Daily streamflow modelling and assessment based on the curve-number technique

Jin-Yong Choi, Bernard A. Engel and Ha Woo Chung

1 Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907-1146, USA.
2 Division of Biological Resources and Material Engineering, Seoul National University, Suwon, Kyonggi 441-744, South Korea

Abstract:
A cell-based long-term hydrological model (CELTHYM) that can be integrated with a geographical information system (GIS) was developed to predict continuous stream flow from small agricultural watersheds. The CELTHYM uses a cell-by-cell soil moisture balance approach. For surface runoff estimation, the curve number technique considering soil moisture on a daily basis was used, and release rate was used to estimate baseflow. Evapotranspiration was computed using the FAO modified Penman equation that considered land-use-based crop coefficients, soil moisture and the influence of topography on radiation. A rice paddy field water budget model was also adapted for the specific application of the model to East Asia. Model sensitivity analysis was conducted to obtain operational information about the model calibration parameters. The CELTHYM was calibrated and verified with measured runoff data from the WS#1 and WS#3 watersheds of the Seoul National University, Department of Agricultural Engineering, in Hwaseong County, Kyonggi Province, South Korea. The WS#1 watershed is comprised of about 35-4% rice paddy fields and 42.3% forest, whereas the WS#3 watershed is about 85-0% forest and 11-5% rice paddy fields. The CELTHYM was calibrated for the parameter release rate, K, and soil moisture storage coefficient, STC, and results were compared with the measured runoff data for 1986. The validation results for WS#1 considering all daily stream flow were poor with $R^2$, $E^2$ and RMSE having values of 0.40, −6.63 and 9.69 (mm), respectively, but validation results for days without rainfall were statistically significant ($R^2_{D}=0.66$). Results for WS#3 showed good agreement with observed data for all days, and $R^2$, $E^2$ and RMSE were 0.92, 0.91 and 2.23 (mm), respectively, suggesting potential for CELTHYM application to other watersheds. The direct runoff and water balance components for watershed WS#1 with significant areas of paddy fields did not perform well, suggesting that additional study of these components is needed. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS watershed modelling; GIS; soil moisture balance; grid-based modelling, paddy field water balance; model calibration; sensitivity analysis; model assessment

INTRODUCTION
Water resources development and watershed management require an understanding of hydrological variations owing to changes in watershed characteristics over long-term periods (Bhaduri et al., 2000), and spatial variability of watershed characteristics that affect hydrological phenomenon also must be evaluated in heterogeneous land-use watersheds. However, hydrological model operations that reflect long-term watershed changes often have limitations owing to difficulties obtaining measured hydrological data and data quantifying land use and soil characteristics. Therefore, simulation of stream flow on a daily basis using a long-term hydrological model that is simple to operate with readily available data is needed.

Continuous models, also called long-term hydrological models or continuous stream flow models, typically are focused on estimating water yield from a watershed. Owing to complications and difficulties related to data preparation and operation, however, these hydrological models largely have been used for daily, 10-day and
monthly time periods, and watersheds have been treated as a single hydrological unit or their characteristics are lumped (Rockwood et al., 1972; Sugawara et al., 1976; Sugawara et al., 1984), even though there have been significant efforts in some cases to apply distributed modelling approaches for stream-flow estimation (Arnold et al., 1998).

Fortunately, recent advances in computer environments, including hardware, software and networks, support the ability to solve the problems of data availability. Current computational speed, storage, software debugging tools and spatial analysis software make distributed hydrological simulation with spatial variation consideration feasible (Arnold et al., 1993). Geographical information systems (GIS) have had an important effect on hydrological modelling and development (Heaney et al., 2001), and GIS tools are now commonly used in hydrological modelling for data preparation (Nageshwar et al., 1992). The value of GIS in this role has been demonstrated for various hydrological models including ANSWERS (Rewerts, 1992), AGNPS (Heaney et al., 2001), SWAT (Arnold et al., 1993) and L-THIA (Bhaduri, 1998).

The CELTHYM (cell-based long-term hydrological model) was developed to provide a simple hydrological model for integration with GIS tools that can use readily available data. It uses the curve number (CN) method on the base of a GIS grid data structure, permits integration with GIS and uses readily available digital maps. This paper describes CELTHYM concepts and its assessment, including sensitivity analysis, calibration and verification for test watersheds.

BACKGROUND

Various watershed models have been developed and introduced in the hydrological literature (Singh, 1989; Singh, 1995). Nevertheless, few of these models have become common planning or decision-making tools, largely because of limitations in availability of measured data to satisfy model input.

Since the development of the Stanford watershed model (Crawford and Linsley, 1966), numerous operational, lumped or conceptual models have been developed. These include the tank model (Sugawara et al., 1976). The tank model is lumped and simulates direct and base flow from a watershed with a simplified tank with bottom and side holes for runoff connected in serial or parallel. The tank model has been used for numerous water resources development projects for reservoir operation planning in the regions of East Asia, Japan and South Korea (Sugawara et al., 1984). Nevertheless, the tank model has limitations, including difficulty in considering watershed spatial variation and a need for calibration (Kim and Park, 1986).

