LARGE-SCALE RUNOFF ROUTING MODELING USING TOPMODEL

Roberto Valmir DA SILVA¹, Yosuke YAMASHIKI² ,Kenichi TATSUMI³ and Kaoru TAKARA⁴

¹Student Member, Graduate student, Dept. of Urban and Environment Eng., Kyoto University (Kyoto 615-8540, Japan) E-mail: roberto.dasilva@flood.dpri.kyoto-u.ac.jp

²Fellow of JSCE, Dr. Eng., Associated Professor, DPRI, Kyoto University (Gokasho, Uji 611-0011, Japan)

³Student Member, Graduate student, Dept. of Urban and Environment Eng., Kyoto University

⁴Fellow of JSCE, Dr. Eng., Professor, DPRI, Kyoto University (Gokasho, Uji 611-0011, Japan)

The objective of this work is to apply a modified TOPMODEL approach in two large-scale basins, the Amazon and Yangtze basins, in order to verify whether TOPMODEL is a feasible model as a Runoff Routing Model (RRM), thus being coupled to Generalized Circulation Models (GCMs) and closing the water cycle. TOPMODEL is also applied for backcasting the period from 1990 to 2008 using the National Centers for Environmental Prediction (NCEP) climate data. This modified approach is based on TOPMODEL-GRASS module code and introduces a multi-velocities parametrization. The modified approach implemented a power law relationship between cumulative area, calculated according to the Global Drainage Basin Database (GDBD) associated dataset, and velocity. This study shows that TOPMODEL can be used as a RRM.

Key Words : Large-scale, GCM, RRM, TOPMODEL, GDBD

1. INTRODUCTION

Rivers are an important component in water cycle, they link the returned water from the land to the oceans¹). In this sense, global scale hydrological modeling is an important issue in order to determine the water quantity and quality inputs into the oceans.

There are many studies suggesting that rivers discharges can affect the ocean circulation, as for example, Campos *et al.*²⁾, cited by Ducharne *et al.*³⁾ suggest that discharge from the Rio de la Plata (Argentina) could be a link between ENSO (El Niño-Southern Oscillation) and the interannual variability of coastal currents along the Atlantic coast of South America.

Generalized Circulation Models (GCMs), representing physical processes in the atmosphere, ocean, cryosphere and land surface, are the most advanced tools currently available for simulating the global climate system. They depict the climate using a three dimensional grid over the globe, typically having a horizontal resolution of between 250 and 600 km, 10 to 20 vertical layers in the atmosphere and sometimes as many as 30 layers in the $oceans^{4)}$.

According to Ducharne *et al.*³⁾, GCMs need the discharges from the rivers to model the water cycling through oceans, atmosphere and land.

GCM have improved with computers and increase knowledge about climate. That is why, in the past decade, GCM modeling using large-scale runoff routing models (RRMs) has received special attention. According to Arora¹⁾, the use of RRMs have basically three purposes: (1) to study the freshwater flux into the oceans, which may affect ocean convection, ocean salinity and ice formation, (2) to evaluate the GCM performance and (3) to study the impact of climate change on water resources.

RRMs are based on grid-cells that represent reservoirs and they route the flow to the outlet using a linear function. RRMs have used Geographical Information Systems (GIS) to support the input data. According to Ngo-Duc *et al.*⁵⁾, in state-of-the-art global routing models, most of the approaches either assume a constant velocity or use simple formulas that use time-independent flow velocities parameterized as a function of the topographic gradient. In general, these approaches are sufficient to model discharges in monthly or longer time scales. However, to model in shorter temporal scales, more sophisticated approach is required.

TOPMODEL⁶⁾⁷⁾ is a hydrological model based on variable source area assumption. TOPMODEL framework has two components: storage component, which is represented by three reservoirs and the routing component, which is derived from a distance-area function and two velocities parameters. Its main parameter is the topography index estimated from a digital elevation model. This index represents the propensity of a cell or region becoming saturated.

The objective of this work is to apply a modified TOPMODEL approach in two large-scale basins, the Amazon and Yangtze basins, in order to verify whether TOPMODEL is feasible as a RRM and therefore to be coupled to GCM and closing the water cycle. TOPMODEL is also applied for backcasting the period from 1990 to 2008 using NCEP climate data. This modified approach is based on TOPMODEL-GRASS module and introduces a multi-velocities parametrization.

