260 82 WO

WATER AND SANITATION FOR HEALTH PROJECT

Operated by
CDM and Associates

ponsored by the U.S. Agency
or International Development

1611 N. Kent Street, Room 1001
Arlington, VA 22209-2111 USA

Telephone: (703) 243-8200
Fax (703) 525-9137
Telex WUI 64552
Cable Address WASHAID

he WASH Project is managed y Camp Dresser & McKee nternational Inc. Principal ooperating institutions and ubcontractors are: Associates n Rural Development, Inc.; nternational Science and echnology Institute, Inc.; esearch Triangle Institute; raining Resources Group; Jniversity of North Carolina t Chapel Hill; University esearch Corporation.

A WORKSHOP FOR THE DESIGN OF LOW COST WATER SYSTEMS IN ECUADOR

WASH FIELD REPORT NO. 63

NOVEMBER 1982

Prepared For:
USAID Mission to the Republic of Ecuador
Order of Technical Direction No. 105

WASH FIELD REPORT NO. 63

ECUADOR

A WORKSHOP FOR THE DESIGN OF LOW COST WATER SYSTEMS IN ECUADOR

Prepared for the USAID Mission to the Republic of Ecuador
Under Order of Technical Direction No. 105

Prepared by:

Donald T. Lauria, Ph.D., P.E.

November 1982

Water and Sanitation for Health Project
Contract No. AID/DSPE-C-0080, Project No. 931-1176
is sponsored by the Office of Health, Bureau for Science and Technology
U.S. Agency for International Development
Washington, DC 20523

TABLE OF CONTENTS

Chapter 1

BACKGROUND

1.1 Initiation

The college of Civil Engineers of Guayaquil (CIG), a society of professional engineers in Ecuador, runs continuing education courses throughout the year for its members. This year, the CIG asked Professor Enrique LaMotta of the Escuela Politecnica Nacional in Quito to teach a short course on wastewater treatment. CIG also asked LaMotta to suggest a teacher for a short course on water supply planning. Knowing of the work of Professor Donald T. Lauria of the University of North Carolina (UNC) in this field, LaMotta suggested his name. CIG requested Lauria's participation through the AID Mission in Quito which in turn asked WASH to make arrangements. The initial request for Lauria's service was made in April 1982 with a proposed date for the course in August. Although the specific details were as yet uncertain, Lauria agreed to give the course.

1.2 Options

As planning developed, two broad options for the course were offered to WASH for consideration. One was to conduct a workshop that would provide participants with an opportunity to use the computer for low-cost water supply planning, and the other was to present a lecture series in which several topics would be covered but the audience would remain essentially passive.

With the workshop format, a two-hour lecture would be given on a particular subject after which the participants would be divided into groups of from five to ten persons. Each group would be assigned a task based on the subject of the lecture. The groups would spend about two hours on the assignment during which the teacher and his assistants would move from one group to another serving as consultants. The assignments would be based on case studies of the participants' choosing using local data, maps and reports. Participants would be asked to bring these materials with them, and the teachers would assist in turning them into manageable projects. After completion of the case studies, a lecture on the next subject would be presented and the process would be repeated. During the workshop, four or five different subjects would be covered, including both theory and applications.

The alternative to a workshop was a course consisting mainly of lectures. This would make it possible to cover more material, but learning would be less thorough because the participants would not be given an opportunity to apply the concepts. Courses of this type are most appropriate for introducing a range of subjects and for stimulating the participants to work on their own. Workshops, by comparison, are aimed at producing new skills which can be rather quickly applied.

Regarding the subject matter of the course, two alternatives were proposed. The first was concerned with the planning and design of facilities other than treatment plants. The title would be "Engineering Project Design" and the major subjects would include:

1. Cost functions
2. Staging of construction
3. Pumping stations
4. Networks
5. Reservoirs.

The alternative focus of the course was "Low-Cost Water Supply Planning" which would be focused on the design of municipal water facilities in developing countries, especially pipe distribution networks, which is the most expensive component. This course would be similar to numerous courses Lauria has given in LDCs for the World Bank. Major subjects for this course would include:

1. Cost functions
2. Branched networks
3. Looped networks
4. Financial feasibility.

The intention with both courses was to use the microcomputer. In the case of a workshop, the computer would be applied to the case studies on which the participants were working. In the case of a lecture series, it would be used for demonstrations. While it was hoped that microcomputers would be available in Ecuador, the possibility existed of bringing a computer from the United States. Because many participants were expected, Lauria recommended that he be assisted by at least one but preferably two doctoral students from his department.

1.3 Final Plan

After extensive discussion between the teachers and WASH, it was decided to adopt a workshop rather than a lecture series format. The "Engineering Project Design" title was selected rather than focusing almost exclusively on distribution networks. Approval was given for Lauria to be assisted by Mr. Keith Little who is a doctoral student working under his direction. In addition, Little has been supported by the World Bank through their contract with UNC to develop software for the microcomputer that would be used in the course. Because the prospects for follow-up work in Ecuador were uncertain, a second assistant was not approved. The teachers were authorized to bring their own microcomputer to Ecuador. In addition, CIG made arrangements for the local distributor of Apple computers to provide two of their machines and several monitor screens. The course was scheduled to start August 11 and end August 14.

Chapter 2

COURSE DESCRIPTION

2.1 Staff

Because of its heavy involvement in offering continuing education courses for its members, CIG has a knowledgeable staff for handling arrangements. The main coordinator for the course was Engineer Carlos Oporto C. who was assisted by Engineer Carlos Assemany A. The coordinators were responsible for liaison between the teachers, CIG, and assisting agencies. Messrs. Lauria and Little were assisted in the classroom by Professors LaMotta and Gonzalo Ordonez of the Escuela Politecnica Nacional. LaMotta's services had been formally arranged through the AID Mission in Quito. Both LaMotta and Ordonez worked as consultants with the teachers during the sessions in which the computer was applied to case studies. In addition, they periodically assisted with explanations and commentary on lectures.

2.2 Description

The course was attended by about 120 participants, all of whom were engineers, although some are not working in the environmental field. The course met six hours daily for the first three days and for about three hours the last day. A session was held each morning form 8:00 to 10:00 am at the university during which Lauria lectured in Spanish on a variety of subjects. The presentations were devoted to theoretical principles and concepts. Each evening from 6:00 to 10:00 pm, a second session was held in the Uni Hotel. The evenings were mainly devoted to developing case studies that provided opportunities to apply the theory to problems using the computer. For these sessions, the participants were arranged in groups of about 12 at tables, each of which was equipped with a monitor linked to a microcomputer.

2.3 Objectives

The overall goal of the course was to introduce concepts and techniques for the improved planning of the major components of community water systems in developing countries. The specific objectives were four-fold:

1. To provide basic instruction on and demonstrate the role of statistical regression analysis, microeconomic theory, mathematical modeling, and computer programming in the design of water systems;

2. To apply these concepts and techniques to real cases in Ecuador;

3. To provide an opportunity for the participants to work with mathematical models and the computer rather than to be merely observers;

4. To investigate interest in a technical assistance program for Ecuador in which these tools could be incorporated in the routine water supply planning work of the country.

2.4 Content

An outline of the subjects covered in the course is shown in Table 1. Most of the topics required use of computer programs for data analysis and optimization which had been developed at the University of North Carolina with research support from the World Bank. Instruction was given in the evening sessions on using these programs. A brief description of the major lectures given in the morning sessions is as follows.

