#1030

WATER RESOURCES RESEARCH, VOL. 30, NO. 3, PAGES 851–854, MARCH 1994

Comment on "Physically based hydrologic modeling, 2, Is the concept realistic?" by R. B. Grayson, I. D. Moore, and T. A. McMahon

R. E. Smith

Agricultural Research Service, U.S. Department of Agriculture, Fort Collins, Colorado

D. R. Goodrich, D. A. Woolhiser,[1] and J. R. Simanton

Southwest Watershed Research Center, Agricultural Research Service, U.S. Department of Agriculture
Tucson, Arizona

Nature has not made it a priority to make it easy for us to discover its laws

Albert Einstein (1901)

In the first of two companion papers, *Grayson et al.* [1992a] described the application of a relatively detailed "terrain-based" model to two quite different catchments. One is in Australia, characterized by both surface and subsurface runoff generation mechanisms, and the other is a catchment in southern Arizona with exclusively surface runoff processes. The second paper [*Grayson et al.*, 1992b], which is the subject of this comment, is an extensive and critical discussion of results of the first paper. It includes a rather pessimistic assessment of the results of this model exercise, but consists primarily of opinions concerning the use of "physically based" models "intended for use as hydrologic components of sediment and nutrient transport models" (p. 2659). Hereafter we refer to these two papers as paper 1 and paper 2.

It is our view that the authors' assessment of their results is overly pessimistic, and many of their rather sweeping generalizations therefrom are unwarranted. While some good points were made by the authors (they have mostly been made before), it is hard not to conclude after reading their philosophizing that the authors came to their exercise of detailed "physically based" modeling with unrealistic expectations.

Much similar philosophy has appeared recently on physically based model applicability [*Beven*, 1989; *Dunne*, 1982; *Klemes*, 1988; *Loague and Freeze*, 1985], and most of that is referenced in the subject manuscript. The work by *Loague and Freeze* [1985] is a classic example of apples versus oranges, in which calibrated empirical models were indicated to be apparently better than uncalibrated physically based models, and has often been cited to cast doubt on physically based models. In reading all the philosophical criticisms of the "failures" and limitations of "physically based" models, it seems clear that some have indeed expected such models to be exact mimics of nature. But one also wonders whether such expectations have resided more in model developers or in model users, and one wonders how many model users have shared such unrealistic expectations?

[1]Retired to Fort Collins, Colorado.

Paper number 93WR03184.

Grayson et al. feel such models have been oversold, but it should be emphasized in that regard that more conceptual or even empirical models have been similarly abused at least as much, if not more. In the United States, in our experience, it is indeed the conceptual and parametric models which represent by far the more egregious examples of overselling. In that light it makes little sense to devote the large part of a manuscript to the perceived overselling of physically based models. Furthermore, the authors have not presented a direct example of the type of overselling or false claims for physically based models against which they appear to argue.

There is evidence in paper 2 of presumptions on the part of the authors which are subjective and color their interpretation of results: "the theoretical rigor of some models is impressive . . . and implies a degree of accuracy that may not exist" (p. 2662). We feel this is a mistaken implication of the authors; theoretical rigor itself makes no promise of accuracy in arbitrary application. It is the responsibility of those applying any theory to ascertain the appropriateness of the assumptions of the theory to the natural condition at hand. Further, they stated, "Process-based models are not the panacea they were once thought to be . . . " (p. 2663). Who thinks of any model as a panacea? Or they stated, "There is a certain arrogance associated with 'physically based' models regarding their superiority over lumped-parameter or empirical models . . . " (p. 2661). These statements may indicate that either the authors' own expectations of such models were inflated, or else they made such characterizations in the form of a straw man, as it were, for their arguments against the use of physically based models for prediction or management purposes and their pessimistic assessment of their results in paper 1.

Some Important Definitions

Before a discussion of hydrologic models it is appropriate to try to define terms. We feel there is a confusion in use of the term "model" by Grayson et al. One needs to distinguish between a fundamental hydrodynamic model/concept (such as the kinematic or diffusive wave equation) based on physical principles, such as momentum and mass conservation, and a computer model such as THALES, in which such physically based concepts are incorporated, but many other assumptions and approximations are also introduced. Both are "models." The first, hereafter denoted type 1 model, is a mathematical abstraction of observed natural behavior, admittedly including simplifications, but verifiable and re-

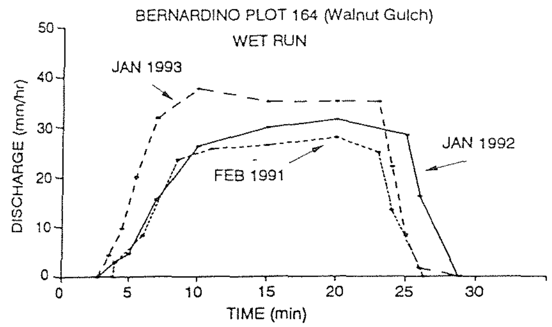

Figure 1. Response hydrographs for three repetitions of a sprinkling plot experiment with very similar initial conditions. The different response of the January 1993 event can hardly be explained by any measured parameters.

peatable in appropriate controlled experiments. The second, type 2, is quite distinct. While including the process or type 1 model such as the kinematic wave (for example), such a type 2 model applies a type 1 model to a complex part of nature with numerical, topographic, and computer logical assumptions along with ever-possible coding mistakes.