2. TOPMODEL BACKGROUND

The TOPMODEL is a rainfall-runoff model that uses the concept of hydrological similarity based on topography⁶⁾⁷⁾. This similarity is defined by the topographic index λ_i :

$$\lambda_j = \ln\left(\frac{a_j}{\tan\beta_j}\right) \tag{1}$$

where a_j [L] is the upslope contributing area per unit contour length for each cell *j* in the catchment and $\tan \beta_j$ [-] is the slope of this cell measured with respect to plan distance.

Once a number of classes with the same hydrological similarity is defined, the storage deficit S_i [L] for each class i is:

$$S_i = S + m \left(\lambda - \lambda_i \right) \tag{2}$$

where *S* [L]is the lumped or mean storage deficit for the entire catchment; λ is the mean topographic index (approximated by a weighed average over the areas with the same hydrological similarity); λ_i is the local topographic index and *m* [L] is a parameter associated with the rate of decline of the catchment recession curve.

For each time step the mean storage deficit is updated following the equation:

$$S_{t} = S_{t-1} + \left[Q_{b_{t-1}} - Q_{v_{t-1}} \right]$$
(3)

where S_t is the updated value of the storage deficit; S_{t-1} is the storage deficit in the previous time step; Q_{bt-1} [LT⁻¹] is the base flow in the previous time step and Q_{vt-1} [LT⁻¹] is the unsaturated zone recharge in the previous time step. This recharge is defined by:

$$Q_{\nu} = \sum_{i=\nu}^{n} \frac{S_{UZ}}{S_i T_D}$$
(4)

where S_{UZ} [L] is the unsaturated zone deficit; T_D is residence time in the unsaturated zone [TL⁻¹] and *n* is the number of classes with the same hydrological similarity.

The baseflow Q_b [LT⁻¹] is defined by:

$$Q_b = Q_s e^{\left(\frac{-S}{m}\right)} \tag{5}$$

where Q_s [LT⁻¹] is the discharge when the catchment is saturated and it is calculated by:

$$Q_S = T_0 e^{-\lambda} \tag{6}$$

where T_0 [L²T⁻¹] is the soil saturated transmissivity, which is constant for the entire catchment.

In the first time step the mean storage deficit is estimated by:

$$S_{t=.} = -m \ln\left(\frac{Q_{.}}{Q_{S}}\right) \tag{7}$$

where Q_0 [LT⁻¹] is the initial discharge at the first time step.

TOPMODEL uses the Dunne⁸⁾ overflow generation mechanism, *i.e.*, when the storage deficit equals to zero.

In TOPMODEL approach there is a reservoir intended to represent root and vegetation storage. This reservoir is called Root Zone reservoir. It can be depleted through the following equation:

$$E_a = E_p \left(1 - \frac{S_{RZ}}{S_{RMAX}} \right) \tag{8}$$

where S_{RZ} [L] is the root zone reservoir deficit; S_{RMAX} [L] is the maximum deficit in the root zone reservoir; E_p [LT⁻¹] is the potential evapotranspiration and E_a is the evapotranspiration rate. Its deficit at the first time step is set through a parameter called S_{RO} .

Flow routing is done through a time-area function. This function is derived from a distance-area function⁹⁾¹⁰⁾ using the following equation:

$$tc_k = \sum_{k=1}^{N} \frac{l_k}{V} \tag{9}$$

where tc_k [T] is the time of concentration of a determined distance-area function class k; V [LT⁻¹] is a velocity parameter, as k = 1, V is equal to the input parameter V_{CH} [LT⁻¹] and for k > 1, V assumes the value of the input parameter V_R [LT⁻¹] . V_{CH} represents the main channel velocity and V_R represents the average velocity of lower order rivers and hillslopes; l_k is the plan flow path length from a class area k to the basin outlet and N is the total number of classes which the area-distance function is composed.

The distance-area function represents the distances from points, cells or areas to the basin outlet. Usually, it is grouped in distances classes. It has great importance when investigating a basin hydrological behavior. **Fig. 1** shows an example of a distance-area function.

3. METHODOLOGY

(1) Studies area and data

For this study were selected two large-scale basins, the Amazon basin and Yangtze basin. The Amazon basin is a forest basin located in South America. Its area is roughly 7.05 million km². According to Beighley *et al.*¹¹⁾, the annual flow of the Amazon basin accounts for approximately one fifth of the all river discharges to the oceans. The Yangtze basin located in China has 1.72 million km² It is the third largest river worldwide in terms of discharge.

