Weather Radar Information and Distributed Hydrological Modelling (Proceedings of symposium HS03 held during IUGG2003 at Sapporo, July 2003). IAHS Publ. no. 282, 2003.

A methodology for assessing the utility of distributed model forecast applications in an operational environment

KONSTANTINE P. GEORGAKAKOS & THERESA M. CARPENTER

Hydrological Research Center, 12780 High Bluff Drive, Suite 250, San Diego, California 92130, USA

KGeorgakakos@hrc-lab.org

Abstract Distributed models have long been viewed as beneficial for hydrological forecasting. However, errors in the precipitation data and model parameters may diminish any gains in prediction accuracy realized by accounting for the spatial variations in precipitation and parameters. Central to these considerations is the resolution of the issue of balancing model complexity with parameter and input estimation uncertainty. The present paper exemplifies a probabilistic methodology to address this issue. A distributed model with components used in the operational forecasting of streamflow in the USA is utilized to produce flow simulations for a catchment in the south-central USA. Models of uncertainty for radar-rainfall data and model parameters are developed. Ensemble flow simulations are produced for a number of input and parameter uncertainty scenarios and for a high and low spatial resolution model configuration. Kolmogorov-Smirnov testing is then performed to assess whether the peak-flow simulations of the low- and high-resolution models can be distinguished with high confidence. Assessments of the implications regarding performance under operational forecast conditions are given.

Key words distributed hydrological modelling; ensemble flow prediction; parameter estimation; radar rainfall uncertainty

INTRODUCTION

The research question this paper addresses is: are ensemble flood flow simulations from models with (a) spatially aggregate parameters and (b) spatially distributed parameters distinguishable on scales of the order of 10^3 km^2 under present-day operational parametric and radar-rainfall uncertainty? Consideration of flow simulations from models is a necessary first step prior to the consideration of lumped and distributed hydrological model operational forecasts that incorporate substantial precipitation forecast uncertainty. Premises of the analysis are:

(a) parametric and radar-rainfall uncertainties are significant under operational conditions;
(b) ensemble flow simulation provides a more complete and useful representation of model response than a single nominal simulation;
(c) the runoff generation component is the same for lumped and distributed models;
(d) the scale of order 10^3 km^2 is significant for operational flow forecasting.

The focus of the paper is to exemplify the approach used to answer the research question. This approach is applicable to any type of distributed hydrological model. In the examples to follow, we use the Sacramento soil moisture accounting model for runoff generation in soil columns for the distributed model and the lumped model, and a kinematic channel routing component for streamflow simulation and time distribution of generated runoff. The distributed model used is a sub-catchment based (rather than grid based) model with spatially distributed parameters and input. The distributed model formulation, validation and sensitivity analyses with a variety of data sets from various basins in the USA are described in Carpenter *et al.* (2001) and Carpenter & Georgakakos (2002).

METHODOLOGY

The target of the methodology is to distinguish the lumped from the distributed model simulations when these simulations are in the form of ensembles that incorporate the uncertainty due to erroneous model parameter values and noisy radar rainfall input.

The methodology consists of the following consecutive steps:

(a) Using long historical records, calibrate lumped and distributed models for the basin of interest.

(b) Develop generic models of uncertainty for parameters based on attributes of spatial data used and on calibration error statistics.

(c) Develop generic models of uncertainty for radar rainfall estimates at the radar-bin scale and upscale uncertainty models to the scale resolved by the distributed and lumped models.

(d) For each model type (lumped *vs* distributed) sample from the respective parameter uncertainty distribution and develop ensemble simulations for several significant events in the test catchment for the catchment outlet (area of order 10^3 km^2).

(e) For each event, assess whether the ensemble of simulated flows by the lumped model at the time of the observed flow peak belongs to the same distribution as the ensemble generated for the same time and spatial scale by the distributed model. A Kolmogorov-Smirnov statistical test is used to compare the distributions.

(f) Repeat (d) and (e), sampling from both the parametric and radar-rainfall uncertainty developed for each model type in (b) and (c).

