

# ベイズ型時系列モデルによる成分分解を用いた ネットワークトラフィックの定量的分析法

生駒 哲一<sup>1</sup>・八名 和夫<sup>2</sup>

(受付 2006年1月16日; 改訂 2006年8月8日)

## 要 旨

インターネットが広く普及した現在, ネットワークトラフィックを分析してこれをネットワーク管理に役立てることが重要となってきた。その分析の一方法として, ネットワークトラフィックをトレンド, 1週間周期の変動, 1日周期の変動, および有色雑音成分に分解するベイズ型時系列モデルを用いる方法を提案する。提案法で用いるモデルは, これらの成分ごとのシステム方程式と, 各成分の和をネットワークトラフィックとする観測方程式とから成る線形ガウス型状態空間モデルで表現される。カルマンフィルタによる状態推定により各成分が求まる。モデルに含まれる分散やAR係数のパラメータは, 尤度の最大化により定める。またモデル間の比較や各成分の要不要の決定は, AIC(赤池情報量規準)に基づいて行う。提案法によるネットワークトラフィックの分析例として, 法政大学小金井キャンパスのダイヤルアップアクセス, および法政大学とSINETを結ぶネットワークに対して成分分解を行った結果について報告する。

キーワード: ネットワークトラフィック, 成分分解, ベイズ型時系列モデル, 状態空間モデル, カルマンフィルタ。

## 1. はじめに

現在, インターネットは広く普及し, 我々の生活に欠かせないものとなりつつある。インターネットを支える技術も日々進歩し続けており, そこを流れるトラフィックも変化し多様になってきている。一方, 企業や大学等の組織においては, 組織内ネットワークを管理運営する必要がある。そのためには, トラフィックを解析してその特性の把握や将来の予測をし, 適切な設備更新やあるいは動的に経路を振り分ける等の対応をとることになる。

図1に, ネットワークトラフィックの一例を示す。これは, 法政大学小金井キャンパスに設置されたダイヤルアップアクセスポイントにおいて, 5分ごとに下り(外部から法政大学へ)のトラフィック量(単位: bit/sec)を記録したものである。期間は2001年5月28日(月)から同年7月8日(日)の6週間である。これは前期授業が定期的に行われている期間である。また, この期間に祝日はない。この期間のうちの最初の1週間は図2に示す。図2は, 2001年5月28日の0時から始まっている。

このトラフィックをみてまず判ることは, 顕著な1日周期の変動があることである。この1日周期の変動は, トラフィックが電話回線を経由したアクセスであることから, 電話料金が格安にな

<sup>1</sup>九州工業大学 工学部電気工学科: 〒804-8550 福岡県北九州市戸畑区仙水町1番1号

<sup>2</sup>法政大学 工学部電子情報学科: 〒184-8584 東京都小金井市梶野町3-7-2

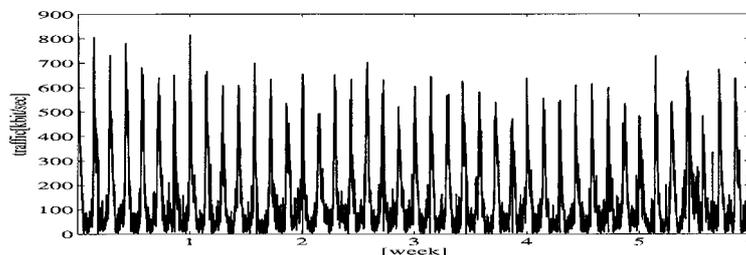


図1. 法政大学小金井キャンパスのダイヤルアップアクセストラフィック：2001年5月28日(月)～7月8日(日)(6週間)。

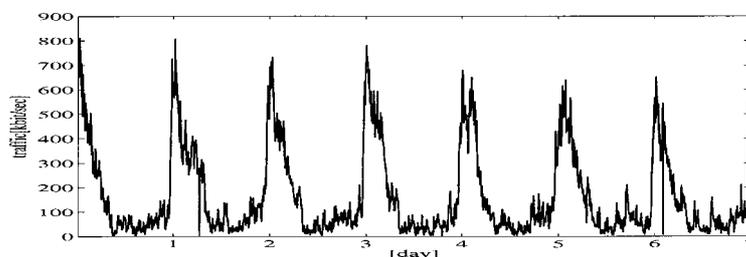


図2. 法政大学小金井キャンパスのダイヤルアップアクセストラフィックの第1週目。

る深夜にアクセスが集中し、昼間は電話料金が低い事や利用者の多くが大学に居てアクセスが少ない事等が、その要因として考えられる。また、日によって、アクセスの多い少ないの違いが見られる。全体的な傾向としても、1~2週目は多いトラフィックが、その後半ではやや少なめになってきているのも判る。そして最終週の中頃にて、トラフィックの一時的な上昇もみられる。これらの変動傾向は、人間の視覚によって定性的に解ることである。これらを定量的に捉えるためには、何らかのモデルに基づく分析を行う必要がある。

ネットワークトラフィック時系列の統計的モデル化に関する研究は、Leland et al. (1994)、Paxon and Floyd (1995)による報告以来数多くあり、大別すると、現象的なモデル化と構成的なモデル化に分類される。現象的なモデル化においては、トラフィック時系列を単一の時系列とみなし、その自己相似性、長相関性などが分析され、それらの特徴を有する FARIMA モデル (Resnick, 1997) 等が提案されている。しかしながら、本論文で取り上げるようにトラフィック時系列を特徴的なマクロな成分に分解して定量的に解析した例はない。構成的なモデル化 (Willinger et al., 1997; Yang and Petropulu, 2001) においては、トラフィックを要素となる成分に分解するが、パケット単位のミクロな成分分解であり、本論文で取り扱う成分分解とは異なる。

本研究では、上述のような特徴をもつネットワークトラフィックを解析の対象とする。このようなネットワークトラフィックを、トレンド、1週間周期の変動、1日周期の変動等の成分に分解する事を目的とする。そのために、時系列をトレンド、循環成分などに分解するベイズ型時系列モデル (Kitagawa and Gersch, 1984) に基づく、新しい分析手法を提案する。

さて、Kitagawa and Gersch (1984) で提案されているベイズ型時系列モデルは、時系列データをいくつかの特徴的な成分に分解することができる。成分としては、トレンド、季節性、曜日効果、AR 成分などがある。このモデルは経済データにおける季節調整を念頭において開発

されたもので、典型的には月次データや日次データを扱い、季節性は12ヶ月や1週間を周期とする成分を意図している。また曜日効果は、月次データの際に使われる成分で、各月に含まれる同一曜日の日数の違いを表すものである。

ネットワークトラヒックに対しては、単に Kitagawa and Gersch (1984) で挙げられている季節調整モデルをそのまま適用すれば良い訳ではない (高橋 他, 2001)。その理由は、一つにはトラヒックの持つ周期性が単一ではなく、1日周期の変動と1週間周期の変動という、二つの周期性を持つことが挙げられる。またもう一つの理由としては、トレンド自体も、日内の短いスパンでのトレンドと、数日に渡るような長期的なトレンドとがあり得るからである。さらに、これら以外の成分として、有色の雑音成分があると考えられる。よって、ネットワークトラヒック分析のためには、以上の成分を含むように Kitagawa and Gersch (1984) のモデルを拡張する必要がある。

関連する類似の研究としては、まず月次データに対する季節調整として、例えば電話料金収入の系列について、料金改定の効果項を入れたモデル (阿部・上田, 1985) や、さらに曜日効果も考慮したモデル (Yada et al., 1994) がある。これらは単一の周期 (12ヶ月) 成分とトレンド成分 (月に関する階差) のみを扱っており、本研究で課題としている二つの周期性やトレンドは考慮していない。また、日次データの系列に関する成分分解としては、電話の通信トラヒックに対して、電話料金改定の効果を考慮した研究 (矢田・北川, 1994) があり、そこでは料金改定をモデル構造の変化点として考え、観測ノイズ分散を大きくしてロバストな推定を試みたり、非ガウス型フィルタを用いた構造変化の推定を試みている。これら三つの研究のうち、最初の二つは電話料金収入の系列に対するモデル化である。三つめはトラヒック系列のモデル化ではあるが、電話料金改正に伴う収入の変化に主眼を置いたものである。また、データ系列のサンプリング間隔も月次や日次と長く、時々刻々変化するトラヒックを扱うには不十分である。