2.4.1 Statistical Analysis

The purpose of the lecture was to show how mathematical equations can be fitted to cost data using linear regression analysis, especially equations with more than a single independent variable. The resulting equations can be used for predicting the costs of water systems. The lecture covered determination of goodness of fit using R2 and F statistics. A summary of the material covered is included in the User Instructions (see Appendix F.)

2.4.2 Economies of Scale

Once equations have been fitted to data, it is easy to determine the economies of scale of the various water system components. If power functions have been fitted, the exponent of the independent variable is a measure of cost elasticity with respect to capacity. For example, an exponent of 0.7 for a water treatment cost function (which is a typical value) implies that by doubling the capacity of a plant, construction costs will increase by only about 70 percent. Using local data, economies of scale were found to be large for pipelines and considerably smaller for treatment plants and pumping stations.

2.4.3 Optimal Design Periods

Components with large economies should be designed with more excess capacity than where economies are small. Optimal design periods were shown through the use of mathematical models to depend on economies of scale and the discount rate. The role of other factors on design periods was also discussed, such as the length of the planning period and the rate at which demand changes over time.

2.4.4 Branched Networks

Piped water distribution networks are generally the most expensive components in commuity water systems. Because of their expense, it is especially important in developing countries that networks be well designed and that costs be minimized. In cases where branched networks can be used (i.e., networks without closed circuits), it is possible to use linear programming for determining optimal pipe sizes. A mathematical model and computer program for designing branched networks using LP was described, as was the basic technique of linear programming. User instructions for the program are Appendix G.

2.4.5 Looped Networks

While branched networks can frequently be used in rural and small communities, they are often unacceptable in larger towns and cities where circuits must be closed. Unfortunately, linear programming cannot be directly used for the optimal design of looped networks. Rather, network design must depend on the trial-and-error use of Hardy Cross or Newton Rathson models, which were described. Also described was a heuristic programming approach for the nearly optimal design of networks that starts first by designing the primary pipes as a branched system followed by a sequence of rules for selecting the secondary links needed to close circuits.

2.4.6 Pumping Stations

Particular attention is usually needed for sizing of wet wells in pumping stations. The principles of wet well design are essentially equivalent to those for water storage tanks and distribution reservoirs. Models were developed for determining the size of wet wells that would minimize costs. These models considered both single and multiple pumping units of both equal and unequal size.

2.4.7 Reservoirs

Where they are needed, water storage reservoirs are a particularly costly component of municipal systems. The conventional approach for sizing reservoirs involves Rippl analysis, which unfortunately suffers from several limitations. A model was presented for the least-cost design of reservoirs using linear programming. The model can handle both single and multiple-purpose units. The heart of the model is a linear decision rule for operation. The model can accommodate probability constraints on demand. It is described in Appendix J.

2.5 Assessment

Arrangements for the course were excellent; CIG was well organized and thoroughly attended to details. In general, facilities were satisfactory. CIG required participants to check in at the beginning of each session to assure attendance; punched time cards were used for this purpose. Only those participants who attended all the sessions were eligible for continuing education credits. This practice worked well and is highly recommended. For the evening sessions, breaks were arranged with coffee and refreshments. CIG is skilled in running short courses, and their operation was very professional and of the highest caliber.

The duration and arrangement of the course were adequate, although six hours of classroom teaching per day borders on the excessive. An arrangement of two hours of theory in the morning followed by three hours of practice in the evening for four and one-half days might have been better. A national holiday midweek prevented this schedule from being adopted.

Professors Enrique LaMotta and Gonzalo Ordonez helped explain some particularly complicated matters and were very useful as consultants during the time the students worked on their case studies. This arrangement worked well and is recommended for future courses. Mr. Little was taxed in trying to teach use of the computer to such a large group. Despite the fact that a few computers were available for the course, he could work with only two of them which limited the rate at which case studies could be solved. A better arrangement would have been to include two doctoral students who could have spent more time with the participants demonstrating computer software and enabling investigation of more alternatives.

There were too many participants in this course for thorough learning in a workshop format. With the objective of hands-on experience in using the computer, it would have been better to limit attendance to 40 persons. With over 100 participants, it was necessary to rely more on demonstrations than solution of cases. However, each participant had the opportunity to design at least one real system using local data.

The maps and basic data supplied by CIG for the case studies were excellent. However, the cost data for water system components were only fair, which made it difficult in some cases to perform regression analyses. This in turn interfered with determining optimal design periods.

Some problems were encountered in making the Apple computers operate satisfactorily. It turned out to be essential that a computer had been brought to the course from the United States.

The course was well received. By the time it ended, the President of CIG asked about steps to provide further instruction in the subjects that had been introduced. Actually, CIG offered to buy the computer that had been used in the course to make it available for its members. The counterpart of CIG in Quito is similarly interested in a course of this type for its own program of continuing education. Finally, interest was expressed by some of the faculty members at the Escuela Politecnica Nacional in Quito for follow-up training in these subjects.

Chapter 3

RECOMMENDATIONS

A course of the type held in Guayaquil only makes engineers aware of new concepts and techniques; it cannot enable the engineers to use these techniques routinely in their work. A goal for the future would be to develop a technical assistance program that would provide practicing engineers with a thorough understanding of the concepts and techniques (including skills for using the computer programs of this course) that are necessary to integrate these approaches into operational planning and design situation. That a demand for such assistance exists was demonstrated by the strong response to the Guayaquil course.

In developing any additional courses one should be guided by the following concepts:

- The next stage of tranferring the technique should be focused on helping practicing engineers in operational agencies to integrate these practices into their daily work. This is a long-term effort in which present day engineers need to be retrained and new engineers need to be indoctrinated in the techniques. The availability of local resources such as computers, etc. will be a major element in the direction of this effort.

- Any future courses should be organized so that one or more of the local professors who attended the Guayaquil course (i.e. LaMotta or Ordonez) would be the organizer as well as an instructor. Professor Lauria's role would be that of a principal instructor and a course advisor. To do this Lauria should work with the local professors to adapt the existing software package to such computing capability as is available in Ecuador.

- The next courses should be sponsored by one or more of the operating agencies (for example: IEOS, Quito Water, etc.). Carefully prepared case studies of typical often-used local problems should be used as the main teaching tool.

- the local universities should try to include the "Ecuadorian" version of these concept into their engineering courses. Professors LaMotta and Ordonez should be encouraged to take a leading role in helping to adapt the concepts to Ecuadorian human, financial, technical and organizational resources.

The discussion that follows focuses on the next steps of how USAID/Ecuador should respond to the request for a course in Quito.

It is important to respond to the request for a course in Quito since the request demonstrates interest in improving water supply planning to meet the goals of the UN Decade, but local engineers feel that their skills need to be enhanced for them to be adequate to this task. In discussions with the AID Mission and Engineer LaMotta, it appears that an appropriate time for this course might be January 1983. Either in connection with the course or as a separate exercise, it would be important to provide in-depth training to a small group of engineers to make them thoroughly familiar with these techniques. This group would be expected to use this technology in their work, and they in turn would assume the responsibility of providing similar training for their colleagues. At this time, it is not entirely clear who should receive the training, but based on discussions with the AID Mission, it appears that the most likely group would be the national water supply planning agency, IEOS.