TEST BASIN AND MODEL PERFORMANCE

The 1232-km^2 Blue River basin with outlet near Blue, Oklahoma, is used as the test basin for this analysis. The basin is one of four basins used at the US Distributed Modeling Intercomparison Project (DMIP) organized by the US National Weather Service. It is in a semiarid environment, receiving much of its warm season rainfall from convective storms. The spatial data used in distributed modelling for the basin are: USGS Digital Elevation (90-m resolution); USGS LULC (200-m resolution); STATGO soils data; EPA Reach File 3 (RF3). The distributed model subdivides the basin into 21 sub-catchments with an average size of 59 km^2. Estimates of parameters

Table 1 Performance of lumped and distributed model for historical period.

	Distributed model	Lumped model
Calibration period	8 May 1993–31 May 1999	8 May 1993–31 May 1999
Hourly cross-correlation	0.86	0.79
Hourly flow bias	1.40%	5.80%

and input were obtained for each of these sub-catchments. The lumped model considers the basin as an aggregate unit and applies parameters and input to the entire area.

The hourly historical hydrometeorological data used consisted of: discharge (USGS site 07332500); NEXRAD (WSR-88D) Stage III Precipitation; and energy forcing provided by DMIP databases (air temperature, air pressure, solar radiation, relative humidity). The historical data period was of a 6-year duration. Table 1 indicates the performance of the lumped and distributed models for the historical calibration period. The performance of the two models is comparable and it generally suggests that both models reproduce the historical flows well (less than 10% bias with an explanation of variance of hourly observed flows that is in the range 63–74%). There is slightly better performance shown for the distributed model. Georgakakos & Carpenter (2002) show simulations of individual events.

MODELS FOR PARAMETER UNCERTAINTY

The generic formulation for parametric uncertainty for both the lumped and the distributed model is:

$$\alpha = \mu_\alpha + \varepsilon_\alpha$$

where μ_α is the mean value of parameter α, and ε_α is a uniformly distributed error in range $[-\alpha_L, \alpha_L]$.

For the distributed model, spatially uniform parameters were estimated with an adjustment made for the upper soil parameters to account for the different spatial scale of the various sub-catchments. The median of the range of the STATSGO database available water content and permeability was then used to distribute the parameter values in space, keeping the basin spatial mean equal to the uniform parameter estimate obtained earlier. In all cases, the ranges of the STATSGO database estimates were used to define the limit α_L of the uniform distribution of the parametric error.

The constant and the exponent parameters of the soil model percolation function were adjusted in space from their uniform estimates using the distribution of texture and a table of association of texture and percolation parameters developed by the Staff of the Hydrology Laboratory, US National Weather Service (reproduced here for convenience in Table 2). The ranges indicated in the table were used to estimate α_L in each case. In the soil model the percolation rate normalized by the maximum possible baseflow, p_n, is given by:

$$p_n = 1 + C\,x_n^{\,M}$$

Table 2 Association of soil texture and percolation parameters.

	Percolation exponent, M	Percolation constant, C
Sand	1.4–1.8	5–20
Silt	1.8–2.5	20–75
Clay	2.5–3.5	75–200

