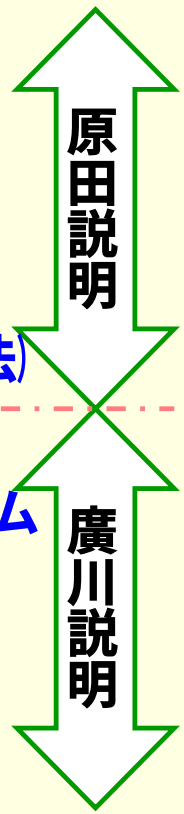
異常検知アルゴリズム紹介

原田　薫明
NTTサービスインテグレーション基盤研究所

これからの流れ

- アルゴリズムへの要求（振り返り）
- 異常検知アルゴリズムいろいろ
  - 波形予測アルゴリズム
    - Holt-Winters法（HW:rrdtool）
    - Kalman-Hoffding アルゴリズム（KH:原田法）
  - 変化点検出アルゴリズム
    - Traffic Leap/sLump Detection アルゴリズム（TLD:廣川法）
- 各アルゴリズムの精度の比較
- まとめ



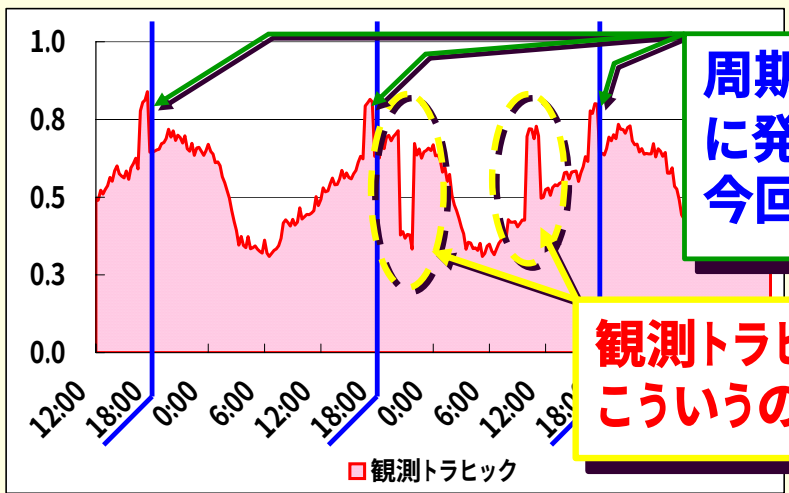
アルゴリズムへの要求（振り返り）

- **設定が簡単で精度の良い**アルゴリズム
  - 誤報が少ない
  - トラヒック量の多寡によらない

手間削減！

- **計算量の少ない**アルゴリズム
  - 処理が軽くて速いプログラム
  - 専用の装置を必要としない

経費節減！



周期的（毎18:00頃）に発生しているので，今回は通報しない．

観測トラヒックによらず，こういうのを見つけたい．

異常検知アルゴリズムいろいろ

波形予測アルゴリズム：予測波形から外れたら通報

- HW法 (rrdtool)
  トラヒックの周期性を考慮する波形予測アルゴリズム.
  同じ形の波の繰り返しとして周期変化する波形を予測.
- KHアルゴリズム (原田法)
  トラヒックの周期性を考慮する波形予測アルゴリズム.
  トラヒックの増加期と減少期の
  定期的な切替りとして周期変化する波形を予測.

変化点検出アルゴリズム：観測値の大きな変化を通報

- TLDアルゴリズム (廣川法)
  周期性を考慮しない簡単な計算で
  通常の昼夜変動と異なる変化を検出するアルゴリズム.
  トラヒックの時刻毎の変化量から異常値を検出.

