DESIGN AND ANALYSIS OF THE CATACHMENT EXPERIMENT

John D. Hewlett and Leon Pienaar

About ten years ago there was so much criticism of the existing experimental watershed programs that it was unpopular for a while to plan catachment experiments of any type. The criticism had come largely from quarters with little experience in the method of experimental watersheds and none at all in the analysis of their results. Nevertheless opinions to the effect that we had learned little from experiments on watersheds, that their cost was far greater than their benefits and that they should be severely curtailed in future, were trumpeted from Washington and other high circles, both domestic and international.

The criticism of these programs did some good and some bad (2). There were (and still are) many small, relatively aimless programs that cost millions of dollars but yielded only local information on certain components of the hydrologic cycle. But somehow the intended attack upon these non-productive programs developed into a much more effective attack upon the method of watershed experimentation. Because it was relatively undeveloped, the method proved an easier target than the programs, the latter rendered largely immune by budgetary entrenchment. As a consequence, theoretical progress on the method almost stopped and the new or independent researcher was discouraged from proposing projects based on experiments with whole watersheds. Indeed, one federal agency (the Office of Water Resources Research) made such discouragement an article of policy. The unfortunate effect has been an almost total lack of progress in the theoretical aspects of catchment experimental design during a period when the whole trend of ecosystem research was converging upon the method as ideal from both a practical and scientific standpoint. In the last ten years the "unit ecosystem" has virtually been defined as a drainage basin or watershed.

Herein lies an odd situation and an opportunity, although those of us most closely associated with the use of experimental basins for studying land management, ecology and hydrology have been slow to see it. There are in existence a mere handful of short papers that deal with the logic, strategy, design and analysis of small watershed research. The only monograph of any weight which deals with the method (10) elaborately reviews techniques of data collection and processing but is painfully brief on research design and the nature of inferences to be drawn from results. It is almost as if we were permanently locked into the second of the three classic phases of science, that is, the collection of data, and are neglecting the formulation of new hypotheses and the testing of them.

There is no doubt in my mind that the watershed experiment has yielded more information about the effect of man's use of the ecosystem on the hydrologic cycle than any other approach. Early summaries of these effects -- for example, those of George P. Marsh in his book, Man and Nature (6) -- contained substantial misinterpretations based

---

Professor of Forest Hydrology and Assistant Professor Biometrics, School of Forest Resources, University of Georgia, Athens, Georgia.

Reprinted from PROCEEDINGS of a Symposium on Use of Small Watersheds in Determining Effects of Forest Land Use on Water Quality; EDITED by E. H. White, University of Kentucky, Lexington, Ky., May 22-23, 1973.

on case histories and some descriptive observations, misinterpretations which have been corrected only in late years by results from experimental watersheds. Not the least of these is the ancient and international controversy over the effects of forest upon the water balance of the land, on flood flows, and more recently upon the mineral balance. Information has accumulated despite the lack of rigor in experimental design.

The experimental watershed method can, like any other, be oversold. I tried to review some of the reasons in a paper given at the Joint FAO/USSR Symposium on Forest Influences, held in Moscow, August, 1970 (3). Toebes and Ouryvaev' monograph (10) also contains evidence for caution against rushing into small watershed research with the innocent idea that a few grab-samples and some rainfall-runoff measurements is all we need to prove a point. The method is powerful but is, as the critics have said, costly and time consuming. It requires careful planning and funding. There is an expertise to it, not necessarily acquired by reading a few papers reporting similar experiments. And finally, because of time and money invested, there develops a keen temptation either to bury the unhappy failure or to liberally fudge the results into a preconceived idea of what they should have been. Almost as bad, there has been some tendency to repeat successful experiments over and over, without securing either true replications or new hypotheses to test. Howard Lull, commenting on the popularity of the Coweeta-type catchment experiment, has wryly termed such habit-forming research as the "calibrate, clear-cut and apologize" syndrome.

New enthusiasts are descending upon the concept of the small basin, bringing in funds normally funneled elsewhere. They assume that the basin's ability to integrate the net effect of myriad internal cycles in the ecosystem is all they need to settle such controversies as the effect of clear cutting on the nitrogen or phosphorus cycle. Some, perhaps stimulated by the Hubbard Brook mineral cycling study, have leaped upon the nearest set of gaged basins unaware of the difficulties associated with the use of whole drainage basins to study massive mineral and hydrological phenomena. It is not that these hasty efforts will fail -- we are bound to learn something from them -- but that the results will be so hard to assimilate into an organized body of knowledge about the ecosystem if each bit of data is offered as a proof and each watershed experiment as a law to itself.

DESIGN PROBLEMS

Before we can approach the problem of how to draw inferences about treatment intensities and responses in general, we must assure ourselves that a response to a particular treatment can be demonstrated. This involves the accuracy with which various components of the hydrological cycle can be measured, the standard error of estimate in each component and what trends in hydrological processes are taking place at the time of treatment. Specifying the exact nature of the treatment is not always easy. There is no space here to go into the various difficulties with selection of sites, measurement of variables and exactness of treatment; others have treated these in some detail elsewhere (10). More attention has been paid to these difficulties than to logic and strategy in the use of watersheds as experimental or observational units. Basically there are four approaches: the correlation

study, the single watershed time-trend analysis, the paired watershed experiment and the multiple watershed experiment. The "nested pair" is considered a variation on the paired watershed approach.

THE CORRELATION STUDY

The correlation study is one in which the responses of interest and the presumed causes of the responses are measured or estimated largely from existing data collected under established program of hydrologic inventory (USGS, TVA, Weather Service, EPA). Little or no additional controls are added and a rather low level of accuracy and precision in basic data is acceptable. The bank of data is analyzed by correlation techniques, multiple regression and factor analysis in hope that the main causative influences will be revealed. The end result is usually a regression model with rather hopelessly interrelated "independent" variables. We often then have a choice between a set of real variables, such as rainfall, temperature or elevation, which we understand but cannot relate independently to response, or, on the other hand, a set of factor complexes which are presumably independently related to response but which we most likely will be unable to interpret in relation to real variables in nature. In any case, multiple correlation coefficients are usually too low for any but crude predictions and 20 to 50 percent of the variation in response is often left unaccounted for.

A well-known example of this type of study is that of Rakhmanov (7). In a correlation analysis of a number of large subbasins of the upper Volga River, he concludes that forest cover increases annual streamflow or water yield, basing his argument on a persistent positive correlation between percent forest cover and streamflow. The inter-correlation between percent forest, elevation, precipitation, size of basin and latitude is not overlooked entirely but Rakhmanov's conclusion seems strongly influenced by a prior conviction that the positive correlation was to be expected. In this case, a prior probability has been applied but one subjectively arrived at. The upshot is a controversy that has diverted a number of research programs.

