Sudhakar Mamillapalli, R. Srinivasan, J. G. Arnold, Bernard A. Engel

Effect of Spatial Variability on Basin Scale Modeling

ABSTRACT

With the integration of GIS and distributed parameter hydrologic models, a watershed can be divided into many subbasins. However, the effect of discretization on the quality of the simulated output has not been studied very widely. Using the concept of virtual basins the different soil and landuses in the subbasin can be simulated to the level of detail desired. Using a 4300km^2 watershed in Texas, the present study was undertaken to study the effect of increasing level of discretization and virtual basins, on the accuracy of the output. The results indicate that in general, increasing level of discretization and increasing the number of soil and landuse combinations simulated within each subbasin increases the accuracy of the simulation. There is a level beyond which the accuracy can't be improved, suggesting that more detailed simulation may not always be necessary. Preliminary investigations have been conducted to determine optimal configuration.

INTRODUCTION

Hydrologic models can be broadly divided into lumped parameter models and distributed parameter models. The lumped parameter approach considers the whole catchment as a single entity and maps the input rainfall excess to an output hydrograph. Though computationally efficient, this approach doesn't explicitly account for spatial variabilities present within the catchment. Chief among this type of model is the USLE(Wischmeier and Smith, 1978). Distributed models divide the catchment into a number of smaller areas (which could be square elements or subcatchments), which are assumed to be uniform with respect to the hydrologic parameters. Hydrology is simulated within each of these elements and the output routed to the outlet. Hence these models take into consideration spatial variability of the watershed. Examples of these include the AGNPS (AGricultural Non-Point Source Pollution) model (Young et al., 1987), ANSWERS (Aerial Nonpoint Source Watershed Response Simulation)(Beasely et al., 1977) and SWAT (Soil and Water Assessment Tool) (Arnold et al., 1993). Considerable time and effort are required to acquire the data, run the models and interpret resulting information. Integration with a GIS can eliminate many of these problems. Several models have been integrated with GIS which include AGNPS and GRASS GIS by Srinivasan and Engel (1994), ANSWERS and GRASS (Rewerts and Engel, 1991), and SPUR and ERDAS (Sasowsky and Gardner, 1991).

As noted before, these models either discretize the watershed into smaller elements by overlaying a square grid(ANSWERS or AGNPS) or into various subbasins (SWAT and SPUR). With the integration of these models with a GIS, it is possible to divide the watershed into a large number of elements since the GIS automatically generates the input. Hence we can consider the spatial variability to the level of detail supported by the data. However, as the number of such elements increase, so does the computation time. It is not clear from studies to date that the effect of increasing input levels of detail improves the accuracy of the simulated output. For effective use of the above tools, it is necessary to be able to discretize the watershed to an appropriate level of detail. A gross discretization may lead to poor simulation results whereas very fine discretization would require far more input data and significantly increased computation time and space (which may be important for large watersheds comprised of hundreds of subbasins) with no or little increase in accuracy. Also using the concept of virtual basins, different soil and landuse combinations within a subbasin can be simulated, instead of considering the dominant landuse and soil to be representative of the subbasin. It may not be necessary to consider all the combinations within a subbasin. The impact of increased detail in accounting for various soil and landuse combinations within a subbasin and effect of level of discretization on accuracy is studied for a watershed in Texas and presented.

OBJECTIVES

1. Quantify the effect that level of discretization has on the accuracy of output obtained from the distributed parameter hydrologic basin model SWAT.
2. Examine the impact of increased detail in simulating the soil and landuse combinations within a subbasin using the concept of virtual subbasins on the accuracy of output using SWAT.

RELEVANT LITERATURE

SWAT(Soil Water Assessment Tool), a continuous daily time step model developed by Arnold et al.(1993), was obtained by adding a new routing structure to the SWRRB model(Arnold et al., 1990, Williams et al.,1985) so as to remove the restriction of only being able to simulate 10 subwatersheds in the case of SWRRB. The new routing structure of SWAT routes and adds flows down through the basin reaches and reservoirs. Apart from this, changes were incorporated to simulate lateral flow, ground water flow, reach routing transmission losses, and sediment and chemical movement through ponds, reservoirs, streams and valleys. SWAT is capable of simulating hundreds of subwatersheds for periods of 100 years or more. The major components of the model include hydrology, weather, sedimentation, soil temperature, crop growth, nutrients, pesticides, ground water and lateral flow, and agriculture management. Additional details about the model can be found in Arnold et al. (1994).

