京都大学防災研究所年報　第51号B　平成 20 年 6 月
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 51 B, 2008

TCBM を利用した洪水流況予測に関する研究

和田健太郎*・小尻利治・原山和也**
田中賢治・浜口俊雄

*京都大学工学研究科
** 株式会社　山武

要　旨

近年，世界各地で津波や洪水といった水災害の発生が報告されており，それらを予測する技術の向上が期待されている。なかでも，河川下流部への人口・資産の集中が進む現代においては，洪水予測の精度を上げることは重要である。一方，データの電子化や計算機技術の向上が今後も進むことを考えると，過去のデータの誤記や欠損は非常に少なくなっていくと考えられる。

そこで本研究では，過去のデータを蓄積した事例ベースを利用して洪水予測を行うことを試みる。具体的には，事例ベースモデル（TCBM）を利用して実時間洪水予測である。TCBM の導入にあたり，適用性の拡大のため分布型流出モデル Hydro-BEAM と時系列予測モデル LLM を利用する。

キーワード：洪水予測，事例ベース，LLM，分布型流出モデル

1.　はじめに

近年，世界各地で主に集中豪雨や台風に起因する水災害の発生が頻繁に報告されている。スマトラ島沖地震津波やアメリカ南東部を襲ったハリケーン・カトリーナ等が記憶に新しいが，日本においても新潟豪雨や福井豪雨，東海豪雨など 2004 年の被害は甚大であり，2005 年も台風 14 号によって宮崎県が多大な被害を受けた。そのような状況を考慮すれば，水災害の予測と対策技術の向上は，極めて重要な課題で，高棹らは，レーダー雨量計を用いた降雨域の移動予測に関して移流モデルを提案し，洪水予測を行った（高棹　他，1983)。小尻はファジイ理論を用いた洪水予測とダム操作を展開し，その後，ニューラルネットワークの利用を提案している（小尻　他，1990)。Smith らも降雨予測の不確実性をランダム現象とみなし，確率論的出水予測として超過確率での対応を提案している(Smith et al., 2003)。さらに，関井等（2007）は AI 手法の洪水予測分野への新たな適用として，SSNN の分布型洪水予測への適用を試みている。この手法は，分布型流出モデルへの時空間的な対応を意味しており，幅広い可能性を含んでいる。ここで，実時間洪水予測の条件を鑑みると，

・計算時間が短いこと
・予測は 1 時間だけでなく 5～6 時間程度先まで得られること
・出来るだけ高い予測精度を維持できること
・必要な地点で予測できること

が挙げられる。そこで本研究では，最近適用されだした数理モデルの一種である事例ベースを取り上げ，その適用可能性について検討するものである。導入する事例ベースモデル（或いはデータ蓄積型モデル）は，洪水予測よりもむしろ流通分野や金融分野，最近は製造プロセス分野で多く用いられてきた。データを蓄積する操作はデータマイニングと呼ばれ，計算機技術の進歩により増大した蓄積データを有効利用するため，データの蓄え方と取り出し方に工夫を凝らしたものである。事例ベースモデルの 1 つである，TCBM（Topological Case-Based Modeling）の洪水事象へ適用は，既に実流域で試みられてきたが，予測時間を長くした時系列とした予測ではないこと，適用範囲が観測データや観測地点・基準地点に限ら

れていたこと，などの問題があった。ここでは，それらを改良し実用性を高めることを目的としている。

2. TCBMによる水位予測

2.1 予測システムの基本構造

事例ベース推論を用いて河川水位の予測を行うが，その構造は知識ベースシステム（knowledge-based system）の一環として展開することができる。知識ベースは一般にエキスパートシステムとも呼ばれ（戸内，1995, 小林, 1997），専門家から獲得した知識を主体として，専門家と同程度の問題解決水準を達成しようとするものである（Fig. 1 参照）。

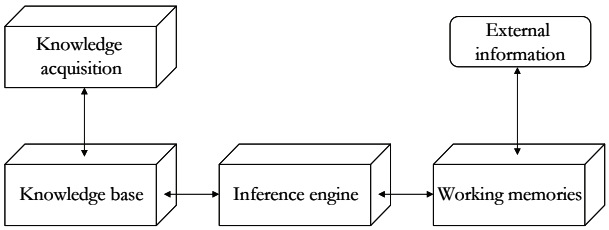


Fig. 1 Basic structure of knowledge-based expert system

(1) 知識ベースの基本構造

一般にプロダクションシステムでは知識は次のような if-then 文の集まりで表現される。

if (条件), then (処理)

if-then 文はプロダクションルールまたはルールと呼ばれ，（条件）が満たされた時，（処理）が実行される。また，知識獲得の際に設定された各条件，それぞれ一つずつの代表値を事実的知識として格納する。知識ベースのプロダクション集合を示すと以下のようになる。

Rule 1: if R1=n11,R2=n12,...,Rp=n1p, then RV=QV(n11,n12,...,n1p), m=QS(n11,n12, ...,n1p)

・・・

Rule i: if R1=ni1,R2=ni2,...,Rp=nip, then RV=QV(nil,ni2,...,nip), m=QS(nll,ni2,...,nip)

・・・

Rule q: if R1=nq1,R2=nq2,...,Rp=nqp, then RV=QV(nql,nq2,...,nqp), m=QS(nql,nql,...,nqp)

このプロダクションシステムでは，ルール i において，変数 *R1* が *ni1*，*Rp* が *nip* をとるとき,*処理 RV* は *QV(nil,ni2,...,nip)*で与えられることを意味している。結局，知識の条件式は q 個，存在することになる。*RV* は各条件式に一つずつ与えられている代表値（Representing Value），*m* は *RV* を設定する基となったデータを表し，*QV, QS* の値は条件式ごとに異なる数値，関数である。もちろん，事例*(nip)*もさらに下部の情報によるプロダクションシステムを構成している場合があり，条件式の総数は全体の和になる。また，no data（値無し）の場合もある。*RV* と *m* は知識獲得の度に更新されるが，条件式の数は常に一定である。*p* はシステム構築時に設定される各変数の分割数を表し，*R1~Rp* は，対象となる p 個の変数（事例内容）を意味している。

(2) ワーキングメモリ

ワーキングメモリ（working memory）は，知識ベースシステム構築の際に外界の環境を基にして設定され，事例内容 *R1~R*p を決定するための離散化条件が記述されている。知識ベースと同じく以下のように条件式で表されてはいるが，知識獲得の際に更新される知識ベースと異なり，一旦，システムが構築されて以降は不変である。

$$
\begin{aligned}
aj &\le Xj < aj + dj \\
aj + dj &\le Xj < aj + 2*dj \\
&\cdots \\
aj+(nj-1)*dj &\le Xj < aj+nj*dj
\end{aligned}
$$

ここに，*aj* (j=1，2，・・・，p) はシステム構築時の変数 *xj* (j=1，2，・・・，p) の最小値:min{ xj}を設定する。*dj* (j=1，2，・・・，p) は変数 *xj* に関する条件式の range, つまり区分単位幅を表し, 式:(max{ *xj*}-min{ *xj*})/ *nj* により設定されている。つまり *aj* + *nj*dj* は max{ *xj*}と同値となる。

