統計数理（1997）
第45巻 第2号 329-342

状態空間接近による季節変動調整に関する一考察

——東京湾データへの適用——

統計数理研究所 柏 木 宣 久

（1997年7月 受付）

1. 経 緯

東京湾沿岸の各自治体は，湾内の水質状況を把握するため，水質汚濁防止法に基づく公共用水域水質測定計画に沿って，毎月1回，測定点を分担し合い，水温，塩分，COD等，多項目の水質測定を実施している．この測定は既に二十数年に亘り実施されており，現在までに収集されたデータは相当量に上る．これらのデータを再活用するため，各自治体の環境保全局や環境部の協力を得て，データスクリーニングを行いながら，過去20年分のデータをデータベース化する作業が進行中である．その内，5年分のデータベースが先行して完成しており，傾向分析を目的に解析を試みている．

東京湾水質測定データには，傾向分析の立場から見ると，2つの特徴がある．ひとつが気象や潮流等の季節変化に起因する季節変動を内在する点，もうひとつが東京湾という地理的広がりを持つ空間上での観測の時系列，すなわち時空間データであるという点である．そのため，傾向分析を行う場合，時空間季節変動調整が解析の基本となる．しかしながら，時空間季節変動調整には膨大な計算量が必要で，解析を始めた当初には，その実行は不可能であった．そこで，データを測定点毎に独立な時系列とみなし，Akaike (1980) の季節変動調整法を適用し，得られた各測定点の推定値を空間補間 (Matusita et al. (1988)) で繋ぎ，各測定項目の時空間変化を観察するという方法をとってきた（二宮 他 (1996a, 1996b, 1997)）．ところが，最近，計算機が一段と発達し，大規模計算を安価に実行できる情勢となってきたため，時空間季節変動調整法の開発に着手した．本稿では，その開発過程で生じた諸問題に対する検討結果の内，時系列の季節変動調整に関する事柄について述べる．

時系列としての東京湾水質測定データには，従来，Akaikeの方法を適用してきた．仮定したモデルを以下に記す．

$$y_k = t_k + s_k + u_k, \qquad u_k \sim \mathrm{NID}(0, \sigma^2), \qquad k=1,\ldots, n \tag{1.1}$$

$$t_k - 2t_{k-1} + t_{k-2} \sim \mathrm{NID}(0, \sigma^2/\alpha), \qquad k=3,\ldots, n \tag{1.2}$$

$$s_k - s_{k-12} \sim \mathrm{NID}(0, \sigma^2/\beta), \qquad k=13,\ldots, n \tag{1.3}$$

$$\sum_{l=0}^{11} s_{k-l} \sim \mathrm{NID}(0, \sigma^2/\beta), \qquad k=12,\ldots, n \tag{1.4}$$

ただし，y_k, t_k および s_k は，特定の測定点における第 k 月の観測，トレンドおよび季節成分を各々表す．Akaike (1980) は(1.4)を(1.3)と異なる分散を用いて

$$\sum_{l=0}^{11} s_{k-l} \sim \mathrm{NID}(0, \sigma^2/\gamma)$$

としたが，$\gamma \to 0$ となった場合，(1.2)と(1.3)が(1.1)を通じて釣り合い，トレンドと季節成分が識別不能となるため，(1.3)と(1.4)の分散を共通にしてある．また，Akaike (1980) で仮定された端点に関する条件は，退化した事前分布を容認すれば不要なため，省略した．Akaikeの方法では，仮定したモデルに含まれる未知変数および変量の内，変量であるトレンドと季節成分を，モデルから演繹されるそれらの事後密度により推定する．特に，各変量の事後密度に関する平均値は，分散成分を既知とした場合の以下の基準による最小2乗推定値に一致する．

$$\sum_{k=1}^{n}(y_k - t_k - s_k)^2 + \alpha\sum_{k=3}^{n}(t_k - 2t_{k-1} + t_{k-2})^2 + \beta\sum_{k=13}^{n}(s_k - s_{k-12})^2 + \beta\sum_{k=12}^{n}\left(\sum_{l=0}^{11} s_{k-l}\right)^2$$

残る未知変数である分散成分は，以下のベイズ型対数尤度を用いて最尤推定する．

$$-2\log L(\alpha, \beta, \tilde{\sigma}^2) = (n-13)\log \tilde{\sigma}^2 - (n-2)\log \alpha - (n-11)\log \beta + \log|E'E| + \text{const.}$$

ただし，

$$\tilde{\sigma}^2 = \frac{1}{n-13}(\boldsymbol{z} - E\tilde{\boldsymbol{x}})'(\boldsymbol{z} - E\tilde{\boldsymbol{x}})$$

であり，$\boldsymbol{z}$ は y_k と0を要素とする $(4n-25)\times 1$ のベクトルであり，$\tilde{\boldsymbol{x}}$ は t_k と s_k の最小2乗推定値を要素とする $2n\times 1$ のベクトルであり，E は最小二乗基準を行列表示するための $(4n-25)\times 2n$ の係数行列である．以上の推定手続きを数値的方法により履行する（例えば，Akaike and Ishiguro (1980)，Ishiguro (1984) を参照）．扱う数値的問題の大きさは，最小2乗基準や対数尤度中の行列式から明らかなように，変量 t_k および s_k の数により規定される．