Other efforts at multiple correlation analysis have been scarcely more successful in getting at the true relationship between cause and response. The correlation approach will seldom be sensitive enough to serve as more than a crude predictor of response within a given hydrologic province, or as supporting evidence for major effects on response revealed by more exacting methods.

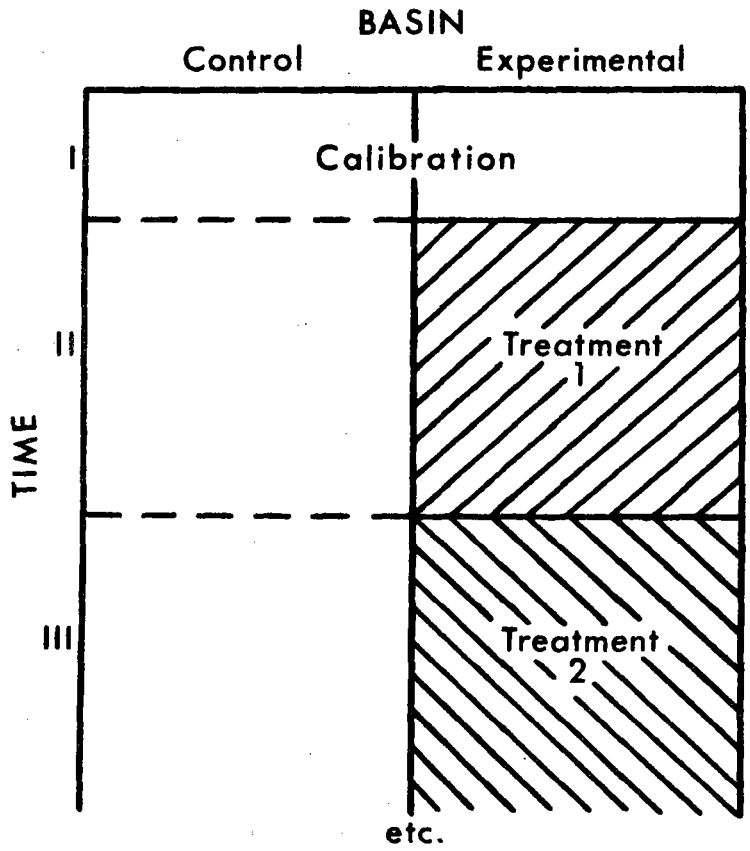
THE SINGLE BASIN APPROACH

The single watershed experiment (that is, one with no control basin) takes two variations. The one most often encountered is an analysis of an existing set of hydrologic record on a basin that has more or less accidentally sustained some "treatment", for example, a wildfire, a blow-down or a gradual change in land use. The response factor of interest is regressed or plotted over one or more "independent" variables that are changing with time, rainfall, temperature, land use and so on. The data are usually crude and incomplete and only part of the basin may be affected by the treatment. Less

often the single basin may be selected in advance and a uniform treatment applied after some effort to "calibrate the watershed on itself." Compared to the first type of single watershed experiment, this type has the limited advantage of a deliberate effort to control treatment and to measure response. Both can be classified as time-trend analyses, a search for some sharp break in a plotting of response over time. The method is extremely weak analytically and yields information only after dozens of similar experiments have been carefully reported. So far there is little backlog of results from single catchment experiments and the fortuitive analysis of representative basins for response to "treatments" is not very convincing. Reigner (8) carried the analytical aspects of the approach as far as it could be taken, but was forced to conclude that "there is no way of knowing if the correlations are valid, if they are the results of chance, or if they stem from overmanipulation of the data". Such are the hazards of the single catchment approach, although for such visible effects as gully formation, drastic changes in sediment production or water temperature studies, the single basin method may suffice.

THE PAIRED BASIN EXPERIMENT

The paired watershed experiment is relatively more costly but quite productive of information. In essence there are two basins, two or more periods of observation, and one or a series of treatments:



The control watershed is simply a barometer and the measurement of a variable of interest on the control replaces the measurement of many extraneous variables whose operation upon the system we really do not understand and cannot predict with any accuracy. For example, rainfall intensity and antecedent soil moisture on the basin combine in a complex manner to produce stormflow. We do not know the function that relates these two things to stormflow in the absence of treatment and therefore cannot solve it. However, the control basin serves as a real-world analog to solve this unknown function for us. The effect of a treatment is evaluated as a difference between expected response (Y_E), a value based on correlation between the two catchments before treatment, and the actual response (Y_A) measured after treatment on the experimental unit.

$$\text{Effect of Treatment} = Y_A - Y_E = Y_A - (b_0 + b_1X_1 + .. + b_nX_n) \quad (1)$$

If the water quantity, quality or timing variable of interest were a linear and time-invariant function of treatment, we would have no difficulty with the analysis. But neither the treatment nor its effect is constant over time; in a clear-cutting experiment regrowth begins immediately and any changes in mineral or water export are not apt to be linearly related to any corollary variable such as basal area, crown coverage, number of stems and so on. So far we have accumulated only limited information on the time distribution of response to a treatment; most take the form of a sigmoid growth or log-linear die-away function with time. This is stable enough in the case of water yield changes after felling to propose the analysis model (4):

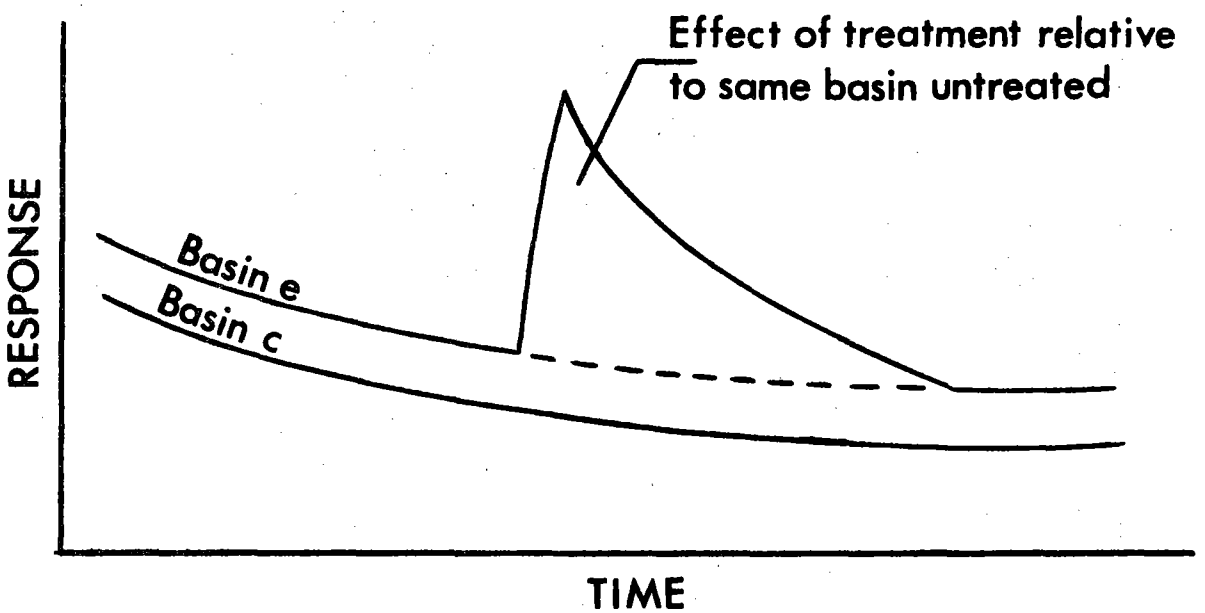
$$\text{Effect of Treatment} = Y_A - (b_0 + b_1X_c - b_2\ln T + b_3P) \quad (2)$$