(3) 推論エンジン

推論エンジン（inference engine）は，与えられた問題に対し，知識ベースとワーキングメモリを利用して，問題解決を達成するための制御を行う部分である。

i) まず，入力（*X1*，*X2*，・・・，*Xp*）を認識する。

ii) 次に，*X1~Xp* それぞれについてワーキングメモリを参照し，事例番号 R1~Rp を決定する。

iii) 知識ベースを利用し，*RV*=no data でなければ以下の式で推定値 Y が決定される。

$$Y = RV = QV(ni1,ni2,\cdots,nip)) \qquad (1)$$

iv) もし，*Y=RV*=no data の場合は，条件式の対象を変数ごとに前後一つずつ広げ，以下の式により推定値を得る。

$$Y_1 = AVE\left[\sum_{k_1=-1}^{1}\sum_{k_2=-1}^{1}\cdots\sum_{k_p=-1}^{1} QV(ni1+k_1, ni2+k_2, \cdots, nip+k_p)\right] \quad (2)$$

ただし，*RV*=no data の条件式の *RV* は，平均化の材料に含まない。

v) 更に *Y*=*RV*=no data の場合は，条件式の対象を各変数ごとに更に前後 1 つずつ広げ，以下の式により推定値を得る。

$$Y_2 = AVE\left[\sum_{k_1=-2}^{2}\sum_{k_2=-2}^{2}\cdots\sum_{k_p=-2}^{2} QV(ni1+k_1, ni2+k_2, \cdots, nip+k_p)\right] \quad (3)$$

$$Y_3 = AVE\left[\sum_{k_1=-3}^{3}\sum_{k_2=-3}^{3}\cdots\sum_{k_p=-3}^{3} QV(ni1+k_1, ni2+k_2, \cdots, nip+k_p)\right] \quad (4)$$

式(4)のように 3 つ範囲を広げても *RV* が存在しなかった場合は，*Y*=[出力無し]とする。iii)～v)より，推定値（*Y*）を出力する。

(4) 知識獲得支援

知識獲得支援サブシステム（knowledge acquisition support）は，新たに得られた観測情報を知識ベースに移植することを支援する役割を担う。本システムでの知識獲得の工程は，得られた観測情報を基に知識ベース内の代表値の値を更新する手法をとる。この手法により，ルールの数が増えると知識ベースの保守が難しくなるというプロダクションモデル特有の欠点が解消される利点がある。代表値の更新に用いる式は以下のようになる。

$$QV_{new}(ni1, ni2, \cdots, nip) = \frac{QV(ni1, ni2, \cdots, nip) + Y_{new}}{m(ni1, ni2, \cdots, nip) + 1} \quad (5)$$

2.2 事例ベース推論モデルでの展開

(1) 事例ベース推論モデルの概要

事例ベース推論モデル（TCBM）は，知識ベースシステムの構造を位相空間モデルとして表したもので，位相（Topology）の概念に基づく入出力関係の連続性が成り立つ対象に適用できる。TCBM を用いる際の工程は大きく 3 つに分けられる（Yamatake, 2004)。この 3 つを知識ベースシステムの類似性から表現すると，「モデリング」はワーキングメモリの設定を主としたシステム全体の構築段階に該当し，「推定」は推論エンジンによる出力，「学習」は知識獲得支援による知識ベースへの知識の格納作業にあたる。なお，以後では“量子化”との言葉が多発するが，これは入力空間を分割して部分空間の集合とすることを意味しており，分割数として p で表現したものに等しい。

TCBM ではモデルの次数やネットワーク構造などのモデリング用パラメータを同定して決めるのではなく，出力許容誤差を指定することで入力空間の位相を同定している。データは固定された事例ベース（位相空間）に事例として蓄えられ，出力推定時には入力と蓄積事例との位相空間内での距離（類似度）によって類似事例を検索し，見つかった事例を基に推定出力値を求めることになる（Fig. 2 参照）。

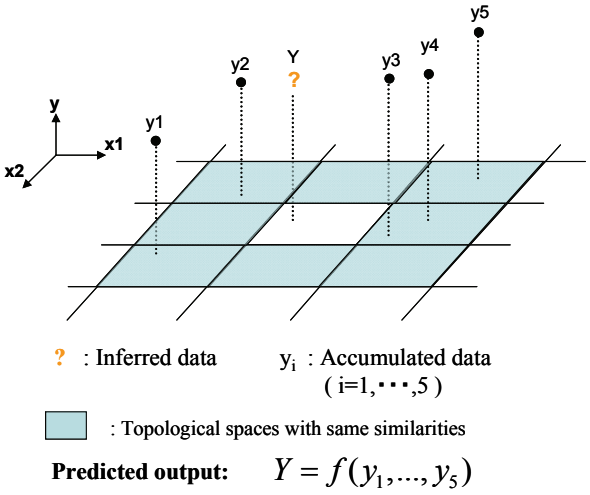


Fig. 2 Inference process of TCBM

図では TCBM の推定値の求め方を暫定的に関数 *f* として表しているが，具体的には代表値の概念を用いる。つまり，検索事例から見て同類似度を有する蓄積データから代表値を算出し，それを推定値として出力に当てるものである。

(2) 入力空間の量子化

入力空間の量子化に用いる二つの評価指標の出力分布条件では，量子化された部分空間それぞれ一つずつにおいて，あらかじめ設定された出力許容誤差（ε）内にその部分空間に蓄積されているデータの出力分布幅がおさまっているかどうかが必要となる。もし，部分空間内の蓄積事例の y 値の最高値と最低値の幅である出力分布幅が出力許容誤差（ε）より小さければ，出力分布条件を満たしていることになる。連続性条件では，部分空間それぞれ一つずつにおいて，その部分空間の代表事例値と周りを囲む部分空間の代表事例値の平均値との差が出力許容誤差（ε）内におさまっているかどうかを調べる。これは隣り合う部分空間で値が大きく異なるデコボコの位相空間とならないようにチェックするものであり，周りの代表値の平均値との差が出力許容誤差（ε）より小

さければ，連続性条件を満たしていることになる。

類似度は量子化された空間で Fig. 3 のように定義する。入力と同じ部分空間内にある代表値を類似度 4 として最高類似度に設定する。その後ひとつ範囲を広げるたびに 3，2，1，0 と値を下げていき，それ以外は全て類似度 0 とする。

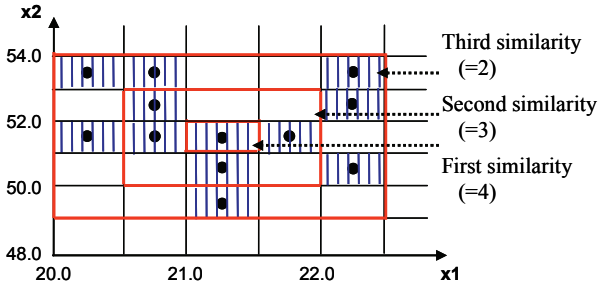


Fig. 3 Definition of similarity

3. TCBM による実時間洪水予測

3.1 TCBM における予測手法

(1) ローカルリニアーモデル

動的システムの将来の動きを予測するために，効果的な近似手法として，過去のデータの中で最も類似した軌跡をもつ曲線のみを使っての局所近似がしばしば用いられる。ローカルモデル（Local Model: LM）による時系列予測の手順は主に以下の 3 つのステップから成る（Babovic, 2001）。

i) 時系列データを位相空間に埋め込む。
ii) その中から，近隣点 k 個を拾い出す。
この局所部分で回帰を行い，予測値を得る。

ローカルモデルの同タイプのモデルとして，本研究で導入するローカルリニアモデル（Local Linear Model: LLM）がある。このモデルではステップ 3 の回帰が 1 次式で行われる。LLM では予測に線形モデルを用いるが，その結果は概して非線形モデルとなる。この非線形モデルは，各近隣点それぞれでの線形近似のつなぎ合わせからなっている。

