PROGRESS REPORT FOR PROJECT NUMBER NA04NWS4620012
Covers the Period 06/01/2006 TO 11/30/2006
Submitted August 13, 2007

TITLE: **Parameterization and Parameter Estimation of Distributed Models For Flash Flood and River Prediction**

PI: **Hoshin V. Gupta[1], Professor**
Co-PI: **Thorsten Wagener[2], Assistant Professor**
[1] **Department of Hydrology and Water Resources, The University of Arizona, Tucson, Arizona, 85721, Tel: 520-626-6974**
[2] **Now at Department of Civil and Environmental Engineering, Penn State Univ., University Park, PA 16802, Tel: 814-865-5673**

---

1. PROJECT OBJECTIVES AND ACCOMPLISHMENTS

1.1. Project Objectives

The shift from lumped to distributed models raises many important questions about the proper choice of model parameterization, including the desirable level of model complexity, while significantly increasing the complexity of the parameter estimation problem. The main objective of this project is to collaborate with and support the Hydrologic Modeling team at the NWS Office of Hydrology in the rapid development of an advanced version of the NWS-OH distributed hydrologic model, with particular attention to the issues of parameter estimation, appropriate model structure, supportable model complexity, and model evaluation, diagnosis and improvement. The unifying theme through this proposal is to focus on improving distributed watershed modeling through addressing issues of model parameterization (specification of model components), and estimation of the model parameters in both gauged and ungauged settings. The following tasks were listed under this contract:

1. Parameterization of semi-distributed and distributed hydrologic models within Hydrology Laboratory-Research Modeling System (HL-RMS) framework,
2. Distributed parameter estimation (automated and/or semi-automated) for the above (this work will build on our experience with lumped models, while introducing novel ideas such as regularization that are specifically tailored to distributed models),
3. A priori methods for parameter estimation in ungauged basins using direct inference from watershed properties and statistical regression analysis (existing work by NWS-HL staff will be extended and used to drive this important area of hydrologic modeling research forward).

This work extends our past collaborative work with the NWS by supporting the development of distributed modeling capabilities with particular attention given to ungauged and poorly gauged watersheds, consistent with the aims and future directions of the NWS. This research is being implemented in the context of the HL-RMS thereby maximizing technology transfer and ensuring that the work outcome is of direct value to the NWS.

1.2. Summary of Work Proposed for the First and Second Years

a. Implement HL-RMS at the University of Arizona as a modeling environment. Incorporate currently available calibration algorithms.
b. Investigate and implement a distributed parameter estimation algorithm based on the concept of regularization
c. Investigate the a priori parameter estimation problem using both bottom-up (incl. the development of relationships between the parameters of the NWS conceptual model components to soil and watershed characteristics) and top-down (regionalization) approaches.
d. Testing of various modifications of the HL-RMS including the comparison to semi-arid specific models.
e. Combining the research on a priori and distributed parameter estimation into a single procedure.
f. Testing the basic equations relating model parameters and watershed properties in a multi-watershed study. Complementing this approach with a statistical regionalisation approach using a minimum of 30 watersheds.
g. Implementing and testing various ensemble-forecasting schemes.
h. Technology transfer through (in person and telephone) meetings.
i. Implementing and testing the new tools for a priori and distributed parameter estimation into the HL-RMS.
j. Extending the Bayesian recursive scheme from lumped to distributed model structures.
k. Continue work on ensemble forecasting schemes.
l. Technology transfer through (in person and telephone) meetings.

1.3. Project Accomplishments During Progress Report Period (06/01/2006 – 11/30/2006)

Our activities during last year focused on: a) developing a strategy for diagnostic evaluation of HL-DHMS model, b) developing regularization relations for HL-DHMS model, c) developing a new approach to predictions in ungauged basins through the regionalization of constraints, d) investigating the sensitivity of the HL-DHMS.

a) A diagnostic model evaluation strategy is being developed to detect causes of HL-DHMS model performance inadequacies and to guide appropriate model parameters/structure adjustments.