where lnT represents the logarithm of time since treatment, X_c is water yield on the control basin and P is seasonal or annual precipitation. The parameter b_2 is not a constant for repeated cuttings on a particular basin but depends primarily on the rapidity of regrowth following cutting (9). It remains to be seen how many other responses will obey this general relationship; it is of course fairly reasonable to expect all "shock" treatments to have a die-away effect and all gradual treatments (such as the planting of grasslands to forest) to have a sigmoid growth response with time. A treatment applied gradually, or one that produces a delayed effect, will most likely generate a response fitting a gamma function, that is, a rapid rise to a peak and then a long die-away recession. In the case of water yield increases following clear felling of mature forest the functional model will most likely be:

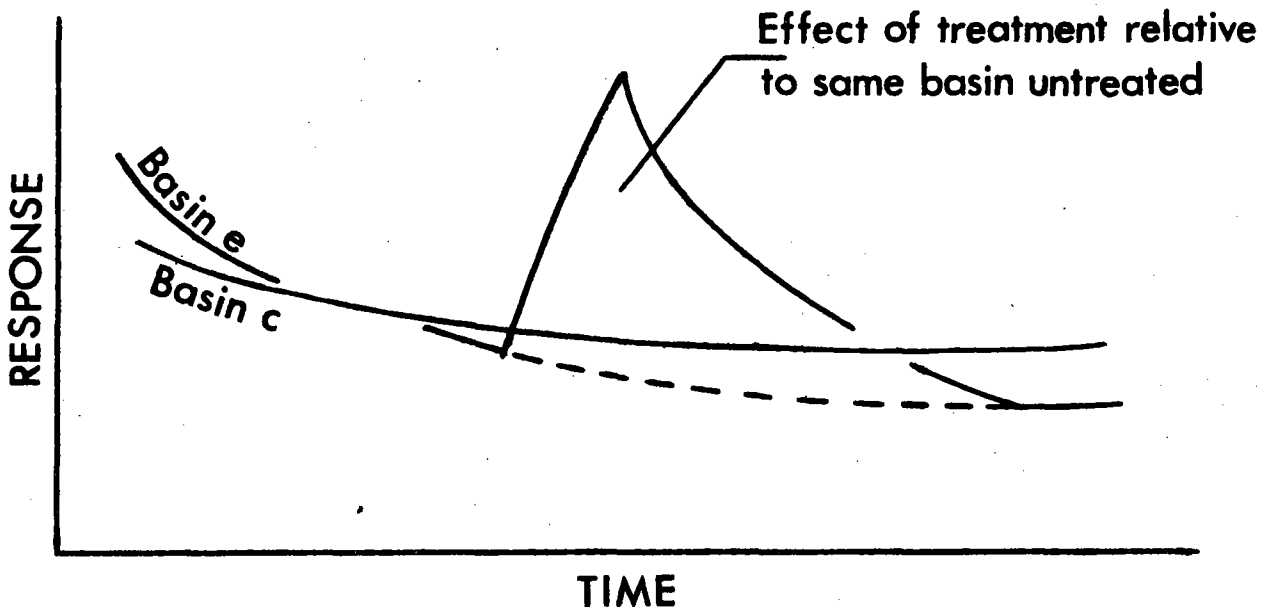
Incremental Yield = f (Phytomass or specific surface removed, log of time since removal, seasonal or annual precipitation) (3)

But the analysis of a single experimental pair does not yield sufficient evidence to permit the size or intensity of treatment (phytomass removed or regrown) to be functionally related to water yield changes. The problem of inference in this respect will be dealt with later.

Wicht (11) and others have been concerned with the possibly crippling effects of long-term changes in soil and vegetation on the analysis of paired watershed experiments. One question is, how do we account for slowly changing effects of response that may be taking place during the calibration period? Many watersheds are slowly changing for reasons not always immediately apparent. Fire protection for the last forty years has engendered slowly changing mineral export rates in wildland watersheds. The death and replacement of the American chestnut has been a long-drawnout ecological process that possibly affects water and nutrient balances. Recovery of abandoned agricultural land certainly is accompanied by steady changes in mineral and water blances. Dust storms, industrialization nearby, massive weed and insect infestations and the chemicals used to control them, as well as the almost universal fertilization of field and forest, all may produce effects that extend into and through the experimental period. Fortunately, the paired watershed approach allows such effects to be largely extracted during analysis. The essential condition that must prevail is a high correlation of the measured variable of interest between the two catchments before treatment. A diagram helps to show why the analysis of the effect of treatment may be valid despite changing levels of response from the two catchments before treatment:



The dotted line represents an expected response as predicted by the control basin during a treatment that was imposed on a gradually changing system. The functional relationship between the control and experimental basin need not be one to one; that is the time-dependent change inherited by the two basins before gaging may be progressing at different rates without damaging the experimental analysis:



The relationship between the control and experimental basin is still linear in this case, but that isn't necessary as long as there is sufficient paired observations before treatment to develop a predicting relationship.

The paired watershed experiment will yield a valid estimate of a change in response on one basin but relating that change to a range in treatment intensity requires more than one experiment on one catachment.