異常検知アルゴリズム紹介

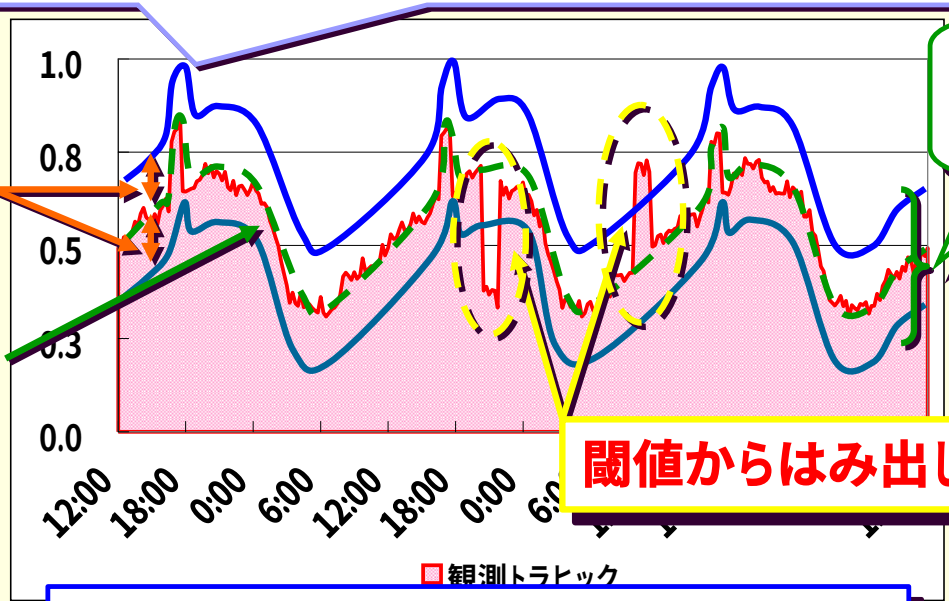
波形予測アルゴリズム

「予測が外れたら通報！」

波形予測アルゴリズム

いつもと同じくらいのトラヒック量かどうかを判定するために，
判断基準となるトラヒック量を計算（波形を予測）する．

基準値と許容量を組合せた閾値を毎時刻設定する．



許容量

基準値

いつもならこのくらい

閾値からはみ出したら通報！

⇒基準値と許容量を計算（予測）する．

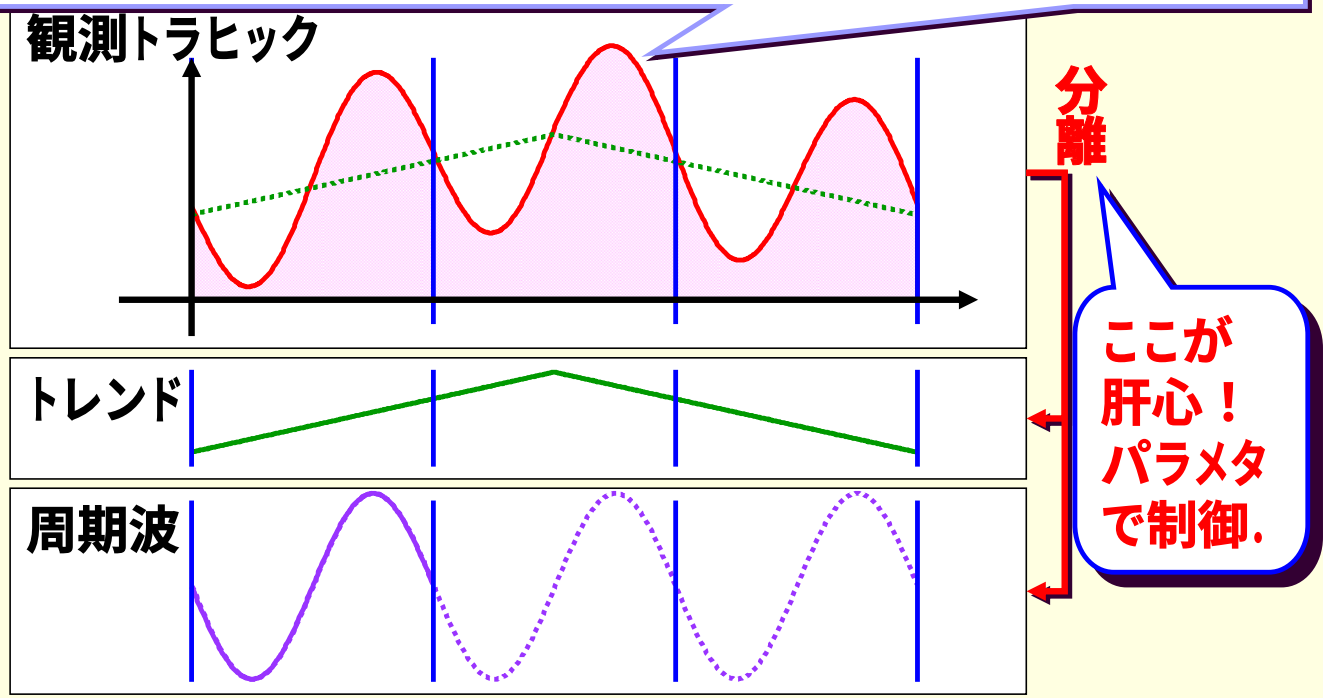
波形予測アルゴリズムその1

Holt-Winters法
(HW法：rrdtool)