The SHE model (Abbott et al., 1986a,b), a distributed and numerical model, simulates surface and subsurface water movement in a basin with the finite difference solution of the partial differential equations describing the processes of overland and channel flow, unsaturated and saturated subsurface flow, interception, evapotranspiration (ET) and snowmelt. The spatial distribution of catchment parameters is achieved by representing the basin with an orthogonal grid network (Arnold et al., 1998). Arnold et al. (1998) indicate that although Jain et al. (1992) successfully applied the SHE model to an 820 km² catchment in central India, the SHE model is too complicated to use for simple applications, and the data requirements are substantial. Jain et al. (1992) also concluded that the strength of differential models such as SHE lies beyond the field of pure rainfall—runoff modelling, for which purpose traditional and simpler hydrological models often perform equally well.

The variable source area concept, based on topographic index concepts and the curve number methods, can be used as an adaptable runoff estimation technique. This approach simplifies spatial variation consideration by using detailed field data and adapting sophisticated hydrological computational algorithms without entire loss of consideration of distributed watershed characteristics. Beven and Kirkby (1979) suggested that a conceptual model that can simulate variable source areas could be used for long-term water yield estimation. Based on this approach, several conceptual models have incorporated soil moisture replenishment, depletion and redistribution for the dynamic variation in areas contributing to direct runoff. The TOPMODEL (Beven and Kirkby, 1979; Beven et al., 1984; Beven et al., 1995) is a representative conceptual continuous stream
flow hydrological model that uses a topographic index and variable source area, but the model does not consider the spatial variability of land use and soil characteristics that are not incorporated directly with the topographic index.

The Soil Conservation Service (SCS) CN method was developed in 1954 by the United States Department of Agriculture Soil Conservation Service, which is now the NRCS (USDA, 1986). The method has been criticized as an obsolete and simple methodology to simulate the sophisticated hydrological system (Ponce and Hawkins, 1995). However, Ponce and Hawkins (1995) indicated that the method is widely used in the USA and other countries, and mentioned that the perceived advantages of the method are (i) its simplicity (ii) its predictability (iii) its stability (iv) its reliance on only one parameter and (v) its responsiveness to major runoff-producing watershed properties including soil type, land use/treatment, surface condition and antecedent condition. They also indicated that the perceived disadvantages are (i) its marked sensitivity to curve number, (ii) the absence of clear guidance on how to vary the antecedent moisture condition factor; (iii) the method’s varying accuracy for different biomes, (iv) the absence of an explicit provision for spatial-scale effects and (v) the fixing of the initial abstraction ratio at 0-2, preempting a regionalization based on geological and climatic settings. In addition, the method does not consider the time distribution of rainfall within a day when it is used to calculate daily direct runoff (Rallison and Miller, 1982), because it was developed for conditions usually encountered in small watersheds for which only daily rainfall is ordinarily available. Further the method does not account for runoff from snowmelt or the effect of frozen ground. It applies only to direct surface runoff and does not consider additional runoff as a result of subsurface flow or high ground water levels. If the curve number is less than 40, Rallison and Miller (1982) suggest that an alternative method be used to determine runoff.

Since the 1990s, GIS tools are commonly used with hydrological and water quality model simulation. The curve number method is readily implemented within GIS and has been placed in hydrological models as a frequent runoff calculator. One of the models that adapts the CN method and GIS tools, SWAT (soil and water assessment tools), is an operational or conceptual model that operates on a daily time-step (Arnold et al., 1998). Although SWAT is a long-term model, it has been limited in application owing to limitations of data availability, including soil characteristics, weather, crop and gauged stream flow (Spruill et al., 2000). To overcome the shortage of available data and difficulty of model selection, the better assessment science integrating point and non-point sources model (BASINS) has been developed to meet the needs of pollution control agencies. It integrates a GIS, national watershed and meteorological data, and state-of-the-art environmental assessment and modelling tools, including SWAT and HSFP (hydrological simulation program—Fortran, Bicknell et al., 1997) into one convenient package.

Hydrological models have trade-offs between efficiency and complexity in terms of quantity/quality of data prepared and model components considered. However, as a key component of hydrological models, the CN method for direct runoff estimation has benefits of simplicity and ease of use in GIS tools with supporting data available from GIS data clearing houses.

Even though the curve number method has the limitations pointed out by Ponce and Hawkins (1995), the method has been widely used in numerous models, including AGNPS for single events and by the long-term hydrological impact assessment model (L-THIA; Harbor, 1994) for tens of years, and stream-flow estimation from small watersheds with no stream flow records (Bhaduri, 1998; Bhaduri et al., 2000).

Consequently, to understand the hydrological phenomena of a watershed, a model is needed that can simulate direct runoff, base flow, soil moisture and evapotranspiration. There is no doubt that the hydrological models have advantages and disadvantages for use in hydrological impact evaluation, but the complexity and demands of detailed field data are still a barrier for use of the models.

Hence, the goal of the CELTHYM development focused on creating a model that is comprised of simple concepts using the CN method and soil water balance, and operated with readily available digital spatial data sets. The CELTHYM also includes a rice paddy field water balance to support application to watersheds in East Asia.
MODEL COMPONENTS

Main concept of the cell-based long-term hydrological model

The CELTHYM (cell-based long-term hydrological model) was developed to simulate stream flow from small rural watersheds. The CELTHYM is a simplified, operational and conceptual model that uses grid data and a daily time-step. It also has pre- and post-processors for interoperation with GIS. The conceptual schematic diagram of the model is shown in Figure 1.

