

COMPARISON OF DISTRIBUTED AND LUMPED HYDROLOGICAL MODELS

Huaxia YAO¹, Michio HASHINO², Akira TERAOKA³ and Toshiro SUZUKI⁴

¹Member of JSCE, Dr. of Eng., Assoc. Professor, Dept. of Civil Engineering, The University of Tokushima (2-1, Minami-josanjima, Tokushima 770, Japan)

²Member of JSCE, Dr. of Eng., Professor, Dept. of Civil Engineering, The University of Tokushima (2-1, Minami-josanjima, Tokushima 770, Japan)

³Member of JSCE, M. Eng., Head, Environmental Planning Division, Public Works Research Institute, Ministry of Construction (1 Asahi, Tsukuba, Ibaraki 305, Japan)

⁴Member of JSCE, M. Eng., Research Engineer, Hydrology Division, Public Works Research Institute, Ministry of Construction (1 Asahi, Tsukuba, Ibaraki 305, Japan)

A physically based distributed model, consisted of grid cell-sub-model of SVAT type and river routing sub-model of linear reservoir type, is proposed. Grid cell-specific meteorological variables are produced by a special regression method and applied to the Fuji River basin divided into 3,376 grid cells. On the other hand, the basin is regarded as one grid and a lumped model is considered, with its input being the averages of distributed data. Then these two models are compared as for differing situations. They give similarly good runoff process when rainfall on basin is large and relatively uniform. However they behave differently when rainfall is not large and very non-uniform in spatial distribution. The lumped model underestimates flood peak flows. And the distributed model is applied to analysis of hydrological response to possible landuse change, while it is difficult to do so with the lumped model. Conclusively distributed model could be more appropriate than lumped model for a large basin.

Key Words : Distributed model, lumped model, comparison, Fuji River basin, spatial data allocation, runoff

1. INTRODUCTION

Hydrological models, either physically-based or conceptual, may be divided into two types: lumped and distributed model¹⁾. The lumped model has a long history of development and is commonly accepted in research and application. However it is pointed out that a lumped model may not well apply to the large inhomogeneous catchment because it does not include spatial difference within the catchment, although it does behave well for a small or rather homogeneous catchment. On the other hand, the distributed model, having a shorter history of research, has been given much attention in recent years, especially when DEM(digital elevation models) and GIS(geophysical information system) are becoming more and more available. But it can not be used to some areas without enough data.

The objective of this paper is to compare these two kinds of models, by using same data in a large-

scale basin of the Fuji River in Japan. Problems such as in what situation two models give different output or give similar output, in what practical usage the distributed model can be applied but the lumped model can not be applied, are examined.

First, the basin (3,432 km²) is separated into many squared grid cells of 1 km and the distributed model is set up, with sub-model for hydrological process in each grid cell and sub-model for river routing between grid cell and basin outlet. Second, the basin is treated as one big square (whole basin) and the lumped model is set up, with same structure as the distributed model. Third, the two models are operated and compared under different precipitation conditions. Finally, hydrological response to a supposed landuse change scenario is simulated by the distributed model.

2. DISTRIBUTED MODEL

(1) Input Data

The Fuji River basin is located at central-southern Japan (Fig. 1), with the flat Kofu plain in center being surrounded by steep mountains high up to 3,000 meters. Most of basin surface is covered by forest and grass or crop plants. The main river length is 128 km. 3,376 grid cells are created for the drainage area of discharge gauge at Kitamatsuno.

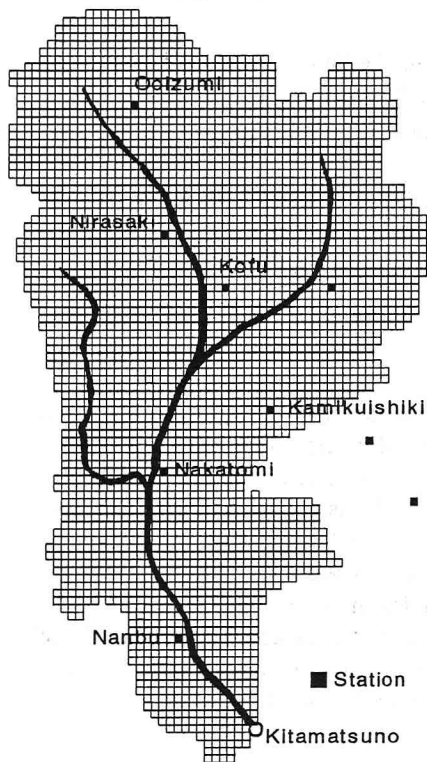


Fig.1 Fuji River basin.