The HL-DHMS model parameterized by Koren apriori parameterization scheme (Koren et al., 2000) is termed as “baseline model” and was used as a benchmark to test the performance improvements achieved by the model evaluation approach. The proposed model diagnostic strategy follows a hierarchy of timescales. At each timescale, measures are formulated to test the ability of the model to represent hydrological processes dominant at that timescale. Considering major functions of a watershed/hydrologic model, these processes can be grouped as those related to: a) water balance, b) partitioning of excess rainfall between fast and slow runoff, and c) flow timing. While model components/parameters related to water balance are tested with measures formulated at long timescales (annual, monthly), shorter timescales are used to test model components/parameters related to water partitioning and timing (daily, hourly). The diagnostic measures selected to summarize the water balance function of the model are percent flow bias (compared to observed flows) and percent cumulative flow bias. The latter measure is a stronger test of the water balance because it measures the maximum water balance deviation throughout the study time period. The partitioning between fast and slow processes in the watershed was summarized by the slope of the mid section of flow duration curve together with the percent bias in high flow segment of flow duration curve. One-at-a-time and random sampling-based perturbation analysis were performed to identify the HL-DHMS model parameters that have dominant control on the aforementioned watershed/model functions. The analysis indicated that water balance function is highly controlled by parameters related to evapotranspiration (UZTWM, LZTWM, PFREE) and flow partitioning function is highly controlled by the parameters affecting the percolation in the model (REXP, LZFSM, ZPERC, LZFPM, LZSK, UZFWM).

Parameter perturbations were performed by assuming that a priori model parameters adequately represent the spatial distribution of hydrological processes. Therefore parameter maps for each parameter was perturbed by using a single coefficient. In general the coefficient is used as a multiplier to the parameter maps and the feasible range of coefficient is chosen subjectively. However in this study a novel non-linear transformation was applied to the spatially distributed parameter values that maintains the parameter values within their range of feasible values, without the requirement for a subjectively selected threshold (see Yilmaz [2007] for details).

An additional measure that is sensitive to flow timing and shape is currently being developed. Current diagnostic strategy is focused primarily on overall model performance (at the outlet) and model deficiencies caused by incorrect spatial distribution of parameters were left for future work.

b) *Regularization relations were developed to constrain the high-dimensional parameter space of HL-DHMS for use within automatic calibration.*

The high dimensionality of the parameter search space can be solved by the introduction of additional information about the parameters. Regularization is a mathematical technique that utilizes additional information or constraints about the parameters to solve over-parameterized problems. In this research the information embodied in the 11 SAC-SMA apriori estimates, derived using the Koren approach (Koren et al. 2000), was utilized to develop the constraints (in the form of regularization relations).

The general procedure adopted to derive the regularization relations was to identify the physically observable watershed characteristics (elevation, slope and soil depth) or properties derived from physically observable characteristics of the watershed (curve numbers, curvature of the ground, topographic index and specific catchment area) that may, in theory, influence spatial distribution of parameters in a watershed. Then analyze the distribution of the apriori parameter with respect to the watershed characteristics and identify trends or relationship between them that could be expressed in the form of simple empirical relations. Finally, calibrate the parameters pertaining to these relations, instead of the original SAC-SMA parameters.

Using the procedure outlined above 11 such regularization relations were identified. The equations take the form of exponential, logarithmic or linear relations. These equations can be expressed in a generalized form as shown below:

$$\theta = \alpha_\theta \times [X]^{\beta_\theta} \pm \gamma_\theta \qquad (1)$$

where, θ is the regular SAC-SMA parameter α, β and γ are the coefficients of the equations which have to be calibrated and X is the observable watershed characteristic whose value can be easily calculated or measured.

The coefficients of the parameters were calibrated using the Multi-objective Shuffled Complex Evolution Metropolis (MOSCEM) algorithm (Vrugt et al. 2003). The model used for the calibration was the Distributed Hydrologic Model University of Arizona (DHM-UA), which was developed for this study. Calibration of these new sets of parameters, reduced the number of parameters to be calibrated from 858 (78 grids x 11 parameters/grid) to 33 (3 regularization equations x 11 parameters). Since calibration of all 33 parameters at the same time can still impose a significant computational burden, the calibration of the parameters was done in three steps. Firstly the intercepts of the equations (Gamma) were calibrated, following that the curvatures (beta), and finally the slopes (alpha) were calibrated. The calibration order was worked out after trying all the six combinations. The objective functions used were the Mean Squared Error (MSE) and the Log Mean Squared error (LMSE). The calibration of the parameters resulted in significant reduction in the function value (Fig. 1a) and improvement of the simulated hydrograph over the apriori model simulations (Fig. 1b).

c) *Developing a new approach to predictions in ungauged basins through the regionalization of constraints*.