The general watershed water balance equation used is shown in Equation (1)

\[ \Delta S = R - Q_{dr} - Q_{bf} - WET \quad (1) \]

where \( \Delta S \) (mm) is watershed storage change, \( R \) (mm) is precipitation, \( Q_{dr} \) (mm) is direct runoff, \( Q_{bf} \) (mm) is base flow and \( WET \) (mm) is watershed evapotranspiration.

A watershed is described in the model as a set of cells or grids. The watershed is comprised of several subwatersheds that are connected to the stream network grid, and each subwatershed is composed of grids. The runoff is estimated by the sum of direct runoff and base flow. The direct runoff is calculated by the curve number method, and the base flow is estimated by the release from groundwater. The difference between precipitation and direct runoff is treated as infiltration. By comparing infiltration and soil moisture depletion, soil moisture and deep percolation is estimated.

The base flow from groundwater is conceptualized as tank release for each subwatershed. The base flow release rate is formulated with depth and release rate. The depth of a tank is varied by summation of deep percolation from grids in a subwatershed, and the release rate depends on the depth of a tank. The total runoff of a subwatershed is calculated by summation of the direct runoff at each grid on days with rainfall and addition of the base flow from the subwatershed without routing of surface runoff, channel and subsurface flow. The ET of a grid in a watershed is estimated by the FAO modified Penman equation (Doorenbos and Pruitt, 1977) and considers soil moisture and a crop coefficient.

Modelling

Direct runoff. The direct runoff of grids was estimated by the CN method of the USDA–SCS (Soil Conservation Service). The CN method has been used within GIS environments in other research and has been verified to work well for direct runoff simulation.

Figure 1. Conceptual diagram of the CELTHYM
The CN is carefully estimated in computing direct runoff, because the CN directly affects the calculation of the runoff quantity. The curve number is selected by considering various factors of a watershed. Land use and treatment, hydrological conditions of plant coverage, hydrological soil group and antecedent moisture condition (AMC) affect the CN selection procedure. The CN table from the USDA National Engineering Handbook (NEH) (USDA, 1971) contains CN values for three AMC conditions. In the NEH, three levels of AMC are used and defined for the selection of CN. The AMC I is the lower limit of moisture representing a dry condition, AMC II is the average condition of moisture, and AMC III is the upper limit of moisture representing a wet condition.

The CN is chosen from the CN table for a specific AMC based on land use and hydrological soil group. In this study, an equation was proposed to estimate CN continuously to allow representation of varying soil moisture conditions. Using the definition of the AMC, AMC II was assumed at 50% of maximum available soil moisture (0.5 × ASM max), AMC I was considered the soil moisture wilting point (WP), and AMC III was soil moisture field capacity (FC), because the soil moisture varies from WP to FC for natural situations after drainage of excess soil water by deep percolation processes. The ASM max value can be estimated as the difference between WP and FC using Equation (13). Estimation of FC values can be difficult owing to conditions that are site specific in terms of soil characteristics, combinations of different soil profiles and even presence of hardpans (Jury et al., 1991). However, reasonable FC values for CELTHYM operation can be obtained from tables, given soil texture and soil characteristics (Beasley et al., 1980).

Although Arnold et al. (1995) suggested that the CN can be varied non-linearly using soil characteristics and moisture condition, such an approach would be difficult to use with readily available soil data owing to the complexity of data items. Therefore, the CN was estimated to vary linearly from the CN I at AMC I to the CN II at AMC II, and the CN III at AMC III with equations shown below. When the available soil moisture is less than 50% of ASM max then

\[
a = \frac{(SM - WP)}{0.5 \times ASM_{\text{max}}} \quad (2)
\]

\[
CN_{\text{adj}} = a \times CN_{\text{II}} + (1 - a) \times CN_{\text{I}} \quad (3)
\]

When the available soil moisture is greater than or equal to 50% of ASM max then

\[
a = \frac{(ASM_{\text{max}} - (SM - WP))}{0.5 \times ASM_{\text{max}}} \quad (4)
\]

\[
CN_{\text{adj}} = a \times CN_{\text{II}} + (1 - a) \times CN_{\text{III}} \quad (5)
\]

where \(a\) is the ratio between soil moisture of above/below 50% of ASM max and 50% of ASM max, \(CN_{\text{adj}}\) is the adjusted CN for the current soil moisture condition and \(SM\) is available soil moisture, which is estimated daily using soil moisture routing.

After CN estimation for each watershed grid, the potential retention parameter, \(S\), is calculated for each grid by Equation (6)

\[
S = \frac{25400}{CN_{\text{adj}} - 254} \quad (6)
\]

The direct runoff of a grid, \(qdr\) (mm), is computed with precipitation, \(P\) (mm) and \(S\)

\[
qdr = \begin{cases} 
(P - 0.2S)^2 & P \geq 0.2S \\
(P + 0.8S) & P < 0.2S 
\end{cases} 
\quad (7)
\]

The subwatershed direct runoff, \(Q_{\text{sub}}\) (mm), is computed as the mean depth of direct runoff from the grids

\[
Q_{\text{sub}} = \frac{\sum_{i=1}^{N} qdr_i}{N} \quad (8)
\]

where \(N\) is number of grids in a subwatershed.
The total direct runoff of the main watershed, $Q_{dr}$ (mm), can be obtained by the area weighted average of the subwatershed direct runoff using equation (9).

$$ Q_{dr} = \sum_{j=1}^{M} (Q_{sub_j} \times AF_j) $$

where $M$ is number of subwatersheds and $AF_j$ is the area ratio of the $j$th subwatershed to the main watershed.

