DISTRIBUTED MODELING FOR IMPROVED NWS RIVER FORECASTS

Seann Reed, Victor Koren, Ziya Zhang, Michael Smith, Dong-Jun Seo
Hydrology Laboratory, Office of Hydrologic Development,
National Weather Service, NOAA
Silver Spring, MD

INTRODUCTION

The Hydrology Laboratory (HL) of the National Weather Service (NWS) Office of Hydrologic Development (OHD) is investigating methods to improve river and stream flow forecasts by taking advantage of multi-sensor, gridded, precipitation products. NEXRAD-based multi-sensor precipitation products are currently produced at NWS River Forecast Centers (RFCs) at a 1-hour temporal and 4-km spatial resolution. Archiving of these products first began in 1993 with the Stage III algorithms (Fulton et al., 1998) and work to improve these algorithms continues (Seo, 2002).

Thirteen National Weather Service RFCs in the United States produce both short and long-term river flow forecasts at about 4,000 forecast points. Current operational methods for hydrologic modeling were designed when the only source of precipitation data was rain gages. Rain gage data can support lumped modeling at 6-hour time steps on basins ranging in size from 300 – 5000 km². Using multi-sensor rainfall products and additional sources of spatial information describing the land surface, there is a potential to (1) improve forecast accuracy at basin outlets and (2) provide hydrologic simulations for smaller ungaged basins (this could improve flash-flood warnings). Improving our understanding of hydrologic processes through modeling is a pre-requisite to achieving these goals. This paper describes distributed modeling research and development being done at HL in an effort to reach these goals. To provide a framework for analysis, HL has recently developed a set of programs referred to as the Research Modeling System (HL-RMS). Some of the main features of the current HL-RMS are:

1. ingests gridded NEXRAD-based products
2. basic modeling unit is the NEXRAD grid cell (~ 4 km)
3. rainfall-runoff calculations are done independently for each grid cell
4. runoff is routed over hillslopes within a model cell
5. channel routing is done from cell-to-cell
6. rainfall-runoff calculations can be done using lumped or distributed rainfall and lumped or distributed parameters
7. uses the Sacramento Soil Moisture Accounting Model (SAC-SMA) (Burnash et al., 1973)
8. uses the kinematic method for both hillslope and channel routing
9. writes output parameter, state, or forcing grids that can be displayed in ArcView GIS

Several factors played a role in designing the features of HL-RMS. Use of the SAC-SMA model for runoff calculations is a practical choice because NWS hydrologists have a large amount of experience with lumped applications of this model. Also, the work of Koren et al. (2000) to estimate SAC-SMA parameters from soil properties makes it possible to run simulations using SAC-SMA parameter estimates that vary within a basin. For simplicity, the 4 km grid cells used...
to map multi-sensor precipitation are used as the basic modeling unit. The 4-km resolution is adequate to resolve important spatial variability in rainfall that can occur across basins modeled by RFCs, while still maintaining acceptable computational requirements. Certainly, there is a lower limit on the size basin that can be modeled with 4-km grid cells, but there is no reason that smaller grid cells (e.g. 2-km or 1-km) cannot be used in future applications.

One reason for maintaining computational simplicity (e.g. grid structure) and familiarity (e.g. use of SAC-SMA) is to facilitate prototype testing and ultimately to simplify the transition from a prototype to an operational application. Initial testing has shown that the computational resources required to run this model are very modest. HL is hoping to learn about the possible benefits of using more complex rainfall-runoff and routing methods in real world applications through participation in and sponsorship of the Distributed Modeling Intercomparison Project (DMIP) (DMIP, 2002) and other cooperative research.

The remainder of this paper includes a description of the HL-RMS model structure, parameter estimation procedures, and some initial results and discussion.

