

# Nearest Neighbor 法を用いたインドネシア Kahayan 川の水位予測

## Prediction of Water Level of the Kahayan River in Indonesia by The Nearest Neighbor Method

室蘭工業大学  
室蘭工業大学  
(株) リブテック  
日本工営株式会社

○学生員 三野 基之 (Motoyuki Mino)  
正 員 中津川 誠 (Makoto Nakatsugawa)  
フェロー会員 板倉 忠興 (Tadaoki Itakura)  
杉本 光 (Hikaru Sugimoto)

### 1. はじめに

Kahayan 川はインドネシアのカリマンタンに位置し、その周辺には泥炭地が広がり、豊かな自然を形成している。本来泥炭は湿地帯であるが、近年の農地開発の影響で地下水位が低下することなどが原因とみられる泥炭地の乾燥化が起き、そこに焼き畑などの火が引火し、大規模な火災を引き起こしている。そこで、地下水位を正常な状態にコントロールし、泥炭地を本来の湿地帯を回復する対策が必要とされる<sup>1)</sup>。

そこで、本研究では泥炭地の環境再生に向け、地下水位との関係が深い Kahayan 川の PalangkaRaya 観測所の水位を予測することを目指した。

水位を予測するには流出量を計算して水位-流量曲線(H-Q 曲線)から水位に換算する方法があるが、観測データが乏しい場合、流量そのものの把握や流出モデルによる推定が困難である。そこで、本研究では水位そのものを予測する Nearest Neighbor 法の適用を試みた。Nearest Neighbor 法については Karlsson et al.<sup>2)</sup>によって水文資料の時系列予測に適用した結果が報告されている。

今回、Nearest Neighbor 法によって、中期、長期である 5 日先、10 日先というリードタイムで日平均水位の予測を試みる。

### 2. Kahayan 川の概要

Kahayan 川は図-1 のように上流に KualaKurun 観測所、

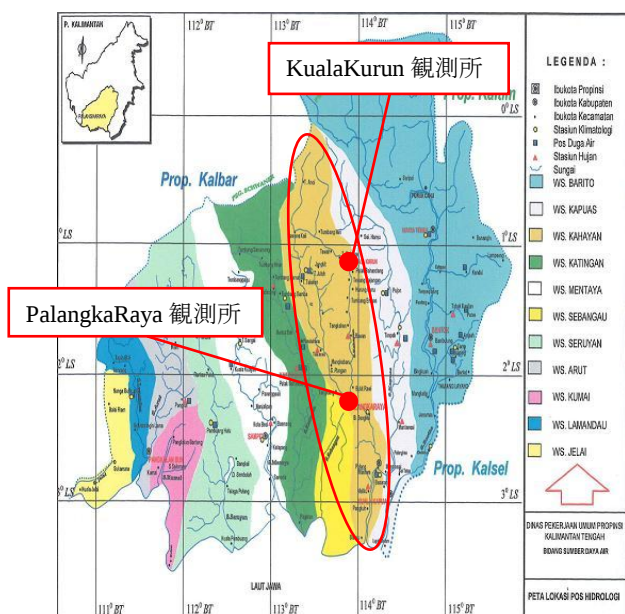


図-1 Kahayan 川概要図

下流に PalangkaRaya 観測所があり、下流では平坦な土地が広がり、上流になるにつれて川の勾配が大きくなるという特徴を有する。また、川沿いには湿原や PalangkaRaya などの町が隣接している。

### 3. Nearest Neighbor 法の概要

#### 3.1 特徴ベクトル

現時刻を含めて A 地点水位、B 地点水位をそれぞれ  $M_1$ ,  $M_2$  個の連続データを解析対象の対象として設定する。

例えば、図-2 のように考え、現時刻を  $n$  とすると、 $H_A$  は  $H_A(n), H_A(n-1), H_A(n-2), H_A(n-3)$  の  $M=4$  個、 $H_B$  は  $H_B(n), H_B(n-1), H_B(n-2)$  の  $M=3$  個となる。

これをベクトル化して表現すると、 $X(n) = (H_A(n), H_A(n-1), H_A(n-2), H_A(n-3), H_B(n), H_B(n-1), H_B(n-2))$  となる。この  $X(n)$  を特徴ベクトルと呼ぶ。

#### 3.2 ユークリッド距離

各因子の過去のデータベースの中から、時刻  $i$  ( $1 \leq i \leq n-1$ ) ごとに  $X(n)$  と同構造の特徴ベクトル  $X(i)$  を形成し、それと  $X(n)$  とのユークリッド距離を逐次計算する。

