COMPARISON OF SIMULATION RESULTS USING SSURGO-BASED AND STATSGO-BASED PARAMETERS IN A DISTRIBUTED HYDROLOGIC MODEL

Ziya Zhang[1], Ziya.Zhang@noaa.gov; Victor Koren[1], Victor.Koren@noaa.gov; Seann Reed[1], Seann.Reed@noaa.gov; Michael Smith[1], Michael.Smith@noaa.gov; and Fekadu Moreda[1], Fekadu.Moreda@noaa.gov

[1]Research Hydrologist, Hydrology Laboratory, Office of Hydrologic Development
NOAA National Weather Service
1325 East West Highway, Silver Spring, MD 20910

Abstract Better estimation of *a priori* gridded parameters is a significant element in the operational implementation of a distributed hydrologic model in the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service. Improved *a priori* parameters have the potential to reduce the cost of calibration of distributed hydrologic models. Calibration costs are a key hurdle to overcome in using distributed models to improve water resources, flood, and flash flood forecasting capabilities for the United States. An initial effort in deriving *a priori* gridded model parameters was based on a relatively coarse-resolution GIS-compatible soils property database, the State Soil Geographic Database (STATSGO). There are limitations in using the STATSGO data because the desired modeling scale has a higher spatial resolution than the data. With the existence and greater availability of finer-scale soils data, the county-level Soil Survey Geographic Database (SSURGO), together with the 30m resolution of land cover and land use within each soil polygon, we are attempting to improve upon our initial estimation methodology. The effects of SSURGO-based model parameters on distributed simulations were analyzed for several basins in the United States through comparisons to the simulations using STATSGO-based model parameters. Simulated hourly flow time series from both SSURGO-based and STATSGO-based cases are compared to hourly observed data to evaluate the relative performance. Some improvements from using SSURGO-based *a priori* instead of STATSGO-based parameters were observed.

INTRODUCTION

The Sacramento Soil Moisture Accounting Model (SAC-SMA) used at NOAA's National Weather Service River Forecast Centers (NWS RFCs) needs to be calibrated in order to get an accurate set of parameters for hydrologic forecasting. Because simulation statistics can be similar with different sets of parameters, a challenge in model calibration is not only to produce a good fit between simulated and observed data but to maintain spatial consistency in parameter derivation and make parameter adjustments that are consistent with the conceptual model. Parameter inconsistency will limit the transferability of parameters across basins in a region. The availability of initial parameter values that reflect the spatial variability of land surface properties in a meaningful way can help to address this problem. The availability of reasonable initial parameter sets is even more critical in the case of distributed parameter modeling where the intra-basin spatial variability of parameters must be specified. The potential benefits of better initial parameter estimates include both reducing the cost of calibration and improving simulations for ungauged basins (Koren, et al., 2000; Carpenter and Georgakakos, 2004). By reducing the subjectivity in calibration, use of gridded *a priori* model parameters will assist with reproducibility and producing reasonable distributions of parameters over a large region or across different regions.

In current operational practice, most hydrological models are lumped (that is, a single set of parameter values apply to the entire geographic extent of a hydrologic basin) and the required manual or automated calibration procedures are well defined. With the availability of spatially detailed data, increased computer processing power, and ever increasing demand for localized information, more distributed hydrological models are being developed and applied for both research and operational uses (Abbott, et al., 1986; Bell and Moore, 1998; Koren et al., 2004; to name a few). One of the challenges with distributed modeling is to derive a set of initial parameters that are based on a basin's physical properties to assist with manual or automatic calibration.

Koren at al. (2000) developed a systematic approach that uses the National Resources Conservation Service (NRCS) State Soil Geographic Database (STATSGO) to derive eleven SAC-SMA parameters nationwide with the grid size of about 16 km^2. The STATSGO data are available at the scale of 1:250,000. The soil polygons defined in the STATSGO often range in size from about 100 to 200 km^2. Although the STATSGO based gridded parameters provide a good estimation of initial values for distributed modeling as shown in the Distributed Model Intercomparison Project (DMIP) (Smith, et al., 2004), there is a potential for improvement.

