Overview of Conference on Hydrologic Data Networks

W. B. Langbein

U.S. Geological Survey, Reston, Virginia 22092

This conference is the latest involving conferences whose research includes work in what is called the design of data networks. The broader scope of this conference reflects, in my view, movement toward greater sophistication in research and in subject matter, and it is good that these kinds of critical inquiry are going on. The rough estimates mentioned at the conference show data collection to be big business, even though data storage and retrieval tend to attract most public attention. The design of data systems, that is, the efficiency and effectiveness, must also be concerned with the manner of collection, the places and numbers of observations, their costs, and their utility. These matters constitute the fundamentals. Yet in the face of the wide range of present and prospective uses, the complex range of objectives, and the diversity of variables the products of research and critical inquiry appear to find few applications. That subject as well as the needs and opportunities were examined at this conference.

WORK AT THE CONFERENCE

The several sessions were organized to address the subject along the following broad lines: basic considerations, role of statistics, economic aspects, worth of hydrologic data, space-time trade-offs, interphenomenon transfers, techniques, and case studies. Much work was reported on each subject and indeed is still in progress. Looked at along different lines, the conference dealt with subject matter in this order: quantity of precipitation, streamflow, water quality, and groundwater, but with virtually no attention paid to soil moisture or to quality of precipitation. In contrast to precipitation and streamflow the treatment of water quality and groundwater appeared rudimentary. This order reflects no more than recent history in working with data programs and not the hydrologic or economic importance. Indeed, network design must broaden its perspective. The rain gage-streamflow gaging station syndrome appears to have constrained the efforts of network design toward the development of ever more sophisticated statistical (and Bayesian) methodology for addressing earlier, but now less relevant, concerns over how many gaging stations to operate and for how long that were dominant when water resources were less encumbered by environmental issues. We need breadth rather than deeper penetration of older subjects.

As to techniques, the conference dealt with them in this order: statistics, physical determinism, economics, and management, again an order that reflects the status of past research and not necessarily of potential usefulness. Greater breadth is needed here too. From the papers presented one might judge the hydrologists are in front of the economists and at least even with the statisticians in their treatment of the problem. The upshot is that hydrologists cannot count on others to solve hydrologic data problems. Hydrologists must do it.

The conference did not consider instruments, sensing devices, telemetering, and the related subject of field costs. These are matters that closely interact with all other aspects of network design.

QUESTIONSPOSED

One of the most rewarding features of the conference was a sharp contrast that developed between the discussions on the floor and the guarded statements in the printed papers and abstracts. Basic elements of uncertainty and disquietude emerged. Conventional wisdoms were queried.

First, there was the assumption that practical network design is constrained by the hydrology, the terrain, and the specific uses for the data so that the few remaining degrees of freedom can be resolved by common sense. Network design is data planning, and planning is concerned with investment of capital and skills: how much to expend for what type of data, of what quality, when, where, and how long. What have been taken as constraints now reappear as planning choices and decisions.

Second, there was the question of the conventional proposition that 'The more data the better.' There was a time when, by almost any measure, all data were useful, and more were better. However, more data may not match the problem. It was shown at the conference by Davis et al. [1979] that the addition of data can worsen a decision that is made using a non-optimal or inadequate technique. It is said, 'A little information is a dangerous thing.' and apparently, so might be an excess of information.

Third, there was the question of the role of optimization techniques. More searching was the question whether efficiency remains the noble objective guiding network design. Other, more simple criteria may be more useful and, indeed, more realistic. Much literature suggests that the political officers and high-level leaders prefer strategies that reduce political or interagency conflict. They seek consensus. Moreover, the higher political and information costs of 'rational' planning tend to minimize comprehensive analysis [Lindblom, 1959]. Indeed, Rodriguez-Iturbe [1978], at this meeting, noted that because of complex objectives, structured analysis can be applied only to special or single-purpose (say, 'captive') networks and rarely to national or even regional networks. He observed that optimization may make no more sense than common sense, which underscores S. J. Burges and D. R. Dawdy's (in a panel discussion) advice at the conference about simplicity in network design.

Networks have been looked at 'in the small' at scales amenable to detailed analysis and to rules of optimization. A consideration of the complexities that emerge as one enlarges the scope must lead one to agree with Rodriguez-Iturbe [1978] that regional design, let alone national designs, is impossible. Rather, this result might suggest a complementary approach from the other direction with particular emphasis on strategies rather than on tactics (see, for example, Herfindahl [1969]).