以上の事柄や背景をふまえて、本論文では、Kitagawa and Gersch (1984) で提案されているベイズ型時系列モデルに基づき、上述したネットワークトラヒックの性質を考慮した新しい分析手法を提案する。提案手法により、法政大学のダイヤルアップアクセス (清水 他, 2003; Ikoma et al., 2003; Shimizu et al., 2003) および SINET のトラヒック (今津 他, 2004) の分析を行う。得られた結果を吟味し、トラヒックの性質について考察する。

## 2. モデル

1時間をサンプリングタイムとするトラヒックデータを想定する。トラヒック量 (単位:kbit/sec) の対数をとったものを解析対象とする。対数を取る理由は、トラヒックデータは正の値を取るため、これを実数域に変換する為である。これらを満たすために、前処理として、まずトラヒックの原系列 (5分サンプリング) 12期の平均を計算して1時間サンプリングの系列を得、その後対数をとるものとする。得られた系列を図3に示す。以下では、時刻 (単位:hour) を整数  $k$  で表し、解析に用いるトラヒックデータを  $y(k)$  と表すことにする。

### 2.1 トrend成分

成分の階差がほぼ0となる成分を「トレンド成分」と呼ぶものとする。ただし階差は日を単位として取るものとする。これは、例えば階差が1の場合には、前日の値と本日の値との差がほぼ0となるものをトレンド成分とすることになる。以下では特に、トレンドの階差が2の場合について説明する。他の階差についても、同様に定義することができる。

トレンド成分は、システム方程式

$$(2.1) \quad x_T(k) = F_T(k)x_T(k-1) + G_T(k)v_T(k)$$

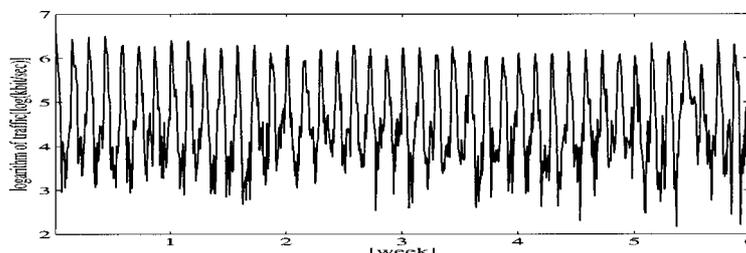


図 3. ダイヤルアップアクセストラヒックの対数系列 .

に従うものとして定義される . ここで  $v_T(k)$  はシステムノイズで , 正規分布  $N(0, \tau_T^2)$  に独立に従うものとする . 式 (2.1) の各要素は , まず

$$(2.2) \quad x_T(k) = [ T_1(k) \ T_2(k) ]'$$

は , トレンド成分を時間順にその階差の日数分だけ並べた状態ベクトルである . 次に行列  $F_T(k)$  は , 時刻  $k$  により 2 通りの値をとり ,  $k = 0(\text{mod } 24)$  の時には

$$(2.3) \quad F_T(k) = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$$

とする . ここで行列の 1 行目は階差 2 を取るための係数であり , 2 行目以下は要素のシフトの役割を果たす .  $k \neq 0(\text{mod } 24)$  の時には  $F_T(k)$  は単位行列とする . そしてベクトル  $G_T(k)$  は , 時刻  $k$  により 2 通りの値をとり ,  $k = 0(\text{mod } 24)$  の時には

$$(2.4) \quad G_T(k) = [ 1 \ 0 ]'$$

であり ,  $k \neq 0(\text{mod } 24)$  の時には  $G_T(k)$  は零ベクトルとする .

行列  $F_T(k)$  および  $G_T(k)$  を , 上述のように時間変化するものとはせず , 時刻  $k$  によらず一定で式 (2.3) (2.4) の値を取るものとする . 階差を時間 (hour) としたトレンド成分を定義することができる . このトレンド成分の状態ベクトルを  $x_t(k) = [ t_1(k) \ t_2(k) ]'$  , システムノイズを  $v_t(k) \sim N(0, \tau_t^2)$  , 上述の行列を  $F_t, G_t$  と表すことにする . 必要に応じて , このような短いスパンのトレンドを用いることができる .

## 2.2 曜日成分

7 日を周期とする周期的変動成分を「曜日成分」と呼ぶものとする . 曜日成分は , システム方程式

$$(2.5) \quad x_D(k) = F_D(k)x_D(k-1) + G_D(k)v_D(k)$$

に従うものとして定義される . ここで  $v_D(k)$  はシステムノイズで , 正規分布  $N(0, \tau_D^2)$  に独立に従うものとする . 式 (2.5) の各要素は , まず

$$(2.6) \quad x_D(k) = [ D_1(k) \ D_2(k) \ \cdots \ D_6(k) ]'$$

は , 曜日成分を 6 日分並べた状態ベクトルである . 次に行列  $F_D(k)$  は , 時刻  $k$  により 2 通り

の値をとり,  $k=0(mod\ 24)$  の時には

$$(2.7) \quad F_D(k) = \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 & 0 \end{bmatrix}$$

であり,  $k \neq 0(mod\ 24)$  の時には  $F_D(k)$  は単位行列とする. そしてベクトル  $G_D(k)$  は, 時刻  $k$  により 2 通りの値をとり,  $k=0(mod\ 24)$  の時には

$$(2.8) \quad G_D(k) = [1\ 0\ \cdots\ 0]'$$

であり,  $k \neq 0(mod\ 24)$  の時には  $G_D(k)$  は零ベクトルとする.

なお, 祝日を含む週については, 次のような扱いが可能である. 祝日のトラヒック量を日曜のトラヒック量と同じと考えることにする. 式(2.7)の行列は, 曜日成分のベクトルに適用すると, 曜日成分のベクトルが 1 日分シフトするという性質がある. よって, 翌日が祝日の場合には, 式(2.7)を複数回適用(行列のべき乗を用いる)し, 曜日成分のベクトルの第 1 要素が日曜日になるようにすればよい. ただし, どの要素が日曜に対応するかを把握しておく必要がある.

### 2.3 日内変動成分

24 時間を周期とする周期的変動成分を「日内変動成分」と呼ぶものとする. 日内変動成分は, システム方程式

$$(2.9) \quad x_s(k) = F_s x_s(k-1) + G_s v_s(k)$$

に従うものとして定義される. ここで  $v_s(k)$  はシステムノイズで, 正規分布  $N(0, \tau_s^2)$  に独立に従うものとする. 式(2.9)の各要素は, まず

$$(2.10) \quad x_s(k) = [s_1(k)\ s_2(k)\ \cdots\ s_{23}(k)]'$$

は, 日内変動成分を 23 時間分並べた状態ベクトルである. 次に行列  $F_s$  は

$$(2.11) \quad F_s = \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 & 0 \end{bmatrix}$$

であり, ベクトル  $G_s$  は

$$(2.12) \quad G_s = [1\ 0\ \cdots\ 0]'$$

とする.