What Is Realistic and What Is Model Failure?

Having distinguished between type 1 models and type 2 models, one can then distinguish between type 1 model failure and type 2 model failure. (Perhaps the meaning of model "failure" also needs to be discussed, but we will omit that here.) The dissatisfaction of Grayson et al. with their ability to match measured values and model values can have a number of causes, but their results do not indict "physically based" models in general, do not necessarily indict their model (THALES), and certainly do not indict type 1 models such as kinematic surface water hydraulics or Dupuit-Forcheimer groundwater behavior. This brings up another point of definition. The title includes the word "realistic." We argue here that perhaps the authors' concepts of "physically based" models and how well they should be expected to perform are both unrealistic. First of all, we suggest certain elements of a "realistic" standard for comparison of type 2 model results to measurement. As *Hillel* [1986] has stated, a computer model should not be expected to simulate a physical event any more precisely than the repeatability of accurate measurements from a controlled experiment. An example of this is the "realistic" scale experiments performed by *Wu et al.* [1982]. Differences in peak discharge for replicate rainfall events on an impervious surface were about 4%. Repeatable responses of a natural system may be even more difficult to obtain, as found in a rainfall simulation study at the Walnut Gulch watershed (the experimental design is given by *Simanton et al.* [1991]). Three rainfall simulations of nearly the same application rate and amount on the same natural plot with very similar antecedent conditions produced very different hydrographs (Figure 1 and Table 1). The point is that a type 2 model should not be asked to match field measurements any better than the field measurements can match themselves (i.e., the best model of a system is the system itself). This seems to be fundamental to their interpretations of results in paper 1 [*Grayson et al.*, 1992a].

Without elaboration, Grayson et al. stated that "the mathematics of the hydrological processes and their boundary conditions are poorly defined and/or poorly used, so the solutions are not unique" (p. 2660). (What do they mean by "unique"?) It appears to us that the mathematics of the type 1 models are on the contrary quite well defined and their boundary conditions quite clear. Perhaps they were referring to statistical uniqueness, insofar as a comparison of model to data could give a similar statistical measure of fit with more than one set of parameters. Perhaps they were discussing type 2 models, and they were perhaps referring to heterogeneity. We discuss this topic below. Much of the disappointment evidenced in their discussion should be traced to the level of a priori expectations. The recognition of the limited conclusions to be drawn from their experience, as well as a clear distinction between type 1 models and type 2 models, was, unfortunately, not made in their discussion. The jump to generalizations is quite reminiscent of that of *Beven* [1989], whose criticisms were apparently instigated by disagreement with several applications of the SHE model by its developers. Applying a model at either an invalid scale or to an invalid physical situation does not invalidate either a type 2 model or the type 1 models therein, and this misuse does not invalidate such models in general.

Tackling Heterogeneity in Nature

Possibly the most significant distinction that needs to be made in response to the Grayson et al. problem is between type 1 model failure and natural heterogeneity. The THALES model application to Wagga Wagga assumed that soil hydraulic properties are stepwise uniform, based on soil surveys, which is traditionally considered quite reasonable. They did not attempt to deal with small-scale heterogeneity in subsurface soils or of surface soil depths or deal with spatial heterogeneity of rainfall input [*Goodrich and Woolhiser*, this issue]. It is well established that small-scale, more or less random heterogeneity characterizes soils wherever statistically large samples have been measured. *Grayson et al.* [1992a] repeatedly pointed to assumptions made in their modeling strategy and also often acknowledged the fact of heterogeneity. It is not appropriate to conclude that the type 1 model has failed because its parameters exhibit natural spatial heterogeneity. Neither is it appropriate to conclude that a parameter has lost its physical significance when a single value has been unsuccessfully assumed for a heterogeneous region. By contrast, one does not see manuscripts by groundwater hydrologists suggesting that Darcy's law is invalid for larger scales, simply because it has been found that saturated hydraulic conductivity varies considerably in

Table 1. Conditions and Response of Natural Plot 164

	Feb. 1991	Jan. 1992	Jan. 1993
Soil moisture, %	NA	11.1	11.1
Precipitation, mm	23.3	25.6	23.9
Precipitation rate, mm/h	61.0	61.6	62.3
Runoff, mm	7.8	9.6	11.4
Runoff rate, mm/h	27.9	31.5	37.7
Bare soil, %	40.8	40.4	43.5
Random roughness, mm	10.0	14.8	11.5

nature. Rather, groundwater hydrologists are working to develop ways to treat such heterogeneities.