Some of the kinds of queries that might be posed are the relative emphasis to be given short time versus long term, the intensive versus extensive investigations, and the role of tech-

1 Retired.

This paper is not subject to U.S. copyright. Published in 1979 by the American Geophysical Union.

Paper number 9W0679.
ology in changing the range and direction of data gathering (water data gathering has been less affected by technology than other aspects of geophysics); detailed analysis then would define the tactics within this structure.

Further, there was the question whether water data systems are to be planned as separate units or as part of an interrelated system that offers opportunities for information transfer. Long practice has centered attention on precipitation, streamflow, groundwater, water quality, etc. networks. This is the segmental approach, usually use specific, which, as Moss [1979] pointed out, requires only operational coordination (concurrency in who operates, avoidance of duplication, etc.) which is already well in hand. However, if networks are to be designed to profit by information transfer or to answer broad questions, such as the effects of various kinds of water and land use on water quantity and quality, then a considerable measure of technological coordination will be required. This may be the direction of network design to be examined in the next conference, together with instrumenal problems.

**TERMINOLOGY**

However, before the next conference, let us clear up some of the terminology. Perhaps the term 'network design' is not an apt description for the enterprise as it developed at this meeting. The views of what constitutes a network or what is network design are apparently as varied as the number of participants at this conference. This situation is not necessarily bad at a state-of-the-art research conference. Matalas and Barnes [1977] suggested data program design. However, we need a term that will accommodate general plans as well as local or otherwise limited schemes and schedules of observations (or stations) and plans for their classification and coordination, including plans for deriving information from those observations by interpolation, transfer, 'interphenomenal' relations, or other means. The plan for the observations is not independent of the plan for their use. I suggest that much of this extended view of the problem is subtended within the notions of technological coordination as presented by Moss [1979].

Considerable emphasis was placed upon a distinction between 'data' and 'information.' Although the usage as developed at the conference might not find universal acceptance, there seemed to be general concurrence that the two terms convey different meanings, data being a set of numerical facts (time varying as well as spatial varying) and information being the result of analysis and interpretation of the numerical facts. Thus hydrological data yield hydrological information upon analysis. Of course, one must recognize that what is information at one level of use may become data at the next higher level.

There was general acceptance of Fisher's [1949] definition of the information content of data (the reciprocal of the error variance), but I believe that Shannon's definition [see Goldman, 1953] should not be neglected for the simple reason that it takes account of the relevance or usefulness of the signal. Data that do not change the probability of an event convey no information either because they are inaccurate, incorrectly interpreted, or irrelevant. This definition incorporates a utility aspect in addition to the strictly 'telegraphic' content of a message or data.

**APPLICATIONS**

However, there are deeper problems. Little of the research has found its way into application, despite the several examples of application that were presented at this conference by the National Weather Service [Crawford, 1979], the U.S. Geological Survey [Tasker and Moss, 1979], and the Institute of Hydrology [O'Connell et al., 1979], among others. An analysis of the worth of data collected by the Science and Education Administration (formerly the Agricultural Research Service) demonstrated that a network of flow and sediment stations should be discontinued, and it was. Maddock [1978] offered a significant example to show that a 30% cut in data resulted in only a 5% loss in predictive information.

One aspect of the data problem is the exact inverse of an earlier day when data programs were sparse and hard to finance. The situation was changed greatly by the environmental movement. Congress, it now seems, is more sensitive to an undersupply than an oversupply of data, with the result that Public Law PL92-500, for example, and the regulations made under that act are prescriptions for great programs of data collection. Congress tends to believe that data will promote fulfillment of the programs. Is there evidence to show that rivers that are more closely monitored have higher compliance to standards (the policeman effect) and therefore are of higher quality? The greater appeal to monitoring should induce greater attention to its utility, but legal codification of the kinds of data, and the place and manner of their collection, in the regulations tends to freeze out inquiries about effectiveness or utility.

The problem may persist because each research paper tends to prescribe its own network objective. Decision makers rarely cooperate by defining their objectives, let alone by providing the a priori probability so eagerly sought by the Bayesian analyst. As we all know, Congress has authorized many data programs but rarely in terms useful for design purposes. Hence the federal agencies try to remedy the situation with formulations of their own. Langford and Kapinos [1979] and Leutner and Ficke [1978] each gave attention to this problem. One of the by-products of research on network design may be that it will suggest to users and decision makers the kinds of questions they need to ask of the data system. For example, recent analysis performed by the federal government in support of plans to reorganize its natural resources management agencies identified three attributes for the data systems: (1) long-term data trends—to look ahead to detect deterioration in water quantity and quality before these are readily apparent and while there is time to do something about them, (2) cause-effect—to detect and understand the future consequences of present day choices, and to evaluate success of management actions, such as for pollution control, and (3) high-quality data—to assure a publicly credible basis for controversial decisions.