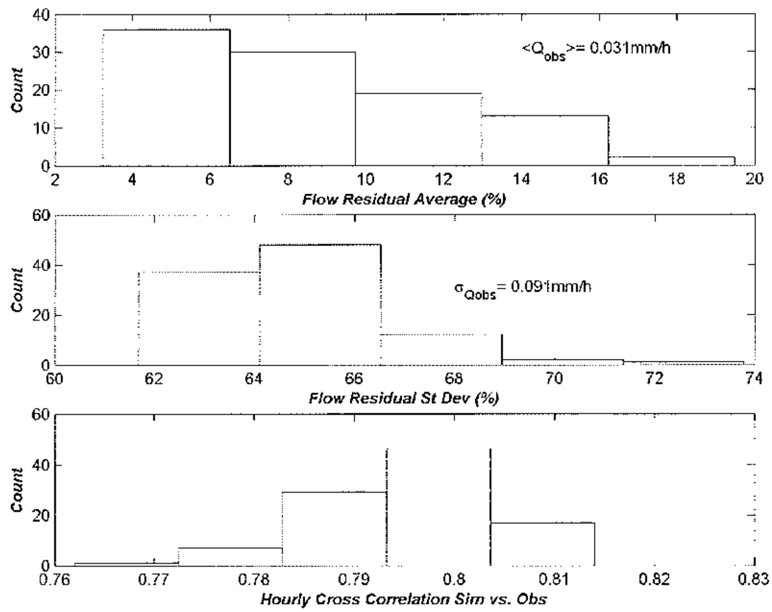

Fig. 1 Performance indices used in calibration of lumped model with Blue River data.

where C and M are the percolation parameters and x_n is the lower soil zone storage available for percolation normalized by lower soil zone capacity (sometimes called the lower zone deficiency ratio).

For the lumped model, the range in parameter error $[-a_L, a_L]$ is determined by degree-of-belief estimates of the hydrologist who calibrated the lumped model for the Blue River basin. A ± 25% error range was used for the uncertain parameters of the upper soil zone. The diagrams of Fig. 1 show the effect on various model performance indices of using this uncertainty range during the calibration period. It is noted that the model performance indices have ranges that are within the range of model performance considered adequate when calibrating with multiple objectives, so the 25% parametric uncertainty estimate appears reasonable for this basin and this model.

MODELS FOR RADAR RAINFALL UNCERTAINTY

The basis of the hourly mean areal rainfall estimates used for sub-catchments in this study is radar rainfall data, Stage III, from the NEXRAD (WSR-88D) radar. The radar rainfall data are available for radar bins of approximate size 3.5×3.5 km^2. For error estimation of the mean areal rainfall for each sub-catchment of the distributed model and for the entire basin for the lumped model, an error model was postulated for radar-bin rainfall estimates. The parameter of the radar-bin error model represents an error half-range that is a function of the radar rainfall estimate magnitude for the radar bin. The postulated model for the relative error half-range (E/R_m: error half-range E divided by observed rainfall R_m) is as follows: largest relative errors correspond to lighter rainfall amounts with a linear decrease of the relative error half-range with increasing observed rainfall amount up to 25 mm h^{-1}. For rainfall amounts greater than 25 mm h^{-1}, the relative error half-range remains constant and equal to 0.5. The distribution of the radar-bin error is assumed uniform within the range limits [$R_m - E$, $R_m + E$], and no space correlation of the radar-bin error is assumed in this initial phase of the study. The latter postulate may result in underestimation of the rainfall error variance for sub-catchments that contain several radar bins, if in fact significant spatial correlation in radar-bin rainfall error exists.

Using Monte Carlo sampling from the radar-bin error model for each of the events, parameters of error functions were developed for each sub-catchment of the distributed model as well as for the entire basin area as used by the lumped model. The error functions assumed uniform mean areal rainfall distribution for each sub-catchment, with a distribution half-range that is given relative to the mean areal rainfall of the sub-catchment in Fig. 2. Dependence of the relative half-range of the error for sub-catchments is still linear for mean areal rainfall rates less than 25 mm h^{-1}, but the slope and intercept of the line decrease with increasing sub-catchment area. The lowest fitted line in Fig. 2 corresponds to the entire Blue River basin area. The errors for each sub-catchment of the distributed model and for the entire basin for the lumped model were