One opportunity for improvement is to use the NRCS Soil Survey Geographic Database SSURGO data which has a resolution approximately 10 times higher than that of STATSGO data in terms of areas of soil polygons. Digitization of these county level data is work in progress. The completion of SSURGO data coverage for whole USA is expected by 2008 (http://www.ncgc.nrcs.usda.gov/products/datasets/ssurgo). A second opportunity is to explicitly use landuse/landcover data in conjunction with soil property data as input to the estimation process. In deriving an initial STATSGO-based parameter set for the conterminous United States, Koren et al. (2000) assumed "pasture or range land use" under "fair" hydrologic conditions.

Prior to embarking on extensive processing necessary to derive SSURGO-based parameters over large areas, it is prudent to evaluate the potential benefits for hydrological modeling that may be achieved using the new higher resolution data set. However, there are only a few published papers that describe the comparisons of the impact of using STATSGO-based parameters and SSURGO-based parameters on simulated results. In comparing outlet streamflow simulations using STASGO-based and SSURGO-based parameter estimates for the Little Washita watershed (600 km^2) in Oklahoma, Reed (1998) found little difference between the two cases. Part of the reason was that the overall surface soil texture distributions defined by STATSGO and SSURGO are similar for this basin. Anderson et al. (2005) ran the SAC-SMA model for several basins in the Ohio River Forecasting Center (OHRFC) and the West Gulf River Forecasting Center (WGRFC) of NWS using STATSGO-based and SSURGO-based parameters. Lumped simulations in which averaged parameter values were used for each basin were created and compared. Anderson et al. (2005) used the 1992 National Land Cover Dataset (NLCD) (30 meter resolution) in combination with STATSGO and SSURGO soil data to estimate SAC-SMA parameters. In these cases Anderson found improvement of outlet flow simulation performance using SSURGO data where there is a considerable difference in soil texture distributions as derived based on STASGO and SSURGO data. Since in their work averaged parameters were used in a lumped simulation mode, much of spatial variation effect of soil texture on hydrological response within a basin was lost. Study of the impact of SSURGO-based gridded parameters for distributed modeling is necessary in order to take full advantage of the fine-scaled

SSURGO data and possibly improve simulations for small basins where STATGO data are too coarse to resolve intra-basin variations.

METHODS

The work presented here is based on gridded SAC-SMA parameters using SSURGO soil data. These gridded parameters were derived using similar approaches to those used by Koren et al. (2000) and Anderson et al. (2005). Initial parameter values were calculated for each soil polygon defined in the SSURGO data set. These parameter maps of polygons were then transformed to gridded data through area-weighted averaging at the resolution desired. The derived SSURGO-based and STATSGO-based *a priori* parameters were then used in the Hydrology Laboratory Research Modeling System (HLRMS) (Koren et al., 2004) to produce distributed simulations. In both cases, a kinematic wave overland flow and channel routing model was used. There were no calibrations for either case. Results from the two cases were compared and analyzed.

TEST BASINS AND RESULTS

Eleven basins located in the states of Oklahoma and Arkansas were selected to run HLRMS in distributed mode at an hourly time step. Table 1 and Figure 1 show the basin information and location map. Some of the basins were studied in the Distributed Intercomparison Project (DMIP) (Smith et al., 2004; Reed et al., 2004). Therefore, extensive hydrological data have been collected and are available for these basins. Hourly gridded radar precipitation data are available from June 1993 on. Hourly observed flow data at each basin's outlet were obtained from the U.S. Geological Survey (USGS). The archived hourly flow data downloaded are provisional with no quality control by the USGS yet. However, scientists in the Hydrology Laboratory of NWS did some rudimentary quality control by comparing to USGS official daily flow data and setting suspicious hourly data to missing. The SSURGO soil data were downloaded from the Geospatial Data Gateway (http://datagateway.nrcs.usda.gov) maintained by the NRCS. Land cover data at a 30 m resolution were obtained from the 1992 National Land Cover Dataset (NLCD) through

