Innovation Approach to the Identification of Causal Models in Time Series Analysis

T. Ozaki
Institute of Statistical Mathematics
Innovation Approach
(Wold, Kolmogorov, Wiener, Kalman, Kailath)
(Box-Jenkins, Akaike etc.)

Causal Model

\[
\begin{align*}
\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N \\
\mathbf{z}_t &= f(\mathbf{z}_{t-1}) + \mathbf{\varepsilon}_t \\
\frac{d\mathbf{z}(t)}{dt} &= f(\mathbf{z}(t)) \\
d\mathbf{z}(t) &= f(\mathbf{z}(t))dt + d\mathbf{w}(t)
\end{align*}
\]

What causes the time-dependency in geophysical time series?

Geophysical Dynamical System
Three topics in time series

1. Nonlinear time series
   - Dynamical Systems

2. Non-Gaussian time series
   - Dynamical Systems
   - Shot Noise

3. Spatial time series
   - Spatial Dynamics

Innovation Approach
Dynamical System & Time Series Model

Ozaki & Oda (1976)

Restoring force \( W \cdot GM \sin x \)

\[
W \cdot GM \sin x \approx x \quad : \text{for small } x
\]

\[
W \cdot GM \sin x \approx x - \frac{1}{6} x^3 \quad : \text{for large } x
\]

\[
\ddot{x}(t) + c\dot{x}(t) + \alpha x(t) = n(t)
\]

\[
\ddot{x}(t) + c\dot{x}(t) + \alpha x(t) + \beta x^3(t) = n(t)
\]
ExpAR Model

1. Smaller prediction errors than AR models.

\[ x_t = \sum_{i=1}^{p} \{\phi_{i,0} + \phi_{i,1} \exp(-\gamma x_{t-1}^2)\} x_{t-i} + \epsilon_t \]

\[ \epsilon_t = x_t - \sum_{i=1}^{p} \{\phi_{i,0} + \phi_{i,1} \exp(-\gamma x_{t-1}^2)\} x_{t-i} \]

2. Mechanical interpretation

Ozaki & Oda (1976)
Natural frequency

\[ \ddot{x}(t) + c \dot{x}(t) + \alpha x(t) = n(t) \]

\[ x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + n_t \]

\[ n(t) \rightarrow x(t) \]

\[ n_t \rightarrow x_t \]

\[ \rho^2 + c \rho + \alpha = 0 \]

\[ \lambda^2 - \phi_1 \lambda - \phi_2 = 0 \]

\[ \omega_0 = \frac{1}{2\pi} \sqrt{\alpha - c^2 / 4} \]

\[ f_0 = \frac{1}{2\pi} \tan^{-1} \left( \sqrt{-4\phi_2 - \phi_1^2} / \phi_1 \right) \]

\[ p(f) = \frac{\sigma^2}{\left| 1 - \phi_1 e^{-i2\pi f} - \phi_2 e^{-i2\pi f/2} \right|^2} \]
Idea: Dynamic eigen-values

1. Duffing equation

\[ \ddot{x}(t) + c \dot{x}(t) + \alpha x(t) + \beta x^3(t) = n(t) \]

2. van der Pol equation

\[ \ddot{x}(t) + c \{ x^2(t) - 1 \} \dot{x}(t) + \alpha x(t) = n(t) \]

\[
\begin{bmatrix}
  x_t \\
x_{t-1}
\end{bmatrix} =
\begin{bmatrix}
  \phi_1(x_{t-1}) & \phi_2(x_{t-1}) \\
  1 & 0
\end{bmatrix}
\begin{bmatrix}
  x_{t-1} \\
x_{t-2}
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_t \\
  0
\end{bmatrix}
\]

\[
\phi_1(x_{t-1}) = \phi_{1,0} + \phi_{1,1} \exp(-\gamma x_{t-1}^2)
\]

\[
\phi_2(x_{t-1}) = \phi_{2,0} + \phi_{2,1} \exp(-\gamma x_{t-1}^2)
\]
Make it non-explosive!