ユークリッド距離は次式により計算される。

$$|X(n) - X(i)| = \sqrt{\{H_A(n) - H_A(i)\}^2 + \{H_A(n-1) - H_A(i-1)\}^2 + \dots} \quad (1)$$

これらの中からユークリッド距離が小さくなるものから順に  $K$  個の特徴ベクトルを選び出し、それらの現象の後に続く  $K$  個の  $H_A(i+l)$  の平均値を  $n+l$  の時点の予測水位  $H_A(n+l)$  とする ( $l$  はリードタイム)。

### 4. Nearest Neighbor 法の適用

#### 4.1 データベース期間

予測に必要な水量はすべて日平均水位である。予測を行うためのトレーニングデータとして設定した期間(データベース期間)は欠測データが多いことから KualaKurun 観測所、PalangkaRaya 観測所ともに連続する水位データが多く存在している 2004 年～2006 年とする。

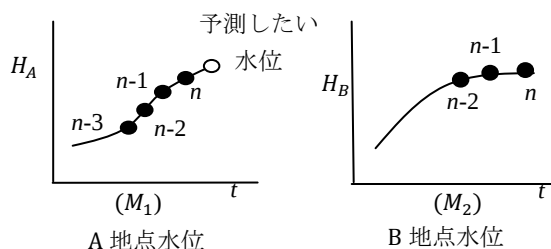


図-2 Nearest Neighbor 法概念図

さらに、2004年～2006年の中でも KualaKurun 観測所、PalangkaRaya 観測所ともに水位データが存在するもののみを使用する。

#### 4.2 対象洪水

予測を行う対象洪水は欠測データが多いことから KualaKurun 観測所、PalangkaRaya 観測所ともに連続する水位データが多く存在している 1995 年、2008 年、2009 年とする。さらに、1995 年、2008 年、2009 年の中でも KualaKurun 観測所、PalangkaRaya 観測所ともに水位データが存在するもののみを対象とする。ここで、1995 年はトレーニングデータよりも過去の事例だが信頼性を確認するため検証の対象とした。

#### 4.3 誤差評価

以下の予測誤差の評価には次式で表される平均二乗誤差平方根（以下、RMSE と表記）を用いる。なお、RMSE は小さくなるほど精度が良好であるといえる式である。

$$RMSE = \sqrt{\frac{\sum (H_{cal} - H_{obs})^2}{N}} \quad (2)$$

ここに、 $H_{cal}$  は予測水位(m)、 $H_{obs}$  は観測水位(m)、 $N$  はデータ数である。

#### 4.4 ずらし時間

河道の流下時間を考慮して水位データの時間ずらしを行い、最も相関係数の高くなるずらし時間を求めた。計算する際、低水時を除外して、PalangkaRaya 観測所の水位が 3 m 以上となるものを対象とする。結果は表-1 に示す。この結果からずらし時間を相関係数が最も高くなる 8 日とする。ずらし時間は構成因子の観測所を 2 個所用いる場合に使用する。例えば図-3 のような場合は下流の観測所の水

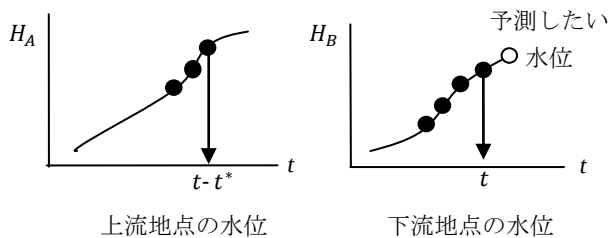


図-3 ずらし時間の適用

表-1 相関係数によるずらし時間

| 被観測所         | 観測所        | ずらし時間(日) | 相関係数  |
|--------------|------------|----------|-------|
| PalangkaRaya | KualaKurun | 4        | 0.422 |
|              |            | 5        | 0.466 |
|              |            | 6        | 0.497 |
|              |            | 7        | 0.516 |
|              |            | 8        | 0.523 |
|              |            | 9        | 0.518 |
|              |            | 10       | 0.505 |
|              |            | 11       | 0.491 |
|              |            | 12       | 0.467 |

位データを現時刻  $t$  から 4 個用い、上流の観測所の水位データを現時刻  $t$  からずらし時間  $t^*$  だけずらした時刻  $t-t^*$  から 3 個用い、ユークリッド距離を計算し、それを小さい順に  $K$  個選び出し、予測を行うこととする。