**Soil moisture.** For continuous hydrological simulation, soil moisture estimation is needed for daily computation of stream flow. Daily soil moisture is used for the estimation of infiltration rate and direct runoff calculation. Furthermore, the soil moisture affects evapotranspiration and deep percolation. The soil water balance equation is shown in Equation (10)

$$ DSM = SM_t - SM_{t-1} = (RAIN + UP_t + HI_t) - (qd_{t} + ET_t + DP_t + HO_t) $$

where $DSM$ is soil moisture change (mm), $RAIN$ is precipitation (mm), $UP$ is capillary rise of water (mm), $HI$ is horizontal inflow (mm), $qd$ is direct runoff (mm), $ET$ is evapotranspiration (mm), $DP$ is deep percolation from effective soil depth (mm) and $HO$ is horizontal outflow (mm).

The CELTHYM uses the soil water balance equation from Equation (10) in a simplified form by neglecting horizontal flow rate and capillary rise. Because these terms are not as important as other terms for long-term simulation, their inclusion would make the model complex beyond the model precision, and obtaining measured field data for these phenomena is difficult. Thus, Equation (10) can be rewritten to form Equation (11), and a descriptive figure of the equation is presented in Figure 2.

$$ DSM = SM_t - SM_{t-1} + SR_t - ET_t - DP_t $$

$$ SR_t = RAIN - qd_{t} $$

Figure 2. Schematic diagram of soil moisture balance in a grid.
In Equation (11), \( SR \) (soil water retention) was adopted in this study to calculate the total infiltrated water from precipitation, and the direct runoff, \( qdr \), was estimated by the CN method. Although in reality \( SR \) includes other initial abstractions including interception, these were neglected to simplify the model and to reduce the input data.

The maximum soil moisture depletion of effective soil depth, \( DF_{\text{max}} \), is equal to the maximum available soil moisture (\( ASM_{\text{max}} \)). The \( DF_{\text{max}} \) value is also the same as the maximum quantity of retained water and can be calculated by Equation (13). Although \( ASM_{\text{max}} \) and \( DF_{\text{max}} \) have the same value, their definitions vary and therefore both terms are included in the CELTHYM

\[
DF_{\text{max}} = ASM_{\text{max}} = FC - WP
\]  

where \( FC \) is the field capacity (mm) and \( WP \) is the wilting point (mm).

The \( DF_{\text{max}} \) value is determined by the soil texture and affects soil moisture and deep percolation. In fact, \( DF_{\text{max}} \) is the most important factor and affects all other hydrological components. Most hydrological models have recommended parameters to calibrate the models. In this case, the CELTHYM needs a parameter that can adjust both surface runoff and deep percolation. Therefore, the \( STC \) (soil moisture storage coefficient) was adopted and used for \( DF_{\text{max}} \) value adjustment because the soil moisture storage can be used to control both surface runoff and deep percolation. The \( DF_{\text{max adj}} \) value is the adjusted value of \( DF_{\text{max}} \) by the \( STC \), and the relationship of \( STC \) and \( DF_{\text{max}} \) is presented in Equation (14)

\[
DF_{\text{max adj}} = STC \times DF_{\text{max}} = STC \times (FC - WP)
\]  

The soil moisture deficit (\( DF \), mm) is changed daily by the soil moisture change and estimated by Equation (15).

\[
DF_i = FC - (SM_i - WP)
\]  

Deep percolation (\( DP \)) is estimated by the relationship of \( SR \) and \( DF \). Depending on the situation, \( DP \) is estimated by Equation (16) or (17), If \( SR_i > DF_i \)

\[
DP_i = SR_i - DF_i
\]  

If \( SR_i \leq DF_i \)

\[
DP_i = 0
\]

**Paddy field water budget.** A rice paddy field in East Asia is normally sited near a stream and cultivated in ponded conditions from immediately before rice transplanting to harvest. The paddy field water balance is significantly different to that of a field in upland or non-ponded conditions. For watersheds that contain paddy fields, the model water balance equation for paddy fields has to be adapted to compute direct runoff, especially in the Asia region, because rice cultivation is a common land use for agriculture in that region. The paddy field water budget equation is

\[
DPD_i = PD_i - PD_{i-1} = RAIN_i + IRRI_i - ET_i - qdr_i - DP_i
\]  

where \( DPD \) is change of pond depth (mm), \( PD \) (mm) is pond depth of the paddy field, \( RAIN \) is precipitation (mm), \( IRRI \) is irrigated water (mm), \( ET \) is evapotranspiration (mm), \( qdr \) is direct runoff (mm) and \( DP \) (mm) is deep percolation. In fact, Equation (18) is a general form of the paddy field water budget equation because direct runoff of a paddy field generally occurs only when the capacity between the pond depth and the outlet height is filled by rainfall or irrigation. The maximum value of \( DPD_i \) is the same as the storage capacity of a paddy field. Therefore, equation (18) can be rewritten as Equation (19)

\[
qdr_i = RAIN_i + IRRI_i - ET_i - DP_i - DPD_{\text{max}}
\]
where $DP_{t_{max}}$ is the maximum value of the storage capacity of a paddy field on day $t$, and $qdr_t$ occurs only when the right-hand side of the equation is positive.