\[ x_t = (\phi_{1,0} + \phi_{1,1} x_{t-1}^2)x_{t-1} + \phi_2 x_{t-2} + \epsilon_t \]

Make it stay inside for large \( x_{t-1} \)!

\[ x_t = \{\phi_{1,0} + \phi_{1,1} \exp(-\gamma x_{t-1}^2)\}x_{t-1} + \phi_2 x_{t-2} + \epsilon_t \]
ExpAR Singular points

Ozaki (1985)

Stable singular point at $x = \pm 0.8355$

Unstable singular point $x = \pm 1.7308$

Stable singular point $x = \pm 0.925$

Unstable singular point $x = \pm 1.1606$
ExpAR-Chaos

Ozaki (1985)

\[ x_t = (0.5 - 9 \cdot e^{-x_t^2}) x_{t-1} \]

\[ x_t = (1 - 15 \cdot e^{-x_t^2}) x_{t-1} - (10.25 - 92.5 \cdot e^{-x_t^2}) x_{t-2} \]

\[ x_t = (0.5 - 16.5 \cdot e^{-x_t^2}) x_{t-1} - (10.25 - 92.5 \cdot e^{-x_t^2}) x_{t-2} \]
ExpAR Models & non-Gaussian distributions

\[ x_t = \phi(x_{t-1}) x_{t-1} + \epsilon_t \]

\[ x_t = \begin{cases} 
0.8 x_{t-1} + \epsilon_t & \text{for } |x_{t-1}| \geq 1 \\
(0.8 + 1.3 x_{t-1}^2 - 1.3 x_{t-1}^4 + \epsilon_t) & \text{for } |x_{t-1}| \leq 1 
\end{cases} \]

- \( \xi_0, \xi_+^-, \xi_-^+ \): Stable singular points
- \( \xi_-^-, \xi_+^- \): Unstable singular points
Distribution of ExpAR process

\[ x_{t+1} = \{1 - 0.2 \exp(-x_t^2)\}x_t + n_{t+1} \]

\[ x_{t+1} = 0.8 + 0.4\exp(-x_t^2)x_t + n_{t+1} \]

\[ x_{t+1} = 0.8 + 0.2\exp(-x_t^2)x_t + n_{t+1} \]
Causal Models in discrete time and continuous time

\[
dx(t) / dt = f(x(t))
\]

\[
dx(t) = f(x(t))dt + dw(t)
\]

\[
x_t = \phi(x_{t-1})x_{t-1}
\]

\[
x_t = \phi(x_{t-1})x_{t-1} + \varepsilon_t
\]
Time discretizations of
\[ dx = f(x)dt + dw(t) \]

Euler scheme
Heun scheme
Runge-Kutta scheme

Explosive nonlinear AR model!

\[ x_{t+\Delta t} = p(x_t)x_t + \sqrt{\Delta t}w_{t+\Delta t} \]

Any non-explosive scheme?
L.L. scheme

\[ x_{t+\Delta t} = \exp(K_t \Delta t)x_t + \sqrt{\Delta tw_{t+\Delta t}} \]

\[ K_t = \log[1 + J_t^{-1} \{ \exp(J_t \Delta t) - 1 \} f(x_t)] / x_t \]

\[ J_t = \left( \frac{\partial f(x)}{\partial x} \right)_{x=x_t} \]

ii). Ozaki(1985)
iii). Biscay et al.(1996)

Characteristics
1. Simple
2. A-stable
Examples of $\text{Exp}(K_t \Delta_t)$

\[ \dot{x}(t) = f(x) + n(t) \]

\[ x_{t+\Delta_t} = \exp(K_t \Delta_t)x_t + \sqrt{\Delta_t} \nu_{t+\Delta_t} \]

\[ x_{t+1} = \{0.8 + 0.2 \exp(-x_t^2)\}x_t + n_{t+1} \]

\[ x_{t+1} = \{1 - 0.2 \exp(-x_t^2)\}x_t + n_{t+1} \]
Innovation Approach
(Wold, Kolmogorov, Wiener, Kalman, Kailath)
(Box-Jenkins, Akaike etc.)

