(14)「時空間データ解析」に関する研究報告

脳機能の計量における時空間データ解析

片野田耕太　杉下守弘
東京大学大学院　医学系研究科　認知言語神経科学

機能的磁気共鳴画像（機能的 MRI）は、陽電子放射断層撮影法（PET）と並んで、ヒトの脳機能の計量によく用いられる手法である。本稿では、機能的 MRI の実験手法とデータの構造について概説し、その解析法を主に統計学的視点から解説する。

（1）機能的 MRI とは？

機能的 MRI は、血中の酸素化ヘモグロビンと脱酸素化ヘモグロビンの磁化率の違いを利用して、脳内の血流動態を調べる手法である。近年、脳の神経活動を非侵襲的に、かつ視覚的に捉えられる手段として注目され、盛んに用いられるようになった。使用される MRI 装置は、通常の臨床で用いられる 1.5 テスラ（15,000 ガウス）の静磁場強度を持つものが一般的で、EPI 法と呼ばれる高速撮像法により頭部の断層画像を撮影する。時間分解能は 1 秒～数秒程度、空間分解能は数 mm 程度である。被験者は頭部を固定された状態で MRI 装置に入り、いくつかの課題を遂行する。その間、頭部の断層画像を連続的に撮像し、脳の各部位から信号を採取する。得られた信号の強度がそれらの課題間で異なるかどうかを統計学的に検定することにより、特定の課題遂行時に脳のどの部位が賦活していたかを評価する。

（2）実験パラダイム

機能的 MRI には大きく分けて block design と event-related design の２つの実験パラダイムがある。両者のデータ構造および解析方法は基本的には同じなので、本稿では block design についてのみ述べる。

block design では、複数の課題を数 10 秒毎に交互に繰り返し行う。図 1 に最も単純な例を示す。この例では、指運動と休みの２種類の課題を 40 秒毎に交互に計 3 回ずつ行っている。機能的 MRI の画像を 4 秒に 1 回撮像したとすると、それぞれの山（谷）につき 10 回の scan、計 60 回の scan を行うことになる。なお、3 種類以上の課題をランダムに繰り返す複雑な実験も可能である。

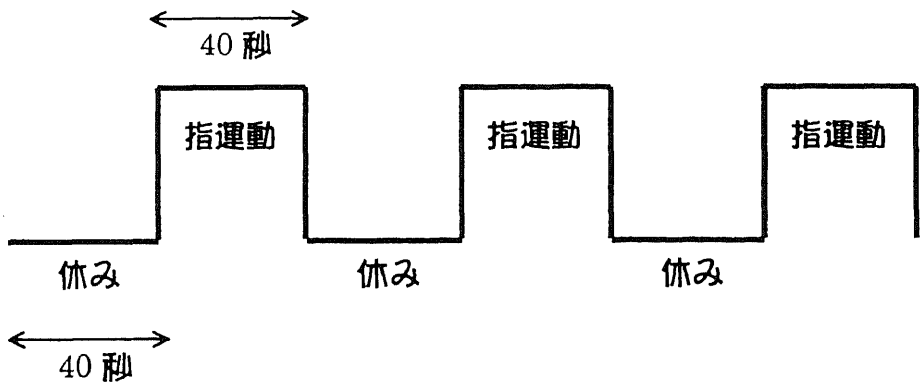


図 1

（3）データの構造

機能的 MRI のデータは、voxel と呼ばれる数 mm 角の直方体から成る。1 つの voxel は 1 つの信号強度（数値）を持ち、それが積み木のように積み重なって 1 つの脳（volume）を構成する。さらにこの volume が時系列方向に scan の数だけ並び（図 1 の例の場合 60 time point）、その結果 voxel 毎に信号変化のタイムコースができる。

（4）統計学的解析 -被験者 1 例の場合-

機能的 MRI の解析方法には、観測された信号変化とモデルカーブとの相関係数を用いるもの、信号の変化率（percent change）を用いるもの等様々な手法が考案されているが、本稿では、最も広く用いられている一般線形モデル(General Linear Model: GLM) による解析法を紹介する。この解析は、ロンドン大学 Wellcome Department の作成した SPM (Statistical Parametric Mapping) と呼ばれる解析ソフトウェアで主に行われる。

解析は原則として各 voxel 毎に独立で行われる。GLM を用いた解析では、voxel 毎に重回帰分析を行うのが主流である。従属変数は voxel 毎に各 time point で得られた信号強度であり、サンプルサイズは time point の数（図 1 の例では 60）に等しい。説明変数は、課題の有無を 1 と 0 のダミー変数化したものである。drift と呼ばれる低周波のノイズをキャンセルするために、いくつかの周波数の正弦波を共変量に加えることもある。また、局所的な信号変化のみに注目するために、頭部の全 voxel を平均した信号変化（global change）を共変量に加えることもある。このようにして作られた重回帰式はベクトルと行列を用いて $Y = X\beta + e$ と表現され、説明変数と共変量をコラムとして持つ X は design matrix と呼ばれる。回帰係数（ベクトル）β は最小二乗法を用いて推定する。

続いて、contrast と呼ばれる重み付け和を用いて、推定された回帰係数に対する検定を行う。例えば図 1 の例では、帰無仮説：「指運動に対応する $\beta = 0$」を設定して voxel 毎に t 値を算出し、この t 値が有意に 0 より大きい voxel が指運動時に休み時と比較して信号強度が増加した部位であると解釈される。こうしてそれぞれの t 値を持つ voxel が積み重なって頭部全体を構成したものを statistical map（t 値の場合 t-map）と呼ぶ。最後に、この t-map に対して予め定めた有意水準（通例 $p < 0.05$)で検定を行い、有意な voxel のみを色付けして解剖学的な画像に overlay して表示する。検定の際、複数の voxel で同時に検定を行うという意味での多重性を考慮するために random field theory と呼ばれる理論に従って多重比較を行う。

（5）統計学的解析 -複数の被験者の場合-

脳の大きさや形は被験者間で異なるが、normalization と呼ばれる、基準脳への標準化を行うことで、複数の被験者をまとめて解析することができる。複数被験者の処理には random effect model を適用し、（4）の処理で被験者毎に推定した各ボクセルの回帰係数 β の重み付け和を対象に、あらためて one-sample t-test を行う。この場合、回帰係数 β の被験者間の分散のみが誤差項となる。また、患者と健常者など2グループ間に差があるかを、同様に各群の回帰係数 β の重み付け和を用いて two-sample t-test で検定することもできる。

発表では、この種の解析法の利点や制限についても言及し、改善点や代替手法を模索する意味で様々な専門分野の方々との討論を行いたい。

空間時系列モデルの検定について

新潟大学　経済学部　松田 安昌
東京大学　経済学部　矢島 美寛

1　はじめに

本報告では、複数の地点で同時に観測された時系列を解析する空間時系列モデルについて論じる。空間時系列を解析する場合、時間をfixして空間データとみなしてモデル化するか、同時点の観測値の空間平均をとって通常の一変量時系列に帰着させて解析されることが多い。ここでは、時間と空間を同時に扱う時空間モデルを考え、特にデータの相関が時間と空間の積で与えられるseparableモデルを扱う。すなわち、$\{Z_{a,t}\}$, $(a=1,\ldots,r,\ t=1,\ldots,n)$ を地点 a、時刻 t において観測された空間定常時系列とし、

$$\mathrm{Cov}(Z_{a,s}, Z_{b,t}) = \sigma_a\sigma_b\rho_1(a,b)\rho_2(s-t) \tag{1}$$

を満たすモデルの検定について以下議論する。

2　separabilityの検定

仮説 (1) を帰無仮説とするseparabilityの検定統計量を提案する。今、r 地点で観測された空間時系列 $\mathrm{Z}_t = (\mathrm{Z}_{1,t},\ldots,\mathrm{Z}_{r,t})'$ の r 次元スペクトル密度行列の存在を仮定して $f(\lambda) = (f_{ab}(\lambda))$ とおく。仮説 (1) の下では、$f(\lambda)$ は、$g(\lambda)$ を $\rho_2(h) = \int \exp(ih\lambda)g(\lambda)d\lambda$ を満たすものとして、

$$f_{ab}(\lambda) = \sigma_a\sigma_b\rho_1(a,b)g(\lambda) \tag{2}$$

となることに注意する。

$Z_1,\ldots,Z_n$ を観測したとして、Z_t の有限フーリエ変換を $W(\lambda_t) = \frac{1}{\sqrt{2\pi n}}\sum_{s=1}^{n} Z_s\exp(-is\lambda_t)$ として、λ_t におけるピリオドグラムを $I_{Z,t} = W(\lambda_t)W(\lambda_t)^*$ と定義しておく。$\hat{f}_{U,t}$ はピリオドグラム $I_{Z,t}$ をスムージングして得られる通常の推定量でその (ab) 成分の値 $\hat{f}_{ab,t}$ は、

$$\hat{f}_{U,ab,t} = \begin{cases} \frac{1}{m/2+t}\sum_{j=-t+1}^{m/2} I_{Z,ab,t+j}, & 0 < t \le m/2, \\ \frac{1}{m+1}\sum_{j=-m/2}^{m/2} I_{Z,ab,t+j}, & m/2 < t \le [n/2]-m/2, \\ \frac{1}{[n/2]+m/2-t+1}\sum_{j=-m/2}^{[n/2]-t} I_{Z,ab,t+j}, & [n/2]-m/2 < t \le [n/2], \\ \overline{\hat{f}_{ab,-t}}, & t < 0. \end{cases}$$

で与えられる。

さて、帰無仮説 (1) の下で有効な推定量 $\hat{f}_{R,t}$ を構成する。

$$\hat{f}_{R,t} = \hat{V}\hat{g}_t,$$

但し、

$$\begin{aligned}\hat{V}_{ab} &= \frac{1}{n}\sum_{t=1}^{n} Z_{at}Z_{bt},\\ \hat{g}_t &= \frac{1}{r}\sum_{a=1}^{r}\frac{\hat{f}_{U,aa,t}}{\hat{V}_{aa}}\\ &= \frac{1}{r}\mathrm{tr}\left[\hat{f}_{U,t}\hat{\Sigma}^{-1}\right],\\ \hat{\Sigma} &= \mathrm{diag}(\hat{V}_{11},\ldots,\hat{V}_{rr})\end{aligned}$$

とする。$I_{Z,t}$ と $\hat{f}_{R,t}$ の2スペクトル間の近さを計る指標を以下の $C_0(I_Z,\hat{f}_R)$ で定義し、検定統計量として利用する。

$$C_0(I_Z,\hat{f}_R) = \frac{1}{n}\sum_{t=-[n/2],t\neq 0}^{[n/2]}\mathrm{tr}(I_{Z,t}\hat{f}_{R,t}^{-1}) - r. \tag{3}$$

以下の仮定のもとで $\hat{T}_0$ の帰無仮説の下での漸近分布が求まる。

(A1) Z_t はガウス過程で $\int \log\det f(\lambda)d\lambda > -\infty$.

(A2) $f(\lambda)$ は $\lambda\in[-\pi,\pi]$ において、2階連続微分可能、

(A3) $f_{aa}(\lambda)>0$、$a=1,\ldots,r,\ -\pi\le\lambda\le\pi$.

(A4) $m=O(n^{\beta})$, $1/2<\beta<3/4$.

Theorem 1. 仮定 *(A1)-(A4)*、帰無仮説 *(1)* のもとで、$n^{1/2}C_0(I_Z,\hat{f}_R)$ は平均 *0*、分散 $S_0=4\pi\int \mathrm{tr}[\alpha_0(\lambda)\mathrm{f}(\lambda)\alpha_0(\lambda)\mathrm{f}(\lambda)]\mathrm{d}\lambda$ の漸近正規性を持つ。ここで、

$$\alpha_0(\lambda) = \left\{\frac{1}{2\pi}-g(\lambda)\right\}\left[f^{-1}(\lambda)-(\Sigma g(\lambda))^{-1}\right].$$

次に、C_0 とは別の尺度を用いることにより、別の検定統計量を構成できる。ここでは、以下の統計量を考える。

$$C_K(\hat{f}_U,\hat{f}_R)=\sum_{t=-[n/2],t\neq 0}^{[n/2]}K(\hat{f}_{U,t}\hat{f}_{R,t}^{-1})-K(I_r),$$

ここで $K(\cdot)$ は、C^{r^2} 上の正則関数である。

Theorem 2. 仮定 *(A1)-(A4)*、帰無仮説 *(1)* のもとで、$n^{1/2}C_K(\hat{f}_U,\hat{f}_R)$ は平均 *0*、分散 $S_K=4\pi\int \mathrm{tr}[\alpha_{\mathrm{K}}(\lambda)\mathrm{f}(\lambda)\alpha_{\mathrm{K}}(\lambda)\mathrm{f}(\lambda)]\mathrm{d}\lambda$ の漸近正規性を持つ。ここで、

$$\begin{aligned}\alpha_K(\lambda) &= \left\{\frac{1}{2\pi}-g(\lambda)\right\}\left[f^{-1}(\lambda)\tilde{K}'-\frac{\mathrm{tr}(\tilde{K})}{r}(\Sigma g(\lambda))^{-1}\right],\\ \tilde{K} &= \left.\frac{\partial K(M)}{\partial M}\right|_{M=I_r}\end{aligned}$$

Theorem 1, 2により、検定統計量 $\hat{T}_0$, $\hat{T}_K$ を提案する。

$$\hat{T}_0 = \sqrt{n}\frac{C_0(I_Z, \hat{f}_R)}{\sqrt{\hat{S}_0}}, \tag{4}$$

$$\hat{T}_K = \sqrt{n}\frac{C_K(\hat{f}_U, \hat{f}_R)}{\sqrt{\hat{S}_K}}, \tag{5}$$

但し、

$$\hat{S}_0 = 4\pi \ \mathrm{tr}\left[\left\{I_r - \hat{V}\hat{\Sigma}^{-1}\right\}^2\right]\left[\frac{\frac{2\pi}{n}\sum_{t=-[n/2],t\neq 0}^{[n/2]}\hat{g}_t^2}{\left\{\frac{2\pi}{n}\sum_{t=-[n/2],t\neq 0}^{[n/2]}\hat{g}_t\right\}^2} - \frac{1}{2\pi}\right]$$

$$= 8\pi\left(\sum_{a<b}\frac{\hat{V}_{ab}^2}{\hat{V}_{aa}\hat{V}_{bb}}\right)\left[\frac{\frac{2\pi}{n}\sum_{t=-[n/2],t\neq 0}^{[n/2]}\hat{g}_t^2}{\left\{\frac{2\pi}{n}\sum_{t=-[n/2],t\neq 0}^{[n/2]}\hat{g}_t\right\}^2} - \frac{1}{2\pi}\right].$$

$$\hat{S}_K = 4\pi \ \mathrm{tr}\left[\left\{\tilde{K}' - \frac{\mathrm{tr}(\tilde{K})}{r}\hat{V}\hat{\Sigma}^{-1}\right\}^2\right]\left[\frac{\frac{2\pi}{n}\sum_{t=-[n/2],t\neq 0}^{[n/2]}\hat{g}_t^2}{\left\{\frac{2\pi}{n}\sum_{t=-[n/2],t\neq 0}^{[n/2]}\hat{g}_t\right\}^2} - \frac{1}{2\pi}\right].$$

例えば、$\mathrm{tr}(\mathrm{M})$、$\mathrm{tr}(\mathrm{M}^{-1})$, $\log\det\{(1-\alpha)\mathrm{I}_r + \alpha\mathrm{M}\}$ for $0 < \alpha < 1$ などが考えられる。最後に、対立仮説の下での漸近的な性質を示す。

Theorem 3. 条件 *(A1)-(A4)* を仮定する。対立仮説の下では、

$$n^{-1/2}\hat{T}_0 \to \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left(\mathrm{f}(\lambda)(\mathrm{Vh}(\lambda))^{-1}\right)\mathrm{d}\lambda - \mathrm{r}}{\sqrt{Q_0}}, \tag{6}$$

$$n^{-1/2}\hat{T}_K \to \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}K\left(f(\lambda)(Vh(\lambda))^{-1}\right)d\lambda - K(I_r)}{\sqrt{Q_K}} \tag{7}$$

in probability as n tends to infinity. 但し、$h(\lambda) = \frac{1}{r}\sum_{a=1}^{r}\frac{f_{aa}(\lambda)}{\sigma_a^2}$ とし、Q_0 と Q_K は S_0 *and* S_K の定義において $g(\lambda)$ を $h(\lambda)$ に置き換えたものである。

よって、(6),(7) の右辺が正になるようなモデルに対しては、それぞれ $\hat{T}_0, \hat{T}_K$ によって漸近的に separability は必ず棄却される。

3 シミュレーション

ここでは、$\hat{T}_0$ と、$K(M) = \mathrm{tr}(\mathrm{M})$, $\mathrm{tr}(\mathrm{M}^{-1})$, $\log\det\{\tau\mathrm{M} + (1-\tau)\mathrm{I}_r\}$, $(0 < \tau \leq 1)$ として構成した $\hat{T}_1, \hat{T}_2, \hat{T}_3(\tau)$ の有限標本での性質を調べる。

$\mathbf{e}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{r,t})' \sim N(0, I_r)$ として、多変量時系列 $\mathbf{Z}_t = (Z_{1,t}, \ldots, Z_{r,t})'$ を下記のモデルで作成する。

$$Z_{t,a} = \phi_{10} Z_{t-1,a} + \phi_{01} Z_{t,a-1} + \phi_{11} Z_{t-1,a-1} + \varepsilon_{t,a}, \quad a = 1, \ldots, r,$$

このモデルは、$\phi 10 \phi_{01} = -\phi_{11}$ のとき、separability をもつモデルとなる。

まず、$\phi_{10} = -0.6$, $\phi_{01} = 0.3, \phi_{11} = 0.18$ で生成したデータに対して、統計量を計算しその分散を評価した結果が以下である。但し、平均、分散は 300 回繰り返し実験を行うことで評価した。

n	$\hat{T}_0$	$\hat{T}_1$	$\hat{T}_2$	$\hat{T}_3(1)$	$\hat{T}_3(\frac{1}{2})$	$\hat{T}_3(\frac{1}{4})$
101	1.53	1.21	223.70	7.93	1.89	1.25
201	1.54	1.17	30.01	4.59	1.66	1.23
301	1.47	1.15	19.03	3.97	1.70	1.28
501	1.24	1.02	10.03	2.70	1.38	1.11
1001	1.18	0.99	5.45	2.04	1.27	1.08

次に、$\phi_{10} = -0.6, \phi_{01} = 0.3, \phi_{11} = 0.08, \ldots, 0.28$ で生成したデータに対して、漸近有意水準 5%で検定を行い、その検出結果を示す。但し、検出力は 300 回繰り返し実験によって評価した。比較のため、Shitan and Brockwell(Commun. Statist.-Theory Meth., 24(8),2027-2040, 1995) の統計量による検定結果も同時に示す。

ϕ_{11}	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28
$\hat{T}_0$	1.00	0.98	0.82	0.55	0.22	0.12	0.19	0.43	0.70	0.84	0.93
$\hat{T}_3(1/4)$	1.00	0.97	0.87	0.55	0.21	0.09	0.17	0.45	0.77	0.92	0.99
S & B	1.00	0.96	0.81	0.51	0.39	0.34	0.39	0.52	0.75	0.90	0.98

4 適用例

次に、日本の 13 都市 (札幌、仙台、東京、新潟、金沢、名古屋、大阪、高松、松江、広島、福岡、鹿児島、那覇) の各気象台において観測された 1950 年から 1998 年までの 49 年分の冬期平均気温 (1、2 各月の平均気温の平均をとったもの) に対して、regression model with separable spatial time series error を用いて温暖化の検証を行った。すなわち、地点 a に於ける t 年の気温を $X_{a,t}$ として、回帰モデル

$$\begin{aligned} X_{a,t} = {} & \phi_0 + \phi_1 t + \phi_2 \times (\text{緯度})_a + \phi_3 \times (\text{経度})_a + \phi_4 \times (\text{高度})_a \\ & + \phi_5 \times (\text{人口密度})_{a,t} + \phi_6 \times (\text{冬期平均降雨量})_a + Z_{a,t} \end{aligned}$$

をあてはめ、その誤差系列 $\{Z_{a,t}\}$ に対し、separable の検定、separable model(1) のあてはめと推定した相関構造の下での回帰係数とその tratio の計算、ガウス性の検定を実行した。

OLS 残差に対する、separability の検定を行った。

m	30	32	34	36	38	40
$\hat{T}_0$	0.201	0.282	0.312	0.206	0.212	0.167
p-value	0.87	0.79	0.70	0.84	0.79	0.81

separability の仮定は妥当であるという結果である。

以下，separable model

$$\mathrm{Cov}(Z_{a,s}, Z_{b,t}) = \sigma_a \sigma_b \exp\{-\psi d(a,b)\}\, \rho_2(s-t, \tau)$$

によるモデル化を以下の手順で行った。

1. OLS を計算.
2. OLS 残差に対し，separable model の MLE を計算.
3. 2. で推定した残差の相関構造の下で，GLS を計算.

GLS 推定結果とその尤度を紹介する。

Noise model		Log	Estimated value with t-ratio						
time	spatial	likelihood	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
indep.	indep.	−978.6	27.7	0.023	−1.31	0.16	−0.008	0.073	0.005
			8.96	7.20	−36.8	5.10	−2.18	4.89	5.69
AR(1)	isot.	−393.9	43.5	0.030	−1.22	0.03	−0.006	0.083	0.005
			18.11	3.01	−56.5	1.42	−4.50	18.32	11.35
AR(2)	isot.	−362.9	43.6	0.031	−1.22	0.03	−0.006	0.083	0.005
			16.29	2.78	−50.8	1.26	−4.07	16.39	10.25
AR(3)	isot.	−350.0	43.6	0.031	−1.22	0.03	−0.006	0.083	0.005
			15.35	2.67	−47.8	1.19	−3.86	15.37	9.68
ARIMA(0,d,0)	isot.	−318.8	43.4	0.031	−1.22	0.03	−0.007	0.081	0.005
			11.33	2.89	−35.6	0.96	−3.01	11.14	7.10
ARIMA(0,d,0)	anisot.	−318.6	42.9	0.032	−1.23	0.04	−0.007	0.080	0.005
			11.37	2.94	−35.4	1.11	−3.23	11.24	6.82

FARIMA(0,d,0) モデルによる時間相関と isotropic な空間相関の積で与えられる相関構造が尤度最大のモデルになっている。

Spatial Smoothing via a Resampling Method: Estimation with Area-based Panel Data

Atsushi Yoshida *
Osaka Prefecture University
1-1, Gakuen-cho, Sakai, Osaka 599-8531

Tatsuhiro Shichijo
Osaka Prefecture University
1-1, Gakuen-cho, Sakai, Osaka 599-8531

(Preliminary)

Abstract

When we use area-based panel data to analyze economic activities in a metropolitan region, we have to treat cluster effects because some economic activities agglomerate in a group of areas adjacent each other and thus form a cluster. We propose a resampling method, namely *leave-one-out cross-validation*, to find how many clusters are there in the region and which area belongs to which cluster. We examine the effectiveness of the method with simulation studies and compare the estimates with the within-class estimates. We also apply our method to find potential demand for houses in Tokyo Metropolitan Area.

Key words: cluster-effects model, housing start, leave-one-out cross-validation, panel data, resampling method

JEL classification:

*This research was supported by a grant-in-aid from the Zengin Foundation for Studies on Economics and Finance.

1 INTRODUCTION

In this paper, we will consider statistical issues when we use area-based panel data models in order to analyze economic activities in a metropolitan region. The metropolitan region consists of officially pre-determined areas like counties or municipalities. Most of the data available for us are based on these areas.

These official borders do not necessarily constrain economic activities of a private sector in a metropolitan region. Economic infrastructure such as railways, subways, highways, roads, canals, ports and so on, which are called the *second nature* by Krugman (1993, 1996), lays across the borders and possibly affects the activities. We often find manufacturing factories agglomerate along a canal, while software firms agglomerate around a university in the metropolitan region. Their activities are across the pre-determined official areas but are concentrated in areas adjacent each other. We will call the group of areas where some economic activities agglomerate a cluster.

Although detecting the clusters may not be difficult when the economic activities of concern are observable, it is a statistical issue to do so when they are unobservable. We will consider the case where area-based panel data are available and the clusters are represented by cluster parameters in a linear regression model, which are the same within a cluster but different between clusters. Thus the model is regarded as a panel data model that has cluster-effects as fixed effects.

If we are not concerned with the cluster-effects but concerned only with the parameters of observable explanatory variables, we can adopt an area-effects model that has area-specific-effect parameters for each area and obtain the within-class estimates of parameters of concern by applying analysis of variance (see Hsiao (1986) for details). Even in this case, the within-class estimates are possibly less efficient than the estimates of the cluster-effects model, as long as we can detect the structure of clusters among the metropolitan areas.

In section 2, we will consider a statistical method how to detect which area belongs to which cluster. From statistical viewpoint, this issue is regarded as a sort of model selection problems. We adopt a resampling method, namely *leave-one-out cross-validation*, since the

method is robust to distributional assumptions and the calculation is easily implemented in a linear model. The method is introduced by Stone (1974) and Geisser (1975) and its application for broad model selection problems is discussed in Davison and Hinkley (1997), though it has not been applied to detect clusters with area-based data. The selection procedure with the method is also explained. Section 3 shows results of simulation studies how well the *leave-one-out cross-validation* works to detect the clusters. We also compare the within-class estimate with the cluster-effects estimates. We find the cluster-effects estimates are more efficient than the within-class estimates. The method is also applied to estimate a housing demand function in Tokyo Metropolitan Area. Housing demand in an area depends upon basically income per household, amenity of an area and the disutility caused by congestion. It also depends on unobservable utility-improving environmental factors that are not capitalized in the land price of an area. Potential housing demand is affected by these factors in an area, which are represented by the cluster-effects in a statistical model and are to be estimated. Section 4 concludes and discusses remaining issues.

2 STATISTICAL MODEL WITH CLUSTERS

Let us consider a model with area-based panel data. Assume that we have observations of m areas for T periods. The area-effects model is expressed as follows:

$$y_{it} = \mu_i + x_{it}\beta + v_{it}, \quad i = 1, \ldots, m; \; t = 1, \ldots, T \tag{1}$$

The $\{y_{it}\}$ and $\{x_{it}\}$, which does not include a constant term, are dependent and independent variables that represent socio-economic properties of the ith area, respectively. The β $(K \times 1)$ represents the relationship between them which is of concern for researchers. The μ_i, one of area-effects, represents unobservable socio-economic characteristics in the ith area. The v_{it} is a error term that is independent and identically distributed for all i and t.

Let us assume there are q $(q \ll m)$ clusters, which are unobservable and thus we have to decide q statistically. The area-effects, $\mu_i, i = 1, \ldots, m$, should be classified into q classes, say

$\mu_1, \ldots, \mu_q$. Then the cluster-effects model is as follows:

$$y_{it} = \mu_q + x_{it}\beta + v_{it}, \quad i = 1, \ldots, m; \ t = 1, \ldots, T; \quad \text{if } i \in q\text{th cluster} \tag{2}$$

The vector form of eq.(2) is

$$y_t = D_0\mu_0 + X_t\beta + v_t, \quad t = 1, \ldots, T. \tag{3}$$

The $\mu_0 = (\mu_1, \ldots, \mu_q)$ is a parameter vector to be estimated. D_0 is an $m \times q$ matrix of dummies that indicates which area belong to which cluster. For example, if sth area and lth area belong to the same cth cluster, then the cth element of the sth and lth rows of D_0 are the same, namely 1, and the other elements of the rows are 0s.

We will consider how we can estimate the rank of D_0 (namely q), μ_0 and β, and identify the structure of D_0 based on the model eq.(3). There are two points to be considered for the estimation. Firstly, we have to find how many clusters are there, which area belongs to which cluster and to estimate the parameters of concern at the same time. Without the classification of area-effects into cluster-effects, we cannot obtain the consistent estimates of μ_0. Secondly, let A be an adjacent matrix of areas in the region, which is an $m \times m$ symmetric dummy matrix indicating which areas are neighbors of an area. For example, if the i, jth element is 1, then i, jth areas are adjacent. The diagonal elements are 1 by the definition. Since D_0 indicates the structure of clusters, it should be a transformed matrix of the adjacent matrix A by using information of which areas are combined together to a cluster.

From the statistical point of view, detecting the rank and structure of D_0 is regarded as a model selection problem. In this case, the largest model is the case where $\mu_1, \ldots, \mu_m$ have different values, that is, they are not classified into fewer classes at all, namely the area-effects model. On the other hand, the smallest model is the case where $\mu_1, \ldots, \mu_m$ have the same value, that is, they all are classified into one class. There are a lot of possibilities of classification between the largest and smallest models.

In a statistical model selection context, there are two major methods, one is using Kullback-Leibler-information-based selection criteria, namely AIC, BIC and SBIC (see Lütkepohl (1991), for example), and the other using a resampling-method-based selection criterion. The former

criteria are easily calculated but they heavily depend upon the assumptions of distributions. On the other hand, the latter criterion needs huge computation time, though they are robust to them. The asymptotic equivalence of *cross-validation* and AIC is proved by Stone (1979).

The model-selection criterion with the resampling methods is *aggregate prediction error*. In a linear regression model, it is defined as

$$\Delta = \frac{1}{n}\sum_{i=1}^{m} E((Y_{+j} - \eta(X_j, \hat{F}))^2 | \hat{F})$$

where Y_{+j} is one of possible realizations at X_j, $\eta(X_j, \hat{F})$ being an estimate of mean response function and $\hat{F}$ is an empirical distribution of Y and X that represents data. One of the estimate of the *aggregate prediction error* is obtained by *leave-one-out cross-validation*, which is defined as

$$\hat{\Delta}_{CV} = \frac{1}{n}\sum_{i=1}^{m}(y_j - \eta(x_j, \hat{F}_{-j}))^2$$

where $\hat{F}_{-j}$ represents the $n-1$ observations $\{(x_k, y_k), k \neq j\}$. In a linear regression model, we have $\eta(x_j, \hat{F}_{-j}) = x_j\hat{\beta}_{-j}$ where $\hat{\beta}_{-j}$ is the estimate using just the data of Y and X excluding the jth sample. To select a model among possible combination of explanatory variables, we calculate $\hat{\Delta}_{CV}$ for all combinations and select the combination which attains the minimum value in principle. It is, however, almost impossible because there are too many possibilities to try.

In general, forward, backward, or stepwise methods are often used for selecting combinations of variables when trying all combinations is impossible. In our model-selection problem, we select the backward method, that is, starting with the largest model, we combine two adjacent areas for all possible cases, selecting the combination that attains the smallest APE and regard a newly integrated area as a cluster. Then the number of the areas decreases by one in every step of the procedure. By repeating this process, we can find the clusters where the APE is the smallest.

Let us explain the procedure stated above more precisely. The matrix form of eq.(3) is

$$y = (\mathbf{1}_T \otimes D_0)\mu_0 + X\beta + v, \tag{4}$$

where $y = (y_1', \ldots, y_T')'$, $X = (X_1', \ldots, X_T')'$ and $v = (v_1', \ldots, v_T')'$. The purpose is to find the structure of clusters which is expressed in D_0 and estimate μ_0 and β. Since we use the

backward method to combine the adjacent areas, we need an $m \times m$ adjacent matrix, $A(m)$, as an initial areal condition, while an initial matrix for D is an $m \times m$ identity matrix. First, we calculate $APEs$ for all possible combinations of two areas adjacent each other. For example, assuming the kth area and lth area are adjacent, we calculate APE with an $m \times (m-1)$ matrix $D(m-1; k=l)$, which created by integrating the kth and lth column vectors. Among $APEs$ for all possible combinations, we can select the minimum-attained combination. We define the value of *APE(m-1)* as $APE^*(m-1)$ and new $(m-1) \times (m-1)$ adjacent matrix as $A(m-1)$. In the next step, we use the adjacent matrix $A(m-1)$ for searching possible combinations of areas. Let us redefine y_j and x_j as the jth element and row vector of y and X for all $j = 1, \ldots, mT$, respectively. Then *APE(k)* is calculated as

$$APE(k) = \frac{1}{mT} \sum_{j=1}^{mT} (y_j - d_j(k)\hat{\mu}_{-j} - x_j \hat{\beta}_{-j})^2$$

where $d_j(k)$ is the jth column of $\mathbf{1} \otimes D(k)$, $\hat{\mu}_{-j}$ and $\hat{\beta}_{-j}$ is the estimates using the data excluding $y_j, d_j(k)$ and x_j. The optimal *APE(k)* is

$$APE^*(k) = \min APE(k)$$

Repeating this procedure from $k = m$ to $k = 2$ and selecting k^* as $k^* = \min_k \mathrm{APE}^*(k)$. Then k^* is the optimal rank of D_0 and its corresponding $D(k^*)$ is the structure of clusters.

3 SIMULATIONS AND EMPIRICAL EXAPMPLE

In this section, we will examine if the method proposed in the previous section would work well and apply it to analyze municipality-based data of housing start in Tokyo Metropolitan Area. In the simulations, we will compare the estimates of cluster-effects model with that of the area-effects model, namely within-class estimates. Note that we are not able to find the structure of clusters or obtain consistent estimates of μ with the area-effects model. The within-class estimate of β is defined as follows:

$$\hat{\beta}_W = (X'QX)^{-1}X'Qy$$

where $Q = I_{mT} - Z(Z'Z)^{-1}Z'$, $Z = I_m \otimes \mathbf{1}_T$.

3.1 SIMULATIONS

In the simulations, we generate the necessary data based on eq.(2) for 3 years (T=3). We use a 6×6 lattice for a total region to be examined, where there are 36 areas ($m = 36$ in eq.(2)). We have to define which are neighbors of an area at first. We assume that left, right, upper and lower adjacent areas of an area are its neighbors. Note that there are just 2 neighbors for 4 corner-areas of this region and 3 neighbors for edge areas. We make sequential numbers for the areas in an order so that the i,jth cell of the lattice should be $6 \times (i-1) + j$ (see Fig. 1). We set three clusters, the upper-left cluster consisting of 9 areas (1,2,3,7,8,9,13,14,15), the upper-right cluster consisting 9 areas (4,5,6,10,11,12,16,17,18), and the rest consisting of 18 areas. The cluster effects are set as $\mu = (2, 5, 10)$ for upper-left, upper-right and the rest clusters, respectively.

The explanatory variable and the errors , x_{it} and v_{it}, are independently drawn from $N(3, 9)$ and $N(0, 4)$, respectively. The parameter β is set to be 2. We conducted 1000-times simulations.

In these simulations, we also consider the other case, where the cluster-effects are random and spatially correlated. The model is specified as

$$y_{it} = x_{it}\beta + u_i + v_{it}, \quad i = 1, \ldots, m; \; t = 1, \ldots, T,$$

where u_i represents random cluster-effects. We specify the conditional density function of the ith variable, u_i, as follows:

$$f(u_i|\{u_j, j \in \mathcal{N}_i\}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(u_i - m(u_j; j \in \mathcal{N}_i))^2/\sigma^2\right], \tag{5}$$

where $\mathcal{N}_i$ is a set of neighbors of the ith area and the conditional mean is defined as

$$m(u_j; j \in \mathcal{N}_i) = \mu_i + \sum_{j \in \mathcal{N}_i} \lambda c_{ij}(u_j - \mu_j)$$

where $c_{ij} = c_{ji}$, $c_{ii} = 0$ and $c_{ik} = 1$ if there is pairwise dependence between area i and area k, otherwise it is 0. The joint distribution of $u = \{u_1, \ldots, u_m\}$ is obtained as follows:

$$u \sim N(\mu, (I - \lambda C)^{-1}M), \tag{6}$$

where $\mu = (\mu_1, \ldots, \mu_m)$, C a $m \times m$ matrix with its i,jth element being c_{ij} and M is a $m \times m$ diagonal matrix with its ith diagonal element being σ^2. The detailed explanation of its properties is discussed in Cressie (1993). The random cluster-effects, u_i, $i = 1, \ldots, m$, are generated from normal distribution of eq.(6), where $\mu = 0$, $\lambda = 1/4$, $M = 3 \times I_m$ and the adjacent matrix is defined above.

Firstly, we evaluate how correctly we can select the number of clusters, the structure of clusters and how efficiently we can estimate the values of μ with the method. For the first point, we examine the distribution of the selected number of clusters in the simulations. We also examine the expectation how many estimated clusters lay across the true clusters. For the second point, we evaluate the efficiency by the mean squared error for each 36 areas, comparing three mean squared errors of the estimates, namely estimates with true clusters, those with selected clusters and the within-class estimates.

In table 1-1 and 1-2, we can see descriptive statistics of the distribution of the estimated number of clusters and the number of estimated clusters laying across the true clusters for the cases in fixed- and random-effects models. The mean and median of estimated number of clusters obtained with 1000 times simulations are 9.58 and 9, respectively, in the fixed-effects model and 14.16 and 14, respectively, in the random-effects model. Its standard deviation is 1.80 and about 80 % of the estimates is in the region from 7 to 12 in the former case. With this simulation, we can see the method tend to select larger number of clusters. Even if the estimated number of the clusters is larger than the true clusters, it does not cause bias of the estimates of the parameters as long as the column vectors of true cluster matrix, D_0 of eq.(3) are expressed as linear combinations of the estimated cluster matrix $\hat{D}$, though it affects the efficiency of them. The expectation of the number of estimated clusters that lay across the true clusters is less than 0.5, its median being 0 and the 90 percentile is 1 in the fixed-effects model. Thus the probability of the estimated clusters lying across the true clusters is extremely small. On the other hand, in the random-effects model, the expectation is more than 1 and the median is 1 so that there is a little possibility that the estimates of parameters are biased.

In table 2-1 and 2-2, we can see means of $\hat{\mu}$ and mean squared errors (MSEs) of the OLS estimates in the estimated cluster model, true cluster model and non-clustered model, namely

the within-class estimates. In both cases of fixed- and random-effects models, the estimates of the cluster-effects in each area are almost unbiased. In the fixed-effects model, the mean squared error of the clustered model is uniformly smaller than the non-clustered model, though they are larger than the true model. Note that the mean squared error of the clustered model consists of three parts, that is, the squared bias, variance of the estimate and the bias caused by misclustering. The third factor of the MSE is negligibly small from table 1-1. Even in the random-effects model, the MSEs of the estimates in clustered model are uniformly smaller than the non-clustered model, though the values are not so different from the non-clustered model.

Secondly, we compare the estimates of β in the estimated clustering structure with the within-class estimate. In table 3, we can see means, standard deviations and MSEs in OLS estimates with estimated clustering structure, in the within-class estimate and in OLS estimates with true clustering structure. In the case of fixed-effects model, both the standard deviation and MSE of the estimates of the clustered model are superior to the within-class estimate, though OLS estimates of the true cluster model is the most efficient among these estimates. In the random-effects model, since the within-class estimate is obtained by eliminating the effects, it is the most efficient estimate among them. Even in this case, the estimate of the clustered model is nearly as efficient as the within-class estimate. The OLS estimate with true clustering structure is the worst.

From the results of the simulations, we are able to conclude as follows: Firstly, the *leave-one-out cross-validation* tends to select larger model than the true model but the estimated clusters seldom lay across the true clusters so that the estimates of the cluster-effects are almost unbiased and also efficient than the within-class estimates. Secondly, in the case where the cluster-effects are fixed, the parameter-estimates of the explanatory variables are more efficient than the within-class estimates. Even in the cluster-effects being random, they are nearly as efficient as the within-class estimates. Thus, the estimates proposed in this paper are more preferable than the within-class estimates when the cluster-effects exist.

3.2 EMPIRICAL EXAMPLE

In this subsection, we apply our method to examine the determinants of the number of housing start per household of the municipalities in Tokyo Metropolitan Area (TMA) from 1996 to 1998, which is defined as a collection of areas that locate within 60-minute-distant from the Tokyo station. There are 87 municipalities in the region.

The explanatory variables are logarithm of income per household, logarithm of average price of residential land and logarithm of population density in an area. The explained variable is also taken logarithm. We expect housing start in an area with higher income per household will be larger than other areas. The average price of residential land in an area represents amenity of the area since amenity is capitalized in the land price. Thus in the area with higher land price demand for houses is larger than the other areas. The population density represents disutility caused by congestion. These three variables are the kernel of the determinants of housing start. At the same time, we are concerned with unobservable factor that affects housing start except for the kernel. The unobservable factor represents potential demand for houses in an area where, for example, housing stock per household is below the standard. Or it may be resistant or improving factor to build new houses by some official restrictions or area-development policies.

In table 4, we can see the estimates with clustered and non-clustered models. The within-class estimates are somewhat different from the clustered model. In both models, the income factor is insignificant but land price and population density are significant, though the values of the estimates are different. The coefficient of the land price is positive because it represents amenity of an area. Population density affects the housing start negatively because of the disutility of congestion. In figure 2, we can see which areas belong to the same cluster. Clusters are found along railways and river. East areas of TMA along Sumida River; Southern areas of Tokyo and western areas of Yokohama along Odakyu line; Western areas of Tokyo along Seibu-Shinjuku line, areas in Chiba along JR Sobu-line and so on. In figure 3, we can see the potential demand for houses represented by the cluster-effects. Potential strong demand for houses is found in northern part of Tokyo-23-districts and a southern area of Yokohama. On the other hand, the potential demands in the center and the border areas of TMA are weak.

Let us compare figure 3 and 4 that is a crude map of logarithm of housing start per household. These two maps give us different impression. In figure 4, the border areas and center of TMA have strong demand for houses per household. After adjusting the data with income, amenity and disutility by congestion, the potential demand is found in the areas where the demand seems to be weak in figure 3.

4 DISCUSSION

We propose a method of deciding how many clusters are there and which area belongs to which cluster and show by simulations that it works well and the estimated parameters of concern are more efficient than the within-class estimates. We also apply our method to examine what are the determinants of the number of housing start in Tokyo metropolitan area and spatial distribution of the unobservable potential demand for houses.

We adopt an *aggregate prediction error* as a model-selection criterion, which is estimated by a resampling method, namely *leave-one-out cross-validation.* It is possible to estimate the criterion with other resampling methods, say bootstrap or a hybrid type of them, *leave-one out bootstrap*, that may be work better than *leave-one-out cross-validation.*

The cluster-detecting procedure proposed in this paper does not search for all possibilities of clusters, because the calculation cost is huge. There may be, however, another efficient procedure to find the optimum among the possibilities.