Table 1 Basin information

No.	USGS No.	Short Name	Station Name	Area (km^2)
1	7195800	SPRINGT	Flint Creek at Springtown AR	36.8
2	7195865	SSILOAM	Sager Creek near West Siloam Springs OK	48.9
3	7196973	CHRISTI	Peacheater Creek at Christie OK	64.7
4	7196900	DUTCH	Baron Fork at Dutch Mills AR	105.1
5	7196000	KNSO2	Flint Creek near Kansas OK	284.9
6	7195000	ELMSP	Osage Creek near Elm Springs AR	336.7
7	7191220	SYCAM	Spavinaw Creek near Sycamore OK	344.5
8	7197000	ELDO2	Baron Fork at Eldon OK	795.1
9	7195430	ISILOAM	Illinois River South of Siloam Springs AR	1489.2
10	7195500	WTTO2	Illinois River near Watts OK	1644.6
11	7196500	TALO2	Illinois River near Tahlequah OK	2483.7

Seamless Data Distribution (http://seamless.usgs.gov/website/seamless/viewer.php) maintained by the USGS. Simulations were compared for two cases where SSURGO-based and STATSGO-based SAC-SMA parameters are used. The model grid cell size is about 4 km^2 (2×2). Results analyzed here are based on simulations from 1996 to 2004. Initial gridded model states were created by running the model from 1993 to 1996 to avoid the effect of initial conditions on the simulations analyzed here.

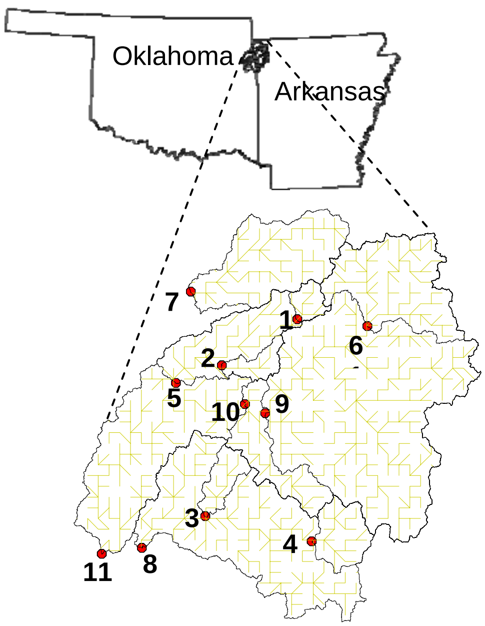

Figure 1 Basin location map. The network inside of basin map shows the grid connectivity at the 2 km scale. It is used for distributed channel routing.

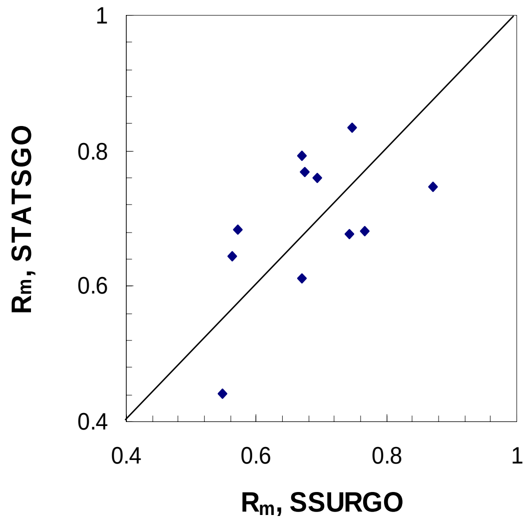

Figure 2 Comparison of R_m of outlet flow for whole time series of 11 basins.

Figure 2 shows the comparison of modified correlation coefficient R_m of discharge between the two cases with R_m values computed relative to the observed discharge time series. R_m is

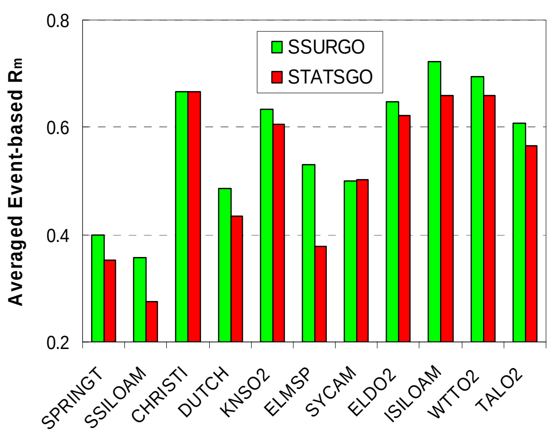

Figure 3 Comparison of averaged R_m based on selected events for 11 basins.