非線形時系列データの分析の大部分は位相空間の近隣点探索である。これらの手法の成果は，採用する近隣点探索手順によって大きく左右される。従って，効果的な近隣点探索手順が選択されるよう，十分注意を払わなければならない。

ローカルリニアモデルの適用手順の予測段階には，近隣点のデータに基づいた回帰も含まれる。最も単純な方法としては，近隣点の動きを全て足し合わせて平均した値を用いるものがある。他には，例えば以下のように近隣点による回帰にファジィ理論を導入することができる。

(2) 遺伝的演算法による最適化

遺伝的プログラミング（Genetic Programming: GP）は遺伝的アルゴリズム（Genetic Algorithm: GA）の拡張という意味で，GA の考え方を多く用いる。GA で扱う情報は，PTYPE と GTYPE の二層構造からなる（伊庭，1996）。GTYPE（遺伝子コードともいい，細胞内の染色体に相当する）は遺伝子型の類似性で，低レベルの局所規則の集合であり，GA のオペレータの操作対象となる。PTYPE は表現型（発言型）であり，GTYPE の環境内での発達に伴う大域的な行動や構造の発現を表す。環境に応じて PTYPE から適合度が決まり，そのため適合選択は PTYPE に依存する。

生殖の際には，GTYPE に対して突然変異，逆位，交叉といった操作が適用され，次の世代の GTYPE を生成する。これらの操作の適用頻度，適用部位は一般にランダムに決定される。GA の基本的な流れをまとめると，次のようになる。

i) ランダムに初期世代の集団 M(0)を生成する。
ii) 適合度計算：現在の集団 M(t)内の各個体 m に対して適合度 *u(m)*を計算する。
iii) 選択：*u(m)*に比例する確率分布を用いて，M(t)から個体 m を選び出す。
iv)生殖：選び出された個体に GA 操作を作用させて，次の世代の集団 M(t+1)を生成する。
v) ii)に戻る

GP は，GA の遺伝子型（GTYPE）を拡張し，構造的な表現を扱えるようにしたものである。ここでの構造的表現とは，グラフ理論におけるグラフや，木構造のことをいう。複雑な数式や概念，関係などは木構造で表現できる。GP では tree と呼ばれる構造表現を使う。tree はサイクルを持たないグラフのことである。tree に対する GA 操作として，以下を導入する。これらはビット列を対象とする従来の GA 操作の自然な拡張である。

Gmutation：ノードのラベルの変更
Ginversion：兄弟の並べ替え
Gcrossover：部分木の取替え

以上の準備のもとに GP のアルゴリズムは次のようになる。

Step1：ランダムに tree:GTYPE{ $g_t(i)$ }を構成する。
Step2：GTYPE{ $g_t(i)$ }の表現型 PTYPE{ $p_t(i)$ }に対して適合度{ $f_t(i)$ }を決める。
Step3：適合度の大きな GTYPE に対して一定数のペアを取り出す。
Step4：取り出したペアに対して Gcrossover を適用し，適合度の小さな GTYPE と置き換える。

Step5：GTYPE に関して，ランダムに Ginversion，Gmutation を適用する。

Step6：以上によって求められた新しい GTYPE を，次の世代の$\{g_{t+1}(i)\}$として，Step2 へ戻る。

ただし，適合度は大きいものほど良いとしている。このアルゴリズムは，実操作の違いを除いて GA のアルゴリズムと同一である。従って，GP では GA の知見の多くをそのまま用いることができる。

3.2 予測・観測データの再構築

(1) 予測降水量の算出

実時間洪水予測に必要な予測降水量の算出には，気象庁が提供している降水短時間予報の値を利用する。降水短時間予報は解析雨量と同じく 30 分間隔で発表され，6 時間先までの各 1 時間雨量を予報している。例えば，9 時の予報では 15 時までの，9 時 30 分の予報では 15 時 30 分までの，各 1 時間雨量を予測する。解析雨量により毎時間の雨量分布が得られる。この雨量分布を利用して雨域を追跡すると，それぞれの場所の雨域の移動速度が分かる。この移動速度を使って直前の雨量分布を 6 時間分移動させて，6 時間後までの雨量分布を作成する。

予測の計算では，雨域の単純な移動だけでなく，山の斜面で雨が強まったり，山を越えて雨が弱まったりする地形の効果も考慮している。また，予報時間が延びるにつれて，次第に雨域の位置や強さのずれが大きくなるので，予報後半には数値予報の結果も加味している。　降水短時間予報では，日本を小さなメッシュで区分けし，それぞれのメッシュごとに 6 時間先までの各 1 時間雨量の予報値が提供されている。メッシュサイズは東西：3.75 分，南北 3.00 分である。

(2) 洪水データの作成

TCBM の適用で問題となるのが，事例として蓄えられている洪水データの数である。少ない場合は，分布型流出モデル（例えば，Hydro-BEAM）を利用し，洪水時の蓄積データを増やすことが出来る。シミュレーションによるデータマイニングの際には，信頼度を添付して蓄積する必要があり，予測の際にはシミュレーションデータに信頼度を掛けた値を使用する。ただし，全てのシミュレーションデータの信頼度は同じとする。信頼度 R の設定には，FFM モデリングの際に設定する出力許容誤差を基準として使用し，式(6)で表される。

$$R = N_p / N_{all} \tag{6}$$

ここでN_pは洪水データのうちでシミュレーション値と真値との絶対誤差が，あらかじめ設定する出力許容誤差より小さかった回数を表す。また，N_{all}は洪水データの全回数を表している。

3.3 洪水予測手順

洪水予測（Flood Forecast Model: FFM）の流れを示すと，Fig. 4 に示すように洪水水位予測用と通常水位予測用の 2 つの位相空間を内包するモデルとなる。それぞれのデータ蓄積位相空間の軸となるモデリング用入力変数の数を一般化のため p 個と q 個とする。