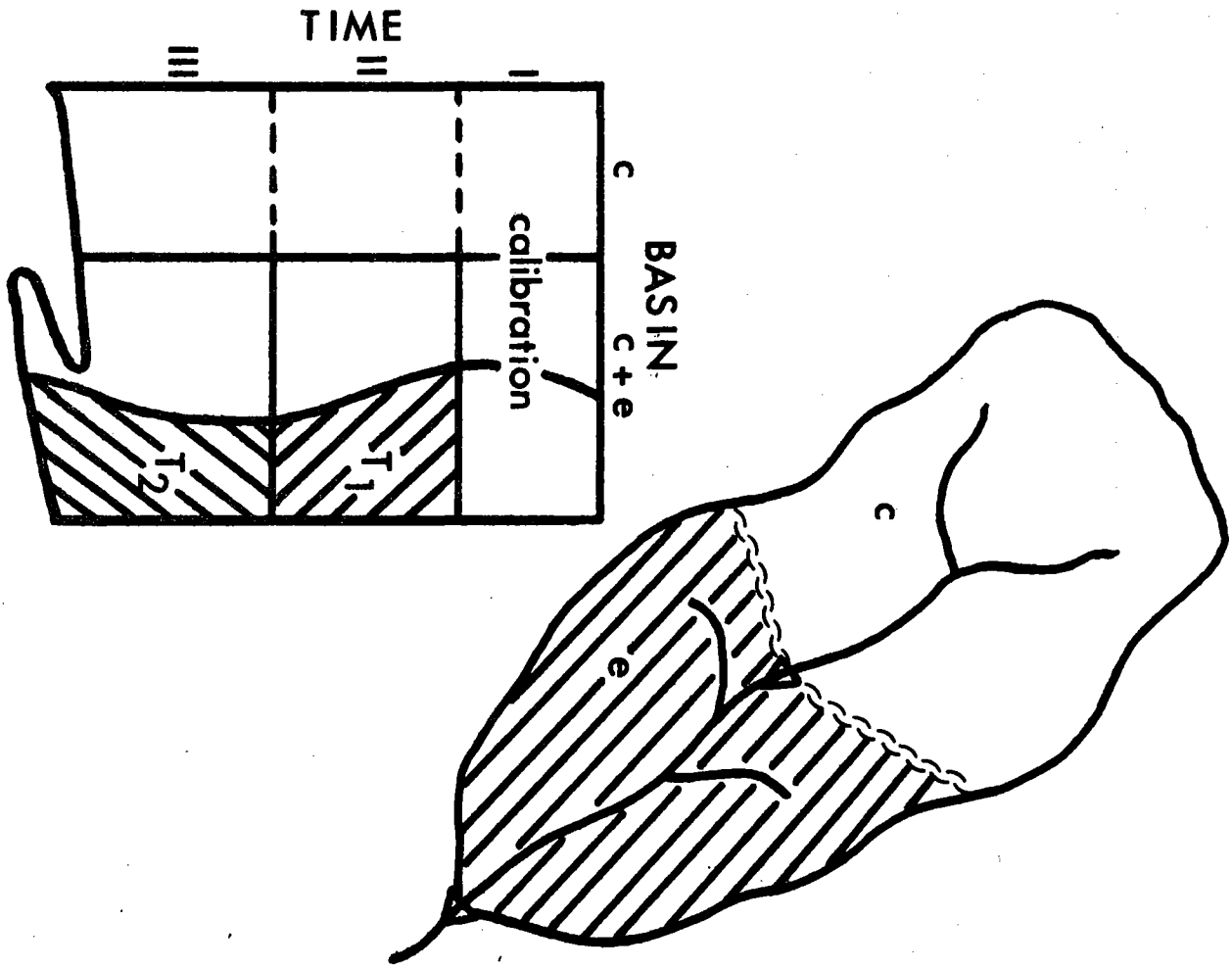
These illustrations highlight the difficulty in the development of a treatment-response model that will be applicable in the general case. Real effects have been demonstrated but the magnitude and duration of these effects have been estimated only as changes relative to the conditions that prevailed when the treatments were imposed. We may draw inferences about the processes involved in the observed responses from one or a few experimental treatments but we should be very cautious about proposing a general model each time a watershed experiment reveals some dramatic effects. In short, we should be honest in reporting these effects, but at the same time should be modest in proposing the model and the explanation for them. Some experimental watershed observations have been overworked in the recent past and the result has been a good deal of unproductive controversy.

THE NESTED PAIR

A variation on the paired catchment experiment is the nested pair or nested group of catchments, wherein a superior or inferior sub-basin of a

watershed is held as a control and a treatment is applied to the remaining portion. Some researchers see an advantage to this arrangement due to the possibility (it was thought) of higher correlations between two successive points on the same stream as compared with two points on separate branches. Thorough analysis of the errors involved in both approaches must await a larger body of clean data from both arrangements than we now have. However, some conclusions can be drawn from present information.

Certainly the inferior placement of the control with respect to the treatment sub-basin can be rejected as a poor design on the simple likelihood that the effects of treatment of the superior basin will influence the inferior control as that effect passes through it. There may be some underflow (flow escaping measurement at the upper weir) which will dilute or concentrate the control discharge in an unknown manner. If we impose a fertilizer treatment on the upper channel, the lower could be contaminated in such a way as to render covariance analysis invalid. The reverse arrangement requires more thought because the control will be beyond the hydrological influence of the treatment. The control-superior, nested pair would look like this:



The analysis of such a nested pair may appear at first to be unaltered by adding the response from Basin c to Basin e and then subtracting the response from Basin c again. If this were the case, there would be no difference between the correlation between c and [(c + e) - c] and the correlation between c and e. In other words, we would be assuming that response c could be substracted <u>without error</u> from the total response at the base of the watershed. Unfortunately the errors in measuring response from both the control and the experimental areas are contained in the measured value at the base of the whole watershed. Thus the measurement error on the control is contained twice in the estimate of treatment effect, which may be represented in this manner:

Paired Basins: $Y_A - Y_E = Y_A - (b_0 + b_1 Y_c)$ (4)