For the Amazon and Yangtze basins the global topographic data were extracted using ETOPO5 data (**Fig. 2**), from the National Geophysical Data Center (NGDC), National Environmental Satellite (NOAA). Basins boundaries (**Fig. 3**) were acquired from the Global Runoff Data Centre (GRDC) and stream networks and cumulative areas from the Global Drainage Basin Database (GDBD). The climate daily data were obtained from the National Centers for Environmental Prediction (NCEP).

Although, ETOPO5 grid has a 5-minute resolution, the resolution was defined to 0.5 degree in order to match the GRDC and NCEP data resolution.

Observed daily data from GRDC for the Amazon data encompass the period from 1990 to 1995 at Obidos station. For the Yangtze basin the period of data used corresponds to the 2004 year at Datong station.

The modified Penman method¹² was used to estimate evapotranspiration. Despite NCEP gives spatial distributed rainfall data, an areal average precipitation and evapotranspiration values for each time step were used for both basins.





Fig.2 ETOPO5 data. Elevations in meters.



Fig.3 GRDC watershed boundaries delineation.

(2) TOPMODEL approach modification

TOPMODEL uses a distance-area function to derive a time-area function for flow routing. At first, a distance-area function must be extracted from topographic data. For each grid cell the distance from this cell to the basin outlet is computed. The function or histogram is estimated grouping the distances in a certain number of class. In the end, the function is changed to its cumulative form.

In this work is assumed that there is a power law relationship between cumulative area and velocity. The meaning of the TOPMODEL velocity parameters is modified, instead of representing velocities, the parameters represent coefficients in the following equation:

$$v_K = V'_{CH} A_K^{V'_R} \tag{10}$$

where v_K is the velocity of the distance-area class *K*; A_K [L²] is the cumulative area of the class *K*; V'_R is a power law exponent [-]; V'_{CH} is a proportionality

constant [L-1T-1].

Rewriting Eq. (10), a time-area function can be derived from a distance-area function applying the following equation:

$$tc_{k} = \sum_{k=1}^{N} \frac{l_{k}}{V'_{CH} A_{K}^{V'_{R}}}$$
(11)

Eq. (11) tries to take into account the spatial distribution of velocities in a basin. However, as well as Eq. (9), Eq. (11) is temporally invariant, the velocity never changes over the simulation time.

(3) TOPMODEL-GRASS module modification

The GRASS GIS software¹³⁾ is distributed with a module called r.topmodel. This module is implemented in C language.

In this work, the module r.topmodel was modified as a mean to simulate discharges from NCEP climate data. The main changes are: (1) improved memory allocation, (2) generation of graphics by GNUPLOT¹⁴, (3) auto-calibration using the Monte Carlo procedure (4) mapping of saturated areas for each time interval and (5) reading precipitation and evapotranspiration raster files as input data. For auto-calibration is possible to choose between two objective functions, the Nash-Sutcliffe coefficient¹⁵⁾ and LogNash¹⁶⁾. As TOPMODEL is a semidistributed precipitation model, and evapotranspiration data were interpolated to produce an areal average value for each time step.

Fig. 4 shows the TOPMODEL-GRASS modified module flowchart.

(4) Simulation of stream discharges

A Monte Carlo procedure was used to find the best set of parameters from pre-defined ranges.

The parameters values were spread according to a uniform distribution of probability. Nash-Sutcliffe coefficient was chosen as an objective function to evaluate the stream-flow efficiency. A Nash-Sutcliffe value of 0.2 was selected to estimate the uncertainty bounds. The uncertainty bounds were delimited as a means to evaluate the model performance and they are useful to identify errors in model structure or errors in input data.

In this way, it was carried out 10,000 simulations for each basin and all simulations with Nash-Sutcliffe values above or equal to 0.2 were selected to estimate uncertainty bounds.

A normal probability distribution was assumed to delineate the lower and upper uncertainty bounds. These bounds, in this work, encompass 90 percent of all discharges for each time step. Therefore, the distance of 1.645 standard deviation from the mean was delimitated.



Fig.4 GTOPMODEL-GRASS module. Modifications and additions are shown in gray color.

(5) Backcasting from 1990 to 2008

In order to verify the model application for longterm simulations and the computational efficiency, the calibrated model was applied to simulated the period from 1990 to 2008. The best sets of parameters were selected to estimated the uncertainty bounds. As there is no observed data for this period, the first observed discharge value from the calibration period was used as initial condition.