What if we cannot use a panel data set? One possible solution is to assume μ is unknown function of location, which is often called an intensity function, and estimate it with a nonparametric method. Though the method in this paper is regarded as spatial smoothing by decreasing parameters related to clusters, the method employing the intensity function is regarded as a nonparametric smoothing method.

References

[1] Cressie, N. A. C. (1993): *Statistics for Spatial Data*, New York: John Wiley and Sons.

[2] Davison, A. C., and D. V. Hinkley (1997): *Bootstrap Methods and Their Application*, Cambridge: Cambridge University Press.

[3] Geisser, S. (1975): "The Predictive Sample Reuse Method with Application", *Journal of the American Statistical Association*, 70, 320-328.

[4] Hsiao, C. (1986): *Analysis of Panel Data* , Cambridge: Cambridge University Press.

[5] Krugman, P. R. (1993):"First Nature, Second Nature, and Metropolitan Location," *Journal of Regional Science*, 33, 293-298.

[6] Krugman, P. R. (1996):*The Self-organizing Economy*, Cambridge, MA: Blackwell.

[7] Lütkepohl, H. (1991): *Introduction to Multiple Time Series Analysis*, 2nd edition, Berlin: Springer-verlag. "Cross-validatory Choice and Assessment of Statistical Predictions," *Journal Of the Royal Statistical Society*, Ser. B, 36, 111-147.

[8] Stone, M (1974): "Cross-validatory Choice and Assessment of Statistical Predictions," *Journal Of the Royal Statistical Society*, Ser. B, 36, 111-147.

[9] Stone, M (1977): "An Asymptotic Equivalence of Choice of Model by Cross-validation and Akaike's Criterion," *Journal Of the Royal Statistical Society*, Ser. B, 39, 44-47.

Table 1-1: Descriptive Statistics of Selecting of Clusters in Fixed-effects Model

	Number of Clusters	Number of Estimated Clusters Laying across True Clusters
Mean	9.58	0.46
Standard Deviation	1.80	0.58
5 percentile	7	0
10 percentile	7	0
Median	9	0
90 percentile	12	1
95 percentile	13	1

Table 1-2: Destrictive Statistics of Selecting of Clusters in Radom-effects Model

	Number of Clusters	Number of Estimated Clusters Laying across True Clusters
Mean	14.16	1.047
Standard Deviation	2.50	0.81
5 percentile	10	0
10 percentile	11	0
Median	14	1
90 percentile	17	2
95 percentile	18	2

Table 2-1: Estimates of the Cluster Effects and Mean Squared Errors in Fixed-effects Model

	Area	Mean	mean squared error in clustered model	mean squared error in true model	mean squared error in non-clustered model		Area	Mean	mean squared error in clustered model	mean squared error in true model	mean squared error in non-clustered model
UPPER-LEFT CLUSTER	1	2.01	1.00	0.16	1.29	LOWER CLUSTER	19	10.02	1.02	0.10	1.32
	2	2.00	1.00	0.16	1.36		20	10.04	1.06	0.10	1.48
	3	2.06	1.33	0.16	1.41		21	10.01	1.08	0.10	1.43
	7	2.02	0.96	0.16	1.34		22	9.96	1.09	0.10	1.33
	8	2.05	1.07	0.16	1.36		23	9.99	1.05	0.10	1.38
	9	2.07	1.26	0.16	1.41		24	10.02	1.09	0.10	1.33
	13	1.99	1.02	0.16	1.34		25	10.05	1.00	0.10	1.36
	14	1.99	1.04	0.16	1.38		26	9.96	1.02	0.10	1.44
	15	2.10	1.25	0.16	1.33		27	10.02	0.99	0.10	1.34
UPPER-RIGHT CLUSTER	4	4.92	1.32	0.23	1.36		28	10.04	0.93	0.10	1.31
	5	4.99	1.16	0.23	1.57		29	10.04	0.90	0.10	1.26
	6	5.02	1.00	0.23	1.32		30	10.00	0.97	0.10	1.32
	10	4.92	1.26	0.23	1.45		31	9.97	1.13	0.10	1.38
	11	5.00	1.12	0.23	1.49		32	10.02	1.02	0.10	1.37
	12	5.03	0.99	0.23	1.36		33	10.00	1.03	0.10	1.41
	16	4.91	1.45	0.23	1.48		34	10.05	1.02	0.10	1.35
	17	5.05	1.43	0.23	1.71		35	10.03	1.11	0.10	1.55
	18	5.00	1.12	0.23	1.32		36	10.00	1.01	0.10	1.32

Table 2-2: Estimates of the Cluster Effects and Mean Squared Errors in Random-effects Model

	Area	Mean	mean squared error in clustered model	mean squared error in true model	mean squared error in non-clustered model		Area	Mean	mean squared error in clustered model	mean squared error in true model	mean squared error in non-clustered model
UPPER-LEFT CLUSTER	1	2.05	4.61	1.59	4.92	LOWER CLUSTER	19	10.03	5.31	1.20	5.44
	2	2.03	4.71	1.59	5.22		20	9.99	6.14	1.20	6.48
	3	2.08	5.19	1.59	5.40		21	9.93	6.58	1.20	6.92
	7	1.97	4.81	1.59	5.31		22	9.72	7.15	1.20	7.37
	8	1.92	5.48	1.59	5.90		23	9.75	6.34	1.20	6.63
	9	2.00	6.10	1.59	6.47		24	9.79	5.34	1.20	5.37
	13	2.04	5.40	1.59	5.58		25	10.02	4.83	1.20	5.25
	14	1.96	6.26	1.59	6.55		26	9.97	5.54	1.20	5.97
	15	2.01	6.45	1.59	6.70		27	9.95	6.30	1.20	6.91
UPPER-RIGHT CLUSTER	4	4.88	5.65	1.72	5.76		28	10.00	6.35	1.20	6.88
	5	5.05	5.19	1.72	5.50		29	9.91	6.19	1.20	6.70
	6	5.00	4.69	1.72	4.97		30	9.88	5.30	1.20	5.67
	10	4.97	5.79	1.72	5.91		31	9.98	4.75	1.20	5.05
	11	4.96	5.22	1.72	5.69		32	9.88	5.13	1.20	5.43
	12	5.06	5.26	1.72	5.65		33	9.97	5.26	1.20	5.66
	16	4.92	6.91	1.72	7.07		34	10.03	5.64	1.20	5.98
	17	4.87	6.44	1.72	6.80		35	9.88	5.25	1.20	5.69
	18	4.91	5.87	1.72	6.07		36	9.98	4.44	1.20	4.81

Table 3: Comaprion of the OLS Estiamte with the Within-class Estimate

	OLS estimate with estimated clustering structure	Within-class estiamte	OLS estimate with true clustering structure
		Fixed-effects Model	
mean	1.997	1.997	1.996
standard deviation	0.076	0.079	0.067
mean squared error	0.0058	0.0062	0.0045
		Random-effects Model	
mean	1.998	2.000	2.002
standard deviation	0.076	0.075	0.084
mean squared error	0.0058	0.0056	0.0071

Table 4: Estimates of the Parameters with Estimated Clustering Structure and Within-class Estimate

	The Number of Areas within the cluster	Estimate with estimated clustering structure			The Number of Areas within the cluster	Estimate with estimated clustering structure			The Number of Areas within the cluster	Estimate with estimated clustering structure			The Number of Areas within the cluster	Estimate with estimated clustering structure	
			t-value				t-value				t-value				t-value
Cluster 1	5	-8.07	-6.28	Cluster 16	2	-8.46	-6.62	Cluster 31	1	-9.00	-7.20	Cluster 46	1	-8.12	-6.26
Cluster 2	2	-7.39	-5.34	Cluster 17	2	-8.22	-6.40	Cluster 32	1	-9.89	-7.95	Cluster 47	1	-7.88	-5.96
Cluster 3	2	-8.17	-6.38	Cluster 18	1	-10.70	-8.67	Cluster 33	1	-8.60	-6.86	Cluster 48	1	-8.71	-6.78
Cluster 4	2	-7.33	-5.68	Cluster 19	1	-7.68	-6.05	Cluster 34	1	-8.35	-6.37	Cluster 49	1	-8.28	-6.39
Cluster 5	4	-7.28	-5.56	Cluster 20	1	-7.12	-5.47	Cluster 35	1	-12.04	-8.10	Cluster 50	1	-9.25	-7.03
Cluster 6	5	-7.66	-5.79	Cluster 21	1	-7.45	-5.88	Cluster 36	1	-7.11	-5.14	Cluster 51	1	-6.20	-4.64
Cluster 7	2	-8.73	-7.01	Cluster 22	1	-8.80	-6.98	Cluster 37	1	-7.86	-5.81	Cluster 52	1	-7.89	-6.10
Cluster 8	4	-7.17	-5.36	Cluster 23	1	-6.68	-5.04	Cluster 38	1	-6.89	-4.93	Cluster 53	1	-6.96	-5.34
Cluster 9	3	-7.31	-5.49	Cluster 24	1	-8.83	-6.86	Cluster 39	1	-8.06	-5.94	Cluster 54	1	-8.43	-6.62
Cluster 10	2	-7.64	-5.84	Cluster 25	1	-9.99	-8.04	Cluster 40	1	-6.28	-4.56	Cluster 55	1	-9.59	-7.29
Cluster 11	2	-8.47	-6.62	Cluster 26	1	-9.50	-7.61	Cluster 41	1	-6.88	-5.06	Cluster 56	1	-7.52	-5.79
Cluster 12	3	-6.91	-5.05	Cluster 27	1	-10.64	-8.53	Cluster 42	1	-6.47	-4.68	Cluster 57	1	-8.52	-6.61
Cluster 13	2	-6.88	-5.16	Cluster 28	1	-11.56	-9.02	Cluster 43	1	-6.73	-4.97	Cluster 58	1	-8.67	-6.81
Cluster 14	2	-8.77	-6.36	Cluster 29	1	-7.78	-6.13	Cluster 44	1	-7.00	-5.21				
Cluster 15	2	-6.45	-4.81	Cluster 30	1	-7.98	-6.29	Cluster 45	1	-8.34	-6.46				

Clustered Model		t-value
Income per Household	-0.302	-0.868
Land Price	1.384	8.257
Population Density	-2.888	-8.990
Adjusted R square	0.644	

Within-Class Estimate		t-value
Income per Household	0.797	0.685
Land Price	1.545	5.040
Population Density	-4.942	-3.904
Adjusted R square	0.654	

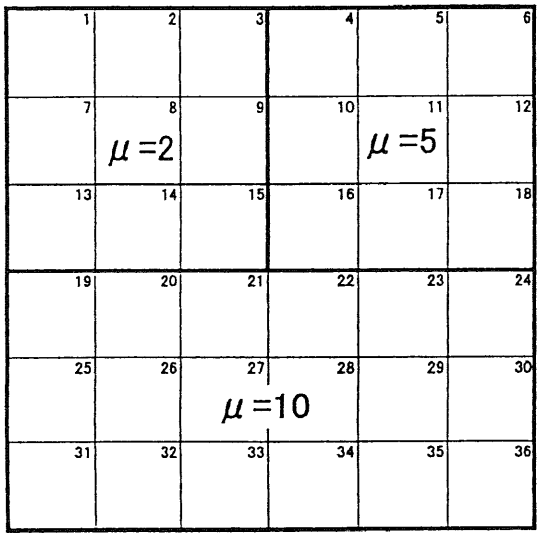


Fig.1 True Structure of Clusters

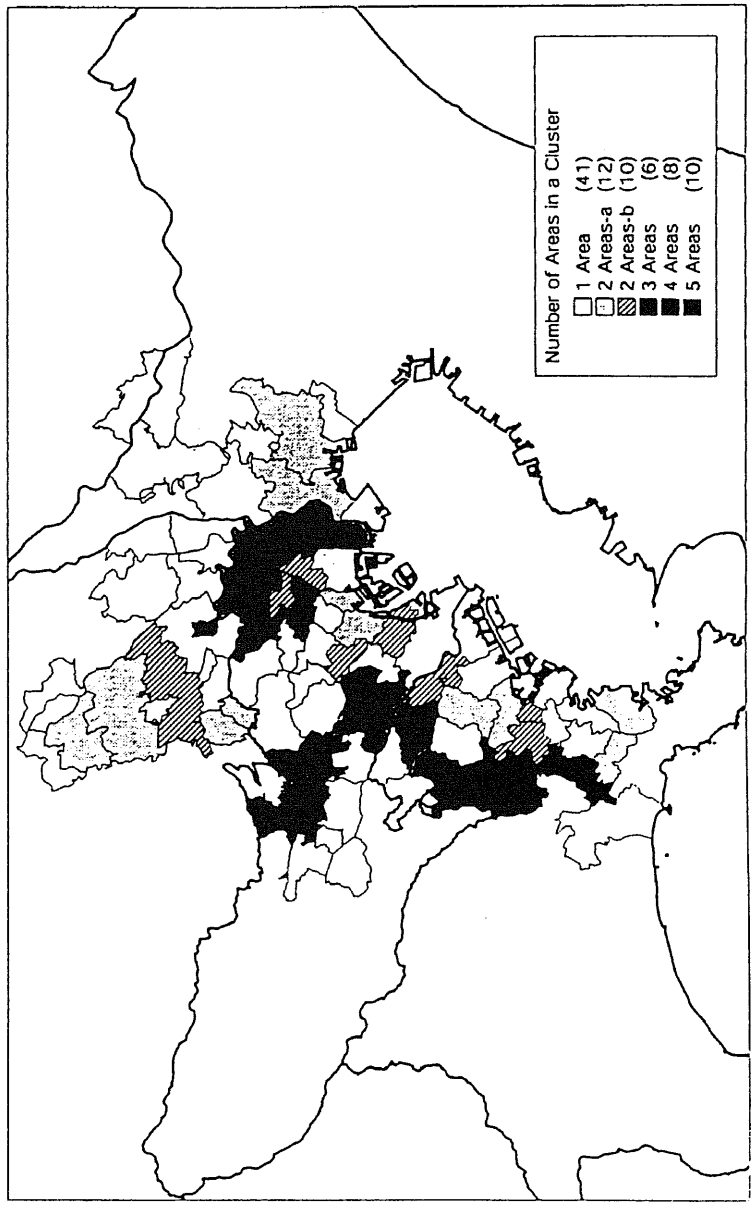


Fig. 2 Clusters in the Number of Housing Start

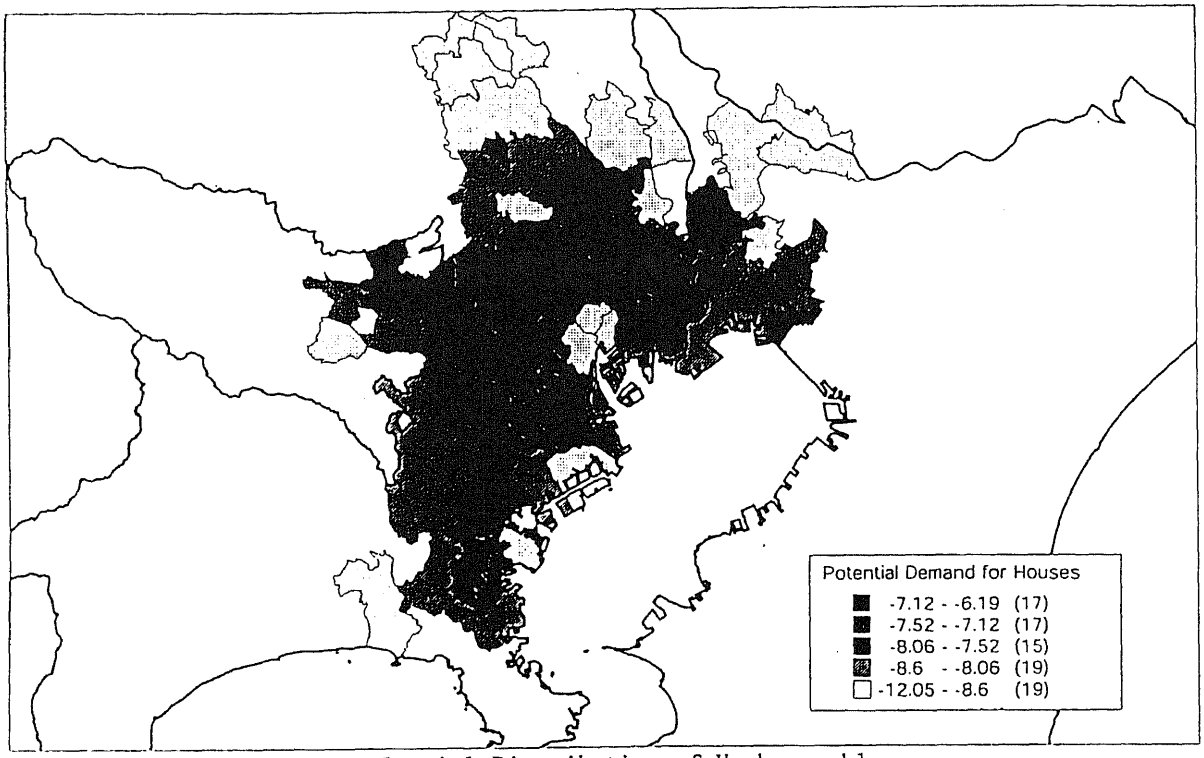


Fig. 3 Spatial Distribution of Unobservable Potential Demand for Houses

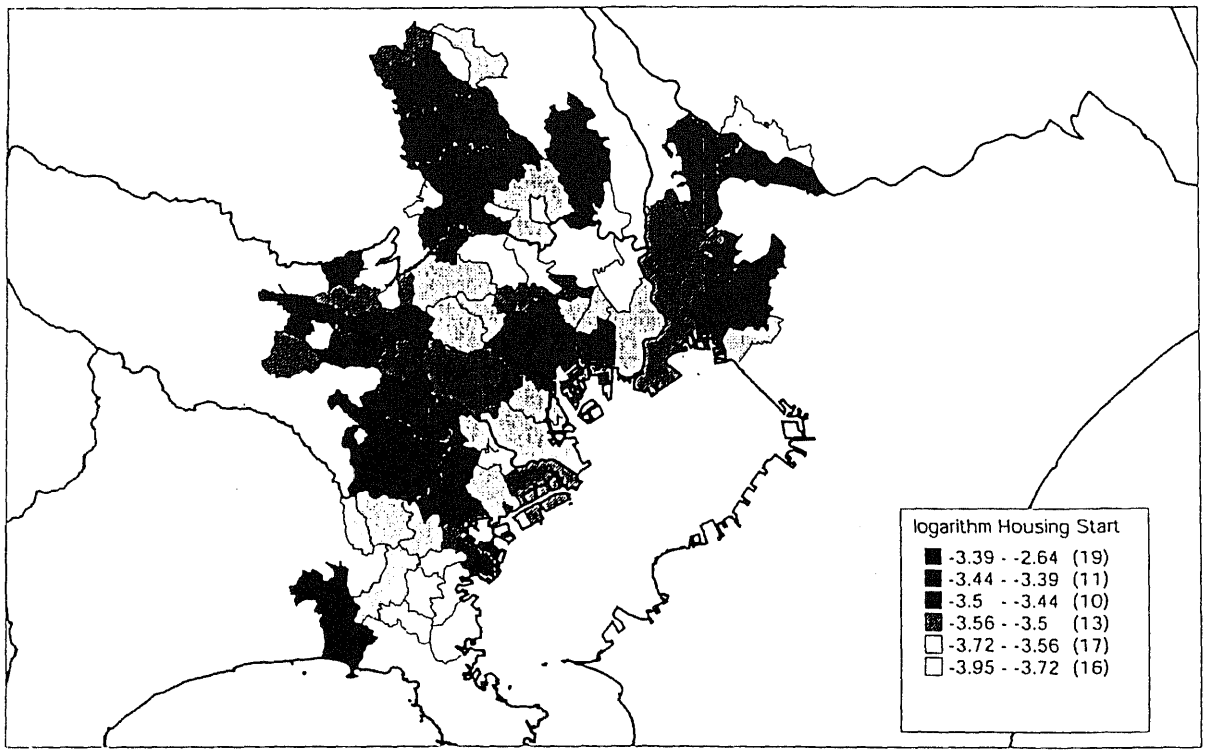


Fig. 4 Crude Map of Logarithm of Housing Start per Household

Testing the One-way Effect and Application

Feng YAO
Faculty of Economics, Kagawa University
2-1 Saiwai-cho, Takamatsu 760-8523, Japan
yao@ec.kagawa-u.ac.jp

This paper provides an approach to testing the measures of one-way effect for cointegrated vector time-series in the presence of trend breaks. We propose the Wald tests of the measures of one-way effect and their computational algorithm. It is an extention of the work of Yao and Hosoya (2000) to the case where trend breaks are explicitly taken into account in the cointegration relationship. The discussions of the one-way effects are based on cointegrated system. Hosoya and Takimoto's (2000) cointegration rank test in the presence of trend breaks is used. On the basis of the proposed method and the derived evidence, the paper presents a causal-structure characterization of the Japanese money and income as well as interest rates in the last forty-four years.

To solve the problems of determining the direction of causality between a pair of time series and also of statistically testing the absence of feedback, Granger (1969) introduced a celebrated definition of causality. Among the earlier representative studies of testing the absence of feedback relation are the Granger test of zero restriction on specific coefficients of a stationary autoregressive representation and the Sims test of the zero restriction on some coefficients in moving-average representation of stationary bivariate processes. For the purpose of quantitative characterization of the feedback relationship between two multivariate time series, Hosoya (1991, 1997) introduced three causal measures which summarizing the interdependency between a pair of time series. Incorporating Johansen's algorithm for the ML estimates and the likelihood ratio tests, in dealing with nonstationary time series processes by error correction model, Yao and Hosoya (2000) proposed the Wald test of the causal measures and a method of confidence-set construction for the causal measures, providing computational algorithm for them. Different from most of the econometric causal analysis literature which seems concerned only with testing Granger's non-causality, Yao and Hosoya (2000) gave a new approach for empirical causal analysis of macroeconomic data based on the concept of the one-way effect measure.

This paper extends the causal inference based on Wald statistics by Yao and Hosoya (2000) to cointegrated processes possessing trend breaks, and exhibits relevant computational procedures. We investigate causal structure among Japanese money, income and interest rates during the period of the first quarter of 1955 through the fourth quarter of 1998. Different from Yao and Hosoya (2000), which dealt with no structural change case, the causal analysis in this paper is based on following cointegrated VAR model which

involves trend break dummy variables,

$$\Delta Z(t) = \alpha\hat{\beta}^* Z(t-1) + \sum_{i=1}^{4} \Gamma(i)\Delta Z(t-i) + \sum_{j=0}^{c} \mu_j D_j(t) + \Phi P(t) + \varepsilon(t),$$

where cointegration rank r and $\hat{\beta}$ ($rank\hat{\beta} = r$) are determined by the test of Hosoya and Takimoto (2000) for $c = 2$ and 3. The reason we chose two or three break points is that in the last forty-four years, Japanese economic growth experienced three major stages of high, medium and low growth. The time points of macro economic structural changes are commonly considered located around the twice of oil crises and the collapse of the "Bubble Economy".

In order to characterize the causal relationship between Japanese money, income and interest rates, we use the one-way effect method based on error correction model with trend breaks. The data used are the quarterly observations of GDP, M2+CD, Call Rates, and Loans & Discouts (LD, all banks and other financial institutions) in the Bank of Japan. Since in our model, the distribution of the likelihood ratio statistic depends upon the location of breaks and the related nuisance-parameters, we present simulation-based estimates of large-sample p-value due to Hosoya and Takimoto (2000)'s algorithm. Our empirical analysis indicates that income notably causes M2+CD; on the other hand, M2+CD causes income very weakly and only in the long-run. The LD causes income but income seems causing the LD only in the very long-run. The interest rates cause both M2+CD and income but not the other way around. Although the estimated causal measure of interest rates to LD is not significant, a small effect of LD to interest rates is observed in the long-run. As regards the policy instrument choice between M2+CD and interest rates to affect the growth of output, our evidence suggests that interest rates might be more effective. The empirical results also show that the effect of the second oil shock is comparatively small.

References

Granger, C.W.J. (1969). Investigating causal relations by cross-spectrum methods, Econometrica, vol.39, no.3, pp.424-38.

Hosoya, Y. (1991). The decomposition and measurement of the interdependency between second-order stationary processes, Probability Theory and Related Fields, vol.88, pp.429-44.

Hosoya, Y. (1997). Causal analysis and statistical inference on possibly non-stationary time series, in:*Advances in Economics and Econometrics: Theory and Application*, Seventh World Congress Vol.III, eds D.M. Kreps and K.F. Wallis, Cambridge: Cambridge University Press, pp.1-33.

Hosoya, Y. and Takimoto, T. (2000). Testing cointegration rank in the presence of structurel changes, Annal Report of the Economic Society, Tohoku University, vol.61, no.4, pp.78-99.

Sims, C.A. (1972). Money, income and causality, American Economic Review, vol.62, no.4, pp.540-52.

Yao, F. and Hosoya, Y. (2000). Inference on one-way effect and evidence in Japanese macroeconomic data, Journal of Econometrics, vol.98, no.2, pp.225-55. *(Free Download* http://www.elsevier.nl/inca/publications/store/5/0/5/5/7/5/)

Spatial resolution enhancement of imagery based on cokriging

Yoji MORISAKI* and Ryuei NISHII**

*Graduate School of Engineering,
Hiroshima University, Kagamiyama 1–7–1, Higashi-Hiroshima 739–8521, Japan
Email: morisaki@mis.hiroshima-u.ac.jp

**Faculty of Integrated Arts and Sciences,
Hiroshima University, Kagamiyama 1–7–1, Higashi-Hiroshima 739–8521, Japan
Email: nishii@mis.hiroshima-u.ac.jp

Abstract — We consider a prediction method based on cokriging for improving quality of images. It is assumed that sets of multivariate data with different spatial resolutions are observed at the same rectangular region such as imagery from satellites. The data of low resolution are corrected by the data of high resolution through cokriging-like method. In this approach, we assume the ordinary covariance structure in cokriging as well as a simple structure called the intrinsic correlation model. Our predictors are applied to the fusion of the multiresolution imaging data from the satellite Landsat. It is seen that they show an excellent performance in comparison with the method in the literature.

I . Introduction

In remote sensing, high spatial resolution images are required. However, in most cases, instruments are not capable of providing such data because of observational limitations. Hence, only one sensor in several sensors may be high-resolution.

As a first example of such situations, suppose that high-resolution panchromatic image and low-resolution multispectral images are given. For instance, the spatial resolution of visible lights of the satellite IKONOS is of 4m, whereas that of panchromatic sensor is of 1m. Also, the resolutions of SPOT are respectively 20m and 10m, and those of LANDSAT 7 are 30m and 15m. Many algorithms for enhancement of lower resolution imagery by combination of high- and low-resolution data are proposed. The most commonly used procedure is the Hue-Saturation-Value (HSV) transform, in which three band spectral data corresponding to red, green and blue at the lower spatial resolution are converted to hue, saturation and value, after which the values are replaced by the panchromatic values of the higher resolution. Then, the result is transformed back to the red-green-blue. Thus, an enhanced multispectral image is obtained. However, the HSV method produces spectral degradation, and this is valid only when the number of images is three.

Here is another example. Suppose that seven bands of the Landsat 5 Thematic Mapper (TM) sensor are given. The Band 6 of the sensor is physically important because it is a measurement on heat radiation. Unfortunately, its resolution is of 120m, while that of the other six bands is of 30m. This difference of resolutions causes many difficulties in analyzing TM data. For example, discriminant analysis on TM data is performed by omitting values of Band 6 in many cases. Such a strategy may lose much information. If one can enhance the resolution of Band 6, 7-dimensional data of high resolution would be helpful in many applications.

Inamura (1988) proposed a deterministic approach for this purpose by estimating proportions of categories. Zhukov, Oertel and Lanzl (1995) and Zhukov et al (1995) took a deterministic and statistical approach. First, the images of high resolution are used for clustering spectrum characteristics. Second, the value of Band 6 at each pixel is replaced by the estimated mean of Band 6 of the corresponding cluster. This method may be powerful, but needs much computation and overfits to the data. See, e.g., Duda and Hart (1973) for clustering techniques, and Zhukov et al (1999) for data fusion. Nishii, Kusanobu and Tanaka (1996) took a fully-statistical approach by employing a multivariate normal distribution for the joint distribution of 7 band values. By assuming a conditional spatial-independence of Band 6 given the high-resolution bands, they correct the values of Band 6 by the conditional expectation.

In this paper we take another statistical approach based on cokriging. The values of low-resolution bands in some pixels are enhanced by the high-resolution bands observed in the first/second order neighborhood of the pixels by taking spatial correlation into account. The proposed procedure is examined by the TM image of Hiroshima City. And then we compare our method with the HSV method and the conditional expectation by Nishii et al (1996), and shows a good performance. See Cressie (1991), Wackernagel (1998), and Chilès and Delfiner (1999) for cokriging methods.

II. Prediction for low resolution bands based on cokriging

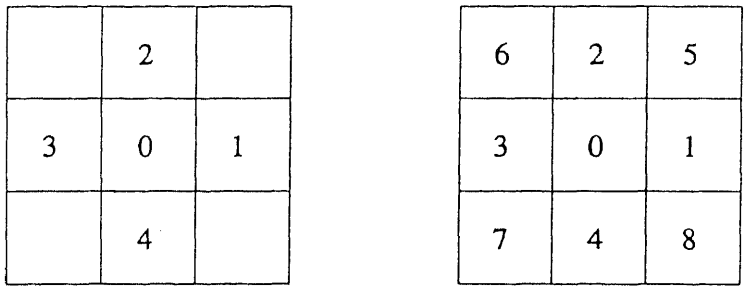


Figure 1: Four-adjacent-pixels window (left) and all-square-pixels window (right) of the center pixel 0.

Figure 1 illustrates two local windows discussed in the article. The values of low resolution bands at the center pixel 0 will be enhanced by the high resolution data at the pixels $0, 1, 2, 3, 4$ or $0, 1, \cdots, 8$. In both cases, we use the same notation for spectrum data. Let

$$Z_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} : (p+q) \times 1, \quad i = 0, 1, \ldots, 8, \tag{1}$$

be spectrum data at the ith pixel, where X_i represents a random vector of low resolution bands, and Y_i represents a random vector of high resolution bands.

We suppose that the random vector Z_i $(i = 0, 1, \ldots, 8)$ have a common mean vector and a common variance-covariance matrix in the respective windows defined as

$$\mathrm{E}[Z_i] \equiv \mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} : (p+q) \times 1, \qquad \mathrm{V}(Z_i) = \Sigma(0) : (p+q) \times (p+q). \tag{2}$$

Corresponding to the following spatial covariance-structures of Z_i, we propose enhancement procedures.

A. Cokriging in four-adjacent-pixels window

Consider the four-adjacent-pixels window. Each pixel is numbered from $i = 0$ to 4, see Figure 1. We suppose the following spatial covariance-structures of Z_i:

$$\mathrm{Cov}(Z_i, Z_j) = \begin{cases} \Sigma(1), & \text{if } d(i,j) = 1 \\ \Sigma(\sqrt{2}), & \text{if } d(i,j) = \sqrt{2} \\ \Sigma(2), & \text{if } d(i,j) = 2 \end{cases} \quad \text{for } i, j = 0, 1, \ldots, 4, \tag{3}$$

where $d(i,j)$ denotes the distance between two pixels i and j. We partition the covariance matrices $\Sigma(k)$ as

$$\Sigma(k) = \begin{array}{c} \\ p \\ q \end{array} \begin{array}{c} \begin{array}{cc} p & q \end{array} \\ \begin{pmatrix} \Sigma_{xx}(k) & \Sigma_{xy}(k) \\ \Sigma_{yx}(k) & \Sigma_{yy}(k) \end{pmatrix} \end{array} : (p+q) \times (p+q) \quad \text{for} \quad k = 0, 1, \sqrt{2}, 2. \tag{4}$$

In our setup, observations on Y_i of high resolution bands, say y_i are available for $i = 0, 1, \ldots, 4$, whereas each X_i of low resolution bands is not observed. We assume that the values of low resolution bands at the center pixel 0 are observations of the average values $\sum_{j=0}^{4} X_j/5 \equiv \overline{X}_5$. Under these conditions, our aim is to predict the values X_0 of low resolution bands based on the observations $y_0, y_1, \ldots, y_4$ and $\overline{x}_5$ in the local window. We shall derive the predictor of X_0 based on cokriging.

The linear predictor of X_0 is defined as

$$\widehat{X}_0 = \overline{X}_5 + B_0 Y_0 + B \sum_{j=1}^{4} Y_j, \tag{5}$$

where B_0 and B are unknown $p \times q$ coefficient matrices.

In the literature, several models in cokriging are proposed, for example, ordinary cokriging, simple cokriging, collocated cokriging and so on. See Cressie (1991), Wackernagel (1998), and Chilès and Delfiner (1999) for these models. Our model is new one in cokriging, which is not discussed in the literature. It is different from the above

models in that $\overline{X}_5$ and Y_0 are available at the center pixel 0.

Also, ask for a predictor that is uniformly unbiased, that is,

$$\mathrm{E}[\widehat{X}_0] = \mu_x + B_0\mu_y + 4B\mu_y = \mu_x. \tag{6}$$

The equation (6) yields the necessary and sufficient condition for unbiasedness as

$$B_0 + 4B = O. \tag{7}$$

Hence, under this condition the generalized mean-square prediction error is given by

$$\sigma_x^2 = \mathrm{E}\left[(\widehat{X}_0 - X_0)'\Sigma_{xx}^{-1}(0)(\widehat{X}_0 - X_0) \right] = \mathrm{tr}\left\{ \Sigma_{xx}^{-1}(0)\left(\frac{4}{25}\Sigma_{xx} + \frac{8}{5}B\Sigma_{yx} + 4B\Sigma_{yy}B' \right) \right\}, \tag{8}$$

where the symbol $'$ denotes the transposition of the vector/matrix, and $\Sigma_{xx} : p \times p$, $\Sigma_{yx} : q \times p$, $\Sigma_{yy} : q \times q$ are submatrices of Σ defined by

$$\Sigma = 5\Sigma(0) - 8\Sigma(1) + 2\Sigma(\sqrt{2}) + \Sigma(2). \tag{9}$$

Hence, the best linear unbiased predictor (BLUP) is obtained by minimizing the generalized mean-square prediction error (8). After differentiating (8) with respect to B, and equating the result to zero, we get the optimal parameters. Consequently, we have BLUP of low resolution bands at the center pixel 0 as

$$\widehat{X}_A = \overline{X}_5 + \frac{4}{5}\Sigma_{xy}\Sigma_{yy}^{-1}(Y_0 - \overline{Y}_4) \quad \text{with} \quad \overline{Y}_4 = \frac{1}{4}\sum_{j=1}^{4} Y_j. \tag{10}$$

B. Cokriging in all-square-pixels window

Next, we consider an all-square-pixels window consisting of 9 pixels, see the right hand side of Figure 1. We suppose that the random vector Z_i $(i = 0, 1, \ldots, 8)$ have the following covariance-structures:

$$\mathrm{Cov}(Z_i, Z_j) = \begin{cases} \Sigma(1), & \text{if } d(i,j) = 1 \\ \Sigma(\sqrt{2}), & \text{if } d(i,j) = \sqrt{2} \\ \Sigma(2), & \text{if } d(i,j) = 2 \\ \Sigma(\sqrt{5}), & \text{if } d(i,j) = \sqrt{5} \\ \Sigma(\sqrt{8}), & \text{if } d(i,j) = \sqrt{8} \end{cases} \quad \text{for } i,j = 0, 1, \ldots, 8. \tag{11}$$

In this case, we suppose that the values of low resolution bands are regarded as observations of the average values $\sum_{j=0}^{8} X_j/9 \equiv \overline{X}_9$. By the unbiased condition, BLUP is of the form:

$$\widehat{X}_B = \overline{X}_9 + B\sum_{j=1}^{4}(Y_j - Y_0) + C\sum_{k=5}^{8}(Y_k - Y_0), \tag{12}$$

where B and C are parameter matrices of size $p \times q$. By minimizing the generalized mean-square prediction error, we estimate them as

$$\widehat{B} = \frac{1}{9}\left\{2\Sigma_{xy}^{**}(\Sigma_{yy}^{**})^{-1}\Sigma_{yy}^{***} - \Sigma_{xy}^{*}\right\}\left\{\Sigma_{yy}^{*} - 4\Sigma_{yy}^{***}(\Sigma_{yy}^{**})^{-1}\Sigma_{yy}^{***}\right\}^{-1} \quad \text{and} \quad \widehat{C} = -(2\widehat{B}\Sigma_{yy}^{***} + \frac{1}{9}\Sigma_{xy}^{**})(\Sigma_{yy}^{**})^{-1},$$

where

$$\begin{aligned}
\Sigma_{xy}^{*} &= 9\Sigma_{xy}(0) - 10\Sigma_{xy}(1) - 2\Sigma_{xy}(\sqrt{2}) + \Sigma_{xy}(2) + 2\Sigma_{xy}(\sqrt{5}), \\
\Sigma_{xy}^{**} &= 9\Sigma_{xy}(0) - 2\Sigma_{xy}(1) - 12\Sigma_{xy}(\sqrt{2}) + 2\Sigma_{xy}(2) + 2\Sigma_{xy}(\sqrt{5}) + \Sigma_{xy}(\sqrt{8}), \\
\Sigma_{yy}^{*} &= 5\Sigma_{yy}(0) - 8\Sigma_{yy}(1) + 2\Sigma_{yy}(\sqrt{2}) + \Sigma_{yy}(2), \\
\Sigma_{yy}^{**} &= 5\Sigma_{yy}(0) - 8\Sigma_{yy}(\sqrt{2}) + 2\Sigma_{yy}(2) + \Sigma_{yy}(\sqrt{8}), \\
\Sigma_{yy}^{***} &= 2\Sigma_{yy}(0) - \Sigma_{yy}(1) - 2\Sigma_{yy}(\sqrt{2}) + \Sigma_{yy}(\sqrt{5}).
\end{aligned}$$

C. Cokriging under the intrinsic assumption

In the hypothesis of the intrinsic correlation model in which the covariance function matrices $\Sigma(k)$ $(k = 1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8})$ are all proportional to a function $\rho(k)$ called a spatial correlation function, that is,

$$\Sigma(k) = \rho(k)\Sigma(0), \quad \text{for } k = 1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}. \tag{13}$$

Under the assumption, the matrix $\Sigma_{xy}\Sigma_{yy}^{-1}$ appearing in the formula is simply given by $\Sigma_{xy}(0)\Sigma_{yy}^{-1}(0)$, which is independent of the spatial correlation function $\rho(\cdot)$. Hence we get the BLUP of $\boldsymbol{X}_0$, say $\widehat{\boldsymbol{X}}_C$, at the center pixel 0 in the four-adjacent-pixels window by

$$\widehat{\boldsymbol{X}}_C = \overline{\boldsymbol{X}}_5 + \frac{4}{5}\Sigma_{xy}(0)\Sigma_{yy}^{-1}(0)(\boldsymbol{Y}_0 - \overline{\boldsymbol{Y}}_4) \quad \text{with } \overline{\boldsymbol{Y}}_4 = \frac{1}{4}\sum_{j=1}^{4}\boldsymbol{Y}_j. \tag{14}$$

Similarly we have the predictor in the all-square-pixels window, say $\widehat{\boldsymbol{X}}_D$ as follows:

$$\widehat{\boldsymbol{X}}_D = \overline{\boldsymbol{X}}_9 + \frac{8}{9}\Sigma_{xy}(0)\Sigma_{yy}^{-1}(0)(\boldsymbol{Y}_0 - \overline{\boldsymbol{Y}}_8) \quad \text{with } \overline{\boldsymbol{Y}}_8 = \frac{1}{8}\sum_{j=1}^{8}\boldsymbol{Y}_j. \tag{15}$$

See, e.g., Wackernagel (1998) for more detail explanation in the intrinsic correlation model.

III. Enhancement of imagery based on a panchromatic image

Many satellites equip panchromatic sensors whose spatial resolution is finer than other sensors, see (a) and (c) of Fig. 2 for example. Concerning the satellite IKONOS, the spatial resolution of visible lights is of 4m, whereas that of panchromatic sensor is of 1m. Also, the resolutions of SPOT are respectively 20m and 10m, and those of LANDSAT 7 are 30m and 15m. The panchromatic images aim to supply supplementary information for the low-resolution images.

The following data fusion technique, called HSV transform, of low-resolution colored imagery and high-resolution panchromatic imagery is widely used. The intensities corresponding to Red, Green and Blue of low-resolution images are transformed into Hue, Saturation, Value. Next, only Value is enhanced by the high-resolution panchromatic image. Then, Hue and Saturation of low resolution and Value of high-resolution are transformed inversely. Thus we obtain pseudo-enhanced colored imagery.

Unfortunately, this method is only available in the case that the dimension of the multispectral imagery is just three. Further, this loses information on mean values of colored images in each pixel of low-resolution.

In this data fusion, we employ the methods proposed in the previous section. The actual Landsat TM data of Hiroshima taken at Oct. 23, 1990 are used for numerical study. We generate a panchromatic image with spatial resolution 30m of size 860×1120 by summing four Bands 1 to 4 corresponding to Blue, Green, Red and Ultra red. Then, we average the visible bands 1 to 3 by 4×4 pixels and get a low-resolution colored image with resolution 120m of the same size, and the inner region of size 250×250 is corrected by of the panchromatic image. The correction is evaluated by the ratio of sums of absolute errors or of square errors:

$$\sum_{i=1}^{250}\sum_{j=1}^{250}\left|\widehat{X}_{ij}^b - X_{ij}^b\right|^\alpha \Big/ \sum_{i=1}^{250}\sum_{j=1}^{250}\left|\overline{X}_{ij}^b - X_{ij}^b\right|^\alpha \quad \text{for } b = 1,2,3 \text{ and } \alpha = 1,2, \tag{16}$$

where X_{ij}^b, $\overline{X}_{ij}^b$ and $\widehat{X}_{ij}^b$ respectively denote original values to be predicted, averaged values by 120m square regions, and the predicted values for the band b at pixels (i,j).