Meteorological variables at each grid cell, which are the input data of hydrological model, have to be derived. It is thought that meteorology and geography (or topography) have a certain relationship within a given region. These relationships can be found from the ten weather stations in and around the basin, as they are representative of basin's meteorology and geography. Supposing Y is any meteorological variable such as temperature, sunshine, wind and precipitation; X_1 , X_2 and X_3 are latitude, longitude and elevation of a station (or grid cell) respectively; then a linear relationship is expressed as,

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \quad (1)$$

where b_1 , b_2 and b_3 are regression coefficients and are derived by a step-wise regression algorithm²⁾.

Data of 16 years are used to derive these four formulas, giving good correlation for monthly and daily regressions. Then the geographical values of all grid cells are taken from the national GIS database³⁾, and daily meteorological values in each grid cell are produced by using these formulas and

observed data in those weather stations. For example, distribution of annual precipitation in 1990 is shown in Fig. 2.

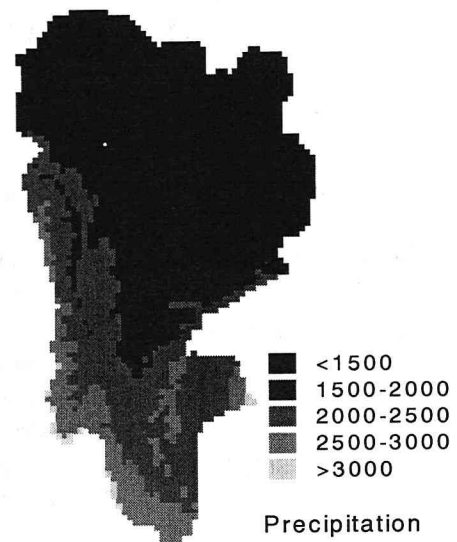


Fig.2 Spatial precipitation distribution.

(2) Grid Cell Sub-Model

The sub-model for any grid cell is formed and simplified based on a former work⁴⁾, and is illustrated in Fig. 3. It is a kind of soil-vegetation-atmosphere-transfer (SVAT) model. Water and energy transfer into and out of the system consisted of boundary atmosphere, plant canopy, surface snow or water, upper soil, lower soil and ground aquifer. Model structure and parameter are briefly explained.

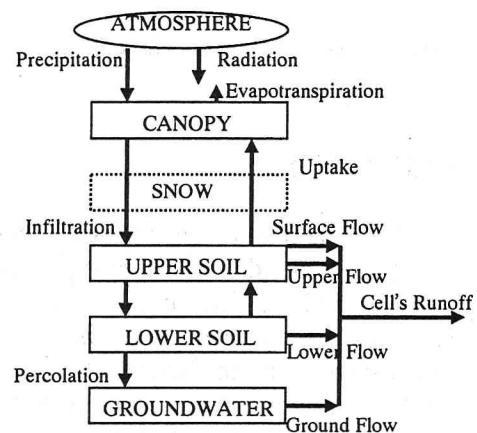


Fig.3 Model frame.

Firstly potential evaporation is estimated with Penman equation, using grid cell meteorological and landuse data. Precipitation in a day is recognized into snowfall or rainfall by means of critical temperature (1.9 °C for Fuji basin). When daily-mean temperature is lower than this critical one, precipitation is in form of snowfall. Then snow accumulation and melting on canopy and ground surface are estimated. Rainfall interception and its

evaporation are calculated from canopy interception capacity S_m and potential evaporation E_p .

$$E_C = E_P \quad (A_i P - A_i S_m > 0) \quad (2)$$

where A_i is the canopy area ratio, P is rainfall.

When there is no intercepted water on canopy, evapotranspiration from plants and soil takes place.

For any time step (a day or an hour), after evapotranspiration is estimated, various outflows of soil layers and aquifer are estimated by flow-storage equations, and soil storage change is made step by step, using water budget equation.

Overland flow is estimated by

$$R_S = P_X + W_U - E_{SU} - W_{US} \quad (3)$$

where P_X , W_U , E_{SU} and W_{US} are the ground rainfall, the water storage, evapotranspiration and saturation storage of upper soil layer.