**MODEL STRUCTURE**

Rainfall-runoff calculations are performed for each 4-km NEXRAD grid cell. The rainfall-runoff parameters are assumed to be uniform within each model cell. The SAC-SMA model defines several types of runoff including fast response impervious, surface, and direct runoff, medium response interflow, and slow response supplemental and primary baseflow. In HL-RMS, the impervious, surface, and direct runoff components are routed over conceptual hillslopes within each NEXRAD cell to a conceptual channel. The interflow and baseflow components are assumed to enter the channel system directly from the soil and therefore bypass the hillslope routing. This differs from a typical lumped application of the SAC-SMA model in which all flow components are routed through a unit hydrograph, and there is no distinction between hillslope and channel flow mechanisms.

Because baseflow is generated and assumed to enter the channel within the same cell, there is no physical connection between the lower zone (groundwater) storages in adjacent model cells. This is perhaps a weakness in the current distributed application of the SAC-SMA model because it seems physically reasonable that more baseflow should be generated in the lower reaches of a basins, due to water that is transferred to the deeper channel banks through subsurface or channel flow from upstream cells. Establishing flow relationships among subsurface storage zones in the HL-RMS conceptualization is one topic for future research.

Because of the relatively large size of the 4-km model cells, the cells are subdivided into conceptual hillslopes to make overland flow distances physically realistic. A drainage density parameter in the model is used to subdivide a cell into equally sized overland flow planes (Figure 1a). These hillslopes drain to a conceptual channel segment within the same cell. Cell-to-cell channel routing is done using flow direction networks like that illustrated in Figure 1b. For the relatively coarse resolution model cells, automatically defining an accurate drainage network that correctly describes the geomorphological character of a basin is not a trivial task. Custom HL-RMS pre-processing tools to derive drainage networks for coarse resolution grid cells using
higher resolution digital elevation models (DEMs) have been developed at HL. To facilitate efficient routing calculations, the drainage network depicted in Figure 1b is translated into a computational sequence file using the scheme of Koren et al. (1992).

Figure 1. (a) Conceptual hillslopes and (b) cell-to-cell drainage network

To maintain computational accuracy in kinematic channel routing calculations, the conceptual channel within each cell is subdivided into several reaches of equal length. Approximating the length of channel flow within each cell as $\sqrt{2*cellarea}$ has given reasonable results in initial simulations.

**PARAMETER ESTIMATION**

**Rainfall-Runoff Parameters:** Calibration procedures to estimate SAC-SMA parameters for lumped model applications are well defined (Burnash, 1995; HRC, 1999; Brazil and Hudlow, 1981; Anderson, 2002, Smith et al., 2002). However, by definition, lumped calibrations do not yield any information about how parameters should vary within a basin, desirable information for the implementation of a distributed model.

To estimate the spatial variability of rainfall-runoff parameters within basins, we use a priori SAC-SMA parameter grids developed by Koren et al. (2000). Koren et al. (2000) developed a set of equations that can be used to derive 11 of the SAC-SMA parameters from the Soil Conservation Service (SCS) curve number (McCuen, 1982), properties that can be inferred from soil texture (e.g. porosity, field capacity, wilting point, and saturated hydraulic conductivity), and soil depth. These equations were developed based on both physical reasoning and empirical relationships. Using 1-km soil property grids derived from STATSGO data (USDA, 1994) by Miller and White (1999), Koren et al. (2000) produced a priori SAC-SMA parameter grids covering the conterminous United States. The a priori values of the Upper Zone Free Water Maximum (UZFWM) parameter for the Arkansas River Basin are shown in Figure 6a.
Koren et al. (2000), Koren et al. (2001), and Duan et al. (2001) all have run lumped simulations using basin averaged a priori parameter estimates. These studies suggest that although these a priori estimates cannot outperform a well calibrated lumped parameter set on a gaged basin, the values are reasonable initial guesses for manual or optimized calibrations and the a priori grids may be beneficial in regional analysis. For example, to estimate parameters for an uncalibrated basin, Koren et al. (2001) used ratios of a priori parameters to scale parameters from a nearby, calibrated basin. We also use this simple idea of parameter scaling in some of our HL-RMS runs to describe spatial variability within a basin.

Simulations using three different approaches to assign SAC-SMA parameters are presented in this paper.

1. **Lumped parameters**: Assign uniform parameters to all of the model cells within a basin. The values for uniform parameters are estimated using manual calibration of a lumped model forced by mean areal precipitation computed from the same hourly precipitation grids used for distributed modeling.