IEOS employs a large number of sanitary engineers for the planning and design of water systems. At present, this agency has no computers but instead performs all design computations by hand. The consequence of this is that engineers are limited to investigating relatively few design alternatives, making it difficult to produce least-cost designs and to tailor them to the affordability of users. The AID Mission at present has given a $6 million loan to IEOS which includes three components, viz. project implementation, training, and equipment. By channeling technical assistance through IEOS, it is likely that an excellent opportunity would exist to apply the computer techniques to the project implementation component of this loan.

In addition to working directly with an agency such as IEOS, it would be desirable to employ the assistance of the Escuela Politecnica Nacional. By involving some professors in this technical assistance project, they would become a more valuable resource capable of responding to requests for assistance in the future. In addition, they would be able to teach these concepts and techniques to students in university courses thereby strengthening their preparation for employment in the water supply sector.

At this point, there appear to be three options for the proposed course in January: (1) offer the course for the civil engineering professional society, as was done in Guayaquil; (2) offer the course exclusively for IEOS; and (3) offer training to both the civil engineering group and IEOS.

An advantage of the first option is that it provides broad exposure to concepts and techniques for the public and the private sectors, both of which are engaged in water supply planning. It also establishes firm contact with the professional society that is likely to play an important role in the water field. A disadvantage is that a course by itself without plans for follow-up will not necessarily make a significant step toward long-term improvement of planning.

The second option is appealing on the grounds that IEOS is the agency that is fundamentally responsible for community water supply, and by training its engineers, positive steps would be taken to strengthen the institution and significantly affect the water sector. Also, as noted above, the AID loan to IEOS would appear to be an excellent vehicle for assuring that the tools and techniques of this assistance find their way into practice. A disadvantage is that consulting engineers and others not employed by IEOS would be excluded from this training.

The third option is a compromise that attempts to combine the first two. Under it, a course similar to the one in Guayaquil would be given to engineers working in both the public and private sectors, and this would be followed by a few days of intense training to a select group in IEOS who would be expected to take the leadership in learning these techniques, applying them routinely in their work, and teaching them to their colleagues. This option would lay the foundation for long-term assistance and would provide considerable flexibility in how the assistance might be channelled. Without considerably more information on how the water sector is organized in Ecuador (which is essential for choosing among the three options), this plan appears to be the most appropriate.

Assuming the third option, work should be started to arrange for a follow-up course in Quito, probably in January. The number of participants should probably be limited to a maximum of 20, all of whom should be sanitary engineers engaged in water supply planning and design. The teaching staff should include help from Professors LaMotta and Ordonez, and it should involve two doctoral students from UNC to assist with computer applications. During the visit for this course, a definite strategy should be developed for providing technical assistance in low-cost water supply planning. Methods for exchange of information and interrelationships between IEOS and private consulting engineers should be determined.

For the course in Quito, it is recommended that AID purchase one or two microcomputers that could later be made available to local engineers and possibly used in connection with its loan to IEOS. UNC is prepared to assist in purchasing this equipment. The U.S. cost of a well equipped microcomputer is roughly $6,000, although an Apple II can be bought for about half this amount.

WATER AND SANITATION FOR HEALTH (WASH) PROJECT
ORDER OF TECHNICAL DIRECTION (OTD) NUMBER 105
July 18, 1982

TO: Dr. Dennis Warner, Ph. D., P.E.
WASH Contract Project Director

FROM: Mr. Victor W. R. Wehman Jr., P.E., R.S. VWW
A.I.D. WASH Project Manager
A.I.D./S&T/H/WS

SUBJECT: Provision of Technical Assistance Under WASH Project Scope of Work for USAID/Ecuador

REFERENCES:
A) QUITO 2443, 7 April 82
B) Memo Olinger (PRE/H)/Wehman (S&T/H), 21 April 82
C) WASH telex no. 207, 27 April 82
D) QUITO 3341, 12 May 82

1. WASH contractor requested to provide technical assistance to USAID/Ecuador as per Ref. A, para. 2, 3 and 4.

2. WASH contractor/subcontractor/consultants authorized to expend up to 28 person days of effort over a 3 month period to accomplish this technical assistance effort.

3, Contractor authorized up to 21 person days of international and/or domestic per diem to accomplish this effort.

4. Contractor to coordinate with LAC/DR/HN (P. Feeney), LAC/DR/ENGR (Rod MacDonald), and Ecuador desk officer (R. Lindsey) and should provide copies of this OTD along with periodic progress reports as requested by S&T/H or LAC bureau personnel.

4. Contractor authorized to provide up to two (2) international round trips from consultants home base through Washington (for briefing and demonstration of software to WASH CIC staff) to Ecuador (Quito and Guayaquil) and return to home base through Washington D.C. (WASH CIC for debriefing) during life of this OTD.

6. Contractor authorized local travel within Ecuador as necessary and appropriate to meet mission needs NTE $400 without prior written approval of AID WASH Project Manager.

7. Contractor authorized to obtain local secretarial, facilitator, professional, graphics, or reproduction services in Ecuador as necessary to accomplish tasks. These services are in addition to the level of effort specified in para. 2 and para. 3 above NTE $1,200 without prior written approval of AID WASH Project Manager.

8. Contractor authorized to expend up to $1,600 for training materials, software demos, supplies, workshop materials, and/or print/support services associated with workshop.

9. Contractor authorized to provide for car or vehicle rental in Ecuador, if necessary, to facilitate effort. Mission is encouraged to provide mission vehicles, if available.

10. Contractor will insure availability of appropriate Apple micro-computer equipment, accessories, software, and training aids for workshop needs. Appropriate shipment as carry-on excess baggage is authorized to facilitate air movement from U.S. to Ecuador if local Ecuadorian Apple micro-computer equipment not available for contractor team and timing.

11. Contractor will insure that a detailed workshop training agenda/curriculum is provided to WASH CIC along with a copy of all pertinent software/ documentation before consultants authorized to travel to Ecuador. These materials should be specifically referred as primary references in WASH Project Official OTD file.

12. Contractor's consultants are authorized to lecture, or to operate Apple micro-computer and various software packages in either demonstrations or training sessions with the Ecuadorian workshop attendees.

13. WASH contractor will adhere to normal established administrative and financial controls as established for WASH mechanism in WASH contract.

14. WASH contractor should definitely be prepared to administratively or technically backstop field consultants and subcontractors. Contractor should secure cable commitment from USAID/Ecuadorindicating that the USAID is prepared to insure and expedite customs clearance in and out of Ecuador of all micro-computer carry-on baggage and associated software. S&T/H/WS WASH Project Manager will then allow consultants to enter into international travel for the purprose of the scope of work defined in Ref. A.

15. Contractor to provide final workshop report within 30 days of return to U.S. with observations, discussions, recommendations, and conslusions.

16. Mission should be contacted immediately and technical assistance initiated as soon as convenient to USAID/Ecuador and GOE.