Table 1 compares the HSV method, conditional expectation due to Nishii et al (1996) and our methods. This table shows that the conditional expectation method is best, and the intrinsic model in all-square-pixels window is the second best.

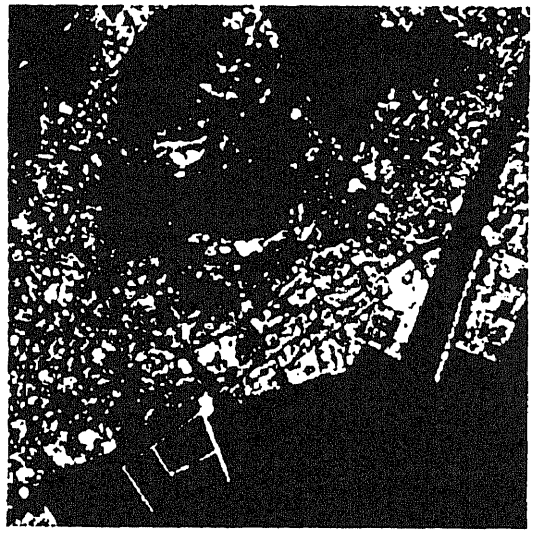
Table 1: The ratios (16) due to the correction methods through the images of size 250×250

Correction methods	Absolute errors ($\alpha=1$)			Square errors ($\alpha=2$)		
	Band1	Band2	Band3	Band1	Band2	Band3
Cokriging (4-adjacent-pixels window)	0.9635	0.9660	0.9696	0.9174	0.9187	0.9190
Intrinsic model (4-adjacent-pixels window)	0.7796	0.7580	0.7533	0.5869	0.5627	0.5568
Cokriging (all-square-pixels window)	0.9518	0.9543	0.9586	0.8911	0.8920	0.8918
Intrinsic model (all-square-pixels window)	0.7155*	0.6805*	0.6767*	0.4794*	0.4432*	0.4373*
Conditional expectation	0.5435**	0.4480**	0.4359**	0.2623**	0.1760**	0.1655**
HSV	1.2646	0.8840	0.8078	0.9520	0.4992	0.4867

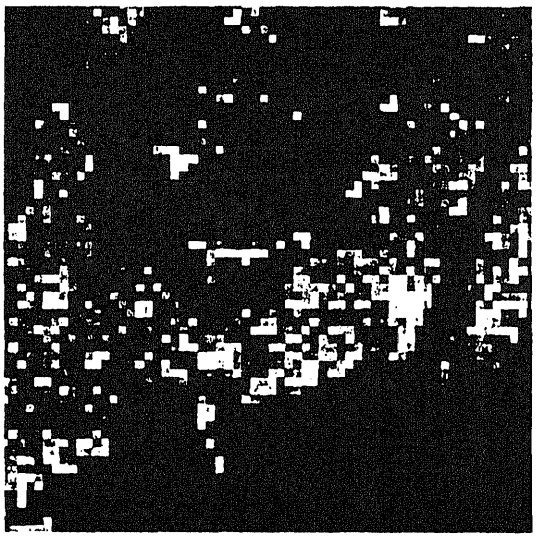
** and * denote the best and the second best values.

The figure (c) of Fig. 2 gives a 4×4 averaged image of the original image (b). Using the panchromatic image (a), (c) is corrected by the three methods: Intrinsic model in all-square-pixels window (d), conditional expectation (e), and HSV (f). In the figure (f), the forest is painted by blue. This point can be also confirmed by Table 1, because the correction of Band 1 (blue) due to HSV is poor. Figures (d) and (e) are clear, but (e) can detect narrow roads.

(a) Panchromatic Image

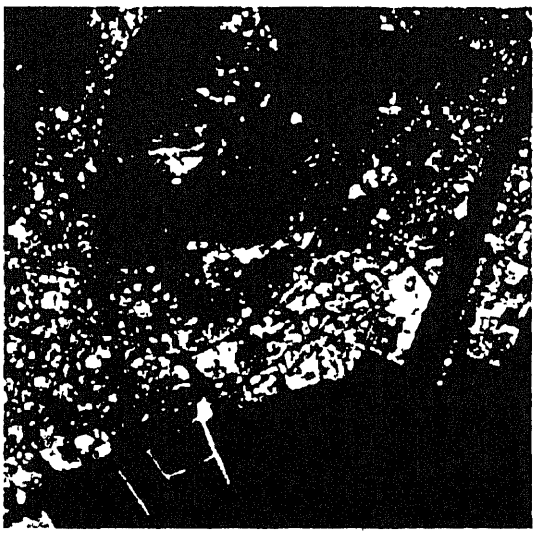


(b) Original color image



(c) Averaged image

(d) Correction by intrinsic model (all-square-pixels window)



(e) Correction by conditional expectation

(f) Correction by HSV

Figure 2: Images of Hiroshima of size 250×250.

IV. Enhancement of infrared images

Next, we consider enhancement of spatial resolution of infrared images. The spatial resolution of Landsat 5 TM sensor except Band 6 is 30m, and the resolution of Band 6 is 120m. Fig. 3 shows a local window consisting of 16 pixels of size 120m×120m at which Band 6 is observed, and Fig. 4 shows gray scale images based on Bands 3 and 6 respectively.

There are some algorithms for enhancement of infrared images based on high-resolution bands as we described in section I . However, these methods seems not to be satisfactory because they can distort the radiometric characteristics of Band 6. In this section, we employ our methods for improvement of the spatial resolution of the infrared images. The proposed methods are applied to Landsat 5 TM images of Hiroshima City, Japan taken at Oct. 23, 1990.

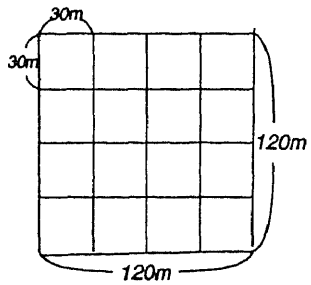


Figure 3: Low- and high-resolutions of the TM sensor

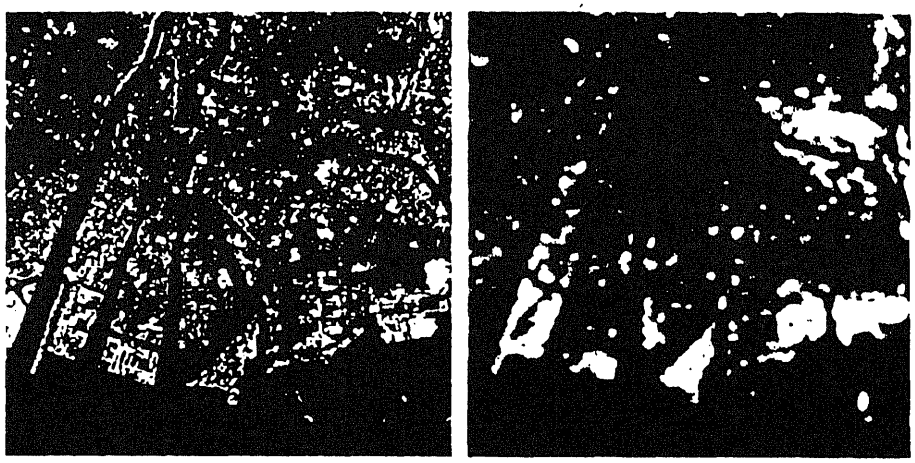


Figure 4: Band 3 (left) and Band 6 (right) TM images of size 300 × 300, Hiroshima, Japan at Oct. 23, 1990

Concerning Band 6, their exact values observed from respective pixels of 30 meters square are not available. Using the wider area of size 2,000 rows × 2,000 columns of the same data used in the previous section, we generate quasi-true values of Band 6 and examine our procedure in the following steps:

(S1) All band data are 4×4 averaged and seven images of size 500×500 are derived.

(S2) Band 6 values are again averaged in 4×4 pixels.

(S3) Choose a subset of the high-resolution bands. Then, Band 6 values generated by (S2) are corrected by our methods based on the selected variables generated by (S1).

(S4) The correction is evaluated in the same way as the previous section.

Table 2 compares conditional expectation due to Nishii et al (1996) and our methods. We see from Table 2 that when all high-resolution bands are used, the cokriging in all-square-pixels window is best, and the cokriging in 4-adjacent-pixels widow comes next. In this case, the remaining methods are poor because the evaluated values are greater than 1. However, the evaluated values are improved in all methods when the high-resolution bands are selected. The minimum value and the selected bands in each method are shown in Table 2. The problem for selection of bands is discussed in Nishii, Kusanobu and Nakaoka (1999).

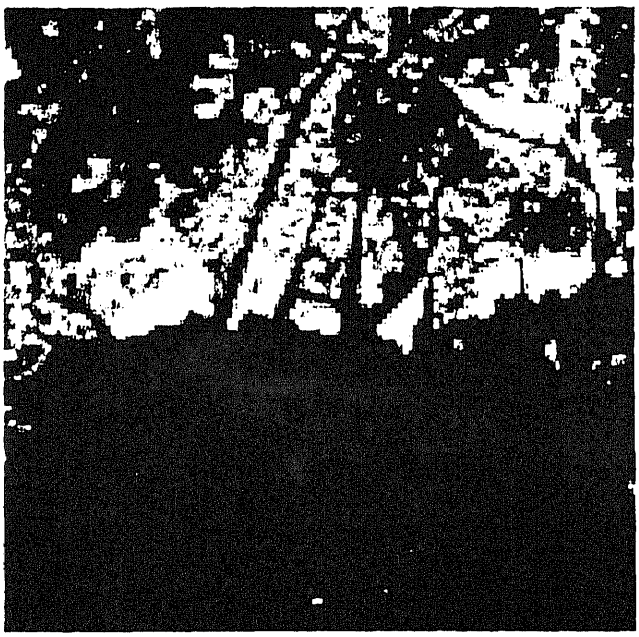
Table 2: Ratios of sums of absolute errors or of square errors due to correction methods

Correction methods	Absolute errors ($\alpha = 1$)		Square errors ($\alpha = 2$)	
	All bands	Selected bands	All bands	Selected bands
Cokriging (4-adjacent-pixels window)	0.9688*	0.9410*(2,3)	0.8696*	0.8275(2,3)
Intrinsic model (4-adjacent-pixels window)	1.1224	0.9464(1)	1.1214	0.8275(1)
Cokriging (all-square-pixels window)	0.9606**	0.9371**(2,3)	0.8577**	0.8199*(2,3)
Intrinsic model (all-square-pixels window)	1.1757	0.9453(1)	1.2105	0.8039**(1)
Conditional expectation	1.2813	1.0076(1)	1.4574	0.9148(1)

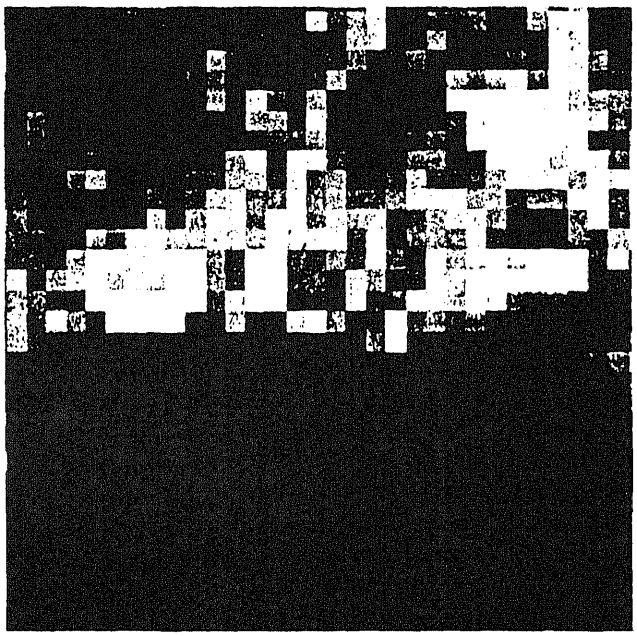
Selected bands are given in parenthesis.

Each figure in Fig. 5 is enlarged to observe in detail. The figure (b) of Fig. 5 gives a 4×4 averaged image of the figure (a). Using all high-resolution bands, (b) is corrected by the two methods: Cokriging in all-square-pixels window (c), conditional expectation (d). Figures (c) and (d) are clearer than the averaged image (b). However,

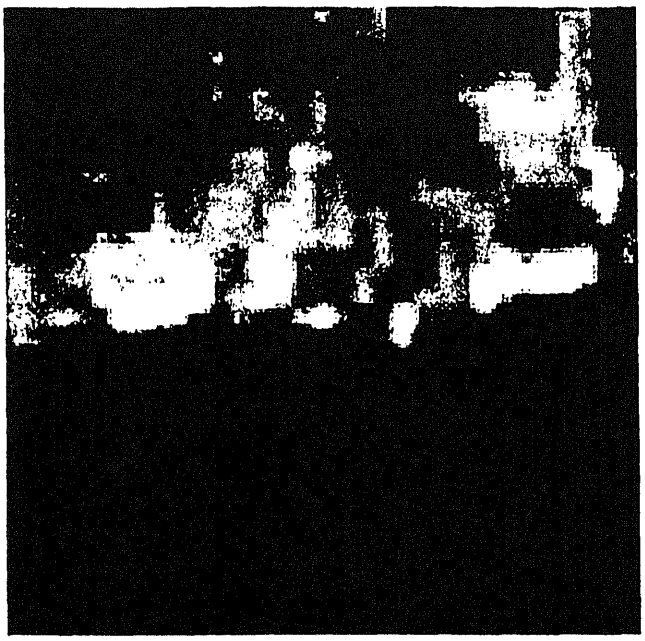
it seems that figure (d) is overfitted by the high-resolution bands, that is, the corrected image loses radiometric information of Band 6.



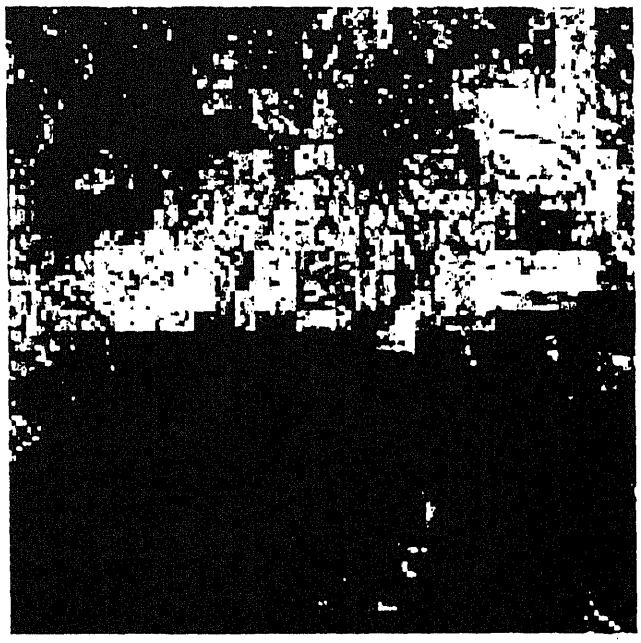
(a) Quasi-true image of Band 6



(b) Averaged image of (a) by 4×4 pixels



(c) Correction by cokriging (all-square-pixels window)



(d) Correction by conditional expectation

Figure 5: Enlarged images of Hiroshima of size 500×500.

V. Concluding Remarks

The purpose of this paper is to enhance low resolution images based on cokriging method. We derived predictors theoretically by considering spatial correlation between pixels.

According to the results of the application to the colored images in section III, the enhanced images by the intrinsic model in all-square-pixels window are better than those by the HSV in the sense that mean values of the colored images in each pixel of low-resolution images are preserved. Thus, our method can be considered as an improvement of the classical HSV method. However, we cannot conclude that our method is best in this case because the results by Nishii et al (1996) are better than those of our method.

In another application to the infrared images in section IV, we could get better results than Nishii et al (1996). In this experiment, we saw that it was difficult to get the high-resolution infrared images as the high-resolution colored images could be obtained in section III. The wavelengths of Band 6 are quite different from those of the other bands, in addition, only Band 6 is a measurement on heat radiation, on the other hand, the other bands are reflective measurements. It can be considered that these differences cause difficulty in enhancement of Band 6 using other 6 bands

Although we applied our methods to only Landsat 5 TM images and showed good performances in this study, we would like to apply those to IKONOS and Landsat 7 TM images, and evaluate the performances in the future.

Finally, the most difficult and important problem is in estimation of the covariance function matrices $\Sigma(k)$ $(k = 0, 1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8})$. The better estimates for the covariance function matrices would improve the predictors. This problem still remains.

References

Chilès, J-P. and Delfiner, P. (1999) *Geostatistics, Modeling Spatial Uncertainty*, New York:Wiley.

Cressie, N. (1991) *Statistics for Spatial Data*, New York:Wiley.

Duda, R. O. and Hart, P.E. (1973) *Pattern classification and scene analysis*, New York:Wiley.

Inamura, M. (1988) "Improvement of spatial resolution for low spatial resolution thermal infrared image using high spatial resolution visible and near-infrared images," *Transactions of the Institute of Electronics, Information, and Communication Engineers*, Part A, vol.J 71-A, 497–504, (In Japanese).

Nishii, R., Kusanobu, S. and Tanaka, S. (1996) "Enhancement of low spatial resolution image based on high resolution bands," *IEEE Transactions on Geoscience and Remote Sensing*, 1151-1158.

Nishii, R., Kusanobu, S. and Nakaoka, N. (1999) "Selection of variables and neighborhoods for spatial enhancement of thermal infrared images," *Communications in Statistics – Theory and Methods* 28 (3 & 4), 965–976.

Wackernagel, H. (1998) *Multivariate Geostatistics*, 2nd completely revised edition, Berlin:Springer.

Zhukov, B., Berger, M., Lanzl, F. and Kaufmann, H. (1995) "A new technique for merging multispectral and panchromatic images revealing sub-pixel spectral variation", *Proceedings of 1995 International Geoscience and Remote Sensing Symposium*, vol. III, 2154–2156.

Zhukov, B., Oertel, D. and Lanzl, F. (1995) "A multiresolution multisensor technique for satellite remote sensing", *Proceedings of 1995 International Geoscience and Remote Sensing Symposium*, vol. I, 51–53.

Zhukov, B., Oertel, D., Lanzl, F. and Reinhäckel, G. (1999) "Unmixing-Based Multisensor Multiresolution Image Fusion", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, No. 3, 1212–1226.

Hybrid expansion for option pricing

東京大学数理　　柏倉賢司

東京大学数理　　吉田朋広

1 Random Limit Expansion for Small Diffusion Processes

Consider a $d = d(1) + d(2)$-dimensional diffusion process
$X^\epsilon = (X^{(1),\epsilon}, X^{(2),\epsilon})_{t\in[0,T]}$ is defined by the stochastic differential equations:

$$\begin{cases} dX_t^{(1),\epsilon} &= V_0^{(1)}(X_t^\epsilon,\epsilon)dt + V^{(1)}(X_t^\epsilon,\epsilon)dw_t^{(1)} \\ dX_t^{(2),\epsilon} &= V_0^{(2)}(X_t^{(2),\epsilon},\epsilon)dt + V^{(2)}(X_t^{(2),\epsilon},\epsilon)dw_t^{(2)}, \\ X_0^{(1),\epsilon} &= x_0^{(1)} \\ X_0^{(2),\epsilon} &= x_0^{(2)}, \end{cases}$$

where:

$w^{(1)}$: $r(1)$-dimensional Wiener process, $w^{(2)}$: $r(2)$-dimensional Wiener process,
coefficients are all in class C_b^∞, (namely, smooth in (x,ϵ) with bounded $\partial_x^i\partial_\epsilon^j$-derivatives for $i \geq 1$ and $j \geq 0$).

Moreover, we assume that

$$V^{(2)}(\cdot,0) = 0 \quad \text{equivalently.}$$

We here consider a functional Z^ϵ defined by

$$Z^\epsilon = \int_0^T \beta(X_t^\epsilon,\epsilon)\nu(dt),$$

where $\beta \in C_\uparrow^\infty(\mathbf{R}^d \times [0,1];\mathbf{R}^k)$ and ν is a random measure on $[0,T]$.

Given a function $r : W^{(1)} \times W^{(3)} \to C_\uparrow^\infty(\mathbf{R}^d \times [0,1];\mathbf{R}_+)$, let

$$G^\epsilon = \exp\left(-\int_0^T r(X^\epsilon,\epsilon)dt\right).$$

Note that time-dependent r can be treated if we extend the original X^ϵ. In order to evaluate the option price, we need to calculate the expected value

$$P\left[G^\epsilon \mathbf{T}(Z^\epsilon)\right]$$

for a measurable function $\mathbf{T}$.
Functionals Z^ϵ and G^ϵ have an asymptotic expansion:

$$Z^\epsilon \sim Z^{(0)} + \epsilon Z^{(1)} + \epsilon^2 Z^{(2)} + \cdots \text{ in } D_\infty(\mathbf{R}^k)$$

as $\epsilon \to 0$ with $Z^{(0)}, Z^{(1)}, \ldots \in D_\infty(\mathbf{R}^k)$ and

$$G^\epsilon \sim G^{(0)} + \epsilon G^{(1)} + \epsilon^2 G^{(2)} + \cdots \text{ in } D_\infty$$

as $\epsilon \to 0$ with $G^{(0)}, G^{(1)}, \ldots \in D_\infty$.
Under the non-degeneracy condition for $Z^{(0)}$, one obtains the asymptotic expansion:

$$P\left[G^\epsilon \mathbf{T}(Z^\epsilon)\right] \sim P\left[\Phi^{(0)}\right] + \epsilon P\left[\Phi^{(1)}\right] + \cdots$$

as $\epsilon \to 0$ for any measurable function $\mathbf{T}$ of at most polynomial growth.

2 Simurations

2.1 Example 1

○モデル

$$\begin{cases} dX_t^\epsilon &= (\mu+\beta v_t^\epsilon)dt + c(v_t^\epsilon)^{\frac{1}{2}}dw_t \\ dv_t^\epsilon &= -\theta(v_t^\epsilon-\alpha)dt + \epsilon(v_t^\epsilon)^{\frac{1}{2}}d\tilde{w}_t \\ X_0^\epsilon &= x_0 \\ v_0^\epsilon &= v_0 \end{cases}$$

ここで、$\alpha,\beta,\theta,\mu,c$は定数、$\tilde{w}=\rho w+\sqrt{1-\rho^2}w^*$, ρは定数,wとw^*は独立
① ヨーロピアンコールオプション ($Z^\epsilon=X_T^\epsilon$ $(T=1),\mathbf{T}(x)=(x-K)_+$)
$G^\epsilon=\exp(-\int_0^T v_t^\epsilon dt)$とする。この時、オプションプライスは次の展開を持つ；

$$P[G^\epsilon(X_T^\epsilon-K)_+]\sim P[\Phi^{(0)}]+\epsilon P[\Phi^{(1)}]+\cdots$$

$$\Phi^{(0)}=G^{(0)}\partial\mathbf{T}(Z^{(0)}),\ \Phi^{(1)}=G^{(0)}\partial\mathbf{T}(Z^{(0)})Z^{(1)}+G^{(1)}T(Z^{(0)}).$$

$$Z^{(0)}=X_T^0,\ Z^{(1)}=D_T,\ G^{(0)}=\exp\left(-\int_0^T v_t^0dt\right),\ G^{(1)}=-G^{(0)}\int_0^T v_t^{(1)}dt.$$

ここで、$D_t:=(\partial_\epsilon)_0X_t^\epsilon$で、次のSDEを満足する:

$$\begin{cases} dD_t &= \beta v_t^{(1)}dt+\frac{1}{2}c(v_t^0)^{-\frac{1}{2}}v_t^{(1)}dw_t \\ D_0 &= 0 \end{cases}$$

また、

$$v_t^{(1)}:=(\partial_\epsilon)_0v_t^\epsilon=\int_0^t e^{-\theta(t-s)}(v_s^0)^{\frac{1}{2}}d\tilde{w}_s$$

◇シミュレーション結果 (μ=0.5,β=0.4,c=2.5,θ=10.0,α=0.5,x_0= 5,v_0=0.1,ρ=0.5)

K=4.5			
ϵ	0.1	0.3	0.5
(1) MC	0.896015	0.892005	0.885187
(2) I	0.900821	0.900821	0.900821
Difference	-0.004806	-0.008815	-0.015633
Diff.rate %	-0.536	-0.989	-1.77
(3) II	0.897841	0.895269	0.885920
Difference	-0.001826	-0.003263	-0.000733
Diff.rate %	-0.204	-0.366	-0.0828
MC=Monte Caro 10000000 回			
I=First order,II=Second order,III=Third order			

K=5.0			
ϵ	0.1	0.3	0.5
(1) MC	0.673349	0.670767	0.664272
(2) I	0.677206	0.677206	0.677206
Difference	-0.003857	-0.006439	-0.012933
Diff.rate %	-0.57	-0.96	-1.95
(3) II	0.674840	0.670108	0.665377
Difference	-0.001419	-0.000659	-0.001105
Diff.rate %	-0.22	-0.09	-0.166
MC=Monte Caro 10000000 回			
I=First order,II=Second order,III=Third order			

K=5.5			
ϵ	0.1	0.3	0.5
(1) MC	0.485223	0.481410	0.479382
(2) I	0.487615	0.487615	0.487615
Difference	-0.002392	-0.006206	-0.008233
Diff.rate %	-0.49	-1.29	-1.72
(3) II	0.486009	0.482797	0.486698
Difference	-0.000786	-0.001387	-0.007317
Diff.rate %	-0.16	-0.29	-1.53
MC=Monte Caro 1000000 回			
I=First order,II=Second order,III=Third order			

② アベレージコールオプション $(Z^\epsilon = \frac{1}{T}\int_0^T X_t^\epsilon dt \ \ (T=1), \mathbf{T}(x) = (x-K)_+)$

G^ϵ は一定とする。

$$P[(\tfrac{1}{T}\int_0^T X_t^\epsilon dt - K)_+] \sim P[\Phi^{(0)}] + \epsilon P[\Phi^{(1)}] + \epsilon^2 P[\Phi^{(2)}] + \cdots$$

$$\Phi^{(0)} = \partial\mathbf{T}(Z^{(0)}),\ \Phi^{(1)} = \partial\mathbf{T}(Z^{(0)})Z^{(1)},\ \Phi^{(2)} = \partial\mathbf{T}(Z^{(0)})Z^{(2)} + \frac{1}{2}\partial^2\mathbf{T}(Z^{(0)})(Z^{(1)})^2.$$

$$Z^{(0)} = \frac{1}{T}\int_0^T X_t^0 dt,\ \ Z^{(1)} = \frac{1}{T}\int_0^T D_t dt,\ \ Z^{(2)} = \frac{1}{T}\int_0^T E_t dt$$

ここで、$E_t := (\partial_\epsilon^2)_0 X_t^\epsilon$ で、次の SDE を満足する:

$$\begin{cases} dE_t &= \beta v_t^{(2)} dt + (\frac{1}{2}c(v_t^0)^{-\frac{1}{2}} v_t^{(2)} - \frac{1}{4}c(v_t^0)^{-\frac{3}{2}}(v_t^{(1)})^2) dw_t \\ E_0 &= 0 \end{cases}$$

また、

$$v_t^{(2)} := (\partial_\epsilon^2)_0 v_t^\epsilon = \int_0^t e^{-\theta(t-s)}(v_s^0)^{-\frac{1}{2}} v_s^{(1)} d\tilde{w}_s$$

◇シミュレーション結果 $(\mu=0.5, \beta=0.4, c=2.5, \theta=2.0, \alpha=1.5, x_0=10, v_0=20, K=15, \rho=0.5)$

ϵ	0.3	0.7	1.0
(1) MC	1.162432	1.184568	1.201118
(2) I	1.145018	1.145018	1.145018
Difference	0.017414	0.039550	0.056100
Diff.rate %	1.5	3.3	4.7
(3) II	1.161097	1.182536	1.198616
Difference	0.001335	0.002032	0.002502
Diff.rate %	0.11	0.17	0.20
(4) III	1.160806	1.180949	1.195375
Difference	0.001626	0.003619	0.005743
Diff.rate %	0.13	0.30	0.47
MC=Monte Caro 10000000 回			
I=First order,II=Second order,III=Third order			

2.2 Example 2

○モデル

$$\begin{cases} dX_t^{(1),\epsilon} &= (\mu_1+\beta_1 v_t^\epsilon)\,dt + c_1\,(v_t^\epsilon)^{\frac{1}{2}}\,dw_t^{(1)} \\ dX_t^{(2),\epsilon} &= (\mu_2+\beta_2 v_t^\epsilon)\,dt + c_2\,(v_t^\epsilon)^{\frac{1}{2}}\,dw_t^{(2)} \\ dv_t^\epsilon &= -\theta\,(v_t^\epsilon-\alpha)\,dt + \epsilon\,(v_t^\epsilon)^{\frac{1}{2}}\,dw_t^{(3)} \\ X_0^{(1),\epsilon} &= x_0^{(1)} \\ X_0^{(2),\epsilon} &= x_0^{(2)} \\ v_0^\epsilon &= v_0. \end{cases}$$

ここで、$\mu_i,\beta_i,c_i,(i=1,2),\theta,\alpha$ は定数。$w^{(1)},w^{(2)},w^{(3)}$ は互いに独立な Wiener 過程。
ペイオフ関数が、$\left(\frac{1}{T}\int_0^T X_t^{(1),\epsilon}dt \vee \frac{1}{T}\int_0^T X_t^{(2),\epsilon}dt - K\right)_+$ で与えられるオプションを考える。即ち,

$$Z^\epsilon = \left(\frac{1}{T}\int_0^T X_t^{(1),\epsilon}dt, \frac{1}{T}\int_0^T X_t^{(2),\epsilon}dt\right), \quad \mathbf{T}(z_1,z_2) = (z_1\vee z_2 - K)_+$$

この時、G^ϵ は一定とするとこのオプションのプライスは次の展開を持つ：

$$P[\mathbf{T}(Z^\epsilon)] \sim P[\Phi^{(0)}] + \epsilon^2 P[\Phi^{(2)}] + \cdots$$

ここで、

$$\Phi^{(0)} = \mathbf{T}(Z^{(0)}), \Phi^{(2)} = \partial^1\mathbf{T}(Z^{(0)})[Z^{(2)}] + \frac{1}{\mathbf{2!}}\partial^2\mathbf{T}(Z^{(0)})[Z^{(1)}]^2$$

$$Z^{(0)} = \left(\frac{1}{T}\int_0^T X_t^{(1),0}dt, \frac{1}{T}\int_0^T X_t^{(2),0}dt\right), Z^{(1)} = \left(\frac{1}{T}\int_0^T D_t^{(1)}dt, \frac{1}{T}\int_0^T D_t^{(2)}dt\right),$$

$$Z^s = \left(\frac{1}{T}\int_0^T E_t^{(1)}dt, \frac{1}{T}\int_0^T E_t^{(2)}dt\right),$$

$D_t^{(i)}, E_t^{(i)}$ $(i=1,2)$ はそれぞれ以下の SDE を満足する,

$$\begin{cases} dD_t^{(i)} &= \beta_i v_t^{(1)}dt + \frac{1}{2}c_i(v_t^0)^{-\frac{1}{2}}v_t^{(1)}dw_t^{(i)} \\ dE^{(i)} &= \beta_i v_t^{(2)}dt + \left(\frac{c}{2}\left(v_t^0\right)^{-\frac{1}{2}}v_t^{(2)} - \frac{c}{4}\left(v_t^0\right)^{-\frac{3}{2}}\left(v_t^{(1)}\right)^2\right)dw_t^{(i)} \\ D_0^{(i)} &= 0 \\ E_0^{(i)} &= 0 \end{cases}$$

また、

$$v_t^{(1)} := (\partial_\epsilon)_0 v_t^\epsilon = \int_0^t e^{-\theta(t-s)}(v_s^0)^{\frac{1}{2}}dw_s^{(3)}$$

$$v_t^{(2)} := (\partial_\epsilon^2)_0 v_t^\epsilon = \int_0^t e^{-\theta(t-s)}(v_s^0)^{-\frac{1}{2}}v_t^{(1)}dw_s^{(3)}$$

◇シミュレーション結果

$(\mu_1=0.5,\beta_1=0.4,c_1=1.5,x_0^{(1)}=10,\mu_2=1,\beta_2=0.8,c_2=1,x_0^{(2)}=10,\theta=1.5,\alpha=5.5,v_0=10)$

K=15			
ϵ	0.5	0.7	0.9
(1) MC	0.387683	0.391336	0.397614
(2) I	0.381100	0.381100	0.381100
Difference	0.006583	0.010236	0.016514
Diff.rate %	1.69	2.62	4.15
(3) II	0.385292	0.389316	0.394682
Difference	0.002391	0.002020	0.002932
Diff.rate %	0.61	0.51	0.73
MC=Monte Caro 10000000 回			
I=First order,II=Second order,III=Third order			

2.3 Example 3

○モデル

$$\begin{cases} dX_t^{(1),\epsilon} &= \mu X_t^{(1),\epsilon} dt + \sigma X_t^{(1),\epsilon}(1+\epsilon\Sigma(X_t^{(1),\epsilon}))dw_t^{(1)} \\ X_0^{(1),\epsilon} &= x_0 \end{cases}$$

ここで、μ,σ は定数、Σ は関数
○ヨーロピアンコールオプション $(Z^\epsilon = X_T^\epsilon \ (T=1), \mathrm{T}(x) = (x-K)_+)$
G^ϵ は一定とする。

$$P[(X_T^\epsilon - K)_+] \sim P[\Phi^{(0)}] + \epsilon P[\Phi^{(1)}] + + \cdots$$

$$\Phi^{(0)} = \partial\mathrm{T}(Z^{(0)}),\ \Phi^{(1)} = \partial\mathrm{T}(Z^{(0)})Z^{(1)}.$$

$$Z^{(0)} = X_T^0,\ \ Z^{(1)} = D_T$$

D_t は次の SDE を満足する；

$$\begin{cases} dD_t &= \mu D_t dt + \sigma(D_t + X_t^0\Sigma(X_t^0))dw_t^{(1)} \\ D_0 &= 0 \end{cases}$$

◇シミュレーション結果

$\Sigma(x) = x^{-\frac{1}{2}}, \mu=0.05, \sigma=0.5, x_0=K=10,$

ϵ	0.1	0.3	0.5
(1) MC	2.374894	2.490895	2.616581
(2) I	2.290992	2.290992	2.290992
Difference	0.073902	0.199903	0.325589
Diff.rate%	3.1	8.0	12.4
(3) II	2.352858	2.476691	2.600323
Difference	0.012036	0.014304	0.016258
Diff.rate %	0.50	0.57	0.62
MC=Monte Caro 1000000 回			
I=First order,II=Second order,III=Third order			

$\epsilon=0.3, \mu=0.1, \sigma=0.3, x_0=K=5,$

$\Sigma(x)$	$\exp(-x)$	$x^{\frac{1}{2}}$	$x^{-\frac{1}{2}}$
(1) MC	0.925646	1.331432	1.006538
(2) I	0.924703	0.924703	0.924703
Difference	0.000943	0.406729	0.081835
Diff.rate %	0.09	30.6	8.24
(3) II	0.926075	1.333830	1.005026
Difference	-0.000429	-0.002379	0.001512
Diff.rate %	-0.04	-0.32	0.15
MC=Monte Caro 1000000 回			
I=First order,II=Second order,III=Third order			

参考文献

[1] Kim, Yong-Jin, Kunitomo, N.: Pricing options under stochastic interest rates: a new approach. Discussion paper F-series, University of Tokyo (1998).

[2] Kim, Yong-Jin, Kunitomo, N.: Effects of stochastic interest rates and volatility on contingent claims. Discussion paper F-series, University of Tokyo (2000).

[3] Kunitomo, N., Takahashi, A.: The asymptotic expansion approach to the valuation of interest rates contingent claims. Discussion paper F-series, University of Tokyo (1995).

[4] Takahashi, A.: An asymptotic expansion approach to pricing financial contingent claims. Asia-Pacific Financial Markets, **6**, 115-151 (1999)

[5] Soerensen, M., Yoshida, N.: Random limit expansion for small diffusion processes. Unpublished manuscript (1998)

[6] Yoshida, N.: Asymptotic expansion for statistics related to small diffusions J. Japan Statist. Soc. **22**, 139-159 (1992)

合成を考慮した衛星搭載レーダの受信波の数学モデル

慶大理工 青木　義充

1　はじめに

衛星に搭載されているレーダーの受信波について考える。受信波を解析する際に従来用いられている手法は、地表面に存在するレーダーの散乱点 (反射点) は離散的であるということを前提にしている (末尾記載の引用文献を参照)。しかし、実際には散乱点は連続的に存在しているはずであり、受信波の解析は、厳密にはこのことを考慮に入れてなされるべきである。

工学の分野では、いわゆる「工学的な直感」にもとづいて問題の処理が行われることが多い。そのため、この問題についても、理論的、数学的にきちんとした議論は、これまでされてこなかった。本研究は、散乱点は連続的に存在するという現実に沿った厳密な議論の手始めとして、より自然な形の受信波を定式化し、その性質について調べることを目的としている。

2　従来の考え方による受信波の形

以下の議論では、地表面の撮影に使われる合成開口レーダーを対象とする。このレーダーの送信波 $s(t)$ は、

$$s(t) = \begin{cases} \exp\left\{2\pi i(f_0 t + \frac{B}{T}t^2)\right\} & , \quad -\frac{T}{2} \le t \le \frac{T}{2} \\ 0 & , \quad \text{その他,} \end{cases}$$

で与えられる。ここで、f_0, B, T はそれぞれ送信周波数、周波数帯域幅、パルス幅と呼ばれ、各レーダーに固有の値になる。いま、衛星の真下を原点にとると、送信波 $s(t)$ に対する地表面上のある点 y からの受信波 $s_r(t,y)$ は、

$$s_r(t,y) = \begin{cases} A(y)\exp\left[2\pi i\left\{f_0(t-\tau(y)) + \frac{B}{T}(t-\tau(y))^2\right\}\right] & , \quad -\frac{T}{2} \le t-\tau(y) \le \frac{T}{2} \\ 0 & , \quad \text{その他,} \end{cases}$$

になる。ここで、$A(y)$ は y での振幅変化を表し、$\tau(y) = 2\sqrt{y^2+h^2}/c$ は衛星と地表面上の点 y との往復時間を表す関数である。$c = 3\times 10^8$ [m/s] は光速で、h はレーダーを搭載した衛星の地表からの高度である (各衛星に固有の値)。$A(y)$ は、一般には何らかの複素関数になる。

従来は、散乱点は観測幅内に離散的に存在するという仮定のもとで、受信波の処理が行われてきている。つまり、観測幅内の散乱点は $y_1, y_2, \ldots,$ であるとし、各散乱点からの波、$s_r(t,y_k)$ それぞれに対して圧縮処理を行なっている。

3　合成を考えた受信波

ここで、時刻 t で観測される受信波は、散乱点がレーダー照射幅内に連続的に存在している為、その時刻を起点とした次のような積分形の合成波 $H(t)$ を考えた方がより厳密な表現になる。

$$\begin{aligned} H(t) &= \int_{\xi(t)}^{\xi(t+T)} s_r(t,y)\,dy \\ &= \int_{\xi(t)}^{\xi(t+T)} A(y)\exp\left[2\pi i\left\{f_0(t-\tau(y)) + \frac{B}{T}(t-\tau(y))^2\right\}\right] dy \\ &= \int_{t}^{t+T} A\left(\xi(u)\right)\exp\left[2\pi i\left\{f_0(t-\tau(y)) + \frac{B}{T}(t-\tau(y))^2\right\}\right]\xi^{(1)}(u)\,du. \end{aligned} \tag{1}$$

ただし、定義域は $t > 2h/c$ であり、$\xi(u) = \tau^{-1}(u)$ と置いている。また、$\xi^{(j)}$ は、ξ の j 次導関数である。

4 受信波の近似式に関して

理論的、数学的には、実際の受信波は (1) で与えられるとみなすのが自然であるが、このままでは $H(t)$ の挙動がよく見えない。そこで、この近似式を考えていく。ここで、受信波 $H(t)$ の近似式は散乱時に起きる振幅の変化を表す y についての関数 $A(y)$ に依存して近似方法が分けられる。ここでは代表的な例として次の二種類を紹介する。

1. 反射率が一定の場合 $A(y)=1$ とした場合に受信波 $H(t)$ は $t>\frac{2h}{c}$ で

$$H(t) \simeq F(t) = \frac{1}{2\pi i f_0}\left[\xi^{(1)}(t) - \xi^{(1)}(t+T)\right]$$

と近似できる。また、その時の誤差は、

$$\left|H(t) - \widehat{H}(t)\right| \leq \frac{(2\pi+3)}{12\pi f_0^2}\left\{\xi^{(2)}\left(t+T-\frac{1}{f_0}\right) - \xi^{(2)}\left(t-\frac{1}{f_0}\right)\right\}$$

と評価できる。ここで、$F(t)$ は単調減少の関数であり、反射率が一定のところからの受信波は波形でないことが分かる。

2. 反射率が不連続に変化する場合 $A(y)$ が地表面上のある点 y_0 を境にして反射率が変化する場合を考える。

$$A(y) = \begin{cases} 1 & , \quad y \leq y_0, \\ \rho & , \quad y > y_0 \end{cases}$$

但し $0 \leq \rho \leq 1$ とする。この条件の下では、受信波 $H(t)$ は $t_0 - T \leq t \leq t_0$ の範囲で

$$\begin{aligned} H(t) &\simeq G(t) \\ &= \frac{1}{2\pi i f_0}\left\{\xi^{(1)}(t) - \rho\,\xi^{(1)}(t+T)\right\} - \frac{1-\rho}{2\pi i f_0}\,\xi^{(1)}(t_0)\exp\left\{2\pi i f_0(t-t_0)\right\} \end{aligned}$$

と近似が行える。実際、その時の誤差は次のように評価できる。

$$\begin{aligned} &|H(t) - G(t)| \\ &\leq \frac{2\pi+3}{12\pi f_0^2}\left[\left\{\xi^{(2)}\left(t+\frac{[(t_0-t)\,f_0]}{f_0}-\frac{1}{f_0}\right) - \xi^{(2)}\left(t-\frac{1}{f_0}\right)\right\}\right. \\ &\qquad \left. +\rho\left\{\xi^{(2)}\left(t+T-\frac{1}{f_0}\right) - \xi^{(2)}\left(t+\frac{[(t_0-t)\,f_0]}{f_0}\right)\right\}\right] \\ &\quad +\frac{(1-\rho)\{(t_0-t)\,f_0 - [(t_0-t)\,f_0]\} + \rho}{f_0^2}\left(1+\frac{1}{2\pi}\right)\left|\xi^{(2)}\left(t+\frac{[(t_0-t)\,f_0]}{f_0}\right)\right| \end{aligned}$$

ここで、$[a]$ は a の整数部分としてある。
$G(t)$ の形より、反射率が変化をする場合では変化を起こしている場所からの波形が返ってくる事が分かる。

発表当日では誤差の評価方法に関して詳細な説明を与える。

参考文献

[1] C. J. Oliver Synthetic-aperture radar imaging. *Applied Physics 22, 871-891,* 1989

[2] 飯坂　譲二　(監修) 合成開口レーダハンドブック. 朝倉書店　, 1998

[3] JERS-1 DATA USERS HANDBOOK. Remote Sensing Technology Center of Japan (RESTEC), 1994

時空間データ解析 講演

「石油鉱業における地球統計学
（地質推計学）の現状紹介」

2000年12月8日

日本オイルエンジニアリング(株)
栗原正典

JAPAN OIL ENGINEERING Co. Ltd.