HW法 (rrdtool) 1/3

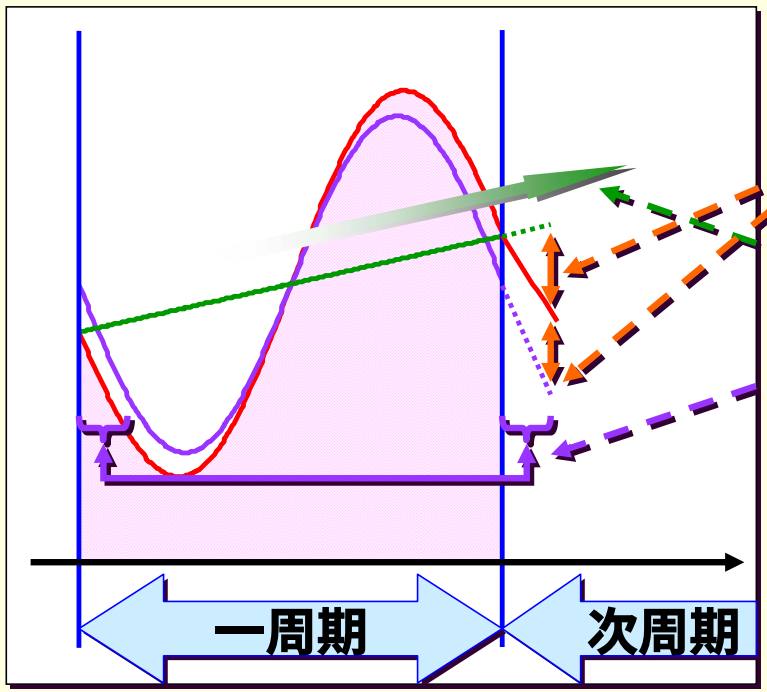
同じ形の波の繰り返しとして周期変化する基準値を予測.

観測トラヒックを大きな波 (トレンド) と小さな波 (周期波) に分離.



HW法 (rrdtool) 2/3

トレンドと周期波に分離するためのパラメータを設定する.