2. **Distributed w/ a priori grids**: Use the a priori SAC-SMA parameter grids with no adjustment. Simulations from this method indicate the degree of success that can be expected in modeling without any calibration.

3. **Distributed with scaled a priori grids**: Multiply each a priori SAC-SMA grid by the ratio of the SAC-SMA parameter from lumped calibration to the parameter estimate derived from averaging the values in the a priori grid.

These methods, as well as scenarios aggregating precipitation to different degrees, are all easy to implement within the HL-RMS framework.

**Routing Parameters**: Similar to the rainfall-runoff parameters, routing parameters are assumed to be constant within each model cell. Thus, spatially variable routing parameters are input as grids with the same spatial resolution as the precipitation and rainfall-runoff parameter data.

**Hillslope Routing Parameters**: Three parameters are defined in each cell for kinematic overland flow routing: hillslope slope, hillslope roughness, and drainage density. Note that in the current model structure, hillslope slope and hillslope roughness may vary from cell to cell, but not among the conceptual hillslopes within a cell. Representative hillslope slopes are estimated using DEM data (initially with 30-m DEM data for basin scale applications and 400-m DEM data for regional scale applications) by first computing the local slope of each DEM cell in the study domain using the Arc/Info slope function (ESRI, 1994), and then averaging all of the DEM cell slopes in each 4-km model cell. In simulation runs presented here, a constant estimate of hillslope roughness (0.15) has been assigned for all model cells. Certainly, spatially variable hillslope roughness values could be related to land use estimates based on a lookup table (e.g. Skahill and Johnson, 1999); however, implementing this option has not yet been a high priority given that (1) within a given land use category, published values of roughness values cover wide ranges of possible values that often overlap with the ranges assigned to other land use categories and therefore offering limited guidance in defining spatial variability, (2) initial outlet simulations using spatially constant hillslope roughness have been satisfactory, (3) HL-RMS has shown more sensitivity to channel routing parameters than hillslope parameters in the basins studied.
For drainage density, Dingman (1993) notes values ranging from 2 \( \text{km}^{-1} \) to 100 \( \text{km}^{-1} \) have been reported in the literature, and that drainage density is related to climate and geology. For areas we are modeling in the Arkansas and Red River Basins, a spatially constant value of drainage density (2.5 \( \text{km}^{-1} \)) has been assumed. Based on model results, this seems like a reasonable assumption.

**Channel Routing Parameters:** In order to solve the continuity equation (Equation 1) using the kinematic routing method, two parameters in the momentum equation describing steady, uniform flow must be prescribed for all points in space (parameters \( a \) and \( b \) in Equation 2).

\[
\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = q \quad (1)
\]

\[
Q = aA^b \quad (2)
\]

Q is flow in \([\text{L}^3 \text{T}^{-1}]\), A is cross-sectional area of flow \([\text{L}^2]\), q is lateral inflow per unit length of channel \([\text{L}^2/\text{T}]\), \(x\) is the distance along the channel, \(t\) is time, and \(a\) and \(b\) are model parameters. We sometimes refer to \(a\) as the channel specific discharge.

The \(a\) and \(b\) parameters can be defined two ways for HL-RMS. Method 1: HL-RMS reads four parameter grids that can be used to derive \(a\) and \(b\). Method 2: HL-RMS directly reads pre-defined grids of \(a\) and \(b\). The four grids required for Method 1 include channel slope (S), channel roughness (n), and two channel shape parameters (\(\alpha\) and \(\beta\)). Pre-processing procedures to derive input grids for either method have been developed and are described here.

The basic idea used to derive distributed routing parameters is to combine point measurements that are available at USGS streamflow stations with the geomorphologic information that can be derived from DEMs. Flow measurement data, including top width (B), flow cross-section (A), and flow (Q) (derived from velocity measurements), are intermittently collected by the USGS at stream gaging stations in order to derive stage-discharge curves. The fact that the USGS has recently added an option to download these data from their national web site makes the parameter estimation process much easier.