### 2.4 AR 成分

有色の雑音成分を表すために, 自己回帰過程に従う成分を考える.  $p$  次の自己回帰過程に従う成分を「AR 成分」と呼ぶものとする. AR 成分は, システム方程式

$$(2.13) \quad x_a(k) = F_a x_a(k-1) + G_a v_a(k)$$

に従うものとして定義される．ここで  $v_a(k)$  はシステムノイズで，正規分布  $N(0, \tau_a^2)$  に独立に従うものとする．式(2.13)の各要素は，まず

$$(2.14) \quad x_a(k) = [c_1(k) \ c_2(k) \ \cdots \ c_p(k)]'$$

は，AR 成分をその次数の分だけ並べた状態ベクトルである．次に行列  $F_a$  は，

$$(2.15) \quad F_a = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 & 0 \end{bmatrix}$$

で，ここで  $a_1, a_2, \dots, a_p$  は自己回帰係数である．そしてベクトル  $G_a$  は，

$$(2.16) \quad G_a = [1 \ 0 \ \cdots \ 0]'$$

とする．

### 2.5 観測方程式

各成分の和をネットワークトラヒックとする．これを観測方程式

$$(2.17) \quad y_k = T_1(k) + D_1(k) + s_1(k) + c_1(k) + w_k$$

で表す．ここで  $w_k$  は観測ノイズで，正規分布  $N(0, \sigma^2)$  に独立に従うものとする．

### 2.6 状態空間表現

これまで定義してきたシステム方程式，観測方程式をまとめて，次の状態空間表現を得る．

$$(2.18) \quad \begin{cases} x_k = F_k x_{k-1} + G_k v_k \\ y_k = h x_k + w_k \end{cases}$$

ここで状態ベクトル  $x_k$  は，各成分の状態ベクトル  $x_T(k)$ ， $x_D(k)$ ， $x_s(k)$ ，および  $x_a(k)$  を並べたものであり，またシステムノイズベクトル  $v_k$  は，各成分のシステムノイズ  $v_T(k)$ ， $v_D(k)$ ， $v_s(k)$ ，および  $v_a(k)$  を並べた列ベクトルである．状態遷移行列  $F_k$  は，各成分の状態遷移行列  $F_T(k)$ ， $F_D(k)$ ， $F_s$ ，および  $F_a$  を対角要素に持つブロック対角行列とする．また行列  $G_k$  は，各成分の列ベクトル  $G_T(k)$ ， $G_D(k)$ ， $G_s$ ，および  $G_a$  を対角要素に持つブロック対角行列である．行ベクトル  $h$  は，式(2.17)が成り立つように定義するものとする．

このように状態空間表現されたモデルを，状態空間モデルと呼ぶ．式(2.18)の状態空間モデルは，全てのノイズが正規分布に従い，また線形式であるので，線形ガウス型状態空間モデルのクラスに属する．

## 3. モデルの推定

### 3.1 状態推定

状態推定とは，状態空間モデルにおいて，観測の系列  $Y_k = (y_1, y_2, \dots, y_k)$  が与えられた下での状態の条件付分布を求めることである．状態推定は，与えられる観測系列の時刻(表記  $Y_k$  の  $k$ )と推定すべき状態の時刻との関係に応じて，「予測」，「ろ波」，「平滑化」の3つに分類することができる．

ろ波とは，観測系列の時刻と状態の時刻が等しい場合をいう．例えば，現在までの全ての観測データが与えられた下で，現在の状態を推定することは，ろ波である．予測とは，状態の時

刻が、観測系列の時刻よりも未来になっている場合である。平滑化とは、予測の逆の状況で、状態の時刻が、観測系列の時刻よりも過去になっている場合である。つまり平滑化は、現在までの観測データを使って、過去の状態を推定していることに相当する。求める条件付分布は、時刻  $k$  の状態の推定に関しては、1 期先予測が  $p(x_k|Y_{k-1})$ 、ろ波が  $p(x_k|Y_k)$ 、そしてラグ  $L > 0$  の平滑化が  $p(x_k|Y_{k+L})$  となる。

本論文で用いるモデルは、式(2.18)のように、線形ガウス型状態空間モデルである。線形ガウス型状態空間モデルは、状態の分布がガウス分布になるので、その平均ベクトルと共分散行列を求めるカルマンフィルタにより、状態推定が可能である。また平滑化についても、カルマンフィルタと同様に、平均ベクトルと共分散行列を求めるアルゴリズムが提案されている。以下にこれらのアルゴリズムを簡潔に示す。

### 3.2 カルマンフィルタ

状態の初期分布を適切に設定し、これを  $p(x_0)$  と表すものとする。初期分布  $p(x_0)$  はガウス分布で、その平均ベクトルが  $x_0$ 、共分散行列が  $V_0$  であるものとする。

状態推定のための平均ベクトル、共分散行列を次のように初期化する。

$$(3.1) \quad \begin{cases} x_{0|0} = x_0 \\ V_{0|0} = V_0 \end{cases}$$

ここで左辺の添え字は、縦線の前が状態の時刻、後が観測系列の時刻を表す。

時刻を  $k=1$  から開始し、これを 1 時刻ずつ進めながら、以下で示す 1 期先予測とろ波の手続きを交互に繰り返す。

#### 1 期先予測

$$(3.2) \quad \begin{cases} x_{k|k-1} = F_k x_{k-1|k-1} \\ V_{k|k-1} = F_k V_{k-1|k-1} F_k' + G_k Q G_k' \end{cases}$$

ここで正方行列  $Q$  は、システムノイズベクトル  $v_k$  の共分散行列である。

観測の 1 期先予測  $p(y_k|Y_{k-1})$  もガウス分布に従い、その平均  $\bar{y}_{k|k-1}$  と分散  $\Sigma_{k|k-1}$  は

$$(3.3) \quad \begin{cases} \bar{y}_{k|k-1} = h x_{k|k-1} \\ \Sigma_{k|k-1} = h V_{k|k-1} h' + \sigma^2 \end{cases}$$

となる。

#### ろ波

$$(3.4) \quad \begin{cases} x_{k|k} = x_{k|k-1} + k_k e_k \\ V_{k|k} = (I - k_k h) V_{k|k-1} \end{cases}$$

ここで、 $I$  は単位行列、 $e_k$  は観測の 1 期先予測(平均値)の誤差

$$(3.5) \quad e_k = y_k - \bar{y}_{k|k-1}$$

であり、また  $k_k$  はカルマンゲインと呼ばれ、

$$(3.6) \quad k_k = V_{k|k-1} h' (\Sigma_{k|k-1})^{-1}$$

と計算される。

### 3.3 平滑化

本論文では、固定区間平滑化と呼ばれる平滑化法を用いる。この方法は、観測データ系列を固定して、全ての時刻の状態の平滑化推定を求める。そのためには、観測データ系列の全ての時刻について、上で述べたる波と 1 期先予測を予め計算し記憶しておく必要がある。

全観測データ系列を  $Y_N = (y_1, y_2, \dots, y_N)$  とする。固定区間平滑化は、データ系列の末尾から開始し、データ系列の先頭方向に 1 時刻ずつ遡りながら、次の手続きを繰り返し適用する。

$$(3.7) \quad \begin{cases} \mathbf{x}_{k|N} = \mathbf{x}_{k|k} + \mathbf{A}_k (\mathbf{x}_{k+1|N} - \mathbf{x}_{k+1|k}) \\ \mathbf{V}_{k|N} = \mathbf{V}_{k|k} + \mathbf{A}_k (\mathbf{V}_{k+1|N} - \mathbf{V}_{k+1|k}) \mathbf{A}_k' \end{cases}$$

ここで行列  $\mathbf{A}_k$  は、次の通り計算する。

$$(3.8) \quad \mathbf{A}_k = \mathbf{V}_{k|k} \mathbf{F}_{k+1}' \mathbf{V}_{k+1|k}^{-1}$$

### 3.4 パラメータの推定

上述のモデルで状態推定を行う際には、観測ノイズの分散  $\sigma^2$  とシステムノイズの分散  $\tau^2 = (\tau_T^2, \tau_D^2, \tau_s^2, \tau_a^2)$ 、及び AR 係数  $\mathbf{a} = (a_1, a_2, \dots, a_p)$  を与える必要がある。これらのパラメータをまとめて  $\theta$  と表すものとする。