Lateral and vertical outflows of upper soil are expressed as

$$R_U = K_{UL}(W_U - W_{UC}) \quad (4)$$

$$V_U = K_{UV}(W_U - W_{UC}) \quad (5)$$

where K_{UL} and K_{UV} are the flow coefficients, W_{UC} is the field capacity of upper soil. And outflows R_L , V_L of the lower soil layer are estimated in a same way.

Groundwater flow is written as

$$R_G = [K_G V_L + (2 - K_G) R_G'] / 2 \quad (6)$$

where K_G is the flow coefficient, R_G' is the flow rate in the former step, and V_L is percolation from soil.

Finally, the sum of overland flow, upper soil flow, lower soil flow and groundwater flow gives the runoff of that grid cell.

$$R = R_S + R_U + R_L + R_G \quad (7)$$

(3) River Routing Sub-Model

The best way of modeling transport and inundation between grid cells and basin outlet may be to draw out the river network from DEM data and then to calculate flow process in the network. This way needs much data and special techniques. Instead, a simpler method is considered for river routing. The detailed pathway from a grid cell to the outlet is neglected, and routing function is expressed by a linear formula.

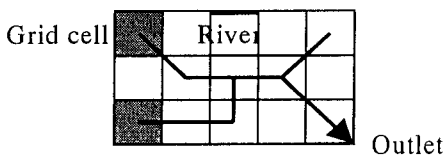


Fig.4 Routing sub-model.

Suppose that there is a river tank (or reservoir) to implement the routing function of the river between a cell and the outlet (Fig. 4). Its inflow is the grid cell-generated discharge I_i (in m^3/s). Its

outflow Q_i is the resulted component of outlet runoff contributed by this grid cell. Obviously there is a tank for each cell. Then continuity equation about tank storage W is written as

$$\frac{dW}{dt} = I_i - Q_i \quad (8)$$

And a linear storage-flow rate relation is hypothesized as

$$Q_i = kW \quad (9)$$

where k is the routing coefficient relating to distance and slope between this grid cell and outlet.

Substituting equation (9) for (8) and writing it in a differential form for an unit time increment give approximately

$$Q_i(t) = k \cdot I_i(t-1) + (1-k)Q_i(t-1) \quad (10)$$

in which $Q_i(t)$ is the discharge at the end of the increment, $I_i(t-1)$ and $Q_i(t-1)$ are inflow discharge and outflow discharge on the starting point of increment.

The sum of routed outflows of all grid cells is the runoff of whole basin at a time step t .

$$Q = \sum_{i=1}^{3376} Q_i(t) \quad (11)$$

This method seems to be conceptual. However, the value of coefficient k varies with each grid cell, therefore the method is of distributed property to some extent. It is explained in more details.

Runoff generated at a grid cell far away from the basin outlet would experience extensive attenuation before reaching the outlet. k might be expressed as an adverse function of the distance D from grid cell to outlet as follows.

$$k = k_{mi} + \frac{D - D_{ma}}{D_{mi} - D_{ma}}(k_{ma} - k_{mi}) \quad (12)$$

in which the farthest grid cell (with largest D_{ma}) corresponds to the smallest k_{mi} ; and the nearest grid cell (with smallest D_{mi}) corresponds to the largest k_{ma} . Of course the thresholds k_{mi} and k_{ma} (0.03 and 0.13 for the basin) should be optimized.

(4) Calibration of Distributed Model

Usually it is difficult to calibrate parameters in a distributed model. If there are lots of experimental and survey data about relations of parameters and affecting factors (soil, geometry, vegetation), parameter values can be assigned onto each grid cell, as did by Kuchment *et al.*⁵⁾ Unfortunately those relations are not available for present basin of study. Therefore another alternative method is tried.

There are twelve parameters in this distributed model: canopy area ratio, canopy water capacity, field capacity and saturation storage of the upper soil layer, field capacity and saturation storage of

the lower soil layer, vertical and lateral outflow coefficients of the upper soil, vertical and lateral outflow coefficients of the lower soil, outflow coefficient of the groundwater aquifer, and routing coefficient of the river.

The canopy area ratio and interception capacity are directly assigned by using land use type obtained from GIS, as shown in **Table 1**.

Table 1 Canopy parameters.