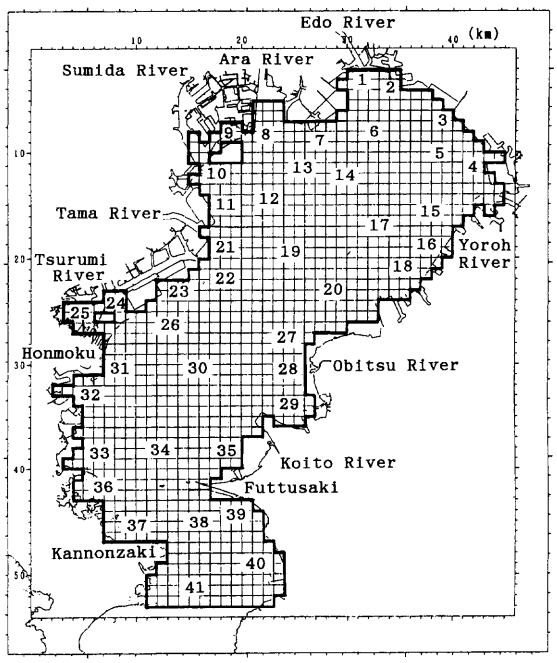

図1． 東京湾を1kmメッシュに分割した図．図中の数字が測定点を示す．

時空間季節変動調整法を構成するには，時間方向の季節変動調整機能に加えて，各変量の空間方向の変動も考慮しなければならない．しかも，東京湾の場合，湾口と湾奥あるいは沿岸と沖合いでは水質状況が異なり，空間方向の変動は非定常と考えられる．そのため，開発中の方法では，東京湾を1 kmメッシュに分割し（図1参照），格子モデルと階差制約による空間平滑化（Kashiwagi (1993)）と同様の方法により空間方向の変動をモデリングしている．メッシュの数は全部で981ある．各メッシュ上でトレンドおよび季節成分が定義されるから，特定の月における変量の数は1962となり，20年分のデータを対象に考えれば，変量の総数は470880となる．この数は，Akaikeの方法を単純に拡張して時空間季節変動調整法を構成した場合に扱う数値的問題の大きさを規定しており，合理的な費用で解ける限界を越えている．また，何らかの方法により測定点に対応しないメッシュ上の変量をモデルから解析的に消去したとしても，現状では測定点を41までしか絞り込めないため，変量の総数は19680までしか減少せず，実用的な大きさにはならない．

状態空間接近を用いると，扱う数値的問題が小さくなる．Akaikeの方法に対応する状態空間接近による季節変動調整法として，Kitagawa and Gersch (1984) の方法がある．彼らは，前述した識別可能性の問題から，2つの季節成分モデル(1.3)および(1.4)の内，(1.4)のみを採用し，季節変動調整のための状態空間モデルを以下のように与えている．

$$y_k = F\boldsymbol{x}_k + u_k, \qquad u_k \sim \mathrm{NID}(0, \sigma^2)$$
$$\boldsymbol{x}_k = G\boldsymbol{x}_{k-1} + \boldsymbol{v}_k, \qquad \boldsymbol{v}_k \sim \mathrm{NID}(\boldsymbol{0}, \sigma^2 H)$$

ただし，$\boldsymbol{x}_k = (t_k, t_{k-1}, s_k, \ldots, s_{k-10})'$，$F = (1, 0, 1, 0, \ldots, 0)$，および

$$G = \begin{bmatrix} 2 & -1 & & & & & \\ 1 & 0 & & & & & \\ & & -1 & \cdots & -1 & -1 & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & & & & 1 & \end{bmatrix}, \quad H = \begin{bmatrix} \alpha^{-1} & & & & & \\ & 0 & & & & \\ & & \beta^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

である．一旦，状態空間モデルが与えられれば，変数および変量はカルマンフィルタ等の逐次公式により推定できる．逐次公式は時点毎の手続きとして与えられるから，逐次公式で扱う数値的問題の大きさは，状態$\boldsymbol{x}_k$の次元のみに依存し，今の場合，その数は13である．Kitagawa and Gershの方法を拡張して時空間季節変動調整法を構成した場合，測定点に対応しないメッシュ上の変量をモデルから解析的に消去すれば，状態の次元は$13 \times 41 = 533$となり，Akaikeの方法を拡張した場合に比べ，扱う数値的問題の大きさが大幅に小さくなる．

上述の事情から，開発中の方法では，状態空間接近を採用している．ただし，季節変動調整のための状態空間モデルは唯一ではなく，Kitagawa and Gershのモデル以外にも，複数のモデルを候補として想定できた．また，現状では僅か5年分のデータしか利用できず，こうした短いデータに対する状態空間接近による季節変動調整法の動作確認も必要であった．そこで，時系列としての東京湾水質測定データに対し，Akaikeの方法とKitagawa and Gershのモデルを含む4つの状態空間モデルによる季節変動調整法を適用し，各々の特性を調べた．本稿では，その結果について述べる．第2章でAkaikeのモデルに対応するKitagawa and Gershのモデルとは異なる状態空間モデルと2つの自己組織化（Kitagawa (1996)）された状態空間モデルを提示し，第3章でシステムノイズの共分散を観測ノイズの分散に対する比の形で与えた場合のカルマンフィルタにおける初期条件について述べ，第4章で時系列としての東京湾水質測定データに対する適用結果を示す．

2. モデル

本章では，時系列としての東京湾水質測定データに適用した季節変動調整のための状態空間モデルの内，Kitagawa and Gersh のモデルを除く3つのモデルを提示する．

2.1 Model A

季節成分モデル(1.3)をバックワードシフトオペレータBを使って表現すると，

$$(1-B^{12})s_k=(1-B)(1+B+\cdots+B^{11})s_k \sim \mathrm{NID}(0,\ \sigma^2/\beta)$$

となる．一方，トレンドモデル(1.2)は，

$$(1-B)^2t_k \sim \mathrm{NID}(0,\ \sigma^2/\alpha)$$

と表現される．両表現に共通因子$(1-B)$が含まれるため，(1.3)のみを季節成分モデルとして採用すると，前章で述べた識別問題が生じる．そこで，Kitagawa and Gersh (1984) は，(1.3)のオペレータ表現の因子でもある$(1+B+\cdots+B^{11})$に対応する(1.4)を季節成分モデルとして採用した．本節では，(1.3)および(1.4)の双方に基づく状態空間モデルを提示する．

季節成分モデル(1.3)および(1.4)をシステムモデルの形で同時に書くと，

$$\begin{pmatrix}1\\1\end{pmatrix}s_k=\begin{bmatrix}0 & \cdots & 0 & 1\\ -1 & \cdots & -1 & 0\end{bmatrix}\begin{pmatrix}s_{k-1}\\ \vdots\\ s_{k-11}\\ s_{k-12}\end{pmatrix}+\begin{pmatrix}v_k\\ w_k\end{pmatrix}$$

となる．ただし，v_k，$w_k \sim \mathrm{NID}(0,\ \sigma^2/\beta)$である．このモデルは，左辺に係数ベクトルがあるため，このままではシステムモデルとして利用できない．そこで，これをMoore-Penrose逆行列を用いて消去すると，

$$s_k=(-1/2,\ldots,\ -1/2,\ 1/2)\begin{pmatrix}s_{k-1}\\ \vdots\\ s_{k-11}\\ s_{k-12}\end{pmatrix}+(v_k+w_k)/2$$