パラメータ  $\theta$  は、尤度を最大化することにより求める。最大化には、数値的最適化手法(グッドサーチ, 準ニュートン法等)を用いる。尤度関数の対数を取ると、

$$(3.9) \quad \begin{aligned} l(\theta) &\equiv \log p(Y_N; \theta) \\ &= \sum_{k=1}^N \log p(y_k | Y_{k-1}; \theta) \\ &= -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{k=1}^N \log \Sigma_{k|k-1} \\ &\quad - \frac{1}{2} \sum_{k=1}^N \frac{(y_k - \bar{y}_{k|k-1})^2}{\Sigma_{k|k-1}} \end{aligned}$$

となる。最後の式の各要素は、カルマンフィルタの 1 期先予測の過程(式(3.3))で得られる。

ここで扱っているように観測が 1 変量の場合には、観測ノイズの分散を 1 としたモデルの尤度を最大化することで、パラメータを推定することができる(北川, 1993)。具体的には、観測ノイズの分散を 1 としたモデルを想定し、これにシステムノイズ分散  $\tau^2$  を与えカルマンフィルタを適用する。その結果から、元のモデルの観測ノイズ分散の推定値  $\hat{\sigma}^2$  及びシステムノイズ分散  $\tau^2 = \hat{\sigma}^2 \tilde{\tau}^2$  と、対数尤度関数値とを得ることができる。この手順では、AR 係数  $\mathbf{a}$  の扱いについては、変更は生じない。

観測ノイズの分散を 1 としたモデルに基づく、観測の 1 期先予測の平均と分散を、それぞれ  $\tilde{y}_{k|k-1}$ ,  $\tilde{\Sigma}_{k|k-1}$  と表記する。ここで、 $\bar{y}_{k|k-1} = \tilde{y}_{k|k-1}$  および  $\Sigma_{k|k-1} = \sigma^2 \tilde{\Sigma}_{k|k-1}$  となることから導かれる( $\sigma^2$  は元のモデルの観測ノイズ分散である)。これらを式(3.9)に代入し、偏微分して 0 とおくと、元のモデルの観測ノイズ分散  $\sigma^2$  の推定値

$$(3.10) \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^N \frac{(y_k - \tilde{y}_{k|k-1})^2}{\tilde{\Sigma}_{k|k-1}}$$

を得る。これを再び式(3.9)に代入すると、対数尤度関数は

$$(3.11) \quad l(\tilde{\tau}^2, \mathbf{a}) = -\frac{N}{2} \log 2\pi \hat{\sigma}^2$$

$$-\frac{1}{2} \sum_{k=1}^N \log \tilde{\Sigma}_{k|k-1} - \frac{N}{2}$$

となる．これを最大化して，パラメータ  $\hat{\tau}^2$ ， $\hat{a}$  を求める．

### 3.5 モデル選択

異なるモデルが存在してそれらの比較をする場合や，モデルの成分の要不要を決定する場合，もしくはモデル中のトレンド階差や AR 次数を決定する場合には，AIC (赤池情報量規準) (Akaike, 1974)

$$(3.12) \quad \text{AIC} = -2l(\hat{\theta}) + 2 \times (\theta \text{の自由度})$$

を用いる．候補モデル中で AIC の値の最も小さいモデルを，最も良いモデルとして選択する．AIC は，パラメータの自由度が比較するモデル間で異なる事を考慮した規準である．これは，解析に用いるモデルと真のモデルとの分布 (観測値の予測分布) の相違をカルバック-ライブラー情報量により計るものとし，適切な仮定の下でその近似値を与えるものである．

## 4. トラヒックデータの解析

ネットワークトラヒックの解析結果について報告する．解析は，提案法で示したモデルの他に，成分の適切な組み合わせから成るモデルについても行った．成分としては，トレンド成分 (日についての階差  $T$ ，および，時間についての階差  $t$ )，曜日成分  $D$ ，日内変動成分  $s$ ，そして AR 成分  $c$  を用いた．トレンド成分の階差は 1 または 2 とし，AR 成分の次数は 1 または 2 とした．トレンド成分の階差が異なる場合や，AR 成分の次数が異なる場合は，別のモデルとみなすものとする．

パラメータの推定では，分散パラメータの数値的最適化法として，2 段階のグリッドサーチを用いた．1 段階では，分散パラメータの指数部分を探索し，2 段階では 1 桁の有効数字を探索した．また AR 係数の推定もグリッドサーチで行い，偏自己相関係数 (PARCOR) を用いて，PARCOR を  $(-1, 1)$  の範囲で 0.05 刻みでグリッドを構成し探索した．分散パラメータのグリッドサーチと PARCOR のグリッドサーチは個別に交互に行い，両グリッドにて最適解が見つかり収束するまで探索を行う方法を探った．ただし分散パラメータには最小値 (ここでは  $10^{-10}$ ) を設定し，これより小さな値については探索しないものとした．

初期分布の平均  $x_0$  と共分散行列  $V_0$  については，その適切な値の求め方はいくつか知られている (1. パラメータとして扱い，尤度最大化により求める．2. 適当な初期分布にてフィルタリングと平滑化を一旦行い，その平滑化推定値を初期分布に用いる．等)．ここでは簡便な方法として，次の通り値を与えることにした．まず，成分は互いに独立と仮定した．次に各成分内では，成分を構成する状態ベクトルの要素ごと独立に正規分布に従うものとした．各成分で用いた正規分布は次の通りである．トレンドについては，平均をデータ系列の先頭 24 時刻 (すなわち 24 時間 = 1 日) の算術平均とし，分散を 1 とした．曜日成分については平均 0 分散 1 とし，AR 成分については平均 0 分散 1 とした．日内変動成分については，尤度に対して比較的大きな影響を与えることが実験的に分かったので，それを考慮し，平均 0 分散  $\tau_{s_0}^{-2}$  と設定し， $\tau_{s_0}^2$  としては適宜 10 や  $10^2$  を与えるものとした．

以下，法政大学小金井キャンパスのダイアルアップアクセストラヒックと，法政大学と SINET を結ぶ SINET トップトラヒックについて，それぞれの解析結果を述べる．

### 4.1 ダイアルアップアクセストラヒック

図 1 (第 1 週目は図 2) の小金井キャンパスのダイアルアップアクセスのトラヒック (5 分間隔

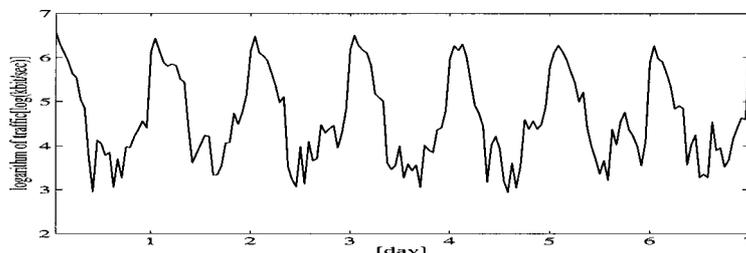


図 4. ダイアルアップアクセストラフィック対数系列の第 1 週目 .

表 1. ダイアルアップアクセストラフィックの最大対数尤度  $l(\hat{\theta})$  と AIC .

Model	$l(\hat{\theta})$	AIC
$T(1)+D+s+c(1)$	-514.64	1039.29
$T(2)+D+s+c(1)$	-518.96	1047.92
$T(1)+D+s+c(2)$	-513.59	1039.17
$T(2)+D+s+c(2)$	-518.11	1048.22
$T(1)+D+s$	-631.69	1269.39
$T(2)+D+s$	-639.43	1284.85
$t(1)+D+s$	-583.25	1172.51
$t(2)+D+s$	-642.14	1290.29
$T(1)+s$	-626.84	1257.69
$T(2)+s$	-644.67	1293.34
$t(1)+s$	-576.33	1156.67
$t(2)+s$	-652.20	1308.41

表 2. ダイアルアップアクセストラフィック :  $T(1)+D+s+c(2)$  モデルにて推定されたパラメータ .