トレンドと周期波の分離法

パラメータ
- トレンドと周期波の混合割合
- トレンドの継続性
- 周期波の形の強さ

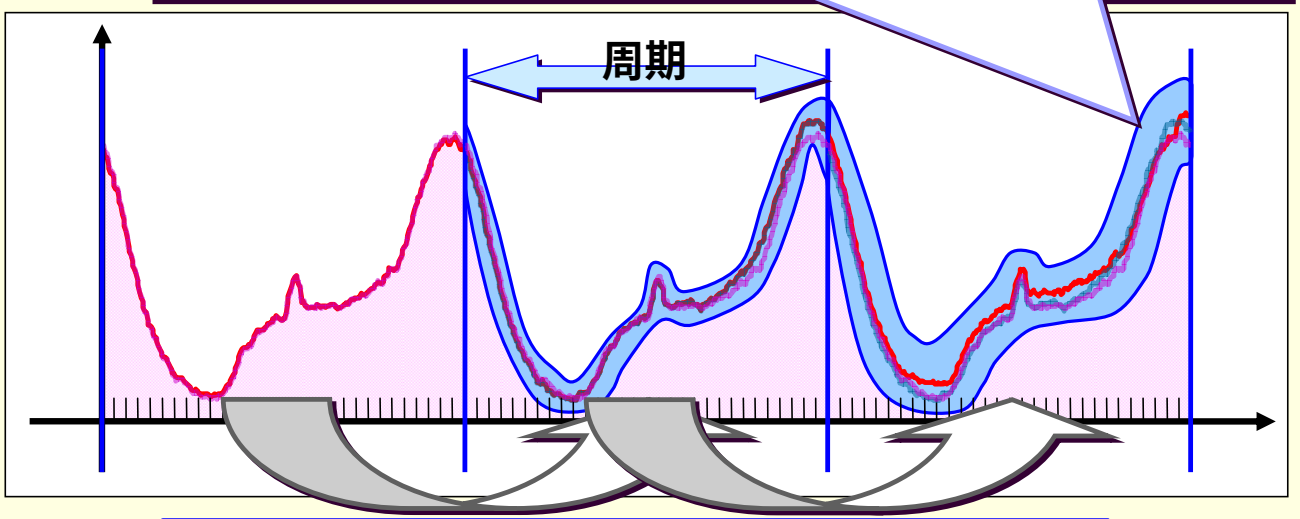
観測波形ごとに適切に設定する

波形予測の成否はパラメータの設定次第.

HW法 (rrdtool) 3/3

観測値を受け取るたびに
トレンドと周期波から次の時刻の基準値を予測する.
受け取った観測値をトレンドと周期波に分離しデータ更新する.

波形のずれに対する感度を設定して許容量とする.

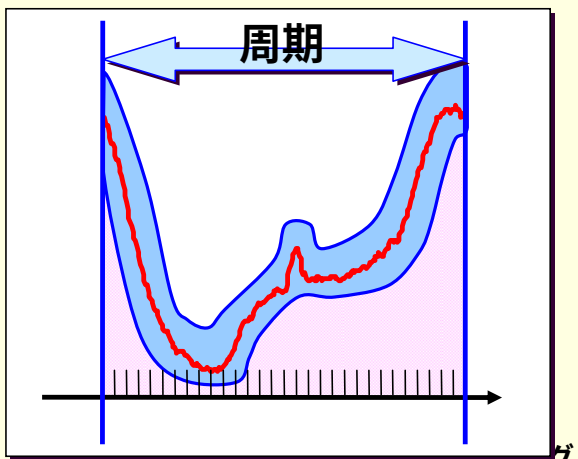


予測波形の“ずれ”を閾値として持ち越す.

HWアルゴリズムの動作条件

アルゴリズムを動作させるために必要な設定：
- 一周期の長さ（例：一日，一週間）
- 観測トラヒック分離のためのパラメータ3つ
- 許容量の感度パラメータ

アルゴリズムを動作させるために必要な記憶容量：
- 一周期分のデータ2セット分（周期波形と閾値）



メモ：

閾値が安定するまでに，
数周期もの時間がかかる．
予測が外れはじめたら，
パラメータの再設定も
必要となってくる．

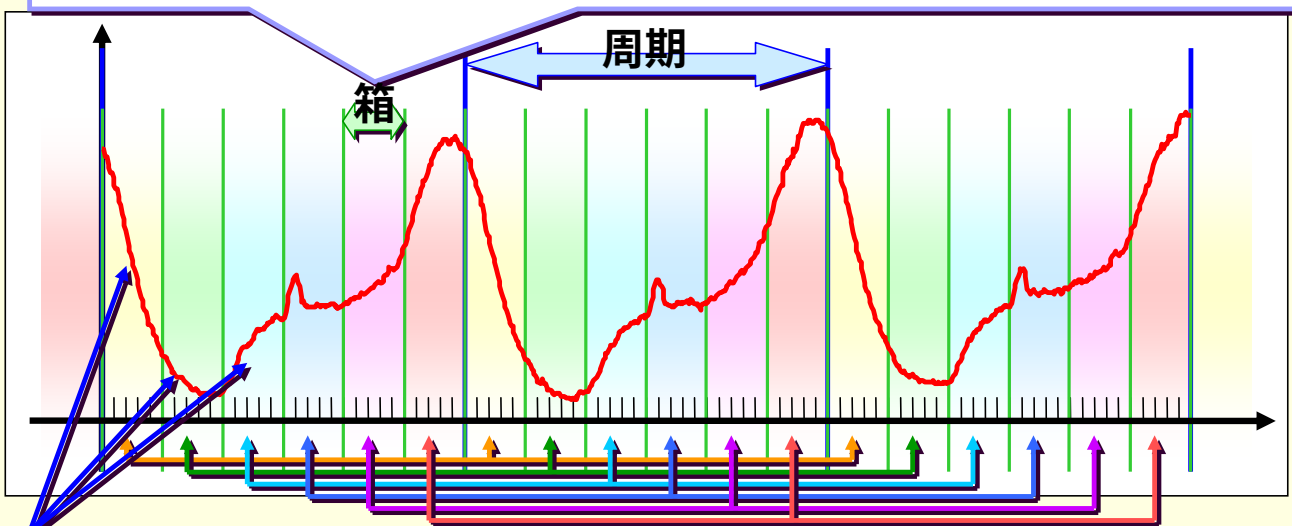
波形予測アルゴリズムその2

Kalman-Hoffding アルゴリズム
(KHアルゴリズム：原田法)

