}^{n} m_{ij} Y_{t-1}(\mathbf{s}_j) + \sum_{k=1}^{n} \sum_{\ell=1}^{n} b_{i,k\ell}\, g(Y_{t-1}(\mathbf{s}_\ell); \boldsymbol{\theta}_g)\, Y_{t-1}(\mathbf{s}_k) + \eta_t(\mathbf{s}_i), \tag{5.22}$$

where m_{ij} are the linear-transition coefficients seen previously, and the quadratic-interaction transition coefficients are denoted by $b_{i,k\ell}$. Importantly, a transformation of one of the components of the quadratic interaction is included through the function $g(\cdot)$, which can depend on parameters $\boldsymbol{\theta}_g$. This function $g(\cdot)$ is responsible for the term "general" in GQN, and such transformations are important for many processes such as density-dependent growth that one may see in an epidemic or invasive-species population processes (see, for example, (5.21) above), and they can keep forecasts from "blowing up" in time.