\[ x_1, x_2, x_3, \ldots, x_N \]

\[
\begin{align*}
  z_t &= f(z_{t-1}) + \varepsilon_t \\
  \frac{dz(t)}{dt} &= f(z(t)) \\
  dz(t) &= f(z(t))dt + dw(t)
\end{align*}
\]

Three types of models

Data

\[ x_1, x_2, x_3, \ldots, x_N \]

Time Series Models (ExpAR, neural net etc.)

\[ x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-k}) + \varepsilon_t \]

Dynamic Phenomena

Dynamical Systems

\[ \frac{dz(t)}{dt} = f(z(t)) \]

Stochastic Dynamical Systems

\[ dz(t) = f(z(t))dt + dw(t) \]
Applications

1. Non-Gaussian time series and nonlinear dynamics

2. Estimation of $dx = f(x)dt + dw(t)$
   $dx/dt = f(x)$

3. RBF-Neural Net vs ExpAR, RBF-AR modeling

4. Spatial time series modeling
Non-Gaussian time series and nonlinear dynamics

Does non-Gaussian-distributed time series mean non-Gaussian prediction errors? Not Necessarily!
Distribution of ExpAR process

\[ n_{t+1} \rightarrow \text{ExpAR} \rightarrow x_{t+1} \]

Gaussian white noise

non-Gaussian

\[ x_{t+1} = \{0.8 + 0.4 \exp(-x_t^2)\} x_t + n_{t+1} \]

\[ x_{t+1} = \{0.8 + 0.2 \exp(-x_t^2)\} x_t + n_{t+1} \]
Same Distribution
Different Dynamics

\[ W(x) = \frac{x^{\alpha-1} \exp(-x/\beta)}{\Gamma(\alpha) \Gamma(\beta)} \]

i)
\[
x(t) = \frac{\beta \ y^2(t)}{2}
\]
\[
\dot{y}(t) = \frac{\alpha / 2}{y} - \frac{y}{2} + n(t)
\]

ii)
\[
x(t) = e^{\sqrt{2\beta} y(t)}
\]
\[
\dot{y}(t) = \frac{\alpha \beta}{\sqrt{2\beta}} - e^{\sqrt{2\beta} y} + n(t)
\]

iii)
\[
x(t) = e^{\sqrt{2\beta} y(t)}
\]
\[
\ddot{y} + a \dot{y} - \frac{\alpha \beta}{\sqrt{2\beta}} + e^{\sqrt{2\beta} y} = n(t)
\]

\[ \alpha=3, \beta=1 \]

Ozaki(1985, 1990)
Mechanism

Gamma-distributed Process (Type II) (Ozaki, 1985, 1992)

\[ \frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} [(\alpha + 1) \beta x - x^2] p + \frac{1}{2} \frac{\partial^2}{\partial x^2} [2\beta x^2 p] \]

\[ p(x) = \frac{1}{\Gamma(\alpha) \Gamma(\beta)} x^{\alpha-1} e^{-x/\beta} \]

\[ x(t) = e^{\sqrt{2\beta}y(t)} \]

\[ dy = \left\{ \frac{\alpha \beta}{\sqrt{2\beta}} - e^{\sqrt{2\beta}y} \right\} dt + dw(t) \]

- \text{x(t) is generated from}
  - i) \text{Gaussian white noise}
  - ii) \text{Nonlinear Dynamics}
  - iii) \text{Variable Transformation}
This implies the validity of innovation approach.

Non-Gaussian time series

\[ x_1, x_2, x_3, \ldots, x_N \]

\[ y_t = h^{-1}(x_t) \]

\[ \varepsilon_t = y_t - f(y_{t-1}) \]

\[ \varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_N \]
When residuals of your model are non-Gaussian looking, what would you do?