#### 5. 特徴ベクトル構成因子の選定

構成因子として水位のほかに降雨量も加味して水位を予測することも考えられるが、事前検討で降雨量を用いて予測した結果、水位データのみを用いた結果より誤差が大きくなる傾向が見られた。そこで本研究では水位データのみを用いた予測について検証することとした。

##### 5.1 観測所の選定

下流の PalangkaRaya 観測所の日平均水位を予測するため、構成因子が 1 個所の PalangkaRaya 観測所のみの水位データを用いる場合と構成因子が 2 個所の KualaKurun 観測所と PalangkaRaya 観測所の水位データを用いる場合が考えられる。その上で構成因子の選定について検証を行う。検証方法は観測所の選定による違いを比較するため、任意の値として  $K=5$ 、 $M=3$  として検証を行った。具体的には PalangkaRaya 観測所のみの水位データを用いる場合、KualaKurun 観測所と PalangkaRaya 観測所の 2 地点の水位データを用いる場合について、それぞれ  $K=5$ 、 $M=3$  と固定して、予測精度を比較した。RMSE による計算結果は表-2 に示す。この結果から 5 日先、10 日先ともに RMSE が比較的安定して小さくなる PalangkaRaya 観測所の水位データのみを用いて以下の検証を行う。

##### 5.2 $K$ の選定

最初にユークリッド距離が小さくなるものから順に選び出した個数である  $K$  について検証する。検証方法は PalangkaRaya 観測所のみの水位データを用い、任意に  $M=3$  と決定し、 $K=1\sim5$  で変化させることで精度を見る。RMSE による計算結果は表-3 に示す。この結果から 5 日先、10 日先

表-2 観測所の選定

| RMSE(m)                     | 1995年 |       | 2008年 |       | 2009年 |       |
|-----------------------------|-------|-------|-------|-------|-------|-------|
|                             | 5日先   | 10日先  | 5日先   | 10日先  | 5日先   | 10日先  |
| PalangkaRaya                | 0.573 | 0.873 | 0.619 | 1.084 | 0.598 | 1.048 |
| PalangkaRaya<br>Kuala Kurun | 0.731 | 1.033 | 0.612 | 0.997 | 0.967 | 1.266 |

表-3  $K$  の選定

| RMSE(m)      | $K$ | 1995年 |       | 2008年 |       | 2009年 |       |
|--------------|-----|-------|-------|-------|-------|-------|-------|
|              |     | 5日先   | 10日先  | 5日先   | 10日先  | 5日先   | 10日先  |
| PalangkaRaya | 5   | 0.573 | 0.873 | 0.619 | 1.084 | 0.598 | 1.048 |
|              | 4   | 0.582 | 0.870 | 0.629 | 1.100 | 0.607 | 1.058 |
|              | 3   | 0.625 | 0.896 | 0.635 | 1.092 | 0.622 | 1.080 |
|              | 2   | 0.606 | 0.939 | 0.658 | 1.149 | 0.656 | 1.166 |
|              | 1   | 0.677 | 1.086 | 0.708 | 1.312 | 0.796 | 1.296 |

表-4  $M$  の選定

| RMSE(m)      | $M$ | 1995年 |       | 2008年 |       | 2009年 |       |
|--------------|-----|-------|-------|-------|-------|-------|-------|
|              |     | 5日先   | 10日先  | 5日先   | 10日先  | 5日先   | 10日先  |
| PalangkaRaya | 5   | 0.536 | 0.886 | 0.663 | 1.145 | 0.633 | 1.099 |
|              | 4   | 0.544 | 0.866 | 0.645 | 1.110 | 0.629 | 1.079 |
|              | 3   | 0.573 | 0.873 | 0.619 | 1.084 | 0.598 | 1.048 |
|              | 2   | 0.603 | 0.901 | 0.635 | 1.046 | 0.592 | 1.023 |

ともに RMSE が比較的安定して小さくなる  $K=5$  を用いて以下の検証を行う。

### 5.3 M の選定

次に現時刻を含めてそれぞれの観測所の連続データの個数である  $M$  について検証する。検証方法は PalangkaRaya 観測所のみのデータを用い、 $K=5$  とし、 $M=2\sim5$  で変化させることで精度を見る。RMSE による計算結果は表-4 に示す。この結果から 5 日先は  $M=2$  と  $M=3$  の優劣の判断が難しく、予測の再現性を可視化して判断する。なお、10 日先は RMSE が比較的安定して小さくなる  $M=2$  を用いる。