アウトライン

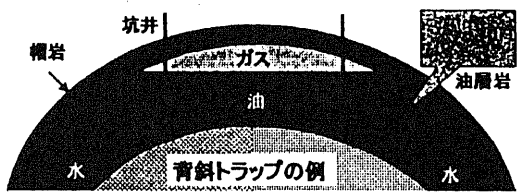
- 石油開発の流れと評価スタディ
- 油層評価スタディにおける油層キャラクタリゼーションと地球統計学の利用
- 地球統計学的手法による複数分野のデータの融合例

JAPAN OIL ENGINEERING Co. Ltd.

石油開発の流れと評価スタディ

JAPAN OIL ENGINEERING Co. Ltd.

石油鉱床(油層)

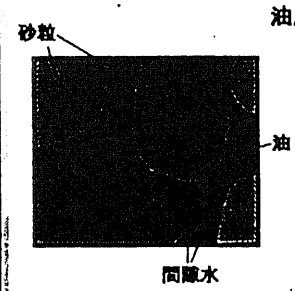


トラップの形態：構造トラップ(背斜、断層)、層位トラップ(尖滅等)
代表的油層深度：500～5,000 m
代表的油層圧力：静水圧力の(1+α)倍
通常の回収法による油回収率：5～40%

JAPAN OIL ENGINEERING Co. Ltd.

油層岩特性

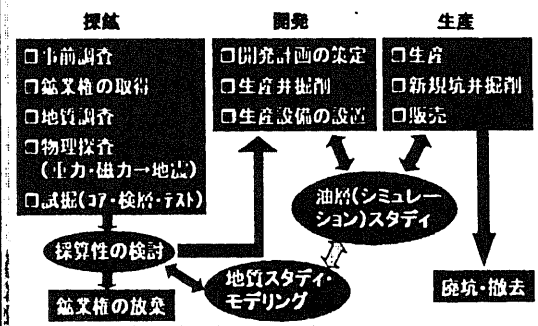
油層岩内の孔隙と流体：油層流体(油、ガス、水)は油層岩孔隙内に貯留



孔隙率(φ)：油層岩内の孔隙の割合→埋蔵量
浸透率(k)：油層岩内流体の流れ易さ(孔隙率と正の相関を示すことが多い)→流動特性(圧力、等)
流体飽和率(S)：孔隙内流体の占有率→埋蔵量・流動特性

JAPAN OIL ENGINEERING Co. Ltd.

石油開発の流れ



JAPAN OIL ENGINEERING Co. Ltd.

石油探鉱・開発における(従来型)評価スタディ

□探鉱・開発・生産の各段階で実施・更新
→埋蔵量の算定、開発計画の策定、油層挙動予測
□ジオロジスト・ジオフィジシスト・レザバーエンジニア・施設エンジニアが参画

データ収集:
コア・検層・坑井テスト
生産・地質・物探

→ データ分析 → 地質モデルの構築

油層シミュレーションスタディ
□油層モデルの構築
□油層モデルの改良
(ヒストリーマッチング)
□将来挙動予測

□地上施設の設計
□経済検討

リスク解析

主としてジオロジスト
ジオフィジシストが従事

シリーズ・モデリング

地質情報と矛盾無きように評価と修正

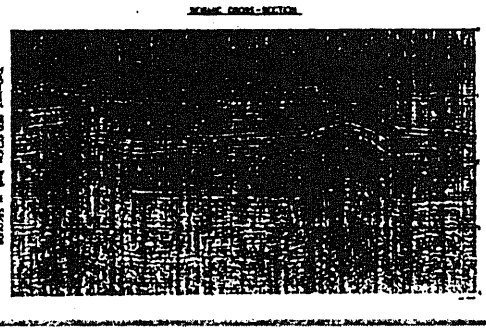
JAPAN OIL ENGINEERING Co. Ltd.

油層評価の基になるデータ

コア・データ
検層データ
坑井試験データ
生産・休止
油層
電波・超音波
地震探査データ
層厚データ

JAPAN OIL ENGINEERING Co. Ltd.

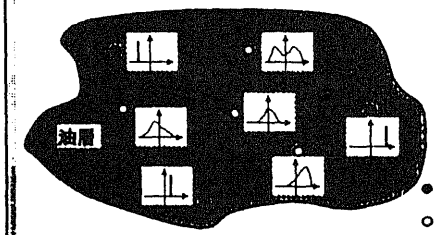
地震探査結果解析例



JAPAN OIL ENGINEERING Co. Ltd.

油層評価(油層モデリング)の概念

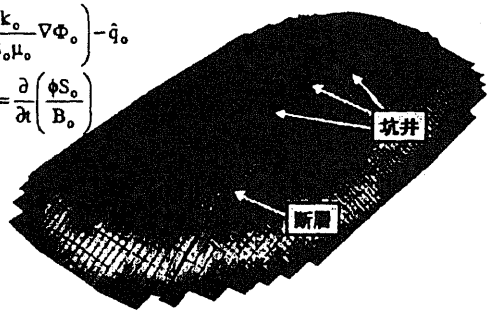
測定点(坑井)におけるデータ値を基に、非測定点における油層特性値(の確率分布)を推定 ← 広域データも参照可能



JAPAN OIL ENGINEERING Co. Ltd.

油層モデル例

$$\nabla \cdot \left(\frac{k_o}{B_o \mu_o} \nabla \Phi_o \right) - \hat{q}_o = \frac{\partial}{\partial t} \left(\frac{\phi S_o}{B_o} \right)$$



JAPAN OIL ENGINEERING Co. Ltd.

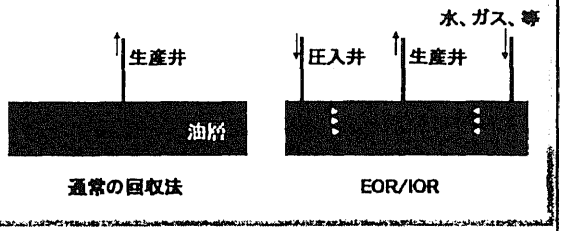
油層評価スタディにおける油層キャラクタリゼーションと地球統計学の利用

JAPAN OIL ENGINEERING Co. Ltd.

石油の回収法

□通常の回収法(1次回収法) → 低回収率

□2・3次回収法、EOR/IOR:
外部から排油エネルギーを油層に注入 → 回収率の増加



JAPAN OIL ENGINEERING Co. Ltd.

地球統計学の導入

1970年代～1980年代後半: EORの研究

良好な回収率・掃効率

実験室での研究/理論研究
→コアレベル、
均質油層を仮定

フィールドに適用→油層不均質性の影響→低い回収率・掃効率

1980年代後半～: 油層キャラクタリゼーション(RC)の研究

□地質学:シーケンス層序、□物理探査:3次元・4次元地震探鉱

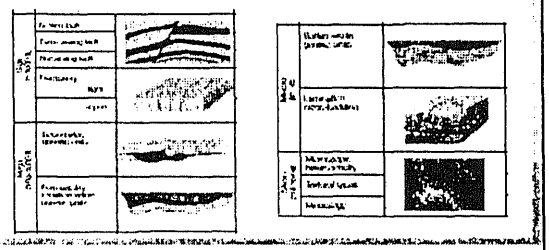
□油層工学:地球統計学→RCの中心的役割

JAPAN OIL ENGINEERING Co. Ltd.

油層の不均質性

種々のスケールの不均質性

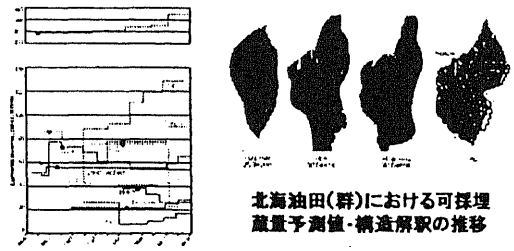
Gigascopic (inter-well level) → Megascopic (grid level)
→ Macroscopic (core level) → Microscopic (pore level)



JAPAN OIL ENGINEERING Co. Ltd.

油層の不確実性(1)

不均質性に基づく不確実性:
開発が進むに連れて解明される予期せぬ油層特性

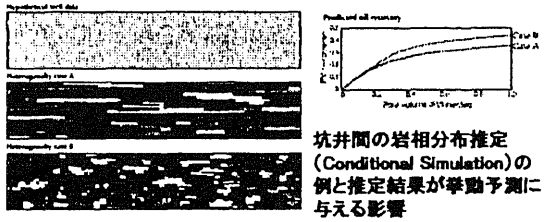


北海油田(群)における可採埋蔵量予測値・構造解釈の推移

JAPAN OIL ENGINEERING Co. Ltd.

油層の不確実性(2)

不確実な不均質性のモデリング(測定点間):
モデルにより異なる油層挙動予測結果
→ 正確な不均質性の推定が重要



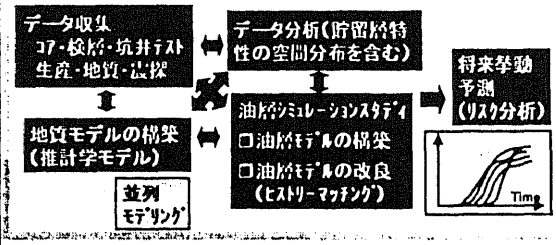
坑井間の岩相分布推定(Conditional Simulation)の例と推定結果が挙動予測に与える影響

JAPAN OIL ENGINEERING Co. Ltd.

地球統計学を適用したRC

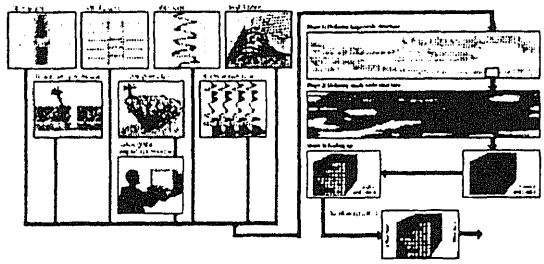
□有効な地質・地球物理・油層・生産データをすべて利用

□ジオロジスト・ジオフィジシスト・レザバーエンジニア・生産エンジニアが同時参画
→ 学際的(multi-disciplinary)アプローチ



JAPAN OIL ENGINEERING Co. Ltd.

油層モデリング



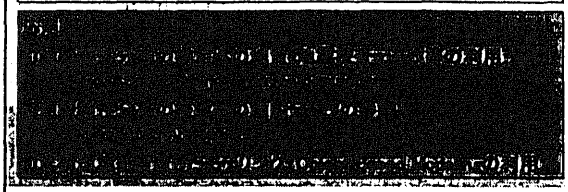
地球統計学を適用したRCにおける油層モデリングの流れ

JAPAN OIL ENGINEERING Co. Ltd.

石油鉱業における地球統計学の利用

基礎編

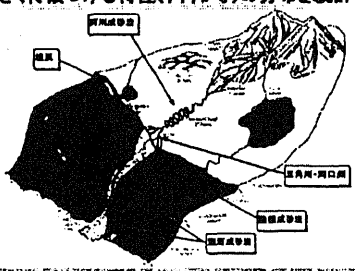
- ロックタイプの特定：検層データを基に岩石特性のタイプ分け
  →孔隙率 vs. 浸透率、多相流動特性等の推定
- 岩相（砂岩、頁岩、等）分布の推定：Indicator Simulation
- 孔隙率・浸透率・流体飽和率分布の推定：Gaussian Simulation



JAPAN OIL ENGINEERING Co. Ltd.

堆積環境を考慮した統計学的モデリング

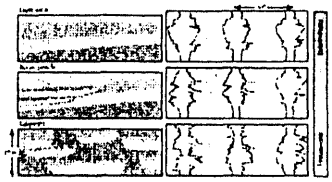
- 油層が堆積した場所、構造運動により油層を大きく特徴づける
- 油層を大きく特徴づける特性（チャネル等）の分布を統計学的に推定



JAPAN OIL ENGINEERING Co. Ltd.

従来の地質モデルと目的関数モデル

- 従来の地質モデル：坑井対比→坑井間のレイヤー特性を推定（GB）
- 目的関数モデル：油層の大きな特徴（頁岩、チャネル、等）の分布を推定



Layer Cake、Jigsaw Puzzle、
Labyrinthの各モデル

Flow Channel分布の推定

JAPAN OIL ENGINEERING Co. Ltd.

地球統計学的手法による
複数分野のデータの融合例

JAPAN OIL ENGINEERING Co. Ltd.

岩相分布データ（地質データ）
を融合した浸透率分布の推定

- 石油鉱業界で実施しているセミナーでの演習問題の1つ
- 石油公団・技術部 難波氏、ジャパン石油開発（株） 桐沢氏の共同研究結果を参考（石油公団・石油開発技術開発センター1998年度年報参照）
- コア・検層・岩相分布データ → 浸透率分布の推定
- ハードデータ：坑井における浸透率
- ソフトデータ：孔隙率分布と岩相分布を融合（ACE法を適用）
- マルコフ・ベイズ法によりハード・ソフトデータを利用

JAPAN OIL ENGINEERING Co. Ltd.

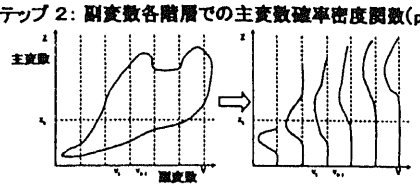
マルコフ-ベイズ法(1)

ソフトデータ(副変数)を参照したインディケータ・シミュレーション

- □主変数間の共分散→副変数間・主変数-副変数間の共分散
- □副変数として利用できる変数は1つのみ

ステップ 1：主変数-副変数間のスキャタグラムの作成

ステップ 2：副変数各階層での主変数確率密度関数(pdf)の作成



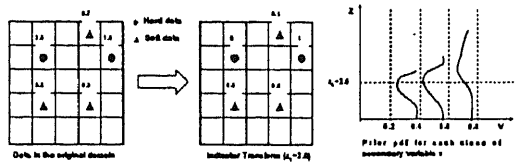
主変数-副変数間のスキャタグラム　副変数の各階層における事前 pdf

マルコフ-ベイズ法(2)

ステップ 3：主変数、副変数のインディケータ変換

主変数：$I(u_1;z_k)=\begin{cases}1, & \text{If } Z(u_1)\le z_k\\ 0, & \text{If } Z(u_1)>z_k\end{cases}$

副変数：$Y(u_2;z_k)=\text{Prob}\{Z(u_2)\le z_k \mid V(u_2)\in(v_l,v_{l+1}]\}$



マルコフ-ベイズ法(3)

ステップ 4：主変数インディケータのバリオグラム作成

ステップ 5：副変数、主変数-副変数(インディケータ)共分散の推定

副変数：$C_Y(h;z_k)=B^2(z_k)-C_I(h;z_k),\quad \forall h>0$

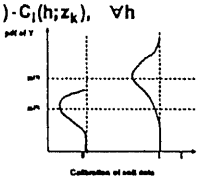
$=|B(z_k)|-C_I(h;z_k),\quad h=0$

主変数-副変数：$C_{IY}(h;z_k)=B(z_k)-C_I(h;z_k),\quad \forall h$

$B(z)=m^{(1)}(z)-m^{(0)}(z)\quad \in[-1,+1]$

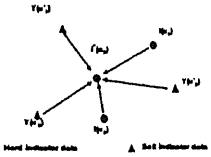
$m^{(1)}(z)=E\{Y(u;z_k)\mid I(u;z_k)=1\}$

$m^{(0)}(z)=E\{Y(u;z_k)\mid I(u;z_k)=0\}$



マルコフ-ベイズ法(4)

ステップ 6：未知主変数インディケータの推定



未知主変数インディケータ $I(u_0;z_k)$のコクリギング法による推定

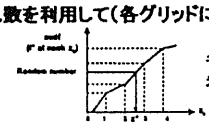
$$I^*(u_0;z_k)=F(z_k)+\sum_{\alpha_1=1}^{n_1}\lambda_{\alpha_1}[I(u_{\alpha_1};z_k)-F(z_k)]+\sum_{\alpha_2=1}^{n_2}\lambda'_{\alpha_2}[Y(u'_{\alpha_2};z_k)-E\{Y(u;z_k)\}]$$

$F(z_k)=E\{I(u;z_k)\}$

マルコフ-ベイズ法(5)

ステップ 7：未知主変数分布の推定

- □あるしきい値z_kに対して推定された主変数インディケータはその値に対応する事後制約累積分布関数(ccdf)値と同値
- □しきい値を変えステップ3～6を再試行し、複数のzに対する(各グリッドにおける)事後制約累積分布関数曲線を作成
- □乱数を利用して(各グリッドにおける)主変数値を推定



モンテカルロ法による未知主変数の推定

ステップ 8：他のリアリゼーションの作成

推定すべきグリッドの順番を変えてステップ6、7を繰り返す

ACE法(1)

Alternating Conditional Expectation：ノン・パラメトリックな回帰法
(J. of the American Statistical Association, Sep. 1985)

- □油層モデリングでは従属確率変数と独立確率変数間の回帰がしばしば試みられる(浸透率 vs. 孔隙率、等)
- □複数の独立確率変数との回帰
  - → 従属確率変数の予測値の精度向上、ソフトデータの利用
  - (検層データ、地震探査データ、等)
- □従属確率変数Yと複数の独立確率変数 X_iを$\theta(Y)$と$\phi_i(X_i)$に変換
- □$\theta(Y)$は平均 0・分散 1、$\phi_i(X_i)$は平均 0
- □$\theta(Y)$と$\phi_i(X_i)$は $\theta(Y)=\sum_{i=1}^{p}\phi_i(X_i)$で関係付けられる

ACE法(2)

□θ(Y)、$\phi_i(X_i)$は誤差の分散

$$e^2(\theta,\phi_1,\ldots,\phi_p)=E\left\{\left[\theta(Y)-\sum_{i=1}^{p}\phi_i(X_i)\right]^2\right\}$$

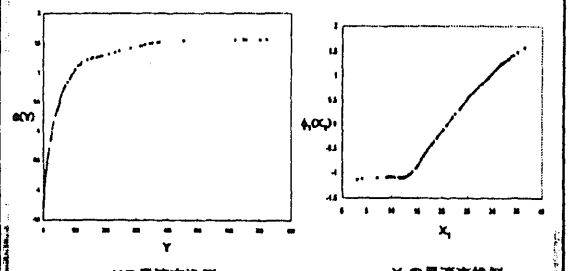
を最小化すべく繰り返し計算により最適化される

$$\phi_k^{(v+1)}(X_k)=E\left[\theta^{v}(Y)-\sum_{i\neq k}\phi_i^{v}(X_i)\middle|X_k\right]$$

$$\theta^{v+1}(Y)=E\left[\sum_{i=1}^{p}\phi_i^{v}(X_i)\middle|Y\right]\Big/\left\|E\left[\sum_{i=1}^{p}\phi_i^{v}(X_i)\middle|Y\right]\right\|$$

JAPAN OIL ENGINEERING Co. Ltd.

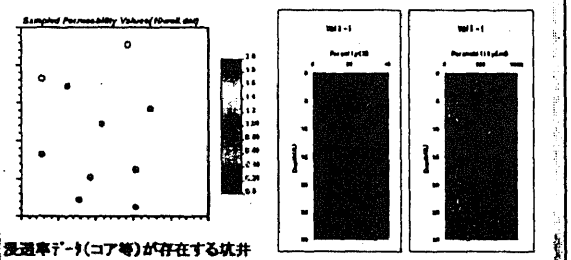
ACE法(3)



Yの最適変換例　　X_iの最適変換例

JAPAN OIL ENGINEERING Co. Ltd.

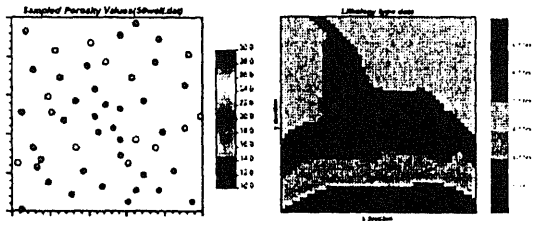
浸透率分布の推定に利用可能なﾃﾞｰﾀ(1)



浸透率ﾃﾞｰﾀ(コア等)が存在する坑井　　坑井-1におけるコア孔隙率・浸透率データ

JAPAN OIL ENGINEERING Co. Ltd.

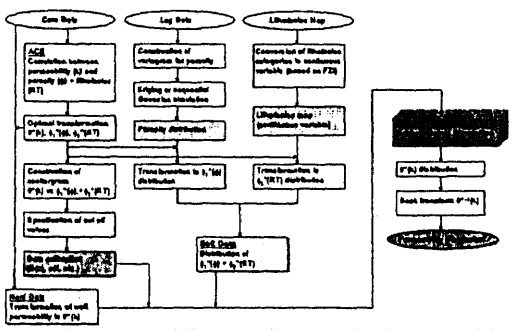
浸透率分布の推定に利用可能なﾃﾞｰﾀ(2)



孔隙率データ(コア・検層等)が存在する坑井　　岩相マップ

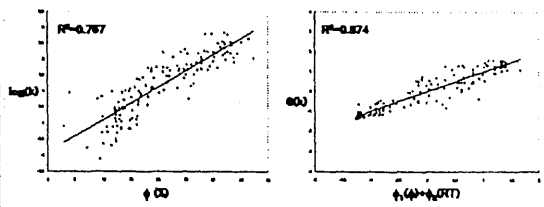
JAPAN OIL ENGINEERING Co. Ltd.

コア・検層・岩相分布ﾃﾞｰﾀ → 浸透率分布



JAPAN OIL ENGINEERING Co. Ltd.

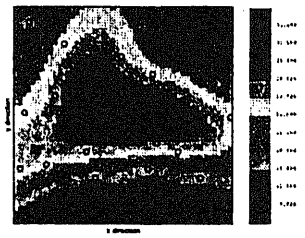
結果(1):浸透率・孔隙率・岩相の最適変換



浸透率 vs. 孔隙率のスキャタグラム　　θ(Y) vs. $\phi_1(\phi)+\phi_2(RT)$のスキャタグラム

JAPAN OIL ENGINEERING Co. Ltd.

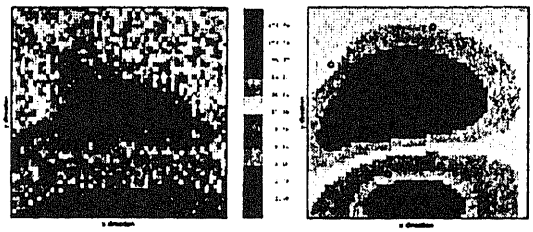
結果(2):孔隙率分布の推定(クリギング)



オーディナリ・クリギングによる孔隙率分布推定値

JAPAN OIL ENGINEERING Co. Ltd.

結果(3):浸透率分布の推定(クリギング)

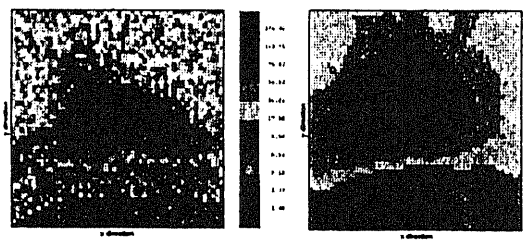


真の浸透率分布

オーディナリ・クリギングによる浸透率分布推定値

JAPAN OIL ENGINEERING Co. Ltd.

結果(4):浸透率分布の推定(ｺｸﾘｷﾝｸﾞ)

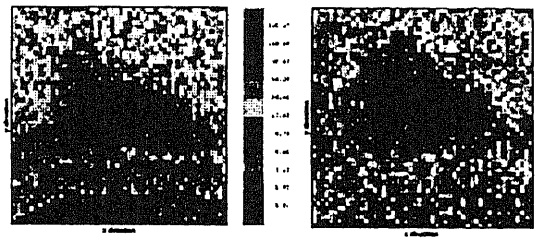


真の浸透率分布

孔隙率分布を副変数としたコクリギングによる浸透率分布推定値

JAPAN OIL ENGINEERING Co. Ltd.

結果(5):浸透率分布の推定(M-B)

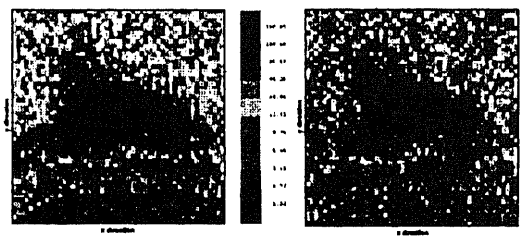


真の浸透率分布

$\phi(\phi)$分布を副変数としたマルコフ-ベイズ法による浸透率分布推定値

JAPAN OIL ENGINEERING Co. Ltd.

結果(6):浸透率分布の推定(M-B)



真の浸透率分布

$\phi_1(\phi)+\phi_2(RT)$分布を副変数としたマルコフ-ベイズ法による浸透率分布推定値

JAPAN OIL ENGINEERING Co. Ltd.

科研費集会講演資料　　2000. 12. 8

確率的シミュレーションと実現像の選択法について

石油公団・石油開発技術センター
不均質炭酸塩岩油層研究プロジェクト
江藤　公治 Ph. D.

石油業界で応用されている地質推計学、それも確率的シミュレーションと実現像の選択法についてテーマを絞ってその概要を紹介する。但し、紹介する手法の選択は私の独断と偏見による。

条件付きシミュレーション

領域内の各点の同時確率分布を求める一手法として逐次シミュレーションを実行する。この方法は、各格子点でその周囲の情報に基づいて条件付き累積確率分布（ccdf）を構築し、乱数を引き、ccdf に基づいてその格子点の値を確定し、それを又新たな拘束条件とする試行法である。全ての格子点に対してこの手続きは繰り返される。試行によって求めた値の集合は実現像と呼ばれる。実現像は単点分布、２点相関分布、及び、観測点での観測値という拘束条件を満足すべきである。乱数によって次の計算点が選択されるから、同じ制約条件の下で数多くの異なった実現像が生成される。その様な数多くの実現像のばらつきの幅が、考慮している事象に対する不確実性を示している。

正規分布系シミュレーションを採るか、非正規分布系シミュレーションを採るか

今その空間分布を追求している変数が、その平均値の周辺で左右対称の正規（ガウス）分布をしているかどうかでこの選択肢がでる。データが正規分布をしていると、その数学的特長を生かして計算手順が簡単になり、手間隙も計算時間も短くてすむ。一方当然ながら平均値から離れた点の出現確率は非常に小さい。標準偏差で３だけ離れると千に三つ実現すれば良いという事で、常識的にはこれは起こらないと解釈しても良い確率である。一方浸透率測定値のヒストグラムをみると、高浸透率側の離れた個所に小さな一群が出ていることがよくある。実際の油層でも高浸透率層の挟みがこの程度の存在確率を持つ事はよくある。しかし油層の流動挙動の点から見ると、この高浸透率層が油層の流動挙動に決定的な影響を及ぼす事が多い。従って、存在確率だけに依存してこの高浸透率層の挟みの実現を期待するのは、守株の様なもので現実的ではない．この様な場合は、確実に高浸透率層の挟みが実現する様な方法を選択する必要がある。非正規分布系シ

ミュレーション、即ち、インディケータ・クリーギングに基礎を置く手法をとる。浸透率分布の様な油層特性パターンの実現については、特性値は岩相（地質学的分類）や岩石型（岩石物性学的分類）といった離散値型のクラス分けを行なった後、次に個々のクラスの中で存在する連結性を表現する為に再度シミュレーションされる事が多い．

主変数だけのシミュレーションを実行するか、副変数と共にコー・シミュレーションを実行するか

主変数だけのシミュレーションは準備する手間が少ないから簡便である。拘束条件は通常坑井位置で与えられるから、その数は余り多くないのが普通である．一方拘束点から離れるに従ってクリーギング誤差分散は急速に増加するから、その点での推定値の確度はそれだけ低くなる．従って、拘束データ点の数が充分多くない限り何らかの副変数の助けを求め、主変数値のばらつきに制限を加える事が有効である。
副変数と共にコー・シミュレーションを実行する場合は、当然主変数と副変数との間の共分散が必要になる．共分散の計算は面倒であるし、副変数の数が増加するに従って計算量も飛躍的に増加する．正直に共分散を全て求めるコー・クリーギングの様な手法と、マルコフ・ベイズ法の様に、或る仮定を受容すれば共分散を全ては計算しなくても良い方法とがある。地震探鉱のデータは質は悪くとも油層全体に満遍なく且つ均質に存在し、副変数としては理想的である。岩相や岩質型の様な副変数となるべき分類変数の分布領域が一定の広がりを持つならば、領域毎に主変数だけのシミュレーションを実行するのも納得できる。

逐次シミュレーション法

領域全体に対する変数の分布を求めるには、ある特定の点uでの ccdf；

$$F(u;z|(n)) = \Pr ob\{Z(u)\le z|(n)\}$$

をN点に拡張した同時確率分布を求めなければならない．ここでzはそのクラスの閾値、(n) は与条件である。即ち、

$$F(u'_1,...u'_N;z_1,...,z_N|(n)) = \Pr ob\{Z(u'_1)\le z_1,...,Z(u'_N)\le z_N|(n)\} \qquad (1)$$

この為に先ず、2点間での同時確率分布を考える．

$$F(u'_1,u'_2;z_1,z_2|(n)) = \Pr ob\{Z(u'_1)\le z_1, Z(u'_2)\le z_2|(n)\}$$

ここで、ベイズの定理を使う．

$$F(u_1', u_2' ; z_1, z_2 | (n)) = F(u_2' ; z_2 | (n+1)) \cdot F(u_1' ; z_1 | (n))$$

この因数分解の意味するものは、先ず、u_1'点での値 $z(u_1')$を ccdf $F(u_1' ; z_1 | (n))$ を使って求め、次にこの値 $z(u_1')$を n 個のデータ点に加えたものを次の拘束条件$F(u_2' ; z_2 | (n+1))$として求める事である。即ち、2 点 ccdf を求める代わりに 1 点 ccdf を続けて 2 回行った事になる。この考え方を N 点迄拡張すると、N 点 ccdf を求める代わりに 1 点 ccdf を続けて N 回行えば良い事になる。

$$F(u_1', \ldots, u_N' ; z_1, \ldots, z_N | (n)) = F(u_N' ; z_N | (n+N-1)) \cdot F(u_{N-1}' ; z_{N-1} | (n+N-2)) \cdots$$

$$F(u_2' ; z_2 | (n+1)) \cdot F(u_1' ; z_1 | (n))$$

この実行の為に現在、次の 4 つの手法がよく使われている.

a．逐次ガウス・シミュレーション法

逐次ガウス・シミュレーション法では、ランダムに各点を訪れてクリーギング推定値(μ)とクリーギング誤差分散 (σ^2) から 1 点 ccdf 模型$N(\mu, \sigma^2)$ を構築した後、その 1 点 ccdf とサンプリング乱数に基づき変数値を推定する。この手法は最も多用されるので、その計算手順を以下に示す。

1　データの正規化を行い、データの正規性を検査する。データの正規性が悪い場合には、逐次ガウス・シミュレーション法は使えない．正規型からの多少の乖離はデータの正規化法（Normal Score Transform）で変数変換を行なう。

2　逐次シミュレーション法を実行する．即ち、

a. 各計算点x_0でクリーギング方程式を解き、クリーギング重み$\lambda_\alpha, \alpha = 1, \ldots, n$を得る.

b. クリーギング推定値とクリーギング分散値を次式から計算する.

$$Z^*(x_0) = \sum_{\alpha=1}^{n} \lambda_\alpha(x_0) \cdot Z_0(x_0)$$

$$\sigma_K^2(x_0) = C(0) - \sum_{\alpha=1}^{n} \lambda_\alpha(x_0) \cdot C(x_0 - x_\alpha)$$

ここで $Z_0(\bullet)$ は計算点 x_0 での変数値、$C(h)$は h をラグとする共分散である。

正規分布はこの 2 つのパラメーターで$N(Z^*(x_0), \sigma_K^2(x_0))$と完全に表現できる。

c. 一様乱数 $r \in [0,1]$ を引き正規分布のクオンタイル関数から実現値を求める.

d. 乱数を引いて次の計算点を選び移動する。

3　全ての点で計算が終了したら、実現値の逆変換を行う．

b．P－フィールド法

この方法では、N点ccdf（式1）をN点での1点ccdf模型で代用する。この時1点ccdfは常に測定点のみを拘束条件点とする。次に、分散模型を実現する為に、ccdfからのサンプリングの際に領域内に前以て求めておいたサンプリング確率値の自己相関パターンを、拘束条件として利用する。この方法は ccdf を各点で計算しないから全体の計算時間が速い．多くの実現像を生成するのに有利である。

c．ベイズ推論的逐次シミュレーション法

ベイズの定理、即ち、

事後分布 ∝ 尤度 × 事前分布

という定理を各格子点で使い事後確率を求め、実現値をサンプリングする方法である。
例として、(n+N) 個の格子模型を考え、各点 $X = (x_1,...,x_{n+N})$ で岩相（砂岩か頁岩）を定義する問題を考える。この内 n 個の $X: x_1,...,x_n$ は測定点であり、N 個の $X: x_{n+1},...,x_{n+N}$ は未知点である。指標変数は次の様に定義する。

$$x_i = \begin{cases} 0 \ , & pixel(i) \in shale \\ & (i = 1,...,n+N) \\ 1 \ , & pixel(i) \in sand \end{cases}$$

指標変数は次の性質を持つとする。

平均：　$E\{x_i\} = \Pr ob(x_i = 1) = \pi_{sand}$

分散：　$E\{x_i - \pi_{sand}] \cdot [x_j - \pi_{sand}]\} = C(h_{ij})$

各格子点には、震探データ $Z = (z_1,...,z_{n+N})$ の様な副変数が存在していて、岩相に対して間接的若しくは不確実な情報を持っているとする。z_iは連続量であり、例えば振幅や音響インピーダンス、その他岩相と何らかの相関のあるものである。従ってこの変量は各岩相に対して周辺確

率分布が定義されているとする。即ち、$f(z_i|x_i=1)$ や $f(z_i|x_i=0)$ は坑井データや震探データから計算されているとする。ベイズ的推論では未知数 $x_{n+1},...,x_{n+N}$ は坑井データや震探データ等の拘束条件が与えられた後の事後確率、即ち、

$$\Pr ob(x_{n+1},...,x_{n+N}|Z,x_1,...,x_n)$$

で推定できる。この手法の目的は、この結合確立分布から多くの実現値を求める事にある。現実にはこの多変数結合確率分布から直接サンプリングを行う事は困難であるから、ベイズの法則

$$\Pr ob(x_{n+1},...,x_{n+N}|Z,x_1,...,x_n) = \prod_{i=n+}^{n+N}\Pr ob(x_i|Z,x_1,...,x_{i-1})$$

を使い、単変数確率分布からサンプリングを連続的にN回行って代用する。

$\Pr ob(x_i|Z,x_1,...,x_{i-1})$ は、インディケーター・コークリーギングでも求められるが、連続量と離散量のコークリーギングには問題があるという指摘もある。ここでは、事後確率を２つの項の積に分解する。

$$\Pr ob(x_i|Z,x_1,...,x_i) \propto f(z_i|x_i)\cdot\Pr ob(x_i|Z,x_1,...,x_{i-1})$$

ここでは比例係数は x_i に対して無関係であるから、サンプリングに際しては不必要である。この関数を x_i に関して見た場合、第１項 $f(z_i|x_i)$ は尤度関数と呼ばれる。特に、$f(z_i|x_i=1)$ と $f(z_i|x_i=0)$ は、各々格子点ｉでの震探データが z_i である場合に砂岩と頁岩である為の尤度関数となる。第２項 $\Pr ob(x_i|Z,x_1,...,x_i)$ は x_i の局所事前確率と呼ばれ、岩相の空間分布の在り様を示すものでインディケーター・クリーギングで求める.

d．マルコフ・ベイズ法

コロケイティッド･コークリーギングとも言う。非正規分布系のシミュレーションも可能である。岩石の孔隙率分布を求めるという問題を考える。数個の実測された孔隙率と孔隙率に関連した他の変数、例えば弾性波の伝播時間、といったデータが手元に有る。従来の手法は、孔隙率と伝播時間の回帰曲線を求め孔隙率を推定することになる。この時、数個の実測された孔隙率は宙に浮く（採用されないから拘束条件を満たさない)。例え、測定点で実測された孔隙率を代入しても、その近傍での整合性は無い。更に進んだ手法は、コー・クリーギング法である。この手法は孔隙率と伝播時間の各々の分散値と共分散値が必要である。又伝播時間から孔隙率を推定する場合の伝播時間のクラスによって推定の信頼性に違いがある時、それが反映されない。更に又、コー・

クリーギング法では局所的補間法の手段を与えない。インディケータ・クリーギング法ならば、伝播時間のクラスによって推定の信頼性に違いがある時、それを反映させる事ができる。この方法なら、ハードデータやソフトデータの存在をその付近の拘束条件とできる。この場合の欠点は、ハードデータ（孔隙率）とソフト・インディケータ・データ（伝播時間）の自己相関を区別出来ない事、従ってソフト・インディケータ・データ（区間だけが記述できるデータ）だけが在る点の情報を更新できない。マルコフ・ベイズ法はインディケータ・クリーギング法の上の欠点を補い、尚且つ、インディケータ・クリーギング法を超える共分散計算の手間を必要としない。マルコフ・ベイズ法の特徴は、

1. 通常の回帰法は基より、コー・クリーギング法より優れた結果を示す。
2. 主変数の局所的平均値の良好な推定値を与える。
3. インディケータ・クリーギング法以上の共分散計算の手間を必要としない。

マルコフ・ベイズ法の主要な結論は、

$$C_{IY}(h;z) = Cov\{I(x+h;z),Y(x;z)\} = B(z)\cdot C_I(h;z) \quad \forall h$$
$$C_Y(h;z) = Cov\{Y(x+h;z),Y(x;z)\} = B^2(z)\cdot C_I(h;z) \quad \forall h>0$$

ここで、

$$C_I(h;z) = Cov\{I(x;z),I(x+h;z)\}$$
$$E\{Y(x;z)\} = F(z)\cdot m^{(1)}(z) + [1-F(z)]\cdot m^{(0)}(z)$$
$$F(z) = \Pr ob\{Z(x)\le z\} = E\{I(x;z)\}$$
$$B(z) = m^{(1)}(z) - m^{(0)}(z) \in [-1,1]$$
$$m^{(1)}(z) = E\{Y(x;z)\,|\,I(x;z)=1\} \in [0,1]$$
$$m^{(0)}(z) = E\{Y(x;z)\,|\,I(x;z)=0\} \in [0,1]$$

であるが、初めの２式は、主変数と副変数との共分散及び副変数の分散が主変数の分散から求められる事を示している。他の式は h に無関係な各閾値毎の定数値である。従って必要な作業量は、単変数のインディケーター・クリーギングと同程度である。

多段グリッド・シミュレーション

バリオグラム（定常状態で $\gamma(h)=C(0)-C(h)$）を構築する拘束点の探索の際、相関距離以遠のデータは計算結果に影響を及ぼさないので、探索半径は相関距離以内に設定する。しかし、バリ

オグラムが近距離構造や遠距離構造等多重構造を持っている場合、多数の近距離点の影響が強過ぎて、遠距離構造の様子が上手く浮き上がって来ない事が往々にしてある。この場合、模型の格子網を細格子と粗格子２段(或いはそれ以上の多段)に設定し、先ず、粗格子について実現像を得、これら粗格子点での値と元々のデータ点とを拘束値として細格子のシミュレーションを行う事が望ましい.

実現像の選択

推計的なシミュレーション手順において、次の推論地点は対象領域でランダムに選択され、局所的な ccdf を構築し、その場所において特性値の推定値を選択するために乱数を引く。即ち、一様分布[0,1]の母数から百分位数として使う為の乱数を引き、これに対応する ccdf 値を確定する。更に、新たに推定された値は、それが本当の（測定された）データセットの一部であるかのように見なされる。それは局所的な構造の枠組みで後続してくる値を拘束するために使われる事になる。始点即ち、乱数の種が変えられる時、推定される次の地点の周りの ccdf は異なり、従って実現値の最終分布図は変化する。しかしながら、実現像構築の過程は同一であるから、その「確実性」は同等である。即ち、他の追加の情報がなければ、何れの実現像がより尤もらしいかを言うことはできない。油層管理の立場から、一連の実現像の選択に於ける最も確かな戦略は、これら等しく尤もらしい地質模型のそれぞれについて流動シミュレーション模型を構築し、それぞれのヒストリーマッチングを実行し、正解を選択する事である。しかし、統計学的に健全な結論を推論するには、恐らく何百回ものシミュレーションを必要とするであろうが、明らかに労力、時間、予算各々の限界の中で実行不可能である。この状況に於ける最善の方法を探るに当たり若干の考察をしてみよう。

51×51×40 の三次元矩形格子上に５点パターンを想定した仮想油田での生産試験を模擬した例を述べる。真の浸透率分布は秘匿し、その統計的性状（バリオグラム）及び生産/圧入レートと観測圧力値及びトレーサー濃度値をデータとして与えておく。模型化の手順は、この細格子模型を 11×11×8 の粗格子にアップスケーリングした上で、四隅の圧入井から４種類のトレーサーを圧入し中心の生産井から生産する。100 個の浸透率分布の実現像の各々に対して流動シミュレーションを行い、その坑底圧力とトレーサー濃度の計算値と観測値のミスマッチの２乗和である目的関数を計算したところ図１の様になった。値域は 1.15～17344、平均は 781 である。ここで、完全マッチならば目的関数値は勿論 0 である。静的な拘束条件だけで正しい実現像を得る事の難しさを示している。

通常実現像の構築法に関して、下に示された選択肢が考慮される：

1）地質模型の（最初の）1つの実現像に基づいて流動シミュレーションを実行する。上の例で分かる様に、この場合に正解値に当たる可能性は非常に低い.