となる．そのオペレータ表現は，

$$(1-B/2)(1+B+\cdots+B^{11})s_k \sim \mathrm{NID}(0,\ \sigma^2/2\beta)$$

であり，(1.2)のオペレータ表現の因子$(1-B)$を含まない．そこで，季節変動調整のための状態空間モデルのひとつとして，以下のモデルを仮定した．

$$\begin{aligned} y_k&=F\boldsymbol{x}_k+u_k, & u_k &\sim \mathrm{NID}(0,\ \sigma^2)\\ \boldsymbol{x}_k&=G\boldsymbol{x}_{k-1}+\boldsymbol{v}_k, & \boldsymbol{v}_k &\sim \mathrm{NID}(\boldsymbol{0},\ \sigma^2H)\end{aligned}$$

ただし，$\boldsymbol{x}_k=(t_k,\ t_{k-1},\ s_k,\ldots,\ s_{k-11})'$，$F=(1,\ 0,\ 1,\ 0,\ldots,\ 0)$，および

$$G=\begin{bmatrix} 2 & -1 & & & & \\ 1 & 0 & & & & \\ & & -1/2 & \cdots & -1/2 & 1/2 \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \end{bmatrix},\quad H=\begin{bmatrix} \alpha^{-1} & & & & & \\ & 0 & & & & \\ & & (2\beta)^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

である．なお，このモデルと同等なモデルがOzaki and Thomson (1994) により提案されていると統計数理研究所の尾崎統氏から個人的に指摘されたので付記する．ただし，導出過程は異なっている．

2.2 Model B

Akaikeのモデルは，各成分の水準が変化する場合を想定して開発されたため，各成分の相対的関係のみにより構成されており，文学的に表現すれば，「柔らかい」モデルと言える．一方，東京湾水質測定データの場合，その長さは短く，水質状況が最近大きく変化した兆候も見られないため，特に季節成分に関しては「より堅い」モデルの方が適切である可能性を排除できない．「より堅い」モデルとしては，季節成分の平均を暦月毎に固定する仕方が考えられる．そして「更に堅い」モデルとして，季節成分自体を暦月毎に固定する仕方もある．ただし，季節成分の平均あるいは季節成分自体を固定すると，標準的な状態空間接近による限り，推定に問題が生じる．本節では，季節成分を暦月毎に固定した場合を取り上げ，問題点を指摘した後，問題点を回避するための自己組織化 (Kitagawa (1996)) された状態空間モデルを提示する．Kitagawa (1996) は自己組織化の方法を非ガウス過程のモデリングに用いているが，この方法はガウス過程でも利用でき，特に外的変数が存在する場合に有用である．

季節成分を暦月毎に固定するには，

$$s_k=\boldsymbol{c}_{[k]}\boldsymbol{a}$$

とすればよい．ただし，$\boldsymbol{a}$ は11ケ月の季節成分からなるベクトル $(a_1,\ldots, a_{11})'$ であり，$[k]$ は k を12で除した余りを表し，$\boldsymbol{c}_i$ は s_k と a_i $(1\leq i=[k]\leq 11)$ を対応させるための0と1からなる 1×11 の係数ベクトルであり，$\boldsymbol{c}_0$ のみ12ケ月の季節成分の和が0となるよう $(-1,\ldots, -1)$ とする．この時，標準的な方法に従えば，状態空間モデルは以下のように書かれる．

$$\begin{aligned} y_k&=F\boldsymbol{x}_k+\boldsymbol{c}_{[k]}\boldsymbol{a}+u_k, \qquad u_k \sim \mathrm{NID}(0, \sigma^2)\\ \boldsymbol{x}_k&=G\boldsymbol{x}_{k-1}+\boldsymbol{v}_k, \qquad \boldsymbol{v}_k \sim \mathrm{NID}(\boldsymbol{0}, \sigma^2 H)\end{aligned}$$

ただし，$\boldsymbol{x}_k=(t_k, t_{k-1})'$，$F=(1, 0)$ および

$$G=\begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix},\qquad H=\begin{bmatrix} \alpha^{-1} & \\ & 0 \end{bmatrix}$$

である．

このモデルに対するカルマンフィルタは以下のように与えられる．

$$\begin{aligned} \text{(prediction)}\quad & \hat{\boldsymbol{x}}_{k/k-1}=G\hat{\boldsymbol{x}}_{k-1/k-1}, \qquad \Omega_{k/k-1}=G\Omega_{k-1/k-1}G'+H\\ \text{(filtering)}\quad & \hat{\boldsymbol{x}}_{k/k}=\hat{\boldsymbol{x}}_{k/k-1}+\Omega_{k/k-1}F'(y_k-F\hat{\boldsymbol{x}}_{k/k-1}-\boldsymbol{c}_{[k]}\boldsymbol{a})/(F\Omega_{k/k-1}F'+1),\\ & \Omega_{k/k}=\Omega_{k/k-1}-\Omega_{k/k-1}F'F\Omega_{k/k-1}/(F\Omega_{k/k-1}F'+1)\\ \text{(log likelihood)}\quad & -2\log L(\boldsymbol{a}, \alpha, \hat{\sigma}^2)=n\log\hat{\sigma}^2+\sum_{k=1}^{n}\log|F\Omega_{k/k-1}F'+1|+\text{const.}\end{aligned}$$

ただし，$\hat{\boldsymbol{x}}_{i/j}$ および $\sigma^2\Omega_{i/j}$ は $y_1,\ldots, y_j$ が与えられた下での状態 $\boldsymbol{X}_i$ の条件付き平均および共分散であり，

$$\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(y_k-F\hat{\boldsymbol{x}}_{k/k-1}-\boldsymbol{c}_{[k]}\boldsymbol{a})^2/(F\Omega_{k/k-1}F'+1)$$

である．初期条件 $\hat{\boldsymbol{x}}_{0/0}$ および $\Omega_{0/0}$ と変数 $\boldsymbol{a}$ および α の値が与えられれば，上記の手続きを実行でき，与えられた $\boldsymbol{a}$ および α の値の良さは対数尤度により評価される．ただし，手続き上，$\boldsymbol{a}$ および α の値を先験的に与える必要があるため，これらの最尤推定値は評価される対数尤度を目的値とした数値的最適化法により求めなければならない．数値的最適化の対象となる変数の次元が高いため，推定には費用が掛かる．