Hyper-parameter	Estimated value	Value for grid search
$\sigma^2$	$5.79 \times 10^{-2}$	1
$\tau_T^2$	$5.79 \times 10^{-12}$	$1 \times 10^{-10}$
$\tau_D^2$	$2.32 \times 10^{-4}$	$4 \times 10^{-3}$
$\tau_s^2$	$1.74 \times 10^{-4}$	$3 \times 10^{-3}$
$\tau_c^2$	$5.79 \times 10^{-2}$	$1 \times 10^0$
$a_1$	0.84	0.7
$a_2$	-0.2	-0.2

で計測)について, 12 期の平均を計算しその対数を取った系列が図 3 である. この系列を解析対象とし(その第 1 週目を図 4 に示す), 提案法のモデル及びその他のモデルを適用した. ここで, 日内変動の初期分散は  $\tau_{s_0}^2 = 10$  とした.

各モデルの最大対数尤度  $l(\hat{\theta})$  と AIC を, 表 1 に示す. 表中のモデル名の箇所にて, 括弧内の数字は, トレンドの階差, および AR 次数を表す. 表 1 を見ると, 提案法のモデル族が他のモデルと比べて良い AIC 値であることが判る. 特に, AR 成分を持つモデルと持たないモデルとで, AIC 値に大きな違いがあることが見てとれる. これは, トラフィックが有色雑音成分を含

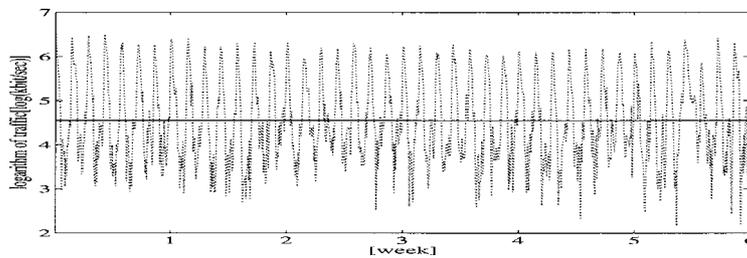


図 5. ダイヤルアップアクセストラフィック：トレンド成分の推定結果(平滑化の平均：実線)と解析系列(破線)。

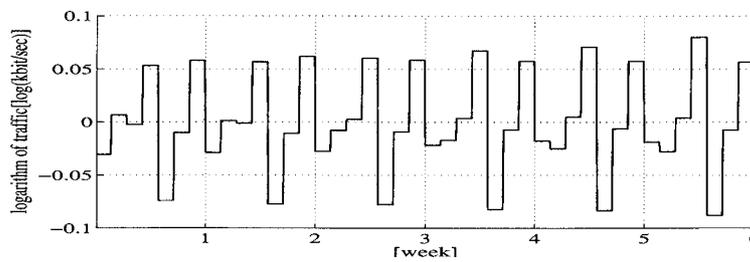


図 6. ダイヤルアップアクセストラフィック：曜日成分の推定結果(平滑化の平均)。

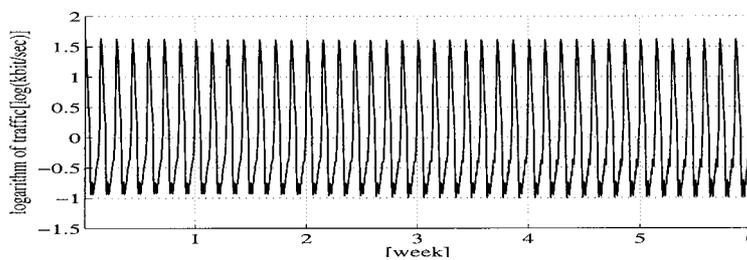


図 7. ダイヤルアップアクセストラフィック：日内変動成分の推定結果(平滑化の平均)。

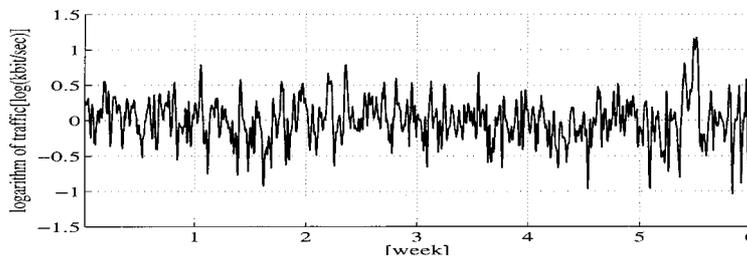


図 8. ダイヤルアップアクセストラフィック：AR 成分の推定結果(平滑化の平均)。

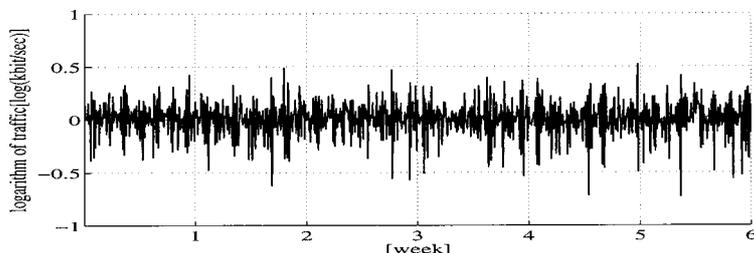


図9. ダイヤルアップアクセストラヒック：推定の結果(平滑化の平均)，得られた残差．

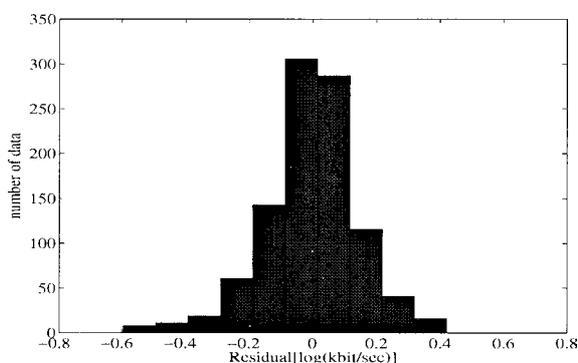


図10. ダイヤルアップアクセストラヒック：推定の結果(平滑化の平均)得られた残差のヒストグラム．

むことを示唆していると考えられる．提案法のモデル族のうち，トレンド成分(日についての階差)  $T$  の階差が1のものが良く，そのうちAR次数が2のモデルが微小ながら良い結果であることが判る．しかしこの微小な差は有意ではないと考えられるので，状況によってはAR次数が1のモデルを選択することもあり得る．

この最も良いモデル  $T(1)+D+s+(2)$  にて得られたパラメータ値を表2に，推定された成分(平均)を図5, 6, 7, および8に示す．また，推定された成分(平均)の和と，トラヒック系列との差(残差)を，図9に示す．

以下，推定結果について考察する．まず，推定されたトレンド成分(図5)は，ほぼ一定値となっている．これは，トレンド項システムノイズ分散  $\tau_T^2$  が，探索グリッドの最小値  $10^{-10}$  となっていることと符合する．すなわち，本トラヒックデータに関しては，トレンド成分は確定的な(確率的要素を含まない)一定値とみなせることがわかる．

次に曜日成分(図6)についてみると，大半の場合，木曜日が最もトラヒックが多く，次にトラヒックが多いのは日曜日である．水曜土曜はその次，月曜火曜がそれに続くトラヒック量を持ち，金曜日が最も少ないことがわかる．

日内変動成分(図7)についてみると，深夜0時の少し前から徐々にトラヒック量が増え始め，その後減ってゆくという変動パターンが得られている．これは，ダイヤルアップアクセスが23時以降の電話料金が格安になる時間帯に集中していることを，成分として抽出したものと考えられる．

そして，AR成分(図8)をみると，他の3つの成分では表しきれなかった微細な変動を，成

分として抽出していることがわかる．特に，最終週中盤の一時的なトラヒック上昇において，その変動が顕著に抽出されている．提案法のモデルでは，推定された AR 成分が，数時間程度の短期的なトレンドを表すものと考えることができる．