KHアルゴリズム（原田法）1/3

観測値の増加期と減少期の
定期的な切替りとして周期変化する基準値を予測．

一周期のトラヒックを小区間（箱と呼ぶ）に区切って考える．

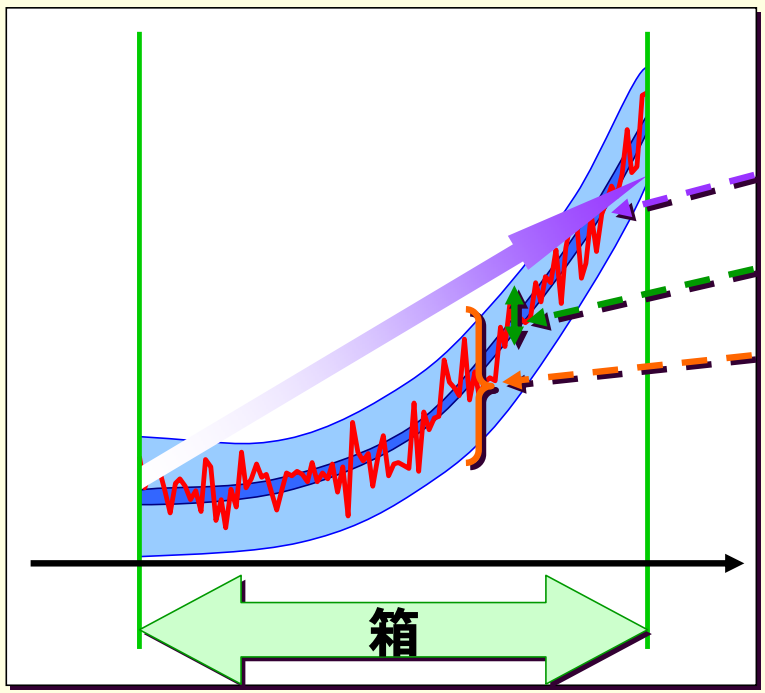


それぞれの箱の中では
増加，減少がはっきりする．

同じ色の箱の中身は似ている
⇒昔の情報を利用できる．

KHアルゴリズム（原田法）2/3

箱ごとに観測トラヒックの特徴量を算出する．



観測トラヒックの特徴量

- 平均的な増加割合
- 平均値のばらつき量
- 観測値のばらつき量

パラメータ

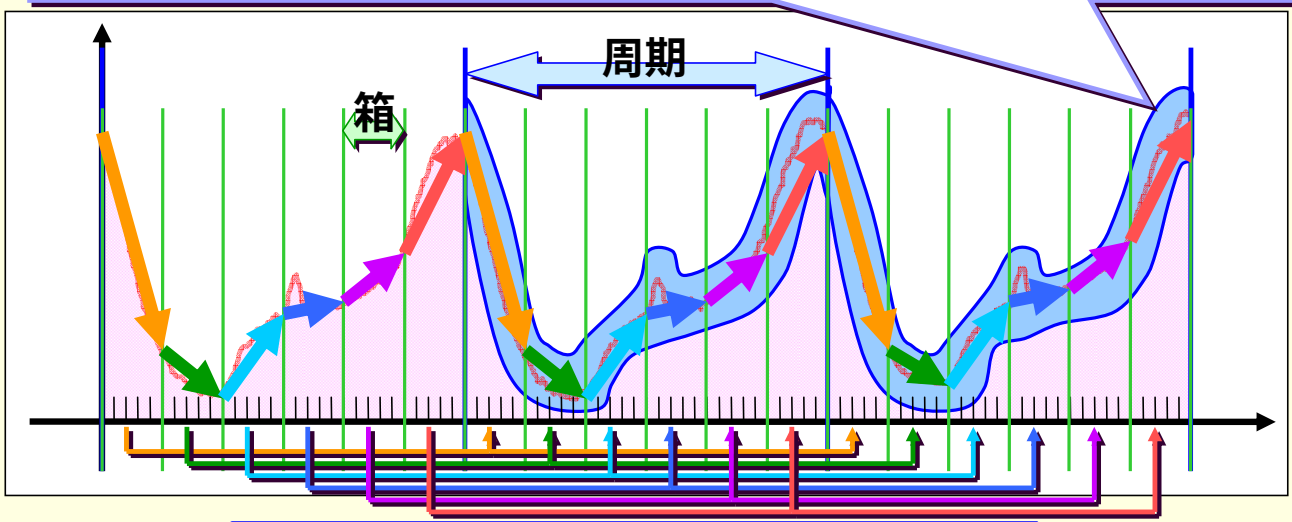
全て
自動計算可能

統計理論に基づくパラメータの自動推定法を適用できる．

KHアルゴリズム（原田法）3/3

計算された特徴量を組合せて，