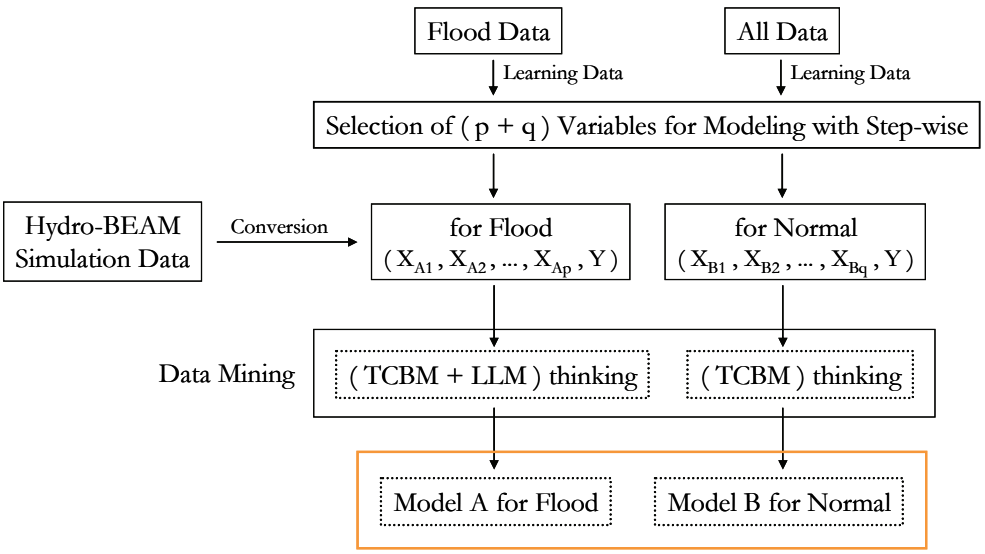


Fig. 4 Flood forecast system

つまり，2 つの位相空間の軸として採用する入力変数はそれぞれ異なり，計(p + q)個のモデリング用入力変数を選択することになる。(p + q)個の変数の選択手法としては，TCBM と同様にステップワイズ法を利用する。洪水データのみでステップワイズ法にかけた結果選ばれた p 個の変数を洪水水位予測用，通常データで選んだ結果選ばれた q 個の変数を通常水位予測用のモデリング変数としてそれぞれ採用する。

ここでは，事例ベース位相空間内にはデータが今全部で n 個あるとして，それらを *Si(ti)*,(i=1,2,…,n)で表すことにする。この n 個から位相距離 r_1， r_2 を基に正式な近隣点として j 個を認定し，それらのデータによって推定値を出力するまでの流れを示している。

i) 時刻 *t* での Input を探索点 *Q(t)*として設定する。
ii) *Si(ti)*, (i=1,2,…,n)のうち，*Q(t)*との位相距離が r_1 より小さいものを k 個選び，それらを *Si(ti)*, (i=1,2,…,k)として iii)に進む。
iii) *Si(ti-1)*, (i=1,2,…,k)に注目し，その中で *Q(t)*との位相距離が r_2 より小さいものを j 個選び，それらを *Si(ti-1)*, (i=1,2,…,j)として iv)に進む。
iv) で選ばれた j 個のデータを正式な近隣点として認定し，*Si(ti)*: *(Xi1, Xi2,..., Xip, Y)*の *Y* を *Y(Si(ti))* と表すことにする。
v) *Y(Si(ti))*, (i=1,2,…,j) を推定値として出力する。位相距離 r_1， r_2 の長さは，代表値の算出に用いた類似度までの部分空間内の値を全て網羅できる長さ，(無次元化した対角線の長さ)/2 に設定する。

以上のようにして LLM 概念の出力を求め，TCBM 概念の出力と合わせた最終的な FFM の出力結果は以下の式で表すことにする。

$$Y(S_i(t_i)) = \frac{RV + \sum_{i=1}^{j1} Y(S_i(t_i)) + R * \left\{ RV_H + \sum_{i=1}^{j2} Y_H(S_i(t_i)) \right\}}{1 + j1 + R * (1 + j2)} \quad (7)$$

ここで，*RV* は TCBM 概念の代表値を表している。*Y(Si(ti))*は LLM 概念の出力を表し，見つかった近隣点の数(j1,j2)だけ足し合わせている。また右下の添え字 *H* は Hydro-BEAM シミュレーションデータであることを表し，*R* はその信頼度である。

本システムは 6 時間先までの水位を予測し，洪水時の水位のようなピーク水位の予測にも対応できるという特徴を持つ。まず，降水短時間予報から 6 時間先までの予測降水量を算出し，FFM の Input と上流水位予測の Input に組み込む。TCBM を用いて上流水位の予測を行い，GP を用いて欠損値の補正を行う。算出された水位予測結果を FFM の Input に組み込み，データセットを作成する。作成されたデータセットを FFM 内に取り入れ，1 時間先のデータセットを判断基準値にかける。判断基準値にかけた結果,「洪水」と判断すれば洪水水位予測用モデルを用いて（TCBM+LLM）概念で推定を行い，「通常」と判断すれば通常水位予測用モデルを用いて（TCBM）概念で推定を行う。その結果を 1 時間先水位予測の結果とする。

FFM の iv)で求められた予測水位を 2 時間先水位予測のためのデータセットに組み込み，再び判断基準値にかけて 2 時間先水位の予測を行う。iv)と v)を繰り返して 6 時間先までの水位を予測し，それらをつなげた結果を 6 時間先までの水位予測の結果とする。

iii)の判断基準値について説明すると，洪水事例として蓄積されているデータの，各入力変数 $x_1,\cdots,x_p$ それぞれの平均を取っておく。そして入力変数 $X_{A1},\cdots,X_{Ap}$ の中で一つでも平均を越えている値があれば，洪水水位となる可能性があると判断し，洪水水位予測用モデルへと送ることにする。その判断基準値を満たさない場合は，通常水位予測用モデルに送ることとする。

i)～iii)が，構築された実時間洪水予測システムによる予測の手順である。次章では，本システムを過去の洪水イベントへ適用し，LLM を導入した効果等について検証・考察を行う。

4. 適用と考察

4.1 適用流域と設定条件

京都府京都市を流れる鴨川の七条大橋以北を対象流域として用いる(Fig. 5 参照)。鴨川は長さ 33km，流域面積 208km^2 の一級河川である。京都の北にある桟敷ヶ岳に源を発し，高野川と合わさって京都の市街地を北から南へ流れたのち，桂川に注いでいる。出町柳付近で高野川と合わさるまでは賀茂川，あるいは加茂川と呼ばれる。また鴨川が注ぐ桂川は，大阪府との境で木津川，宇治川と合流して淀川となるので，鴨川は淀川水系である。鴨川は時に洪水を起こす川としても知られ，1935 年の洪水では京都市内の橋のうち約 9 割が流され，死傷者 83 名を出している。