Approaches to modeling the continuous hydrologic response of ungauged basins use observable physical characteristics of watersheds to either directly infer values for the parameters of hydrologic models, or to establish regression relationships between watershed structure and model parameters. Both these approaches still have widely discussed limitations, including impacts of model structural uncertainty. In this paper we introduce an alternative, model independent, approach to streamflow prediction in ungauged basins based on empirical evidence of relationships between watershed structure, climate and watershed response behavior. Instead of directly estimating values for model parameters, different hydrologic response behaviors of the watershed, quantified through model independent streamflow indices, are estimated and subsequently regionalized in an uncertainty framework. This results in expected ranges of streamflow indices in ungauged watersheds. A pilot study using 30 UK watersheds shows how this regionalized information can be used to constrain ensemble predictions of any model at ungauged sites. Dominant controlling characteristics were found to be climate (wetness index), watershed topography (slope), and hydrogeology. Main streamflow indices were high pulse count, runoff ratio, and the slope of the flow duration curve. This new approach provided sharp and reliable predictions of continuous streamflow at the ungauged sites tested. See Yadav et al. (2007) for details.

d) *Investigating the sensitivity of the HL-DHMS.*

This study component provides a step-wise analysis of the Hydrology Laboratory Distributed Hydrologic Modeling System (HL-DHMS). It evaluates model parameter sensitivities for annual, monthly, and event time periods with the intent of elucidating the key parameters impacting the distributed model's forecasts. This study demonstrates a methodology that balances the computational constraints posed by global sensitivity analysis with the need to fully characterize the HL-DHMS's sensitivities. The HL-DHMS's sensitivities were assessed for annual and monthly periods using distributed forcing and identical model parameters for all grid cells at 24-hour and 1-hour model time steps respectively for two case study watersheds within the Juniata River Basin in central Pennsylvania, USA. This study also provides detailed spatial analysis of the HL-DHMS's sensitivities for two flood events based on 1-hour model time steps selected to demonstrate how strongly the spatial heterogeneity of forcing influences the model's spatial sensitivities. Our verification analysis of the sensitivity analysis method demonstrates that the method provides robust sensitivity rankings and that these rankings can be used to significantly reduce the number of parameters that should be considered when calibrating the HL-DHMS. Overall, the sensitivity analysis results reveal that storage variation, spatial trends in forcing, cell-connectivity, and cell proximity to the gauged watershed outlet are the four primary factors that control the HL-DHMS's behavior. See Tang et al. (2007) for details.

2. SUMMARY OF RESEARCH AND EDUCATIONAL EXCHANGES

Scientific exchanges between UA/PSU researchers and NWS-HL personnel have taken place in the form of phone calls and e-mails. Victor Koren, Seann Reed and Zhengtao Cui of HL provided technical assistance for HL-DHMS model. Victor Koren and Mike Smith of HL provided feedback for the posters presented in scientific meetings.

3. PRESENTATIONS AND PUBLICATIONS

Tang, Y., Reed, P., Van Werkhoven, K. and Wagener, T. 2007. Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research*, 43, doi:10.1029/2006WR005813.

Tang, Y., Reed, P., Wagener, T. and Werkhoven, K. van 2006. Advancing watershed model identification and evaluation: A comparison of sensitivity analysis methods. AGU Fall Meeting, 11-15th December 2006, San Francisco, USA. (Poster)

Tang, Yong, Reed, P. M., Wagener, T., and van Werkhoven, K. 2006. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Science Discussions*, 3, 3333-3395.

Wagener, T., Yadav, M., and Gupta, H. 2006. Hydrologic ensemble predictions in ungauged basins. 2nd International Symposium on Quantitative Precipitation Forecasting and Hydrology, 4th-8th June 2006, Boulder, CO. (Invited Talk) (Oral)

Werkhoven, K. van, Wagener, T., Tang, Y. and Reed, P. 2006. Advancing watershed model identification and evaluation: A hydro-meteorological calibration strategy. AGU Fall Meeting, 11-15th December 2006, San Francisco, USA. (Poster)

Yadav, M, Wagener, T. and Gupta, H.V. 2006. Regionalization of dynamic watershed response behavior. *Eos Trans. AGU, 87*(36), Jt. Assem. Suppl., Abstract H23D-15. (Poster)

Yadav, M., Wagener, T. and Gupta, H.V. 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8), 1756-1774. doi:10.1016/j.advwatres.2007.01.005.

Yilmaz, K., H.V. Gupta and T. Wagener, Diagnostic Evaluation of a Distributed Watershed Modeling Approach, Presented at the Fall Meeting of the American Geophysical Union, San Francisco, CA, USA, December 11–15, 2006

Yilmaz, K., H.V. Gupta and T. Wagener, Diagnostic Evaluation of a Distributed Watershed Modeling Approach, Presented at the PUB USA Meeting, Corvallis, OR, USA, Oct 16-19, 2006.