**Baseflow of a watershed.** The baseflow of a watershed is the groundwater release from a catchment to a stream. In this study, the base flow of a watershed was calculated from each subwatershed of a main watershed. The baseflow release from a subwatershed is conceptualized as a tank and release. The baseflow detention depth ($Z$, mm), the same meaning as the depth of a tank, is calculated by adding the average of deep percolation of each grid in a subwatershed to the depth for the prior day. The average deep percolation of a subwatershed is calculated by Equation (20), and the baseflow detention depth on day $t$ is calculated by Equation (21)

$$\text{DP}_{\text{avg}, t} = \frac{\sum_{i=1}^{N} DP_{si}}{N}$$

$$dZ = Z_i - Z_{i-1} = DP_{\text{avg}, t}$$

In Equations (20) and (21), $DP_{\text{avg}, t}$ is average deep percolation of a subwatershed. To calculate the baseflow from a subwatershed, the calibration parameter, $K$, was introduced and the baseflow was estimated by multiplying $K$ with $Z$. The equation for baseflow calculation for a subwatershed is shown in Equation (22)

$$qbf_t = K \times Z_t$$

where $qbf_t$ (mm) is baseflow of a subwatershed, $K$ is baseflow release rate coefficient and $Z$ (mm) is baseflow detention depth of a subwatershed. The value of $K$ affects the quantity of base flow directly. The total baseflow at the outlet of a watershed was estimated by area-weighted averaging of all baseflow of subwatersheds. The baseflow of a watershed is summarized by Equation (23)

$$Qbf = \sum_{j=1}^{M} (qbf_j \times AF_j)$$

where $Qbf$ (mm) is mean baseflow in a watershed expressed as a depth unit, $M$ is number of subwatersheds and $AF$ is the fraction of the jth subwatershed area to watershed area.

**Computation of evapotranspiration.** The actual evapotranspiration ($ETa_t$) of a grid can be obtained by multiplying the cell crop coefficient ($Kc$) and soil moisture coefficient ($Ks$) with potential evapotranspiration ($PET$). In this study, the $ETa$ was estimated by consideration of topographical characteristics, land use and soil moisture condition using GIS data. The $PET$ was calculated by the modified FAO Penman equation suggested by Doorenbos and Pruitt (1977). Equation (24) is the modified Penman equation

$$PET = C[W \times Rn + (1 - W) \times f(u) \times (ea - ed)]$$

where $PET$ is potential evapotranspiration of the reference crop (mm/day), $W$ is a weighting factor, $Rn$ is net radiation (mm/day), $f(u)$ is a wind function, $(ea - ed)$ is the difference of saturated vapour pressure and actual vapour pressure at average daily temperature, and $C$ is a correction factor for day and night-time weather condition differences. The $ETa$ of a watershed grid is calculated by Equation (25) considering the crop coefficient and soil moisture condition

$$ETa = Kc \times Ks \times PET$$
where $ETa$ is actual evapotranspiration, $Kc$ is a crop coefficient based on land use and time that changes for the land use and growing stage of a crop, and $Ks$ is the soil moisture coefficient. The watershed $ETa$ (mm) is estimated by averaging grid $ET$.

The soil moisture coefficient, $Ks$, can be calculated by Equation (26) proposed by Pruitt et al. (1983)

$$Ks = \frac{\ln(Aw + 1)}{\ln(101)}$$

$$Aw = \frac{SW - WP}{FC - WP} \times 100$$

where $Aw$ is the percentage of available soil moisture.

SITE SELECTION AND DATA PREPARATION

Description of application site

The CELTHYM sensitivity analysis, calibration and verification were conducted on the test watersheds of the Department of Agricultural Engineering, Seoul National University that are located in Hwaseong County, Kyonggi Province, South Korea as shown in Figure 3. Development in the watersheds has been restricted by the Greenbelt Act, and thus the watersheds have been maintained in agricultural conditions. The main watershed contains several small watersheds, and the daily runoff data from the WS#1 and WS#3 watersheds within this larger watershed were used in this study. The meteorological data from the national weather station about 8 km away were used. Watershed characteristics and the data used are described in Tables I and II. One big difference between the two watersheds is land use. The WS#1 watershed is about 35-4% rice paddy fields and 42-3% forest, whereas the WS#3 watershed is about 85-0% forest and 11-5% rice paddy fields as described in Table II. The DEM and pre-processed input data of the CELTHYM for watershed WS#3 is shown in Figure 4.

Digital map construction

The CELTHYM application requires grid maps of elevation, soil characteristics and land use. For the data construction, the 1:25,000 scale national topographical map and a detailed soil map were used as spatial data sources. The selected maps were digitized into vector format and converted to grid format. The data are described further in Table III.

SENSITIVITY ANALYSIS

Sensitivity analysis was conducted to develop a comprehensive understanding of runoff response for changes of calibration parameter values. The results of the sensitivity analysis represent quantitative indices and hydrological responses for the variation of calibration parameters.

Furthermore, sensitivity analysis can be used to check the range of parameter values and assist in selection of parameters for model calibration, operations and improvement of model capability. In this case, sensitivity analyses for the soil moisture storage coefficient (STC) in Equation (14) and the baseflow release rate ($K$) in Equation (22), both CELTHYM calibration parameters, were analysed for three rainfall conditions: daily runoff of all days, days with rainfall and days without rainfall. Sensitivity analyses on the effect of the calibration parameters on the direct runoff and baseflow were conducted as well.