使用する観測情報は，上賀茂，大原，荒神橋の観測水位と荒神橋の H-Q 式である。TCBM や FFM の

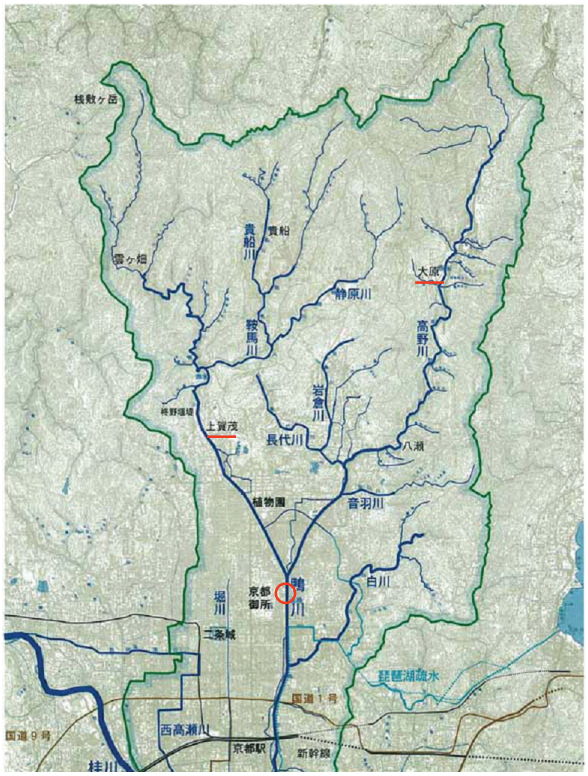


Fig. 5 Applied area of the Kamo River

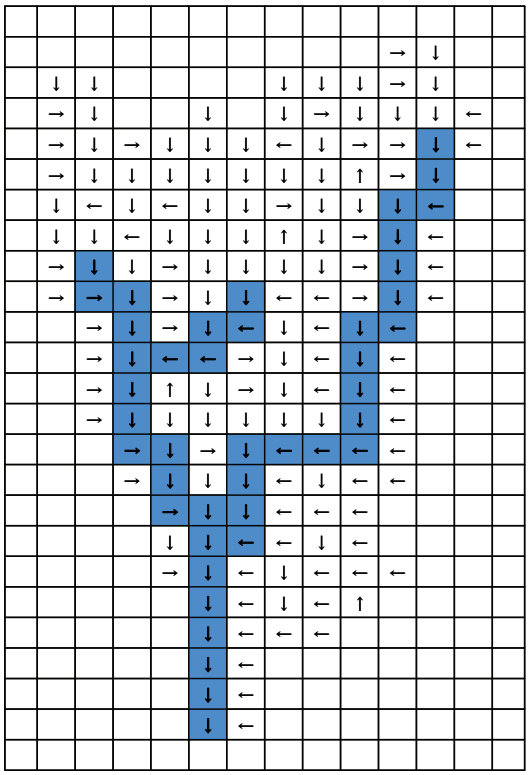


Fig. 6 Flow map of meshed river basin

適用では，荒神橋の水位を予測対象とする。Fig. 6 は適用時におけるメッシュ化された流域図と鴨川の位置を示している。

位相空間の量子化などの際に重要となる出力許容誤差の値は 5cm とする。モデリングの際に必要なモデル変数の数は，計算時間やデータの充足度を考慮して 4 つとする。FFM の一般化のため p と q で表現している入力変数の数もそれぞれ 4 に設定する。分布型洪水予測用の TCBM に関しては，「緯度」と「経度」の両方が選ばれた時点での変数の数をモデル変数の数とする。類似度に関しては類似度 1 まで使用し，0 からは値無しとして出力する。FFM 内の洪水水位予測用モデルには，水位が 80cm 以上のデータを蓄積することとする。

データ欠損のリスクを減らすため，主として雨量は積算雨量として入力変数候補としている。積算雨量データは鴨川流域の AMeDAS 雨量観測地点 4 個(京都，大原，上賀茂，府庁)分を用意し，上流の水位データは 2 地点（大原，上賀茂）分を用意した。積算雨量は積算時間を変えて 19 種類，上流水位は 1，2，3，6 時間前の 4 種類作成したので，合計 84 個のデータセットを入力変数候補として作成したことになる。荒神橋の実時間観測水位も候補に加えたステップワイズ法で変数選択をする際の多重共線性の判断基準値はデータの分布の割合を上げることを考え，0.7 とやや低めに設定した。

気象データとして，AMeDAS 観測値（1979～2000 年）を使用した。気象データは，ティーセン法により AMeDAS 観測所の位置をもとに，それぞれの観測値を各メッシュに割り当てた。気温は AMeDAS 観測所の標高と各メッシュの平均標高の標高差を用い，気温の逓減率を 6.5(℃/km）として補正を行った。

パラメータ同定は，1999 年と 2000 年を対象に行った。続いて FFM を中心とする実時間洪水予測システムの適用結果を，次に示す。予測対象とする洪水イベントは 1999 年 6 月 27 日 5 時～6 月 28 日 0 時，2001 年 6 月 19 日 15 時～6 月 20 日 3 時，2001 年 8 月 21 日 19 時～8 月 22 日 11 時の 3 種類である。簡単のため，それぞれイベント 1，イベント 2，イベント 3 として表現する。

4.2 適用結果と考察

鴨川の荒神橋地点における 1 時間先の水位を予測するが，予測期間は 1 年間とし，2000 年 1 月 1 日 0:00～12 月 31 日 23:00 を対象とする。2000 年の 1 年間を時間に換算すると 8784 時間になり，観測値の欠損期間を除くと 8565 時間となった。よって，2000 年以外のデータから TCBM のモデリングを行い，2000 年において 1 時間先予測を 8565 回行った結果を予測値として出力し，観測値と比較し検証することと

する。

84 個の入力変数候補からステップワイズ法により変数を選択した結果を，選択変数 1 つの場合，2 つの場合，・・・，6 つの場合に分けて Table 1 に示す。これは選択変数の数それぞれについてステップワイズ法が選んだ，最適な入力変数の組み合わせである。設定条件で述べたように，モデル変数を 4 つとして TCBM のモデリングを行うことにする。選ばれた入力変数は，上賀茂-水位(1 時間前)，京都-積算降雨(3 時間)，大原-積算降雨(24 時間)，大原-積算降雨(120 時間)の 4 つ(以後それぞれを変数 x1, x2, x3, x4 と呼ぶ)であった。

続いて，量子化数を「出力分布条件」と「連続性条件」の 2 つの指標を用いて決定し，それぞれの条件を満たした割合を評価指標充足率として算出する。変数 x1～x4 それぞれについて量子化数を 1 から順に増やしていき，評価指標充足率の増加に収束が見られる部分付近の量子化数をその変数の量子化数として採用することにする。Fig. 6 に，変数 x1 の量子化数を 1 から増やしたときの評価指標充足率の変遷を示す。横軸が量子化数，縦軸が評価指標充足率(%)である。

Table 1 Obtained parameters on points

78					
78	3				
78	3	28			
78	3	28	32		
78	3	26	32	58	
78	3	26	32	58	68

Number	Name	Attribute	Conditions
78	Kamigamo	WL	1 hour delay
3	Kyoto	AP	3 hours
28	Ohara	AP	24 hours
32	Ohara	AP	120 hours
58	Prefecture hall	AP	1 hour
68	Prefecture hall	AP	48 hours
26	Ohara	AP	15 hours

WL: water level,
AP: Accumulated precipitation

Fig. 7 を見ると，量子化数を増やすほど評価指標充足率は増加していくが，量子化数が 45 を過ぎたあたりから増加傾向に収束がみられる。そこで充足率が初めて 2 つとも 80%を超えた量子化数（49）を変数 x1 の量子化数として採用する。変数 x2，x3，x4 についても同様の方法で量子化数を決定し，結果として変数 4 つの量子化数は x1：49，x2：46，x3：40，x4：17 となった。

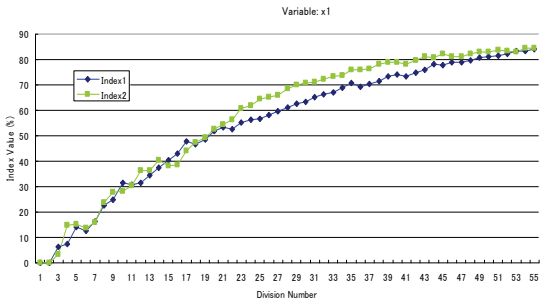
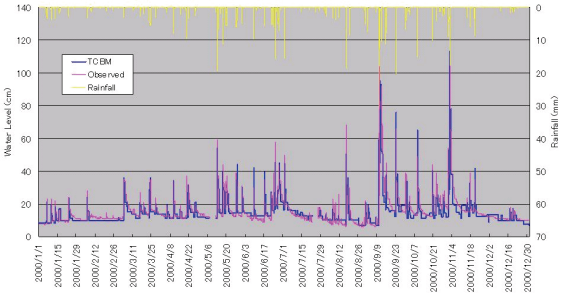
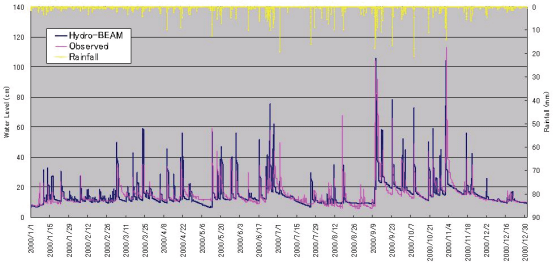