In the context of natural heterogeneity we should also ask what is "realistic" to expect of a physically based model. What is a better model of an experimental plot on a "uniform" soil than a neighboring plot? While we expect some natural random variability, is it realistic to expect a type 2 model, no matter how "physically based," to outperform nature? A revealing set of data on this point is obtained from *Hjelmfelt and Burwell* [1984]. For rainfall events over one season ranging from 6 to 96 mm, the coefficient of variation of runoff volumes among forty 0.01-ha plots ranged from 0.071 (a moderate event) to 1.09 (another event in the middle range). While in general those plots with lower total runoff tended toward low runoff in all events (and vice versa), there was considerable temporal scatter in behavior, confounding any physically based model, which would presumably be consistent (or, for that matter, any simpler model as well). Why should we ask more of a type 2 model than we can expect from nature?

Remarkably, the authors seem to expect (p. 2662) that spatial heterogeneity should result in catchment behavior that is not "well behaved." Why, indeed, should stationary spatial variation result in a system that is not well behaved? Here we are assuming that most users of the language would distinguish between "well behaved" and spatially uniform. In the application of a similar type 2 model to the Lucky Hills data, *Goodrich* [1990] showed that there are some strategies [*Woolhiser and Goodrich*, 1988] for applying type 1 models to heterogeneous areas, and that indeed the use of those strategies produced good simulation of measured runoff over quite a range of storms, when use of spatial averages of parameters would not.

Natural heterogeneities provide perhaps the most important current challenge in hydrologic research. Further, it has been demonstrated that there may not exist a useful average single type 1 model parameter that allows assumption of uniform equivalence on a scale containing significant heterogeneities [*Woolhiser and Goodrich*, 1988; *Binley et al.*, 1989; *Smith et al.*, 1990]. This is certainly true for nonlinear processes. That there is no single average value of, say, Manning roughness for a whole watershed, however, does not imply that the Manning roughness relation is inapplicable. We suggest, rather, that the Manning roughness relation or any type 1 model must be applied in explicit acknowledgement of heterogeneity, without the implicit a priori assumption that an effective parameter exists in heterogeneous conditions. Moreover, Grayson et al. argued (p. 2665) that the challenges of heterogeneity imply that more simple type 2 models are better for "management" (presumably predictive) purposes. This argument has been heard before. They have not shown (nor has anyone, to our knowledge) that heterogeneous conditions can be handled as well or better by a conceptual or parametric type 2 model. Failing that demonstration, it is misleading to make such a suggestion.

Grayson et al. stated (p. 2662),

> The most important conclusion that can be made from those simulations is that the underlying assumptions relating to representation of the surface flow have as large an effect on the flow characteristics as do the parameter values.

They go on to assert that models incorrectly assume "broad sheet flow" for surface runoff. This is another mistaken assumption. One can show, as in KINEROS [*Wu et al.*, 1982; *Woolhiser et al.*, 1990], that the depth term in such flow equations can be quite properly taken as the mean or effective flow depth, even though the point depth may be quite variable across the flow path, without loss of value of the equations. Thus uniform, sheet-type flow is not necessarily assumed (see also *Goodrich* [1992]).

Moreover, we disagree that the above statement is the "most important conclusion to be drawn" from this exercise. What should be concluded from the work of Grayson et al. is that one cannot in advance presume what is the most important source of heterogeneities. Their model uses a relatively detailed representation of the surface topography. This may have been the reason for their unrealistic expectations from simulation. But their model does not represent the rainfall, infiltration, soil, or subsoil variability in sufficient detail, and this seems to have been shown to be significant. Indeed the difficulty they had in representing the "base flow" component illustrates this. Yet they tended to criticize the surface component in their conclusion. They pointed to their ability to roughly simulate the Wagga Wagga outflow hydrograph with a Horton-type mechanism as an example of the untrustworthiness of physically based models. This is an unwarranted conclusion, insofar as the runoff in either case is from surface flow, the difference being the physical cause of the surface flow (saturation or Horton mechanisms). Their spatial segregation of hydraulic conductivities in the Wagga Wagga case leads to the possibility that for either mechanism the area in the swale would be the area of runoff generation. It is not difficult to show that for certain topographies and conditions, saturation excess flow and infiltration excess (Horton) flow mechanisms can produce similar surface runoff patterns, but in general the soil properties would differ between the two types of response. For a case with lower saturated conductivity on more upland areas, the possibility of such similarity is exceedingly more remote.