1. Introduce non-Gaussian noise model

2. Improve the Causal Model so that it produces Gaussian residuals
ExpAR model is not sufficient

\[ x(t) = \sum_{k=1}^{p} \phi_k(x(t-1))x(t-k) + \varepsilon_t \]

More complicated dynamics, i.e. \( \phi_k(x(t-1)) \)
\[ \rightarrow \]
RBF-AR

y-dependent system characteristics, i.e. \( \phi_k(x(t-1), y(t-1)) \)
\[ \rightarrow \]
RBF-ARX
RBF-AR & RBF Neural Net

More complicated $\phi_k(x(t-1))$

RBF-AR($p,d,m$)

\[ x(t) = \phi_0(X(t-1)) + \sum_{i=1}^{p} \phi_i(X(t-1))x(t-i) + \epsilon(t) \]

\[ X(t-1) = [x(t-1), \ldots, x(t-d)]' \]

\[ \phi_i(X(t-1)) = c_{i,0} + \sum_{k=1}^{m} c_{i,k} \exp\{ -\lambda_k \|X(t-1) - Z_k\|_2^2 \} \]

RBF- Neural Net ($p,d,m$)

\[ x_t = \phi_0(X(t-1)) + \epsilon_t \]

\[ \phi_0(X(t-1)) = c_{0,0} + \sum_{k=1}^{m} c_{0,k} \exp\{ -\lambda_k \|X(t-1) - Z_k\|_2^2 \} \]
Application - (2)

Estimation of

\[
\frac{dz}{dt} = f(z(t))
\]

\[
dz(t) = f(z(t))dt + dw(t)
\]

Numerical examples

1. Rikitake chaos, (Geophysics)
2. Zetterberg Model (Brain Science)
3. Dynamic Market model (Finance)
How to identify?

\[ dz(t) = f(z(t))dt + dw(t) \]

from

\[ x_1, x_2, x_3, \ldots, x_N \]
State Space Formulation

\[ x_1, x_2, x_3, \ldots, x_N \]

\[ dz = f(z)dt + g(z)dw(t) \]

\[ x_t = H z_t + e_t \]
Frost & Kailath (1971)'s theorem

\[ dz = f(z | \theta) dt + g(z | \theta) dw(t) \]

\[ x_t = C z_t + \varepsilon_t \]

\[ v_k = x_k - E[x_k | x_{k-1}, \ldots, x_1] \]

\[ v_k \rightarrow v(t) \quad \Delta t \rightarrow 0 \]

Non-Gaussian time series

\[ x_1, x_2, x_3, \ldots, x_N \]

Gaussian white noise

\[ \mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3, \ldots, \mathcal{N}_N \]

Nonlinear Filter

\[ \mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3, \ldots, \mathcal{N}_N \]

: Gaussian white noise
Likelihood Calculation

Innovation Approach

\[ v_k = x_k - E[x_k \mid x_{k-1}, \ldots, x_1] \]

\[ \Delta t \to 0 \quad v_k \to v(t) \]

: Gaussian white noise

Frost & Kailath (1971)

\[
\log p(x_1, \ldots, x_N \mid \theta) = \sum_k \log p(x_k \mid x_{k-1}, \ldots, x_1, \theta) \\
= \sum_k \log p(v_k \mid x_{k-1}, \ldots, x_1, \theta) \\
= \left( -\frac{1}{2} \right) \sum_{t=1}^{N} \left\{ \log \sigma_{\nu_t}^2 + \frac{\nu_t^2}{\sigma_{\nu_t}^2} \right\} 
\]

How to obtain \( \nu_t \) and \( \sigma_{\nu_t}^2 \) ?

\[
v_k = x_k - E[x_k \mid x_{k-1}, \ldots, x_1] \\
= x_k - \int \xi_k p(\xi_k \mid x_{k-1}, \ldots, x_1) d\xi_k
\]