Fig. 7 Sufficient rate against division number on variable x1 where Index1 is output condition and Index2 is continuous condition



(a) At Kamogawa point



(b) At Kojin bridge point

Fig. 8 One hour ahead prediction of water level with observed in 2000; (MAE=2.08, RMSE=3.50 (cm))

Fig. 8 に，鴨川と荒神橋地点での一時間先水位予測の再現結果を示す。つまり，1 時間先の降水量には TCBM の精度を確認するため観測値を用いている。降水量には上賀茂と京都，大原の平均値をとっている。図を見て分かるように，かなり良い精度で再現できている。このことから，データ蓄積型モデルが河川水位の予測にも適応できることが分かった。本

研究では 1 時間先の降雨量に観測値を使ってはいるが，入力変数を積算降雨としていることで，もし予測降雨量に多少のずれがあったとしてもそれほど大きな誤差にはつながらないであろうことが推測できる。図からは全体の予測値の変遷を見ることができる。ピーク予測値にある程度の大小はあるが，平均誤差:2.08(cm)や RMSE:3.50(cm)の数値を見ても分かるように，全般的に基底水位も含め良い精度で予測できている。月平均流量に関しては，梅雨時の 6 月よりも 9 月や 11 月のほうが大きくなっている。9 月に関しては 9 月 8 日～13 日にかけて台風 14 号が日本の西の海上を北上しており，その期間の降水量が非常に大きかった影響で流量が増大し，月平均流量も大きくなっていると考えられる。この期間の AMeDAS 京都地点の 72 時間降雨観測量は 1980 年以降では 4 番目に大きな記録となっている。11 月に関しても 11 月 2 日に発生した豪雨などによる影響のため，月平均流量も大きくなっていると思われる。流量の増大や減少の様子をより如実に見られるように一時間ごとの流量分布の変化の様子を示した。5 月 13 日 19 時～21 時の空間分布図の変化の様子を見れば，ある大きな雨により上流で流量が一気に増え，それが下流へと伝わっていく様子がよく分かる。

Table 2 Obtained parameters for FFM

Name of point	Attribute	Conditions
Kibune	AP	480 hours
Kamigamo	AP	48 hours
Kibune	AP	15 hours
Kamigamo	AP	120 hours

Table 3 Obtained parameters for water level prediction at Kamigamo

Model and points	Attribute	Conditions
Model A for flood		
Kamigamo	AP	6 hours
Prefecture hall	AP	3 hours
Prefecture hall	AP	24 hours
Kamigamo	WL	6 hour ahead
Model B for no flood		
Kamigamo	WL	Real-time
Kyoto	AP	3 hours
Ohara	AP	24 hours
Ohara	AP	120 hours

以下では実時間洪水予測システム（以下，本システムと呼ぶ）の適用結果を示す。イベント 1 では 3 時間先までの水位予測，イベント 2 とイベント 3 では 6 時間先までの水位予測を行う。これは降水短時間予報で 6 時間先までの降雨予報が発表されるようになったのは 2001 年 4 月以降であり，それ以前は 3 時間先までの予報値しか存在しないためである。例としてイベント 3 における京都地点の 6 時間先までの降水予報値と AMeDAS 観測値との比較を行う。予測精度が良いとは言いがたい結果であった。やはり，より先の時間を予測すると精度は悪くなるという印象である。FFM 内の洪水水位予測用と通常水位予測用の 2 つのモデルの変数選択結果を Table 2 に示す。ここに，AP は積算降雨，WL は水位を表す。モデル A が洪水水位予測用モデル，モデル B が通常水位予測用モデルである。

本システムは 6 時間先までの水位予測を行うため，モデル A とモデル B の積算降雨データは観測データと先の予測降水量を組み合わせて作成する。また，モデル B には上流の上賀茂の実時間水位がモデル変数として採用されているため，上賀茂水位を予測するためのモデルを作成する。Table 3 に上賀茂水位予測のためのモデル変数を示す。同時に，荒神橋における洪水イベント時は上賀茂水位も高くなると予想され，データの少なさから値が欠損となることを考えて GP による値の補正も同時に行う。学習データから GP で導かれた上賀茂水位予測式は次のようになった。

$$Q1 = -\left(\left(\begin{array}{l}\sqrt{\sqrt{c}\times\left(39.3\times\sqrt{b}\right)} \\ -\left(44.1+\sqrt{exp\left(\sqrt{\sqrt{d+47.1+\sqrt{d\times c\times b}+b}\right)\times\sqrt{a}}\right) \\ -\sqrt{39.2}\end{array}\right)-\sqrt{b}\right) \tag{8}$$

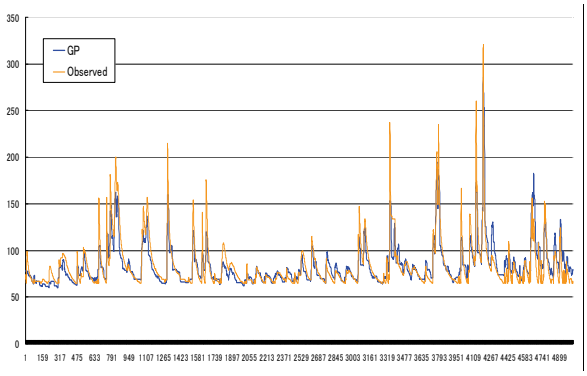
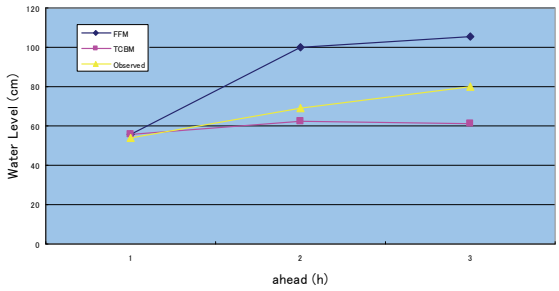


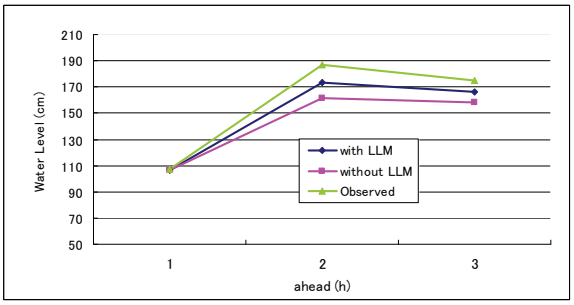
Fig. 9 Predicted water level with GP at Kamigamo point