17. Appreciate your prompt attention to this matter. Good Luck.

Department of State

INCOMING TELEGRAM

QUITO 02443 072209Z 3170 061289 AID2398

ACTION AID-35

ACTION OFFICE LASA-03
INFO LADP-03 LADR-03 PPCE-01 PDPR-01 PPPB-03 HO-04 AAST-01
STUD-01 ENGR-02 HLG-01 HUD-02 RELO-01 DAEN-01 MAST-01
/028 A4

INFO OCT-00 ARA-16 AMAD-01 /CRE-100 W
------------------035160 072221Z /38

R 071726Z APR 82
FM AMEMBASSY QUITO
TO SECSTATE WASHDC 3910

UNCLAS QUITO 2443

AIDAC

E.O. 12065: N/A
SUBJECT: WAS PROJECT ASSISTANCE

1. MISSION HAS RECEIVED A REQUEST FROM THE CIVIL ENGINEERING SOCIETY OF THE PROVINCE OF GUAYAS TO PROVIDE ASSISTANCE IN THE PRESENTATION OF ONE OF ITS EIGHT WEEK-LONG SEMINARS. BASED ON MISSION URBAN DEVELOPMENT OFFICER MILLER CONVERSATIONS IN JANUARY IN WASHINGTON WITH PRE/H.OLINGER AND WASH JIM BEVERLY, MISSION REQUESTS THE WASH PROJECT AS THE SOURCE OF THIS ASSISTANCE;

2. AS PART OF ITS PROFESSIONAL DEVELOPMENT PROGRAM, THE SOCIETY HAS SCHEDULED A TRAINING SEMINAR FOR THE WEEK OF AUGUST 11-17 ON THE SUBJECT OF "COST EFFICIENCY IN THE DESIGN AND IMPLEMENTATION OF SANITARY ENGINEERING FACILITIES IN DEVELOPING COUNTRIES." THE SOCIETY HAS SPECIFICALLY REQUESTED THE PARTICIPATION OF A DR. DONALD T. LAURIA FROM THE DEPARTMENT OF ENVIRONMENTAL SCIENCE AND ENGINEERING OF THE UNIVERSITY OF NORTH CAROLINA IN CHAPEL HILL. WE UNDERSTAND DR. LAURIA IS AVAILABLE THROUGH WASH, THAT AN ECUADOREAN COLLEAGUE OF DR. LAURIA HAS RECOMMENDED HIM, AND THAT THE SOCIETY WOULD BE PARTICULARLY PLEASED IF DR. LAURIA COULD CONDUCT THE SEMINAR.

3. THE SOCIETY IS A LARGE, PROFESSIONAL ASSOCIATION WHOSE MEMBER ENGINEERS ARE ACTIVE IN ALL PHASES OF THE CONSTRUCTION BUSINESS THROUGHOUT THE COUNTRY, INCLUDING LOW-COST HOUSING PROJECTS OF GOE HOUSING INSTITUTIONS. IN THE CONTEXT OF USAID'S STRATEGIES OF TECHNOLOGY TRANSFER, INTEGRATED URBAN DEVELOPMENT, AND SUPPORT FOR THE PRIVATE SECTOR, MISSION WOULD LIKE TO BE RESPONSIVE TO THE SOCIETY. CONSEQUENTLY, MISSION REQUESTS AID/W PURSUIT OF WASH SUPPORT AND THE PARTICIPATION OF DR. LAURIA.

4. THE PARTICIPANTS IN THE SEMINAR WOULD BE APPROXIMATELY 200 VICIL ENGINEERS FROM THE GUAYACUIL AREA. THE LOCALE AND LOGISTICS WOULD BE PROVIDED BY THE SOCIETY, INCLUDING REPRODUCTION OF COURSE MATERIAL. THE TRAINER WOULD BE EXPECTED TO CONDUCT SEMINARS OF FOUR HOURS A DAY ON THE SUBJECT OF EFFICIENT SANITARY ENGINEERING DESIGN AND TECHNOLOGIES. DR. LAURIA APPARENTLY HAS GIVEN SUCH AS COURSE AND SHOULD BE CONTACTED FOR FURTHER INFORMATION ON SPECIFICS OF TOPICS TO BE COVERED.

5. MISSION IS NOT FAMILIAR WITH DR. LAURIA, BUT BASED ON THE SOCIETY'S RECOMMENDATION, WOULD BE SAITSFIED WITH HIS PARTICIPATION. IF HE IS NOT AVAILABLE, A SIMILARLY QUALIFIED SPANISH SPEAKING EXPERT WOULD BE WELCOME.

6. PLEASE ADVISE AS SOON AS POSSIBLE.
YOHLE

UNITED STATES INTERNATIONAL DEVELOPMENT COOPERATION AGENCY
AGENCY FOR INTERNATIONAL DEVELOPMENT
WASHINGTON D.C. 20523

April 21, 1982

MEMORANDUM

TO: S&T/HEA, Victor Wehman

FROM: PRE/H, David Olinger

SUBJECT: WASH Technical Assistance to USAID/Quito

Attached hereto is a telegram from USAID/Quito requesting WASH assistance managing a week long seminar on cost efficiency in design and implementation of sanitary engineering facilities.

The services of Dr. Don Lauria of the University of North Carolina are specifically requested since his work is known to the sponsoring Civil Engineering Society and he has worked on WASH assignments previously.

In a phone conversation today, John Miller, Housing and Urban Development Officer in Quito requested an early indication as to whether it will be possible to meet this request, since the Civil Engineering group is quite eager to finalize their schedule. I have, therefore, taken the liberty of discussing this matter with David Donaldson of WASH.

Although the dates proposed are in mid-August, an early response as to the availibility of Dr. Lauria would be appreciated by all concerned.

Attachment: as stated

cc: USAID/Ecuador, J. Miller
WASH, D. Donaldson
PRE/H, P. Vitale

ACTION COPY

UNCLASSIFIED

Department of State

INCOMING TELEGRAM

PAGE Ø1 QUITO Ø3341 171543Z 5332 Ø93356 AID51Ø7
ACTION AID-35

ACTION OFFICE STHE-Ø1
INFO LASA-Ø3 LADP-Ø3 LADR-Ø3 PPCE-Ø1 PDPR-Ø1 PPPB-Ø3 PPEA-Ø1
HO-Ø4 AAST-Ø1 ENGR-Ø2 RELO-Ø1 TELE-Ø1 MAST-Ø1
/Ø26 A1 417

INFO OCT-ØØ INR-1Ø EB-Ø8 ARA-16 AMAD-Ø1 CCO-ØØ PASO-ØØ
/Ø7Ø W
------------------ØØ117Ø 171612Z /51 38

R 121954Z MAY 82
FM AMEMBASSY QUITO
TO SECSTATE WASHDC 43Ø4

UNCLAS QUITO 3341

AIDAC

DIRECT RELAY

C O R R E C T E D C O P Y (FOR DIRECT RELAY, PASSING)

FOR: WEHMAN, ST/HEA CWSS; OLINGER, PRE/H; VITALE, PRE/H

E. O. 12Ø65: N/A
SUBJECT: TELEGRAM RELAY

TO: DONALDSON
WASH COORDINATION AND INFORMATION CENTER
1611 N. KENT STREET, ROOM 1ØØ2
ARLINGTON, VIRGINIA 222Ø9
(PHONE 7Ø3 (243-82ØØ))

REF: COMMERCIAL TELEX 4/27/82

1. SUBJECT: YOUR CABLE 2Ø7, RE DR. LAURIA.

2. MISSION HAS DISCUSSED WITH GUAYAS ENGINEERING SOCIETY THE POSSIBILITY OF MOVING THE DATES OF LAURIA'S COURSE FROM AUGUST 11-17. UNFORTUNATELY, DUE TO PREVIOUS SOCIETY COMMITMENTS, IT IS NOT POSSIBLE. MISSION THEREFORE RECONFIRMS OUR REQUEST FOR AUGUST 11-17 AND ASKS WASH PROJECT FOR CONFIRMATION.

3. LAURIA SHOULD PLAN TO MEET WITH MISSION IN QUITO PRIOR TO AND AFTER COURSE. FOR YOUR INFORMATION, AUGUST 1Ø IS AN ECUADOREAN HOLIDAY SO THAT LAURIA SHOULD MAKE TRAVEL ARRANGEMENTS TO BE IN QUITO AUGUST 9.