最後に，残差(図 9)が正規分布に従うかどうかを調べた．残差のヒストグラムを図 10 に示す．図 10 をみると，やや左右非対称な単峰の分布であることが判る．残差の平均は  $-0.0056$ ，標準偏差は  $\sigma = 0.1468$  である． $-4\sigma$  の以下の範囲にも残差値が 7 点存在している．これら 7 点を外れ値とみなしてカイ二乗適合度検定を行ったところ，値  $Q = 65.10$  が得られた．ヒストグラムの階級数 10，正規分布のパラメータ数 2 より，カイ二乗分布の自由度は  $(10 - 1) - 2 = 7$  である．自由度 7 のカイ二乗分布の上側 5% 点は 14.067，上側 1% 点は 18.475 であることから，得られた  $Q$  の値では残差が正規分布に従うとは言えない．

残差が正規分布に従わない原因を考察してみると，提案法で用いたモデルには，トレンド成分( $T$  成分)，曜日成分( $D$  成分)といった，日付変更時に値が切り換わる成分を持ち，これが原因の一つではないかと考えられる．すなわち，日付変更時にこれら成分の値が急に变化する為，AR 成分が日付変更時にレベルシフトを伴うこととなり，また残差成分も日付変更時に大きな値を取るようになる．実際，図 8 の AR 成分を見ると，日付変更時にレベルシフトが生じており，図 9 の残差成分を見ると，日付変更時に大きな値を取っていることがわかる．別の原因としては，線形ガウスモデルを適用する為に原データに適用している対数変換が考えられる．対数変換ではなく，別の変換，例えば Box-Cox 変換(Box and Tiao, 1973)などを用いることで，これを改善できる可能性がある．

#### 4.2 SINET トップトラヒック

図 11 に，法政大学と SINET とを結ぶネットワークのトラヒックを示す．これもダイヤルアップアクセストラヒック同様，5 分ごとに下り(外部から法政大学へ)のトラヒック量(単位: bit/sec)を記録したものであり，期間は 2001 年 5 月 28 日(月)時から同年 7 月 8 日(日)の 6 週間である．この期間のうちの最初の 1 週間を図 12 に示す．

乱れの多いデータ系列ではあるが，顕著な 1 日周期の変動と，1 週間周期の変動パターンを見ることができる．この 1 日周期の変動は，利用者の多い昼間にアクセスが集中し，夜間は少なくなっている．また 1 週間の変動パターンについては，日曜日が顕著に少ないトラヒックであることがわかる．全体的な傾向としても，2~3 週目でトラヒックが一旦少なめになり，その後回復してやや多くなっているのが判る．

この SINET トップのトラヒック系列を，ダイヤルアップアクセストラヒック同様，12 期の平均を計算した後対数を取り，得られた系列を解析の対象とする．解析対象の系列と，その最初の 1 週間を，それぞれ図 13，図 14 に示す．

この解析系列に，提案法で示したモデルと，適切な成分の組み合わせから成るモデルを適用した．ここで，日内変動の初期分散としては， $T+D+s+c$  モデル族(提案法のモデル族)については  $\tau_{s_0}^2 = 10^2$  とし，その他のモデルについては  $\tau_{s_0}^2 = 10$  とした．

得られた最大対数尤度と AIC の値を，モデルごとにそれぞれ，表 3 に示す．表を見ると，提案法のモデル族が良い AIC 値を得ている．特に，ダイヤルアップアクセストラヒックと同様に，AR 成分を持つモデルと持たないモデルとで AIC 値に大きな違いがあり，トラヒックが有色雑音成分を含むことを示唆していると考えられる．提案法のモデル族の中でも，トレンド階差 1 のモデルが良い AIC 値となっていて，そのうち，AR 次数が 2 のモデル  $\mathcal{I}(1)+D+s+(2)$  が最も良い AIC 値を得ている．最良の  $\mathcal{I}(1)+D+s+(2)$  モデルについて，2 段階のグリッド

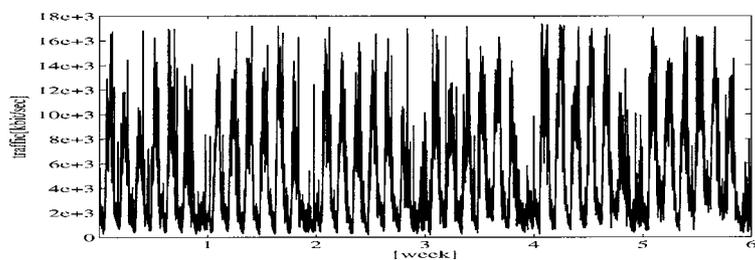


図 11. 法政大学 SINET トップトラフィック : 2001 年 5 月 28 日(月)~7 月 8 日(日)(6 週間).

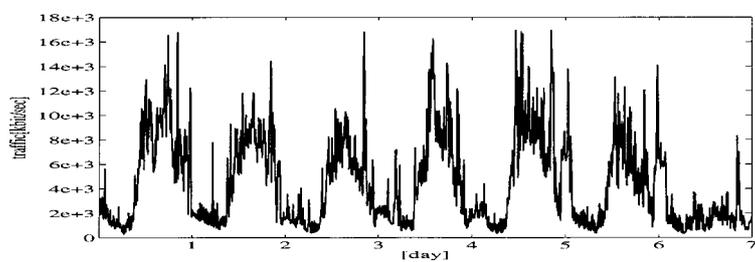


図 12. 法政大学 SINET トップトラフィックの第 1 週目 .

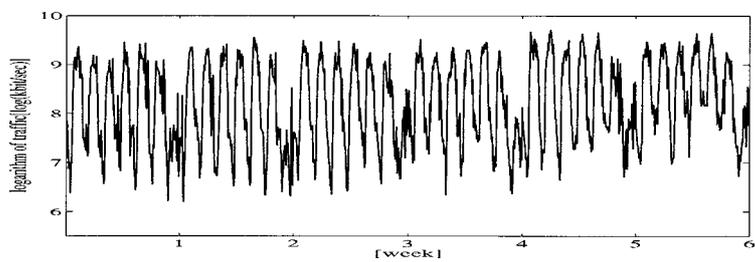


図 13. SINET トップトラフィックの対数系列 .

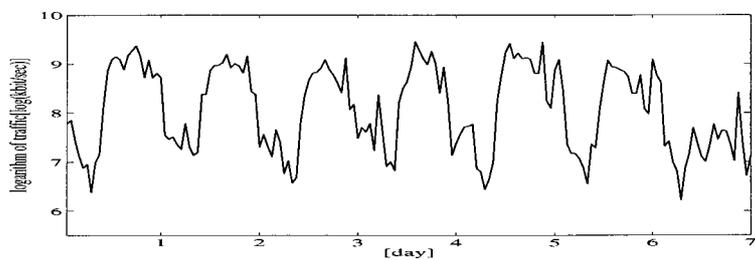


図 14. SINET トップトラフィック対数系列の第 1 週目 .

表 3. SINET トップトラヒックの最大対数尤度  $l(\hat{\theta})$  と AIC .

Model	$l(\hat{\theta})$	AIC
$T(1)+D+s+c(1)$	-384.33	778.65
$T(2)+D+s+c(1)$	-386.84	783.69
$T(1)+D+s+c(2)$	-382.17	776.34
$T(2)+D+s+c(2)$	-384.78	781.57
$T(1)+D+s$	-647.13	1300.27
$T(2)+D+s$	-650.14	1306.29
$t(1)+D+s$	-410.42	826.84
$t(2)+D+s$	-512.36	1030.72
$T(1)+s$	-684.79	1373.59
$T(2)+s$	-704.15	1412.30
$t(1)+s$	-422.01	848.02
$t(2)+s$	-528.63	1061.26

表 4. SINET トップトラヒック :  $T(1)+D+s+c(2)$ モデルにて推定されたパラメータ .