観測値を受け取るたびに次の時刻の基準値を予測する．
それぞれの箱でトラヒックの特徴量を適宜計算し更新する．

観測値のばらつきに対する感度を設定して許容量とする．

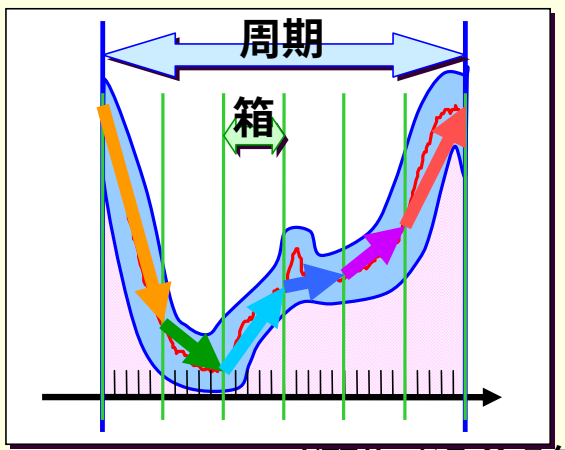


過去の情報を利用して再計算する．

KHアルゴリズムの動作条件

アルゴリズムを動作させるために必要な設定：
- **一周期の長さ（例：1日，1週間）と箱長（例：1時間，2時間）**
- **許容量の感度パラメータ**

アルゴリズムを動作させるために必要な記憶容量：
- **一周期の箱数分の特徴量（3つのパラメータ）**
- **一箱分のトラヒックデータ（特徴量計算のため）**



メモ：
異常検知は2周期目から．
箱の大きさは，
不揃いでも問題なし．
箱毎に特徴量を計算すると
計算処理量はやや多め．

しばらくお待ち下さい

つづく