Nonlinear Filter
Relations to Jazwinski (1970)’s scheme

\[
\log p(x_1, \ldots, x_N | \theta) = \sum \log p(x_k | x_{k-1}, \ldots, x_1, \theta)
\]

\[
p(x_k | x_{k-1}, \ldots, x_1, \theta) = \int p(x_k | z_k) p(z_k | x_{k-1}, \ldots, x_1, \theta) dz_k
\]

\[
p(z_k | x_{k-1}, \ldots, x_1, \theta) = \int p(z_k | z_{k-1}) p(z_{k-1} | x_{k-1}, \ldots, x_1, \theta) dz_{k-1}
\]

\[
p(z_{k-1} | x_{k-1}, \ldots, x_1, \theta) = \frac{p(x_{k-1} | z_{k-1}) p(z_{k-1} | x_{k-2}, \ldots, x_1, \theta)}{\int p(x_{k-1} | \zeta_{k-1}) p(\zeta_{k-1} | x_{k-2}, \ldots, x_1, \theta) d\zeta_{k-1}}
\]

Innovation Approach

Calculate \( p(x_k | z_k) \) \& \( p(z_k | z_{k-1}) \) etc. by Local Gauss model.

\[
z_t = F_{t-1} z_{t-1} + G_{t-1} n_t
\]

\[
x_t = H z_t + e_t
\]

\[
dz / dt = f(z | \theta) + n(t)
\]

\[
x_t = C z_t + e_t
\]
Two Choices for Approximation

1) Local Gauss

\[
\begin{align*}
  z_t &= F_{t-1}z_{t-1} + G_{t-1}n_t \\
  x_t &= Hz_t + e_t
\end{align*}
\]

\[
p(z_k \mid z_{k-1}, \theta) = \mathcal{N}(F_{t-1}z_{t-1} + G_{t-1}G_{t-1}'n, G_{t-1}'G_{t-1})
\]

\[
p(x_k \mid z_k, \theta) = \mathcal{N}(Hz_t, \sigma_e^2)
\]

2) Local non-Gauss

Use of Fokker-Planck equation.

\[
\frac{\partial p}{\partial t} = \sum f_i(z) \frac{\partial p}{\partial z_i} + \frac{1}{2} \sum \sum \sigma_{i,j} \frac{\partial^2 p}{\partial z_i \partial z_j}
\]

\[
p(z_k \mid z_{k-1})
\]

Computationally 1) is super more efficient than 2)
Advantages of the L.L. Scheme

See

B.L.S. Prakasa Rao (1999)
*Statistical Inference for Diffusion Type Processes*

H. Schurz (1999)
*A Brief Introduction To Numerical Analysis of (Ordinary) Stochastic Differential Equations Without Tears*

Essence is

Stability & Efficiency
Numerical examples

1. Rikitake chaos (Geophysics)
2. Zetterberg Model (Brain Science)
3. Dynamic Market model (Finance)
Identification of the chaotic Rikitake model  
(Ozaki et al. 2000)

Stochastic Rikitake model
\[ d\xi = (-\theta_1 \xi + \xi \eta)dt + \sigma_w^2 dw \]
\[ d\eta = (\theta_2 \xi - \theta_1 \eta + \xi \xi)dt \]
\[ d\xi = (1 - \xi \eta)dt \]

observation equation
\[ z_{tk} = \xi(t_k) + e_{tk} \]

Rikitake(1957), Ito(1982)

<table>
<thead>
<tr>
<th>Chaos</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \sigma_w^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>124.8</td>
<td>0</td>
</tr>
</tbody>
</table>
Simulated and estimated states

<table>
<thead>
<tr>
<th></th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\sigma_w^2$</th>
<th>$\xi(0)$</th>
<th>$\eta(0)$</th>
<th>$\zeta(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>actual</td>
<td>5</td>
<td>124.8</td>
<td>0.0025</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>estimated</td>
<td>4.95</td>
<td>123.1</td>
<td>0.0060</td>
<td>1.002</td>
<td>0.087</td>
<td>0.003</td>
</tr>
</tbody>
</table>
Innovation of the estimated model