3. RESULTS AND DISCUSSION

For Amazon basin 17 iterations from 10,000 iterations stayed above the threshold of 0.2. The best Nash-Sutcliffe coefficient for the calibration period was 0.39.

For Yangtze basin after 10,000 iterations, a number of 190 iterations stayed above the threshold. The best simulation obtained a 0.65 Nash-Sutcliffe coefficient.

 Table 1 summarized the best set of parameters for each basin.

Table 1 Best set of parameters.		
Parameter	Basin	
	Amazon	Yangtze
$\frac{\ln T_0}{(\ln(m^2 h^{-1}))}$	-1.29	-6.25
<i>m</i> (m)	0.106	0.007
S_{RMAX} (m)	0.0083	0.0018
$\frac{T_D}{(h m^{-1})}$	426.55	1202.11
V'_{CH} (m h ⁻¹)	67.05	4.87
$\frac{V'_{R}}{(\mathrm{m}\ \mathrm{h}^{-1})}$	0.13	0.31
Nash-Sutcliffe	0.39	0.65

Analyzing **Table 1**, it is possible to notice that parameter T_D is quite different between the two basins. The reason for this parameter being higher in Yangtze basin could be the presence of many dam reservoirs. Thus, as dam reservoirs effect is not modeled in TOPMODEL approach, the calibration approach tries to compensate this effect increasing the T_D parameter value and therefore, increasing the water delay time to reach the basin outlet.

Parameter m is also quite different. This difference could be associate to the soil type and therefore producing different recession curves.

Velocities in Yangtze basin are larger than Amazon river. This can be noticed through parameters V_{CH} and V_{R} . This is associate to steeper slopes in Yangtze basin and to water storage of reservoirs in this basin.

Figs. 5 and 6 show the hydrograph simulation for each basin.

Through **Figs**. **5** and **6** it is possible to notice that TOPMODEL simulated the discharges for both basin in a satisfactory way. However, for the Amazon basin the model does not represent rather well some peaks discharges and lower discharges at the end of the simulation period.

This could be associated to data errors and or model limitations.



Fig.5 Hydrography simulation for Amazon basin. Calibration period from 1990 to 1995, at Obidos station.



Fig.6 Hydrography simulation for Yangtze basin. Calibration period 2004, at Datong station.

For the Yangtze basin, TOPMODEL carried out a better simulation with a Nash-Sutcliffe coefficient of 0.65 than the Amazon basin (Table 1). It is noticed that discrepancies between observed discharge and simulated discharge are larger in higher discharges. It could be explained because in Yangtze basin there are many reservoirs and they are not modeled in TOPMODEL approach. Also, there is a significant difference between the dischargers during the simulation time from 50 to 150 days. This could be associated to the snow melt, which is a phenomena not modeled in TOPMODEL approach. Observing the uncertainty limits, it is observed that Yangtze basin has a wider interval than Amazon basin. It is due to many parameters sets relied above the Nash-Sutcliffe threshold.

Figs. 7 and **8** show the backcasting simulation for the period from 1990 to 2008.

An important comment regarding backcasting simulation is about parametrization. The parameters chosen in the calibration period were able to simulate the entire backcasting period. It means that the parameters assumed values with some physical sense. Otherwise, whether the parameters assume values to compensate model structure deficiency or data errors, the seasonality over the backcasting period would not be viewed.



Fig.7 Hydrography simulation for Amazon basin. Backcasting simulation from 1990 to 2008.



Fig.8 Hydrography simulation for Yangtze basin. Backcasting simulation from 1990 to 2008.

The Yangtze basin was more susceptible to produce such effect than Amazon basin, because its calibration period encompassed only one year.

The model run 6935 daily data, corresponding to the period from 1990 to 2008. For the entire period 17 iterations for Amazon basin and 190 iterations were carried out. For each time step a precipitation and evapotranspiration raster maps were read and interpreted. Taking into account this amount of data, the model performed computationally quite well. Both time simulations were done in less than 2 hours in a personal computer. It is considered a feasible computational efficiency.

4. SUMMARY AND CONCLUSIONS

In this study, two large-scale basins, Amazon and Yangtze basins, were selected, according to daily data availability, to apply a TOPMODEL modified approach.