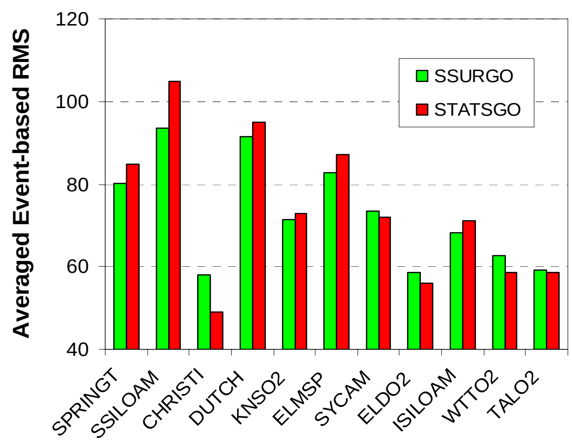

Figure 4 Comparison of averaged RMSE based on selected events for 11 basins.

calculated by reducing normal correlation coefficient by the ratio of the standard deviations of the observed and simulated hydrographs (McCuen and Snyder, 1975). The calculation of R_m was based on the time series from the whole simulation period. It shows that overall flow simulation performance is similar for the two cases as the points fall on both sides of the diagonal line.

For flood forecast purposes, the ability to have better flow simulations during heavy rainfall events is critical. Therefore the largest events (15-17) within the simulation period (from 1996 to 2004) for each basin were selected for event analyses. Some statistics characterizing the flow simulations, peak flow values and peak timing based on selected events are presented in Figures 3-6. Note that the basins in these figures are listed from left to right in order of increasing size. Figure 3 shows the comparison of R_m for flows of the selected events. Values of R_m for the SSURGO-based case are higher than those for the STATSGO-based case for 9 of 11 basins and nearly identical for the remaining two basins, indicating that use of SSURGO-based parameters can improve flow simulations for big events. The RMSE statistics shown in Figure 4 do not show improvement in as many basins as the R_m statistics (6 of 11).

While Figures 3 and 4 show statistics for all flows during large events, Figure 5 and 6 show comparisons of statistics for peak flow values and timing of these events. Comparing simulated peak flow values against observed flow values, peak flow errors were calculated for each basin for the two cases as shown in Figure 5. Peak flow errors associated with the SSURGO-based case are smaller than or equal to the STATSGO-based case for 8 of the 11 studied basins. Similarly, improvements for 9 of 11 basins can be seen in the comparison of peak flow time error shown Figure 6. These results based on selected large events have shown that using SSURGO-based *a priori* parameters can improve flow simulations over using STATSGO-based *a priori* parameters for most statistics and most selected basins.

We plan further examination of the spatial patterns and statistical variability of soil properties within the modeled basins. We hope that these additional analyses will help to explain cases when use of SSURGO data did not show improvement. It will also be insightful to more carefully examine the role of scale in these results. An interesting observation about the RMSE statistics is that 4 of the 5 basins where STATSGO-based simulations outperformed SSURGO-

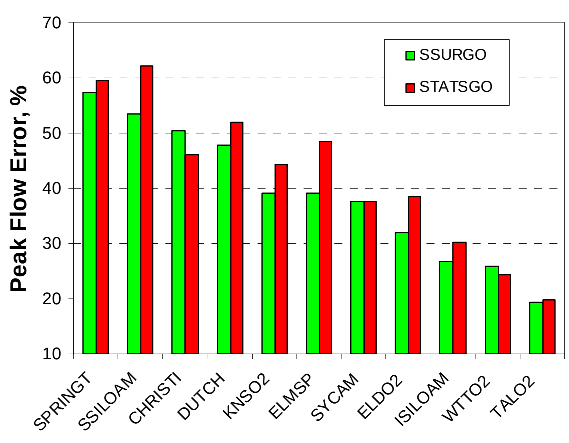

Figure 5 Peak flow error comparison based on selected events for 11 basins.