P > 0.95
Three types of parameters

1. $\theta_1, \theta_2$

2. $\sigma_w^2, \sigma_e^2$

3. $\xi(0), \eta(0), \zeta(0)$

Stochastic Rikitake model

$\frac{d\xi}{dt} = (-\theta_1 \xi + \xi \eta) dt + \sigma_w^2 \, dw$

$\frac{d\eta}{dt} = (\theta_2 \xi - \theta_1 \eta + \xi \xi) dt$

$\frac{d\xi}{dt} = (1 - \xi \eta) dt$

Observation equation

$z_{tk} = \xi(t_k) + e_{tk}$

Structured-parameter optimization for M.L.E. method

Peng & Ozaki (2001)
Initial values & Estimated States

\[ \theta_1, \theta_2 \quad \sigma^2_w, \sigma^2_e \quad : \text{optimized} \]
\[ \xi(0), \eta(0), \zeta(0) \quad : \text{not optimized} \]

\[ \theta_1, \theta_2 \quad \sigma^2_w, \sigma^2_e \quad : \text{optimized} \]
\[ \xi(0), \eta(0), \zeta(0) \quad : \text{optimized} \]
Initial values & Innovations

\( \theta_1, \theta_2 \quad \sigma^2_w, \sigma^2_e \): optimized
\( \xi(0), \eta(0), \zeta(0) \): not optimized

\( \xi(0), \eta(0), \zeta(0) \): optimized
Reality in Data Analysis

1. Zero prediction errors are not possible to attain even with nonlinear causal (chaos) models.

2. Residuals of ARIMA models are usually almost Gaussian, but not always.

- Time - inhomogeneous residuals
- Gaussian white residuals + a few outliers

- Generally distributed residuals are rare! (We don’t see bi-modally distributed residuals)
A question is whether it is possible to find a perfect deterministic model for the data $x_1, x_2, \ldots, x_N$
Application - (4)

Spatial time series modeling
Example: Data assimilation in meteorology

\[ x_t^{(i,j)} \quad (t = 1, 2, \ldots, n) \]

States on the lattice points \((i,j)\)

Observations

\[ y_t = H\{x_t^{(1,1)}, \ldots, x_t^{(N,N)}\} + \varepsilon_t \quad (t = 1, 2, \ldots, n) \]

Estimation

\[ p = \text{dim}(y_t) \ll \text{dim}(x_t) = N \times N \]
Mutual understanding : on the way

♦ Variational method (4D-Var)

\[
I(x_0) = \frac{1}{2} \sum_{i=0}^{n} \{ y_i - H(x_{i|0}) \}' R_i^{-1} \{ y_i - H(x_{i|0}) \} \\
+ \frac{1}{2} \{ x_0 - x_{0|0} \}' B_0^{-1} \{ x_0 - x_{0|0} \}
\]

♦ Penalized Least Squares method

\[
I(x_0) = \frac{1}{2} \sum_{i=1}^{n} \{ y_i - H(x_{i|i}) \}' R_i^{-1} \{ y_i - H(x_{i|i}) \} \\
+ \frac{1}{2} \{ x_0 - x_{0|0} \}' V_0^{-1} \{ x_0 - x_{0|0} \} \\
+ \sum_{i=1}^{n} \{ x_i - x_{i|i-1} \}' Q_i^{-1} \{ x_i - x_{i|i-1} \}
\]

Rabier et al (1993)
Ide & Ghil (1997)

* Closed system
* Perfect model with Observation errors

* Open system
* Model errors as well as Observation errors
Similar principles

- **Variational method** (4D-Var)
  - Open system
  - Model errors as well as observation errors

- **Penalized Least Squares method**
  - Open system
  - Model errors as well as observation errors

- **Sequential method** (Extended Kalman Filter → MLE)
  - Open system
  - Model errors as well as observation errors
  (similar to P.L.S. but not the same!)
Hidden approximations behind perfect-model assumptions

- Infinite dimensional state → Finite dimension
  (→ Model error)
- Open universe → Closed system
  (→ Model error)
- Nonlinear dynamics → Numerical Approximation
  (→ Model error)
Experience in Chaos
(Smoothing the trajectory)

Penalized L.S. method didn’t work!

\[ I(x_0) = \frac{1}{2} \sum_{i=1}^{n} \{y_i - H(x_{il})\}' R_i^{-1} \{y_i - H(x_{il})\} \]
\[ + \frac{1}{2} \{x_0 - x_{0|0}\}' V_0^{-1} \{x_0 - x_{0|0}\} \]
\[ + \sum_{i=1}^{n} \{x_i - x_{il|1}\}' Q_i^{-1} \{x_i - x_{il|1}\} \]

Farmer & Sidorowich (1991)
Kostelich & York (1988)

Many local minimums!
Non-penalized L.S. method (4D-VAR) is even worse!