Although the flow measurement data can be used directly to derive required channel parameters at the basin outlet, the goal is to get parameter estimates at upstream grid cells. For Method 1, both channel slope and drainage area information in each cell are used along with the outlet parameter information to derive routing parameter grids, while in Method 2 only the drainage area is used. The drainage area for each model cell is implicit in the network connectivity (Figure 1b), and representative estimates of channel slope are derived using high resolution (e.g. 30-m) DEM data. To derive channel slope estimates, the average of the slopes of the largest streams branches that pass through each 4-km model cell are used. As should be the case, channel slope estimates for a cell are significantly different than the hillslope slope values.

In Method 1, a simple parabolic channel shape defined by Equation 3 is assumed
\[ B = \alpha H^\beta \]  

(3)

where \( H \) is flow depth, \( B \) is top width, and \( \alpha \) and \( \beta \) are model parameters. Given Equation 3, it is easily shown that

\[ H = \alpha (\beta + 1) H_{avg} \]

(4)

and

\[ B = \alpha^{\beta+1} [(\beta + 1) A]^{\beta+1} \]

(5)

Parameter estimates at a basin outlet, \((\alpha_o, \beta_o)\) are determined by fitting a curve to a plot of \( A \) vs. \( B \) data. With this simple channel model, it is not necessarily possible to get a good fit to the \( A \) vs. \( B \) plot at all flow levels; therefore, we aim to get good fits at high flow levels because we are interested in modeling floods. In our applications, the shape parameter, \( \beta \), is assumed to be constant within a basin, but spatially variable values of the width parameter, \( \alpha \), are derived.

Channel roughness \((n_i)\) values for each cell are calculated using an empirical equation (Tokar and Johnson, 1995),

\[ n_i = n_o S^{0.272} F^{-0.00011} \]

(6)

where \( F \) is upstream drainage area and \( S \) is channel slope. Locally applicable values for the coefficient, \( n_o \), are derived using USGS flow measurements at the basin outlet.

Two basic geomorphologic assumptions and the \( n_i \) values are used to estimate \( \alpha \) values at upstream cells. The first assumption is that channel forming flow is a known function of drainage area, and the second is that cross-sectional area is a known function of stream order. In the current HL-RMS, the following relationships are assumed:

\[ \frac{Q_i}{Q_o} = \frac{F_i}{F_o} \]

(7)

\[ r_i = \frac{A_i}{A_o} = 0.013^{(0.83i - 0.83k)} \frac{R_{i-k}}{R_{i-k}} \]

(8)

where the subscript \( i \) denotes any upstream cell and the subscript \( o \) denotes the value at the outlet, \( R_i \) is Horton’s length ratio, and \( k \) is stream order. Gorbunov (1971) suggests the cross-sectional area relationship in Equation 8. One advantage of using this stream order relationship rather than a relationship that is strictly a function of drainage area is that the \( R_i \) parameter can be varied to reflect local characteristics. In our analysis, we use \( R_i = 2.1 \).

The procedure to estimate distributed \( \alpha_i \) values is as follows. For a selected flow level at the outlet, estimate the upstream flow \((Q_i)\) and cross sectional area \((A_i)\) using Equations 7 and 8.
From this information, a representative velocity \( (v_i) \) at each upstream cell is estimated, and the average depth \( (H_{avg_i}) \) is back-calculated using Manning’s equation. The top width \( B_i \) is then calculated as \( A_i/H_{avg_i} \) and Equation 4 is solved for \( \alpha_i \). With known channel geometry at each cell, values of channel specific discharge \( (a) \) are estimated and used for kinematic channel routing calculations.