4. PLEASE SEND COURSE OUTLINE AND LAURIA RESUME, IN SPANISH. TO JOHN MILLER, USAID, QUITO, APO MIAMI 34Ø39.
YOULE

NOTE BY OC/T: PASSED ABOVE ADDRESSEE.

Received ST/H (Wehman) 5-21-82
Passed to WASH 5-21-82

UNCLASSIFIED

APPENDIX B

Itinerary

Date	Day	Place
10 August	Tuesday	Chapel Hill to Guayaquil
15 August	Sunday	Guayaquil to Quito
17 August	Tuesday	Quito to Chapel Hill

APPENDIX C

Officials Contacted

Colegio de Ingenieros del Guayas

Carlos Balladares V., President IX Jornadas
Pablo Baquerizo N., President CIG
Carlos Oporto C., Course Coordinator

AID-Quito

Herbert Caudill, Jr., Coordinator AID-IEOS
Kenneth R. Farr, Chief, Office of Health

Escuela Politecnica Nacional

Enrique LaMotta, Professor Environmental Engineering
Gonzalo A. Ordonez, Professor Civil Engineering

APPENDIX D

Course Outline
Guayaquil, Ecuador

Wednesday, August 11, 1982

Morning (2 hrs.)

Analysis of Cost Data
Regression Analysis
Determination/Interpretation Economies of Scale

Evening (4 hrs.)

Regression Analysis of Data
Optimal Design Periods

Thursday, August 12, 1982

Morning (2 hrs.)

Design of Branched Pipe Networks
Linear Programming

Evening (4 hrs.)

Use of Computer for Designing Branched Networks

Friday, August 13, 1982

Morning (2 hrs.)

Design of Looped Pipe Networks
Design of Pumping Stations

Evening (4 hrs.)

Use of Computer for Designing Looped Networks

Saturday, August 14, 1982

Morning (3 hrs.)

Design of Multipurpose Reservoirs
Illustrative Reservoir Example

JOURNAL OF
THE ENVIRONMENTAL
ENGINEERING DIVISION

Models for Capacity Planning of Water Systems

By Donald T. Lauria,[1] M. ASCE, Donald L. Schlenger,[2] and Roland W. Wentworth,[3] A. M. ASCE

APPENDIX E

Introduction

It is common practice in the field of sanitary engineering for designers to provide capacity in water systems beyond that needed for satisfying immediate requirements. Facilities are usually sized with sufficient capacity to meet anticipated flows several years in the future. The period during which facilities are expected to have excess capacity is called the design period which is sometimes as short as 10 yr for treatment plants and as long as 50 yr or more for pipelines and conduits.

Sanitary engineers are well equipped with standards for selecting design periods. In the past 10 yr–15 yr, however, these standards have been seriously questioned, and a number of mathematical models have been developed in the search for more nearly optimal designs. The underlying concept of these models is that the amount of excess capacity to be provided is a function of the tension between economies of scale and social time preference as reflected by the discount rate. On one hand, economies of scale make it attractive to build beyond immediate needs as incremental costs are proportionately small; on the other, society is disinclined to tie up valuable resources in facilities that remain unproductive for long periods of time. The models reveal that only after careful consideration of these two factors can proper design periods be selected.

Among sanitary engineers, Lynn (5) was one of the first to address the problem of optimal scale. His work was preceded, however, by Chenery (1) who developed a simple model for determining the optimal scale of capacity expansions. Chenery's model was refined and extended by Manne (6) whose work has received

Note.—Discussion open until September 1, 1977. To extend the closing date one month, a written request must be filed with the Editor of Technical Publications, ASCE. This paper is part of the copyrighted Journal of the Environmental Engineering Division, Proceedings of the American Society of Civil Engineers, Vol. 103, No. EE2, April, 1977. Manuscript was submitted for review for possible publication on February 12, 1976.

[1] Assoc. Prof., Dept. of Environmental Sci. and Engrg., Univ. of North Carolina, Chapel Hill, N.C.

[2] Special Projects Engr., Hackensack Water Co., Weehawken, N.J.

[3] Environmental Engr., Camp, Dresser and McKee, Inc., Boston, Mass.

much attention from sanitary engineers. Indeed, Muhich (8), Rachford, et al. (9), Scarato (12), and Thomas (13) all apply Manne's model or their extensions of it to the water engineering field: recently, the model has appeared in a basic sanitary engineering text by Rich (10). Outside of sanitary engineering, Manne's model has generated considerable interest; two of the many texts in which it appears are those by Rudd, et al. (11) and Zimmerman, et al. (14).

A basic problem with Manne's model is that the mathematical expression for the optimal design period is an implicit function. In order to calculate optimal capacities, trial and error or numerical techniques are necessary. Because this seriously limits the usefulness of the model, an approximating equation is presented in this paper by which optimal design periods can be calculated directly. A second approximating equation is presented herein for the optimality condition developed by Thomas (13) in an extension of Manne's model. Whereas Manne's work is limited to capacity expansions, Thomas is concerned with the optimal scale of a project for which the level of demand exceeds the capacity of supply facilities at the beginning of the planning horizon. Thomas' model, also presented by Rudd, et al. (11), is referred to in this paper as the *Initial Deficit Model*.

The main thrust of this paper is an extension of the Manne and Thomas models for a type of problem commonly encountered in water engineering. A model is developed for determining the optimal waiting period prior to construction of a system for which an unsatisfied demand exists at the beginning of the planning horizon. This is called the Waiting Period Model, and like the models of Manne and Thomas, the optimality condition is an implicit function that cannot be solved directly for the decision variable. Consequently, an approximating equation is presented. Although the Waiting Period Model is applicable to situations in the United States, it is perhaps more useful for water supply planning in developing countries.

The appeal of Manne's model is its simplicity, stripping away peripheral considerations in order to focus on the essential elements of the problem. However, this results in a certain oversimplification that makes the model of questionable value for real planning purposes. For example, demand is assumed to increase linearly with time, facilities are assumed to have infinite economic life, costs are assumed to remain constant over time, and the planning horizon is assumed to be infinite. Inasmuch as these assumptions apply to all the models of this paper, questions of applicability can be raised. Manne, however, has responded to many of the criticisms of his assumptions in Ref. 6, to which the reader is referred; this defense is not repeated herein.

Capacity Expansion Model

In this and the following section, only brief summaries of the models are presented. For more detail, the reader should see Refs. 7 and 13.

Manne's Capacity Expansion Model assumes that demand increases linearly with time t into the indefinite future, as shown in Fig. 1; the annual rate of demand increase is D, which has typical units of millions of U.S. gallons per day per year. At the beginning of the planning horizon (when $t = 0$), the capacity of supply facilities is assumed to equal the rate of demand. If it is required that capacity never be less than demand, the first expansion is needed when $t = 0$. Assuming an excess capacity period (i.e., a design period) of x yr, the expansion will have capacity of xD mgd, and its cost will be $C(xD)$ dollars.

At time $t = x$, excess capacity of the first expansion will be exhausted; by then, demand will have grown equal to capacity. With conditions at $t = x$ essentially identical to those at $t = 0$ (i.e., demand increasing linearly to an infinite time horizon and construction costs and the discount rate unchanged), another expansion of scale xD with cost $C(xD)$ will be required. Repeating this pattern for each point of zero excess capacity, the following expression of total present value construction cost for the infinite planning horizon can be written as

$$C(xD)\left[\sum_{n=0}^{\infty} \exp(-rnx)\right] \quad \text{(1)}$$

in which $\exp(-rnx)$ = present worth factor for year nx; and r = annual discount rate expressed as a decimal. In this and succeeding expressions of present value costs, continuous rather than discrete discounting is assumed to facilitate use of the calculus in deriving optimality conditions.