Yilmaz, K., K., 2007. Towards improved modeling for hydrologic predictions in poorly gauged basins, Ph.D. Dissertation, Univ. of Arizona.

4. FUTURE WORK

In the light of the experience we have gathered from the above analysis, the following studies will be performed:

a) Measures that are powerful in diagnosing HL-DHMS model inadequacies will be researched.

b) Parallel processing techniques will be investigated for faster model runs required by the optimization algorithms.

c) Extending the regionalization approach to utilize multi-objective optimization (rather than Monte Carlo analysis), and implementation of regionalization approach in US watersheds.

d) Connecting sensitivity analysis and model calibration in combined procedure.

5. SUMMARY OF BENEFITS AND PROBLEMS ENCOUNTERED

Benefits that have been experienced during the last year

a. UA/PSU researchers are becoming familiar with the HL-DHMS distributed hydrologic model developed by NWS-HL in an effort to contribute and share new ideas. Students are becoming familiar with NWS software, methods and procedures.

b. Project provided research subjects for two master theses (Prafulla Pokhrel, Univ. of Arizona & Maitreya Yadav, PSU) and in part a Ph.D. dissertation (Koray K. Yilmaz, Univ. of Arizona). Sub-projects have been derived from this project for elements of the PhD dissertations by Yong Tang and Katie van Werkhoven (PSU).

c. Fruitful communication between UA/PSU researchers and HL personnel has continued.

Problems encountered

a. No significant problems were encountered during the last 6 months.

6. REFERENCES

Koren, V. I., Smith, M., Wang, D. and Zhang, Z. (2000). Use of soil property data in the derivation of conceptual rainfall-runoff model parameters. 15th Conference on Hydrology, Long Beach, American Meteorological Society, Paper 2.16, USA.

Vrugt J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, S. Sorooshian, 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39 (8), 1214, doi:10.1029/2002WR001746.

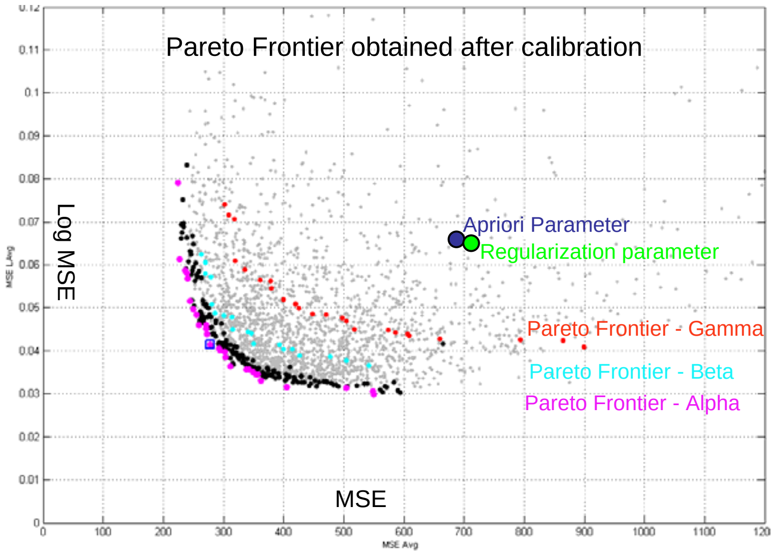

Figure 1a. Pareto frontier obtained after successive calibration of regularization parameters (gamma followed by beta and finally alpha). Function values obtained by model simulation due to apriori parameter estimate and the apriori regularization parameters are also shown.