These attributes are the most up to date available and define objectives more closely; in that respect they mark a considerable change from former general notions, such as data for the efficient management and planning of water projects, data for the evaluation and prediction of hazards, and data for the evaluation of the water resources.

**USERS**

Because of the lack of prescribed objectives from high authority, one seeks guidance from the principle that the data programs must serve users. The relationship is clear with respect to the captive network, the data network established and operated by the user, such as that laid out for a research project. The relationship is not clear with respect to the general purpose network, where the collectors are not identical with the users. There are several reasons for this separation, the main ones being the economy of specialization, the needs for
impartiality in the collection and reporting of data, and Herfindahl’s [1969, p. 197] comment that ‘Indeed, in a large government there are so many operating agencies eager to generate their own hydrologic data that a separate data agency might be useful simply to impose a desirable unity on the collection of this type of data.’

Data agencies have long familiarity with establishing networks of data in response to expressed needs, especially those accompanied by offers of funds. They have a long history of contacts with data users, public or private. The problems arise because data users are not sensitive to redundancies (as long as their needs are supplied), to long-term uses, or to neglect of needs other than their own. They evince little interest in schemes of information transfer or in measures of accuracy (preferring to assume that data are error free and do not therefore add to the many errors that may be entailed by their uses or applications).

Meetings and conferences among data collectors and representative data users have long been part of the national scene, but these have their limits as a sufficient tool in designing a network, largely because, as stated at the meeting (remarks by the session chairman, J. C. Schaake) users are not in fact always sufficiently perceptive to identify actual and prospective needs. They emphasize their ad hoc purposes, whereas data may serve many repeated diverse uses over time. Usually, the most searching critique about use of data originates with those who collect them.

A more direct signal from data users may be needed. As a step in that direction, the principles and standards of the Water Resources Council [1971] contain a section calling for formal tests of the significance of data in each water planning exercise. Sensitivity tests and the like were to provide user signals to the data programs. So far as I know, the section is not observed in practice.

Alternatively, it was suggested, for example, that all that is needed to get the right signals from the users is to make data marketable. Indeed, a market may be emerging. The contractors now engaged by the federal data services find it profitable to collect data for direct sale to specialized users—hence a market. The firm profits by repeated use of the same records, provided there is no ‘leakage.’ The problem here, one long recognized, is that the market forces give low value to long-term, continuous records of water quality and quantity. Moreover, as the example given by O’Connell et al. [1979] showed, user-dominated networks lead to redundancies.

Economics may be the bridge between the supplier and the users of data to show the value of particular networks to society. Several authors emphasized this role, but McGuire [1978] noted that economists do not appear to brood about the economic value of their own data systems. Economists might well direct their talents toward the possible relation between the needs for data created by alternate schemes of water management, for example, permits versus charges (for withdrawals of water and discharges of effluent) and regulation of river quantity and quality versus control of obstructions and returns.

**INFORMATION TRANSFER**

One of the most important subjects presented at the conference dealt with information transfer—spatial, temporal interchange between space and time and between different but related phenomena. These methods multiply the effectiveness of a data program. In fact, J. Stahl (in a panel discussion) noted that in Illinois, regional patterns of stream geometry, streamflow, precipitation, low flows, lake evaporation, and storm patterns have been developed so that regional estimates may be better than short-term data at individual sites. Methods of information transfer need to be examined and developed further. For example, we have not yet explored sufficiently the interrelations between surface water and groundwater—the use of groundwater observations to predict low flows, streamflow behavior to infer groundwater resources, and for that matter, inferences about low flows obtained from the concentration of conservative mineral constituents.

Even though it was observed at the conference that decision makers (political leaders) may not care much for answers from monitoring unless related to enforcement, this incredible world will ask some troublesome questions about the results of the large investments being made in water, air, and the rest of the environment. There will be questions about cause-effect and dose-response relations, the effectiveness of regulations upon water quality and upon human health, and the effectiveness of land regulations on flood damages and soil erosion. Answers to these kinds of questions will require cause-effect and dose-response models. The tools are exactly of the same kind that are part of information transfer.