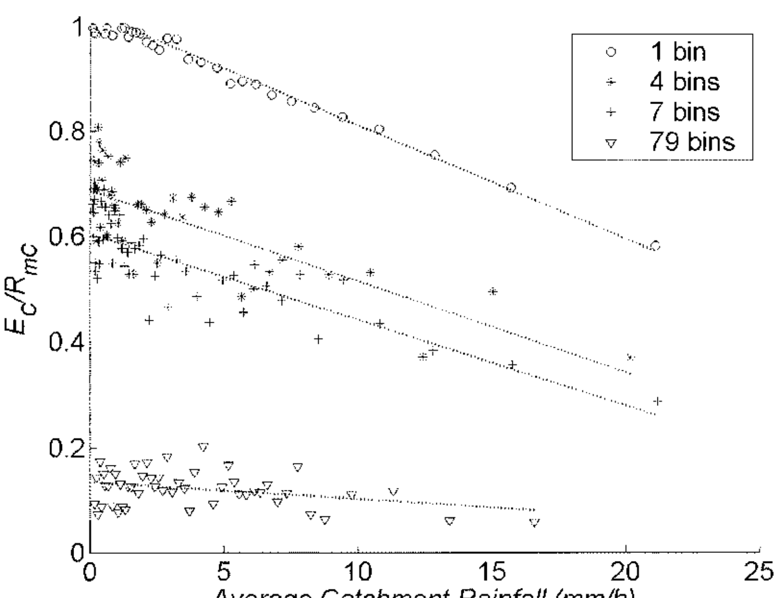

Fig. 2 Sub-catchment relative error half-range E_c/R_{mc} as a function of sub-catchment rainfall R_{mc}. The legend shows the number of bins for the sub-catchments of the Blue River.

assumed uniformly distributed in the range $[R_{mc} - E_c, R_{mc} + E_c]$ with R_{mc} representing the sub-catchment rainfall mean (obtained by averaging the enclosed radar-bin observations), and E_c representing the error half-range.

RESULTS OF APPLICATION

A total of 25 events were selected from the historical database that spanned the period from December 1993 through March 1999. The methodology was applied as described above for each of these events. For each event the models to run in a Monte Carlo fashion (sampling from parametric and input uncertainty distributions) for 2 months prior to the event to establish a stable distribution of the initial soil moisture field at the beginning of the event and then the Kolmogorov-Smirnov test was applied during the event period. The Kolmogorov-Smirnov test statistic D_o is determined by:

$$D_o = \max \left| P_{\mathrm{Distr}}(Q_p) - P_{\mathrm{Lump}}(Q_p) \right|$$

$$\mathrm{Prob}\,(D > D_o) = F(N_e, D_o)$$

$$N_e = \frac{N_D N_L}{N_D + N_L}$$

with $N_D = N_L = 100$ and with the following nomenclature:
N_D, number of distributed model ensemble members; N_L, number of lumped model ensemble members; Q_p, simulated discharge at the time of the observed peak flow for each event; P_{Distr}, cummulative frequency distribution of distributed model ensemble flows at peak time; P_{Lump}, cummulative frequency distribution of lumped model ensemble flows at peak time.

The Kolmogorov-Smirnov test does not depend on the type of distribution but it does require that the distributions involved be continuous. It is most effective for measuring shifts in distributions (most sensitive around the median value of distributions). The null hypothesis in this application is that the two sets of flow samples, resulting from the distributed and the lumped models at the time of the observed flow peak are drawn from the same distribution (and therefore are indistinguishable). The significance level probability for accepting the null hypothesis is 1%. For parametric uncertainty only, the test indicates that the distributions at the time of the observed peak are different at the 1% significance level. That is, $Prob(D > D_o)$ is less than 1% for all 25 events. When radar rainfall input uncertainty is added to the parametric uncertainty, then in all but one event $Prob(D > D_o)$ is less than 1%. These results imply that for this catchment and intercomparison scale, the distributed and lumped model ensemble flow simulations of high flows are statistically different with high confidence.

Given that the distributed and lumped model simulations of event peaks are statistically different for virtually all events under parametric and input uncertainty, the question arises as to which model has better performance. This question requires the definition of a suitable assessment metric that explicitly takes into consideration that the flow "simulation" is in the form of an ensemble due to parametric and input uncertainty. The sample frequency that the simulated flow is within ± 20% of the