To implement Method 2, channel specific discharge \( (a) \) and the exponent \( (b) \) in Equation 2 are estimated at the outlet by fitting a curve to the \( A \) vs. \( Q \) data. The same geomorphologic assumptions are used to estimate \( a \) and \( b \) at upstream cells and from Equations 7 and 8, it can be shown that:

\[
a_i = \frac{F_i}{F_o} a_o \left( \frac{1}{r_i} \right)^{b_o} \quad \text{and} \quad b_i = b_o \quad (9)
\]

As discussed in the next section, both methods for channel parameter estimation yield reasonable results in simulation runs. As an independent check on the procedure, Figure 2 shows that the pattern of predicted channel width variations along the main channel is similar to the pattern of actual channel top width measurements made in the field. What is interesting here is the pattern rather than the absolute widths. Top widths measured in the field are based on a channel bank approximation and there is no channel bank in the simplified conceptual channel defined by Equation 3. The estimated widths shown in Figure 2 are generated for a flow level that produces similar overall widths.

\[
\begin{align*}
\text{Estimated} \\
\text{Bank width from c.-s.}
\end{align*}
\]

Figure 2. Comparison of estimated and observed top widths

**RESULTS AND DISCUSSION**

Initial testing of HL-RMS is being carried out for basins within the Arkansas and Red River Basins (Figure 3). The main reason for this is that the Arkansas-Red Basin River Forecast Center (ABRFC) has the longest archive of 4-km multi-sensor precipitation grids (May 1993 – present), and these rainfall grids have been evaluated more thoroughly than grids produced in other parts of the country. Also, these basins are not regulated. Others have also used this area
in distributed modeling studies (e.g., Carpenter et al., 2001). Although all of the basins shown in Figure 3 are being studied as part of DMIP, specific results for only a few of these basins will be discussed here. It should be noted that the following simulations were derived by HL scientists and do not correspond to the forecasts derived at the ABRFC. RFC forecasts include a significant amount of human quality control not present in simulations shown here. Ongoing developments on assimilation/updating techniques (Seo et al., 2002) should also improve the performance of HL-RMS.

![Figure 3. Location reference for the Arkansas, Red, and DMIP basins](image)

A critical question to consider in evaluating the potential benefits of distributed modeling for NWS operations is whether a distributed model can produce simulations that are comparable to or better than simulations from existing lumped models. Of course, producing better simulations is ideal, but the ability to produce comparable simulations is also a positive result because there are other potential benefits from running a distributed model, like the ability to simulate flows at small, ungaged sites within a basin, and the ability to incorporate future sources of spatial data that describe hydrologic variables (e.g. satellite observations).

In general, distributed model runs for the Blue River at Blue, OK, have shown noticeable improvements over lumped simulations, while distributed model simulations for the Illinois River at Watts, OK, and the Baron Fork at Eldon, OK, have yielded results comparable to lumped simulations.

Figures 4a and b show hydrographs for two events in the Blue River. These two events have been selected from a 6 year continuous simulation (June 1993 – May 1999) to illustrate results that can be obtained in events with relatively non-uniform and relatively uniform rainfall. Several other events during the calibration period show results similar to those shown here. Figure 4a shows that a distributed model outperforms a lumped model when the rainfall
distribution is relatively non-uniform. Figure 4b shows that the lumped simulation results are more comparable for cases with less spatial rainfall variability. Based on visual and statistical comparisons of hydrographs at the basin outlet, the use of scaled, a priori SAC-SMA grids produces better overall simulations than the use of unscaled a priori grids.

Figure 4. Representative events for the Blue River. (a) Case with highly non-uniform rainfall. (b) Case with relatively uniform rainfall.

Initial distributed model simulations in the Illinois River Basin above Watts (1645 km²) and the Baron Fork at Eldon (795 km²) show results comparable to lumped simulations. In these basins, little to no gain from distributed modeling is seen in the overall simulation quality at the basin outlet. The difference from the Blue River results is likely due to a number of factors including differences in basin shape and orientation, and the possible dampening affects of deeper soils in Watts and Eldon basins. These results agree with those of Obled et al. (1994) who suggested that the spatial variability of the precipitation was not sufficiently organized to overcome the effects of dampening present in the basins studied.
Further study of the Tahlequah (2484 km$^2$) and Eldon basins is of interest because of the potential to verify modeling results at interior gaged locations. The Watts, Savoy (433 km$^2$), and Flint Creek (295 km$^2$) basins are all within Tahlequah and Peacheater Creek (65 km$^2$) is within the Eldon basin.