### 5.4 水位予測の再現結果

図-4 および図-5 は構成因子を PalangkaRaya 観測所のみのデータを用い  $K=5$  とし、それぞれ  $M=2$ 、 $M=3$  として 5 日先を予測した結果であり、図-6 は構成因子を PalangkaRaya 観測所のみのデータを用い、 $K=5$ 、 $M=2$  として 10 日先を予測した結果である。図-4、図-5 および図-6 のグラフで表記されていない箇所は欠測期間である。なお、ユークリッド距離に欠測データが代入される際、予測は行わず、グラフには表記されない。また、実測、予測ともにデータがある箇所のみを表記している。

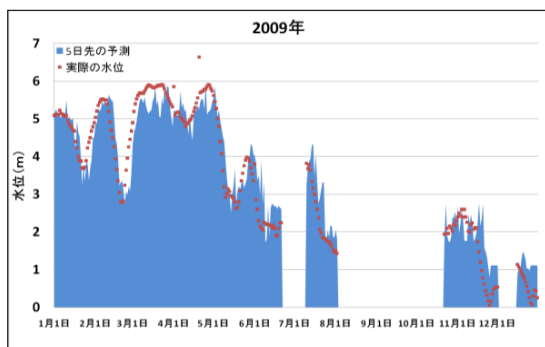
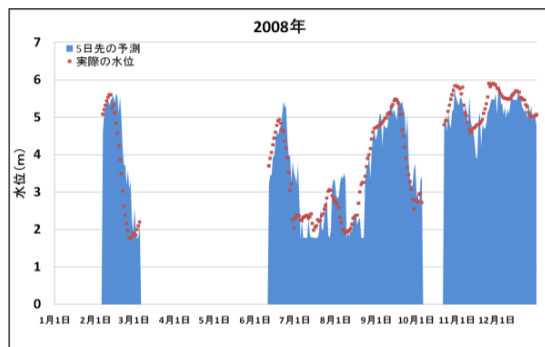
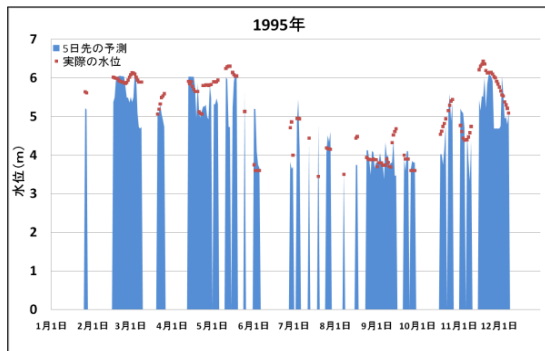


図-4 5 日先の予測  
(構成因子が PalangkaRaya 観測所,  $K=5$ ,  $M=2$ )

### (1) 5 日先の予測

5 日先の構成因子は図-4 および図-5 より予測結果がほぼ同じであるので予測を行う場合、必要なデータがより少ない  $M=2$  を用いることとする。したがって、5 日先の構成因子は PalangkaRaya 観測所のみのデータを用い  $K=5$ 、 $M=2$  が最適であるとし、以下の精度評価でこの構成因子を用いる。

### (2) 10 日先の予測

10 日先の構成因子は PalangkaRaya 観測所のみのデータを用い、 $K=5$ 、 $M=2$  が最適であるとし、以下の精度評価でこの構成因子を用いる。