2）地質模型の平均（例えば、10 の像）の実現像に基づいて流動シミュレーションを実行する。特に（地質の）シミュレーションの早い段階で、拘束条件になるハードデータの数がほとんどなく、（ハードデータが存在する）既知の点の近くに選択される計算点の数がわずかである時、結果の実現像は恐らく特定の選択された点の位置や計算の結果に偏っているであろう。この場合いくつかの実現像の平均をとることは非常に重要である。

3）多数の地質学的実現像から、先ず簡単なしかし計算の速い伝達関数を使ってスクリーニングを行い（cdf1)、ある特定の百分位数、例えば、P5、P25、P50、P75、と P95 に対応する実現像を選択し、これらの実現像を流動シミュレーション用の模型構築に使う。流動シミュレーションを実行した後、興味の対象変数の挙動（油の累積回収、油産出レートなど）の特性値に関した cdf が構築できる（cdf2)。この曲線上で、その平均値（P50）やランダムに選択された確率に従って最もありそうな特性値が選択される。
油層の流動解析には通常三相圧縮性流体の三次元模型が使用されるが、計算速度重視の伝達関数としては、次のものが考えられる.

1．トレーサー模型（単相非圧縮性流体)。
2．粗格子を使った本格模型。
3．流線模型（単相非圧縮性流体 $+\alpha$)。
4．パーコレーション理論を使ったウォーター・ブレイクスルーの時期の迅速計算。
この手法の問題点は、cdf1 と cdf2 内の順序関係が当該パラメーターについて維持できるかという事である（一般的には “できない “)。

4）地質模型の実現像の全ての場合について流動シミュレーションを実行する。流動シミュレーションを全て実行した後、興味の対象変数の挙動（油の累積回収、油産出レートなど）の特性値に関した cdf を構築し、この曲線上で、その平均値（P50）やランダムに選択された確率に従って最もありそうな特性値が選択される。圧入／生産のヒストリー・マッチングを行うのなら、実挙動に最も近い挙動を示す地質模型及び流動模型を選択する。しかし、この方法は、対象油層が余程小さくない限り実行不可能に近い。

5) 逐次変形法
この手法は、1．出発値の模型を、乱数空間の構造（平均値、共分散、ヒストグラム等）を変えずに徐々に変化させる。2．地質模型に動的な拘束条件を効率的に課する事を目的としている。以下の手順に見る様に、常にそれ迄の最適解を保持しているから、収束状況は単調非増加である。

多くの最適化問題がそうである様に、最初の数ステップでの収束速度は大きい。

手順

1- 正規空間で第1乱数$z_1^{(0)}$を発生させる。

2- 正規空間で乱数$z_2^{(0)}$を発生させる。これは第1乱数$z_1^{(0)}$に対する摂動乱数である。次式に従い$z_1^{(0)}$と$z_2^{(0)}$の線形結合を作る。

$$z(\rho) = z_1^{(0)} \cos(\pi\rho) + z_2^{(0)} \sin(\pi\rho)$$

（この段階では助変数ρは未確定であるから、数個仮定する必要がある）

3- 目的関数J(z)を最小化する助変数ρを決定する。第一ステージの終了。

4-目的関数値が収束閾値に達していなければ、次ステージの実行の為にステップ2に戻る。

乱数値の更新過程は図2に示してある。数値実験によると、摂動乱数は数個同時に発生した方が収束速度は速い。

手順2の式を一般化すると、

$$z(\rho) = \prod_{i=1}^{n-1} \cos(\pi\rho_i) z_i + \sum_{i=1}^{n-1} \sin(\pi\rho_i) \prod_{j=i+1}^{n-1} \cos(\pi\rho_j) z_{i+1}$$

即ち、合成した実現像Zはn-1個の独立した助変数$\{\rho_1, \ldots, \rho_{n-1}\}$の関数となる。

頭書の例題の場合では、一つづつ実現像を比較更新した場合（1助変数）は目的関数値が1.1位で殆ど減少しなくなった。この例（問題の性質、計算速度コントロール・パラメーターの組み合せ等）では、一度に二つづつ実現像を取込み比較更新した場合（2助変数）が最も小さい値に収束していった（図3）。

図4と図5は51×51×40細格子の浸透率値の対数プロットであり、正解地質模型（図4）と最適化された地質模型（図5）の対角断面図である。両端に圧入井が、中央に生産井がある。模型の周辺部と違って、この断面は最も流体が通過するので、両者の一致度は最も良い筈である．想起すべきは、流動模型は11×11×8にアップスケールされた粗格子上で行なわれているから、この細格子模型中の個々のグリッド・ブロックをマッチさせた訳ではない事である。しかし両者は浸透率分布の特徴を良く捕まえていて、その一致度は非常によい。研究すべき点は未だ多々あるが、アップスケーリングの方向性としては的を射ている様に思える。

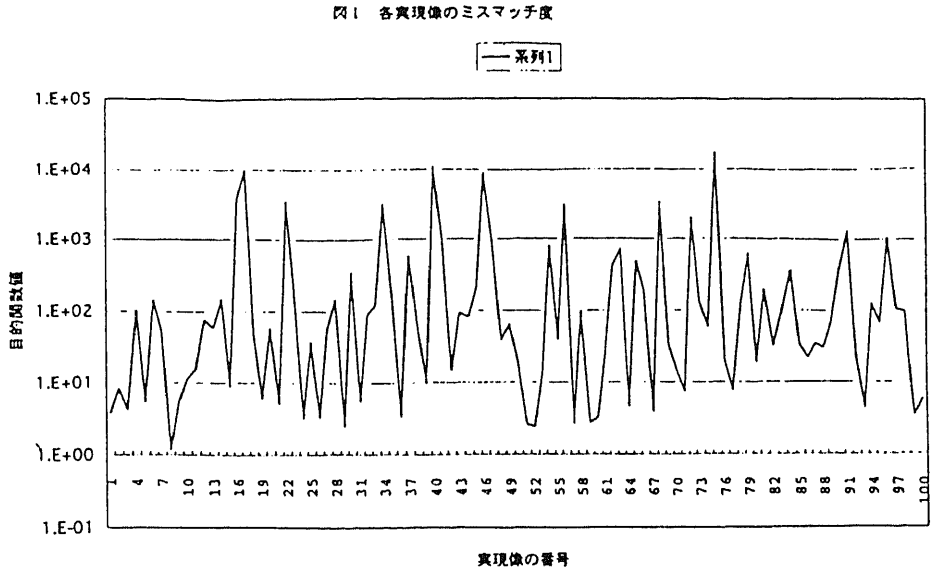


図1 各実現像のミスマッチ度
系列1
1.E+05
1.E+04
1.E+03
1.E+02
1.E+01
1.E+00
1.E-01
目的関数値
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97 100
実現像の番号

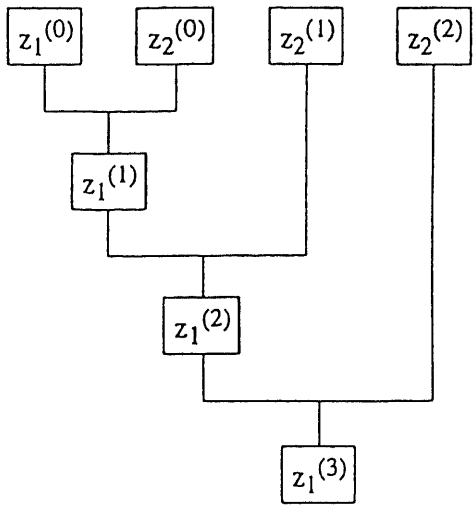

$z_1^{(0)}$
$z_2^{(0)}$
$z_2^{(1)}$
$z_2^{(2)}$
$z_1^{(1)}$
$z_1^{(2)}$
$z_1^{(3)}$

図2　乱数値の更新過程

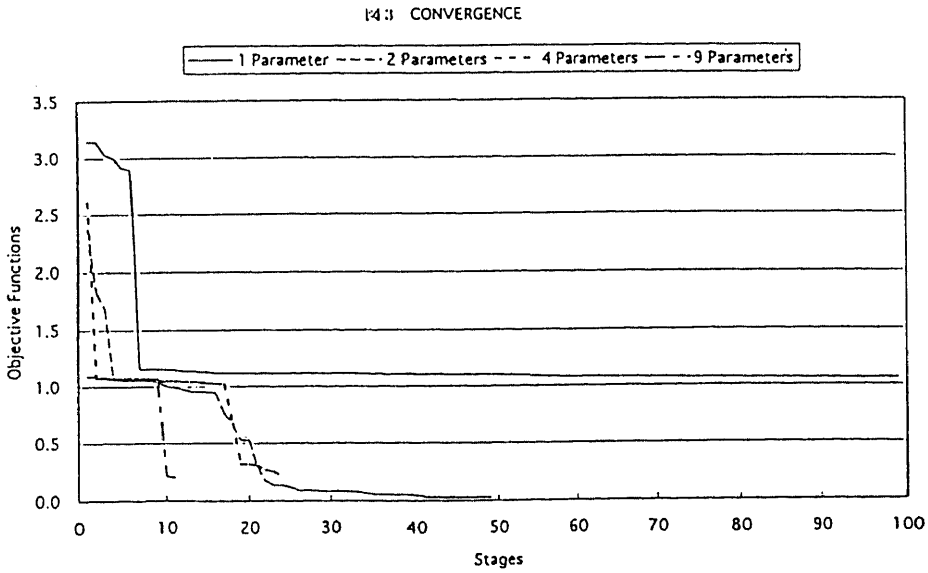


図3 CONVERGENCE
1 Parameter
2 Parameters
4 Parameters
9 Parameters
Objective Functions
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0 10 20 30 40 50 60 70 80 90 100
Stages

講演者の横顔

江藤　公治

学歴

東京大学工学部でBS及びMS、米国テキサス大学（オースチン）でPh.D.を授与さる。専攻は全て石油工学。又、ワイオミング大学にも在籍。1969年度フルブライト奨学生。

職歴

1972年から1973年迄テキサス大学地球科学及び工学研究所（研究員）。1973年からダラスのコア・ラボラトリー社のエンジニアリング＆コンサルティング部（油層工学スペシャリスト）。続いてヒューストンのガルフ・オイル社、ヒューストン・テクノロジー・センター、EOR部（上級エンジニア/コンサルタント）に所属。1989年から1992年迄サウジアラビア王立キングファハド石油鉱物大学付属研究所（教授/上級エンジニア）勤務．功労賞を授与さる。1992年から今日に至る迄、石油公団・石油開発技術センター（招聘研究員、プロジェクト・エンジニア、リサーチ・コンサルタント）。1999-2001 年度 SPE, Technical Committee on Fluid Mechanics and Oil Recovery Process 論文選考委員。

経験

油層工学全般、油層キャラクタリゼーション、各種油層シミュレーション法（ブラック・オイル、コンポジショナル・オイル、ミッシブル攻法、ケミカル攻法、ガス/不活性ガス圧入法）、過渡的圧力解析、岩石物性（ウェッタビリティ、毛細管圧力、相対浸透率等）の研究、開発、応用、教育に従事。地質推計学は油層キャラクタリゼーションの一分野としてこの10年間に亘り関心を持っている。

断層周辺の岩石データの解析

大阪女子大学・理　綿森　葉子

1　はじめに

標本空間が球面（$\mathbb{S}^2$ であらわす。）の場合、通常の統計的手法を適用したのでは問題がおこる場合がある。あるいは、これまでに用いてきた統計手法をそのままでは適用できない場合もある。ここでは、2年前に府大の4年生であった隅田剣生さんが卒論のために大変な労力を費やして採取してきた岩石のデータについて考察する。記憶が不確かなのと、当時は色々と忙しくて本人に詳しい話があまり聞けなかったために、背景についてほとんど説明できないのが残念である。

この岩石データは、ある断層（長野県の辺り？）の周辺の6つの地点から岩石を採取しその磁力の方向を測定したものである。岩石採取の過程やその後のデータ処理については、詳しくはわからない。データは6つに分類されており、それぞれ6個前後の方向（偏角と俯角）から構成されている。このデータ数でなにか統計的な結論を導くことにはかなり無理があるのだが、当てはめられそうな手法をいくつか適用してみることにする。データが方向であるので、方向統計の手法を用いるのはごく自然であるはずだが、どうも現場ではあまり利用されていないような印象をうける。というのは、方向統計理論の整備がまだ不充分でその手法を利用することは色々な意味で難しいのではないかと思われる。ここでは、理論的な背景を簡単に述べておく。

2　座標

$\mathbb{R}^3$ には、自然に座標が入っていて、それ以外に特に統計上考慮の対象となるような座標系はない。しかし、$\mathbb{S}^2$ にもっとも自然と思われる座標は角度であろうが、これはいろいろな点でやっかいである。角度を極座標変換により $\mathbb{R}^3$ の元とみなせば線形性は保てるが、ノルムの制約が入り成分間の独立性が最初から崩れてしまう。また、次元が本来の次元より1つ上がってしまうことも問題である。ここで極座標表示は

$$x = \sin\theta\cos\phi, \quad y = \sin\theta\cos\phi, \quad z = \cos\theta$$

$$0 \leqq \theta \leqq \pi, \ \ 0 \leqq \phi < 2\pi$$

であり、ヤコビアンは $J_2 = 1$ である。ノルムの制約から導かれるモーメント制約は、Σ を分散共分散行列、μ を平均ベクトルとするとき

$$\mathrm{trace}\Sigma + ||\mu||^2 = 1$$

となる。ここでは、角度を対象とするときはθを用い、$\mathbb{R}^3$の元とみるときは$\mathbf{x}$を用いることにする。

3 分布の特性量

いったん確率変数と確率測度が定義されれば、母集団平均、分散を含むモーメントはΩが球面如何に関わらず定義される。それに対応する座標系の下で得られたデータに対する標本平均や分散も同様である。しかし、Ωが$\mathbb{R}^3$のときと、Ωが$\mathbb{S}^2$のときでは、その意味、解釈が異なってくる。例えば、$\mathbb{R}^3$での平均は位置の中心であり重心であった。$\mathbb{S}^2$の極座標をもとに得た平均は、一般に$\mathbb{S}^2$の外にありもはや重心の意味を持ち得ない。つまり量的な中心と位置的な中心は別の概念として捉えなければならない。ここに至って「中心」のさすところを再度考え直さなければならなくなる。

4 相関と回帰

極座標を用いて$\mathbb{R}^3$の特異な分布として捉えた場合、相関係数はそのまま定義できるが、回帰関数はそうはいかない。特異性の$\mathbf{x}$のノルム制約が大きく影響する。では、Θをもとにしてはどうか？今度は新しい概念を導入する必要に迫られる。$\mathbb{S}^2$上では例えば

$$\rho_{FL} = \frac{\det E[\mathbf{x}\mathbf{y}']}{\sqrt{\det E[\mathbf{x}\mathbf{x}']\det E[\mathbf{y}\mathbf{y}']}}$$

が提案されている。Fisher & Lee (1986) この他にもいくつか提案されているが、それぞれ長所短所があって決めてになるものは今のところない。

回帰に関しては、$\mathbb{S}^2$では適当な回転行列Aを用いて

$$\hat{\mathbf{y}} = A\mathbf{x}$$

と表せるかどうかというが、最も自然であるように思われる。そして、これを利用したモデルが研究されている。

5 参考文献

N.I.Fisher, T.Lewis and B.J.J.Embleton,
Statistical analysis of spherical data, 1987 Cambridge Universi ty.
K.V.Mardia and P.E.Jupp, Directional Statistics, 2000 John Wiley & Sons.

A parametric model for forecasting time varying spectral density function of the sea surface motion in the wave developing process

Tsukasa Hokimoto(Graduate School of Fisheries Sciences, Hokkaido University)

1. Introduction

In physics, natural science, economics and the other various fields, the study on the statistical inference of the spectral density function of the nonstationary periodic motion has been progressed until now. In this paper, we focus on the natural phenomena, among them, on the sea surface motion when the sea condition is becoming stormy, and then propose a statistical method for forecasting time-varying spectral density function. The phenomena on the wind wave (i.e. the sea surface motion developing by the kinetic energy from wind) have become important subjects for study from various standpoints, such as physics, navigation safety etc, because their studies might contribute to the inference on flude dynamics of the sea surface movement. Our goal in this paper is to explain the complicated relationship among the movements of sea surface, wind direction and wind speed from the statistical standpoint, and then to develop a statistical model for forecasting the change of spectral density function in the case of developing wind wave.

In this paper, we analyse the time series data on the changes of sea surface and wind, which have been observed in Funka-Bay, Hokkaido, Japan. These series have been measured by using a microwave waveheight meter in our research ship. The purpose of the above measurement is to research the limit of the sea condition for the fishing vessels to keep on working safely. Figure 1 shows the measured time series data. Measured items are relative sea surface level(m), wind direction(deg.) and wind speed(m/sec). They have been obtained by a measurement for 90 minutes, at sampling time 0.2 seconds and the sample size is 27000. Here, the origin of the sea surface movement is the mean level of the sea surface movement in the past 10 minutes. Also, the origin of the wind direction means north and the positive values means the deviation to east from north. We can see from Figure 1 that the amplitude of the sea surface movement develops gradually over time. On the other hand, it may look that wind speed becomes higher and then slightly lower, and wind direction changes slowly in the range roughly from -100 deg. to 0 deg. (i.e. from north-west to north).

The study on the estimation and forecasting problem of spectral density function of nonstationary time series has been progressed from both of nonparametric and parametric approaches. For our problem, it may be neccesssary for effective forecasting to take into account of the wind's effects as well as the sea surface movement. To include the information of exogenous time series into the change of spectra, it might be better to construct a parametric model to explain the statistical effect to the change. From the above reason, in this paper, we approach this problem from the parametric standpoint.

The outlines of this paper is as follows. In the next section, we analyze the statistical structure on the sea surface movement in the wave developing process. In Section 3, we propose a nonstationary statistical model for forecasting the time-varying spectral density function in the above situation. To exaimine the availability of the presented method, we evaluate the forecasting performance by numerical experiments. The results and their analyses are shown in Section 4.

2. Statistical structure on the sea surface movement in developing wind wave and the estimation of the spectral density function

In this section, we show the motivation of our method for forecasting spectral density function in the wave depeloping process. In the followings, let t be a discrete parameter on sampling time point, $\{Z_t\}$ be the stochastic process which the sea surface movement follows and $Z_1, \ldots, Z_T$ be T samples from $\{Z_t\}$. When we estimate the spectral density function of the sea surface motion practically, the nonparametric methods such as Periodgram and Blackman-Tukey's method are frequently used. For example, the estimator of spectral density function based on Blackman-Tukey's method is given by

$$\hat{P}(\lambda) = \sum_{k=-T+1}^{T-1} w(k)\hat{C}(k)e^{-i2\pi k\lambda}, \tag{1}$$

and

$$\hat{C}(k) = \frac{1}{T}\sum_{t=k+1}^{T}(Z_t - \bar{Z})(Z_{t-k} - \bar{Z}), \qquad \bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t,$$

where λ is frequency and $w(k)$ is a window function for smoothing the raw spectrum. It is well-known that if $\{Z_t\}$ has stationarity, then the estimator of the above type has the consistency as T becomes sufficiently large. It means that the estimator has an advantage in the sense that it can estimate spectral density function without assuming any models to time series data. However, when it is necessary to develop the method for forecasting dynamic changes of the spectral density function, it might be more effective to consider the method which is based on a parametric model, because the forecasting performance may be expected to be improved by taking account of the information of physical factors such as wind direction and wind speed. From this reason, in the followings, we focus on a parametric model to forecast the spectral changes.

Here, let us look at the statistical structure of the sea surface movement to develop a parametric model. First, we investigate the short-term movement. Figure 2 displays an example of the time series on the sea surface motion for 200 seconds (sample size is 1000). Here, the vertical axis means the relative sea surface level(m) and its origin means the mean level in the past 10 minutes. It may look that the average level and the amplitude of the sea surface motion does not change over time. Now we regard the above series to be stationary and

then obtain the autocorrelation function and partial autocorrelation function, which are well-known as the preliminary analysis for model identification proposed by Box and Jenkins(1970). Their results are shown in Figures 3(a) and 3(b), respectively. The former result shows that it decays slowly as time lag increses and the latter dumps rapidly. When it is necessary to assume a statistical model to the sea surface data in Figure 2, we may think of various time series models. According to the identification procedure proposed by Box and Jenkins(1970), the above features on autocorrelation function and partial autocorrelation function may suggest the possibility that this time series follows an autoregressive model

$$Z_t = m_t^Z + \sum_{j=1}^{q} a_j Z_{t-j} + \delta_t \tag{2}$$

where m_t^Z is the mean of $\{Z_t\}$, q is the order of the model, $\{a_j; j = 1, \ldots, q\}$ are unknown parameters and $\{\delta_t\}$ is a random variable which follows a white noise process, say WN(0,σ^2). If $\{Z_t\}$ follows (2), the theoretical spectral density function, $f(\lambda)$ is given by

$$f(\lambda) = \frac{\sigma^2}{|1 + a_1 e^{-i2\pi\lambda} + \cdots + a_q e^{-i2q\pi\lambda}|^2} \tag{3}$$

Hence, if the assumption that (2) is reasonable as a statistical model of the sea surface movement is correct, then the estimator of (3) is reasonable as a spectral estimator. But, we do not know the true model of the movement, and therefore, it is necessary to examine whether this estimator is really reasonable as a spectral estimator for the sea surface movement. Figure 4 shows the simultaneous plot of the spectral density functions estimated by using Blackman-Tukey's method (1) and (3). The solid line means the spectral estimates obtained by (1) and the dotted line means the one by (3). It looks that the spectral estimates obtained by (3) can approximate the estimates by nonparametric estimator (1) fairly well, which suggest that we can regard (3) as a basis of our model.

Now we look at the long-term movement of the sea surface movement in the wave developing process. The important point is that it is necessary to examine whether the sea surface movement in the wave developing process keeps stationary statistical structure; (2) is reasonable if the movement has stationarity, but it is not clear whether or not the sea surface movement in the wave developing process always keeps the stationarity. Figure 5 shows the estimation results of the spectral density functions obtained by (1) and (3), using time series data in the past 300 seconds (sample size is 1500). From the comparison between Figures 4 and 5, we see that the characteristics of the spectral estimates differs depending on the time interval of the time series used for the estimation. It suggests that the statistical structure of the sea surface motion in the wave developing process has nonstationarity. But clearly, the degree of the nonstationarity of the sea surface movement is not so drastic; it might be natural to suppose that the structural change of the above spectrum to be slow. From the above reasons, when we need to forecast the spectral density function in the wave developing process, it is reasonable to approach from the statistical structure that the spectrum of (3), which has been estimated regarding the sea

surface movement to be locally stationary, changes gradually over time. Also, it means that we can assume to the sea sea surface movement in this situation the statistical structure such that each parameter and noise variance of (2) changes gradually over time.

3. A model for spectral forecasting based on the locally stationary autoregressive model

As stated before, we can regard the stochastic process of the sea surface movement as a nonstationary process which changes slowly its statistical structure. From the results in the previous section, instead of (2), we assume the following time-varying coefficients autoregressive model,

$$X_t = m_t + \sum_{j=1}^{p} \beta_{j,t} X_{t-j} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma_t^2) \tag{4}$$

to the sea surface movement in the wave developing process, where m_t is the mean of $\{X_t\}$, $\{\beta_{j,t}; j = 1, \ldots, p\}$ are unknown autoregressive coefficients which might change with t. Here, we focus on the structural changes for every certain unknown time interval, say M, because the speed of changing statistical structure of the sea surface movement is slow. Now let n be a new time parameter which takes positive integer and then define each time point $n = k(k = 1, \ldots, N)$ corresponds to $t = kM$. In the followings, we focus on the change with respect to n. Here, it is necessary to estimate the value of M using the samples from $\{X_t\}$, because it is unknown. Also, the model order p is assumed to be unknown and to be constant with n. The method for choosing M and p will be shown later. Now from (4), the theoretical spectral density at the time parameter n, say $f(\lambda, n)$, is given by

$$f(\lambda, n) = \frac{\sigma_n^2}{|1 + \beta_{1,n} e^{-i2\pi\lambda} + \cdots + \beta_{p,n} e^{-i2p\pi\lambda}|^2} \tag{5}$$

Note here that we use M samples at the sample time point $t \in [(n-1)M + 1, nM]$ for the estimation of the parameters $(\{\beta_{1,n}, \ldots, \beta_{p,n}\}, \sigma_n^2)$. For long-term forecasting of $f(\lambda, n)$ with respect to the time parameter n, it is necessary to obtain forecasted values on the parameters $\{\beta_{j,n}; j = 1, \ldots, p\}$ and the innovation variance σ_n^2.

Here, let us investigate the statistical features on the behaviors of the estimates of parameters. Figure 6 shows an example on behaviors of the estimates of $\beta_{1,n}$ and σ_n^2 with respect to n, when we fit an autoregressive model (2) under the order $q = 2$. To obtain the above series, we fixed the total sample size as 500 and, by updating every 200 samples(40 seconds), an autoregressive model under the constant order was fitted sequentially to the newly updated data. From the above results, we might find out the followings. First, the change of the estimate of σ_n^2 has the tendency to increase with n, and in this sense it has nonstationary statistical structure. From (5), it may be possible to regard σ_n^2 as the parameter which give impacts on "magnitude" of the spectrum. From physical standpoint, this increase of the magnitude of spectrum may cause by the supply of the wind's energy. Therefore, for forecasting the future values of σ_n^2, it

may be reasonable to take the changes on wind's direction and speed into consideration. On the other hand, it looks that each behavior of $\{\beta_{j,n}\}$ also exhibits nonstationarity, because its trend changes clearly with n. The parameter $\{\beta_{j,n}\}$ affects the dominant frequency (i.e. the frequency maximizing the spectral density function). Physically, this change is also caused by the wave development. Hence, for forecasting the changes of $\{\beta_{j,n}\}$, it will be effective to take into account of the changes of wind direction and speed. In addition, from the above relationship on wind effect, σ_n^2 and $\{\beta_{j,n}\}$ are not independent with each other. Hence, we also take into account of the change on wind's direction and speed, say $WD(n)$ and $WS(n)$, as well as time histories of σ_n^2 and $\{\beta_{j,n}\}$, to construct a model for forecasting σ_n^2 and $\{\beta_{j,n}\}$. Figure 7 shows an example on the behaviors of the differenced series $\nabla\sigma_n^2$, $\{\nabla\beta_{j,n}\}$, $\nabla WD(n)$ and $\nabla WS(n)$, where $\nabla\sigma_n^2 = \sigma_n^2 - \sigma_{n-1}^2$ and so forth. It looks from their behaviors that their means and variances are constant to some extent and therefore we might regard the above series as stationary processes. Hence, it might be natural to suppose that the simultaneous changes on the parameters $(\{\beta_{j,n}\}, \sigma_n^2)$ and parameters on motion of wind ($WD(n)$, $WS(n)$) follow a multivariate autoregressive model,

$$\boldsymbol{\theta}_n = A_1\boldsymbol{\theta}_{n-1} + A_2\boldsymbol{\theta}_{n-2} + \ldots + A_m\boldsymbol{\theta}_{n-m} + \boldsymbol{\delta}_n, \tag{6}$$

where

$$\boldsymbol{\theta}_n = (\nabla WD_n, \nabla WS_n, \{\nabla\beta_{j,n}; j = 1, \ldots, p\}, \nabla\sigma_n^2)^t \tag{7}$$

m is the model order, $\{A_i; i = 1, \ldots, m\}$ are unknown coefficient matrices, and $\boldsymbol{\delta}_n$ is a white noise vector satisfying $E(\boldsymbol{\delta}_n) = 0$, $E(\boldsymbol{\delta}_n\boldsymbol{\delta}_n^{'}) = (\sigma_{ij})$ and $E(\boldsymbol{\delta}_n\boldsymbol{\delta}_{n'}^{'}) = 0$ $(n \neq n')$. The method to choose m will be described later. For identifying this model, the above matrices are estimated using the least squares method(for example, see Kitagawa(1993)) under m is chosen. To forecast the future values $\boldsymbol{\theta}_{n+l}$ $(l = 1, \ldots, L)$, we use the linear predictor $\tilde{\boldsymbol{\theta}}_{n+l}$, which is defined by

$$\tilde{\boldsymbol{\theta}}_{n+l} = \tilde{A}_1\boldsymbol{z}_{n+l-1} + \tilde{A}_2\boldsymbol{z}_{n+l-2} + \ldots + \tilde{A}_m\boldsymbol{z}_{n+l-m} \tag{8}$$

where $\boldsymbol{z}_{n+l-m} = \boldsymbol{\theta}_{n+l-k}$ $(l \leq k)$ and $\boldsymbol{z}_{n+l-m} = \tilde{\boldsymbol{\theta}}_{n+l-k}$ $(l > k)$. Also, $\{\tilde{A}_i\}$ are the estimates of the coefficient matrices. From forecasting results on the L steps ahead on $\boldsymbol{\theta}_n$ using (8), the forecasted values on $\tilde{\sigma}_{n+l}^2$ and $\{\tilde{\beta}_{j,n+l}\}$ are obtained. Thus, the forecasts on spectral density function at l steps ahead can be obtained by

$$\tilde{f}(\lambda, n+l) = \frac{\tilde{\sigma}_{n+l}^2}{|1 + \tilde{\beta}_{1,n+l}e^{-i2\pi\lambda} + \cdots + \tilde{\beta}_{p,n+l}e^{-i2p\pi\lambda}|^2} \tag{9}$$

Finally, we show the method to choose the time interval M and optimal orders of the models (4) and (6). The choices of the orders p, m and the time interval M affect to the forecasting accuacy of $\tilde{\boldsymbol{\theta}}_{n+l}$, because the number of parameters to be estimated and the samples to use for estimation differs depending on their values and then it may give bad effects on the estimation

accuracy of each parameter in coefficients matrices $\{A_i; i = 1, \ldots m\}$. It might be natural to select these values so that the forecasted spectrum gives the best agreement with the locally stationary spectrum estimated using real data, viz. we choose their values such that the sum of squared forecasting errors over all frequencies, which has been obtained at every time point in the period $[1, n-l]$,

$$S(p(l), m(l), M(l)) = \sum_{j=1}^{n-l} \int_{\lambda} (\hat{f}(\lambda, j; M(l)) - \tilde{f}(\lambda, j; p(l), m(l), M(l)))^2 d\lambda$$

is minimized for every forecast step l.

4. Statistical evaluation on the forecasting accuracy

To show the validity of the model proposed in the previous section, it is necessary to evaluate the forecasting accuracy to the actual changes of the spectral density function. But it is difficult to prove the validity from the theoretical standpoint. So in this section, we investigate the statistical features on forecasting by applying the proposed model to observed sea surface data. In the followings, first, we show an example on long-term forecasting of the change of spectral density function in the wave developing process. And next, we show the numerical results on evaluation of the forecasting performance.

4.1 Example of the application to the observed time series in the wave developing process

Let us look at an example on the change of spectral density function, which has been forecasted by using the proposed model. First, we show an example of a bird's-eye view in Figure 8. Here, top of the figure shows the change of the spectral estimates obtained by using actual observations, and the bottom shows the forecasting result using presented model. Note here that these results were obtained under $n = 119$ (about 79 minutes) and $L = 10$ (400 seconds). It looks that both results have the similar tendency; the maximum value of the spectral density increases gradually with the forecast step. It might be explained physically that the energy by the wave motion increases rapidly in the wave developing process.

Next, the simultaneous plots of several spectra of Figure 8 is shown in Figure 9. Here, the top shows a simultaneous plot of the estimates obtained at n+1 (40 seconds later), n+5 (200 seconds later) and n+9 (360 seconds later), and the bottom is a simultaneous plot of the corresponding forecasts. It looks that both of these results have the following tendencies;

(1) the maximum value of the spectrum becomes larger over time. From physical standpoint, this phenomenon will be explained that the wave energy increases in the wave developing process by the supply of wind energy.

(2) the dominant frequency becomes lower. Physically, it will be explained that the wave period gets longer as wave develops, because it needs long period for the periodic motion with large amplitudes to oscillate.

4.2 Evaluation on forecasting accuracies from numerical experiments

Now we examine forecasting performances of the proposed model by means of the numerical experiments. When we forecast the spectral changes in the wave developing process, the forecasting features may change depending on the time point to start forecasting, because of the nonstationarity of the sea surface movement in the above situation. It is necessary, therefore, to examine whether the presented model can give good forecasting performances at any time point. We examine this point by numerical experiments. In our experiment, we examine the effectiveness of forecasting using our model by numerical comparisons of forecasting performances among the other several methods which are expected to give reasonable forecasting.

Our method of numerical experiment is the following way. First, by changing the time point to start forecasting n, we forecast the future change on spectral density function up to 10 steps ahead, and then evaluate the forecasting performance. Here, we also used three predictors, $\tilde{f}_A(\lambda, n+l)$, $\tilde{f}_B(\lambda, n+l)$ and $\tilde{f}_C(\lambda, n+l)$, to compare the forecasting performance. The details of these predictors are the followings;

(A) $\tilde{f}_A(\lambda, n+l) = \hat{f}_L(\lambda, n)$, where $\hat{f}_L(\lambda, n)$ is the estimates at the time point n using data in the time interval which is regarded to be locally stationary (See Appendix). If the statistical structure of the sea surface movement in the forecasting period does not change over time, this predictor's performance will give the best.

(B) The proposed predictor which does not include the information on wind direction and wind speed. Our intention of introducing this predictor is to examine whether or not the information of wind is really effective for the improvement of forecasting performances. This predictor is constructed by replacing (7) with $\boldsymbol{\theta}_n = (\{\nabla\beta_{j,n}; j = 1, \ldots, p\}, \nabla\sigma_n^2)^t$.

(C) It may also be possible that the changes of coefficients $\beta_{j,n}(j = 1, \ldots, p)$ with respect to n do not necessarily follow a multivariate autoregressive model. To examine that our construction of the multivariate autoregressive model is reasonable, we also forecast spectra from the values which have been obtained by forecasting $\{\beta_{j,n+l}\}$ for every j, using an univariate statistical model. Here, as shown before, their changes exhibit nonstationarity, so we construct two predictors using the following models.

(C01) ARIMA(P,1,0)(Autoregressive Integrated Moving Average) model (for example, Box and

Jenkins(1970))

$$\gamma_{j,n} = \sum_{r=1}^{P} z_r \gamma_{j,n-r} + \zeta_{j,n}, \qquad \gamma_{j,n} = \beta_{j,n} - \beta_{j,n-l}$$

(C02) Time varying coefficients autoregressive model (for example, Kitagawa and Gersch(1985))

$$\begin{aligned} \beta_{j,n} &= \sum_{r=1}^{R} \kappa_{r,n}\beta_{j,n-r} + \xi_{j,n}, \qquad \xi_{j,n} \sim WN(0, \sigma_\xi^2) \\ & \kappa_{r,n} - \kappa_{r,n-1} = v_{r,n}, \qquad r = 1, \ldots, R \\ & (v_{1,n}, \ldots, v_{R,n})^t \sim N(0, diag(\tau^2, \ldots, \tau^2)) \end{aligned}$$

where z_r is unknown coefficient, $\zeta_{j,n}$ and $\xi_{j,n}$ are random variables which follow white noise processes. For compuation of (C02), it is necessary to estimate the unknown variances of white noise variance σ_ξ^2 and system noise τ^2. In this experiment, we used the methodology shown in Kitagawa(1993).

Now we define the evaluation criterions of the forecasting performances. It is not easy to evaluate "the goodness on forecasting spectral density function" using single statistics, because there are several aspects to evaluate. One of basic evaluation criterions may be the sum of squared errors between estimates and forecasts of the spectrum. We define the following statistics,

$$SSE(l) = \frac{1}{K} \sum_{k=1}^{K} \int_\lambda \left(\tilde{f}^{(k)}(\lambda, n+l) - \hat{f}_L^{(k)}(\lambda, n+l) \right)^2 d\lambda$$

where l is the forecast step($l = 1, \ldots, 10$), $k(k = 1, \ldots, K)$ is the experimental times, $\hat{f}_L^{(k)}(\lambda, n+l)$ is the estimates of spectral density at $n+l$ of the kth experiment using the nonparametric estimator (1), $\tilde{f}^{(k)}(\lambda, n+l)$ is the corresponding forecasts, using each method stated above (i.e. $\tilde{f}_A^{(k)}(\lambda, n+l)$, $\tilde{f}_B^{(k)}(\lambda, n+l)$ and $\tilde{f}_C^{(k)}(\lambda, n+l)$). Here for estimation of $\hat{f}_L^{(k)}(\lambda, n+l)$, we used the time series data in the time interval which is regarded to be locally stationary. The methodolgy to determine this time interval is shown in Appendix. Also, the reason why we used the nonparametric estimator for $\hat{f}_L^{(k)}(\lambda, n+l)$ is to examine whether or not our assumption that the model (2) is practically reasonable as a statistical model of the sea surface is true. $SSE(l)$ evaluates the degree of spectral form's agreement between $\hat{f}_L^{(k)}(\lambda, n+l)$ and $\tilde{f}^{(k)}(\lambda, n+l)$. But it may not be guaranteed that the maximum value and the dominant frequency of the forecasts of spectral density, which minimizes $SSE(l)$, are optimized. So, we also define $SFE(l)$ and $SME(l)$ to examine the degree of agreement between extreme values of $\hat{f}_L^{(k)}(\lambda, n+l)$ and $\tilde{f}^{(k)}(\lambda, n+l)$; $SFE(l)$ and $SME(l)$ are defined to examine the degree of accordance with the dominant frequency and maximum value of the spectral density, respectively.

$$SFE(l) = \frac{1}{K} \sum_{k=1}^{K} \left(\tilde{\lambda}_{max}^{(k)}(n+l) - \hat{\lambda}_{L,max}^{(k)}(n+l) \right)^2$$

$$SME(l) = \frac{1}{K}\sum_{k=1}^{K}\Big(\max_{\lambda}(\tilde{f}^{(k)}(\lambda, n+l)) - \max_{\lambda}(\hat{f}_L^{(k)}(\lambda, n+l))\Big)^2$$

where $\tilde{\lambda}_{max}^{(k)} = \arg\max_\lambda \tilde{f}^{(k)}(\lambda, n+l)$ and $\hat{\lambda}_{L,max}^{(k)} = \arg\max_\lambda \hat{f}_L^{(k)}(\lambda, n+l)$, respectively. We can evaluate as the best predictor when both of the above statistics take the smallest value.

Step	(A)	(B)	(C01)	(C02)	Our Method
1	110.4	42.6	219.7	90.5	36.1
2	212.9	71.2	334.5	149.0	58.7
3	248.3	79.9	357.0	122.8	67.7
4	265.9	94.2	339.5	108.0	75.0
5	329.6	123.2	373.6	120.3	96.3
6	342.9	137.3	427.3	137.8	112.1
7	340.8	143.8	507.2	131.5	115.9
8	412.0	194.2	681.1	182.6	166.1
9	595.7	337.3	1012.0	425.8	259.1
10	764.1	480.0	1364.0	639.6	317.7

Table 1(a). Numerical Results on $SSE(l)$

Step	(A)	(B)	(C01)	(C02)	Our Method
1	0.00024	0.00012	0.00035	0.00025	0.00009
2	0.00048	0.00027	0.00381	0.00088	0.00083
3	0.00063	0.00025	0.01140	0.00113	0.00052
4	0.00074	0.00028	0.01333	0.00116	0.00018
5	0.00081	0.00041	0.01267	0.00057	0.00027
6	0.00074	0.00042	0.01236	0.00075	0.00029
7	0.00061	0.00037	0.01836	0.00084	0.00023
8	0.00055	0.00034	0.02604	0.00058	0.00024
9	0.00052	0.00030	0.03265	0.00042	0.00024
10	0.00049	0.00088	0.04064	0.00057	0.00019

Table 1(b). Numerical Results on $SFE(l)$

Step	(A)	(B)	(C01)	(C02)	Our Method
1	14.9	5.1	28.3	10.2	3.7
2	27.6	6.8	43.2	16.3	6.2
3	34.4	7.4	44.9	13.8	6.9
4	34.4	8.9	41.3	10.6	8.8
5	49.3	13.3	46.2	12.1	11.4
6	52.4	15.7	55.9	15.0	14.2
7	56.3	19.3	73.8	16.2	15.9
8	83.8	36.5	125.9	32.9	31.7
9	152.3	89.6	250.9	113.1	59.5
10	234.7	149.6	392.4	195.3	84.8

Table 1(c). Numerical Results on $SME(l)$

We show the numerical results in the followings. Tables 1(a)~1(c) show the numerical results on $SSE(l)$, $SFE(l)$ and $SME(l)$, respectively. Here "Step" shows the forecast step l, where 1 step corresponds to 40 seconds. Also, the experimental times K is 1350.

First, we focus on the result of $SSE(l)$ in Table 1(a). The predictors (B),(C02) and our predictor take smaller values than (A) at any forecast step, which suggests that constructing time-varying statistical structures are essentially effective for the improvement of forecasting performances. Also, from the comparison between (B) and (C01), it suggests that it is effective to construct the multivariate autoregressive model to $(\beta_{1,n}, \ldots, \beta_{p,n})$. This becomes more evident when we compare the results between the pairs of predictors (C01,C02) and (B,our predictor); both of C01 and C02 take larger values than (B) and our predictor. Moreover, the comparison between (B) and our predictor shows that wind's motion give good effects for forecasting especially for long-term forecasting. We see similar tendencies on $SFE(l)$ and $SME(l)$ in Tables 1(b) and 1(c), although (C02) becomes worse than (A) in $SFE(l)$. Our predictor gives the best forecasting performances among 5 predictors and we can say from our experiments that this predictor is practically available for forecasting the tendencies of forms, dominant frequency and maximum value of time varying spectral density. We can find out the similar tendency on the results of $SFE(l)$ and $SME(l)$ in Tables 1(b) and 1(c). These results show that our model also give good forecasting performances on the maximum value and dominant frequency.