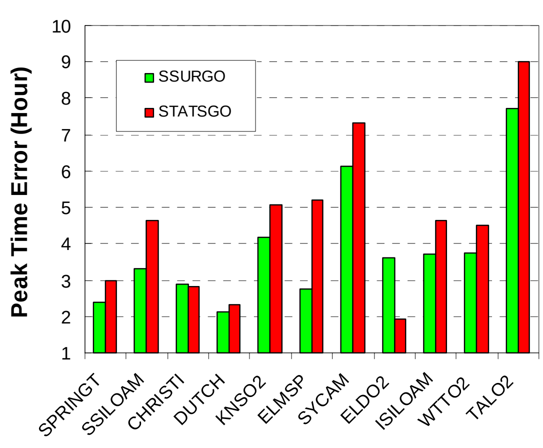

Figure 6 Peak flow time error comparison based on selected events for 11 basins

based simulations are large basins.

When hydrographs are plotted in log-scale as shown in Figures 7 and 8, an interesting result exists for base flow comparison during recession. Each plot shows one water year of flow time series with Figure 7 for a small basin and Figure 8 for a large basin. For the small basin, when compared to observed flow data, base flow based on SSURGO soil data consistently outperforms

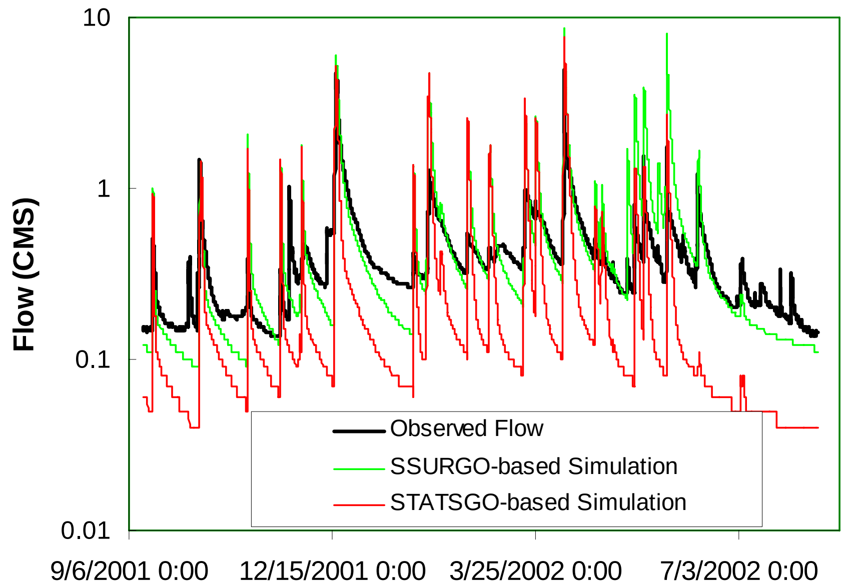

Figure 7 Hydrograph comparison in log-scale for a small basin, SPRINGT.

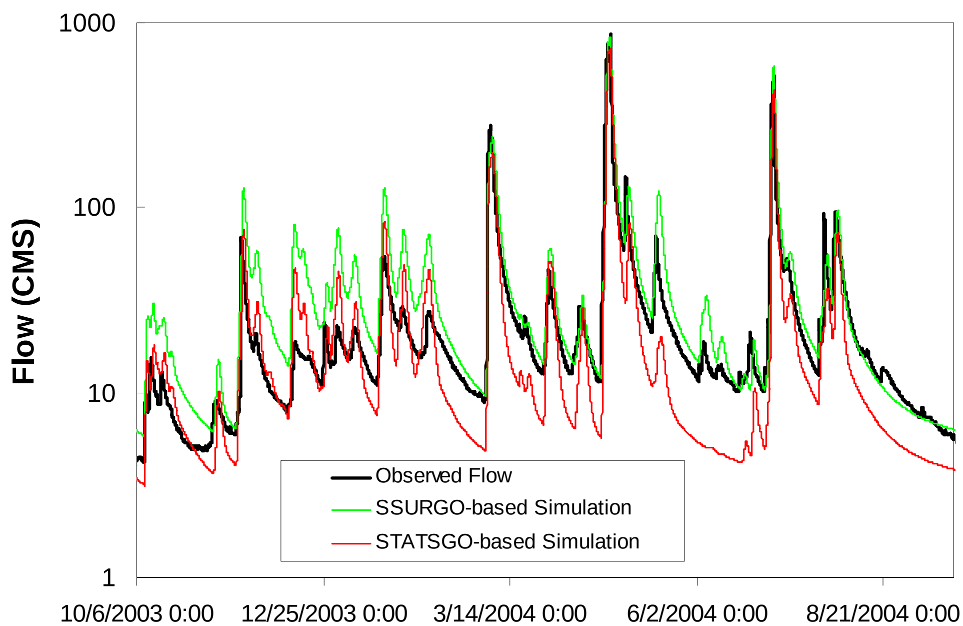

Figure 8 Hydrograph comparison in log-scale for a large basin, TALO2.