### 6. 水位予測の精度評価

精度評価はピーク誤差、相対誤差を用いる。ピーク誤差、相対誤差の計算について、観測、予測ともにデータがある箇所を対象とし行う。ピーク誤差は次式より計算される。

$$J_{PE} = \frac{|H_P^* - H_P|}{H_P^*} \quad (3)$$

ここに、 $H_P^*$  は年最大の観測水位(m)、 $H_P$  は年最大の観測水位時の予測水位(m)である。

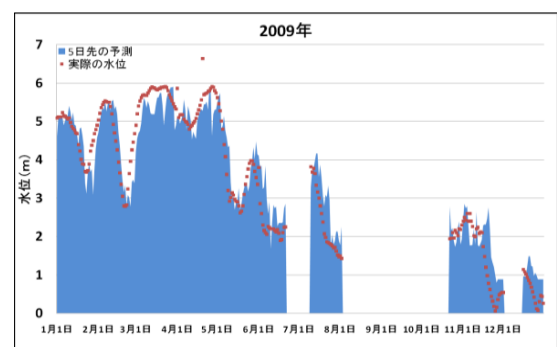
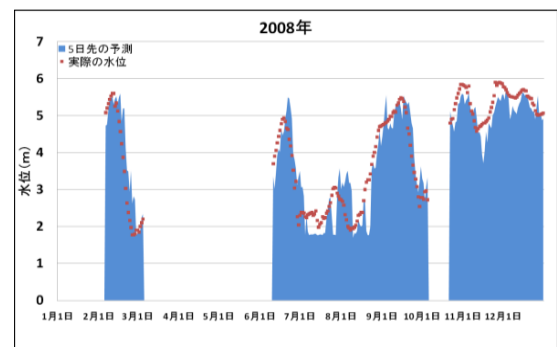
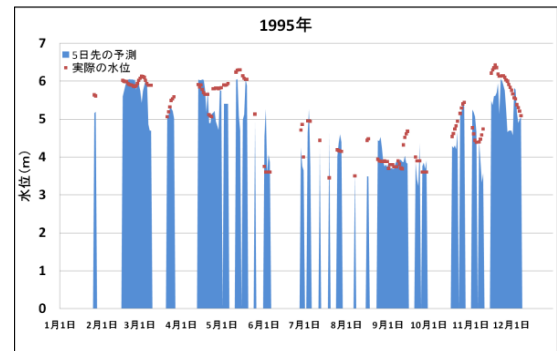


図-5 5 日先の予測  
(構成因子が PalangkaRaya 観測所  $K=5$ ,  $M=3$ )

相対誤差は次式により計算される。

$$J_{RE} = \frac{1}{N} \sum_{i=1}^N \frac{|H_{si}^* - H_{si}|}{H_{si}^*} \quad (4)$$

ここに、 $N$ はデータ数、 $H_{si}^*$ は観測水位(m)、 $H_{si}$ は予測水位(m)である。

表-5 はピーク誤差、表-6 は相対誤差、表-7 は水位 3m 以上の相対誤差の計算結果を表わす。表-5 での予測対象日は観測水位が年最大値となる日としている。予測基準日からのリードタイム分を足したものが予測対象日となる。しかし、1995 年の 10 日先予測については対象日から 10 日前のデータが欠測であったため、次善の策として 11 月 21 日を予測対象日として精度評価を行っている。

### 6.1 予測の精度評価

表-5、表-6 および表-7 より、5 日先の予測、10 日先の予測ともにピーク水位について概ね妥当な予測結果が得られたが、全体として若干低く予測される傾向にある。また、5 日先の予測、10 日先の予測ともに相対誤差は 1995 年、2008 年では比較的良好な予測となっているが、2009 年は他の 2 年と比べると精度が悪い。これは低水位での予測が過大で

あることに起因している。しかし、水位 3m 以上の高い水位であれば 5 日先の予測では相対誤差が 0.1 以内となり、10 日先の予測では相対誤差が 0.2 以内の精度で予測できることが示された。

### 7. まとめ

本研究で得られた知見を以下にまとめる。

- 5 日先の予測の構成因子は PalangkaRaya 観測所のみデータをを用い、 $K=5$ 、 $M=2$  が最適である。
- 10 日先の予測の構成因子は PalangkaRaya 観測所のみデータをを用い、 $K=5$ 、 $M=2$  が最適である。
- 5 日先の予測、10 日先の予測ともにピーク水位は低く予測される傾向にある。また、10 日先の予測では水位の立ち上がりが遅い傾向にある。
- 全水位を対象とした相対誤差は低水位部分で精度が落ちる傾向にある。
- 水位 3m 以上となる高い水位であれば 5 日先の予測では相対誤差が 0.1 以内となり、10 日先の予測では相対誤差 0.2 以内と良い精度で予測できる。
- 5 日先の予測は誤差が見られるが、水位の立ち上がりが良好であり、水位の将来の傾向を見ることは可能である。

### 参考文献

- 1) asahi.com 記事 (2007 年 10 月 6 日 21 時 56 分)  
「インドネシア、大量 CO<sub>2</sub> 焼き畑で泥炭火災相次ぐ」  
<http://www.asahi.com/special/070110/TKY200710060259.html>
- 2) M.Karlsson and S.Yakowitz: Water Resources Research, vol.23. No. 7, pp. 1300-1308, 1987.
- 3) 満倉真, 貞本均, 中津川誠: Nearest - Neighbor 法による感潮河川の水位予測, 水文・水資源学会誌, No.4, pp.371 - 380, 2002.