The velocity parameters in TOPMODEL were modified with the objective to give a more realistic representation of the velocities in the basin rivers, instead of using only one velocity parameter, which is an usual approach in RRMs. The method presented here implemented a power law relationship between cumulative area and velocity.

The global topographic data, basins boundaries stream network were acquired and from Geophysical National Data Center, Global Runoff Data Center (GRDC) and Global Drainage Basin Database (GDBD), respectively. The climate daily data were obtained from the National Centers for Environmental Prediction (NCEP). Observed daily data from GRDC for the Amazon data encompass the period from 1990 to 1995. For the Yangtze basin the period of data used. As TOPMODEL is a semidistributed model. precipitation and evapotranspiration data were interpolated to represent an areal average value for each time step.

Monte Carlo simulations were used to find the best set of parameters according to the Nash-Sutcliffe coefficient.

Besides the model calibration, the model also was tested for backcasting the period from 1990 to 2008. The conclusions can be summarized as:

- 1. TOPMODEL simulated (calibration) period, the discharges in a satisfactory way for both basins
- 2. Long-term simulations were coherent and with good computational efficiency.

Further studies should be carried out in order to identify source of errors and/or improve TOPMODEL approach for large-scale runoff modeling. The principal improvement to be done will be to consider the spatial distribution of precipitation.

REFERENCES

- 1) Arora, V. K.: Streamflow simulations for continental-scale river basins in a global atmospheric general circulation model, *Advances in Water Resources*, Vol. 24, pp. 775-791, 2001.
- Campos, E., Lentini, C., Miller, J., Piola, A.: Interannual variability of the sea surface temperature in the South Brazil Bight, *Geophys. Res. Lett*, Vol.26, pp. 2061-2064, 1999.
- Ducharne A., Golaz, C., Leblois, E., Laval, K., Polcher, J., Ledoux, E., Masily, G.: Development of a high resolution runoff routing model, calibration and application to assess runoff from the LMD GCM, *Journal of Hydrology*, Vol. 280, pp. 207-228, 2003.
- IPCC, Intergovernmental Panel on Climate Change. Data Distribution Centre. High resolution observation climatologies. Available in <<u>http://www.ipccdata.org/obs/cru_ts2_1.htm</u>> Accessed in April 2009.
- Ngo-Duc, T., Oki, T., Kanae, S.: A variable streamflow velocity method for global river routing model: model description and preliminary results, *Hydrol. Earth Syst. Sci. Discuss.*, Vol. 4, pp. 4389-4414, 2007.
- Beven, K.J. and Kirkby, M.J.: A physically-based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.* Vol. 24, pp. 43 – 69, 1979.
- Beven, K.J., Lamb, R, Quinn, P., Romanowicz, R. and Freer, J.: Topmodel. In: *Computer Models of Watershed*, Singh VP (ed). Water Resources Publication, pp. 627 – 668. 1995.
- Dunne, T., Black, R.D.: Partial-area contributions to storm runoff in a small New England watershed. *Water Resources Research*, Vol. 6, pp. 1296–1311, 1970.
- Rinaldo, A., Vogel, G., Rigon, R. and Rodriguez-Iturbe, I.: Can one gauge the shape of a basin? *Water Resources Research*, Vol. 31, pp. 1119 – 1127, 1995.
- Rodriguez-Iturbe, I. and Rinaldo, A.: *Fractal river-basins*. *Chance and self-organization*. Cambridge University Press, New York. 547 p., 1996.
- Beighley, R.E., Eggert, K.G., Dunne, T., He, Y., Gummadi, V. and Verdin, K.L.: Simulating hydrologic and hydraulic processes throughout the Amazon River Basin. *Hydrological Process*, Vol. 23, pp. 1221-1235, 2009.
- 12) Doorenbos, J., Pruit, W.O.: Crop water requirements. Roma: FAO, 144 p., 1992.
- 13) GRASS Development Team. Geographic Resources Analysis Support System (GRASS) Software, Version 6.4.0. http://grass.osgeo.org. 2009.
- Williams, T. and Kelley, C.: GNUPLOT: An Interactive Plotting Program. Version 4.2. <u>http://www.gnuplot.inf</u>, 2008.
- Nash, J. E. and Sutcliffe J. V.: River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, Vol. 10 (3), pp. 282–290, 1970.
- 16) Wagener, T., Wheater, H.S., Gupta, H.V.: Rainfall-runoff modelling in gauged and ungauged catchments. Imperial College Press. 306 p., 2004.