Prediction errors

\[ \nu_k = x_k - E[x_k \mid x_{k-1}, \ldots, x_1] \]

with \( \sigma_n^2 = 0 \)
Prediction errors with assumptions

\[ \sigma_n^2 = 0.1 \times 10^{-7} \]

\[ \sigma_n^2 = 0.1 \times 10^{-6} \]

\[ \sigma_n^2 = 0.1 \times 10^{-4} \]
M.L.E. with L.L. Filtering


Maximum Likelihood Method works for Lorenz chaos & Rikitake chaos

\[ (-2) \log p(y_1, y_2, \ldots, y_N) = \sum_{t=1}^{N} \left[ \log \sigma^2_{t|t-1} + \frac{(y_t - H(x_{t|t-1}))^2}{\sigma^2_{t|t-1}} \right] \]

Remained problem:
Computational burden from huge dimensional states

1. Optimization
2. Objective function
3. Filtering

Innovation Approach
Innovation Approach to Spatial Time Series:

Numerical Example

fMRI data

147,456- dimensional time series
fMRI Machine
fMRI data

Sampling frequency: 3 sec  (3T)
Resolution: 64×64×36 (slices)

147,456 dimensional time series

147,456-dim AR(1) is impossible?

(147,456 ×147,456 transition matrix ?)
Innovations in spatial dynamics

\[ \varepsilon_t^{(i,j)}(x) = x_t^{(i,j)} - E[x_t^{(i,j)} | x_{t-1}^{(*,*)}, x_{t-2}^{(*,*)}, ...] \]

\[ \varepsilon_t^{(i,j)}(x | s) = x_t^{(i,j)} - E[x_t^{(i,j)} | x_{t-1}^{(*,*)}, x_{t-2}^{(*,*)}, ..., s_{t-1}^{(*,*)}, s_{t-2}^{(*,*)}, ...] \]

Why not start from the simplest linear model.
Space-Temporal Model with stimulus

Simplest example:

\[ x_t^{(i,j)} = \mu^{(i,j)} + \alpha^{(i,j)} x_{t-1}^{(i,j)} + \beta^{(i,j)} \xi_{t-1}^{(i,j)} + \gamma^{(i,j)} s_{t-1} + \varepsilon_t^{(i,j)} \]

\[ \xi_{t-1}^{(i,j)} = (x_{t-1}^{(i,j+1)}, x_{t-1}^{(i,j-1)}, x_{t-1}^{(i-1,j)}, x_{t-1}^{(i+1,j)}) \]

\[ \beta^{(i,j)} = (b_N^{(i,j)}, b_S^{(i,j)}, b_W^{(i,j)}, b_E^{(i,j)}) \]

Solve a linear equation for each space point (i,j).

\[ s_{t-\tau} \quad \text{: Stimulus} \]

\[ \xi_{t-1}^{(i,j)} \quad \text{: Neighbour vector} \]

\[ \beta^{(i,j)} \quad \text{: Neighbour coefficient} \]

\[ \hat{\mu}^{(i,j)}, \hat{\alpha}^{(i,j)}, \hat{\beta}^{(i,j)}, \hat{\gamma}^{(i,j)}, \hat{\sigma}_\varepsilon^{(i,j)} \]
Estimated Model tells you something