自己組織化の方法を用いると，推定に掛かる費用を大幅に軽減できる．季節成分を暦月毎に固定した場合の自己組織化されたモデルは以下のように与えられる．

$$\begin{aligned} y_k&=F_k\boldsymbol{x}_k+u_k, \qquad u_k \sim \mathrm{NID}(0,\ \sigma^2)\\ \boldsymbol{x}_k&=G\boldsymbol{x}_{k-1}+\boldsymbol{v}_k, \qquad \boldsymbol{v}_k \sim \mathrm{NID}(\boldsymbol{0},\ \sigma^2H)\end{aligned}$$

ただし，$\boldsymbol{x}_k=(t_k,\ t_{k-1},\ a_1,\ldots,\ a_{11})'$，$F_k=(1,\ 0,\ \boldsymbol{c}_{[k]})$，および

$$G=\begin{bmatrix} 2 & -1 & & & & \\ 1 & 0 & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & & 1 \end{bmatrix},\quad H=\begin{bmatrix} \alpha^{-1} & & & & & \\ & 0 & & & & \\ & & 0 & & & \\ & & & \ddots & & \\ & & & & & 0 \end{bmatrix}$$

である．このように，自己組織化の方法では $\boldsymbol{a}$ も状態に組み込む．そのため，$\boldsymbol{a}$ は最尤推定における数値的最適化の対象ではなくなり，推定の費用が大幅に軽減される．ただし，$\boldsymbol{a}$ は変数から変量へと変質する．

季節成分を暦月毎に固定する仕方は Kitagawa and Gersh のモデルで $\beta=\infty$ としても実現できる．ただし，その場合のモデルは上記の自己組織化されたモデルと同等である．また，外的変数 $\boldsymbol{\theta}$ を含むガウス過程の状態空間モデル，例えば，

$$\begin{aligned} y_k&=F\boldsymbol{x}_k+D\boldsymbol{\theta}+u_k, \qquad u_k \sim \mathrm{NID}(0,\ \sigma^2)\\ \boldsymbol{x}_k&=G\boldsymbol{x}_{k-1}+\boldsymbol{v}_k, \qquad \boldsymbol{v}_k \sim \mathrm{NID}(\boldsymbol{0},\ \sigma^2H)\end{aligned}$$

というモデルを自己組織化すれば，推定が容易になるのは述べるまでもない．

2.3 Model C

本節では，季節成分の平均を暦月毎に固定した場合の自己組織化されたモデルを提示する．平均を暦月毎に固定する仕方は，データに依存した Stein 型推定の一種と言える．データに依存した Stein 型推定は意図通りに機能しない場合が多いという経験があり，特に動作確認が必要であった．

季節成分の平均を暦月毎に固定した場合の季節成分モデルは，前節の記号を用いて，

$$s_k-\boldsymbol{c}_{[k]}\boldsymbol{a} \sim \mathrm{NID}(0,\ \sigma^2/\beta)$$

と書ける．この時，自己組織化されたモデルは以下のように与えられる．

$$y_k=F\boldsymbol{x}_k+u_k\,,\qquad u_k\sim \mathrm{NID}(0,\ \sigma^2)$$
$$\boldsymbol{x}_k=G_k\boldsymbol{x}_{k-1}+\boldsymbol{v}_k\,,\qquad \boldsymbol{v}_k\sim \mathrm{NID}(\boldsymbol{0},\ \sigma^2H)$$

ただし，$\boldsymbol{x}_k=(t_k,\ t_{k-1},\ s_k,\ a_1,\ldots,\ a_{11})'$，$F=(1,\ 0,\ 1,\ 0,\ldots,\ 0)$，および

$$G_k=\begin{bmatrix} 2 & -1 & & & & \\ 1 & 0 & & & & \\ & & 0 & & \boldsymbol{c}_{[k]} & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix},\quad H=\begin{bmatrix} \alpha^{-1} & & & & & \\ & 0 & & & & \\ & & \beta^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

である．

3. 初期条件

前章までのように，システムノイズの共分散を観測ノイズの分散に対する比の形で与えた場合，第2.2節のカルマンフィルタに記述してある通り，観測ノイズの分散の最尤推定値が陽に求まり，最尤推定における数値的最適化の対象となる変数が1つ減少する．ただし，初期条件に関し問題が生じる．本章では，システムノイズの共分散を比の形で与えた場合のカルマンフィルタにおける初期条件について述べる．

以下の状態空間モデルを考えよう．

$$\boldsymbol{y}_k=F\boldsymbol{x}_k+\boldsymbol{u}_k,\qquad \boldsymbol{u}_k\sim \mathrm{NID}(\boldsymbol{0},\ \sigma^2I)$$
$$\boldsymbol{x}_k=G\boldsymbol{x}_{k-1}+\boldsymbol{v}_k,\qquad \boldsymbol{v}_k\sim \mathrm{NID}(\boldsymbol{0},\ K)$$

ただし，$\dim\boldsymbol{y}_k=m$ および $k=1,\ldots,\ n$ とする．このモデルに対するカルマンフィルタは以下のように与えられる．

$$\text{(prediction)}\quad \hat{\boldsymbol{x}}_{k/k-1}=G\hat{\boldsymbol{x}}_{k-1/k-1}\,,\qquad \Sigma_{k/k-1}=G\Sigma_{k-1/k-1}G'+K$$
$$\text{(filtering)}\quad \hat{\boldsymbol{x}}_{k/k}=\hat{\boldsymbol{x}}_{k/k-1}+\Sigma_{k/k-1}F'(F\Sigma_{k/k-1}F'+\sigma^2I)^{-1}(y_k-F\hat{\boldsymbol{x}}_{k/k-1})\,,$$
$$\Sigma_{k/k}=\Sigma_{k/k-1}-\Sigma_{k/k-1}F'(F\Sigma_{k/k-1}F'+\sigma^2I)^{-1}F\Sigma_{k/k-1}$$
$$\text{(log likelihood)}\quad -2\log L(\sigma^2,\ K)=\sum_{k=1}^{n}\log|F\Sigma_{k/k-1}F'+\sigma^2I|$$
$$+\sum_{k=1}^{n}(\boldsymbol{y}_k-F\hat{\boldsymbol{x}}_{k/k-1})'(F\Sigma_{k/k-1}F'+\sigma^2I)^{-1}(\boldsymbol{y}_k-F\hat{\boldsymbol{x}}_{k/k-1})+\text{const.}$$