Land use	Area ratio	Capacity (mm)
Paddy land	1.0	0.3
Dry farm	0.8	0.5
Fruit tree	0.8	0.8
Other tree	0.8	0.8
Forest	1.0	2.0
Waste land	0.5	0.3
Urban land	0.0	0.0
Rural house	0.1	0.1
Traffic land	0.0	0.0
Lake	0.0	0.0
River-beach	0.1	0.1
Sea beach	0.0	0.0

Nine soil and one river-routing parameters are important, controlling runoff generation, separation and inundation. Their calibration is carried out by two steps. The spatial variation patterns of them are firstly hypothesized, then the critical thresholds in these patterns are optimized by trial-and-error method, minimizing deviation of estimated runoff from the observed one. For example, the upper soil layer becomes generally deeper and more condensed as the surface slope gets flatter from mountain to plain. As a result, the water storage capacity W_{US} is supposed to be adversely proportional to the slope S of a grid cell as follows.

$$W_{US} = W_{USmi} + \frac{S - S_{ma}}{S_{mi} - S_{ma}} (W_{USma} - W_{USmi}) \quad (13)$$

in which W_{USmi} is the smallest storage capacity at the grid cell with the largest slope S_{ma} ; W_{USma} is the largest capacity at the grid cell with smallest slope S_{mi} . On contrast, water storage capacity of the lower soil is positively proportional to the slope.

In general, soil matrix is more porous and permeable in higher areas than in lower areas since forest and macro-pores exist mainly in higher place. Then soil flow parameters are supposed to be positively proportional to grid cell elevation.

$$K_U = K_{Umi} + \frac{Z - Z_{mi}}{D_{ma} - D_{mi}} (K_{Uma} - K_{Umi}) \quad (14)$$

where K_U is the flow coefficient of upper soil for a grid cell having an elevation Z , K_{Uma} is the largest threshold value occurring at the grid cell with the largest elevation Z_{ma} , and K_{Umi} is the smallest threshold value at the grid cell with the smallest

elevation Z_{mi} .

In that way parameter values are allocated to each grid cell by means of GIS data.

Three years (1990-1192) of hydrological data are used to calibrate the model. Daily runs give good annual water budgets and runoff volume. Annual runoff errors against the observed ones are 1.1%, -0.1% and 10.5% for 1990, 1991 and 1992 respectively.

Parameter value ranges (thresholds) of hourly-run model are listed in **Table 2**. For instance, storage threshold W_{USmi} and W_{USma} for the upper soil are 250 mm and 310 mm respectively; Flow coefficient threshold K_{Umi} and K_{Uma} for the upper soil are 0.0085 hr⁻¹ and 0.025 hr⁻¹ respectively.

Table 2 Parameter values.

Parameter	Value
Upper soil	
field capacity (mm)	200-250
saturation capacity (mm)	250-310
flow coefficient (1/hr)	0.0085-0.025
Lower soil	
field capacity (mm)	400-600
saturation capacity (mm)	460-700
flow coefficient (1/hr)	0.003-0.006
Groundwater	
flow coefficient (1/hr)	0.02-0.035
River	
routing coefficient (1/hr)	0.03-0.13

Estimated hourly processes for two periods of 15 days, Periods A and B, are shown in **Fig. 5** and

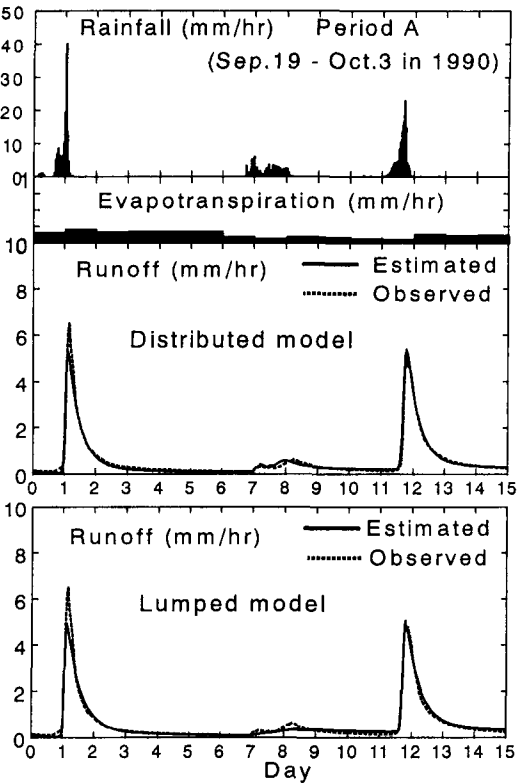


Fig.5 Simulation of two models for Period A.