Hyper-parameter	Estimated value	Value for grid search
$\sigma^2$	$4.9 \times 10^{-3}$	1
$\tau_T^2$	$1.5 \times 10^{-3}$	$3 \times 10^{-1}$
$\tau_D^2$	$9.7 \times 10^{-3}$	$2 \times 10^0$
$\tau_s^2$	$9.7 \times 10^{-5}$	$2 \times 10^{-2}$
$\tau_c^2$	$9.7 \times 10^{-2}$	$2 \times 10^{+1}$
$a_1$	0.675	0.75
$a_2$	0.1	0.1

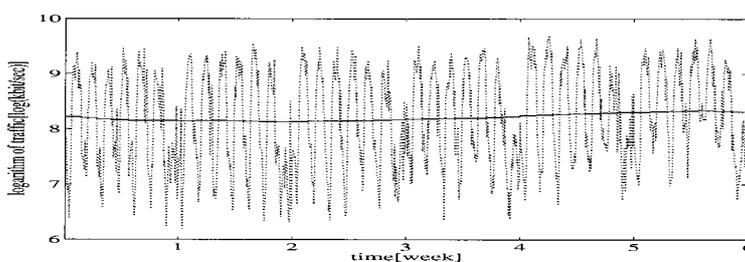


図 15. SINET トップトラヒック : トレンド成分の推定結果(平滑化の平均)と解析系列 .

サーチにより得られたパラメータの値を表 4 に示す .

$T(1)+D+s+c(2)$ モデルについて, 推定された各成分(平均)を図 15, 16, 17, および 18 に示す . また, 推定された成分の和と, トラヒック系列との残差を図 19 に示す .

以下, 結果について考察する . まず, 推定されたトレンド成分(図 15)は, ややなだらかではあるが, 2~3 週目でやや減少し, その後増加に転じるという, トラヒック系列の長いスパンでの傾向を抽出している . 次に曜日成分(図 16)は, 日曜日が最もトラヒックが少なく, 続いて土曜日が少なくなっていることがわかる . その他の曜日は週によって異なるが, 週の前半で

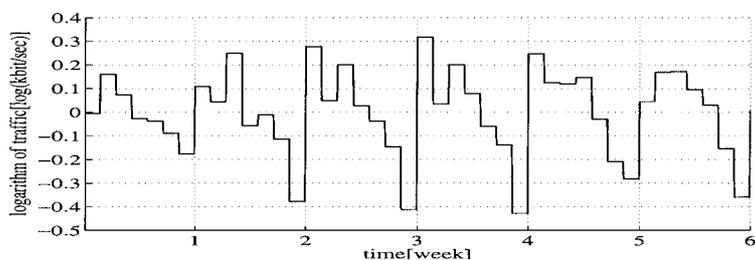


図 16. SINET トップトラフィック：曜日成分の推定結果(平滑化の平均)。

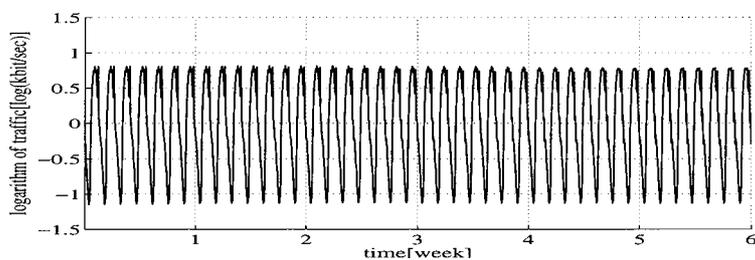


図 17. SINET トップトラフィック：日内変動成分の推定結果(平滑化の平均)。

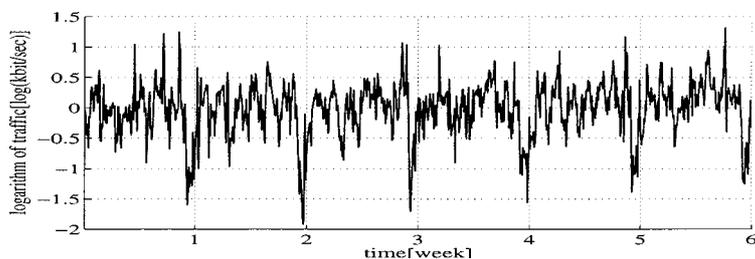


図 18. SINET トップトラフィック：AR 成分の推定結果(平滑化の平均)。

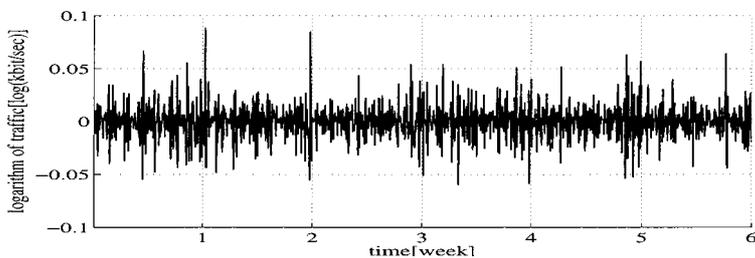


図 19. SINET トップトラフィック：推定の結果(平滑化の平均)，得られた残差。

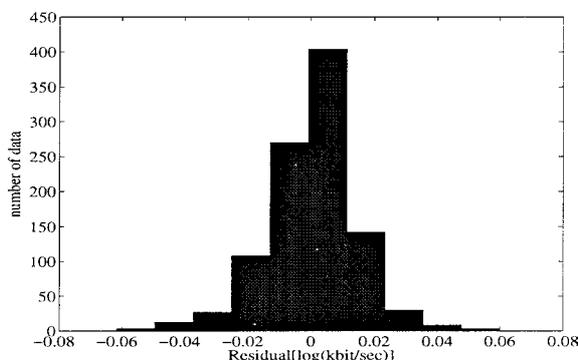


図 20. SINET トップトラヒック：推定の結果(平滑化の平均)得られた残差のヒストグラム。

はトラヒックが多く、後半では少ない傾向が見てとれる。日内変動成分(図 17)についてみると、昼間トラヒック量が多く夜少ないという変動パターンが抽出されている。そして、AR 成分(図 18)をみると、微細な変動を成分として抽出しているが、特に日曜日のトラヒック量減少をこの成分でも表していることがわかる。

残差(図 19)が正規分布に従うかどうか調べた。残差のヒストグラムを図 20 に示す。やや左右非対称な単峰の分布をしている。残差の平均は  $3.7837 \times 10^{-4}$ 、標準偏差は  $\sigma = 0.0139$  である。 $\pm 4\sigma$  の範囲外に残差値が 4 点存在しており、これら 4 点を外れ値とみなして、カイニ乗適合度検定を行った。その結果、階級数 10 に対して  $Q = 53.92$  が得られ、正規分布に従うとは言えないという結論となった。

## 5. おわりに

Kitagawa and Gersch (1984) で提案されているベイズ型時系列モデルに基づき、ネットワークトラヒックの系列をトレンド、1 週間周期の変動、1 日周期の変動、および有色雑音成分に分解する新しい分析手法を提案した。提案法を、法政大学のダイヤルアップアクセス、および SINET トップのトラヒックに適用し、成分分解の結果を得、結果について若干の考察を行った。トラヒックへの適用に際しては、提案法で示したモデルの持つ成分を適切に組み合わせをしたモデルをいくつか挙げ、それらモデルと提案法のモデルとを AIC により比較した。比較の結果、提案法のモデルが他のモデルよりも良いことが判った。