Total daily runoff

The sensitivity of daily runoff for $K$ and STC is shown in Figure 5. The total daily runoff variation for changes in $K$ is not very sensitive as shown in Figure 5. However, increasing STC causes a proportional
Figure 3. Location map of study watersheds. The watersheds are located in the middle of the Korean (peninsula).

Table I. Summary of runoff and weather data used for model application

<table>
<thead>
<tr>
<th>Watershed (Hwaseong County)</th>
<th>Weather data measuring station</th>
<th>Runoff data period used</th>
<th>Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banwol WS#1</td>
<td>Suwon</td>
<td>1 April to 30 September 1986</td>
<td>13 March to 30 September 1987</td>
<td></td>
</tr>
<tr>
<td>WS#3</td>
<td>Suwon</td>
<td>1 April to 30 September 1986</td>
<td>13 March to 30 September 1988</td>
<td></td>
</tr>
</tbody>
</table>

decrease in daily runoff and indicated that STC is the more sensible parameter to adjust for daily runoff estimation.

Daily runoff of days with rainfall

The sensitivity for the daily runoff of days with rainfall for $K$ and STC is shown in Figure 6. Daily runoff for days with rainfall increased proportionally for $K$ and decreased for STC. The phenomena above can be
Table II. Study watershed characteristics and land use

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Watershed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WS#1</td>
</tr>
<tr>
<td>Area (ha)</td>
<td>274.1</td>
</tr>
<tr>
<td>Stream length (km)</td>
<td>3.83</td>
</tr>
<tr>
<td>Shape factor</td>
<td>0.19</td>
</tr>
<tr>
<td>Relief ratio (m/m)</td>
<td>0.021</td>
</tr>
<tr>
<td>Channel slope (m/km)</td>
<td>13.77</td>
</tr>
<tr>
<td>Channel frequency</td>
<td>0.36</td>
</tr>
<tr>
<td>Land use (%)</td>
<td></td>
</tr>
<tr>
<td>Paddy field</td>
<td>35.4</td>
</tr>
<tr>
<td>Upland field</td>
<td>14.5</td>
</tr>
<tr>
<td>Forest</td>
<td>42.3</td>
</tr>
<tr>
<td>Urban</td>
<td>7.8</td>
</tr>
</tbody>
</table>

explained considering the physical meaning of $K$ and STC. Increased $K$ increases baseflow by the means of the release rate rising on days with rainfall. On the other hand, larger values of STC increase soil moisture capacity and infiltration rate that in turn reduce direct runoff on days with rainfall.

Daily runoff of days without rainfall

The sensitivity analysis results of the daily runoff on days without rainfall for $K$ and STC are shown in Figure 7. The total daily runoff of days without rainfall for $K$ and STC showed slight proportional decreases, but $K$ was more sensitive than STC for the daily runoff of days with no rainfall, because $K$ controlled the release rate of base flow that is the main contribution to the amount of stream flow of days without rainfall.

Increasing STC reduced runoff and affected the direct runoff and baseflow directly. This indicates that the model properly simulated these phenomena of a watershed. For example, watersheds with high infiltration rates and high moisture capacity soils produce less runoff than those with lower infiltration rates and lower soil moisture capacities. Even though the calibration parameter $K$ affected baseflow rate, it can be used only to control flow rate for baseflow, not the total runoff of a watershed. Therefore, STC was used as the control parameter for the total runoff, direct runoff and baseflow of the model.

MODEL CALIBRATION

Initial conditions

When a model is run, initial conditions are needed and usually are supplied from field data. In general, measured initial condition field data, such as soil moisture, may not be available for the start time of simulations because such data would require measurement long before their use for modelling purposes. However, in a long-term hydrological model simulation, the hydrological components have characteristics that can be automatically initialized or converged as a result of large rainfall events. The CELTHYM needs initial conditions for soil moisture and baseflow release depth. In this case, the initial conditions were selected by comparing estimated runoff results with historical runoff data. The initial soil moisture condition used was 50% of total available soil moisture. This value was selected based on the soil moisture seasonal average for the spring, which seemed reasonable to use in this instance. For the initial condition of release depth of baseflow, the median value of annual flow rate was used.
Figure 4. Digital elevation model (DEM) and pre-processed CELTHYM input data for watershed WS#3; (a) WS#3 watershed DEM; (b) spatial variation of CN for WS#3 watershed; (c) spatial distribution $ASM_{\text{max}}$ based on soil characteristics; (d) WS#3 subwatersheds

<table>
<thead>
<tr>
<th>Map</th>
<th>Data source</th>
<th>Scale</th>
<th>Vector map</th>
<th>Grid data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topographic map</td>
<td>National Geographic Institute (Republic of Korea)</td>
<td>1:25 000</td>
<td>Contour Land cover/use</td>
<td>DEM Land use</td>
</tr>
<tr>
<td>Soil map</td>
<td>National Plant Environmental Institute (Republic of Korea)</td>
<td>1:25 000</td>
<td>Soil characteristics</td>
<td>Soil texture Drainage condition Effective soil depth</td>
</tr>
</tbody>
</table>