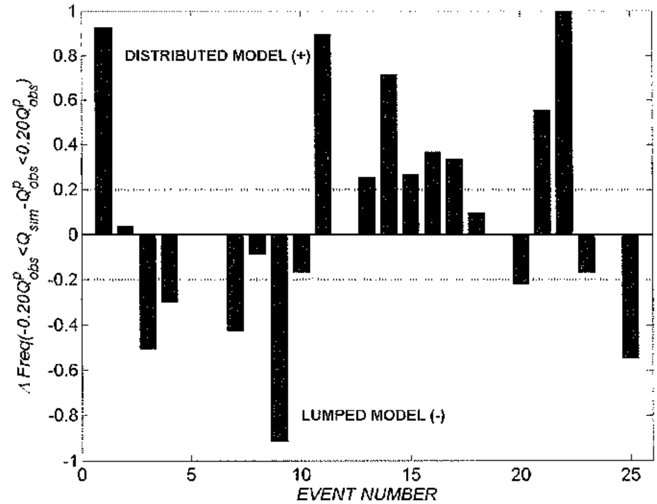

Fig. 3 Differences in sample frequencies between distributed and lumped model of the event that the simulated ensemble flows at the time of the observed peak flow is within ±20% of the observed peak flow.

observed peak flow is used as our assessment metric for this analysis. Figure 3 shows the differences in these sample frequencies between the distributed and the lumped model. Positive numbers indicate higher sample frequency for the distributed model, which is interpreted to mean better performance. Negative numbers indicate better performance for the lumped model. There are 100 samples in each ensemble of simulated flows for each of 25 events. Because the assessment metric depends on the number of samples, only values outside the interval [− 0.2, + 0.2] are considered indicative of better performance. Values of the assessment metric within the interval were considered uncertain. The results in Fig. 3 show that in 44% of the events the distributed model had a better performance, in 36% of the events the lumped model had a better performance and in 20% of the events best performance could not be determined with the given assessment metric. Study of the dependence of the model performance with respect to magnitude of the observed peak flow indicated that the distributed model is clearly better in the medium peak-flow range. For the highest peak flows there is indication that the lumped model performs better, but more data is necessary for a more definitive statement in this peak-flow range.

CONCLUSIONS AND RECOMMENDATIONS

A new methodology is described and applied in this paper for determining whether distributed model performance is different from lumped model performance under parametric and input uncertainty. A probabilistic metric is also proposed to assess which model has better performance, once it is established that the model simulations are distinguishable with high confidence. The methodology was applied to the 1232 km^2 Blue River basin with outlet near Blue, Oklahoma. The results of the application show that the ensemble flow simulations from distributed and lumped models are statistically different with a high degree of confidence, and that for medium flow events, the distributed model has a better performance than the lumped model.

It would be useful for future research to improve the input error models used herein. Allowing spatio-temporal dependence on the radar-bin error model is an especially important research area. If such dependence is high, then it is likely that in terms of being able to distinguish lumped from distributed model simulations in the present operational environment the results presented in this paper are rather optimistic. The use of various levels of spatial aggregation for the distributed hydrological model and application of the proposed methodology is another area of future research, which would help determine the spatial scale of model discretization that is most suitable for use with operational estimates of parameters and input. Lastly, application of the methodology to other basins and different types of distributed models should also be made.

Acknowledgements This research was supported by the Office of Hydrological Development (OHD) of the US National Weather Service, National Atmospheric and Oceanic Administration under Award no. NA07WH0368. The authors wish to thank Drs Michael Smith and Victor Koren of OHD for several useful comments and suggestions. The ideas and opinions expressed herein are those of the authors and do not necessarily reflect those of NOAA and of its sub-agencies.

REFERENCES

Carpenter, T. M. & Georgakakos, K. P. (2002) Distributed model flow sensitivity to uncertainty in radar-rainfall input. In: *Information from Weather Radars and Distributed Hydrological Modelling* (ed. by Y. Tachikawa *et al.*) (Proc. Symp., Sapporo HS03, July 2003), 114–120. IAHS Publ. no. 282 (this volume).

Carpenter, T. M., Georgakakos, K. P. & Sperfslage, J. A. (2001) On the parametric and NEXRAD-radar sensitivities of a distributed hydrological model suitable for operational use. *J. Hydrol.* **253**, 169–193.

Georgakakos, K. P. & Carpenter, T. M. (2002) Probabilistic analysis of model complexity versus parametric and input uncertainty in operational streamflow forecasting. HRC Tech. Note no. 20. Hydrological Research Center, San Diego, California, USA.