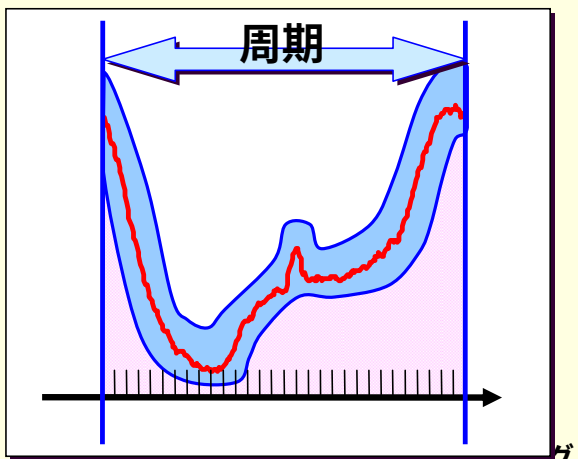
付録

波形予測アルゴリズムの実データへの適用結果

HWアルゴリズムの動作条件

アルゴリズムを動作させるために必要な設定：
- 一周期の長さ（例：一日，一週間）
- 観測トラヒック分離のためのパラメータ3つ
- 許容量の感度パラメータ

アルゴリズムを動作させるために必要な記憶容量：
- 一周期分のデータ2セット分（周期波形と閾値）

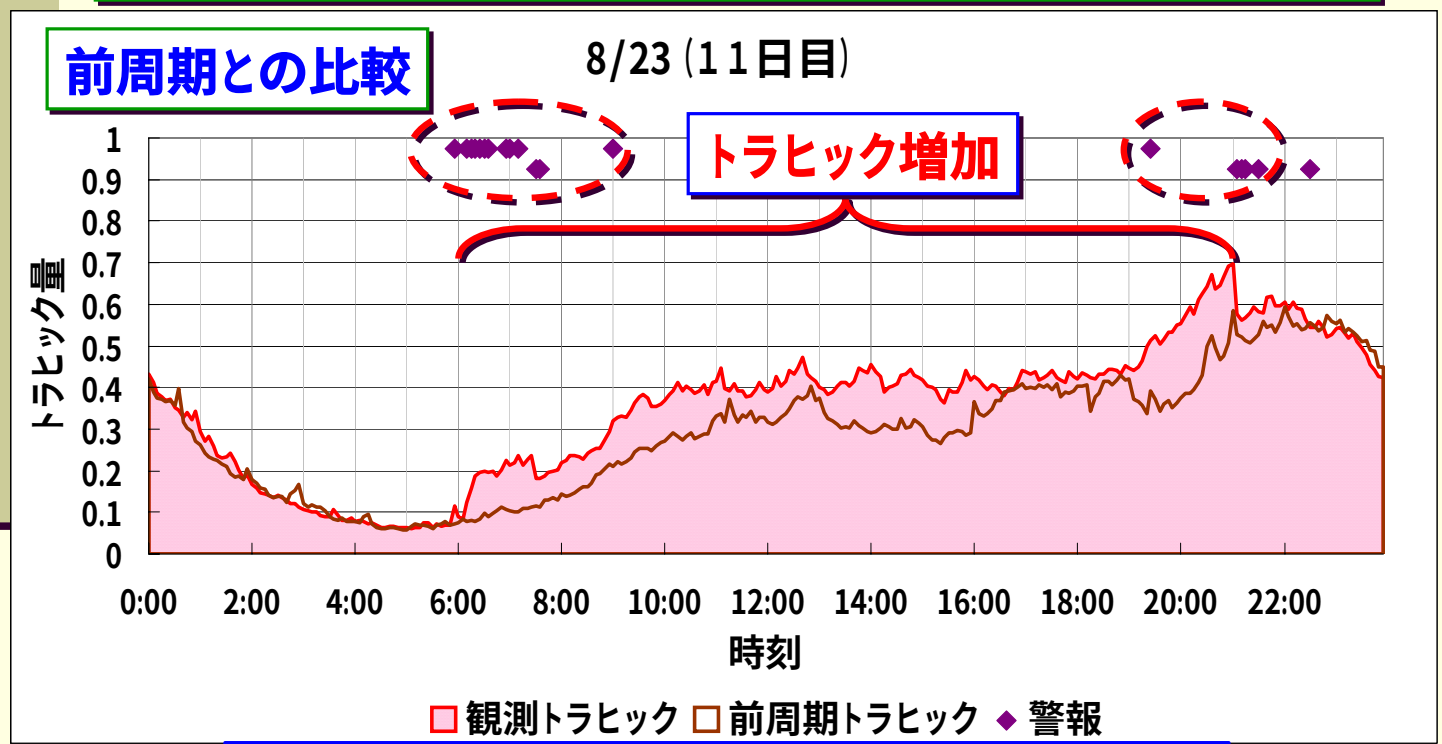


アルゴリズム評価：
（5分毎計測データを使用）

周期 の長さは1日
分離パラメータは適宜
感度パラメータは適宜
として，適用してみる．

HW適用結果

トラヒックデータがほぼ完全な周期傾向を持っている場合.

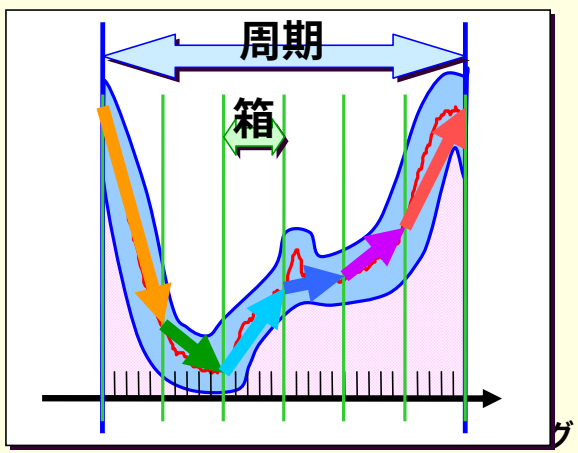


周期的な増減ではないトラヒック増加を検出.