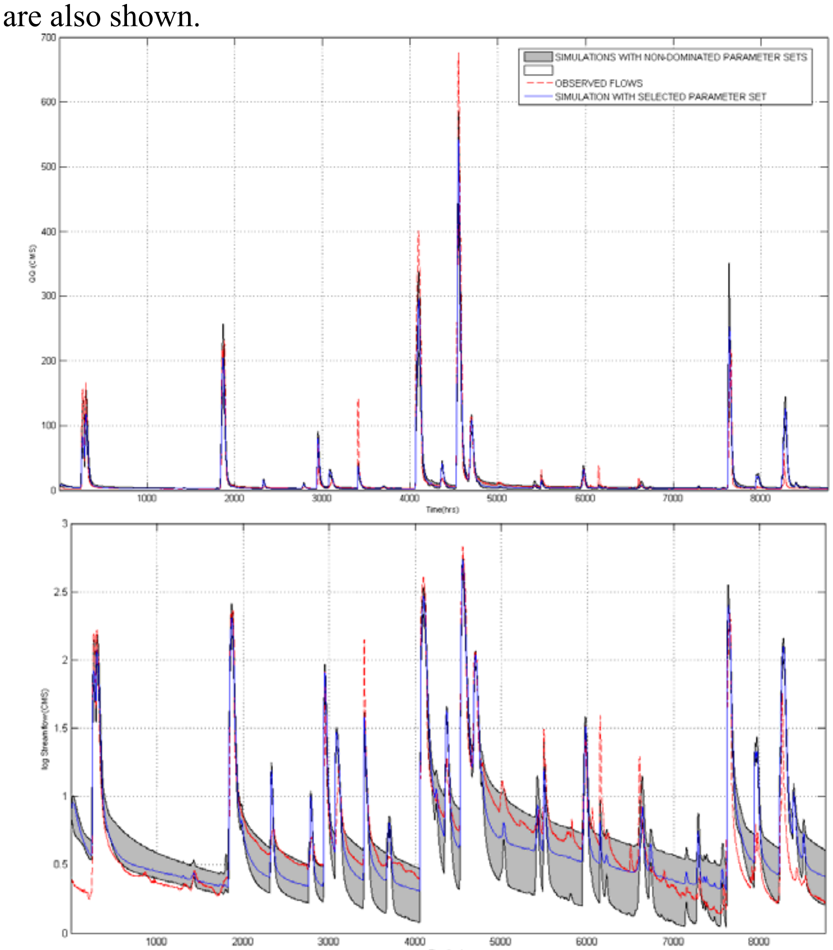

Figure 1b. Observed and simulated hydrograph after calibration of the parameters (Top) linear Y axis, (Bottom) log Y axis.

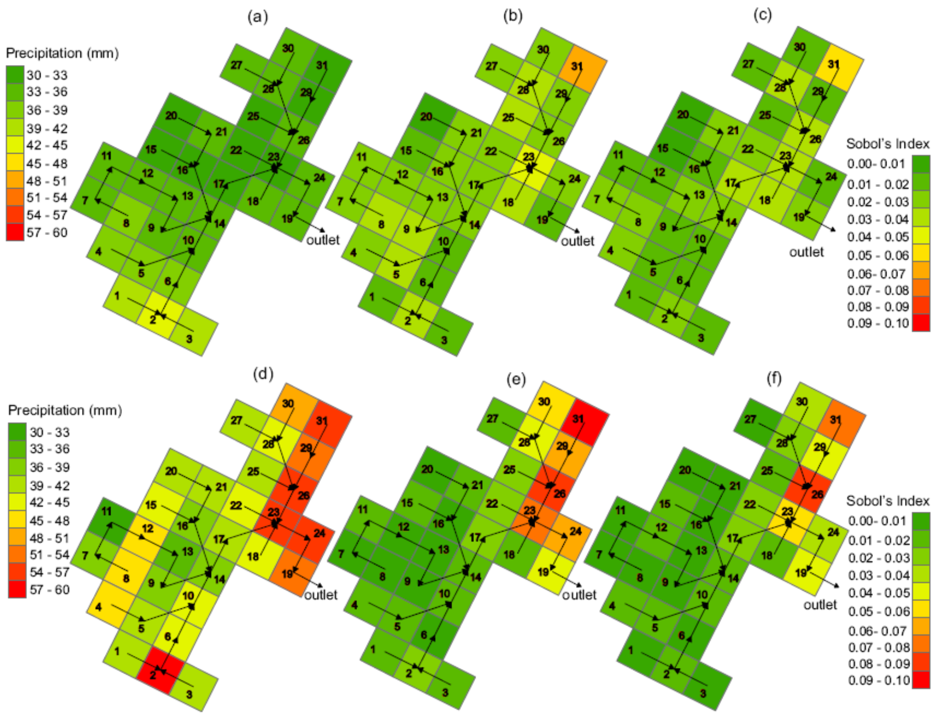

Figure 2. Spatial distribution of the total event precipitation and cell-level sensitivities for the SPKP1 watershed. The Sobol first order indices shown are the % of the model's ensemble variance contributed by a single parameter, while total order indices include all interactions involving a parameter. The May 2002 event is represented by (a) its spatial precipitation distribution, (b) the first order Sobol's indices for each model cell, and (c) the cell level interactions. The September 2003 event is represented by (d) its spatial precipitation distribution, (e) the first order Sobol's indices for each model cell, and (f) the cell level interactions (cell level interactions were computed as difference between each cell's total order and first order indices). The cell-level Sobol's indices were computed by summing over all of individual parameter indices analyzed in each cell. The arrows in the cells designate surface flow directions (from Tang et al., 2007).