SWAT allows for considerable flexibility in watershed discretization. The watershed can be divided into cells and/or subwatersheds. Different parts of the watershed can be divided differently. The dominant soil and landuse within each subbasin is considered to be the soil and landuse of the subbasin. However, in order to account for multiple soil and landuse combinations, the concept of virtual subbasins was incorporated into SWAT. Instead of assuming the dominant soil and landuse to be the soil or landuse of the subbasin, each subbasin is discretized into virtual areas (referred to as virtual basins), each having a unique soil and landuse combination without reference to their spatial positioning within the subbasin. This is similar to the concept of Hydrologic Response Units (HRU's) given by Maidment (1991). The hydrologic response is generated within each of these virtual areas and then the weighted average (by area) of the response from these virtual subbasins is taken to be the output of the subbasin. Since there can be large numbers of such combinations, a threshold is set. Only soil and landuse combinations forming a proportion larger than that of the threshold are considered. The threshold is arbitrary and is set by the user.

The way the dominant soil and landuse are chosen in SWAT is that, first the dominant landuse is determined within the subbasin, and within this landuse the dominant soil prevalent is determined. This soil and landuse are taken to be representative of the subbasin. It should be noted this may not be the same as the dominant soil and landuse combination. For example this may occur within a subbasin where there are two different landuses A and B each occupying say 60% and 40% and landuse A has 10 different soils occupying 10% of the area while landuse B has just one soil occupying the whole area. In this case the soil landuse combination chosen will be the A landuse with one of the soils even though the other combination may occupy a larger proportion of the area. Since between soil and landuse, landuse affects stream flow more, this approach seems logical. In the case of the virtual basin approach, based on the threshold on landuse set by the user all the landuses occupying an area greater than the threshold are selected. Within these landuses, soils forming a proportion greater than that of the soil threshold set by the user is used to select the soils. Both the thresholds are set by the user. The effect of varying the threshold on the output obtained is studied.

A GRASS GIS interface to the SWAT model was developed by Srinivasan and Arnold (1994). Given the appropriate data layers and data bases the interface extracts the data and writes the SWAT input file. A number of tools were incorporated allowing for automatic extraction of various inputs. These include tools for accessing the appropriate databases, hydrologic tools to retrieve the topographic and other attributes including automatic generation of the routing structure and aggregarion tools to aggregate the inputs at the subwatersed level. Once the input files are generated the model can be run and then the results visualized using the output interface. The GIS interface, facilitated to a large extent the study conducted.

Among studies made to determine the impact of level of discretization on output of basin scale models are that of Wood et al. (1988) and that of Sasowsky and Gardner (1991). Wood et al. introduced the concept of the representative elementary area (REA) in hydrologic modeling. They refer to REA as a fundamental building block of catchment modeling. They argue that at smaller scales, actual patterns of variability of topography, soil or rainfall lead to differences in the output even though the underlying distribution is the same. As larger and larger scales are considered, more and more of the variability is sampled and then finally an area is obtained whose hydrologic response can be considered to be the net effect of the individual point hydrologic responses within the subbasin or basin. So a basin with all its variation in soils, topography, and weather can be represented by these REAs without much loss in quality of the output. To prove the existence of the REA, Wood et al. discretized the Coweta River experimental catchment in North Carolina which had an area of 17km^2, into 3, 19, 39 and 89 subcatchments by the method described by Band and Wood (1986). In order to emulate point hydrologic response which can be then averaged to form the basin hydrologic response, they applied TOPMODEL (Beven and Kirby, 1979; Beven, 1986) within each 30m pixel comprising the catchment. Then pixel output was aggregated to form the subbasin response. The subbasin responses where then arranged in increasing order of their areas and a running average of 15 subcatchments, moving in steps of 5, was taken. The mean area within each window was plotted against the mean average response. The graphs indicated that the areal response stablized at around 1km^2 area. The size was the same for all the outputs studied. Thus they concluded the REA for this catchment was 1km^2. They made further studies and remarked that the size of the REA is governed primarily by the topography. Soil and rainfall variability didn’t have a big role in determining the size of the REA. In this study, variability in only soil, rainfall and topography were studied. In general, large catchments in addition to the above have land use variability to consider.

Sasowsky and Gardner (1991) used three different configurations of the 146km$^2 Walnut Gulch watershed in Arizona: >= 2nd, >= 4th, >= 13th stream order with 28, 15, and 1 channel segments and 66, 37, and 3 contributing areas and made SPUR runs for each of these configurations. The runoff at the outlet was compared with observed data and then the results seemed to imply the >= 4th order stream network gave as good as results as the >= 2nd and hence they concluded that REA exists for the basin under consideration. However, again this study doesn’t consider land use variability and they admit the evaluation criteria used affect the conclusions that could be drawn. Also calibration (the curve number) needed to be done on the model and again the the model evaluation results depended on the changes made in the curve number. Due to ambiguities, it is difficult to conclude that the study establishes the existence of the REA.