**Parameter calibration**

The model calibration was conducted manually by referencing of selected statistical indices, because CELTHYM has only two calibration parameters. Initial conditions of soil moisture and release depth were used, and the release rate ($K$) and soil moisture storage coefficient (STC) were calibrated.
Criteria for comparing model output with historical data and analysing statistical similarity were needed. Three statistical variables, $R^2$, $E^2$ and RMSE, were selected for the criteria. The determination coefficient ($R^2$) was selected for analysis of statistical relationship and significance. To check precision and model correctness, model efficiency ($E^2$) proposed by Nash and Sutcliffe (1970) was used. The RMSE (root mean square of errors) were used for the quantitative analysis of residuals.
The daily runoff data of WS#1 and WS#3 in 1986 were used for calibration. In this study, the calibration was accomplished by changing release rate, \( K \), from 0.06 to 0.9, and the soil moisture storage coefficient, STC, from 0.6 to 4.0. The optimal values of \( K \) and STC were selected by checking \( R^2 \), \( E^2 \) and RMSE. The calibrated value selection procedure was to first set the value of STC at 1.0 and change the value of \( K \), checking \( R^2 \), \( E^2 \) and RMSE and select the value of \( K \) when \( R^2 \) and \( E^2 \) are the largest and RMSE is the smallest. Once the value of K is selected, the value of STC was changed to obtain the largest \( R^2 \) and \( E^2 \) and the smallest RMSE, and the procedures were repeated until the statistical indices were stable.

In the simulation of long-term runoff processes, the runoff discharge for days with no rainfall is important for the purposes of water resources planning and design. Hence, the value of \( R^2 \), \( E^2 \) and RMSE were checked for the no rainfall days, and more emphasis was placed on the values of \( R^2 \), \( E^2 \) and RMSE for those days. The calibrated value of baseflow release rate, \( K \), was 0.5 for both WS#1 and WS#3 following the above procedure, and the value of soil moisture storage coefficient, STC, was 0.6 for watershed WS#1 and 1.8 for watershed WS#3. The descriptions of calibrated results are presented in Table IV and in Figures 8 and 9.

Table IV. Calibrated CELTHYM results for the watersheds in 1986

<table>
<thead>
<tr>
<th>Watershed</th>
<th>Total runoff</th>
<th>Runoff of no rainfall days</th>
<th>Runoff of days with rainfall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 )</td>
<td>( E^2 )</td>
<td>( R^2 )</td>
</tr>
<tr>
<td>WS#1</td>
<td>0.58</td>
<td>−0.17</td>
<td>3.28</td>
</tr>
<tr>
<td>WS#3</td>
<td>0.71</td>
<td>0.71</td>
<td>2.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Watershed</th>
<th>Rainfall (mm)</th>
<th>Total runoff</th>
<th>Runoff (mm)</th>
<th>ET (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rainfall days</td>
<td>Observed</td>
<td>Simulated</td>
<td>Observed</td>
</tr>
<tr>
<td>WS#1</td>
<td>720.60</td>
<td>275.20</td>
<td>326.67</td>
<td>83.73</td>
</tr>
<tr>
<td>WS#3</td>
<td>720.60</td>
<td>382.14</td>
<td>319.70</td>
<td>136.17</td>
</tr>
</tbody>
</table>

Figure 8. Hydrograph of model calibration results for watershed WS#1 (1986, \( R^2 = 0.58 \), \( E^2 = −0.17 \), RMSE (mm) = 3.28)
MODEL VERIFICATION

Model verification was conducted as a consistency check of the model simulation results without any other adjustment of calibrated parameters. The daily runoff data of WS#1 in 1987 and WS#3 in 1988 were used for verification, and $R^2$, $E^2$, RMSE and verification results were compared with those for calibration. The verification results are presented in Table V and Figures 10 and 11.

In the verification results of WS#1, $R^2$ is slightly smaller than values for calibration, and $E^2$ has a negative value. This indicates that the verification results were statistically significant but quantitatively overestimated runoff. The simulation results of days without rainfall, however, showed good agreement with observed values of runoff as described in Table V. The cause of the negative value of $E^2$ for watershed WS#1 is due to the simulation results of direct runoff on days with rainfall that were quite different from the observed values, and affect the value of $E^2$ as shown in Figure 10. Watershed WS#1 has a large portion of paddy fields,

Table V. The CELTHYM verification results for the watersheds for 1987 and 1988

<table>
<thead>
<tr>
<th>Watershed</th>
<th>Total runoff</th>
<th>Runoff of no rainfall days</th>
<th>Runoff of days with rainfall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$ $E^2$ RMSE</td>
<td>$R^2$ $E^2$ RMSE</td>
<td>$R^2$ $E^2$ RMSE</td>
</tr>
<tr>
<td>WS#1(1987)</td>
<td>0.40 -6.63 9.70</td>
<td>0.66 -0.60 1.48</td>
<td>0.30 -8.45 16.53</td>
</tr>
<tr>
<td>WS#3(1988)</td>
<td>0.92 0.91 2.23</td>
<td>0.96 0.78 1.40</td>
<td>0.93 0.92 3.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Watershed</th>
<th>Rainfall (mm)</th>
<th>Total runoff</th>
<th>Runoff of no rainfall days</th>
<th>Runoff of days with rainfall</th>
<th>ET (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Observed</td>
<td>Simulated</td>
<td>Observed</td>
<td>Simulated</td>
</tr>
<tr>
<td>WS#1(1987)</td>
<td>1205-90</td>
<td>320-17</td>
<td>817-97</td>
<td>101-34</td>
<td>181-80</td>
</tr>
</tbody>
</table>
about three times the area of paddy fields in WS#3. Watershed WS#1 has 35.4% of land use in paddy fields, whereas WS#3 has just 11.5% of land use in paddy fields. Furthermore, the culture and water management for paddy fields is very complex to maintain optimal pond depth, and the drainage outlets are operated randomly by farmers. The paddy fields may act like ponds during some storms and decrease runoff, and in other instances farmers may release water from paddy fields during rainfall events to maintain optimal pond depth. Hence, more detailed consideration of paddy fields by introducing parameters that can reflect water