On one point made by Grayson et al. we can heartily agree: "In the present scientific climate, the collection and analysis of field data is undervalued," (p. 2663). We would add to that, the data should be comprehensive and accurate. Their expressed doubts concerning their base flow data provide a good example of data accuracy problems. Getting data from only one rain gage and one runoff gage is not enough, as they also demonstrated, to evaluate a model at the catchment scale. We need to have enough data on enough types of runoff events so that when a type 2 model simulation is considerably different from a measurement, we can answer the question, Why? Very few if any of our present data sets allow this analysis.

Apparently, Grayson et al. expect (or require) hydrologists to provide or discover a set of equations, equivalent to our current set (kinematic surface flow/Darcy's law/Richards's equation/etc.), that are equally parsimonious, with parameters of a physical nature (rather than conceptual) that can somehow be measured, and where a representative elementary area can be quite large. Certainly, their complaints concerning existing type 2 models may be interpreted as frustration that such a large-scale type 1 model or set of equations is not available. However, is there reason to expect that such equations will ever exist? If not, we need to get on with developing approaches to dealing with heterogeneity using the physical laws we know, and not throw out or abuse the only baby we have because she/he requires the effort of a (heterogeneous) bath.

References

Beven, K. J., Changing ideas in hydrology—The case of physically-based models, *J. Hydrol.*, *105*, 157–172, 1989.

Binley, A., K. Beven, and J. Elgy, A physically based model of heterogeneous hillslopes, 2, Effective hydraulic conductivities, *Water Resour. Res.*, *25*(6), 1227–1233, 1989.

Dunne, T., Models of runoff processes and their significance, in *Scientific Basis of Water Resource Management*, pp. 17–30, National Academy Press, Washington, D. C., 1982.

Goodrich, D. C., Geometric simplification of a distributed rainfall-runoff model over a range of basin scales, Ph.D. dissertation, 361 pp., Univ. of Ariz., Tucson, 1990.

Goodrich, D. C., Discussion of "The kinematic wave controversy," by V. M. Ponce, *J. Hydraul. Eng.*, *118*(9), 1334–1335, 1992.

Goodrich, D. C., and D. A. Woolhiser, Comment on "Physically based hydrologic modeling, 1, A terrain-based model for investigative purposes" by R. B. Grayson, I. D. Moore, and T. A. McMahon, *Water Resour. Res.*, this issue.

Grayson, R. B., I. D. Moore, and T. A. McMahon, Physically based hydrologic modeling, 1, A terrain-based model for investigative purposes, *Water Resour. Res.*, *28*(10), 2639–2658, 1992a.

Grayson, R. B., I. D. Moore, and T. A. McMahon, Physically based hydrologic modeling, 2, Is the concept realistic?, *Water Resour. Res.*, *28*(10), 2659–2666, 1992b.

Hillel, D., Modeling in soil physics: A critical review, in *Future Developments in Soil Science Research*, pp. 35–42, Soil Science Society of America, Madison, Wis., 1986.

Hjelmfelt, A. T., and R. E. Burwell, Spatial variability of runoff, *J. Irrig. Drain. Eng.*, *110*(1), 46–54, 1984.

Klemes, V., A hydrologic perspective, *J. Hydrol.*, *100*, 3–28, 1988.

Loague, K. M., and R. A. Freeze, A comparison of rainfall runoff modeling techniques on small upland catchments, *Water Resour. Res.*, *21*(2), 229–248, 1985.

Simanton, J. R., M. A. Weltz, and H. D. Larson, Rangeland experiments to parameterize the water erosion prediction project model: Vegetation canopy cover effects, *J. Range Manage.*, *44*(3), 276–282, 1991.

Smith, R. E., D. C. Goodrich, and D. A. Woolhiser, Areal effective infiltration dynamics for runoff of small catchments, *Trans. Int. Congr. Soil Sci.*, *14th*, I-22 to I-27, 1990.

Woolhiser, D. A., and D. C. Goodrich, Effect of storm rainfall intensity patterns on surface runoff, *J. Hydrol.*, *102*, 335–354, 1988.

Woolhiser, D. A., R. E. Smith, and D. C. Goodrich, KINEROS, A kinematic runoff and erosion model: Documentation and user manual, *Rep. ARS-77*, Agric. Res. Serv. U. S. Dep. of Agric., Tucson, Ariz., 1990.

Wu, Y.-H., D. A. Woolhiser, and V. Yevjevich, Effects of spatial variability of hydraulic resistance on runoff hydrographs, *J. Hydrol.*, *59*, 231–248, 1982.

D. R. Goodrich, D. A. Woolhiser, and J. R. Simanton, USDA Agricultural Research Service, Southwest Watershed Research Center, Tucson, AZ 85719.

R. E. Smith, USDA Agricultural Research Service, AERC, Colorado State University, Fort Collins, CO 80523.

(Received July 6, 1993; revised October 12, 1993;
accepted November 9, 1993.)