ただし，$\hat{\boldsymbol{x}}_{i/j}$ および $\Sigma_{i/j}$ は $\boldsymbol{y}_1,\ldots,\ \boldsymbol{y}_j$ が与えられた下での状態 $\boldsymbol{X}_i$ の条件付き平均および共分散である．初期条件 $\hat{\boldsymbol{x}}_{0/0}$ および $\Sigma_{0/0}$ と変数 σ^2 および K の値が与えられれば，上記の手続きを実行できる．初期条件としては，無情報的な条件

$$\hat{\boldsymbol{x}}_{0/0}=\bar{\boldsymbol{x}}_{0/0}\,,\qquad \Sigma_{0/0}=\lambda I$$

を仮定するのが一般的である．ただし，$\bar{\boldsymbol{x}}_{0/0}$ は適当な定数，例えば，$\bar{\boldsymbol{x}}_{0/0}=\boldsymbol{0}$ であり，λ は十分大きな数とする．変数 σ^2 および K は，与えられた値に応じて評価される対数尤度を目的値とした数値的最適化法により最尤推定する．

ここで，システムノイズの共分散を以下のように書き換える．

$$K=\sigma^2H$$

この共分散を上記のカルマンフィルタに代入し，$\Sigma_{i/j}=\sigma^2\Omega_{i/j}$ として σ^2 について整理すると，以下のようになる．

$$\begin{aligned}
&\text{(prediction)} \quad \hat{\boldsymbol{x}}_{k/k-1}=G\hat{\boldsymbol{x}}_{k-1/k-1}, \qquad \Omega_{k/k-1}=G\Omega_{k-1/k-1}G'+H\\
&\text{(filtering)} \quad \hat{\boldsymbol{x}}_{k/k}=\hat{\boldsymbol{x}}_{k/k-1}+\Omega_{k/k-1}F'(F\Omega_{k/k-1}F'+I)^{-1}(y_k-F\hat{\boldsymbol{x}}_{k/k-1}),\\
&\qquad\qquad \Omega_{k/k}=\Omega_{k/k-1}-\Omega_{k/k-1}F'(F\Omega_{k/k-1}F'+I)^{-1}F\Omega_{k/k-1}\\
&\text{(log likelihood)} \quad -2\log L(\hat{\sigma}^2, H)=mn\log\hat{\sigma}^2+\sum_{k=1}^{n}\log|F\Omega_{k/k-1}F'+I|+\text{const.}
\end{aligned}$$

ただし，

$$\hat{\sigma}^2=\frac{1}{mn}\sum_{k=1}^{n}(\boldsymbol{y}_k-F\hat{\boldsymbol{x}}_{k/k-1})'(F\Omega_{k/k-1}F'+I)^{-1}(\boldsymbol{y}_k-F\hat{\boldsymbol{x}}_{k/k-1})$$

である．このように，システムノイズの共分散を比の形で与えると，σ^2 の最尤推定値が陽に求まり，数値的最適化の対象となる変数が1つ減少する．ただし，初期条件に関し問題が生じる．

対数尤度は，上記最初のカルマンフィルタで明らかなように，初期条件，すなわち $\boldsymbol{X}_0$ の初期分布 $\mathrm{N}(\hat{\boldsymbol{x}}_{0/0}, \Sigma_{0/0})$ に依存している．従って，未知変数を最尤推定するには，同一条件下での対数尤度の比較を実現するため，$\mathrm{N}(\hat{\boldsymbol{x}}_{0/0}, \Sigma_{0/0})$ を一定に保つ必要がある．上記最初のカルマンフィルタの場合，この条件を満たすのは容易である．一方，上記2番目のカルマンフィルタの場合，$\boldsymbol{X}_0$ の初期分布は $\mathrm{N}(\hat{\boldsymbol{x}}_{0/0}, \sigma^2\Omega_{0/0})$ と書かれるが，初期条件として与えられるのは $\hat{\boldsymbol{x}}_{0/0}$ および $\Omega_{0/0}$ だけであり，σ^2 は事後的に推定されるから，$\mathrm{N}(\hat{\boldsymbol{x}}_{0/0}, \sigma^2\Omega_{0/0})$ を先験的には一定にできない．初期分布が変化すれば，対数尤度は未知変数ばかりでなく初期分布の関数としても振る舞い，最尤推定に支障が生じる．東京湾水質測定データを用いた数値実験では，$\hat{\boldsymbol{x}}_{0/0}$ および $\Omega_{0/0}$ のみを一定にした場合，初期状態の尤度が高くなるよう $\hat{\sigma}^2\to 0$ となる H が推定される傾向が強く見られた．

従来，対数尤度に対する初期分布の影響を無視できない場合，その影響を減少させるため，状態の共分散の挙動が安定する時点を主観的に選び出し，その時点以降から尤度計算を開始する方法がしばしば採られてきた．状態の次元が低くデータが十分に長い場合に有効な方法と言える．ただし，状態の次元が比較的高くデータが短い場合，時点選出の任意性やデータの希少性により，そうした方法は取り難い．そこで，東京湾水質測定データの解析では，上記2番目のカルマンフィルタを用いる場合，初期分布をほぼ一定にするため，以下の手続きを採用している．

1. λ として十分大きな数を採り，$\hat{\sigma}_0^2\leftarrow 1$ および $\log L_0\leftarrow\delta$ とする．
2. $\hat{\boldsymbol{x}}_{0/0}\leftarrow\bar{\boldsymbol{x}}_{0/0}$ および $\Omega_{0/0}\leftarrow(\lambda/\hat{\sigma}_0^2)I$ とする．
3. カルマンフィルタにより推定値 $\hat{\sigma}^2$ および対数尤度 $\log L$ を求める．
4. $|\log L-\log L_0|<\epsilon$ ならば手続きを終了し，そうでなければ $\hat{\sigma}_0^2\leftarrow\hat{\sigma}^2$ および $\log L_0\leftarrow\log L$ としてステップ2に戻る．