5. Concluding Remarks

The numerical results shown in the previous section suggests that our predictor is practically reasonable for forecasting future changes on the time-varying spectral density function in the wave developing process.

Appendix

In this paper, we estimate the spectral density function based on a locally stationary autoregressive model at the time point $t = nM$ in the following way. First, it is necessary to estimate the time interval, M_L, in which we can regard the movement $\{X_t\}$ in the time interval $[nM - M_L + 1, nM]$ as a stationary time series. We choose the value of it, say M_L^*, such that the sum of squared residual errors of X_t when the autoregressive model (2) is fitted,

$$SSR(M_L) = \frac{1}{M_L} \sum_{t=nM-M_L+1}^{nM} (X_t - \hat{X}_t)^2$$

is minimized with repect to M_L, where, $\hat{X}_t$ is estimated model from the data in $[nM - M_L + 1, nM]$. Next, using the data in the time interval $[nM - M_L^* + 1, nM]$, we obtain the estimates of each parameter of (2), $\hat{a}_{M_L^*,1}, \ldots, \hat{a}_{M_L^*,q}$, and the innovation variance, $\hat{\sigma}^2_{M_L^*}$. Here, we used

AIC (Akaike Information Criterion) for order determination. Hence, from the above estimates, we obtain the spectral estimates,

$$\hat{f}(\lambda) = \frac{\hat{\sigma}^2_{M_L^*}}{|1 + \hat{a}_{M_L^*,1} e^{-i2\pi\lambda} + \cdots + \hat{a}_{M_L^*,q} e^{-i2q\pi\lambda}|^2}$$

References

[1]. Box, G.E.P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.

[2]. Kitagawa, G. (1993). FORTRAN77 Time Series Programming, The Iwanami Computer Science Series(in Japanese).

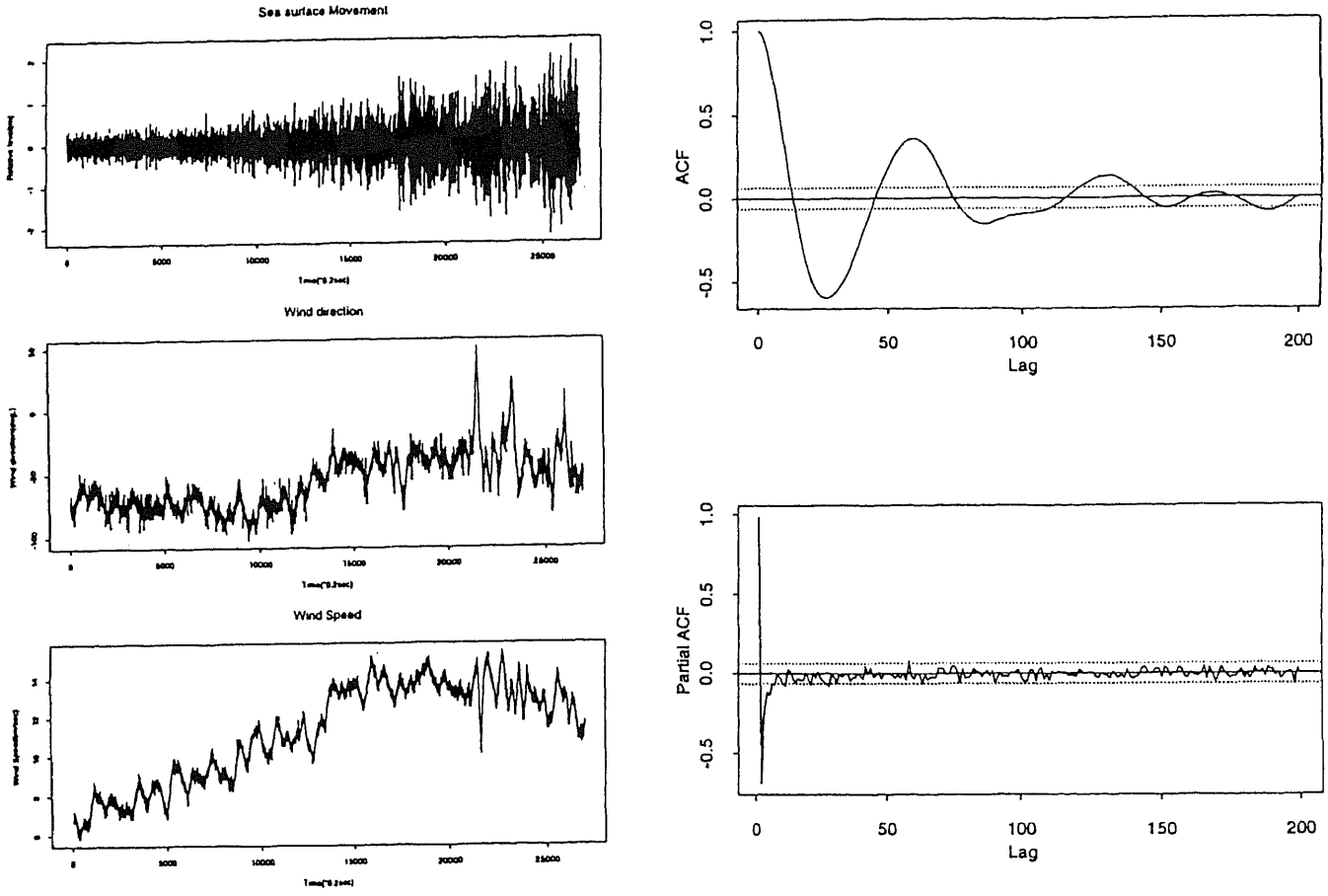


Fig. 1 (*left*) Observed time series ((a) sea surface level, (b) wind direction, (c) wind speed)
Fig. 3 (*right*) (a) autocorrelation function and (b) partial autocorrelation function (T=1000)

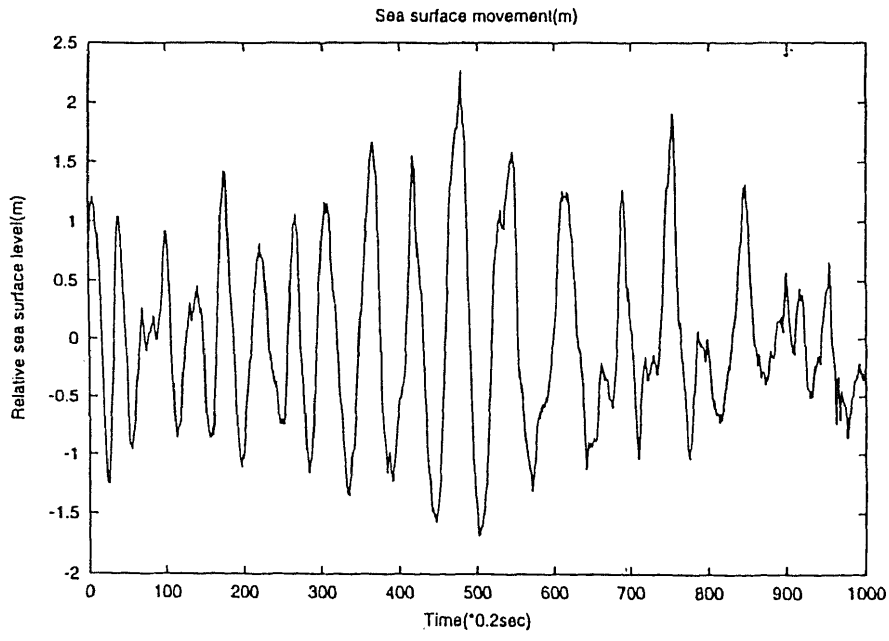
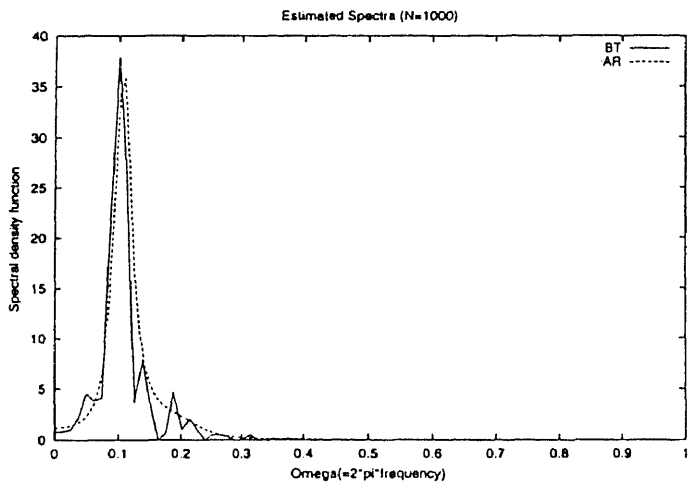


Fig. 2 An example of sea surface movement (T=1000)



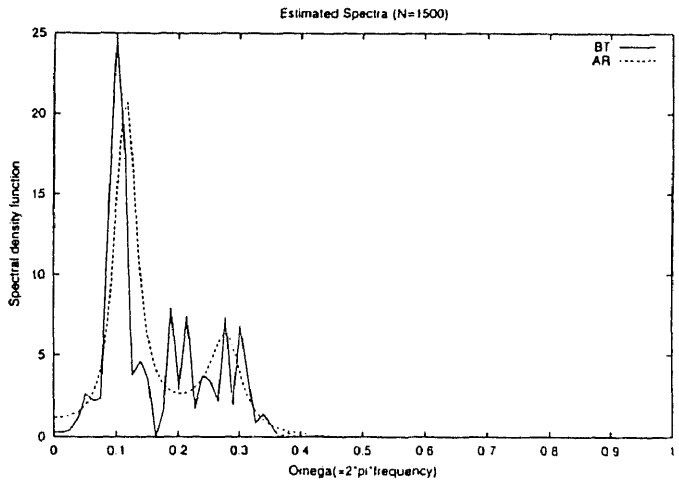
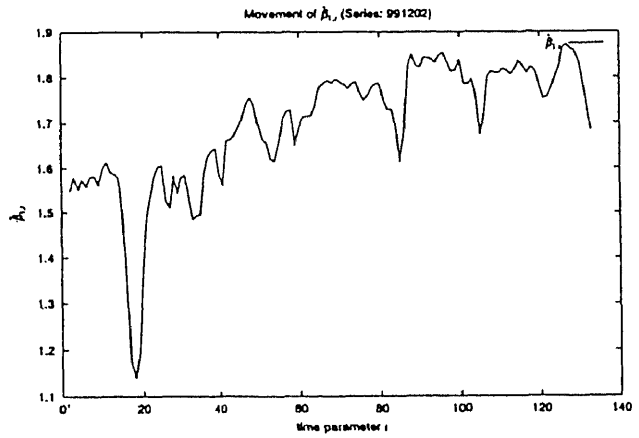
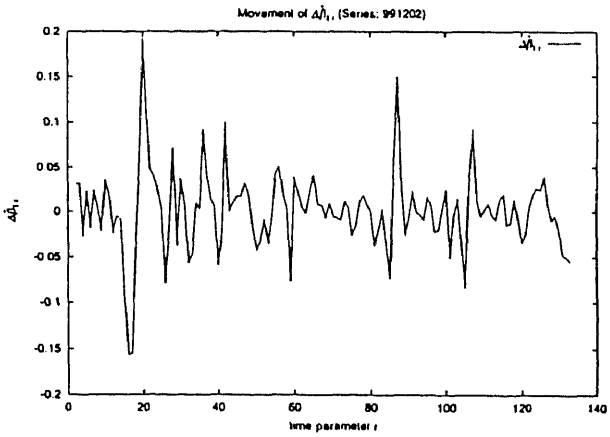
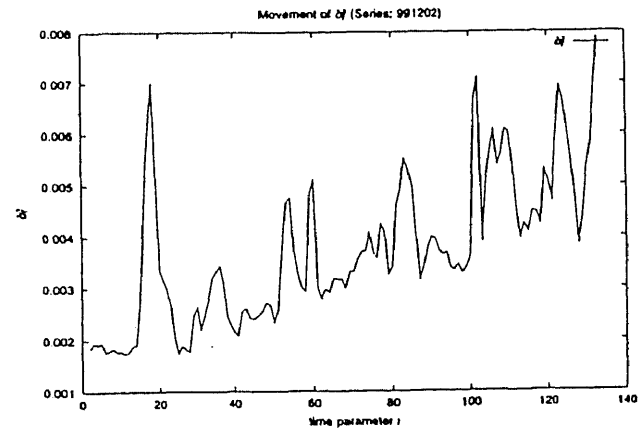


Fig. 4 (*left*) Estimates of spectral density function using the time sereis of Fig. 2 (T=1000)
Fig. 5 (*right*) Estimates of spectral density function at the same time point as in Fig. 4 (T=1500)







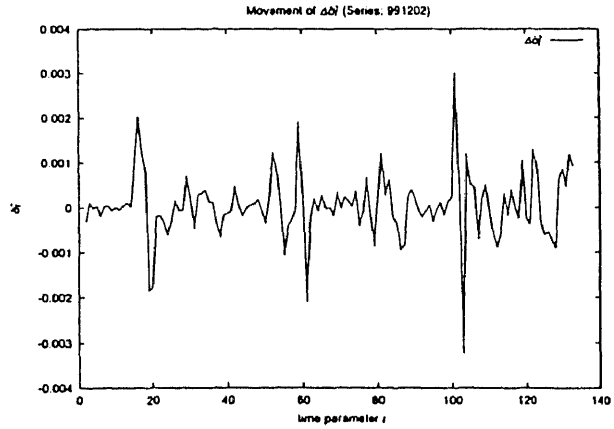
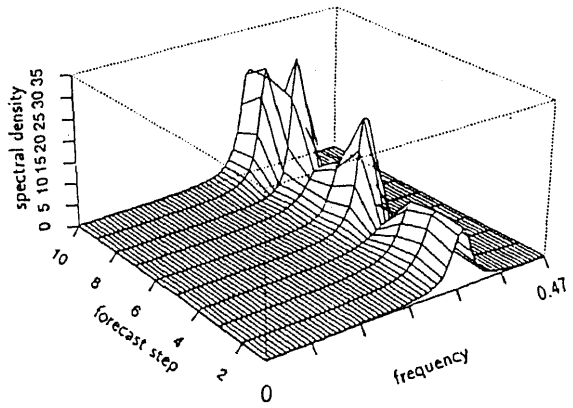
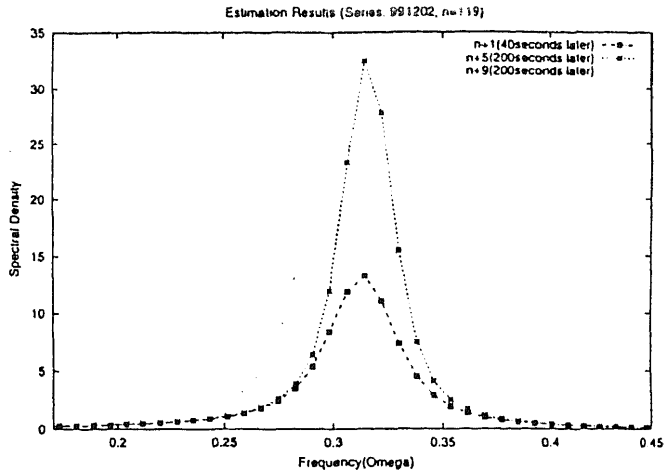
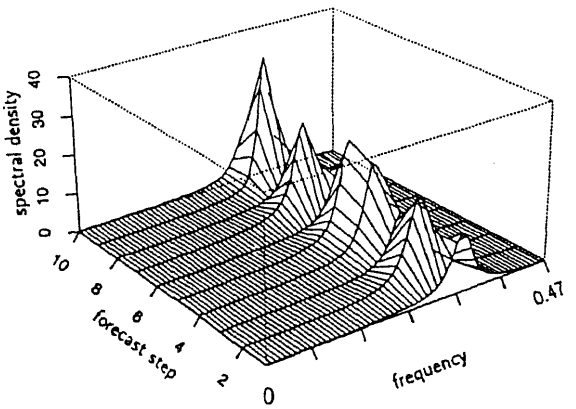


Fig. 6 (*left*) The behaviors of $\hat{\beta}_{1,n}$ and $\hat{\sigma}_n^2$
Fig. 7 (*right*) The differenced series $\nabla\hat{\beta}_{1,n}$ and $\nabla\hat{\sigma}_n^2$







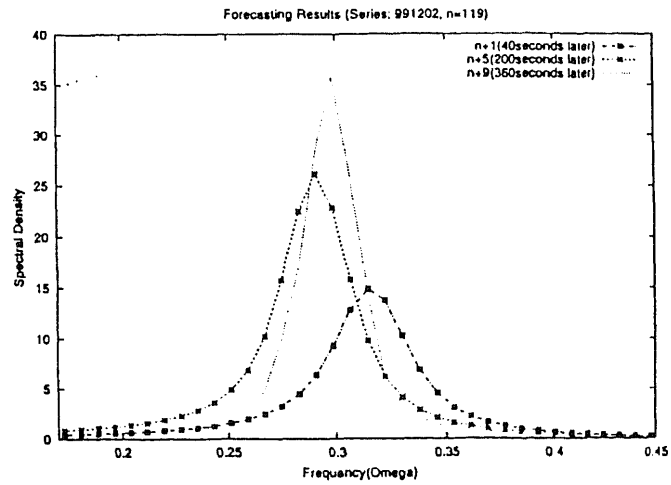


Fig. 8 (*left*) Bird's-eye views on time-varying spectra ((*top*) estimates, (*bottom*) forecasts)
Fig. 9 (*right*) Simultaneous plots on time-varying spectra ((*top*) estimates, (*bottom*) forecasts)

K-Asymptotics Associated with Deterministic Trends in the Integrated and Near-Integrated Processes

一橋大学大学院経済学研究科　田中　勝人

いわゆる単位根検定は，時系列データに含まれるトレンドあるいは非定常性が確率的か確定的かを判断するものであるが，その場合，確率トレンドと確定トレンドはデータに基づいて明確に区別できることが前提となっている．本論文では，この二分法が無効になるような状況を設定して，そこにおける統計的な諸性質を考察する．

そのために，DGP としては確率トレンドを表す I(1) 過程

$$y_t = y_{t-1} + u_t, \qquad y_0 = 0, \qquad (t = 1, \ldots, T),$$

$$u_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j|\alpha_j| < \infty, \quad \alpha \equiv \sum_{j=0}^{\infty} \alpha_j \neq 0,$$

を想定する．ここで，$\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$ であり，短記憶過程 $\{u_t\}$ は長期分散

$$\sigma_L^2 = \lim_{T\to\infty} \frac{1}{T} V\left(\sum_{t=1}^{T} u_t\right) = \sigma^2 \left(\sum_{j=0}^{\infty} \alpha_j\right)^2,$$

を持つものとする．また，長期分散 σ_L^2 に対応して，短期分散 $V(u_t)$ を σ_S^2 と表すことにする．さらに，ある $p\,(>2)$ に対して，$E\left(|\varepsilon_t|^p\right) < \infty$ であるとする．

以上の設定のもとで，まず，$\{y_t\}$ を一次独立な K 項からなる三角関数を使った確定トレンド $\phi_1, \cdots, \phi_K$ へ回帰した回帰式

$$y_t = \sum_{k=1}^{K} \tilde{b}_k \phi_k\left(\frac{t}{T}\right) + \tilde{u}_t = \tilde{b}(K)'\phi(K, t/T) + \tilde{u}_t$$

を考える．このとき，大標本のもとで $(T \to \infty)$，回帰の項数 K を大きくするとき（K-**asymptotics**）の統計量の挙動について次のことが成り立つ．

(a) $c(K)'\tilde{b}(K)/\sqrt{T} \quad \Rightarrow \quad \mathrm{N}(0, \sigma_0^2), \quad \|c(K)\| = 1, \sigma_0^2 = \sigma_L^2 \sum_{n=1}^{\infty} c_n^2 / \left((n-1/2)^2\pi^2\right),$
(b) $\sum_{t=1}^{T} \tilde{u}_t{}^2/T^2 = O_p(1/K),$
(c) $t_{c(K)'\tilde{b}(K)}/\sqrt{T} = O_p(\sqrt{K}),$
(d) $R^2 \to 1$ in probability,
(e) $T \times DW = O_p(K).$

上記の諸性質が示唆することは，回帰の係数が全体として有意であり，決定係数が 1 に近づくという意味で，確率トレンドが確定トレンドで十分に表現されるということである．すなわち，K-asymptotics においては，確率トレンドと確定トレンドは区別できない．しかも，そのために使われる確定トレンドは，必ずしも多項式に限らないことが示される．

次に，単位根検定に使われるモデル，すなわち，確定トレンドの他に I(1) のラグも含む回帰関係

$$y_t = \hat{\rho} y_{t-1} + \sum_{k=1}^{K} \hat{b}_k \phi_k \left(\frac{t}{T}\right) + \hat{u}_t = \hat{\rho} y_{t-1} + \hat{b}(K)' \phi(K, t/T) + \hat{u}_t, \quad (t = 2, \ldots, T).$$

を考えると，その K-asymptotics として次の結果が得られる．

$$ADF_\rho \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi^2 K}{2}, \frac{\pi^4 K}{6}\right),$$

$$ADF_t \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi\sqrt{K}}{2}, \frac{\pi^2}{24}\right),$$

$$\sqrt{T} c(K)' \hat{b}(K)/\sigma_S \quad \Rightarrow \quad \mathrm{N}\left(0, \pi^2 K^2 \sum_{k=1}^{K} \frac{c_k^2}{(2k-1)^2}\right),$$

$$t_{c(K)'\hat{b}(K)} \quad = \quad O_p(\sqrt{K}).$$

ここで，ADF_ρ は単位根検定のための Dickey-Fuller 係数統計量であり，ADF_t は $t-$統計量である．以上より，K-asymptotics においては，単位根検定統計量は正規分布に従うこと，DGP には含まれない確定トレンド項が無数に追加されても，それらは有意であり，しかも，ラグ変数の係数の推定量は一致性をもっていることがわかる．

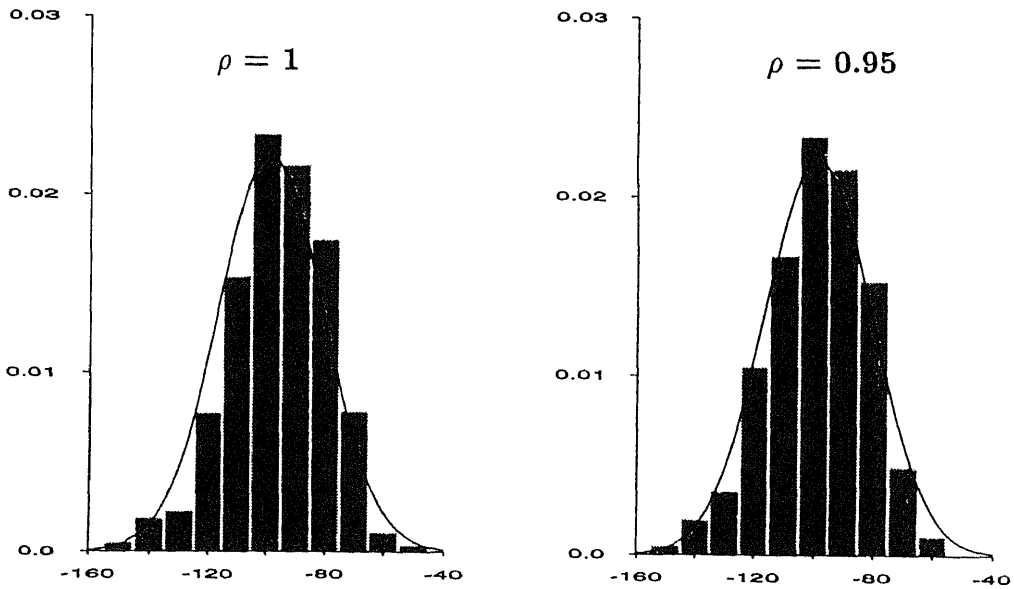
以上の結果は，純粋な I(1) だけではなく，局所的に I(1)（**near-integrated processes**）に従う DGP

$$y_t = \rho y_{t-1} + u_t, \qquad y_0 = 0, \qquad \rho = 1 - (c/T), \qquad (t = 1, \ldots, T),$$

についても，同様に成り立つことが示される．

最後に，K-asymptotics における単位根検定は，局所対立仮説に対しては無効になることが示される．その理由は，単位根検定統計量の局所対立仮説のもとでの分布は，下の図の例のように帰無分布と変わらないからである．

単位根検定統計量 $T(\hat{\rho} - 1)$ の分布（標本サイズ $= 400$，項数 $= 20$）



KRIGING法による画像修復の試み

東工大 情報理工　間瀬 茂
東芝 (東工大 情報理工)　池添 禎孝

クリギング法 (Kriging) 連続確率場からの離散的観測値で、確率場全体を連続的予測

地球統計学 (geostatistics) の中心的手法

統計理論の立場からは、最小自乗法による空間予測

一次元標準理論との相違点
- コバリアンス ⇒ バリオグラム
- 弱定常性 ⇒ 本質的定常性

本来ボーリングデータから鉱区全体の鉱石含有量を見積もる手法
⇒ 鉱山学、石油探索、環境科学、地球科学、気象学等多くの分野で使用

J. Besag の主張:　「空間統計学の多くの問題は画像解析！」
⇒ Geman 流の画像解析理論を考古学・疫学データの解析に応用

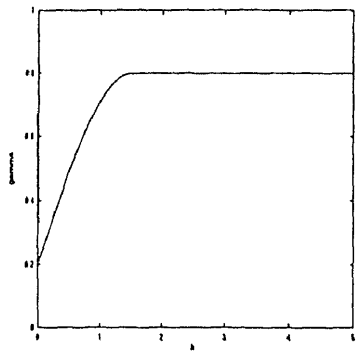
逆問題:「空間統計学の手法は画像解析の手法たり得るか？」
⇒ クリギング法を画像解析の道具として使用！

低解像度画像を高解像度画像の一部である標本と考える
低解像度画像 ⇒ もとの高解像度画像を予測 (画像補間)

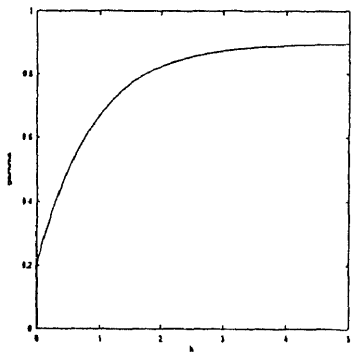
注意: 順序を逆に考えれば「画像圧縮」の問題とみなせる

本質的定常性　$\boldsymbol{E}\{Z(x) - Z(y)\} = 0$
$\boldsymbol{E}\{Z(x) - Z(y)\}^2$ は $x - y$ だけの関数
semivariogram　$\gamma(h) \equiv \boldsymbol{E}\{Z(x+h) - Z(x)\}^2/2$

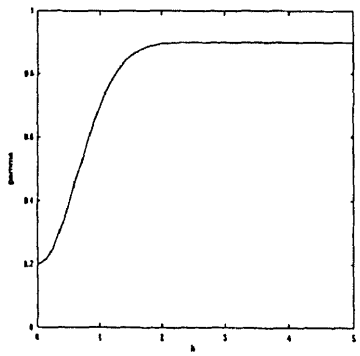
代表的セミバリオグラムモデルの形状:



球型



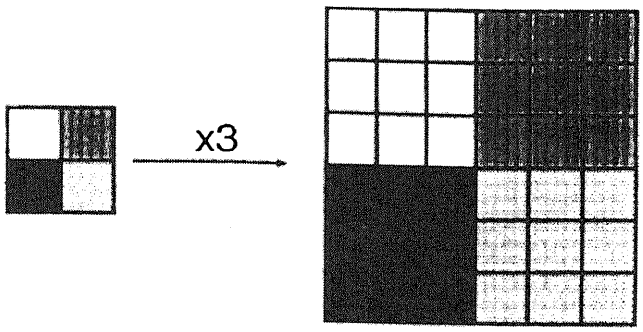
指数型



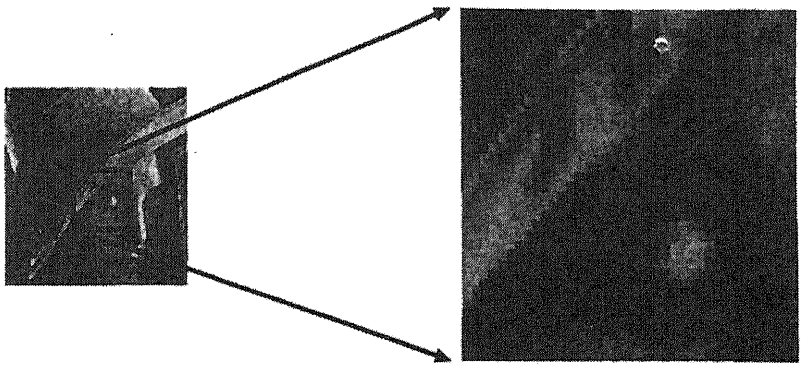
ガウス型

真の画像 I ⇒ $\mathbb{R}^2$ 上の確率場
ディジタル画像 J ⇒ $\mathbb{Z}^2$ 上の離散確率場
実際の画像 K ⇒ J を水平・垂直に $1/n$ に間引いた画像とみなす

間引き画像を単純に複製拡大 ⇒ モザイク状画像



複製による拡大



モザイク状拡大画像

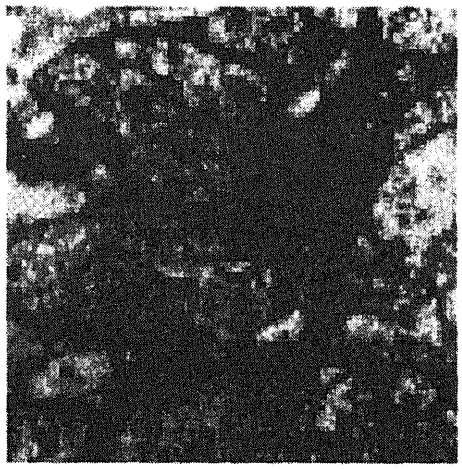
予想される困難とその対策:

画像のサイズは極めて大で、場所により性格が大きく異なる。
定常性、計算量、画像の境界 (エッジ)

定常性？⇒単純に無視
⇒semivariogram を天下りに与える
⇒cross validation によるパラメータ調整
⇒空間変動項 (普遍クリギング)
計算量！⇒moving kriging(局所的な区画毎に予測)

実験に使用した画像:

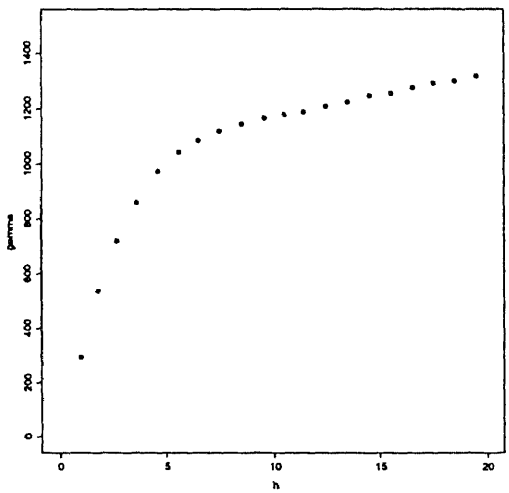
ブタ神経細胞画像 J
サイズは 258×258



$(1/3)^2$ に間引いた画像 K を
3^2 倍に単純に拡大した画像

真の画像 J とその補間画像 $\hat{J}$ の類似性の尺度
⇒ 画素毎の誤差の自乗平均 (MSE)

$$\mathrm{MSE} = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}\left|J(x,y) - \hat{J}(x,y)\right|^2 \quad (X, Y \text{ は画像のサイズ})$$

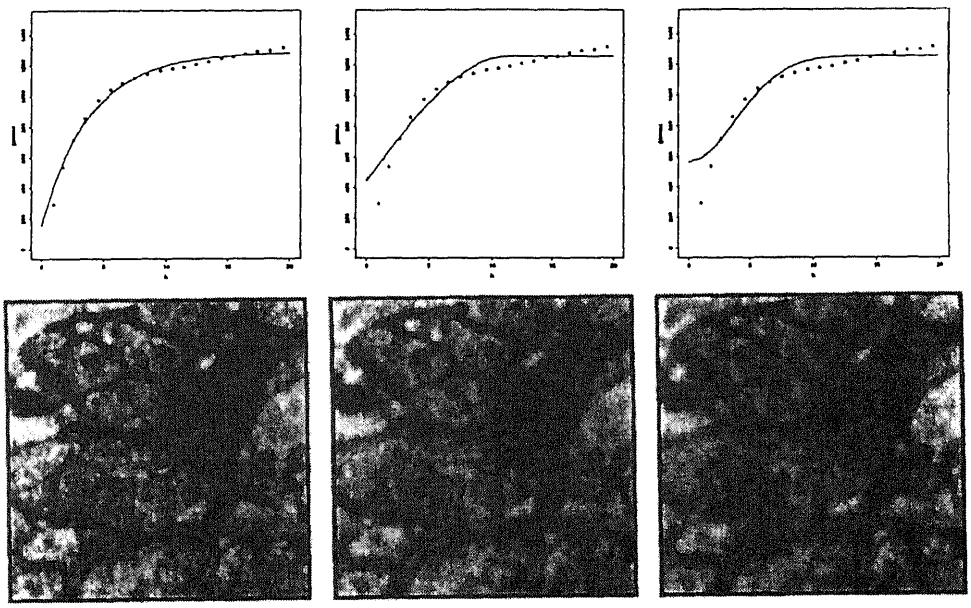


テスト画像の経験セミバリオグラム

	指数型	球型	ガウス型
θ_0(ナゲット効果)	156	446	565
θ_1	1.13×10^3	809	689
θ_2	3.90	11.2	5.40
$\theta_0 + \theta_1$(シル)	1.29×10^3	1.26×10^3	1.26×10^3
WRSS	913	3.21×10^3	2.99×10^3

Table 1: 各モデルのパラメータの推定値とその時の WRSS

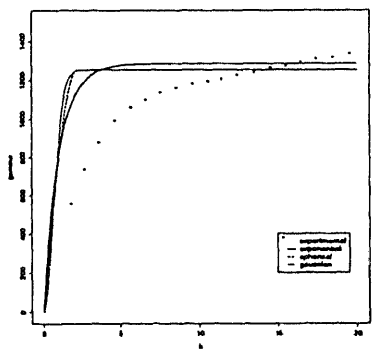
各種バリオグラムモデルと通常クリギングによる画像補間



指数型セミバリオグラム
通常クリギング (MSE 320)

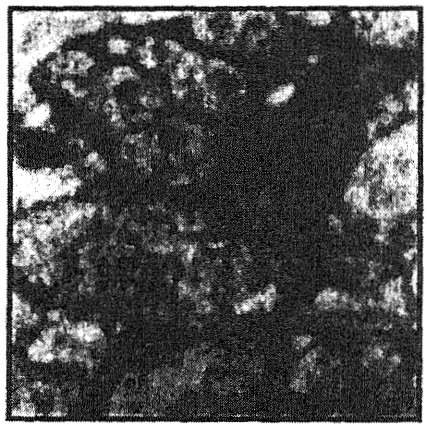
球型セミバリオグラム
通常クリギング (MSE 445)

ガウス型セミバリオグラム
通常クリギング (MSE 550)

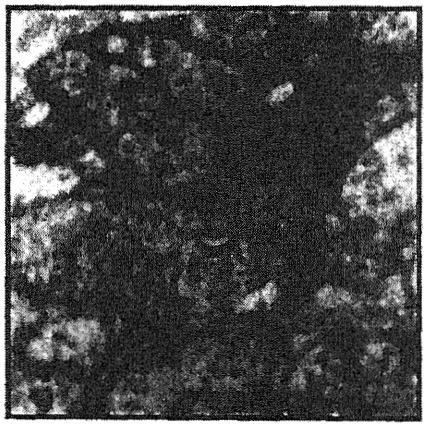


cross validation によるセミバリオグラム

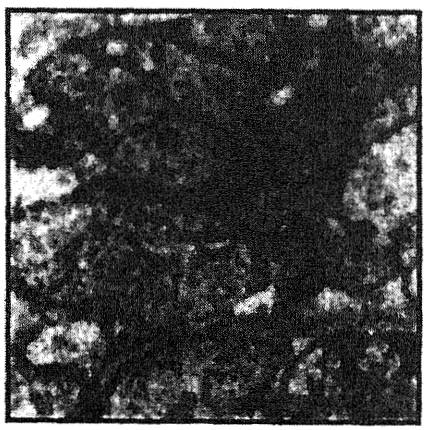
対応する修復画像



指数型セミバリオグラム
＋通常クリギング (MSE = 277)



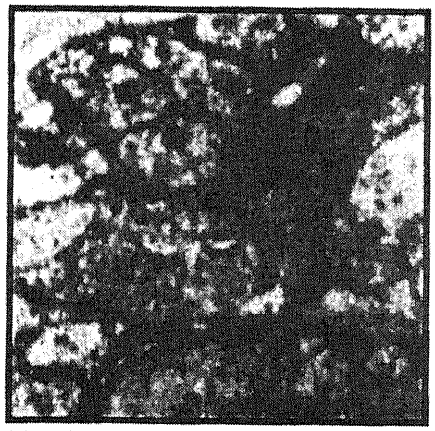
球型セミバリオグラム
＋通常クリギング (MSE = 257)



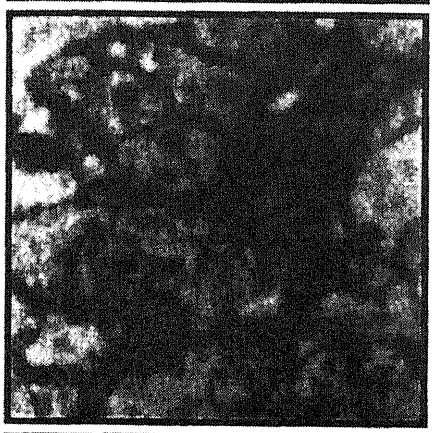
ガウス型セミバリオグラム
＋通常クリギング (MSE = 229)

普遍 kriging を用いた実験 ⇒ 失敗 (画像の局所的変動を追えない)

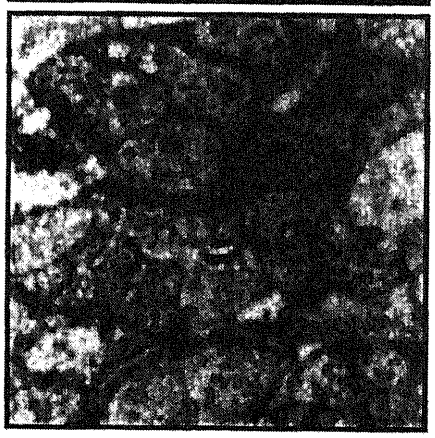
他の方法との比較 (kriging は有効？)



bicubic(MSE = 237)



cubic B-spline (MSE = 342)



cross val. + gauss (MSE = 229)

Discriminant Analysis for Locally Stationary Processes

KENJI SAKIYAMA[a] and MASANOBU TANIGUCHI [b]

Department of Mathematical Science, Osaka University

1 Introduction

There has been a series of works on discriminant analysis for stationary processes. For the problems of discriminanting two Gaussian processes by linear filtering, Shumway and Unger (1974) gave certain spectral approximations of Kullback-Leibler discrimination information rate, J-divergence rate and detection probabilities. They introduced linear discriminant filters maximizing these approximations and applied them to seismic records from selected earthquakes and nuclear explosions. Shumway (1982) gave an extensive review of various discriminant problems in time series. Using an approximation of the Gaussian likelihood ratio (GLR), Zhang and Taniguchi (1994) disscussed the parametric discriminant problems for non-Gaussian vector linear processes. They showed that the classification statistic based on a Gaussian likelihood ratio has some good properties, for example, asymptotic normality and non-Gaussian robustness. For discrimination between such non-Gaussian multivariate time series, Kakizawa, Shumway and Taniguchi (1998) have introduced a disparity measure, which includes the Kullback-Leibler discrimination infomation and the Chernoff information measure, and given applications to the problems of classifying earthquakes and mining explosions (see also Shumway and Stoffer (2000), Taniguchi and Kakizawa (2000)).

Dahlhaus (1997) has introduced a class of locally stationary processes (non-stationary processes), and established the asymptotic theory of statistical inference. In this paper we investigate the problems of classifying a multivariate non-Gaussian locally stationary process $\left\{X_{t,T} = (X_{t,T}^{(1)}, \ldots, X_{t,T}^{(d)})'\right\}$ into one of two categories described by two hypotheses: $\pi_1 : f(u,\lambda)$, $\pi_2 : g(u,\lambda)$, where $f(u,\lambda)$ and $g(u,\lambda)$ are time varying spectral density matrices. It is well known that the classification by the likelihood ratio (LR) gives the optimal classification (see Anderson (1984)). However, in time series situation, if the sample size T is large, LR is intractable even if we assume Gaussianity and stationarity of the process. Dahlhaus (1997) gave an approximation of the log-likelihood by

$$\frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[\log|f(u,\lambda)| + \mathrm{tr}\left\{I_N(u,\lambda)f^{-1}(u,\lambda)\right\}\right]d\lambda.$$

where M and N satisfy $T = NM$, $|f(u,\lambda)|$ is the determinant of $f(u,\lambda)$, and $I_N(u,\lambda)$ is a sort of periodgram of $\{X_{t,T}\}$. Although we do not assume Gaussianity of the process, we

[a] e-mail:sakiyama@sigmath.es.osaka-u.ac.jp
[b] e-mail:taniguti@sigmath.es.osaka-u.ac.jp

use the following approximated Gaussian LR

$$D(f:g)=\frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[\log\left\{\frac{|g(u_j,\lambda)|}{|f(u_j,\lambda)|}\right\}+\mathrm{tr}\left[I_N(u_j,\lambda)\left\{g^{-1}(u_j,\lambda)-f^{-1}(u_j,\lambda)\right\}\right]\right]d\lambda$$

as a classification statistics for our problem. That is, if $D(f:g)\gtrless 0$ we choose category π_1. Otherwise we choose category π_2.