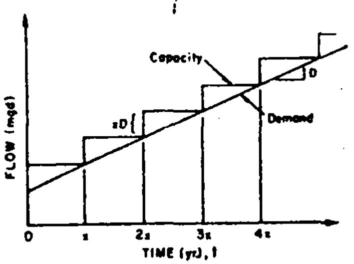


FIG. 1.—Capacity Expansion Model (1,000,000 gal/day = 37,800 m³/day)

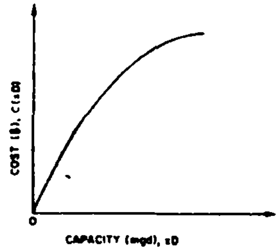


FIG. 2.—Typical Construction Cost Function (1,000,000 gal/day = 37,800 m³/day)

The term in brackets in Eq. 1 is the sum of a geometric progression that has the value of

$$\sum_{n=0}^{\infty} \exp(-rnx) = \frac{1}{1 - \exp(rx)} \quad \text{(2)}$$

In the water engineering field, the expansion cost in Eq. 1 commonly is a concave power function of the form shown in Fig. 2. The equation of this function is

$$C(xD) = k(xD)^a \quad \text{(3)}$$

in which k and a = constants, the latter called the economy of scale factor; k = cost of a 1,000,000-gal/day system, seen by substituting $xD = 1$ into Eq. 3; and a = percentage change in cost per percent change in scale, or equivalently, the ratio of marginal to average cost. Substituting Eqs. 2 and 3 into Eq. 1 results in Eq. 4, an expression of total present value cost in terms of excess capacity period x (the decision variable) and parameters k, D, r, and a:

$$\frac{k(xD)^a}{1 - \exp(-rx)} \quad \ldots \quad (4)$$

To find the optimal design period, x^*, that minimizes total present value cost,

TABLE 1.—Comparison of x^* Values from Eqs. 5 and 6

Discount rate, r (1)	Economy-of-scale factor, a (2)	x^* from numerical solution Eq. 5 (3)	x^* from approximating Eq. 6 (4)	Error as a percentage (5)
0.05	0.5	25.128	23.925	4.79
	0.6	18.948	18.634	1.66
	0.7	13.509	13.501	0.06
	0.8	8.616	8.573	0.49
	0.9	4.142	3.945	4.77
0.10	0.5	12.564	11.962	4.79
	0.6	9.474	9.317	1.66
	0.7	6.754	6.751	0.05
	0.8	4.308	4.287	0.49
	0.9	2.071	1.972	4.77
0.15	0.5	8.376	7.975	4.79
	0.6	6.316	6.211	1.66
	0.7	4.503	4.500	0.06
	0.8	2.872	2.858	0.49
	0.9	1.380	1.315	4.72
0.20	0.5	6.282	5.981	4.79
	0.6	4.737	4.659	1.66
	0.7	3.377	3.375	0.05
	0.8	2.154	2.143	0.49
	0.9	1.035	0.986	4.72

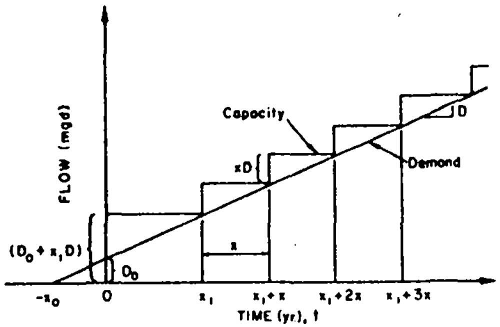


FIG. 3.—Initial Deficit Model (1,000,000 gal/day = 37,800 m³/day)

the derivative of Eq. 4 with respect to x is set equal to zero; the resulting optimality condition is

$$a = \frac{rx^*}{\exp(rx^*) - 1} \quad \ldots \quad (5)$$

Note that throughout this paper optimal values are marked with an asterisk.

Eq. 5 shows that the optimal design period depends only on the economy of scale factor a and the annual discount rate r under the assumptions of this model. Given values for each, Eq. 5 reduces to an expression containing the single unknown, x^*. It is impossible to solve Eq. 5 explicitly for x^*; trial-and-error methods, or preferably numerical techniques such as Newton's method, must be used. To avoid this necessity, cross plots of Eq. 5 have been reported by Manne (6), the first writer (4), Rachford, et al. (9), and others for selected values of r and a. More useful, however, is the approximation

$$x^* = \frac{2.6(1 - a)^{1.12}}{r} \quad \ldots \quad (6)$$

In Eq. 6, the parameters, 2.6 and 1.12, were obtained using linear regression techniques. In this analysis, Eq. 5 was solved for 20 values of x^* using Newton's method with a-values of 0.5, 0.6, 0.7, 0.8, and 0.9 and r-values of 0.05, 0.10, 0.15, and 0.20. The stopping criterion for Newton's method was 10^{-3} yr. The values of a and r were selected as being particularly relevant in the field of sanitary engineering.

The form of the model proposed for regression analysis derived from Eq. 5 in which a series expansion was made of exp (rx^*). Additionally, it was known that $x^* = 0$ when economies of scale are absent (i.e., when $a = 1$), which suggested inclusion of $(1 - a)$ as an independent variable in the regression model.

The standard error of estimate based on the 20 residuals obtained from the difference between x^* values from Eqs. 5 and 6 is 0.34 yr, and the correlation coefficient is 0.998. The excellent agreement between values of x^* from Eq. 5 and approximation from Eq. 6 is shown in Table 1.

Initial Deficit Model

The previous model is for expansions only; it assumes that demand and capacity are equal at the beginning of the planning horizon. It is common, however, for demand to exceed capacity at the start of planning, giving rise to the Initial Deficit Model of this section. This situation exists when users switch from an existing supply facility to an entirely new one, when a new demand suddenly presents itself, or when supply facilities simply have not been provided to satisfy demands. Examples of initial deficit include the installation of treatment plants for housing developments formerly served by septic tanks, the construction of water and wastewater systems for new towns, and the abandonment of local community facilities in favor of regional systems. In most of the towns of developing countries and in many places in the United States where public water and wastewater systems are lacking, capacity deficits are encountered when planning to provide these facilities is finally begun.

The Initial Deficit Model of this section retains the assumptions of the previous model, but in addition assumes that D_o, the rate of demand at $t = 0$, is unsatisfied. As shown in Fig. 3, the project to be constructed at the beginning of the planning horizon will have excess capacity for x_1 yr, at the end of which time a planning situation identical to that of the Capacity Expansion Model will be encountered.

The planning problem is to determine the optimal value of the initial design period, x_1^*, the methodology for which has been developed by Thomas (13).

Thomas' approach is similar to Manne's. An expression is written for total present value cost which includes the construction cost of the initial project (the first term in Eq. 7) plus the present value cost of an infinite number of future expansions (the second term in Eq. 7). The resulting objective function is

$$k(D_o + x_1 D)^a + \exp(-rx_1)\frac{k(xD)^a}{1 - \exp(-rx)} \quad \ldots\ldots (7)$$

Note from Fig. 3 that D_o can be replaced by $x_o D$, in which x_o = hypothetical elapsed period, in years, from the point of zero excess capacity prior to the start of planning to $t = 0$, assuming demand has increased at rate D.