ただし，δ は計算機が許容する最小の数であり，ϵ は対数尤度に対し要求される精度に応じて決める定数である．

上記2番目のカルマンフィルタを用いると，最尤推定における数値的最適化の対象となる変数が1つ減少するが，上述の手続きを併せて用いなければならないため，上記最初のカルマンフィルタを用いた場合に比べ，最尤推定の費用が常に軽減されるとは限らない．軽減されるか否かは，対数尤度に要求される精度に依存する．要求精度が小数点以下1ないし2桁程度であれば，手続きは比較的早く収束し，ほとんどの場合，費用が軽減される．しかし，要求精度が高くなると，手続きの収束が遅くなり，費用が増加する可能性すらある．例えば，可変計量法により最適化を行う際に，最適解近傍等で対数尤度が変数の変化に対し極端に鈍感になると，

より正確な微係数が必要となり，要求精度を高くしなければならなくなる．そうした場合，ϵ の値を適応的に変化させても対処できるが，費用の問題や，希に探査方向の自由度を増すと可変計量法の動作が改善される場合があるため，状況に応じて上記最初のカルマンフィルタの利用も考慮した方が良い．

なお，最後に，$\boldsymbol{x}_{0/0}$ および λ について付言する．一般に，初期値 $\boldsymbol{x}_{0/0}$ としては $\mathbf{0}$ が採られるが，適当な値を想定できる場合には，そうした値を採用した方が良い．初期分散 λ を極端に大きくする必要がなくなるためである．初期値が適当な値であれば，そう大きくない λ によって初期分布を無情報化できる．他方，初期値が適当な値から乖離していると，その分 λ を大きくしなければならなくなり，λ が大き過ぎると，計算機は有限桁しか扱えないため，カルマンフィルタの実行中，初期の時点で状態の共分散が数値的に退化してしまう．著者の時系列季節変動調整プログラムでは，トレンドの初期値としては最初の 1 周期の標本平均，自己組織化された場合も含めた季節成分の初期値としては $\mathbf{0}$，λ としては最初の 1 周期の標本分散の 10000 倍を採っている．

4. 適用結果

短いデータに対する状態空間モデルによる季節変動調整法の動作確認を目的に，前章までに述べた方法を時系列としての東京湾水質測定データに適用した．本章ではその結果について述べる．

使用したデータは，湾内 41 個所の測定点（図 1 参照）における上層および下層の塩分の時系列である．塩分は主として河川から流入する真水の混入量により変化し，混入量は降雨量や河川からの距離ばかりではなく他の気象や潮流によっても変化する．そのため，使用した 82 組の時系列データは，測定点や層の違いにより様々な変動の様相を示しており，動作確認に都合の良いデータと言える．一方，最尤推定における数値的最適化法としては，動作確認が目的のため，格子探査法を採用した．その探査範囲は各々 $2^i (i=-5,\ldots,15)$ である．探査範囲を粗く採ったため，前章 2 番目のカルマンフィルタを初期分布を一定にする手続きと共に用いた．ただし，$\epsilon=0.1$ とした．そして，トレンドと季節成分は，逐次公式における平滑化の公式により推定した．

結果を総評すれば，各モデルは，ほぼ予想された通りの動作を示した．自己組織化されたモデルも季節成分の推定に関しては問題なく動作した．ただし，Model C では，予想された通り，一部のデータの組で $\hat{\sigma}^2 \to 0$ となり，残差の分離に失敗した．データに依存した Stein 型推定でよく経験する現象で，データとモデルの相性の問題から，観測の繰り返しを捉えきれないのが原因である．もうひとつの自己組織化されたモデルである Model B は安定して動作した．ただし，僅か 5 年分のデータとはいえ，残差が大きくなり過ぎる場合があり，モデルが「堅すぎる」欠点も見られた．残る Kitagawa and Gersh のモデルと Model A は，データが短いにもかかわらず，順調に動作した．「柔らかさ」という点では，状態の次元が高い分，Model A の方が Kitagawa and Gersh のモデルより「柔らかく」，季節成分をより多く見積もる傾向が見られた．そうした傾向は Akaike の方法にも見られ，Model A による結果と Akaike の方法による結果は概ね類似していた．図 1 中に数字 6 で示す測定点のデータに対する Akaike の方法と各モデルの適用結果を図 2 から 6 に示す．各々の微妙な特性が現れている．特に 1985 年 7 月に見られる異常値に対する反応は特徴的で，Model B は別として，Akaike の方法と Model A および C が類似の反応を示す一方，Kitagawa and Gersh のモデルがやや頑健な反応を示した．ちなみに，この異常値は測定数日前に接近した台風に起因する．また，何れのトレンドも上に凸の形をしているが，これは年間降雨量の推移に対応している．実際，関東地方は 1986 年度から 1987 年度にかけて少雨であった．

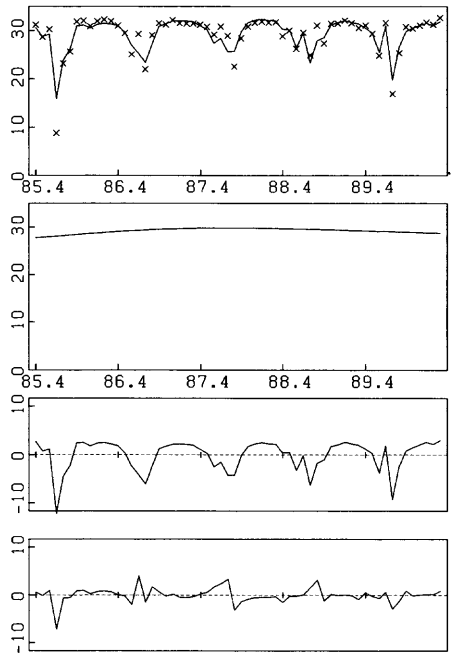

図2. Akaikeの方法の適用結果．上段から，データ（×印）とトレンド＋季節成分の平均値（実線），トレンドの平均値，季節成分の平均値，および各成分の平均値から見た残差を示す．各図の縦軸は塩分濃度，横軸は年月を表す（以下同）．