that based on STATSGO soil data. Whereas for the large basin, a better performance for SSURGO-based case was observed only during the spring and the summer periods. Further investigation will be required to understand the differences between the small basin and large basin results and how the improvements relate to each model parameter.

CONCLUSIONS

Characterizing the spatial variability of model parameters is a central challenge in distributed hydrologic modeling. Coarse resolution soil data of STATSGO were initially used to derive gridded model parameters for a gridded SAC-SMA model implementation. While STATSGO helps a great deal in running a distributed model with distributed parameters, its coarse resolution limits the parameter variation over small areas in which the local hydrological response may be important, particularly for flash floods. We investigated whether or not the finer-scale soil data of SSURGO can improve model simulation results over use of STATSGO data. Several statistics were presented based on the analyses of two simulated discharge time series; one generated using SSURGO-based parameters and the other using STATSGO-based parameters. It was found that although overall statistics for two cases were similar, there was improvement in event flow simulation, peak flow values and peak flow timing for most basin-statistic combinations when SSURGO-based *a priori* parameters were used. When hydrographs were plotted in log-scale, it was observed that SSURGO-based case had better base flow simulations. For the small basin the better base flow performance can be seen throughout the year while for the large basin the better baseflow performance was seen mainly during the spring and the summer. The reasons for this need to be further studied. These preliminary results have shown the benefit of SSURGO soil data in distributed modeling. Use of these finer scale data can help to derive better *a priori* parameters. Improved, higher resolution, *a priori* parameters can provide a better starting point for calibration of distributed models and a more accurate characterization of parameters in small, ungauged basins.

REFERENCES

Abbott, M., Bathurst, J., Cunge, J., O'Connell, P., Rasmussen, J., 1986. An introduction to the European Hydrological System—Systeme Hydrologique European (SHE), 1, History and philosophy of a physically-based, distributed modeling system. Journal of Hydrology, 87, 45-59.

Anderson, R. Koren, V.I., Reed, S.M., 2005. Using SSURGO data to improve Sacramento model *a priori* parameter estimates. Submitted to MOPEX Special Issue of 'MODEL PARAMETER ESTIMATION' in Journal of hydrology.

Bell, V.A., Moore, R.J., 1998. A grid-based distributed flood forecasting model for use with weather radar data: Part 2. Case studies. Hydrology and Earth System Sciences 2(3), 283-298.

Carpenter, T.M., Gerorgakakos, K.P., 2004. Impacts of parametric and radar rainfall uncertainty on the ensemble simulations of a distributed hydrologic model. Journal of Hydrology, 298, 202-221.

Koren, V.I., Smith, M., Wang, D., Zhang, Z., 2000. Use of soil property data in the derivation of conceptual rainfall-runoff model parameters. Procededings of the 15th Conference on Hydrology, AMS, Long Beach, CA, 103-106.

Koren, V.I., Reed, S.M., Smith, M., Zhang, Z., Seo, D.J., 2004. Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. Journal of Hydrology, 291, 297-318.

McCuen, R.H., Snyder, W.M., 1975. A proposed index for comparing hydrographs. Water Resources Research, 11(6), 1021-1024.

Reed, S.M. 1998. Use of digital soil maps in a rainfall-runoff model. Doctoral Dissertation, The University of Texa at Austin, Austin, Texas.

Reed, S.M., Koren, V.I., Smith, M.B., Zhang, Z., Moreda, F., Seo, D.-J., DMIP participants, 2004. Overall distributed model intercomparison project results. Journal of Hydrlogy, 298, 27-60.

Smith, M., Seo, D.J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., Cong S. 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. Journal of Hydrology, 298, 4-26.