METHDOLOGY

A watershed in Texas of size 4297 km^2 was used in this study. It has originally been subdivided into 40 subbasins, composed of agriculture and range land. Figure 1 shows the watershed with the 40 subbasins. Using the "r.watershed" tool within the GRASS GIS and the 1:250,000 DEM, the watershed was discretized into 4, 8, 14, 20, 24, 29, 35, 40 and 54 subbasins. Measured stream flow data was available at two locations within the watershed. Since both these gages are not located at the outlet of the subbasin, the simulated flow draining into the basin where the gage was located was extracted and compared with the output. Statistics used in the comparison are the coefficient of determination, and the coefficient of efficiency of Nash and Sutcliffe (1971). A coefficient of efficiency of 1 indicates perfect agreement. If the results are highly correlated but biased, then the coefficient of efficiency will be less than the coefficient of determination (Aitken, 1973).

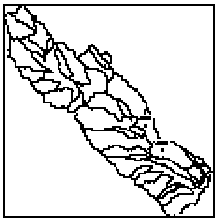


Figure 1: The Bosque Watershed

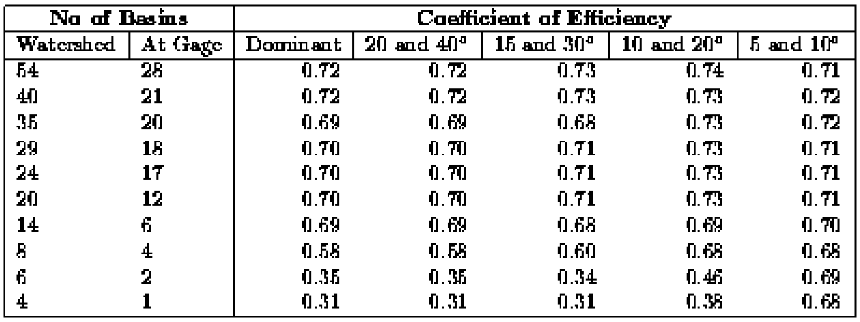
Simulations were made both for the dominant case where the dominant soil and landuse within the basin was considered to be the soil and landuse of each subbasin and the virtual basin approach with thresholds ranging from 5% to 20% for landuse and 10% to 40% for soil. For example a threshold of 10% for soil and 5% for landuse indicates that landuses which form at least 5% of the subbasin area and soils which form at least 10% of the area within each of the selected landuses will be taken as virtual basins. Results for all these cases are presented here. In the table of results the different thresholds are mentioned as "landuse threshold and soil threshold". For example "10% and 20%" in the results table indicate that 10% is the landuse threshold and 20% the soil threshold used for the corresponding results. Observed flow data was available for years 1965 to 1974 and 1975 to 1984 for two different USGS gages 5000 and 5200 within the basin. Simulations were made for these time periods and the results compared. A single simulation was not done for both these time periods since the rain gage data available changed in 1975. No calibration whatsoever was attempted throughout the study, so that the impact of spatial variability alone can be studied. Also to remove the impact of weather variability, the Thiessen polygon average of all the weather gages present in the basin was taken to be the rainfall and temperature data for all the subbasins within the basin.

RESULTS

Statistics, including mean, standard deviation, coefficient of efficiency, and coefficient of determination were computed for the simulations described previously. Results are presented for both USGS gages 5000 and 5200 within the study basin for years 1965-1974 and 1975 to 1984.

These gages are upstream of the outlet, so the number of basins draining into the basin having these gages is different from the total number of basins in the watershed. In the result tables, the number of basins draining into these gages, as well as the total number of basins in the watershed are presented. The coefficient of efficiency (COE) is used as a measure of accuracy of the simulated results and presented here, though other measures generally followed the same trend.

Table 1: Results Obtained when Runoff Simulated using Various Approaches for Different Basin Configurations for Years 1965 to 1974 for Gauge 5000



No of Basins		Coefficient of Efficiency				
Watershed	At Gage	Dominant	20 and 40[a]	15 and 30[a]	10 and 20[a]	5 and 10[a]
54	28	0.72	0.72	0.73	0.74	0.71
40	21	0.72	0.72	0.73	0.73	0.72
35	20	0.69	0.69	0.68	0.73	0.72
29	18	0.70	0.70	0.71	0.73	0.71
24	17	0.70	0.70	0.71	0.73	0.71
20	12	0.70	0.70	0.71	0.73	0.71
14	6	0.69	0.69	0.68	0.69	0.70
8	4	0.58	0.58	0.60	0.68	0.68
6	2	0.35	0.35	0.34	0.46	0.69
4	1	0.31	0.31	0.31	0.38	0.68

---

[a] Landuse and Soil Thresholds