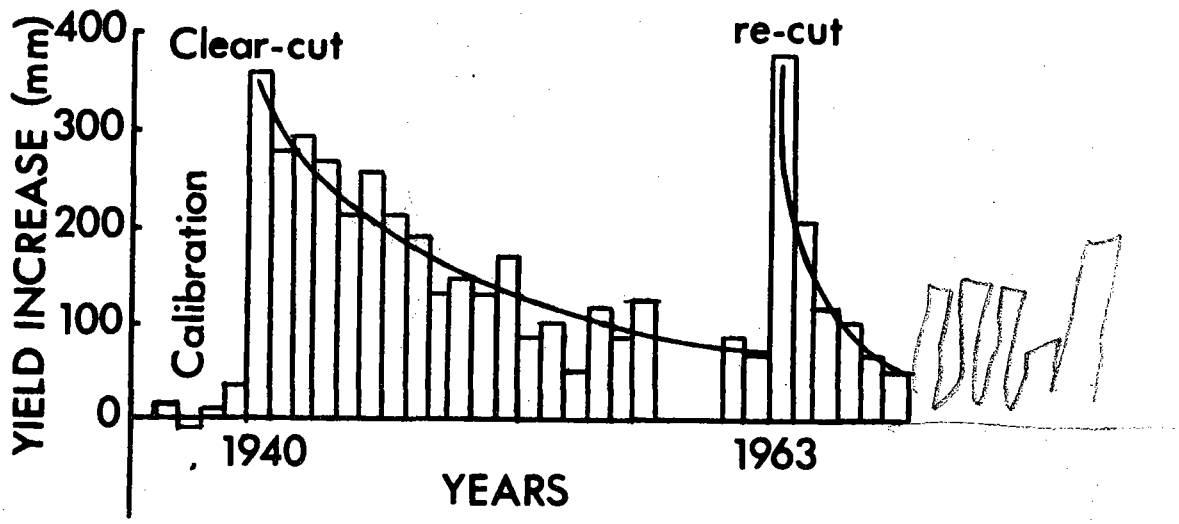
Nested Basins: $Y_A - Y_E = (Y_c + Y_e) - (b_0 + b_1 Y_c)$ (5)

All other factors set aside, the nested pair is an inferior design to the independent pair. However, it is entirely possible that particular nested pairs might be more highly correlated than some independent pairs. The fact being already in evidence, one cannot quarrel with the choice of the pair that demonstrates the highest pre-treatment correlation, as long as this is also accompanied by the lowest absolute standard error of estimate. Suffice to say, there is an a priori advantage in the independent pair as compared with the nested pair in catchment experimentation.

We will be reminded at this point that in common use of regression analysis of a paired watershed experiment, we normally assume that Y_c in the estimating equation (4) is plugged in without error to estimate Y_E for a single determination of treatment effect $(Y_A - Y_E)$. There is only one value for Y_c associated with the one estimate we wish to make. If successive observations, (say, ten years of annual water yield after clear-cutting) could be assumed to come from the same treatment-response population, then we would not have to assume no error in Y_c but could proceed with covariance analysis as if the years were sampled independently of treatment.

However, the successive measurements of response may be expected to correlate strongly with time because the "treatment" is time dependent -- each year sees an altered vegetal cover occupying the treated basin. Even drastic treatments (short of complete paving of the basin) are apt to be steadily modified by time insofar as their effects upon water and mineral exports are concerned. This difficulty has been and remains a weakness in the analysis of paired watershed experiments.

But we should not be satisfied with this situation -- there must be better ways to analyse the effects of treatment than we have used in the past. Figure 5 from Swank and Helvey (9), reproduced below, represents responses in annual water yield to hardwood clear-cutting on a pair of Coweeta catchments . Standard errors of estimate, based on single-value solution of the calibration estimating equation for each successive



year, indicate that the .95 percent confidence limit on these estimates is on the order of plus or minus 4 inches (100 mm). Yet the time trend (the variation around the fitted line) gives visible evidence that these error estimates surely cannot be applied to the over-all experiment. Kovner (4) fitted a log-linear curve to these responses but it is evident from the second clear-cutting that the coefficient (b_2 in equation 2) to be applied to the logarithmic decline in yield must be some function of the rate of reoccupation of the site by vegetation (it was faster the second time than the first). If we may assume that the treatment responses are to be transformed in accordance with the logarithm of time since treatment, then we may have part of the answer to the dilemma whether to handle observations as a sample from a treatment-response population or as a series of independent items representing a varying response to a varying treatment.

Of course, this treatment effect as a function of time is not a "law" but rather a reasonable assumption based on past experience with the same type of treatment. The conversion of forest land to agriculture will lead to complex cycles in sediment and mineral export which may exhibit an altogether different response to time. Planting agricultural land to trees should produce a sigmoid return to sediment, mineral and water export rates normal under forest cover.

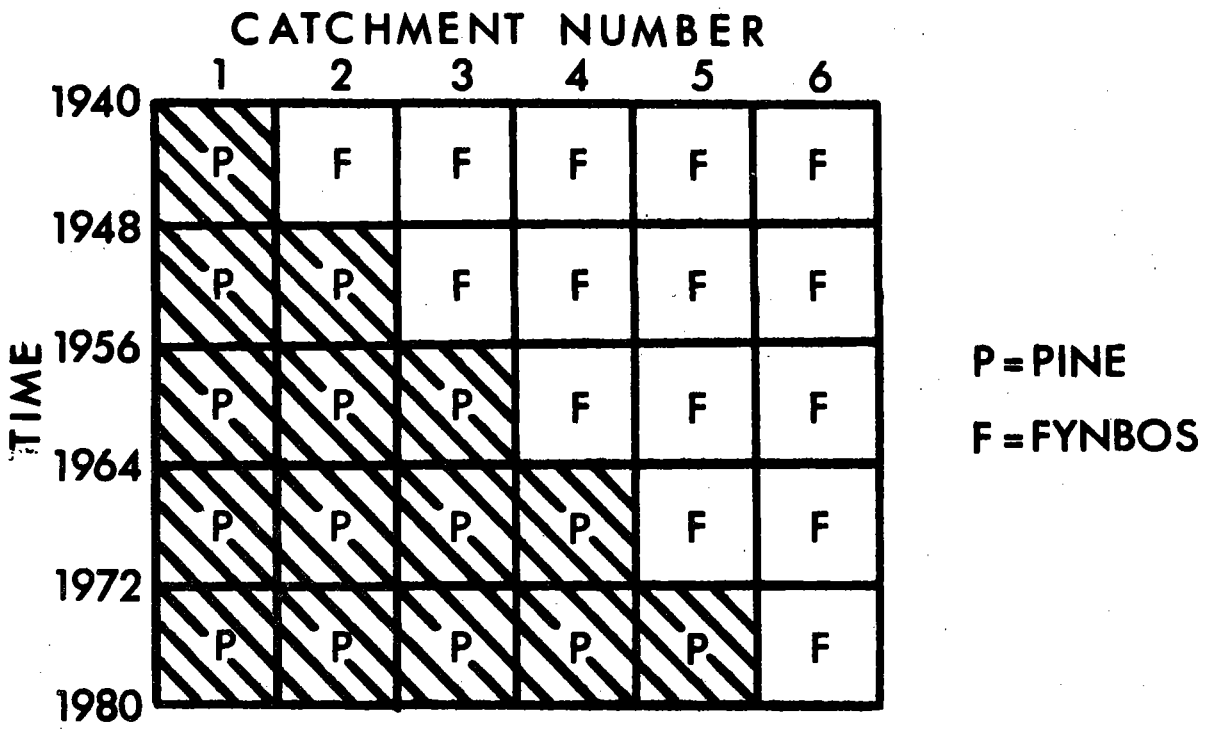
One cannot escape the impression that much remains to be done in the theory of analysis of paired watershed experiments, and in view of our evident plans to continue to use the method, the subject should be one of some interest to biometricians.