KHアルゴリズムの動作条件

アルゴリズムを動作させるために必要な設定：
- **一周期の長さ**（例：1日，1週間）**と箱長**（例：1時間，2時間）
- **許容量の感度パラメータ**

アルゴリズムを動作させるために必要な記憶容量：
- **一周期の箱数分の特徴量**（3つのパラメータ）
- **一箱分のトラヒックデータ**（特徴量計算のため）



アルゴリズム評価：
（5分毎計測データを使用）

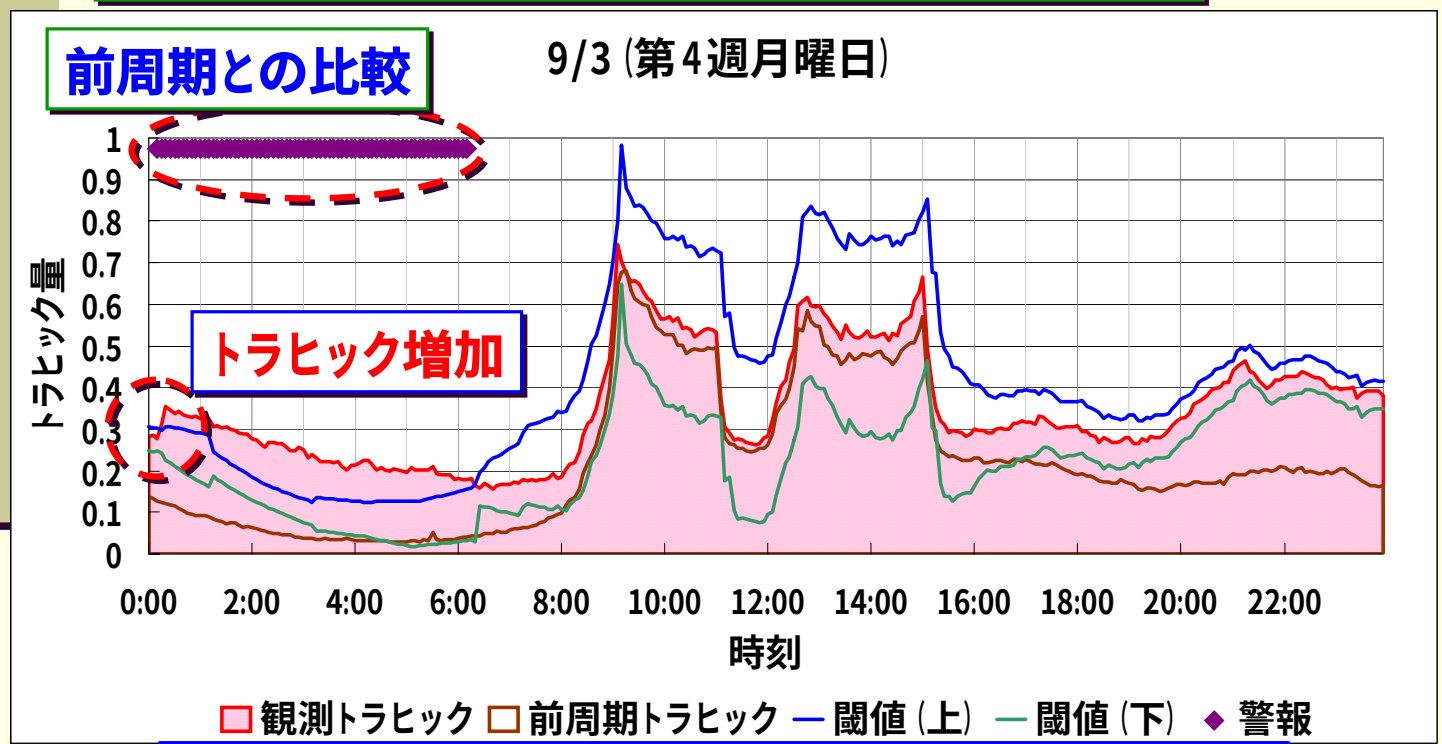
周期 **の長さは1週間**

箱 **の長さは2時間**

感度パラメータは適宜として，適用してみる．

KH適用結果

巨大な周期的トラヒックが観測される場合の結果.



周期的トラヒックを学習しつつ，特異な増加を検知.