Fig. 6. The rainfall in two figures is basin-averaged values. Runoff is that of basin outlet, and is well fitted with the observed.

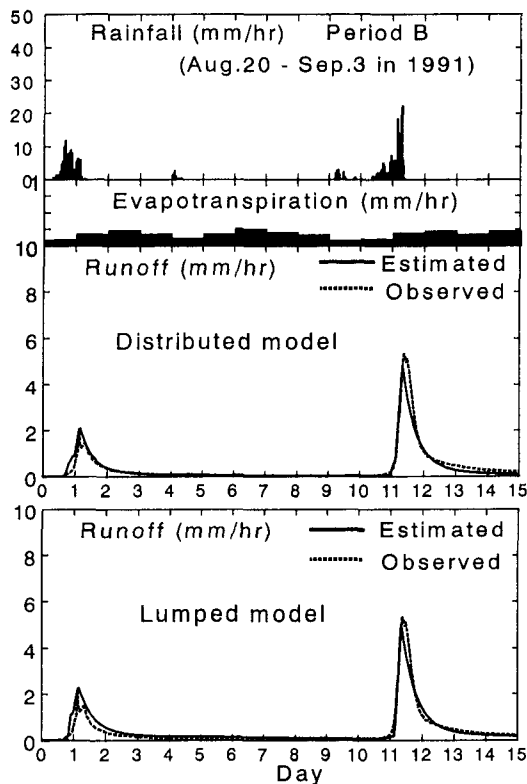


Fig.6 Simulation of two models for Period B.

4. LUMPED MODEL

Physical structure of the lumped model is same as that of distributed model. The only difference is that whole basin is regarded as only one grid (itself). Rainfall-runoff calculation on the averaged basin is carried on, and an integrated river routing is similarly carried to give the final runoff at the outlet.

All inputs of this lumped model are produced by averaging meteorological and geographical data distributed in all grid cells. Parameter values are also assigned by averaging those distributed values in the distributed model. Then the lumped model is operated toward the same period as done by the distributed model.

5. COMPARISONS

Behavior and estimation accuracy of the two models are compared under several situations.

(1) Large and Uniform Rainfall

The rainfall events during Period A (Sept. 19 – Oct. 3 in 1990) and Period B (Aug. 20 – Sept. 3 in 1991) are relatively heavy and uniform in spatial

distribution, as shown by the rainfall amounts of six stations from south to north in **Table 3**.

Table 3 Station rainfall for period A and B .

Station	Rainfall (mm)	
	A	B
Nanbu	573	366
Nakatomi	395	304
Kamikuishiki	317	299
Kofu	283	223
Nirasaki	328	217
Ooizumi	211	204

Results of lumped model for these two periods are also plotted in **Fig. 5** and **Fig. 6**. By comparing the results, it is seen that the lumped model may produce good simulations, almost same as the results of distributed model. That is because rainfall has no evident spatial difference. In other words, the lumped model can also work well in that situation.

(2) Moderate and Non-uniform Rainfall

Selecting another two periods, C (Aug. 9 – 23 in 1990) and D (Jul. 30 – Aug. 13 in 1993). Rainfall events are moderate or small but not uniform within the basin (see **Table 4**).

Table 4 Station rainfall for period C and D.

Station	Rainfall (mm)	
	C	D
Nanbu	327	293
Nakatomi	206	116
Kamikuishiki	397	189
Kofu	113	82
Nirasaki	184	86
Ooizumi	94	101

The runoff results of two models are shown in **Fig. 7** and **Fig. 8**. They are different. And the distributed model behaves better than lumped model in that case. The averaging treatment causes lumped model to underestimate flood runoff. That also means distribution of rainfall have a key role in runoff generation.

(3) Landuse Change Scenario

A scenario of landuse change is supposed. The lower part of basin, possessing 1/3 area of whole basin, be changed into urban and suburban district in future. Then former plants area loses and interception capacity becomes zero at these 1125 grid cells, ground reflectance of radiation becomes 0.16, soil saturation storage becomes probably decreased by 20%, and soil flow coefficient is also decreased by 20%. Applying daily-run distributed model to this future scenario, and comparing the results with that of present (1991) situation, responses in annual water budget can be drawn out

as Table 5.