THE MULTIPLE CATCHMENT EXPERIMENT

There have been only one or two serious attempts at a truly multipe catchment experiment -- those of Professor Wicht, Stellenbosch University, Republic of South Africa. Elsewhere the cost of establishment, control and maintenance of a multiple set of basins and treatments to serve one research goal has discouraged even the hardiest of institutional programs. Research basins such as Coweeta, San Dimas, Hubbard Brook and H. J. Andrews are collections of paired watersheds, each with its own research objective. Despite the more or less accidental coexistence of several basin treatments, these stations do not constitute in any way replicated multiple basin experiments. There are a few examples of grouped or nested basin experiments in which one basin served as a control for two or three experimental basins, but the experimental treatments have invariably been efforts to reveal responses to several levels and intensities of treatment, not to replicate the experiment. The error involved in response to one type of treatment, the experimental error, is still unaccounted for in the analysis of a grouped or nested set of basins treated at various levels and times.

The paired catchment experiment focuses attention on a calibration period in advance of treatment. The few papers that have dealt with the statistical analysis of the watershed experiment (5;12;13) have concentrated on determination of the length of the calibration period necessary to test the significance of a change in response from the treated basin. In the early days of catchment experimentation, this was a valuable precaution against hasty conclusions based on new methods. However, the calibration problem may have been over-emphasized; those seeking evidence of land-use effects on water have been discouraged by what appears to be a general conviction that long, costly calibration periods in advance of watershed experimentation are absolutely necessary.

This conviction is largely unfounded, as is visibly evident in the response in water yield revealed in Figure 5 from Swank and Helvey (9). Furthermore, Wicht was not concerned with the calibration problem when he established a 40-year multiple catchment experiment in 1940 to determine the effect of replacing fynbos vegetation with Pinus radiata. The design might best be described as a sliding replication of paired catchment experiments, with the calibration period telescoped into the treatment period:



Pine was planted on one basin the first year and for eight years the developing pine stand was matched aginst five control basins under the more slowly developing fynbos vegetation. In the ninth year, another basin was planted to pine, and in the 17th year still another, and so on. One control basin remains as an index to changing climate and developing fynbos to the end of the experiment. The multiple controls decrease in number while the treatment is replicated through time. One clear advantage is the built-in check upon the quality of control; if one control basin is for any reason a renegade (perhaps a slow subsurface leak is developing) the interrelation among the controls in the absence of treatment will revel it. This advantage, however, is gained at considerable cost and it is difficult to see any other design advantage over a series of paired catchment experiments.

It is tempting to think of this design as a fractional factorial but there is no indication that such was the intent. Nor can it be analysed factorially -- of the three major factors (time, catchment and treatment) all are confounded, time with treatment, catchment with treatment and catchment with time. In addition practically all of the interactions of interest, such as basin-treatment and time-treatment, seem also to be confounded. We have not made a thorough analysis to determine the full extent of confounding in

this particular design but it appears that it amounts to a series of paired watershed experiments in which 1-to-8-year-old pine is compared to developing fynbos five times at different points in time; 9-to-16-year-old pine is compared to developing fynbos four times; 17-to-24-year-old pine to fynbos three times; and so on to one comparison of 32-to-40 year-old pine and relatively mature fynbos.

Nevertheless, this ambitious experimental design does serve to throw further doubt on the absolute necessity for long pre-treatment calibration periods. As Wicht has point out, the design yields information of practical value to watershed managers progressively during the experiment, beginning with the first year.

The question suggests itself, what is the minimal design which would both afford good control and eliminate the need for long calibration periods in advance of experimentation? Possibly a three-catchment, four-year experiment would do the trick; one year to compare responses on an event basis before treatment, then one treated catchment the second year, two the third and perhaps all three treated by the fourth year. As with Wicht's experiment, the result would be a sliding replication with a telescoped control period. Furthermore, it could be an open-ended plan with the option of continuation or termination at any time after the first treatment. Such a design might serve many research purposes and should be studied more rigorously to determine confounded effects.