Figure 10. Observed and simulated runoff of model verification results for watershed WS#1 (1987, $R^2 = 0.40$, $E^2 = -0.63$, RMSE (mm) = 9.70)

Figure 11. Observed and simulated runoff of model verification results for watershed WS#3 (1988, $R^2 = 0.92$, $E^2 = 0.91$, RMSE (mm) = 2.23)
management on the paddy fields is needed for model application for watersheds that have large areas of paddy fields.

The precipitation in 1987 on watershed WS#1 was about 1200 mm, and measured runoff discharge is about 320 mm, with the remaining moisture used for evapotranspiration, soil moisture increases and recharge of groundwater. The estimated total discharge of that year is 617 mm and the difference between estimated and observed discharge is large; however, the difference mainly occurred on the days of storms. Management of paddy fields probably resulted in this difference. The hydrographs for the recession phase and no rainfall days are similar for estimated and observed runoff. The CELTHYM properly simulated the baseflow of runoff and therefore can be used for water use planning and design for these parameters.

The verification results of WS#3 were improved compared with calibrated results. The values of $R^2$ and $E^2$ are 0.92 and 0.91, respectively, and the runoff discharge is nearly the same—380 mm of estimated runoff and 339 mm of observed runoff. Based on these results, the CELTHYM probably is more suitable for use on typical small agricultural watersheds such as WS#3.

Because the runoff of days without rainfall is more important than that on days with rainfall as mentioned before, the results of days without rainfall were given more significance than those of days with rainfall. The values of $R^2$ and $E^2$ on no rainfall days are larger than those for results for all days, and the RMSE is smaller on days of no rainfall than for all days. The results show that the CELTHYM performed well for baseflow simulation.

The estimated values and observed values are presented in scatter diagrams in Figures 12 and 13 for each watershed. For watershed WS#1, the scattering of verification results was similar to the calibrated results. In watershed WS#3, the scattering of verification results differed somewhat from the calibrated results, but most points are dispersed around the 1:1 line. Based on the simulated results, the model should be useful.

**SUMMARY AND CONCLUSIONS**

The CELTHYM (cell-based long-term hydrological model) was developed for estimation of runoff, evapotranspiration and soil moisture of small agricultural watersheds on a daily time-step using weather data.
and spatial grid data from GIS. The CELTHYM computes daily runoff from direct runoff and baseflow for each grid of a watershed. The direct runoff of each upland grid is computed by the SCS curve number method, runoff for each rice paddy field grid is estimated by water balance, and the baseflow is computed by release depth and release rate of subwatersheds that belong to stream grids. The potential evapotranspiration of each watershed grid is computed by the FAO modified Penman equation, and actual evapotranspiration of each grid of watersheds is calculated with crop coefficients and soil moisture coefficients.

Sensitivity analysis was conducted to obtain information about the model responses to calibration parameters. Based on the results of the sensitivity analysis, the soil moisture storage coefficient (STC) seems more important than release rate (K), because increasing STC reduced runoff and affected the direct runoff and baseflow directly. Release rate affects baseflow, but changes total runoff of a watershed only slightly. Therefore, the parameter STC is more important than K, because increasing STC reduced the runoff of all days including days without rainfall and days with rainfall, and also affected direct runoff and base flow.

An assessment study of the CELTHYM was conducted on watersheds WS#1 and WS#3 of the Seoul National University, Department of Agricultural Engineering, in Hwaseong County, Kyunggi Province, South Korea. The model, supported with GIS data, was calibrated and verified for these watersheds. The CELTHYM was calibrated for release rate and soil moisture storage coefficient, and results were compared with the measured runoff data for 1986. The verification results for WS#1 were not so good, and $R^2$, $E^2$ and RMSE were 0.40, −6.63 and 9.70 (mm), respectively, but results for days without rainfall were statistically significant with an $R^2$ of 0.66. The verification results of WS#3 were much better, and $R^2$, $E^2$ and RMSE were 0.92, 0.91 and 2.23 (mm), respectively.

The runoff results for watershed WS#1, which has a significant portion of its land use in paddy fields (35-4% of the watershed area), were overestimated, indicating that further study of paddy field hydrology is needed. Therefore, the CELTHYM seems best suited for typical small agricultural watersheds that are dominated by forest and upland land uses such as those in WS#3. The CELTHYM has potential for use in water resources planning and design with integrated GIS and should be useful for the understanding of long-term hydrological phenomena.
ACKNOWLEDGEMENT

This work was supported by a post-doctoral fellowships programme from the Korea Science and Engineering Foundation (KOSEF).

REFERENCES


Arnold JG, Williams JR, Srinivasan R, King KW. 1995. SWAT: Soil Water Assessment Tool. Texas A&M University, Texas Agricultural Experiment Station, Blackland Research Center: 808 East Blackland Road, Temple, TX.


Bhaduri B. 1998. A geographic information system based model of the long-term impact of land use change on non-point source pollution at watershed scale. PhD thesis, Purdue University, West Lafayette, IN.