One may anticipate, however, that such models may create demands for data of kinds not within the present programs. For example, these cause-effect models may require validation through observations on internal processes that are now inferred or simulated (for example, observations of soil moisture and groundwater levels now only inferred in rainfall-runoff modeling).

**EFFICIENCY OR EFFECTIVENESS**

In reporting on efficiency, several authors argued that because data costs are low in relation to other costs, one need not be concerned about tests of efficiency or cost effectiveness. However, that is hardly a prescription for lack of a design. The major problem that forces one to a consideration of efficiency or cost effectiveness is the potential of the cost of neglect, benign or not, to overcollect in various ways, to the neglect of observations that may prove to be of greater need. As stated at the meeting, data programs are usually found to be too ambitious and inadequate at the same time.

Of course, one may recognize that efficiency may be only one criterion in a mix. It may be necessary to relax efficiency in order to achieve other attributes, such as flexibility and contingency, which may in turn be provided through planned redundancy in a given parallel or related network.

As the hydrologic system departs more and more from the natural, through increased regulation of river flows, increased land use effects on quantity and quality of water, and increased use and management of groundwater, the data plan shifts more toward ‘metering-monitoring and surveillance’ rather than resource evaluation. This usually leads to operation-research type studies such as those explained by Ward and Loftis [1978]. One might expand these studies to examine the effects of the management or regulatory scheme on the data plan, as previously suggested.

**OPTIMIZATION**

It was frequently noted that network design need not necessarily be based on formal schemes of optimization, such as the minimum cost of attaining some accuracy standard. A design may be erected upon judgmental analyses (‘back-of-the-envelope’) to accommodate a mix of design criteria. What distin-
guishes a design from a mere classification plan is the explicit statement of the criteria used to make decisions about the composition of the network and to achieve the objectives. It should be noted that some of the more instructive papers at the conference were those that showed the relative contributions to the error variance of such factors as the number of observations, the frequency or length of the period of observations, and the nature of the model. This analysis provides insight as to tactics for improving the network even though an explicit optimum is not sought.

RECOMMENDATIONS

Although several papers illustrated actual or potential applications of the results of research, there was general understanding that these applications are still few in the general order of affairs. One researcher expressed the view that analysis of data networks found greater applicability during retrenchment than during periods of growth, when operational details were controlling. Unless greater use is found, network design may have little purpose other than to furnish learned journals with articles by researchers to be read by their colleagues. I suggest a line of effort that might yield some useful results before the subject entirely dissolves into abstract research.

Research on data transfer. Several papers employed tactics of transfer, but generally, they were submerged within some scheme of optimization. The results usually suggested that against more stations or longer periods of record, better models would yield the greater payout, since they provide a considerable multiplication of information from a data base. The possible techniques are as numerous as the scope of hydrologic science, from simple spatial interpolation and correlation to regression, physical models, simulation schemes, and so on, including relations with and among various species or series of data.

The general subject of information transfer deserves examination on its merits and not necessarily as a selected method embedded within a particular scheme of network design. Again, this inquiry should broaden the scope of network design because it confronts centrally the relations among the various hydrologic elements.

Information transfer is congruent with the need to define cause-effect relations and to predict trends. As water is increasingly affected by land use, water management, and water use and control, current statistics will have limited use as predictors of the future. It will be necessary to understand to what degree these reflect man or nature before one can properly estimate the future trends.

Broaden scope. The scope of the subject matter needs to be broadened, hydrologically as well as in techniques. Increased attention should be paid to groundwater and water quality, as well as other elements of the water environment. Techniques of network design should indeed use the best of statistical analysis, but more attention must be given to physical, chemical, and biological relations.

The long-standing familiarity with statistical analysis by the surface water community explains, in part, why most progress in network design has been in that field and not in groundwater, and even less in quality of water. Perhaps greater attention upon processes will automatically broaden use of the techniques. In any case, researchers should avoid increasing specialization upon popular lines of statistical analysis in one field, which lead away from simplicity—one of the messages conveyed at the conference.