THE INFERENCE PROBLEM

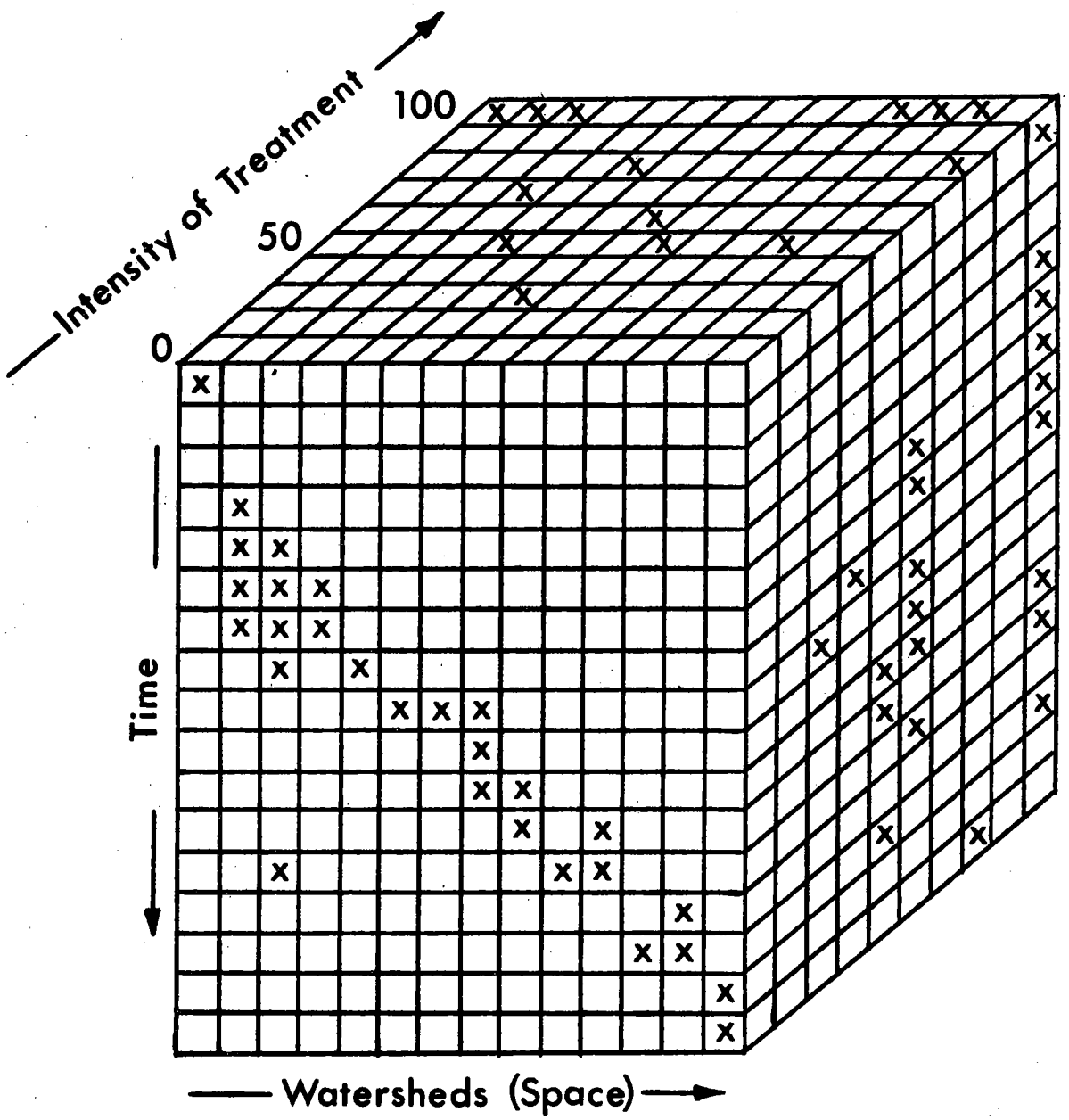
After half a century of catchment experimentation to learn the effects of land use and vegetal cover on mineral, water and energy balances, it behooves us to take stock of the amount and quality of the information we have amassed. This brings up the second phase in the use of experimental watersheds to develop knowledge about hydrological and mineralogical impacts following natural, deliberate or accidental changes in land use or cover. The first phase was the demonstration and proof of a change following some treatment on a particular basin; we might call this the statistical inference problem. But catchment experiments do not stand alone -- every time a definite response is demonstrated, a new benchmark for comparison has been established and the interpretation of other experimental results is reinforced or contradicted. Therefore, the second phase is the establishment by inference that there exists a general relationship between specified treatments and the responses observed. We may call this the general inference problem, in which we attempt to generalize the application of the results.

Until recent years we were in real trouble in trying to prove that the responses we measured were representative of all such responses, or that either the treatment of the basin were representative of all such treatments of basins. We are in better shape now for two reasons. One, an increasing number of independent experiments have been reported, most tending to show that a response of some kind can be measured. Two, non-parametric approaches to the analysis of experimental results are becoming more accepted in our search for the truth. We no longer must carry out all experiments in a Fisher-type random block design in order to proceed with a rigorously tested hypothesis in hand. We can come to valuable conclusions

with regard to some influences and can proceed with model building on the basis of provisional functions between treatment and response.

We have seen how cost and various logistical problems have all but prevented replication of watershed experiments. The restriction on replication puts a severe burden upon the practitioner of catchment research. Each experimenter is in fact one member of a team -- a disassembled team, if you will -- widely scattered in both time and space. More so than in other lines of research, the watershed experimenter must undertake the responisbility to design, collect and analyse with more than ordinary care, for the ultimate analysis of his experiment will most likely be performed not by him but by a member of the team elsewhere or at another time. It will be pointed out that this is not necessarily unique to catchment research but we would argue that it is in the particular sense that each catchment researcher's whole experiment constitutes a single sample unit.

Some of what we shall say about the catchment experiment from here on will be rather obvious to most of you, but it will aid in our perspective if we review some elementary points. Using water yield changes following forest clearing as an example, it must be apparent at the start that we have never really had a replicated experiment, either in space or in time. That is, even where grouped or multiple catchments were treated, or where a single one had been retreated, no two treatments were ever subjected to exactly the same criteria or conditions. The nature of the treatments were different, they were imposed in different seasons, the stands and sites were different, and so on. The experimental evidence, therefore, has built up in this fashion:


Intensity of Treatment
100
50
0
Time
Watersheds (Space)

Using water yield changes as an hypothetical example, forty or more experiments have been carried out between 1930 and 1970, widely scattered about the world, and varying from 10 to 100 percent reduction in basal area of forest stands. First, second, third and higher year responses of water yield have been estimated, represented in the diagram by runs in the time direction. These are not replications in time, for the "treatment" in successive years is not the same. It is also apparent that there have been few if any replications in space.

The diagram illustrates our problem with inference; there are huge gaps in time, space and intensity of treatment. But we wish to be in a position to estimate the response in the space-time-intensity boxes not checked.

The response in this case is change in water yield. In logical order we are concerned with first the sign of the change, whether it is an increase or a decrease, and second with the magnitude of the change, which will determine whether the change is important and whether we can evaluate functional relations. Finally, we want to know the time trend in the change so that we can "build a model" that will accurately predict what will happen over time under varying degrees of cutting on other watersheds.

First let us deal with the sign of the change. It will be said that we knew all along that cutting increased streamflow. But we didn't really, because it isn't absolutely certain even yet that the elimination of forest increases water yield in all circumstances; it is to be expected with a high degree of probability but not absolutely certain. The first catchment experiment that indicated a positive response of water yield to forest removal was reported in 1928 (1). At that point in the developing historical record, we had no basis for predicting the outcome of the second experiment. But the next carefully conducted test of forest cutting on water yield, reported from the Coweeta Watershed No. 17 in 1944, also demonstrated a positive response. As more tests of a similar nature were added, it became possible to apply a form of Bayesian inference to indicate that a positive response to forest cutting has a high degree of probability.

It might have seemed reasonable to assume in 1928 that it was equally likely to observe among untreated watersheds an increase as a decrease in streamflow over some period of time. That is, the probability of a positive change was one-half and of a negative change the same (we will ignore zero on the unlikelihood of exactly no change):

$$P(+) = 0.5 = P(-) \tag{6}$$

Not knowing any more about it, we may also have assumed that the probability of a positive change following forest cutting was also equal to the probability of a negative change:

$$P(t/+) = 0.5 = P(t/-) \tag{7}$$

The positive result from the Wagon Wheel Gap experiment didn't alter these basic probabilities but the 1940s and 1950s added a number of positive and a few zero or negative changes in yield following various experimental cuttings. Within a hypothetical set of 40 experimental watersheds, observed with some degree of care and control, let's assume that 32 have shown increases and 8 have shown decreases in water yield. Among those showing increases, 23 were partially or wholly cut, and among those showing decreases, only one sustained any cutting. Therefore, the prior probabilities under forest cutting are:

$$P(t/+) = 23/32 = .719 \text{ and } P(t/-) = 1/8 = .125 \qquad (8)$$

that is 71.9% of those watersheds showing increases were partially or wholly cut and 12.5% of those showing decreases were also cut in some manner.