本手法は、ネットワークトラヒックの定量的な成分分解を可能とするものである。これには次のような意義がある。まず、分析手法の妥当性が、情報量規準(AIC)の意味において保証されていることが挙げられる。各種の分析手法が工夫され提案されているが、どれが妥当な手法であるか決めるのは容易なことではない。本論文で提案するような、統計モデルの枠組みでの分析手法では、情報量規準による手法の比較が可能であり、その意味で妥当な分析法を採用できることになる。本論文では、成分の組み合わせに関して、この妥当性を得ている。次に、成分分解が自動的かつ定量的に行えることが、利点として挙げられる。扱っているトラヒックデータは、目視であっても、1 日の周期的変動や 1 週間の周期的変動傾向が見て取れるものではある。しかし、それは人間の感覚に基づくものであって、数値化するのには別の客観的合理的の方法が必要である。また、人手による手作業でそのような成分分解を行うのは現実的ではない。提案手法はこれを自動的に行えるので、ネットワークトラヒックに対する予測や制御を自

律的自動的に行うシステムの開発に利用できるというメリットがある。そして、成分分解を継続的に実施することで、長期にわたる定量的データが蓄積でき、ネットワークの更新設計などに際し有益な指針となり得る点も強調しておく。

今後の課題としては、以下の項目が挙げられる；

1) 提案法で示したモデルは、カルマンフィルタを適用可能とする為、線形ガウス型状態空間モデルの枠内で、モデルが立てられている。しかし、実際のトラヒックの性質は、必ずしも線形性やガウス性を満たすとは限らない。トラヒックの性質を考慮した非線形項や、非ガウスなノイズ項の導入により、より良いモデルが開発できる可能性がある。その際、状態推定には、最近盛んに研究されている非ガウス型フィルタ(Doucet et al., 2001)を用いることになる。

2) 本論文で行ったネットワークトラヒックの分析では、残差が白色ガウス雑音にはならなかった。その原因としては、トラヒック原系列に対して行う変換や、日付変更時のレベルシフトなどが考えられることを、4章にて考察した。これを改善するため、今後の課題として、まずは、トラヒック原系列に対して行う変換としてより適切なもの(例えば Box-Cox 変換)を用いることが挙げられる。次に、日付変更時のレベルシフトに対応できるモデルを採用することが考えられる。現時点では、カルマンフィルタの枠組みで推定を行う為、モデルを線形ガウスに限定していたが、たとえば裾の重い分布を持つ非ガウスモデルに変更することで、値の急激な変化にも対応可能なモデルが構築できる可能性がある。

3) 本論文では、トラヒック系列を 1 時間サンプリングタイムの系列に変換した上で、ベイズ型時系列モデルを適用している。その理由は、提案手法のモデルの枠組みでは、あまり長い周期の日内変動を記述しづらい事による。より具体的には、モデル中の日内変動成分の周期 24 時間を表すために状態ベクトルのうち 23 次元を使用しているが、仮にサンプリングタイムが原系列のまま 5 分であるとすると、日内変動を表すのに  $(24 \times 12 - 1)$  次元のベクトルが状態に現れることになる。このように高次元の状態ベクトルは、状態推定に困難を生じると考えられる。この問題点を解消するアイデアが望まれる。

4) トラヒックの短期予測が可能であれば、予測に基づいたルーティングの逐次修正などを可能としトラヒック制御の立場から興味深い。そのためには、2)で述べたように、原系列のサンプリングタイム 5 分のままでモデル化を行うのが望ましい。

## 謝 辞

本研究は、統計数理研究所共同研究プログラム(16-共研-0010)による。トラヒックデータの整理と予備的な解析は、法政大学大学院の高橋兼基君(平成 14 年修士課程修了。現在、日本テレコム株式会社勤務)、清水威臣君(平成 16 年修士課程修了。現在、株式会社電通勤務)、今津和之君(平成 17 年修士課程修了。現在、日本電気株式会社勤務)による。

## 参 考 文 献

- 阿部威郎, 上田 徹(1985). 状態空間モデルを用いた電話収入予測, 電子通信学会論文誌(A), J68-A(5), 437-443.
- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley.
- Doucet, A., de Freitas, N. and Gordon, N. J. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*, Springer, New York.

- Ikoma, N., Shimizu, T., Imazu, K. and Yana, K. (2003). Decomposition of network traffic of dialup access into trend and periodic components based on nonstationary Bayesian model, *Proceedings of SICE Annual Conference 2003*, Fukui, Japan, 359–362.
- 今津和之, 清水威臣, 八名和夫, 生駒哲一 (2004). トレンド, 周期変動および AR の特徴を抽出する状態空間モデルによる SINET トラフィック系列の成分分解, 第 20 回ファジィシステムシンポジウム講演論文集, 545–548.
- 北川源四郎 (1993). 『FORTRAN77 時系列解析プログラミング』, 214–217, 岩波書店, 東京.
- Kitagawa, G. and Gersch, W. (1984). A smoothness prior state space approach to the modeling of time series with trend and seasonality, *Journal of the American Statistical Association*, **79**(386), 378–389.
- Leland, W. E., Taqqu, M. S., Willinger, W. and Wilson, D. V. (1994). On the self-similar nature of Ethernet traffic, *IEEE/ACM Transactions on Networking*, **2**, 103–115.
- Paxon, V. and Floyd, S. (1995). Wide-area traffic: The failure of Poisson modeling, *IEEE/ACM Transactions on Networking*, **3**, 226–244.
- Resnick, S. I. (1997). Heavy tail modeling and teletraffic data, *Annals of Statistics*, **25**(5), 1805–1848.
- 清水威臣, 今津和之, 生駒哲一, 八名和夫 (2003). 非定常ベイズモデルによるネットワークトラフィックの成分分解, 2003 年春季電子情報通信学会総合大会講演論文集, A-4–47, 136–136.
- Shimizu, T., Imazu, K., Ikoma, N. and Yana, K. (2003). Nonstationary Bayesian time series analysis for network traffic by decomposition model into trend, periodic and AR components, *Proceedings of Science of Modeling — The 30th Anniversary of the Information Criterion (AIC) —*, Yokohama, Japan, Dec. 14–17, 408–409.
- 高橋兼基, 生駒哲一, 八名和夫 (2001). 非定常ベイズモデルによるネットワークトラフィックの成分分解, 第 17 回ファジィシステムシンポジウム講演論文集, 153–156.
- Willinger, W., Taqqu, M. S., Scherman, R. and Wilson, D. V. (1997). Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level, *IEEE/ACM Transactions on Networking*, **5**, 71–86.
- 矢田 健, 北川源四郎 (1994). 競合環境下における通信トラフィックデータの分析, 第 62 回日本統計学会講演報告集, 34–35.
- Yada, T., Kitawaga, G. and Inoue, M. (1994). Prediction of telephone revenue by using a Kalman filter, *Proceedings of the 14th International Teletraffic Congress (ITC)*, **14**, 1165–1174.
- Yang, X. and Petropulu, A. P. (2001). The extended alternating fractal renewal process for modeling traffic in high-speed communication networks, *IEEE Transactions on Signal Processing*, **49**, 1349–1363.

Quantitative Analysis Method of Network Traffic  
by Component Decomposition Using Bayesian Time Series Model

Norikazu Ikoma<sup>1</sup> and Kazuo Yana<sup>2</sup>

<sup>1</sup>Faculty of Engineering, Kyushu Institute of Technology

<sup>2</sup>Faculty of Engineering, Hosei University

The ever-increasing expansion of the Internet constantly increases the importance of analyzing network traffic and making use of those results. To analyze network traffic in this study, we used a Bayesian time series model, which decomposes network traffic into trends, weekly and daily cycles, and colored noise components. The model is represented in a linear-Gaussian state space model that consists of system equations for each component and an observation equation, in which sums of the components yields the observed network traffic. State estimation by Kalman filter is used to obtain each component. Variance and AR parameters in the model are obtained to maximize the likelihood. Then a comparison among candidate models and selection of the components proceed based on the Akaike Information Criterion (AIC). To provide an example of an analysis using this proposed method, we report on the decomposition results of traffic data of dialup access at Koganei Campus and those at the top network domain of Hosei University, Tokyo which connects to the SINET.