Audit. While addressing ourselves to the design of new networks as a research objective, we might more profitably examine or audit existing networks to see how these fulfill their objectives as stated at their creation and the objectives as now perceived. We have been examining the error structure of proposed new schemes for networks before having a proper notion of how much error the present networks contain when applied to certain question, for example, the mean variance and other statistics of ungauged streams, the error in estimating the change over time in the regional (spatial) variance in some measure of water quality, and so on. These are proper questions even if they are not yet common. Such questions as the error of estimate in evaluating the effects of land use (non-point sources) on water quality may yet await answers to simpler queries. The point to be emphasized is that an audit is not research but an accounting of the results that one seeks to obtain from a data base. The little work that has been done in evaluating stream gaging and precipitation networks falls considerably short of the necessary job.

An outline of an audit follows:

1. Describe the network and its financing, and identify its uses (the latter as documented in actual operations, project reports, environmental impact statements, etc.).
2. Attempt to define the network objectives as perceived over time (flow frequency, water supply, pollution control, etc.).
3. Assess errors of estimate of parameters as applied in use (measurement, processing records, extrapolation errors, sampling, and model error).
4. Analyze results, efficiency of network, redundancies and gaps, and use of data transfer.
5. Study the implications of the network on public policy and the effectiveness of water programs. How valid are claims of the lack of data?

An audit may provide the practical and faster road to network design, since results may disclose redundancies, especially those that detract from getting more useful data, mismatches between current and prospective objectives, or errors too large to resolve the public policy questions that are asked.

There is more to be achieved from an audit. There was little reference at the conference to quality of data. Yet one must concur that increasing attention will be paid to this credibility as results become more marginal and controversial. While most attention will be paid to details of laboratory practices and protocols and less to field measurements and sampling. I surmise that the critical view might better include the data network itself as a scientific instrument. Credibility can be maintained only through frank assessment of the inherent errors and limitations. Error assessment is part of the information provided by a data network.

Data needs created by alternate schemes of water management. Particularly to be examined are the effects of management principles upon need for data. It is often noted that the free market minimizes the need for data, whereas regulated schemes dependent on regulatory tools, such as permits and charges, tend to multiply the need for data, especially when legally codified in such a way that questions of efficiency of data collection are irrelevant. Is this true? To what extent would alternate regulatory methods affect needs for data?

Technological coordination. At present, very considerable progress has been made toward operational coordination. Most parties concerned know who is collecting what and are aware of the standards to be applied. Data sources are in-
ded, catalogued, and warehoused. Most parties should have knowledge of the data available from other networks, thereby permitting them to take advantage of the efforts of others and to avoid obvious inefficiencies such as duplication. There is more to coordination than is shown by the work at the conference.

A first step in technological coordination would lead to the incorporation of the technical and hydrological relations among the various networks of data in order to increase their joint usefulness. Some of these have already been mentioned; for example, the information available from a groundwater level network might be incorporated into a streamflow network that is used to estimate low flows.

Further development would lead to the coordination among the data factors used in the several models used to extend, to simulate, or otherwise to apply the data base to various problems. (For example, where models are limited to simple regressions, then are the data among the several networks so distributed in time and place so as to sample factors in proportion to their effectiveness? Can the data for one model help another?)

A third element is a bit more complex, since it involves combining the first and second elements. In this exercise one seeks the opportunities for information transfer among the various networks as well as their uses in the several specific working models. There are still further stages in technological coordination that involve (1) the plans for networks that serve multiple objectives and (2) means for balancing or offsetting the work or utility of one kind of data against another kind. Formal tools for these processes are as yet in their formative stages, but in the meantime it is suggested that consideration of the first two elements, separately or jointly, would go far toward the purposes of network design in its entirety.

CONCLUSIONS

In short, the state of the art together with current research was well exposed at the conference. Serious questions were posed that reached deeper than the details of the several papers. These questions revealed the uncertainties that need to be resolved and the gaps in the research that need to be repaired if the research is to find significant applications. These deficiencies centered on the limited breadth of hydrologic subject matter and the limited range of the techniques used. Queries were raised about whether more data are always better, about the difficulties in defining the objectives of a data network, about the roles of optimization and simplicity, about the use of information transfer, about the definition of network design, and about why applications of research results have as yet been few.

The many different techniques that were espoused and discussed at the conference made it clear that network design does not involve a single formulation. The overriding message was also clear that some sort of systematic analysis is needed to assure that proliferating data networks achieve their scientific and social objectives. For this reason, recommendations were offered in the form of audits of existing data systems and of schemes (or plans) of technological coordination to supplement ongoing operational coordination.

REFERENCES


(Received December 18, 1978; revised March 30, 1979; accepted April 23, 1979.)