With these prior probabilities based on watershed experiments in various climates, we can use Bayes' Theorem to estimate the "posterior probability" that cutting will increase streamflow. The theorem states:

$$P(+/t) = \frac{P(t/+)\ P(+)}{P(t/+)\ P(+) + P(t/-)\ P(-)} \qquad (9)$$

The probability of a positive response, given a [redacted] treatment, becomes:

$$P(+/t) = \frac{(.719)\ (.50)}{(.719)\ (.50) + (.125)\ (.50)} = .853 \qquad (10)$$

Whereas before the experiments we had to accept the probability of an increase after cutting ($P(t/+) = .50$, we now have a basis in Bayes' Theorem for a statement of some confidence regarding the inverse relation between forest cover and water yield. If we predict an increase after cutting, we should be right 85 percent of the time. This analysis is, of course, hypothetical; I suspect the real probability is higher but we will not know for sure until each experiment is evaluated and weighted in accordance with its experimental error.

By writing the probabilities around increases greater than one inch per year, two inches per year, and so on, the estimated of response can be further refined quantitatively. The same applies to the intensity of treatment. In fact, the reasoning applies equally well to any form of quantitative, qualitative or timing response to watershed treatments.

Further points can be made about the diagram above. The large gaps among years, watersheds and intensities of treatment are only partially narrowed by inference based on prior experiments. Even without a complete inventory of current experiments, we may assume that more simultaneous treatments, more levels of intensity and more catchments are being added in a rather helter-skelter manner. Is any consideration being given to the inference

problem? Are we wasting effort on time trends that should be devoted to establishing more watersheds? What are the relative benefits of testing responses from small differences in treatment intensities versus longer time trends on extreme treatments? For example, it scarcely seems necessary to test all levels of basal area reductions from 0 to 100 percent in, say, one Appalachian hardwood type. Nor will it prove efficient to test for small differences between two similar types or species. More information should be forthcoming if more sites or basins are tested, or if longer time trends are observed on existing experiments. At this point in time, the inference problem with respect to space (watersheds) is greater than with either of the other two dimensions. Given a choice between time (years) and levels of treatment, longer experiments in time will pay off better than more levels of treatment. In the latter case, three levels of treatment, perhaps the 33, 66 and 100 percentiles, should suffice for most research objectives; if only one is possible, the 100 percent reatment intensity will probably yield the most information.

SUMMARY

No one should be entirely satisifed with the current analytical technology in watershed experimentation. This review is far from comprehensive, but we have shown that there is a need for a more rigorous approach to such research. The ore is rich but the cost of extraction is high. Under these circumstances, careful and logical planning is called for and we cannot afford to waste much of the precious metal in the slag heap. The older work is not yet purely refined and much of the newer work is not built carefully on the old. Some general recommendations can be made.

All things considered, there seems to be no better basic design than the paired catchment experiment. Although some advantage in control is gained by grouping the pairs, there is a concommitant loss in the sampling of space (watersheds). The ideal design may be the three-basin, four-year study.

Effort will be wasted testing trivial hypotheses or those which have been sufficiently tested before. Designing to test differences between small levels in intensity for one treatment is a waste of money when prior inference indicates that variability among basins or treatments is the main bottleneck to progress.

Wrong again John

Experiments should not be continued after trends have been sufficiently established to answer the main questions. Nature is conservative and very little unusual is apt to happen after the form of the trend is revealed.

If at least a ten-percent change in the variable of interest cannot be expected as a response to treatment, time is wasted on watershed experimentation. One alternative is to redesign the treatment to produce a substantial change and interpolate the results back to the lesser levels of treatment. In this connection, beware of losing the treatment effect among compounded errors when the variable of interest is a sum or a product.

And finally, care in supervising the hydrological and analytical controls in field and lab is paramount, lest all the time and money invested in a massive outdoor experiment be lost for trivial reasons.

LITERATURE CITED

1. Bates, C. G. and J. Henry. 1928. Forest and streamflow experiment at wagon wheel gap. Colo., Monthly Weather Review, Supplement No. 30, GPO, Wash., 79 p.

2. Hewlett, J. D., H. W. Lull and K. G. Reinhart. 1969. In defense of experimental watersheds, Water Resour. Res., 5(1), 306-316.

3. Hewlett, J. D. 1970. Review of the catchment experiment to determine water yield. Proc. of the Joint FAO/USSR Intern. Symp. on Forest Influences and Watershed Management, Moscow, Aug.-Sept. 1970, FAO (Rome), 145-155.

4. Kovner, J. L. 1956. Evapotranspiration and water yields following forest cutting and natural regrowth. Proc. Soc. Am. Foresters, 106-110.

5. Kovner, J. L. and T. C. Evans. 1954. A method for determining the minimum duration of watershed experiments. Trans. Am. Geophys. Union, 35(4), 608-612.

6. Marsh, G. P. 1863. Man and Nature. (Later published as The Earth as Modified by Human Action. Charles Scribner's Sons, New York, 1907).

7. Rakhmanov, V. V. 1970. Effect of forest on runoff in the Upper Volga Basin. Proc. of the Joint FAO/USSR Intern. Symp. on Forest Influences and Watershed Management, Moscow, Aug.-Aept., 1970, FAO (Rome), 187-208.

8. Reigner, I. C. 1964. Calibrating a watershed by using climatic data. Northeastern Forest Exp. Sta. Research Paper NE-15, 45 p.

9. Swank, W. T. and J. D. Helvey. 1970. Reduction of streamflow increases following regrowth of clearcut hardwood forests. In: Symp. on the Results of Research on Repres. and Exp. Basins. UNESCO-AIHS Publ. 96, 346-360.

10. Toebes, C. and V. Ouryvaev. 1970. Representative and Experimental Basins: An International Guide for Research and Practice. Studies and Reports in Hydrology, No. 4, UNESCO, Paris, 348 p.

11. Wicht, C. L. 1967. The validity of conclusions from South African multiple watershed experiments. In: Intern. Symp. on Forest Hydrology. edited by W. L. Sopper and H. W. Lull, Pergamon Press (Oxford), 749-760.

12. Wilm, H. G. 1944. Statistical control of hydrologic data from experimental watersheds. Trans. Am. Geophys. Union, Part II, 618-622.

13. Wilm, H. G. 1949. How long should experimental watersheds be calibrated. Trans. Am. Geophys. Union, 30, 272-278.

hewlett-73-catchment