ただし，a は貴船 AP(480h)， b は上賀茂 AP(48h)， c は貴船 AP(15h)， d は上賀茂 AP(120h) であり，Q1 が上賀茂水位である。学習データとしては，データが不足しがちな高水位時の補正に重点を置き，65cm 以上のデータを学習データとして学習を行った。 Fig. 9 に上賀茂用 TCBM と GP による上賀茂 6 時間先水位予測結果の中で，実際に GP による値の補正を行ったケースを示す。

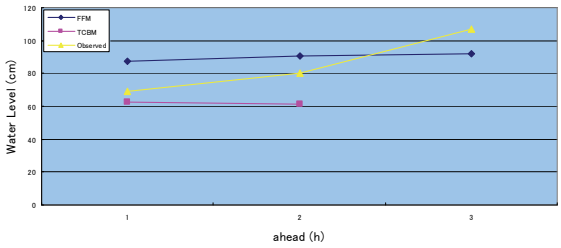
求められた上賀茂水位と予測降水量，観測降水量を用いて FFM のインプットデータを作成した。さらに，Hydro-BEAM シミュレーションデータによる FFM の洪水データの作成を行った。具体的には，1999 年～2000 年において雨を 1.5 倍，1.8 倍，2.1 倍，2.4 倍， 0.5 倍に変化させ，その雨を用いて行った Hydro-BEAM シミュレーションの結果のうちで，モデル A 内にデータとして入れる基準に設定した水位



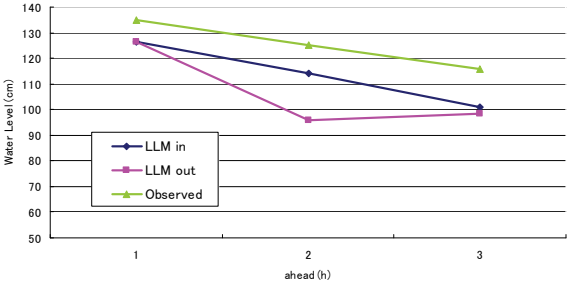
(a) At 4:00 in June 27, 1999



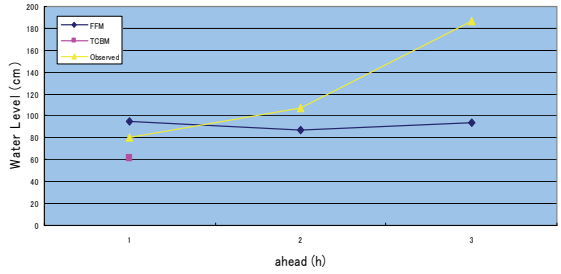
(a) At 4:00 in June 27, 1999



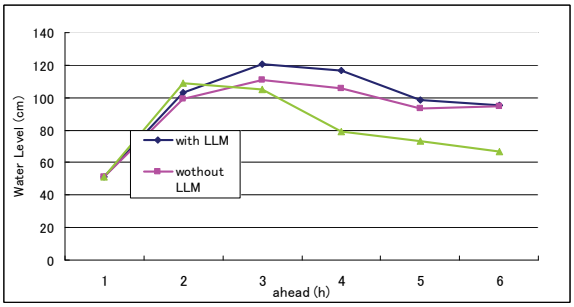
(b) At 5:00 in June 27, 1999



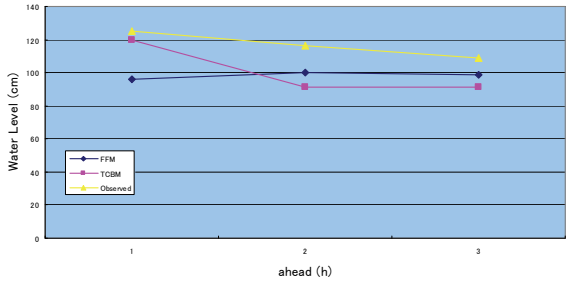
(b) At 5:00 in June 27, 1999



(c) At 6:00 in June 27, 1999

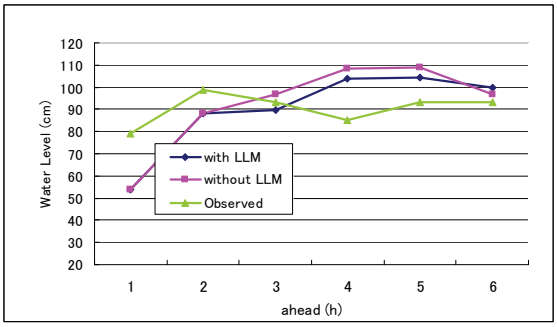


(c) At 6:00 in June 27, 1999



(d) At 13:00 in June 27, 1999

Fig. 10 Predicted results



(d) At 13:00 in June 27, 1999

Fig. 11 Comparison between predicted results with

80cm 以上のデータを蓄積させた。合計 820 個のデータが作成され，信頼度 R とともにモデル A 内の事例ベースに蓄積された。信頼度 R は式(6)を計算した結果，R=0.16 となった。この信頼度 R は予測の際に利用する。

以上によって構築された FFM にインプットデータを与えて実時間洪水予測を行った。イベント 1 の予測結果を Fig. 10 に示す。(b)と(c)の図から，系列的な水位変動については，データを増やした効果が見てとれる。効果割合の定量的把握を行い，実用性の確認を行う必要がある。図(c)の 4～6 時間先の予測や(d)の 2 時間先予測で値に大きな改善が見られる。LLM 概念による値の改善はピーク値の立ち上がり部分に特に効果を発揮すると言えそうである。ただ，(b)では途中に観測値と真逆の経緯をたどっている。また(b)や(d)を見ると，いったん予測値を外すとその後の予測値に悪い影響を与えてしまうケースがある。イベント 3 の洪水予測においては大きな改善効果があったことが分かる。(a)～(d)のどの図においても FFM が TCBM よりも観測値にかなり近い挙動をとっており，大きく外れる値もとっていない。これらの結果から，ピークの立ち上がりだけでなく，だらだらと高い水位を維持しているケースに対して時系列情報を用いた予測が効力を発揮するということが分かる。図を見ると，ピーク時に適切に洪水と判断してモデル A に予測を委ねており，導入の効果があったことが分かる。別の視点として，Hydro-BEAM シミュレーション値を事例ベースに加えたことによる効果である。洪水イベントの推定値出力割合が 75.0(%)から 94.4(%)に向上したことで，数値として表された。つまり，「推定値なし」が出力されるという問題が約 20(%)改善されたことになる。

LLM 有りと LLM 無しの結果を観測値と共に Fig. 11 (a) ~ (d)に示す。LLM の導入により，(a)ではピークの立ち上がり，(b)ではピーク後の減衰の表現において精度に改善が見られる。(d)でも高水位時の挙動の追跡に精度の改善が見られる。一方，(c)では 2 時間先の予測値には改善が見られたが，3 時間先の予測で値を外して以降，LLM 有りの方で精度が悪くなっている。やはり一度予測を外した後に値が正値と離れていってしまう場合があるようである。まとめると，LLM の導入はピークの立ち上がりやピーク後の減衰に効果を発揮するが，一旦， 値を外した後の予測に弱さを見せることがあると分かった。

続いて，TCBM による河川メッシュのみを対象とした分布型洪水予測の結果について検証する。「緯度」と「経度」として選んだモデル変数を選んだ結果を Table 4 に示す。先に緯度が選ばれ，経度が選ばれたのは 5 番目であった。