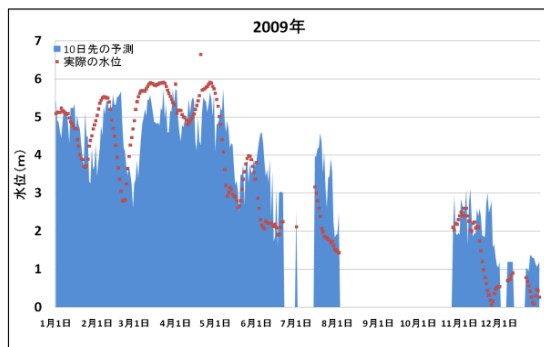
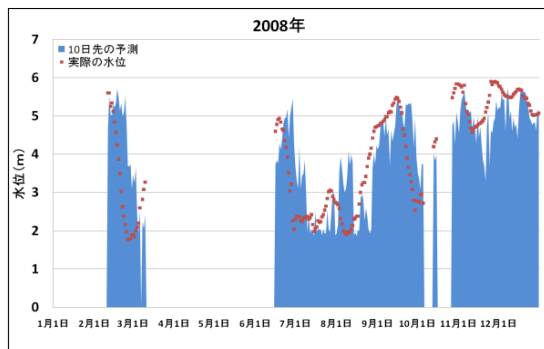
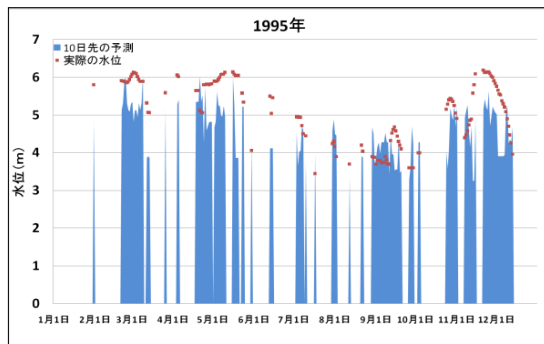


図-6 10 日先の予測

(構成因子は PalangkaRaya 観測所,  $K=5$ ,  $M=2$ )

表-5 ピーク誤差

| 予測リードタイム | 年     | 予測基準日  | 予測対象日  | 年最大の観測水位 $H_p^*$ | 年最大の観測水位時の予測水位 $H_p$ | ピーク誤差 $J_{PE}$ |
|----------|-------|--------|--------|------------------|----------------------|----------------|
| 5日       | 1995年 | 11月14日 | 11月19日 | 6.43m            | 5.50m                | 0.145          |
|          | 2008年 | 11月20日 | 11月25日 | 5.90m            | 5.22m                | 0.115          |
|          |       | 11月23日 | 11月28日 | 5.90m            | 5.49m                | 0.070          |
|          | 2009年 | 4月15日  | 4月20日  | 6.64m            | 5.26m                | 0.208          |
| 10日      | 1995年 | 11月11日 | 11月21日 | 6.19m            | 5.16m                | 0.166          |
|          | 2008年 | 11月15日 | 11月25日 | 5.90m            | 4.63m                | 0.215          |
|          |       | 11月18日 | 11月28日 | 5.90m            | 4.94m                | 0.163          |
|          | 2009年 | 4月10日  | 4月20日  | 6.64m            | 4.43m                | 0.333          |

表-6 相対誤差

| 予測リードタイム | 年     | 相対誤差 $J_{RE}$ |
|----------|-------|---------------|
| 5日       | 1995年 | 0.088         |
|          | 2008年 | 0.142         |
|          | 2009年 | 0.440         |
| 10日      | 1995年 | 0.142         |
|          | 2008年 | 0.263         |
|          | 2009年 | 0.796         |

表-7 水位 3m 以上の相対誤差

| 予測リードタイム | 年     | 相対誤差 $J_{RE}$ |
|----------|-------|---------------|
| 5日       | 1995年 | 0.088         |
|          | 2008年 | 0.092         |
|          | 2009年 | 0.087         |
| 10日      | 1995年 | 0.142         |
|          | 2008年 | 0.155         |
|          | 2009年 | 0.151         |