This paper is organized as follows. In Section 2 we provide a limit theorem for an integral functional of $I_N(u,\lambda)$, which is useful to describe the asymptotics of $D(f:g)$ under π_1 and π_2.

In Section 3, by using the result in Section 2, we show that the classification statistic $D(f:g)$ gives a consistent classification rule. We also evaluate the misclassification probabilities of $D(f:g)$ when $g(u,\lambda)$ is contiguous to $f(u,\lambda)$. Then the problem of non-Gaussian robustness is addressed.

In Section 4, some numerical studies for the misclassification probabilities are given.

Throughout this paper we write $\boldsymbol{N}=\{1,2,\ldots\}$, and denote Kronecker's delta by $\delta(\cdot,\cdot)$.

2 A limit theorem for multivariate locally stationary processes

Dahlhaus(1997) developed asymptotic theory for univariate locally stationary processes. Since we discuss the discriminant analysis for maltivariate locally stationary processes, we extend some of his results to the case when the process concerned is multivariate.

We start with the following definition.

Definition 1 *A sequence of multivariate stochastic processes* $X_{t,T}=(X_{t,T}^{(1)},\ldots,X_{t,T}^{(d)})'$ $(t=1,\ldots,T)$ *is called locally stationary with transfer function matrix* $A_{t,T}(\lambda)=\{A_{t,T}(\lambda)_{a,b}:a,b=1,\ldots,d\}$ *and mean* $\mathbf{0}$ *if there exists a representation*

$$X_{t,T}=\int_{-\pi}^{\pi}\exp(i\lambda t)A_{t,T}(\lambda)d\xi(\lambda) \tag{1}$$

where the following holds.

(i) $\xi(\lambda)=(\xi^{(1)}(\lambda),\ldots,\xi^{(d)}(\lambda))'$ *is a complex valued vector process on* $[-\pi,\pi]$ *with* $\overline{\xi_a(\lambda)}=\xi_a(-\lambda)$, $E\xi_{a_j}=0$ *and*

$$cum\{d\xi_{a_1}(\lambda_1),\ldots,d\xi_{a_k}(\lambda_k)\}=\eta\left(\sum_{j=1}^{k}\lambda_j\right)g_{a_1,\ldots,a_k}(\lambda_1,\ldots,\lambda_{k-1})d\lambda_1,\ldots,d\lambda_k$$

where $cum\{\cdots\}$ *denotes the cumlant of k-th order,* $g_a(\lambda)=0$, $g_{a,b}(\lambda)=\delta(a,b)$, $|g_{a_1,\ldots,g_{a_k}}(\lambda_1,\ldots,\lambda_{k-1})|\leq C(k)$ *(*$C(k)$ *is constant) for all* $a_1,\ldots,a_k\in\{1,\ldots,d\}$ *and* $\eta(\lambda)=\sum_{l=-\infty}^{\infty}\delta(\lambda+2\pi l)$ *is the period* 2π *extension of the Dirac delta function.*

(ii) There exsits a constant K and a 2π-preiodic matrix valued function $A(u,\lambda) = \{A(u,\lambda)_{a,b} : a,b = 1,\ldots,d\} : [0,1] \times \boldsymbol{R} \to \boldsymbol{C}^{d\times d}$ with $\overline{A(u,\lambda)} = A(u,-\lambda)$ and

$$\sup_{t,\lambda} \left| A_{t,T}(\lambda)_{a,b} - A\left(\frac{t}{T},\lambda\right)_{a,b} \right| \leq KT^{-1}$$

for all $a,b \in \{1,\ldots,d\}$ and $T \in \boldsymbol{N}$, where $A(u,\lambda)$ is assumed to be continuous in u.

We call $f(u,\lambda) \equiv A(u,\lambda)\overline{A(u,\lambda)}'$ the time varying spectral density matrix of $\{X_{t,T}\}$. Letting

$$d_N^{(a)}(u,\lambda) = \sum_{s=0}^{N-1} X_{[uT]-N/2+s+1,T}^{(a)} \exp(-i\lambda s) , \tag{2}$$

we introduce the periodogram matrix $I_N(u,\lambda) = \{I_N(u,\lambda)_{a,b} : a,b = 1,\ldots,d\}$ over a segment of length N with midpoint $[uT]$, where

$$I_N(u,\lambda)_{a,b} = \frac{1}{2\pi N} d_N^{(a)}(u,\lambda) d_N^{(b)}(u,-\lambda) . \tag{3}$$

The shift from segments to segment is denoted by N. $I_N(u_j,\lambda)$ is calculated over segments with midpoints $u_jT = t_j = N(j-1/2), (j = 1,\ldots,M)$ where $T = NM$.

For $\phi : [0,1] \times [-\pi,\pi] \to \boldsymbol{C}^{d\times d}$, define

$$J_T(\phi) \equiv \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \mathrm{tr}\left\{\phi(u_j,\lambda) I_N(u_j,\lambda)\right\} d\lambda \tag{4}$$

and

$$J(\phi) \equiv \int_0^1 \int_{-\pi}^{\pi} \mathrm{tr}\left\{\phi(u,\lambda) f(u,\lambda)\right\} du d\lambda . \tag{5}$$

We set down the following assumption.

Assumption 1 *(i) The functions $A(u,\lambda)$ and $\phi_j(u,\lambda)$, $(j = 1,\ldots,k)$ are 2π-periodic in λ and the periodic extensions are differentiable in u and λ with uniformly bounded derivative $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda} A(u,\lambda)$ and $\frac{\partial}{\partial u}\frac{\partial}{\partial \lambda}\phi_j(u,\lambda)$. The fourth-order cumulant spectral density $g_{a_1,a_2,a_3,a_4}(\lambda_1,\lambda_2,\lambda_3)$ is continuous with respect to λ_1,λ_2 and λ_3.*

(ii) All the eigenvalues of $A(u,\lambda)\overline{A(u,\lambda)}'$ are bounded from below by some $C > 0$ uniformly in u and λ.

(iii) The parameters N and T fulfill the relations $T^{1/4} \ll N \ll T^{1/2}/\ln T$.

The following lemma is a multivariate version of Theorem A.2 of Dahlhaus (1997).

Lemma 1 *Suppose Assumpton 1 holds. Then,*

(i) $EJ_T(\phi) = J(\phi) + o\left(T^{-\frac{1}{2}}\right)$

(ii) $Tcov\{J_T(\phi_i), J_T(\phi_j)\} = V_{i,j} + o(1)$, *where*

$$\begin{aligned} V_{i,j} &= \int_0^1 \Big[4\pi \int_{-\pi}^{\pi} tr\{\phi_i(u,\lambda) f(u,\lambda) \phi_j(u,\lambda) f(u,\lambda)\}\, d\lambda \\ &\quad + 2\pi \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{a_1,a_2,a_3,a_4=1}^{d} \phi_i(u,\lambda)_{a_1,a_2} \phi_j(u,\mu)_{a_4,a_3} \\ &\qquad \times \sum_{b_1,b_2,b_3,b_4=1}^{d} A(u,\lambda)_{a_2,b_1} A(u,-\lambda)_{a_1,b_2} A(u,-\mu)_{a_4,b_3} A(u,\mu)_{a_3,b_4} \\ &\qquad \times g_{b_1,b_2,b_3,b_4}(\lambda,-\lambda,-\mu) d\lambda d\mu \Big] du. \end{aligned}$$

(iii) The quantities

$$\sqrt{T}\,[J_T(\phi_j) - J(\phi_j)]\,,\ j = 1,\ldots,k,$$

have, asymptoticaly, a normal distribution with zero mean vector and covariance matrix $V = \{V_{i,j} : i,j = 1,\ldots,k\}$.

We have placed the proofs of Lemmas and Theorems in Section 4.

3 Discriminant analysis for multivariate locally stationary processes

In this section we study the problems of classifying a multivariate locally stationary process $\{X_{t,T}\}$ into one of two categories described by two hypotheses:

$$\pi_1 : f(u,\lambda)\ \ ,\ \ \pi_2 : g(u,\lambda)$$

where $f(u,\lambda)$ and $g(u,\lambda)$ are $d \times d$ time varying spectral density matrices. For this discriminant problem, we use

$$D(f:g) = \frac{1}{4\pi M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left[\log \left\{ \frac{|g(u_j,\lambda)|}{|f(u_j,\lambda)|} \right\} + \mathrm{tr}\left[I_N(u_j,\lambda) \left\{ g^{-1}(u_j,\lambda) - f^{-1}(u_j,\lambda) \right\} \right] \right] d\lambda \tag{6}$$

as a classification statistic. That is, if $D(f:g) > 0$ we choose category π_1. Otherwise we choose category π_2. We set down the following further assumption.

Assumption 2 *There exists* $C > 0$ *such that the minimum eigenvalues of* $f(u,\lambda)$ *and* $g(u,\lambda)$ *are greater than* C *for all* u *and* λ.

The following theorem describes the asymptotics of $D(f:g)$ under π_1 and π_2.

Theorem 1 *Suppose that Assumptions 1 and 2 hold. Then, as* $T \to \infty$, *under* π_1

$$\sqrt{T}\,[D(f:g) - E\{D(f:g)|\pi_1\}] \xrightarrow{\mathcal{D}} N\left\{0, \sigma^2(f,g)\right\} \tag{7}$$

and under π_2

$$\sqrt{T}\left[D(f:g) - E\left\{D(f:g)|\pi_2\right\}\right] \xrightarrow{\mathcal{D}} N\left\{0, \sigma^2(g,f)\right\} \tag{8}$$

where

$$\begin{aligned}\sigma^2(f,g) = \int_0^1 \Bigg[& \frac{1}{4\pi}\int_{-\pi}^{\pi} tr\left\{f(u,\lambda)g^{-1}(u,\lambda) - I_d\right\}^2 d\lambda \\ & + \frac{1}{8\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{b_1,b_2,b_3,b_4=1}^{d} \left[A(u,\lambda)^*\left\{g^{-1}(u,\lambda) - f^{-1}(u,\lambda)\right\}A(u,\lambda)\right]_{b_2,b_1} \\ & \times \left[A(u,\mu)^*\left\{g^{-1}(u,\mu) - f^{-1}(u,\mu)\right\}A(u,\mu)\right]_{b_3,b_4} \\ & \times g_{b_1,b_2,b_3,b_4}(\lambda,-\lambda,-\mu)d\lambda d\mu\Bigg] du,\end{aligned}$$

$[M]_{a,b}$ *is the* (a,b) *element of matrix* M, I_d *is the identity matrix, and* $*$ *denotes complex-conjugate transpose.*

If we use $D(f:g)$ as a classification criterion, the misclassification probabilities are

$$P(2|1) = P\{D(f:g) \le 0|\pi_1\} \ , \ P(1|2) = P\{D(f:g) > 0|\pi_2\}.$$

The following theorem shows that our classification statistic is asymptotically consistent.

Theorem 2 *Under Assumptions 1 and 2,*

$$\lim_{T\to\infty} P(2|1) = 0 \ , \ \lim_{T\to\infty} P(1|2) = 0.$$

To evaluate the goodness of $D(f:g)$ we assume that $g(u,\lambda)$ is contiguous to $f(u,\lambda)$. Now we set the spectral densities as

$$\pi_1 : f(u,\lambda) = f_0(u,\lambda) \ , \ \pi_2 : g(u,\lambda) = f_{\theta+h/\sqrt{T}}(u,\lambda) \tag{9}$$

where $\theta \in \Theta \subset \boldsymbol{R}^q$ and $h = (h_1,\ldots,h_q)'$.

Assumption 3 *(i) We observe the realization* $X_{1,T},\ldots,X_{T,T}$ *of a* d*-dimensional locally stationary process with mean* $\mathbf{0}$ *and transfer function matrix* $A_{t,T}(\lambda)$*. The time varying spectral density matrix is* $f_\theta(u,\lambda) = A_\theta(u,\lambda)\overline{A_\theta(u,\lambda)}'$, $\theta \in \Theta \subset \boldsymbol{R}^q$*, where* Θ *is compact.*

(ii) All the eigenvalues of $f_\theta(u,\lambda)$ *are bounded from below by some constant* $C > 0$ *uniformly in* θ*,* u *and* λ*.*

(iii) The components of $f_\theta(u,\lambda)$*,* $\nabla f_\theta(u,\lambda)$ *and* $\nabla^2 f_\theta(u,\lambda)$ *are continuous on* $\Theta \times [0,1] \times [-\pi,\pi]$ *(*∇ *denotes the gradient with respect to* θ*).*

(iv) N *and* T *fulfill the relations* $T^{1/4} \ll N \ll T^{1/2}/\ln T$*.*

Theorem 3 *Under (9), we suppose Assumption 3. If we use $D(f:g)$ as a classification criterion, then*

$$\lim_{T\to\infty} P(2|1) = \lim_{T\to\infty} P(1|2) = \Phi\left[\frac{-\frac{1}{2}F(\theta)}{\{F(\theta)+D(\theta)\}^{\frac{1}{2}}}\right] \tag{10}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution,

$$F(\theta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\left\{\sum_{i=1}^{q} h_i \nabla_i f_\theta(u,\lambda) f^{-1}(u,\lambda)\right\}^2 du d\lambda,$$

$$\begin{aligned} D(\theta) \;=\; \frac{1}{8\pi}\int_0^1 \sum_{b_1,b_2,b_3,b_4=1}^{d} \sum_{i=1}^{q} h_i \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \Big[A_\theta(u,\lambda)^*\Big\{f_\theta^{-1}(u,\lambda)\nabla_i f_\theta(u,\lambda) f_\theta^{-1}(u,\lambda)\Big\}A_\theta(u,\lambda)\Big]_{b_2,b_1} \\ \times \sum_{j=1}^{q} h_j \Big[A_\theta(u,\mu)^*\Big\{f_\theta^{-1}(u,\mu)\nabla_j f_\theta(u,\mu) f_\theta^{-1}(u,\mu)\Big\}A_\theta(u,\mu)\Big]_{b_3,b_4} \\ \times\, g_{b_1,b_2,b_3,b_4}(\lambda,-\lambda,-\mu)d\lambda d\mu\Big] du. \end{aligned}$$

References

[1] Anderson, T. W.(1984). *An Introduction to Multivariate Statistical Analysis.* New York: Wiley

[2] Brillinger, D. R.(1975). *Time Series: Data Analysis and Theory.* New York: Holt, Rinehart and Winston.

[3] Dahlhaus, R.(1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1-37

[4] Dahlhaus, R.(1998). A likelihood approximation for locally stationary process. Preprint.

[5] Hosoya, Y. and Taniguchi, M.(1982). A central limit thorem for stationary processes and parameter estimation of linear processes. *Ann. Statist.*, **10**, 135-153. *Correction: Ann. Statist.* **7**, 490-506.

[6] Kakizawa, Y., Shumway, R.H., and Taniguchi, M.(1998). Discrimination and clustering for multivariate time series. *J. Amer. Statist. Assoc.* **93** 328-340

[7] Magnus, J.R. and Neudecker, H.(1988). *Matrix Differential Calculus with Apprications in Statistics and Econometrics.* Chichester: Wiley.

[8] Shumway, R.H. and Unger, A.N.(1974). Linear discriminant functions for stationary time series. *J. Amer. Statist. Assoc.* **69** 948-956.

[9] Shumway, R. H.(1982). Discriminant analysis for time series. In *Handbook of Statistics.* Vol.2 ed. P. R. Krishnaiah and L. N. Kanal. Amsterdam: North-Holland, 1-46.

[10] Shumway, R. H. and Stoffer, D. S.(2000). *Time Series Analysis and ItsApplications.* New York: Springer-Verlag.

[11] Taniguchi, M. and Kakizawa, Y.(2000). *Asymptotic Theory of Statistical Inference for Time Series.* New York: Springer-Verlag.

[12] Zhang, G.Q. and Taniguchi, M.(1994). Discriminant analysis for stationary vector time series. *J. Time Ser. Anal.*, **15**, 117-126.

Prediction problems for square-transformed stationary processes

BY IN-BONG CHOI AND MASANOBU TANIGUCHI

Department of Mathematical Science, Osaka University , Toyonaka, 560, Japan
choi@sigmath.es.osaka-u.ac.jp taniguti@sigmath.es.osaka-u.ac.jp

SUMMARY

This paper discusses the prediction problems for square-transformed process, $Y_t = X_t^2$, where $\{X_t\}$ ia a stationary process with spectral density $g(\lambda)$. First, we evaluate the mean square prediction error for square-transformed process when the predictor is constructed from the true spectral density $g(\lambda)$. However, actually, it is often that the true structure $g(\lambda)$ is not completely specified. Hence, we consider the problem of misspecified prediction when a conjectured spectral density $f_\theta(\lambda)$, $\theta \in \Theta$, is fitted to $g(\lambda)$. Then, constructing the best linear predictor based on $f_\theta(\lambda)$, we can evaluate the prediction error for square-transformed process. Also, we consider a bias adjusted prediction problem for the above two cases. Furthermore, for non-Gaussian process we evaluate the mean square prediction errors when the best linear predictor is constructed by the true spectral density $g(\lambda)$ and the conjectured spectral density $f_\theta(\lambda)$, respectively. Since θ is usually unknown we estimate it by a quasi-MLE $\hat{\theta}_Q$. The second-order asymptotic approximations of the mean square errors of the predictors based on $g(\lambda)$ and $f_{\hat{\theta}_Q}(\lambda)$ are given. Finally we provide some numerical examples, which show some unexpected features.

Some key words: Transformation; Stationary process; Misspecified prediction; Spectral density; Conjectured spectral density; Best linear predictor; Quasi-MLE.

1. INTRODUCTION

There has been much discussion, in recent years, in the statistical prediction of time series analysis. However, there are many unsolved problems. Suppose that $\{X_t\}$ is a stationary process with zero mean and spectral density $g(\lambda)$. However, it is often that the true structure $g(\lambda)$ is not completely specified. If we fit an autoregressive model to a set of data by using some information criterion, the true order of the model is often incorrectly estimated, hence it is likely to be misspecified. This leads us to a misspecified prediction problem when a conjectured structure $f(\lambda)$ is fitted to $g(\lambda)$. It is Grenander & Rosenblatt(1957) that first evaluated the prediction error of the best linear predictor $\hat{X}_t$ which is computed on the basis of a conjectured spectral density $f(\lambda)$ although the true one is $g(\lambda)$. Then they showed $E|X_t - \hat{X}_t|^2 \propto \int_{-\pi}^{\pi} \frac{g(\lambda)}{f(\lambda)} d\lambda$. To recognize importance of the misspecified prediction problem, suppose that $g(\lambda) = (2\pi)^{-1}|1 - 0.5e^{i\lambda}|^2$, and $f(\lambda) = (2\pi)^{-1}|1 - (0.5 + \theta)e^{i\lambda} + 0.5\theta e^{2i\lambda}|^2$, $|\theta| < 1$. In this case $I = \int_{-\pi}^{\pi} \frac{g(\lambda)}{f(\lambda)} d\lambda = \frac{2\pi}{1-\theta^2}$, hence $I \to \infty$ if $|\theta| \nearrow 1$. Therefore investigation of the misspecified prediction problem seems important.

Choi & Taniguchi(2000) evaluated the asymptotic mean squared prediction error for the regression model with linear long-memory residual process in the case of misspecified prediction.

In this paper, we are concerned with the mean square error of the misspecified prediction for square-transformed process, $Y_t = X_t^2$. Here, the square transform problem is very important to predict the volatility for autoregressive heteroscedasticity(ARCH) and generalized ARCH(GARCH) models which introduced by Engle(1982). So our problem is not special. Regarding general polynomial transformations Hannan(1970) evaluated the autocovariance function for Hermite polynomials of a Gaussian process. Granger & Newbold(1976)(for short G-N) addressed the problems of prediction for a class of nonlinear transformations $T = T(\cdot)$ of a Gaussian process. It is assumed that T can be approximated by Hermite polynomials. This excellent paper provided the mean square prediction errors of various predictors. Here we note the distinction between G-N and ours. G-N deals with more general transformation, but is restricted to Gaussian processes. Our paper is restricted to the case of square-transformation. But we do not assume Gaussianity of the process concerned, and consider the misspecified prediction problems.

This paper is organized as follows. In Section 2, under the assumption that $\{X_t\}$ is a first order autogressive process we evaluate the mean square prediction error for transformed process, $Y_t = X_t^2$, using a naive predictor i.e., (the best linear predictor of $X_t)^2$ and the best linear predictor of Y_t in terms of $Y_{t-1}, Y_{t-2}, \ldots$. Also, in the naive prediction, we consider a bias adjusted prediction problem. Futhermore, we give the mean square prediction error and the bias adjusted mean square prediction error for transformed scalar linear processes when the best linear predictors are constructed by the true spectral density and a conjectured spectral density, respectively. In Section 3, for non-Gaussian linear process we evaluate the mean square prediction error when the best linear predictors are constructed by the true spectral density and a conjectured spectral density, respectively. In Section 4, we derive the asymptotic mean squared prediction error for transformed linear processes when the predictor is constructed by a parametric spectral density model. Here the parameter is estimated by the quasi-MLE. Section 5 illuminates some unexpected aspects of our prediction problems nurmerically.

2. Transformed prediction problem for Gaussian scalar processes

Suppose that $\{X_t\}$ is generated by

$$X_t = \theta X_{t-1} + \epsilon_t \tag{1}$$

where $|\theta| < 1$ and the innovation process ϵ_t is independent identically distributed $\epsilon_t \sim N(0, 1-\theta^2)$. We now consider a naive prediction for the transformation of the form $Y_t = X_t^2$. As the naive predictor of Y_t we use $(\theta X_{t-1})^2$. Then the mean square prediction error is evaluated as follows.

$$\begin{aligned} E[\{Y_t - \hat{Y}_t\}^2] &= E[\{(\theta X_{t-1} + \epsilon_t)^2 - (\theta X_{t-1})^2\}^2] \\ &= 4\theta^2 E[X_{t-1}^2]E[\epsilon_t^2] + E[\epsilon_t^4] = (1-\theta^2)(3+\theta^2) = E_1, (\text{say}). \end{aligned} \tag{2}$$

It is seen that $E(Y_t) = 1$, $E\{(Y_t - 1)(Y_{t+l} - 1)\} = \theta^{2l}$, and that $\theta^2(Y_{t-1} - 1)$ is the best linear predictor of $Y_t - 1$ based on $Y_{t-1}, Y_{t-2}, \ldots$. Hence the mean square prediction error(MSPE) is given by

$$\begin{aligned} E[\{Y_t - 1 - \theta^2(Y_{t-1} - 1)\}^2] &= E[\{X_t^2 - 1 - \theta^2(X_{t-1}^2 - 1)\}^2] \\ &= 2(1 - \theta^4) = E_2, (\text{say}). \end{aligned} \tag{3}$$

Therefore, since $E_1 - E_2 = (\theta^2 - 1)^2 > 0$, this implies that the predictor $\theta^2(Y_{t-1} - 1)$ is better than the naive predictor.

Next we consider the bias adjusted prediction problem in (2). Note that $E[X_t^2] = 1$ and $E[(\theta X_{t-1})^2] = \theta^2$. Then the mean square prediction error is

$$E[\{X_t^2 - 1 - (\theta X_{t-1})^2 + \theta^2\}^2] = 2(1 - \theta^4) = E_3, (\text{say}). \tag{4}$$

Therefore, from the above result, we see that $E_2 = E_3 < E_1$.

Now we discuss the prediction problem in more general setting. Let $\{X_t\}$ be a Gaussian stationary process with zero mean, $EX_t^2 = 1$ and spectral density

$$g(\lambda) = \frac{1}{2\pi}|c(e^{-i\lambda})|^2, \quad |c(0)|^2 = \sigma^2. \tag{5}$$

We write the spectral representation of $\{X_t\}$ as

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} dz(\lambda). \tag{6}$$

Then we can construct the best linear predictor $\hat{X}_t$ based on $g(\lambda)$ as

$$\hat{X}_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{c(e^{-i\lambda}) - c(0)}{c(e^{-i\lambda})} dz(\lambda). \tag{7}$$

The mean square prediction error of the naive predictor $\hat{X}_t^2$ for X_t^2 is then

PROPOSITION 1.

$$\begin{aligned} E[\{X_t^2 - \hat{X}_t^2\}^2] &= 4\sigma^2 \int_{-\pi}^{\pi} g(\lambda) d\lambda - \sigma^4 \\ &= \text{MSPE}_1, \text{ (say)}. \end{aligned}$$

We place the proofs of Propositions of this paper in the Appendix.

Next we consider the bias adjusted prediction problem in Proposition 1. Solving $E[\hat{X}_t^2] - b = 0$ with respect to b, we have $b = \int_{-\pi}^{\pi} g(\lambda) d\lambda - \sigma^2$. The mean square prediction error is

PROPOSITION 2.

$$E[\{X_t^2 - 1 - \hat{X}_t^2 + b\}^2] = \text{MSPE}_1 - (1 - b)^2 = \text{MSPE}_2, \text{ (say)}.$$

Since the true spectral density $g(\lambda)$ is often unknown or misspecified, it is important to see what happens when a predictor is computed on the basis of a conjectured

spectral density $f(\lambda)$. Now we assume that the conjectured spectral density $f(\lambda)$ satisfies

$$\int_{-\pi}^{\pi} |\log f(\lambda)| d\lambda < \infty.$$

Writing

$$f(\lambda) = \frac{1}{2\pi} |c(e^{-i\lambda})|^2, \tag{8}$$

we can construct the best linear predictor of X_t based on $f(\lambda)$ as

$$\hat{X}_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{c(e^{-i\lambda}) - c(0)}{c(e^{-i\lambda})} dz(\lambda). \tag{9}$$

Another condition necessary to give integral (9) meaning is

$$\int_{-\pi}^{\pi} \frac{g(\lambda)}{f(\lambda)} d\lambda < \infty.$$

The mean square prediction error of $\hat{X}_t^2$ for X_t^2 is then

PROPOSITION 3.

$$\begin{aligned} E[\{X_t^2 - \hat{X}_t^2\}^2] &= \frac{2}{\pi^2} \left\{ \int_{-\pi}^{\pi} \overline{c(e^{-i\lambda})} c(0) \frac{g(\lambda)}{f(\lambda)} d\lambda \right\}^2 + \frac{3\sigma^4}{4\pi^2} \left\{ \int_{-\pi}^{\pi} \frac{g(\lambda)}{f(\lambda)} d\lambda \right\}^2 \\ &\quad - \int_{-\pi}^{\pi} \frac{g(\lambda)}{f(\lambda)} d\lambda \left\{ \frac{5\sigma^2}{2\pi^2} \int_{-\pi}^{\pi} \overline{c(e^{-i\lambda})} c(0) \frac{g(\lambda)}{f(\lambda)} d\lambda \right. \\ &\quad \left. + \frac{\sigma^2}{2\pi^2} \int_{-\pi}^{\pi} c(e^{-i\lambda}) \overline{c(0)} \frac{g(\lambda)}{f(\lambda)} d\lambda - \frac{2\sigma^2}{\pi} \int_{-\pi}^{\pi} g(\lambda) d\lambda \right\} \\ &= \text{MSPE}_3, \text{ (say)}. \end{aligned}$$

Next we consider the bias adjusted misspecified prediction problem in Proposition 3. Let $E[\hat{X}_t^2] - b = 0$, then

$$\begin{aligned} b &= E\left[\int_{-\pi}^{\pi} e^{it\lambda} \frac{c(e^{-i\lambda}) - c(0)}{c(e^{-i\lambda})} dz(\lambda) \int_{-\pi}^{\pi} e^{it\mu} \frac{c(e^{-i\mu}) - c(0)}{c(e^{-i\mu})} dz(\mu) \right] \\ &= \int_{-\pi}^{\pi} g(\lambda) d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} c(e^{-i\lambda}) \overline{c(0)} \frac{g(\lambda)}{f(\lambda)} d\lambda - \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{c(e^{-i\lambda})} c(0) \frac{g(\lambda)}{f(\lambda)} d\lambda \quad (10) \\ &\quad + \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \frac{g(\lambda)}{f(\lambda)} d\lambda. \end{aligned}$$

Then the mean square prediction error is given by the following proposition.

PROPOSITION 4.

$$E[\{X_t^2 - 1 - \hat{X}_t^2 + b\}^2] = \text{MSPE}_3 - (1-b)^2 = \text{MSPE}_4, \text{ (say)}.$$

Let $\{X_t\}$ be a Gaussian process with spectral density $g(\lambda)$. We next consider the misspecified prediction for the transformation $Y_t = X_t^2 - 1$. Write the spectral representation of Y_t as

$$Y_t = \int_{-\pi}^{\pi} e^{it\lambda} d\omega(\lambda).$$

It is seen that the spectral density of $\{Y_t\}$ is given by $h(\lambda) = 2g^{*2}(\lambda)$(see Hannan(1970, p. 83)). Denote a conjectured spectral density for $g(\lambda)$ by

$$q(\lambda) = \frac{1}{2\pi}|a(e^{-i\lambda})|^2.$$

Based on $q(\lambda)$ we can construct the best linear predictor of Y_t by

$$\hat{Y}_t = \int_{-\pi}^{\pi} e^{it\lambda}\frac{a(e^{-i\lambda}) - a(0)}{a(e^{-i\lambda})}d\omega(\lambda).$$

Then the mean square error of the prediction is

$$\begin{aligned} E[\{Y_t - \hat{Y}_t\}^2] &= E\left[a(0)\int_{-\pi}^{\pi} e^{it\lambda}\frac{1}{a(e^{-i\lambda})}d\omega(\lambda)\right]^2 \\ &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{h(\lambda)}{q(\lambda)}d\lambda. \end{aligned}$$

3. Transformed prediction problem for non-Gaussian scalar processes

Untill now, we restricted to the case when the process concerned is Gaussian. In this section, we consider the case where $\{X_t\}$ is a non-Gaussian process. Let $\{X_t\}$ be a non-Gaussian stationary process with spectral density

$$g(\lambda) = \frac{1}{2\pi}|c_g(e^{-i\lambda})|^2, \quad |c_g(0)|^2 = \sigma^2, \tag{11}$$

and spectral representation

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}dz(\lambda). \tag{12}$$

Also, we assume that the process $\{X_t\}$ has the fourth-order cumulant spectral density $g_4(\lambda_1, \lambda_2, \lambda_3)$. For the fundamental properties of the cumulant, see Brillinger(1975). First we construct the best linear predictor $\hat{X}_t$ based on $g(\lambda)$ by

$$\hat{X}_t = \int_{-\pi}^{\pi} e^{it\lambda}\frac{c_g(e^{-i\lambda}) - c_g(0)}{c_g(e^{-i\lambda})}dz(\lambda). \tag{13}$$

Then we have the following proposition.

Proposition 5.

$$\begin{aligned} E[\{X_t^2 - \hat{X}_t^2\}^2] &= \mathrm{MSPE}_1 + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{c_g(0)}{c_g(e^{-i\lambda_1})}\frac{2c_g(e^{-i\lambda_2}) - c_g(0)}{c_g(e^{-i\lambda_2})}\frac{c_g(0)}{c_g(e^{-i\lambda_3})} \\ &\quad \times \frac{2c_g(e^{i(\lambda_1+\lambda_2+\lambda_3)}) - c_g(0)}{c_g(e^{i(\lambda_1+\lambda_2+\lambda_3)})}g_4(\lambda_1, \lambda_2, \lambda_3)d\lambda_1 d\lambda_2 d\lambda_3 \\ &= \mathrm{MSPE}_5, \text{ (say)}. \end{aligned}$$

Next we consider the misspecified prediction problem when $\{X_t\}$ is non-Gaussian. Denote a conjectured spectral density for $g(\lambda)$ by

$$f(\lambda) = \frac{1}{2\pi}|c_f(e^{-i\lambda})|^2. \tag{14}$$

The best linear predictor of X_t based on $f(\lambda)$ is given by

$$\hat{X}_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{c_f(e^{-i\lambda}) - c_f(0)}{c_f(e^{-i\lambda})} dz(\lambda). \tag{15}$$

The mean square prediction error of $\hat{X}_t^2$ for X_t^2 is then

PROPOSITION 6.

$$\begin{aligned} E[\{X_t^2 - \hat{X}_t^2\}^2] &= \mathrm{MSPE}_3 + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{c_f(0)}{c_f(e^{-i\lambda_1})} \frac{2c_f(e^{-i\lambda_2}) - c_f(0)}{c_f(e^{-i\lambda_2})} \frac{c_f(0)}{c_f(e^{-i\lambda_3})} \\ &\quad \times \frac{2c_f(e^{i(\lambda_1+\lambda_2+\lambda_3)}) - c_f(0)}{c_f(e^{i(\lambda_1+\lambda_2+\lambda_3)})} g_4(\lambda_1,\lambda_2,\lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 \\ &= \mathrm{MSPE}_6, \text{ (say).} \end{aligned}$$

4. PREDICTION PROBLEMS FOR ESTIMATED PREDICTOR

In actual situation we fit a parametric spectral density f_θ for the process concerned, where θ ia an unknown parameter. Then we estimate θ by some estimator $\hat{\theta}$. In this section we discuss the prediction problems for the predictors constructed from $f_{\hat{\theta}}$. Let $\{X_t\}$ be a general linear process generated by

$$X_t = \sum_{j=0}^{\infty} G(j) e_{t-j}, \qquad t \in J, \tag{16}$$

where $\{e_t\}$ is a sequence of independent identically distributed random variables with $E(e_t) = 0$ and $\mathrm{var}(e_t) = \sigma_e^2$ and $G(j)$'s satisfy

$$\sum_{j=0}^{\infty} G(j)^2 < \infty.$$

Then the process $\{X_t\}$ has the spectral density

$$g(\lambda) = \frac{\sigma_e^2}{2\pi} |\sum_{j=0}^{\infty} G(j) e^{i\lambda j}|^2. \tag{17}$$

For $g(\lambda)$, we fit a parametric specteal density $f_\theta(\lambda)$, $\theta = (\theta_1, \ldots, \theta_q)' \in \Theta \subset R^q$, where θ is an innovation-free parameter. Write the spectral representation of $\{X_t\}$ as

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} dz(\lambda). \tag{18}$$

Now we assume that the conjectured spectral density $f_\theta(\lambda)$ satisfies

$$\int_{-\pi}^{\pi} |\log f_\theta(\lambda)| d\lambda < \infty.$$

Writing

$$f_\theta(\lambda) = \frac{1}{2\pi}|c_\theta(e^{-i\lambda})|^2, \tag{19}$$

we can construct the best linear predictor of X_t based on $f_\theta(\lambda)$ by

$$\hat{X}_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{c_\theta(e^{-i\lambda}) - c_\theta(0)}{c_\theta(e^{-i\lambda})} dz(\lambda). \tag{20}$$

The following assumption is imposed.

Assumption 1. (i) The parameter θ is innovation-free.
(ii) f_θ is continuously three times differentiable with respect to $\theta \in \Theta$.

In Proposition 3, denote the mean square prediction error of $Y = X_t^2 - 1$ by $\mathrm{MSPE}_3(f_\theta)$ if the conjectured spectral density f is f_θ. Suppose that we have an observed stretch $X_1', \ldots, X_T'$ which has the same probability structure as $\{X_t\}$ and is independent of $\{X_t\}$. The unknown parameter θ is estimated by a quasi-MLE $\hat{\theta}_Q = (\hat{\theta}_{n,1}, \ldots, \hat{\theta}_{n,p})'$ which minimizes

$$\int_{-\pi}^{\pi} \{\log f_\theta(\lambda) + I_T(\lambda) f_\theta(\lambda)^{-1}\} d\lambda,$$

with respect to $\theta \in \Theta$, where

$$I_T(\lambda) = \frac{1}{2\pi T} |\sum_{t=1}^{T} X_t' e^{it\lambda}|^2.$$

Then the estimated predictor is given by

$$\hat{\hat{X}}_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{c_{\hat{\theta}_Q}(e^{-i\lambda}) - c_{\hat{\theta}_Q}(0)}{c_{\hat{\theta}_Q}(e^{-i\lambda})} dz(\lambda), \tag{21}$$

and the mean square prediction error of $\hat{\hat{X}}_t^2$ is

$$E[\{X_t^2 - \hat{\hat{X}}_t^2\}^2] = \mathrm{MSPE}_3(f_{\hat{\theta}_Q}).$$

Expanding $\mathrm{MSPE}_3(f_{\hat{\theta}_Q})$ at $\theta = \underline{\theta}$ in a Taylor series we obtain

$$\begin{aligned} &\mathrm{MSPE}_3(f_{\hat{\theta}_Q}) \\ &= \mathrm{MSPE}_3(f_{\underline{\theta}}) + \frac{\partial}{\partial \theta'} \mathrm{MSPE}_3(f_\theta)(\hat{\theta}_Q - \underline{\theta}) \\ &\quad + \frac{1}{2}\left\{(\hat{\theta}_Q - \underline{\theta})' \frac{\partial^2}{\partial \theta \partial \theta'} \mathrm{MSPE}_3(f_{\underline{\theta}})(\hat{\theta}_Q - \underline{\theta})\right\} + \text{lower order terms}, \end{aligned} \tag{22}$$

where $\underline{\theta} = \arg\min_{\theta\in\Theta}\int_{-\pi}^{\pi}[\log f_\theta(\lambda)+\{g(\lambda)f_\theta(\lambda)^{-1}\}]d\lambda$. To evaluate the second term of the right hand side of (22), we make the following assumption.

Assumption 2. For $\alpha = 1,\ldots,p$

$$E[\sqrt{T}(\hat{\theta}_{n,\alpha} - \theta_\alpha)] = T^{-1/2}\mu^\alpha + o(T^{-1}), \tag{23}$$

where μ^α's are constants.

This assumption is reasonable. To see this, let $\{X_t : t = 0,\pm1,\pm2,\cdots\}$ be a Gaussian autoregressive moving average(ARMA) process with spectral density $f_\theta(\lambda)$ which depends on an unknown parameter $\theta \in R^1$. Then it is known that

$$E_\theta[\sqrt{T}(\hat{\theta}_Q - \theta)] = -\frac{B(\theta)}{I(\theta)\sqrt{T}} - \frac{J(\theta)+K(\theta)}{2\{I(\theta)\}^2\sqrt{T}} + o\left(\frac{1}{\sqrt{T}}\right),$$

where

$$I(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta}\log f_\theta(\lambda)\right\}^2 d\lambda,$$

$$B(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta}f_\theta(\lambda)\right\} b_\theta(\lambda)\{f_\theta(\lambda)\}^{-2}d\lambda,$$

$$b_\theta(\lambda) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|\gamma(n)e^{in\lambda}, \quad \text{with } \gamma(n) = E_\theta X_t X_{t+n},$$

$$J(\theta) = -\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta}f_\theta(\lambda)\right\}^3\{f_\theta(\lambda)\}^{-3}d\lambda + \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial^2}{\partial\theta^2}f_\theta(\lambda)\right\}\left\{\frac{\partial}{\partial\theta}f_\theta(\lambda)\right\}\{f_\theta(\lambda)\}^{-2}d\lambda,$$

$$K(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta}f_\theta(\lambda)\right\}^3\{f_\theta(\lambda)\}^{-3}d\lambda,$$

(see Taniguchi(1983)).

Furthermore, Taniguchi and Watanabe(1994) evaluated $E[\sqrt{T}(\hat{\theta}_{n,\alpha}-\theta_\alpha)]$ in the form of (23) for generalized curved probability densities. To describe the asymptotics of $\hat{\theta}_Q$ we need the following regularity conditions.

Assumption 3. (A.1) The spectral density $g(\lambda)$ is square-integrable;
(A.2) $g(\lambda) \in \mathrm{Lip}(\alpha)$, $\alpha > 1/2$;
(A.3) The fourth-order moment of e_t exists;
(A.4) $\underline{\theta}$ is exists uniquely and $\underline{\theta} \in \mathrm{Int}\Theta$;
(A.5)The matrix

$$M_f = \int_{-\pi}^{\pi}\left[\frac{\partial^2}{\partial\theta\partial\theta'}f_\theta(\lambda)^{-1}g(\lambda) + \frac{\partial^2}{\partial\theta\partial\theta'}\log f_\theta(\lambda)\right]_{\underline{\theta}} d\lambda,$$

is nonsingular.

By means of (22) we evaluate $\overline{E}\{\mathrm{MSPE}_3(f_{\hat{\theta}_Q})\}$ where $\overline{E}\{\cdot\}$ is the expectation with respect to the asymptotic distribution of $\sqrt{T}(\hat{\theta}_Q - \underline{\theta})$.