Table4　Obtained parameters for distributed runoff model

Name of point	Attribute	Condition
Kyoto	AP	6 hours
Latitute		
Kojin briudge	WL	Real-time
Keihoku	AP	6 hours
Longitude		

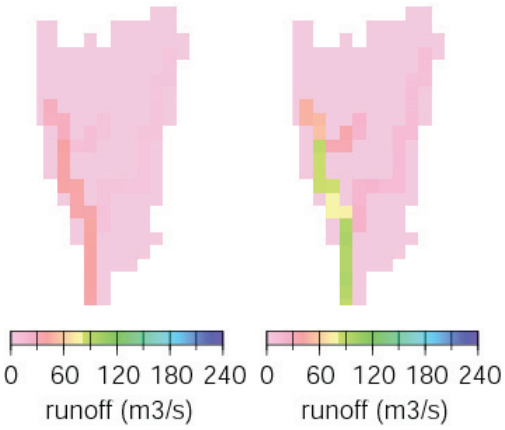
前述のように，緯度と経度の量子化数は単位幅が Hydro-BEAM のメッシュサイズと同じになるように設定する。メッシュサイズを 1km に設定したので，位相空間内の緯度・経度方向の単位幅が 1km となるように量子化を行った。その他の 3 つの変数については今までと同様に「出力分布条件」と「連続性条件」を用いて決定した。結局，決定された 5 変数の量子化数は変数の 1 つ目から順に 34，23，28，21，17 となった。

最後に，実時間洪水予測システムによって求められた，荒神橋地点の実時間水位を用いて，6 時間先までの実時間分布型洪水予測の結果を Fig. 12 の(a)~(f)に示す。2001 年 6 月 19 日 19 時において，6 時間先までの鴨川流域の河川メッシュの洪水分布を予測したものである。ピーク部分での予測精度に課題は残るものの，分布型で出力が可能になったことは TCBM の汎用化という観点からは意義が大きい。

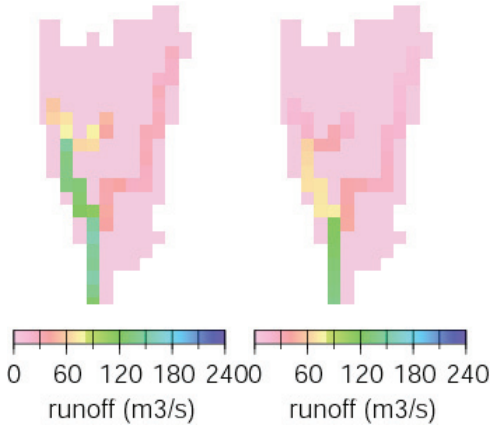
5. 結語

本研究では事例ベースモデル（TCBM）をより汎用性のあるモデルにすることを研究目的とし，時系列情報を重視したモデルへ拡張した TCBM による実時間洪水予測と，分布型での流量情報の提供を可能にした TCBM による分布型洪水予測を行った。またその際，実時間で 6 時間までの洪水予測を行うため，時系列予測に優れた LLM の概念を導入して TCBM の拡張を行った。さらに，分布型流出モデルである Hydro-BEAM を利用して洪水特性の把握を行い，汎用性拡大のために利用した。ここで，得られた研究成果をまとめると以下のようになる。

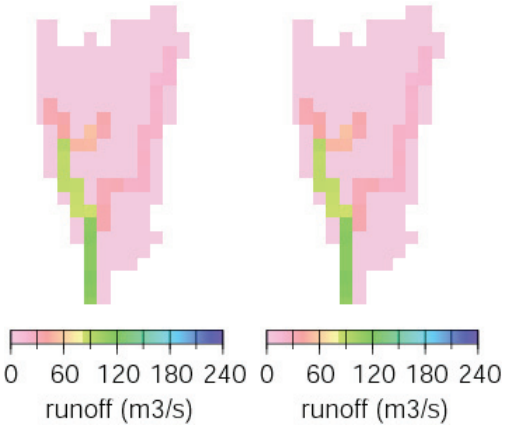
- TCBM を知識ベースシステムの一手法と解釈し，事例ベース内の位相空間をプロダクションシステムの形式で表現できることとした。
- TCBM での実時間予測にローカルリニアー(LLM)概念による時系列予測を組み込むことが可能であり，ピーク値への立ち上がり挙動やピーク水位の緩やかな減衰の予測に効果を発揮できた。



(a) One hour ahead (b) Two hour ahead



(c) Three hour ahead (d) Four hour ahead



(e) Five hour ahead (f) Six hour ahead

Fig. 12 Real-time flood prediction results at 18:00 of June 19

・観測情報が少ない地点での予測に AI 手法の一つである GP を利用することで欠損値の補正が可能になり, TCBM による水位予測の汎用性を上げた。
・対象流域の洪水特性を把握した分布型流出モデルによるシミュレーションデータを蓄積データとして格納することで，事例の希少ない場合出力データを安定にした。
・分布型流出モデルによる流量情報を蓄積することで河川メッシュを対象とした分布型洪水予測が可能となり，危険地域や避難可能地域の空間的分布を把握できるようになった。

参考文献

伊庭斉志（1996）：遺伝的プログラミング，東京電機大学出版局，pp.13-27.

小尻利治・藤井忠直(1990)：知識ベースを用いたファジイ貯水池操作に関する研究，土木学会水工論文集，34，pp. 601−606.

小尻利治・東海明宏・木内陽一（1998）：シミュレーションモデルでの流域環境評価手順の開発，京都大学防災研究所年報，第 41 号 B-2, pp.119-134.

小林重信（1997）：知識工学，昭晃堂.

関井勝善・　スミス，J.ポール・小尻利治（2007）：分布型を考慮した AI 手法による実時間流出予測モデルの構築，水文水資源学会誌，Vol.20, No.6, pp.329-340.

高棹琢馬・椎葉充晴・中北英一(1983)：レーダー雨量計による短時間予測の検討，京都大学防災研究所年報，第 26 号 B-2, pp.165-180.

戸内順一（1995）：人工知能入門，日本理工出版会.

Babovic, V. and Fuhrman, D. R. (2001) : Data assimilation and error prediction using local models, D2K Technical Report 0401-2, pp.25-30.

Smith, P. J., and Kojiri, T., (2003) : Probabilistic short-term distributed flood forecasting, Annuals of Disaster Prevention Research Institute, Kyoto University, No.46B, pp.885-897.

Yamatake Corporation (2004): dataFOREST Version 4.0.0 マニュアル

A Study on Flood Forecasts using TCBM

Kentaro WADA*, Toshiharu Kojiri , Kazuya HARAYAMA**
Kenji TANAKA and Toshio HAMAGUCHI

* Graduate School of Engineering, Kyoto University
** Yamatake, Ltd., Japan

Synopsis

The aim of this study is to improve TCBM (Topological Case-Based Modeling) and to apply it for flood forecast by introducing other two concepts. One is LLM (Local Linear Model) thinking to use time-series information which is generally thought to be important for flood forecast. The other is to increase the data volume of flood events using the distributed runoff model (Hydro-BEAM) to cover the weak point of limited flood-case data. This enhanced TCBM is called FFM (Flood Forecast Model) for real-time flood forecasting system with six hour ahead forecast in this study. As a result, the effect of improving TCBM is found and the versatility of TCBM thought to be expanded.

Keywords: Flood prediction, Topological case-based modeling, Local linear model, Distributed runoff model