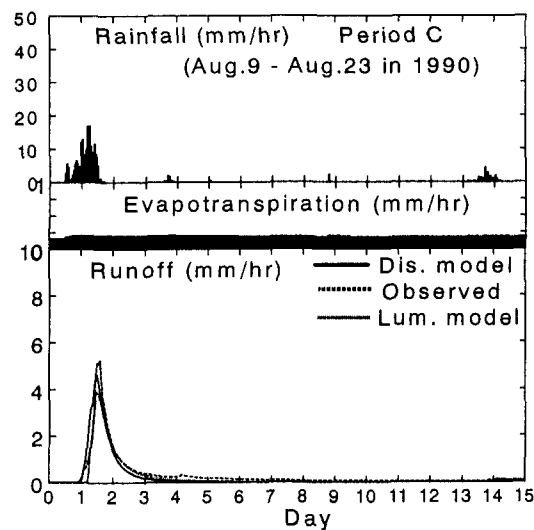


Fig.7 Comparison of two models for period C.

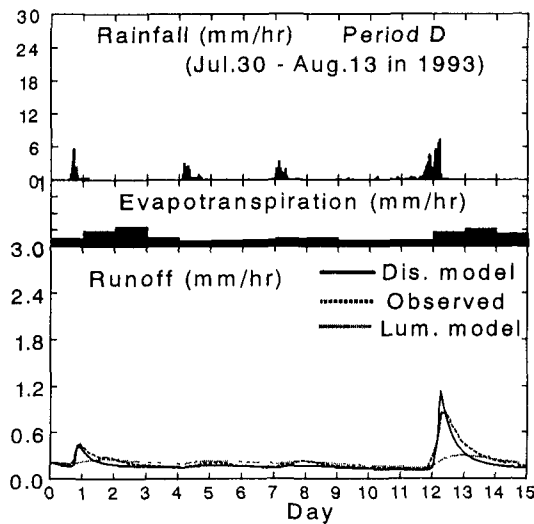


Fig.8 Comparison of two models for period D.

Table 5 Hydrological changes (unit: mm) .

	present	future	change
Evapotranspiration	846.9	792.3	-6.4%
From interception	233.6	147.5	-36.9%
From soil	613.3	644.7	5.1%
Runoff	838.1	1012.1	20.8%
Surface	462.3	810.8	75.4%
Upper soil	104.6	59.1	-43.5%
Lower soil	44.2	21.4	-51.6%
Groundwater	227.6	121.6	-46.6%

Annual evapotranspiration would decrease by

6.4%, because the evaporation from canopy interception decreases by 36.9%. Runoff would increase by 20.8%, because the surface flow gets remarkably larger (75.4% increase). But soil flow and groundwater flow get decreased. Although such kind of prediction could not be tested directly at present stage, it shows the possibility of using distributed model to analyze hydrological response to landuse change.

Unfortunately, it is difficult to apply the lumped model for this purpose, as the adjustment of lumped parameters under partial-area change has not been well understood.

6. CONCLUSION

A distributed model and a lumped model are studied and compared with each other as for the Fuji River basin. The distributed model is proved applicable to the basin. When rainfall is heavy and relatively uniformly distributed in the basin, both of the models may regenerate hourly runoff quite well, without clear difference. When rainfall is moderate or small and spatially non-uniform, the two models behave differently, distributed model is better while lumped model tends to underestimate the flood peaks. Furthermore the usage of distributed model to landuse change problem is explained, it might be an advantage over the lumped model.

REFERENCES

1) Maidment, D. R.: *Handbook of Hydrology*, McGraw-Hill, New York, 1993.
2) Yao, H., Hashino, M. and Yoshida, H.: Prediction of future changes in climate and water cycle by using step-wise regression method, *J. of Japan Society of Hydrology and Water Resources*, Vol. 8, pp.574-582, 1995.
3) Geographical Survey Institute: *Users Guide of Digital Maps*, Japan Map Center Press, 1994.
4) Yao, H., Hashino, M. and Yoshida, H.: Modeling energy and water cycle in a forested headwater basin, *J. of Hydrology*, Vol. 174, pp.221-234, 1996.
5) Kuchment, L. S., Demidov, V. N., Naden, P. S., Cooper, D. M. and Broadhurst, P.: Rainfall-runoff modelling of the Ouse basin, North Yorkshire: an application of a physically based distributed model, *J. of Hydrology*, Vol. 181, pp.323-342, 1996.

(Received September 30,1997)