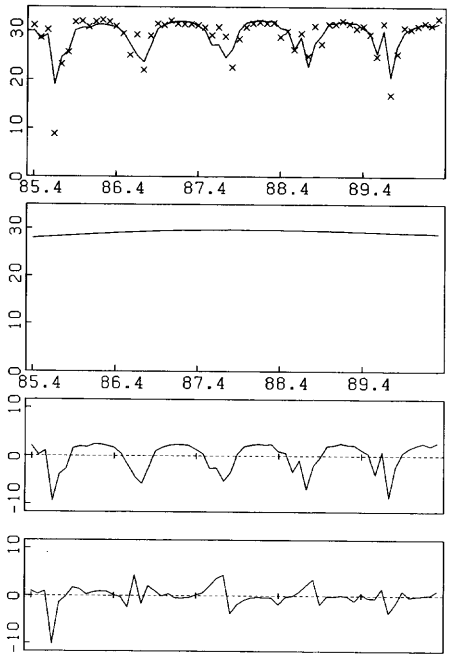

図3. Kitagawa and Gershのモデルの適用結果．

なお，今回は，動作確認のみを目的とし，方法やモデルの優劣の比較までは企画しなかった．東京湾水質測定データの場合，その長さは短く，水質状況が最近大きく変化した兆候も見られないため，方法が機能しさえすれば，季節成分モデルに多少の違いがあっても，ほぼ同様の結果が得られると予想されたからである．実際，そのような結果となり，モデリングの巾が広がった．勿論，各々の方法やモデルには特性がある．例えば，季節成分の水準が変化する場合には，Model BおよびCは利用できない．そして，Kitagawa and Gershのモデルは残る2つに比べやや頑健であり，Akaikeの方法とModel Aは初期分布の取扱い方と推定法に違いがある．一方，平均的な暦月毎の季節成分を仮定し推論したい場合には，当然ながら，Model BあるいはCが適している．両者の違いは，その「柔らかさ」にある．更に，こうした特性によってではなく，データ毎にモデルの優劣を客観的に評価したい場合には，何れも対数尤度が求まるので，対数尤度に基づく推論を組み立てれば良い．ただし，その場合，対数尤度をより厳密に扱う必要がある．例えば，最大対数尤度は，格子探査法のような粗い方法によってではなく，可変計量法等により精度良く求めなければならない．また，本稿では，異なるモデル間で初期分布を揃えなかったが，対数尤度が初期分布に依存する以上，初期分布を揃えるか，あるいは初期分布に依存しない対数尤度を求める工夫が必要となる．

現在，実際に稼働する時空間季節変動調整法の開発を優先しているため，モデルの厳密さを必ずしも追求してはいない．ただし，査読者からモデルの厳密さについて指摘があり，読者も同様の疑問を抱く可能性があると思われたので，以下に，これらの点について付記する．

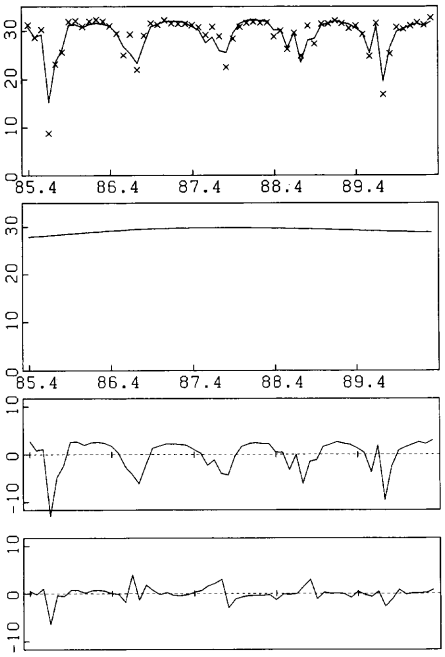

図 4. Model A の適用結果.

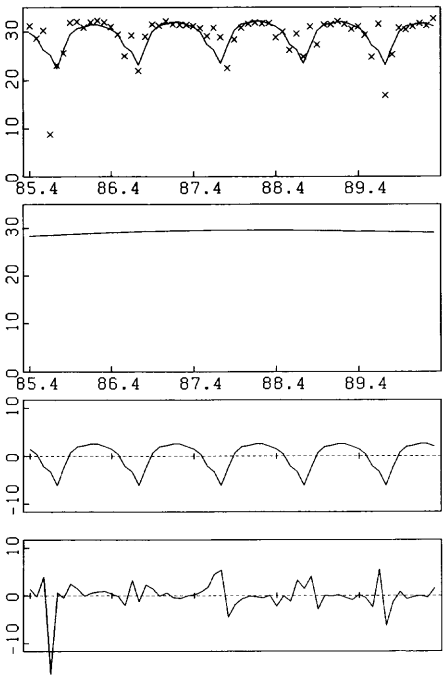

図 5. Model B の適用結果.

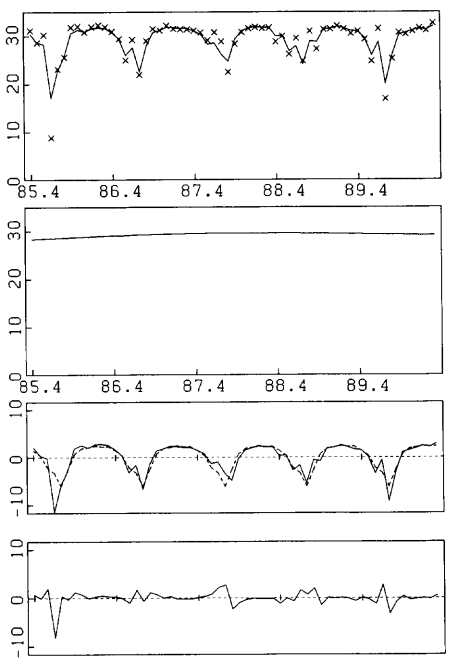

図 6. Model C の適用結果. ただし, 季節成分の図中の破線は季節成分の平均の平均値を示す.