In this model, the two decision variables are x_1 and x. The optimal value of x, the design period for expansions, is found from the derivative of Eq. 7 with respect to x set equal to zero; the optimality expression is identical to Eq. 5 for which approximating Eq. 6 can be used. The optimal value of x_1, the excess capacity of the initial project to be constructed at $t = 0$, results from the derivative of Eq. 7 with respect to x_1 set equal to zero. The optimality condition is

$$D\,a[k(D_o + x_1^* D)^{a-1}] = \exp(-rx_1^*)\frac{r\,k(xD)^a}{1 - \exp(-rx)} \quad \ldots\ldots (8)$$

The interpretation of this expression is as follows. The term in brackets on the left-hand side is the average cost of the optimally scaled initial project. Previously it was mentioned that a is the ratio of marginal to average cost, from which it follows that a times the term in brackets [ ] is the marginal cost of the initial project. Multiplying the product a[ ] by D, the amount by which demand increases every year, results in the approximate incremental cost of adding an extra year's capacity to the optimally scaled initial project. In mathematical symbols, the left side of Eq. 8 is roughly equal to $k[D_o + (x_1^* + 1)D]^a - k[D_o + x_1^* D]^a$.

On the right-hand side, $r\,k(xD)^a$ = annual interest cost of any future expansion. The denominator is the same present worth factor of Eq. 4; it discounts the infinite series of annual interest costs to year x_1. These costs are in turn discounted to year zero by the term, $\exp(-rx_1^*)$. Thus, Eq. 8 states that the initial project is optimally scaled when the incremental cost of providing one more year's capacity is equal to the present value of annual interest costs of all expansions. Alternatively stated, the initial project should be scaled so that its cost is equal to the difference between the cost of a project with an extra year's capacity and the present value annual interest cost of all future expansions. The optimality condition of Eq. 8 can be simplified by replacing D_o with $x_o D$ and dividing through by kD^a to obtain

$$a(x_o + x_1^*)^{a-1} = \exp(-rx_1^*)\frac{r\,x^a}{1 - \exp(-rx)} \quad \ldots\ldots (9)$$

Eq. 9 shows that x_1^* is a function of r, a, x_o, and x, but x itself (or more exactly, the optimal value of x) depends on r and a and can be replaced in

TABLE 2.—Comparison of x_1^* Values from Eqs. 9 and 10

Discount rate, r (1)	Economy-of-scale factor, a (2)	Elapsed period, x_o, in years (3)	x_1^* from Eq. 9 (4)	x_1^* from Eq. 10 (5)	Error, as a percentage (6)
0.05	0.5	10	29.70	28.73	3.28
		40	36.22	35.36	2.37
		70	39.88	38.95	2.33
	0.6	10	23.50	23.18	1.36
		40	29.32	28.94	1.31
		70	32.44	31.94	1.54
	0.7	10	17.85	17.69	0.90
		40	22.72	22.40	1.39
		70	25.22	24.78	1.75
	0.8	10	12.44	12.20	1.91
		40	16.11	15.67	2.71
		70	17.90	17.38	2.92
	0.9	10	6.96	6.60	5.24
		40	9.08	8.58	5.47
		70	10.06	9.56	4.94
0.10	0.5	10	16.24	15.80	2.71
		40	20.41	19.93	2.33
		70	22.54	22.00	2.40
	0.6	10	13.02	12.86	1.24
		40	16.62	16.35	1.63
		70	18.40	18.07	1.80
	0.7	10	10.01	9.91	1.01
		40	12.93	12.69	1.84
		70	14.32	14.05	1.86
	0.8	10	7.06	6.90	2.30
		40	9.17	8.90	2.92
		70	10.15	9.89	2.58
	0.9	10	3.98	3.76	5.50
		40	5.15	4.91	4.69
		70	5.66	5.49	3.04
0.15	0.5	10	11.52	11.23	2.51
		40	14.62	14.27	2.40
		70	16.14	15.76	2.36
	0.6	10	9.29	9.18	1.22
		40	11.93	11.72	1.76
		70	13.19	12.96	1.71
	0.7	10	7.18	7.09	1.20
		40	9.28	9.11	1.85
		70	10.26	10.10	1.58
	0.8	10	5.08	4.95	2.57
		40	6.58	6.40	2.75
		70	7.26	7.12	1.91
	0.9	10	2.86	2.71	5.41
		40	3.68	3.54	3.71
		70	4.03	3.97	1.40
0.20	0.5	10	9.05	8.84	2.32

TABLE 2.—*Continued*

(1)	(2)	(3)	(4)	(5)	(6)
		40	11.54	11.26	2.40
		70	12.72	12.45	2.14
	0.6	10	7.33	7.23	1.40
		40	9.42	9.25	1.81
		70	10.40	10.24	1.53
	0.7	10	5.68	5.60	1.47
		40	7.33	7.20	1.82
		70	8.08	7.99	1.13
	0.8	10	4.02	3.92	2.58
		40	5.19	5.07	2.29
		70	5.71	5.65	1.00
	0.9	10	2.37	2.14	9.65
		40	2.89	2.81	2.60
		70	3.16	3.16	−0.12

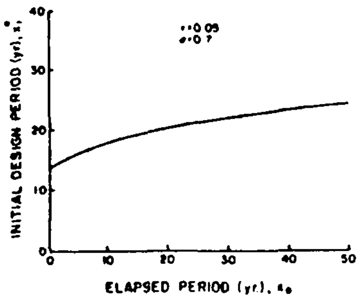


FIG. 4.—Optimal Initial Design Period Versus Elapsed Period

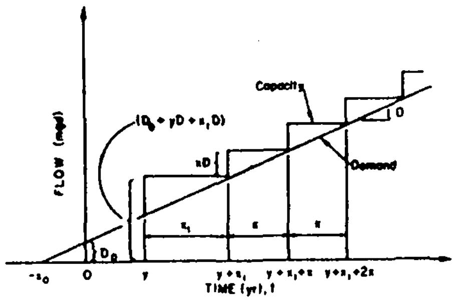


FIG. 5.—Waiting Period Model (1,000,000 gal/day = 37,800 m³/day)

Eq. 9 by values from Eqs. 5 or 6. The model thus shows that the optimal design period for initial construction x_1^* is a function of the annual discount rate r, the economy of scale factor, a, and the elapsed period, x_o. Given values for these parameters, Eq. 9 is an expression with the single unknown, x_1^*. As in the case of the Capacity Expansion Model, however, it cannot be solved explicitly for x_1^*; trial and error or numerical methods must be used.

Newton's method with a stopping criterion of 10^{-3} yr was employed for solving Eq. 9 for 140 values of x_1^*. To obtain these solutions, the discount rate, r, was varied from 0.05-0.20 in 0.05 increments, the economy of scale factor, a, was varied from 0.5-0.9 in 0.1 increments, and the elapsed period, x_o, was varied from 10 yr-70 yr in 10-yr increments; 60 of the x_1^* values are shown in Table 2. Graphical analysis of the 140 values suggested a model that was used in linear regression analysis to develop an approximating equation solved explicitly for x_1^*; the result is

$$x_1^* = x + \left(\frac{1-a}{r}\right)^{0.7} \frac{x_o^{0.9}}{(x_o + x)^{0.6}} \quad \text{(10)}$$

Note that $x_1^* > x$ for $x_o > 0$, i.e., the optimal design period of the initial project is greater than the design period of expansions when an initial capacity deficit exists. Optimal initial design period versus elapsed period is shown in Fig. 4.