An interesting result from this type of study is presented in Figure 5. In Figure 5a, the distributed model with scaled SAC-SMA parameters produces results comparable to a lumped simulation and better than a distributed model with unscaled a priori parameters at Eldon. However, for the same storms at Peacheater Creek (Figure 5b) the model with unscaled a priori parameters produces better results than runs with both lumped and scaled SAC-SMA parameters. Note that the lumped parameters and scaling factors used in this example are based on
calibration at Eldon only, without calibration at Peacheater Creek. It is perhaps not difficult to accept this result because the lumped parameters and scaling factors derived from manual calibration at Eldon are scale dependent (Koren et al., 1999) and they implicitly account for a certain scale of rainfall averaging. Although this result indicates that there are scaling issues, it should not deter continued efforts to refine modeling and calibration procedures to produce better results at small, nested basins while maintaining good hydrologic performance at the outlet.

Modeling results for an interior point within the Watts basin, the Illinois River at Savoy (433 km²), do not show the same trend as Figure 5 when moving from a larger to a smaller basin. That is, results using scaled grids based on a Watts calibration and unscaled a priori grids yield comparable simulation results at Savoy. Perhaps the difference in these results is due to the fact that the Eldon basin is 12 times as large as Peacheater Creek, while the Watts basin is only about 4 times as large as the Savoy basin. The reality is that defining a distributed parameter model calibration strategy that is robust across many spatial scales remains a challenge.

Strategies to calibrate several basins in a hydrologic region have long relied on qualitative assessments of soil and vegetation properties. The scaling of a priori parameter grids used here and the similar procedures described by Koren et al. (2001) for adjacent basin are attempts to approach this type of analysis in a quantitative and consistent manner. These procedures can continue to be refined, both with new theory and the use of additional data sources.

In order to improve our understanding of regional variability in both rainfall-runoff and routing parameters, HL-RMS is also being applied to wider areas. Given the simple structure of the model, it is possible to run simulations for large areas such as the Arkansas River (408,939 km²). From a computational standpoint, this type of run is feasible. To produce a 10 day simulation in quasi-forecast mode for the entire Arkansas River area only takes 5-6 minutes of CPU time on an HP9000/J5000 workstation.

Figures 6b-d show example results from this type of demonstration run. For this demonstration, a priori parameter grids are used for rainfall-runoff modeling and routing parameter grids are derived using data from only a few stations on the main stem of the Arkansas River. This demonstration of running the model on a large area is considered a beginning and not an end. Certainly, we do not initially expect accurate hydrograph simulations in many parts of the Arkansas River, because there are numerous processes that are not represented in HL-RMS (e.g. reservoirs and the potential for dynamic flow variations and backwater effects on the larger rivers). However, running HL-RMS over a large area provides a framework for further study. Studies on specific subbasins can begin to populate the HL-RMS input grids with more accurate parameter estimates and refine the regional picture.

**SUMMARY AND CONCLUSIONS**

HL-RMS provides a framework to run distributed models that will improve our understanding of hydrologic process. HL-RMS shows the potential to improve the accuracy and resolution of river forecasts. The basic conclusions from initial simulation runs are:
• HL-RMS produces simulations that are comparable to or better than the simulations that are produced by lumped models.

• The improved results were found in the Blue River, OK, for events where spatial variability of rainfall is significant.

• Schemes developed to estimate distributed routing parameters produce reasonable results for moderate to large storm events without any calibration.

• Progress has been made in quantitatively estimating spatially variable rainfall-runoff parameters, but a robust method to calibrate the distributed parameter model is still not well defined.

The simple gridded structure of the model facilitates both scientific research and prototype testing in field offices. HL-RMS results will be compared with more complex models through the DMIP project to identify possible improvements. Future work should also include exploring the use of additional data sources (e.g. NDVI, soil moisture observations) to help improve both model formulation and parameter estimation procedures.

Plans to run the model over large areas will also help to refine calibration procedures and identify model components that can be improved. More applications of HL-RMS will improve our understanding of subbasin processes and scale issues important for flash-flood modeling.

Figure 6. HL-RMS demonstration for the Arkansas River Basin
REFERENCES