PROPOSITION 7. *Under the Assumptions 1, 2 and 3,*

$$\begin{aligned}\overline{E}\{\mathrm{MSPE}_3(f_{\hat{\theta}_Q})\} &= \mathrm{MSPE}_3(f_{\underline{\theta}}) + \frac{1}{T}\sum_{\alpha=1}^{q}\mu^{\alpha}\left\{\frac{\partial}{\partial\theta_\alpha}\mathrm{MSPE}_3(f_{\underline{\theta}})\right\} \\ &\quad + \frac{1}{2T}\mathrm{tr}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\mathrm{MSPE}_3(f_{\underline{\theta}})M_f^{-1}\tilde{V}M_f^{-1}\right\} + o(T^{-1}),\end{aligned}$$

where

$$M_f = \int_{-\pi}^{\pi}\left[\frac{\partial^2}{\partial\theta\partial\theta'}f_\theta(\lambda)^{-1}g(\lambda) + \frac{\partial^2}{\partial\theta\partial\theta'}\log f_\theta(\lambda)\right]_{\underline{\theta}} d\lambda,$$

$$\begin{aligned}\tilde{V} &= 4\pi\int_{-\pi}^{\pi}\left[g(\lambda)\frac{\partial}{\partial\theta}\{f_\theta(\lambda)\}^{-1}g(\lambda)\frac{\partial}{\partial\theta}\{f_\theta(\lambda)\}^{-1}\right]_{\underline{\theta}} d\lambda \\ &\quad + 2\pi\sum_{rtuv=1}^{s}\int\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta}f_\theta^{(r,t)}(\lambda_1)\frac{\partial}{\partial\theta}f_\theta^{(u,v)}(\lambda_2)\right\}_{\underline{\theta}} \\ &\quad \times \tilde{Q}^X_{rtuv}(-\lambda_1,\lambda_2,-\lambda_2)d\lambda_1 d\lambda_2.\end{aligned}$$

Next, assuming that $\{X_t\}$ is a Gaussian process with spectral density $h(\lambda)$, we consider the misspecified prediction for the transformation $Y_t = X_t^2 - 1$. We write the spectral representation of Y_t as

$$Y_t = \int_{-\pi}^{\pi} e^{it\lambda}d\omega(\lambda). \tag{24}$$

The spectral density of $\{Y_t\}$ is given by $g(\lambda) = 2h^{*2}(\lambda)$. Denote a conjectured spectral density by

$$f_\theta(\lambda) = \frac{1}{2\pi}|a_\theta(e^{-i\lambda})|^2. \tag{25}$$

Based on $f_\theta(\lambda)$ we can construct the best linear predictor of Y_t by

$$\hat{Y}_t = \int_{-\pi}^{\pi} e^{it\lambda}\frac{a_\theta(e^{-i\lambda}) - a_\theta(0)}{a_\theta(e^{-i\lambda})}d\omega(\lambda). \tag{26}$$

Then the mean square error of the prediction is

$$\begin{aligned}E[\{Y_t - \hat{Y}_t\}^2] &= E\left[a(0)\int_{-\pi}^{\pi}e^{it\lambda}\frac{1}{a_\theta(e^{-i\lambda})}d\omega(\lambda)\right]^2 \\ &= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{g(\lambda)}{f_\theta(\lambda)}d\lambda \\ &= \mathrm{MSPE}(f_\theta),\ (\text{say}).\end{aligned} \tag{27}$$

Here θ is actually an unknown parameter. Suppose that we have an observed stretch $Y_1', \ldots, Y_T'$ which has the same probability structure as $\{Y_t\}$ and is independent of $\{Y_t\}$. Then θ is estimated by a quasi-MLE $\hat{\hat{\theta}}_Q$ which minimizes

$$\int_{-\pi}^{\pi}\{\log f_\theta(\lambda) + I_T(\lambda)f_\theta(\lambda)^{-1}\}d\lambda,$$

with respect to $\theta \in \Theta$, where

$$I_T(\lambda) = \frac{1}{2\pi T}|\sum_{t=1}^{T} Y_t' e^{it\lambda}|^2.$$

Then the estimated predictor is given by

$$\hat{Y}_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{a_{\hat{\theta}_Q}(e^{-i\lambda}) - a_{\hat{\theta}_Q}(0)}{a_{\hat{\theta}_Q}(e^{-i\lambda})} d\omega(\lambda), \tag{28}$$

and the prediction error is

$$E[\{Y_t - \hat{\hat{Y}}_t\}^2] = \mathrm{MSPE}(f_{\hat{\hat{\theta}}_Q}). \tag{29}$$

Expanding $\mathrm{MSPE}(f_{\hat{\hat{\theta}}_Q})$ at $\theta = \underline{\theta}$ in a Taylor series we obtain

$$\begin{aligned} \mathrm{MSPE}(f_{\hat{\hat{\theta}}_Q}) &= \mathrm{MSPE}(f_{\underline{\theta}}) + \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\mathrm{MSPE}(f_{\underline{\theta}})(\hat{\hat{\theta}}_Q - \underline{\theta})(\hat{\hat{\theta}}_Q - \underline{\theta})'\right\} \\ &\quad + \text{lower order terms.} \end{aligned} \tag{30}$$

Here it should be noted that

$$\frac{\partial}{\partial\theta}\mathrm{MSPE}(f_\theta)\Big|_{\underline{\theta}} = 0,$$

because θ is an innovation-free parameter.

PROPOSITION 8. *Under the Assumptions 1 and 2,*

$$\overline{E}\{\mathrm{MSPE}(f_{\hat{\hat{\theta}}_Q})\} = \mathrm{MSPE}(f_{\underline{\theta}}) + \frac{1}{2T}\mathrm{tr}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\mathrm{MSPE}(f_{\underline{\theta}})M_f^{-1}\tilde{V}M_f^{-1}\right\} + o(T^{-1}).$$

REFERENCES

BRILLINGER, D. R. (1975). *Time Series; Data Analysis and Theory.* New York: Holt.

CHOI, I. B. & TANIGUCHI, M. (2000). Misspecified prediction for time series. *J. Forecasting.* To appear.

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica.* **50**, 987-1007.

GRANGER, C. W. J. & NEWBOLD, P. (1976). Forecasting transformed series. *J. Roy. Stat. Soc.* B **38**, 189-203.

GRANGER, C. W. J. & NEWBOLD, P. (1977). *Forecasting Economic Time Series.* New York: Academic Press.

GRENANDER, U. & ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time*

Series. New York: Wiley.

HANNAN, E. J. (1970). *Multiple Time Series.* New York: Wiley.

NASSIUMA, D. K. & ORDOUKHANI, N. (1991). Forecasting of some transformed series. *Commun. Statist. -Theory Meth.* **20(10)**, 3207-3220.

TANIGUCHI, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.* **11**, 157-169.

TNIGUCHI, T. & WATANABE, Y. (1994). Statistical analysis of curved probability densities. *J. Multivariate. Anal.* **48**, 228-248.

TANIGUCHI, M. & KAKIZAWA, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* New York: Springer-Verlag.

Third-Order Asymptotic Expansions of the Power of Test Statistics for Mixing Processes with Applications to Diffusion Models

阪本雄二*

名古屋大学工学研究科

Abstract

In the case where the observation is a mixing process with a continuous time parameter, we will obtain asymptotic expansions of the local power for test statistics included in a large class in the multi-parameter setting. Its application to the diffusion process will be also presented.

1 Mixing process with a Markovian property

Let Θ be an open convex subset in $\mathbb{R}^p$, and for any θ, let $Y^\theta = (Y^\theta(t))_{t\in\mathbb{R}_+}$ be an $\mathbb{R}^d$-valued càdlàg process defined on a probability space $(\Omega, \mathscr{F}, P)$. Suppose that for any $T > 0$, the probability measure induced by $Y_T^\theta = (Y^\theta(t))_{t\in[0,T]}$ has a positive density $p_{T,\theta}$ with respect to some reference measure. For an unknown $\vartheta \in \Theta$, the observation is assumed to be a realization of Y_T^ϑ. In this article, we will consider asymptotic expansions of test statistics, based on the log-likelihood $\ell_T(\theta; Y_T^\vartheta) = \log p_{T,\theta}(Y_T^\vartheta)$, for the hypothesis $\vartheta = \vartheta_0$ against $\vartheta \neq \vartheta_0$. For simplicity, we will often omit ϑ from the function of θ evaluated at $\theta = \vartheta$, e.g., $Y(t) = Y^\vartheta(t)$. We will also omit the phrase "for any ϑ" if there is no confusion.

Let us describe the probabilistic structure of the observation Y and the log-likelihood function $\ell_T(\theta; Y_T)$. Fix $\epsilon \geq 0$. Suppose that for any $\vartheta \in \Theta$,

[A1] there exists a positive constant a such that

$$E\left[\left|E[f\ |\mathscr{B}^Y_{[s-\epsilon,s]}] - E[f]\right|\right] \leq \frac{1}{a}e^{-a(t-s)} \sup_{\omega\in\Omega} |f(\omega)|$$

for any s, $t \in \mathbb{R}_+$, $s \leq t$, and for any bounded function $f \in \mathscr{F}\mathscr{B}^Y_{[t,\infty)}$, where $\mathscr{B}^Y_I = \sigma(Y(t) : t \in I)$ for any interval $I \subset \mathbb{R}_+$, and $\mathscr{F}\mathscr{B}$ denote the total of mesurable functions with respect to $\mathscr{B}$ for any sub σ-field $\mathscr{B}$.

Such a process Y is referred to as a (geometric) mixing process. In addition, Y is supposed to have a Markovian property as follows. Let $X = (X(t))_{t\in\mathbb{R}_+}$ be an $\mathbb{R}^r$-valued càdlàg process with independent increments, i.e., $\mathscr{B}^{X,Y}_{[0,t]}$ is independent of $\mathscr{B}^{dX}_{[t,\infty)}$ for any $t \in \mathbb{R}_+$, where $\mathscr{B}^{X,Y}_I = \sigma(X(t), Y(t) : t \in I)$ and $\mathscr{B}^{dX}_I = \sigma(X(t) - X(s) : t, s \in I)$. Note that if X depends on θ, then it is supposed to be evaluated at $\theta = \vartheta$. Suppose that Y is an ϵ-Markov process driven by X, i.e., $Y(t) \in \mathscr{F}(\mathscr{B}^Y_{[s-\epsilon,s]} \vee \mathscr{B}^{dX}_{[s,t]})$ for any $s > 0$ and $t > 0$ with $\epsilon \leq s < t$. Moreover, assume that for any $t > 0$, the log-likelihood function $(\ell_t(\theta; Y_t))_{t\in\mathbb{R}_+}$ has all of the third-order derivatives with respect to θ, and for any $\theta \in \Theta$ let $Z^\theta = (Z^\theta(t))_{t\in\mathbb{R}_+}$ be a $\mathbb{R}^{p+p(p+1)/2+p(p+1)(p+2)/6}$-valued stochastic process defined by

$$Z^\theta(t) = (\ell_{t,a}(\theta; Y_t), \ell_{t,ab}(\theta; Y_t), \ell_{t,abc}(\theta; Y_t))_{a,b,c=1,\dots,p},$$

where $\ell_{t,a} = \delta_a\ell_t$, $\ell_{t,ab} = \delta_a\delta_b\ell_t$, $\ell_{t,abc} = \delta_a\delta_b\delta_c\ell_t$, $\delta_a = \partial/\partial\theta^a$ for any $a = 1, \dots, p$. For this process Z^θ, suppose that $Z^\theta(t) - Z^\theta(s) \in \mathscr{F}\mathscr{B}^{dX,Y}_{[s,t]}$ for any $\theta \in \Theta$ and $0 \leq s \leq t$, where $\mathscr{B}^{dX,Y}_I = \sigma(X(t) - X(s), Y(t) : s, t \in I)$ and that $Z^\theta(0) \in \mathscr{F}\mathscr{B}^Y_{[0]}$. Such a process Z is called additive functional of X and Y.

For example, X is the the standard Wiener process $w = (w(t))_{t\in\mathbb{R}_+}$ and Y^θ is the stationary diffusion process satisfying

$$dY(t) = V_0(Y(t), \theta)dt + V(Y(t))dw(t), \tag{1.1}$$

for some functions $V_0 : \mathbb{R}^d \times \Theta \to \mathbb{R}^d$ and $V : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^r$. It is clear that for any $\theta \in \Theta$, Y^θ is an ϵ-Markov process with $\epsilon = 0$, and for the geometric mixing property of the diffusion process we can use the sufficient conditions given by Strook(1994), Roberts-Tweedie(1996) and Kusuoka-Yoshida(2000). Furthermore, under

*Fax:81-52-789-3724. E-mail: yuji@na.cse.nagoya-u.ac.jp

some regularity condition (the Novikov condition), the probability measure induced of Y_T^θ has a density and the log-likelihood function is given by

$$\ell(\theta; Y_T) = \log \frac{d\nu_\theta}{dx}(Y(0)) + \int_0^T V_0'(VV')^{-1}(Y(t),\theta)dY(t) - \frac{1}{2}\int_0^T V_0'(VV')^{-1}V_0(Y(t),\theta)dt; \tag{1.2}$$

where V' is the transpose of V, and ν_θ is the probability measure of the stationary distribution. Therefore, the additivity of the derivatives of the log-likelihood function is obvious.

Another example is a non-linear time series $\tilde{Y}^\theta = (\tilde{Y}_n^\theta)_{n=1,2,\dots}$ defined by

$$\tilde{Y}_n^\theta = f_n(\tilde{Y}_{n-1}^\theta, \dots, \tilde{Y}_{n-k}^\theta; \theta) + \xi_n$$

for an $\mathbb{R}^r$-valued i.i.d. sequence $\xi = (\xi_n)$ and some function $f_n : \mathbb{R}^{kd} \times \Theta \to \mathbb{R}^d$. Obviously, it is an ϵ-Markov process with $\epsilon = k - 1$, and under some conditions it has a geometric mixing property(see Tweedie(1983), Tong(1990)). The Markovian property leads to the additivity of the derivatives of the log-likelihood. These processes Y and ξ are those with discrete time parameter, but they can be embedded into continuous-time-parameter process Y and X satisfying the above properties.

Under these settings, the asymptotic expansions of estimators were given by Kusuoka-Yoshida(2000) and Sakamoto-Yoshida(1999), and the information criteria was discussed by Uchida-Yoshida(1999).

2 A class of test statistics

In this section, we will introduce a large class of test statistics, based on the log-likelihood, for the hypothesis $\vartheta = \vartheta_0$ against $\vartheta \neq \vartheta_0$.

In order to define the class, let us prepare some notations. For any a, $b = 1, \dots, p$, and any index sets A, A_1, $\dots$, A_j whose elements run from 1 to p,

$$\begin{aligned}
\bar{Z}_A^\theta =& \frac{1}{\sqrt{T}}(\ell_{T,A}(\theta; Y_T) - E[\ell_{T,A}(\theta; Y_T)]),\\
g_{ab}(\theta) =& \frac{1}{T}E[\ell_{T,a}(\theta; Y_T)\ell_{T,b}(\theta; Y_T)]\\
\bar{\nu}_A(\theta) =& \frac{1}{T}E[\ell_{T,A}(\theta; Y_T)], \quad \bar{\nu}_{A_1,\dots,A_j}(\theta) = \frac{1}{T}E[\ell_{T,A_1}(\theta; Y_T)\cdots\ell_{T,A_j}(\theta; Y_T)]
\end{aligned}$$

As mentioned before, $\bar{Z}_A$, g_{ab} and $\bar{\nu}_{A_1,\dots,A_j}$ denote $\bar{Z}_A^\vartheta$, $g_{ab}(\vartheta)$ and $\bar{\nu}_{A_1,\dots,A_j}(\vartheta)$, respectively. Note that $\bar{Z}$, g and $\bar{\nu}$ depend on T, but we will hereafter write them without T. Suppose that $\bar{\nu}_a = 0$ for any $a = 1, \dots, p$ and that g_{ab} is non-singular for sufficiently large T. Let $(g^{ab}) = (g_{ab})^{-1}$ and $g^{A,B} = g^{a_1b_1}\cdots g^{a_jb_j}$ for any index sets $A = a_1\cdots a_j$ and $B = b_1\cdots b_j$. The orthogonalized random variables $\tilde{Z}_a$ and $\tilde{Z}_A$ for any index set A whose length is greater than 2 is defined by $\tilde{Z}_a = \bar{Z}_a$, $\tilde{Z}_A = \bar{Z}_A - \bar{\nu}_{A,b}g^{ba}\bar{Z}_a$. The contravariant representations of $\bar{\nu}_{A_1,\dots,A_j}$, Z_A and $\tilde{Z}_A$ are denoted by $\bar{\nu}^{A_1,\dots,A_j} = g^{A_1,B_1}\cdots g^{A_j,B_j}\bar{\nu}_{B_1,\dots,B_j}$, $\bar{Z}^A = g^{A,B}\bar{Z}_B$ and $\tilde{Z}^A = g^{A,B}\tilde{Z}_B$.

By using these notations, we define a class $\mathscr{T}$ as the total of test statistics V_T admitting a stochastic expansion $V_T = S_T + R_T$, where

$$\begin{aligned}
S_T =& g_{ab}\tilde{Z}^a\tilde{Z}^b + \frac{1}{\sqrt{T}}\left(a_{1;cd}^{ab}\tilde{Z}_a\tilde{Z}_b\tilde{Z}^{cd} + a_2^{abc}\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c\right)\\
&+ \frac{1}{T}\left(b_1^{ab}\tilde{Z}_a\tilde{Z}_b + b_{2;cd,ef}^{ab}\tilde{Z}_a\tilde{Z}_b\tilde{Z}^{cd}\tilde{Z}^{ef} + b_3^{abcd}\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c\tilde{Z}_d + b_{4;de}^{abc}\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c\tilde{Z}^{de}\right.\\
&\left.+ b_{5;def}^{abc}\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c\tilde{Z}^{def} + b_6^{abcdef}\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c\tilde{Z}_d\tilde{Z}_e\tilde{Z}_f\right)
\end{aligned}$$

for some constants $a_{1;cd}^{ab}$, a_2^{abc}, b_1^{ab}, $+b_{2;cd,ef}^{ab}$, b_3^{abcd}, $b_{4;de}^{abc}$, $b_{5;def}^{abc}$, b_6^{abcdef}, and R_T is a random variable satisfying

$$P_{\vartheta_0}\{|R_T| \geq CT^{(2+\epsilon)/2}\} = o(T^{-1})$$

for some $C > 0$ and $\epsilon > 0$. This class $\mathscr{T}$ includes many statistics, e.g. the likelihood-ratio, the Wald, the modified Wald, and the Rao's score statistic as is shown below.

Note that in the case where the observation is a time series and the dimension p of the parameter space Θ is one, this class was introduced by Rao-Mukerjee(1997), while its origin was given in Taniguchi(1991). They obtained formal asymptotic expansions for the class, made some comparisons of the local powers, and gave some applications to i.i.d. observations and linear time series.

Example 1. *(Modified Wald statistic)* Suppose that there exists the maximum likelihood estimator $\hat{\vartheta}_T$ for ϑ. Then the modified Wald statistic MW_T is defined by $g_{ab}(\hat{\vartheta}_T - \vartheta_0)^a(\hat{\vartheta}_T - \vartheta_0)^b$. Under some condition for smoothness of log-likelihood function ℓ, it is easy to show that it has the stochastic expansion

$$\begin{aligned}\mathrm{MW}_T =& g^{ab}\tilde{Z}_a\tilde{Z}_b\blacktriangledown + \frac{1}{\sqrt{T}}((\bar{\nu}^{abc} + 2\bar{\nu}^{ab,c})\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c + 2g^{ab}g^{cd}\tilde{Z}_{bc}\tilde{Z}_a\tilde{Z}_d)\\ &\blacktriangledown + \frac{1}{T}\left(\frac{1}{3}\bar{\nu}^{abcd} + \frac{5}{4}\bar{\nu}^{abe}g_{ef}\bar{\nu}^{fcd} + 4\bar{\nu}^{abe}g_{ef}\bar{\nu}^{fc,d} + \bar{\nu}^{abc,d} + 3\bar{\nu}^{ae,b}g_{ef}\bar{\nu}^{cf,d}\right)\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c\tilde{Z}_d\\ &+ \frac{1}{T}(4\bar{\nu}^{abc}g^{de} + 6\bar{\nu}^{ab,c}g^{de})\tilde{Z}_{ad}\tilde{Z}_b\tilde{Z}_c\tilde{Z}_e + \frac{1}{T}g^{ab}g^{cd}g^{ef}\tilde{Z}_{ace}\tilde{Z}_b\tilde{Z}_d\tilde{Z}_f\\ &+ \frac{3}{T}g^{ab}g^{cd}g^{ef}\tilde{Z}_{ac}\tilde{Z}_{be}\tilde{Z}_d\tilde{Z}_f + R_{MW,3},\end{aligned}$$

where $R_{MW,3}$ is some remainder term. When $\sup_\theta E_{\vartheta_0}|\tilde{Z}_A^\theta|^p < \infty$ for some $p > 1$, we can estimate the rate of convergence of $R_{MW,3}$, and then we see that MW_T belongs to the class above. If the Fisher information $g_{ab}(\theta)$ is smooth, the Wald statistic $g_{ab}(\hat{\vartheta}_T)(\hat{\vartheta}_T - \vartheta_0)^a(\hat{\vartheta}_T - \vartheta_0)^b$ can be expanded in the same form, and it turns out that the Wald statistic also belongs to the class $\mathscr{T}$ under the moment condition for $Z_A(\theta)$.

Example 2. *(Likelihood Ratio statistic)* Suppose that there exists the maximum likelihood estimator $\hat{\vartheta}_T$ for ϑ. Then the likelihood ratio statistics LR_T is defined by $2(\ell_T(\hat{\vartheta}_T; Y_T) - \ell_T(\vartheta_0; Y_T))$. If the Bartlett identity $\bar{\nu}_{a,b} + \bar{\nu}_{ab} = 0$ holds true, it is easy to show that under some condition for smoothness of log-likelihood function ℓ, it can be expanded as follows.

$$\begin{aligned}\mathrm{LR}_T =& g^{ab}\tilde{Z}_a\tilde{Z}_b\blacktriangledown + \frac{1}{3\sqrt{T}}((\bar{\nu}^{abc} + 3\bar{\nu}^{ab,c})\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c + 3g^{ab}g^{cd}\tilde{Z}_a\tilde{Z}_c\tilde{Z}_{bd})\\ &\blacktriangledown + \frac{1}{12T}(\bar{\nu}^{abcd} + 3\bar{\nu}^{abe}g_{ef}\bar{\nu}^{fcd} + 12\bar{\nu}^{abe}g_{ef}\bar{\nu}^{fc,d} + 12\bar{\nu}^{a,be}g_{ef}\bar{\nu}^{fc,d} + 4\bar{\nu}^{abc,d})\tilde{Z}_a\tilde{Z}_b\tilde{Z}_c\tilde{Z}_d\\ &+ \frac{1}{T}g^{ab}(\bar{\nu}^{cde} + 2\bar{\nu}^{cd,e})\tilde{Z}_{ac}\tilde{Z}_b\tilde{Z}_d\tilde{Z}_e + \frac{1}{T}g^{ab}g^{cd}g^{ef}\tilde{Z}_{ac}\tilde{Z}_{be}\tilde{Z}_d\tilde{Z}_f + \frac{1}{3T}g^{ab}g^{cd}g^{ef}\tilde{Z}_{ace}\tilde{Z}_b\tilde{Z}_d\tilde{Z}_f + R_{LR,3},\end{aligned}$$

where $R_{LR,3}$ is some remainder term. It also belongs to the class of test statistics under the moment condition of $\tilde{Z}_A^\theta$.

3 Asymptotic expansion of the local power

For any constant $\varepsilon \in \mathbb{R}^p$, we consider the test based on V_T for the null-hypothesis $\vartheta = \vartheta_0$ against $\vartheta = \vartheta_\varepsilon (:= \vartheta_0 + \varepsilon/\sqrt{T})$. In order to obtain the asymptotic expansion, we assume the moment condition of the process $Z^\theta = (Z^\theta(t))$:

[A2] For any $\theta \in \Theta$ and $\Delta > 0$, $\sup_{t\in\mathbb{R}_+, 0\le h\le\Delta} E\left|Z^\theta(t+h) - Z^\theta(t)\right|^p < \infty$ for any $p > 1$, and $E[Z^\theta(t+\Delta)] = E[Z^\theta(t)]$. Moreover, $E|Z^\theta(0)|^p < \infty$ for any $p > 1$ and $E[Z^\theta(0)] = 0$.

In addition, we assume the regularity condition [A3] in Kusuoka-Yoshida(2000) and Sakamoto-Yoshida(1999). It ensures the regularity (or continuity) of the distribution of $(Z(t))$, and is the crucial one for the validity of the asymptotic expansion, but we here omit to state it because it is necessary to prepare many notations in terms of the Malliavin calculous.

For the representation of the asymptotic expansion, the following notations are necessary. Let $K_{abc} = E[Z_aZ_bZ_c]$, $K^{abc} = g^{aa'}g^{bb'}g^{cc'}K_{a'b'c'}$, $J_{ab,c} = E[Z_{ab}Z_c]$, $J^{ab,c} = g^{aa'}g^{bb'}g^{cc'}J_{a'b',c'}$ $M_{ab,cd} = E[Z_{ab}Z_{cd}]$, $M^{ab,cd} = g^{aa'}g^{bb'}g^{cc'}g^{dd'}M_{a'b',c'd'}$, $\tilde{M}^{ab,cd} = M^{ab,cd} - J^{ab,e}g_{ef}J^{f,cd}$, $N_{ab,c,d} = E[Z_{ab}Z_cZ_d]$, $L_{abc,d} = E[Z_{abc}Z_d]$, $\tilde{N}^{a,b,cd} = N^{a,b,cd} - J^{cd,f}g_{ef}K^{abe}$, $H_{abcd} = \mathrm{Cum}[Z_a, Z_b, Z_c, Z_d]$, and $H^{abcd} = g^{aa'}g^{bb'}g^{cc'}g^{dd'}H_{a'b'c'd'}$. Moreover, let

$$g^{(\varepsilon)}_{a_1\cdots a_k,j} = \sum_{(a_1,\cdots,a_{k-2j},a_{k-2j+1}a_{k-2j+2},\cdots,a_{k-1}a_k)} g_{a_1a'_1}\cdots g_{a_{k-2j}a'_{k-2j}}\varepsilon^{a'_1}\cdots\varepsilon^{a'_{k-2j}}g_{a_{k-2j+1}a_{k-2j+2}}\cdots g_{a_{k-1}a_k},$$

$$\tilde{g}^{(\varepsilon)}_{a_1\cdots a_k,j} = \sum_{\substack{s+l=j \\ 0\le l\le s\le [k/2]}} \binom{s}{l}(-1)^l g^{(\varepsilon)}_{a_1\cdots a_k,s},$$

$$H^{\varepsilon,g}_{a_1\cdots a_k,p}(x) = \sum_{j=0}^{[k/2]} g^{(\varepsilon)}_{a_1\cdots a_k,j}\chi_{p+2k-2j,\delta}(x),$$

and

$$\tilde{H}^{\varepsilon,g}_{a_1\cdots a_k,p}(x) = \sum_{j=0}^{2[k/2]} \tilde{g}^{(\varepsilon)}_{a_1\cdots a_k,j}\chi_{p+2k-2j,\delta}(x),$$

where $\delta = g_{ab}\varepsilon^a\varepsilon^b$, $\chi_{p,\delta}(x)$ is the non-central χ^2 distribution function of degree p with non-central parameter δ.

By using these notations, we obtain the valid asymptotic expansion of the power function for any test statistics V_T in the class $\mathscr{T}$.

Theorem 1.

$$P_{\theta_\varepsilon}[V_T < x] = \chi_{p,\delta}(x) + \frac{1}{\sqrt{T}}U_1(x) + \frac{1}{T}(U_2^{(1)}(x) - \partial_x U_2^{(2)}(x) - \partial_x U_2^{(3)}(x) + \frac{1}{2}\partial_x^2 U_2^{(4)}(x)) + o(\frac{1}{T})$$

where

$$U_1(x) = \frac{1}{6}K^{abc}\tilde{H}^{\varepsilon,g}_{abc,p}(x) + \frac{1}{2}J_{ab,c}\varepsilon^{ab}g^{cd}\tilde{H}^{\varepsilon,g}_{d,p}(x) + \frac{1}{6}\bar{\nu}_{abc}\varepsilon^{abc}\chi_{p,\delta}(x) - a_2^{abc}\partial_x H^{\varepsilon,g}_{abc,p}(x)$$

$$\begin{aligned}U_2^{(1)}(x) &= \frac{1}{24}H^{abcd}\tilde{H}^{\varepsilon,g}_{abcd,p}(x) + \frac{1}{72}K^{abc}K^{def}\tilde{H}^{\varepsilon,g}_{abcdef,p}(x) + \frac{1}{12}K^{abc}g^{de}\varepsilon^{fg}J_{fg,e}(\tilde{H}^{\varepsilon,g}_{abcd,p}(x) + 3g_{ad}\tilde{H}^{\varepsilon,g}_{bc,p}(x))\\ &+ \frac{1}{36}K^{abc}\varepsilon^{def}\bar{\nu}_{def}\tilde{H}^{\varepsilon,g}_{abc,p}(x) + \frac{1}{4}\tilde{N}^{c,d,ef}g_{ae}g_{bf}\varepsilon^{ab}\tilde{H}^{\varepsilon,g}_{cd,p}(x)\\ &+ \frac{1}{24}\varepsilon^{abcd}(3\tilde{M}_{ab,cd}\chi_{p,\delta}(x) + 4L_{abc,d}\chi_{p+2,\delta}(x) + \bar{\nu}_{abcd}\chi_{p,\delta}(x) + 3J_{ab,e}J_{cd,f}g^{ee'}g^{ff'}H^{\varepsilon,g}_{e'f',p}(x))\\ &+ \frac{1}{72}\bar{\nu}_{abc}\varepsilon^{abcdef}(6J_{de,f}\chi_{p+2,\delta}(x) + \bar{\nu}_{def}\chi_{p,\delta}(x))\end{aligned}$$

$$\begin{aligned}U_2^{(2)}(x) &= \frac{1}{12}a_2^{abc}K^{def}(H^{\varepsilon,g}_{abcdef,p}(x) - 3g_{ef}H^{\varepsilon,g}_{abcd,p}(x)) + \frac{1}{4}a^{ab}_{1;cd}\tilde{N}^{cd,e,f}(H^{\varepsilon,g}_{abef,p}(x) - g_{ef}H^{\varepsilon,g}_{ab,p}(x))\\ &+ \frac{1}{4}a^{ab}_{1,cd}g_{ee'}g_{ff'}\varepsilon^{ef}\tilde{M}^{cd,e'f'}H^{\varepsilon,g}_{ab,p}(x) + \frac{1}{4}a_2^{abc}J_{ef,d'}g^{dd'}\varepsilon^{ef}H^{\varepsilon,g}_{abcd,p}(x) + \frac{1}{6}a_2^{abc}\bar{\nu}_{def}\varepsilon^{def}H^{\varepsilon,g}_{abc,p}(x)\end{aligned}$$

$$U_2^{(3)}(x) = b_1^{ab}H^{\varepsilon,g}_{ab,p}(x) + b^{ab}_{2;cd,ef}\tilde{M}^{cd,ef}H^{\varepsilon,g}_{ab,p}(x) + b_3^{abcd}H^{\varepsilon,g}_{abcd,p}(x) + b_6^{abcdef}H^{\varepsilon,g}_{abcdef,p}(x)$$

$$U_2^{(4)}(x) = a^{ab}_{1;a'b'}a^{cd}_{1;c'd'}\tilde{M}^{a'b',c'd'}H^{\varepsilon,g}_{abcd,p}(x) + a_2^{abc}a_2^{def}H^{\varepsilon,g}_{abcdef,p}(x).$$

4 Diffusion process

Let $Y = (Y(t))_{t\in\mathbb{R}_+}$ be a d-dimensional stationary diffusion process defined by (1.1) with $V_0 = (V_0^i)_{i=1,\cdots,d} \in C^\infty_\uparrow(\mathbb{R}^d\times\Theta;\mathbb{R}^d)$, $V = (V_j^i)_{i=1,\cdots,d,j=1,\cdots,r} \in C^\infty_\uparrow(\mathbb{R}^d;\mathbb{R}^d\otimes\mathbb{R}^r)$. Assume that

(i) for any $\theta\in\Theta$ ν_θ has a density with respect to the Lebesgue measure, and Y_T also has the density $p_{T,\theta}(Y_T) = e^{\ell(\theta;Y_T)}$, where $\ell(\theta;Y_T)$ is given by (1.2), with respect to some reference measure.

(ii) for any $p\ge$ and $t\in\mathbb{R}_+$, $E|Y(t)|^p < \infty$

(iii) Y has a geometric strong mixing property[A1];

(iv) $(Z(t))$ satisfy the condition[C3] in Kusuoka-Yoshida(2000).

Let $b = V'(VV')^{-1}V_0$ and $b_{a_1\cdots a_k} = \delta_{a_1}\cdots\delta_{a_k}b$, where $\delta_a = \partial/\partial\theta^a$. For functions f satisfying $\nu(f) = 0$, denote by G_f the Green function such that

$$\mathcal{A}G_f = f,$$

where $V_0 = (V_0^i)$, $V = (V_j^i)$ and

$$\mathcal{A} = \sum_{i=1}^{d} V_0^i \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{i,j}^{d} \sum_{k=1}^{r} V_k^i V_k^j \frac{\partial^2}{\partial x^i x^j}.$$

The $\mathbb{R}^r$-valued function $[f]$ is defined by

$$[f] = -V' \nabla G_{f-\nu(f)},$$

where $\nabla = (\partial/\partial x^1, \ldots, \partial/\partial x^d)'$. For index sets $A = a_1 \cdots a_j$, $B = b_1 \cdots b_k$, $C = c_1 \cdots c_l$, $D = d_1 \cdots d_m$, let

$$F_{A,B} = \nu(b_A \cdot b_B), \qquad F_{[A,B],C} = \nu([b_A \cdot b_B] \cdot b_C),$$

$$F_{[[A,B],C],D} = \nu([[b_A \cdot b_B] \cdot b_C] \cdot b_D), \qquad F_{[A,B],[C,D]} = \nu([b_A \cdot b_B] \cdot [b_C \cdot b_D]).$$

Moreover, let $\rho_{ab} = F_{a,b}$, $(\rho^{ab}) = (\rho_{ab})^{-1}$, $\rho^{[3]}_{abcd} = \sum^{[3]}_{(ab,cd)} \rho_{ab}\rho_{cd}$, $\rho^{[15]}_{abcdef} = \sum^{[15]}_{(ab,cd,ef)} \rho_{ab}\rho_{cd}\rho_{ef}$, $J^*_{ab,c} = F_{ab,c} - F_{[a,b],c}$, $J^{*ab,c} = \rho^{aa'}\rho^{bb'}\rho^{cc'}J^*_{ab,c}$

$$K^{*abc} = \rho^{aa'}\rho^{bb'}\rho^{cc'} \sum_{(a'b',c')} F_{[a',b'],c'}, \quad K^{**} = \sum^{[15]}_{(ab,cd,ef)} \rho^{ab}\rho^{cd}\rho^{ef} \sum_{(ab,c)} F_{[a,b],c} \sum_{(de,f)} F_{[d,e],f}$$

$$M^{*ab,cd} = \rho^{aa'}\rho^{bb'}\rho^{cc'}\rho^{dd'}\Big(F_{[a'b',c'd']} - F_{[a',b'],c'd'} - F_{[c',d'],a'b'} + F_{[a',b'],[c',d']}\Big),$$

$$N^{*a,b,cd} = \rho^{aa'}\rho^{bb'}\rho^{cc'}\rho^{dd'}\Big(\sum^{[3]}_{(a'b',\{c'd'\})} F_{[a',b'],c'd'} - F_{[[c',d'],a'],b'} - F_{[[c',d'],b'],a'} - F_{[a',b'],[c',d']}\Big),$$

$$\tilde{M}^{*ab,cd} = M^{*ab,cd} - J^{*ab,e}\rho_{ef}J^{*f,cd}, \quad \tilde{N}^{*a,b,cd} = N^{*a,b,cd} - J^{*cd,f}\rho_{ef}K^{*abe}$$

and

$$H^* = \sum^{[3]}_{(ab,cd)} \rho^{ab}\rho^{cd}\Big(\sum^{[6]}_{(ab,c,d)} (F_{[[a,b],c],d} + F_{[[a,b],d],c}) + \sum^{[3]}_{(ab,cd)} F_{[a,b],[c,d]}\Big).$$

Using these notations, we can obtain the asymptotic expansion of the local power for the diffusion process Y. For simplicity, we here present the asymptotic expansion under the null-hypothesis ($\epsilon = 0$).

Theorem 2.

$$P[V_T < x] = \int_0^x q_{T,2}(z)dz + o(T^{-1}), \tag{4.1}$$

where $q_{T,2}(z) = g_p(z) + T^{-1}\sum_{i=0}^{3} C_i^* g_{p+2i}(z)$, *and*

$$\begin{aligned}
C_0^* =& \frac{1}{24}H^* - \frac{1}{72}K^{**} + \frac{1}{4}a^{ab}_{1;ef}\tilde{N}^{*c,d,ef}\rho_{ab}\rho_{cd} - \frac{1}{2}(b_1^{ab} + b^{ab}_{2;cd,ef}\tilde{M}^{*cd,ef})\rho_{ab} \\
&+ \frac{1}{8}a^{ab}_{1;ef}a^{cd}_{1;gh}\tilde{M}^{*ef,gh}\rho^{[3]}_{abcd}, \\
C_1^* =& -\frac{1}{12}H^* + \frac{1}{24}K^{**} - \frac{1}{4}a^{ab}_{1;ef}\tilde{N}^{*c,d,ef}(\rho_{ab}\rho_{cd} + \rho^{[3]}_{abcd}) + \frac{1}{4}a_2^{abc}K^{*def}\rho_{ef}\rho^{[3]}_{abcd} \\
&+ \frac{1}{2}(b_1^{ab}\rho_{ab} + b^{ab}_{2;cd,ef}\tilde{M}^{*cd,ef}\rho_{ab} - b_3^{abcd}\rho^{[3]}_{abcd}) + \frac{1}{8}(a_2^{abc}a_2^{def}\rho^{[15]}_{abcdef} - 2a^{ab}_{1;ef}a^{cd}_{1;gh}\tilde{M}^{*ef,gh}\rho^{[3]}_{abcd}), \\
C_2^* =& \frac{1}{24}H^* - \frac{1}{24}K^{**} - \frac{1}{12}a_2^{abc}K^{*def}(3\rho^{[3]}_{abcd}\rho_{ef} + \rho^{[15]}_{abcdef}) + \frac{1}{4}a^{ab}_{1;ef}\tilde{N}^{*c,d,ef}\rho^{[3]}_{abcd} \\
&+ \frac{1}{2}(b_3^{abcd}\rho^{[3]}_{abcd} - b_6^{abcdef}\rho^{[15]}_{abcdef}) + \frac{1}{8}(-2a_2^{abc}a_2^{def}\rho^{[15]}_{abcdef} + a^{ab}_{1;ef}a^{cd}_{1;gh}\tilde{M}^{*ef,gh}\rho^{[3]}_{abcd}), \\
C_3^* =& \frac{1}{72}K^{**} + \frac{1}{12}a_2^{abc}K^{def}\rho^{[15]}_{abcdef} + \frac{1}{2}b_6^{abcdef}\rho^{[15]}_{abcdef} + \frac{1}{8}a_2^{abc}a_2^{def}\rho^{[15]}_{abcdef}
\end{aligned}$$

Example 3*(Mean reverse Ornstein-Uhlenbeck process)* Consider the stationary diffusion process Y satisfying

$$dY(t) = (\theta_1 - \theta_2 Y(t))dt + \sigma dw(t),$$

where θ_1 are θ_2 are unknown parameter, but σ is supposed to be known. If $\theta_2 > 0$, then all of the conditions are fulfilled automatically. Furthermore, the stationary distribution is the normal distribution with mean θ_1/θ_2 and variance $\sigma^2/(2\theta_2)$ and the basic elements F's for the coefficients of the asymptotic expansion have explicit representation as follows:

$$F_{1,1} = \frac{1}{\sigma^2}, \quad F_{1,2} = -\frac{\theta_1}{\theta_2\sigma^2}, \quad F_{2,2} = \frac{1}{\sigma^2}\left(\frac{\sigma^2}{2\theta_2} + \frac{\theta_1^2}{\theta_2^2}\right)$$

$$F_{[1,1],1} = F_{[1,1],2} = 0, \quad F_{[1,2],1} = -\frac{1}{\theta_2\sigma^2}, \quad F_{[1,2],2} = \frac{\theta_1}{\theta_2^2\sigma^2},$$

$$F_{[2,2],1} = \frac{2\theta_1}{\theta_2^2\sigma^2}, \quad F_{[2,2],2} = -\frac{1}{\theta_2\sigma^2}\left(\frac{\sigma^2}{2\theta_2} + 2\frac{\theta_1^2}{\theta_2^2}\right),$$

$$F_{[1,1],[a,b]} = 0, \ a,b = 1,2, \quad F_{[1,2],[1,2]} = \frac{1}{\theta_2^2\sigma^2}, \quad F_{[1,2],[2,2]} = -\frac{2}{\theta_2^2\sigma^2}\frac{\theta_1}{\theta_2}$$

$$F_{[2,2],[2,2]} = \frac{1}{\theta_2^2\sigma^2}\left(\frac{\sigma^2}{2\theta_2} + 4\frac{\theta_1^2}{\theta_2^2}\right),$$

$$F_{[[1,1],a],b} = F_{[[1,2],1],a} = 0, \ a,b = 1,2, \quad F_{[[1,2],2],1} = F_{[[2,2],1],1} = \frac{1}{\theta_2^2\sigma^2},$$

$$F_{[[1,2],2],2} = F_{[[2,2],1],2} = -\frac{1}{\theta_2^2\sigma^2}\frac{\theta_1}{\theta_2}, \quad F_{[[2,2],2],1} = -3\frac{1}{\theta_2^2\sigma^2}\frac{\theta_1}{\theta_2},$$

$$F_{[[2,2],2],2} = \frac{1}{\theta_2^2\sigma^2}\left(\frac{\sigma^2}{2\theta_2} + 3\frac{\theta_1^2}{\theta_2^2}\right).$$

From these, we can obtain the representaion of the asymptotic expansion.

References

[1] Kusuoka, S., and Yoshida, N.(1997): Malliavin calculus, strong mixing, and expansion of diffusion functionals. Cooperative Research Report **111**, The Institute of Statistical Mathematics. to appear in *Probab. Theory Relat. Fields.*

[2] Rao, C.R., Mukerjee, R. (1997): Comparison of LR, Score, and Wald Tests in a Non-IID Setting. *J. Multivariate Anal.*, **60**, 99-110.

[3] Roverts, G.O., Tweedie, R.L. (1996): Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, **2-4**, 341-363.

[4] Sakamoto, Y., Yoshida, N. (1998): Third order asymptotic expansion for diffusion process, In: Theory of statistical analysis and its applications, Cooperative Research Report **107**, 53–60, The Institute of Statistical Mathematics.

[5] Sakamoto,Y. and Yoshida,N.(1999): Higher order asymptotic expansion for a functional of a mixing process with applications to diffusion processes. *submitted*

[6] Stroock, D.W.(1994): Probability theory, an analytic view. Cambridge.

[7] Taniguchi, M. (1991): Third-Order Asymptotic Properties of a Class of Test Statistics under a Local Alternative. *J. Multivariate Anal.*, **37**, 223-238.

[8] Tong, H.(1990): Non-linear Time Series: A Dynamical System Approach. Oxford Univ. Press.

[9] Tweedie, R.L. (1983): Criteria for rates of convergence of Markov chains, with applications to queueing theory. In *Papers in Probability, Statistics and Analysis* (J.F.C. Kingman and G.E.H. Reuter, eds.) 260–276. Cambridge Univ. Press.

[10] Uchida, M., and Yoshida, N.(1999): Information criteria in model selection for mixing process. *submitted*