This is a special type of $N_p \text{dim ARX model}$

$N_p = 64 \times 64 \times 36 = 147,456$)

\[
x_t^{(i,j)} = \hat{\mu}^{(i,j)} + \sum_{k=1}^{r_1} \hat{\alpha}_k^{(i,j)} x_{t-k}^{(i,j)} + \sum_{k=1}^{r_2} \hat{\beta}_k^{(i,j)} \xi_{t-k}^{(i,j)} + \sum_{k=1}^{r_3} \hat{\gamma}_k^{(i,j)} s_{t-k} + \epsilon_t^{(i,j)}
\]

\[
X_t = \hat{M} + \sum_{k=1}^{r_1} \hat{A}_k X_{t-k} + \sum_{k=1}^{r_3} \hat{\Gamma}_k S_{t-k} + E_t
\]
Looking through

1. Mean field map \( \hat{\mu}^{(i,j)} \)
2. Cross-spectrum field map \( \hat{p}(f) = \hat{A}(f)\hat{\Sigma} \hat{A}(f) \)
3. Causality field map \( \hat{p}^{(n_{ij},n_{ij})}(f) = \sum_{k=1}^{N_p} |\hat{a}^{(n_{ij},k)}(f)|^2 \hat{\sigma}_k^2 \)
4. Impulse response function \( h(i, j, \tau) \)
5. Innovation field map \( \hat{\epsilon}_t^{(i,j)} \)
6. Response field map \( \sum_{\tau_1}^{\tau_2} \hat{\gamma}^{(i,j)}(i,j) \)
7. Innovation + Response field map \( \hat{\epsilon}_t^{(i,j)} + \sum_{\tau_1}^{\tau_2} \hat{\gamma}^{(i,j)}(i,j) S_{t-k} \)

\[
X_t = \hat{\mu} + \sum_{k=1}^{r_1} \hat{\Lambda}_k X_{t-k} + \sum_{k=1}^{r_2} \hat{\Gamma}_k S_{t-k} + \hat{\epsilon}_t
\]
Example – A

fMRI data 3T

Task: Visual stimulus by black and white shuffled check board

Sampling frequency: 3s
Resolution: $64 \times 64 \times 36$ (slices)

($=147,456$)
Slice 12

Time Series plots of special line (j=25) of the slice k=12.
\[ x_t(i,j) = \sum_{k=\tau_1}^{\tau_2} \gamma_k S_{t-k} + \varepsilon_t(i,j) \]
\[ x_t(i, j) \]

\[ \sum_{k=\tau_1}^{\tau_2} \gamma_k^{(i, j)} S_{t-k} \]

\[ \epsilon_t^{(i, j)} + \sum_{k=\tau_1}^{\tau_2} \gamma_k^{(i, j)} S_{t-k} \]

\[ i=41 \sim 45 \]

\[ k=12 \]

\[ j=25 \]
\[
q^{(i,j)}_t \quad \sum_{k=\tau_1}^{\tau_2} \gamma_k S_{t-k} \\
\epsilon_t^{(i,j)} + \sum_{\tau_1}^{\tau_2} \gamma_k S_{t-k}
\]
Innovation Maps

\[ \varepsilon_{t(i,j)} + \sum_{k=\tau_1}^{\tau_2} \gamma_{k}^{(i,j)} S_{t-k} \]

\[ \sum_{k=\tau_1}^{\tau_2} \gamma_{k}^{(i,j)} S_{t-k} \]

\[ x_t^{(i,j)} \]
Spatial Impulse Response
Spatial ARX Simulation-1

\[ x_t^{(i,j)} = 0.7 x_{t-1}^{(i,j)} + 0.1 x_{t-1}^{(i+1,j)} + 0.1 x_{t-1}^{(i-1,j)} + 0.1 x_{t-1}^{(i,j+1)} + 0.1 x_{t-1}^{(i,j-1)} + \varepsilon_t^{(i,j)} \]
Innovation Approach could be useful in Space-Time

\[ x_1^{(i,j)}, x_2^{(i,j)}, x_3^{(i,j)}, \ldots, x_N^{(i,j)} \]

\[ \varepsilon_1^{(i,j)}, \varepsilon_2^{(i,j)}, \varepsilon_3^{(i,j)}, \ldots, \varepsilon_N^{(i,j)} \]

- Time Series Models
- Dynamical Systems
- Stochastic Dynamical Systems
Thank you
(ozaki@ism.ac.jp)