先ず,「図 2 から 6 を見ると, 季節成分の変化の様相が 0 に対し非対称であり, しかも夏期の季節成分の変動が大きく, 他の推定値がこれに引きずられるので, 季節成分が対称となるようデータを非線形変換しては」との指摘があった. 季節成分の対称性について述べれば, 季節成

分が三角関数のように変化しなければならない理由はなく，モデルは対称性を前提にしていないし，対称となるようモデルを設定してもいない．季節成分モデル(1.3)では今期と前年同期の季節成分の差が平均的に0であり，(1.4)では1周期に渡る季節成分の和が平均的に0であるとしているだけである．もっとも，塩分の場合，外洋濃度をトレンドの基準に採れば，1周期に渡る季節成分の和が平均的に0となるようにしているので，その分トレンドがシフトしてはいる．一方，夏期の季節成分の変動が大きいのは事実で，それがシステムノイズの分散の推定値に影響を及ぼし，同時に他の変数や成分の推定値にも影響を及ぼしているのは確かである．そして，変動が大きくなるのは観測値が小さい時に限られているので，データを非線形変換すれば，見掛け上，変動を小さくできる．東京湾水質測定データの解析では，クロロフィルのデータで非線形変換の利用も検討しており，非線形変換の有用性を否定するつもりはない．ただ，非線形変換にはいくつかの問題がある．例えば，どの様な非線形変換を施すか，施した非線形変換の妥当性をどう判断するか，あるいは変換前と変換後に当てはめたモデル同士をどう比較するかといった問題である．また，原数値に興味の主体がある場合に非線形変換を利用すると，推定後，推定値を逆の非線形変換を用いて原数値に引き戻さなければならないが，ベイズ的方法や状態空間接近による場合，技術的な問題が生じる．例えば，対数変換後に正規モデルを当てはめた場合，原数値での平均値を求めるには，変換後の観測ノイズの分散の値も必要となるが，観測ノイズの分散は観測モデルだけでなくシステムモデルの残差も加えて推定されるため，推定値は過大評価される傾向にあり，対応して原数値での平均値も過大評価される．こうした問題と今後の解析の推移を考慮しながら，非線形変換の利用について判断したい．ただ，現時点では，モデルを工夫して対処する方法に興味を抱いている．

更に，「日本の土地環境を考えた場合，様々な要因の効果は，積分効果としてより測定日直前の効果として現れるのではないか．特に夏期にそうした傾向が見られる．そうした場合，降雨量を外的変数として取り込んだ方が良いのではないか．また，柔らかい季節成分モデルを用いる必要もないのではないか．」との指摘があった．確かに，1985年7月の異常値は台風に起因しており，直前の効果を否定できない．一方，図2から6のトレンドの推定値は期間降雨量の変化を捕捉しており，積分効果も一概に否定できない結果となっている．実際，晴天が続いたからといって，江戸川，荒川，多摩川といった大きな河川は滅多に干上がらない．低下したとはいえ土地にはまだ貯水機能があり，水利ダム等の影響もあるからである．決定論的に考えれば，月1回の測定頻度で直前の効果を無視できない状況では，単純な傾向分析には無理があるようにも思われるが，月1回とはいえ長期間継続的に測定していれば，観測の繰り返しと同様の効果が生じてくる．それ故，今回の不自然ではない適用結果が得られたと理解している．勿論，外的変数を導入すれば，モデルの当てはまりは良くなり，異常値も説明できるようになる．東京湾水質測定データの解析でも，降雨量を含めた気象との関係を把握するため，既に気象データの解析を開始している．ただし，傾向分析と回帰では解析の目的が異なる．東京湾水質測定データの解析では，塩分以外にも多くの測定項目を解析の対称としているため，汎用的な方法の開発を優先している．また，塩分だけに注目したとしても，東京湾を巡る水文は降雨量だけで規定されるわけではなく，降雨量にしても，関東地方とその周辺を含む広い地域を対称に考えなければならず，問題は極めて複雑である．更に，現在の計算機環境下では，時空間季節変動調整モデルを現在想定している以上に複雑にできないという事情もある．状況に応じ，解析を進めていきたい．なお，著者は必ずしも柔らかい季節成分モデルに固執していない．実際に稼働する時空間季節変動調整法を開発するため，特定のモデルに不具合が生じた場合に他のモデルに切り替えられるよう，手持ちのモデルを増やすのが今回の検討の目的のひとつでもあった．

謝　辞

最後に，有益かつ建設的なコメントを下さった査読者ならびに統計数理研究所の尾崎統氏に感謝の意を表します．

参 考 文 献

Akaike, H. (1980). Likelihood and the Bayes procedure, *Trabajos de Estadistica*, **31**, 143-166.

Akaike, H. and Ishiguro, M. (1980). BAYSEA, a Bayesian seasonal adjustment program, *Comput. Sci. Monographs*, No. 13, The Institute of Statistical Mathematics, Tokyo.

Ishiguro, M. (1984). Computationally efficient implementation of a Bayesian seasonal adjustment procedure, *J. Time Ser. Anal.*, **5**, 245-253.

Kashiwagi, N. (1993). On use of the Kalman filter for spatial smoothing, *Ann. Inst. Statist. Math.*, **45**, 21-34.

Kitagawa, G. (1996). A self-organizing state-space model, 平成7年度科学研究費補助金（一般研究C）研究成果報告書「時系列解析における数値的方法の研究」, 107-121.

Kitagawa, G. and Gersh, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality, *J. Amer. Statist. Assoc.*, **79**, 378-389.

Matusita, K., Kashiwagi, N., Aki, S. and Kuboki, H. (1988). Statistical analysis of air pollutant data with graphical methods; location characteristics of the monitoring stations, *Statistical Theory and Data Analysis II* (ed. K. Matusita), 59-70, North-Holland, Amsterdam.

二宮勝幸，柏木宣久，安藤晴夫(1996a)．東京湾における水温と塩分の空間濃度分布の季節別特徴，水環境学会誌，**19**，480-490．

二宮勝幸，柏木宣久，安藤晴夫(1996b)．東京湾におけるCODとDOの空間濃度分布の季節別特徴，水環境学会誌，**19**，741-748．

二宮勝幸，柏木宣久，安藤晴夫，小倉久子 (1997)．東京湾における容存性無機態窒素およびリンの空間濃度分布の季節別特徴，水環境学会誌，**20**，457-467．

Ozaki, T. and Thomson, P. J. (1994). A dynamical systems approach to X-11 type seasonal adjustment, Research Memo., No. 498, The Institute of Statistical Mathematics, Tokyo.

A Study on Seasonal Adjustment by State Space Approach: An Application to Tokyo Bay Data

Nobuhisa Kashiwagi

(The Institute of Statistical Mathematics)

Three models for seasonal adjustment are proposed. One is a state space model which corresponds to Akaike's Bayesian seasonal adjustment model, and the others are self-organizing state space models. The treatment of the initial state in the Kalman filter is also discussed. The proposed models as well as Akaike's and Kitagawa-Gersh's models are applied to time series of measurements on water quality in Tokyo Bay.

Key words: Akaike's Bayesian information criterion, Kalman filter, likelihood, seasonal adjustment, self-organizing state space model, smoothing.