Table 2 includes a comparison of exact and approximate x_1^* values obtained from solution of Eqs. 9 and 10, respectively; the differences are seen to be small. From the 140 residuals obtained by subtracting the x_1^* values from the two equations, the standard error of estimate was found to be 0.34 yr, and the correlation coefficient was 0.999.

Waiting Period Model

The previous models are preliminary to that of this section developed by the writers primarily for, but not limited to, the case of water supply planning in developing countries. In earlier models, the restriction that capacity equal or exceed demand essentially eliminates timing as a decision variable; construction is to take place at the beginning of the planning horizon and at subsequent points of zero excess capacity. In the model presented in this section, however, an unsatisfied demand exists at $t = 0$ as in the initial deficit case, but the decision is made to let demands go unsatisfied from local facilities for y yr before constructing the first system, as shown in Fig. 5. This is the situation that normally prevails in developing countries. A comparable case in the United States is one where excess demands beyond the capacity of local supply facilities are satisfied by importing water from a neighboring system. Once construction or expansion of the local system is made, however, capacity must always equal or exceed demand, as before. The three decision variables of this model are y, the waiting period before constructing the first system; x_1, the excess capacity period of the initial project; and x, the excess capacity period of future expansions.

During the waiting period from the beginning of the planning horizon to the time of initial construction, costs are incurred. In the case of water supply in developing countries, demands are usually only partially satisfied from natural sources or vendors resulting in health and other social costs. Where demands are met by importing as in the United States, fees must be paid for service from the neighboring system. For purposes herein, it is assumed that a cost of p dollars is associated with each gallon of water demanded but not supplied from the local system. A general analysis of this penalty cost for developing countries is presented by the first writer (4).

Proceeding as in previous models, optimality conditions can be developed from an expression of total present value cost in terms of decision variables. At any time t in the waiting period, the rate of unsatisfied demand is $(D_o + Dt)$; the rate of penalty cost accrual is p times this amount, convertible to present value through use of the present worth factor, $\exp(-rt)$, in which r = annual discount rate, as before. Integrating from time 0 to y results in the following expression of present value penalty costs:

$$\int_{t=0}^{y} \exp(-rt)\, p(D_o + Dt)\, dt = \frac{p}{r}\left[\left(D_o + \frac{D}{r}\right) - \exp(-ry)\left(D_o + \frac{D}{r} + Dy\right)\right] \quad \text{(11)}$$

Assuming a construction cost function of the type in Fig. 2 represented by Eq. 3, the expression of total present value penalty and construction costs for an infinite time horizon is

$$\int_{t=0}^{y} \exp(-rt)\, p(D_o + Dt)\, dt + \exp(-ry)\, k(D_o + yD + x_1 D)^a$$

$$+ \exp[-r(y + x_1)] \frac{k(xD)^a}{1 - \exp(-rx)} \quad \ldots \quad (12)$$

the second and third terms of which are present value initial construction and expansion costs, respectively. Optimal values of the decision variables can be found by setting the appropriate partial derivatives of Eq. 12 equal to zero and solving.

The derivative with respect to x results in an expression identical to Eq. 5 of Manne's expansion model. The derivative with respect to x_1 results in Eq. 13 which is equivalent to Eq. 8 obtained by Thomas for the Initial Deficit Model but slightly different in symbols:

$$D\,a[k(D_o + Dy + Dx_1^*)^{a-1}] = \exp(-rx_1^*) \frac{r\,k(xD)^a}{1 - \exp(-rx)} \quad \ldots \quad (13)$$

Note on the left side that Dy has been added to the term in parentheses to account for increasing demand during the initial years of deficit. The interpretation of Eq. 13 is unchanged from that presented for Eq. 8, but it applies at time $t = y$; i.e., the initial project is optimally scaled when the incremental cost of providing one more year's capacity is equal to the present value of annual interest costs of all expansions discounted to year y.

The optimal waiting period is determined from the derivative of Eq. 12 with respect to y. The resulting expression is cumbersome:

$$p(D_o + Dy^*) = r\,k(D_o + Dy^* + Dx_1)^a + \exp(-rx_1) \frac{r\,k(xD)^a}{1 - \exp(-rx)}$$

$$- D\,a[k(D_o + Dy^* + Dx_1)^{a-1}] \quad \ldots \quad (14)$$

Note, however, that the last two terms on the right-hand side are identical to the right and left-hand terms of Eq. 13, except for the asterisks. If the initial project is optimally timed and scaled (i.e., if $y = y^*$ in Eq. 13 and $x_1 = x_1^*$ in Eq. 14), then the last two terms of Eq. 14 are equal and can be eliminated because they are different in sign. The resulting optimality condition is

$$p(D_o + Dy^*) = r[k(D_o + Dy^* + Dx_1^*)^a] \quad \ldots \quad (15)$$

This completes development of the optimality conditions for the Waiting Period Model. The optimal design period of expansions x^* can be calculated from Eq. 5, or approximated from Eq. 6 given values for r and a. The optimal timing, y^*, and design period of the initial project, x_1^*, can be determined by solving Eqs. 13 and 14 simultaneously as shown in Appendix I given values for x^*, p, D_o, D, and k. This, however, is difficult and requires use of numerical techniques and the electronic computer.

Before presenting an approximating equation for the optimal waiting period, it is useful to examine Eq. 15, the condition that must be met if the waiting and design periods of the initial project are to be optimal. Note that the term in parentheses on the left side of Eq. 15 is the rate of unsatisfied demand at $t = y$, the time of initial project construction; multiplying by p results in the rate of penalty cost accrual at $t = y$, with typical units in dollars/year. The right side of Eq. 15 is the annual interest cost of the initial project. Thus, Eq. 15 says that construction of the first project should be delayed until its annual interest cost is equal to the rate of penalty cost accrual.

Eq. 15 can be rearranged to provide useful insights into capacity planning. Solving for y^* results in

$$y^* = \frac{r\,k(D_o + Dy^* + Dx_1^*)^a}{pD} - \frac{D_o}{D} \quad \ldots \quad (16)$$

which shows that the optimal waiting period decreases as the penalty price, p, increases, i.e., if the cost of importing water from a neighboring community is high, or if the social loss from letting demands go unsatisfied is high, then local facilities should be constructed sooner than if p is low. The value of p for which no waiting is optimal (denoted $\hat{p}$) can be found by setting $y^* = 0$ in Eq. 16 and solving; the resulting expression is

$$\hat{p} = \frac{r\,k(D_o + Dx_1^*)^a}{D_o} \quad \ldots \quad (17)$$

Note that Eq. 17 applies to the Initial Deficit Model of the previous section since that model is identical to the Waiting Period Model with $y^* = 0$. Should the planner, then, use Thomas' model for determining the optimal scale of a project and decide to build at $t = 0$, a penalty price of $\hat{p}$ would be implicitly assigned to each gallon of unsatisfied demand. Eq. 17 provides the means for imputing $\hat{p}$; note that it is simply the product of the discount rate and initial project cost divided by the unsatisfied rate of demand at the time of construction.

Values of p greater than $\hat{p}$ result in negative waiting periods. If y^* is negative, it would have been optimal for construction to have taken place sometime in the past (i.e., before the start of the planning period). Since this is obviously impossible, implementation should take place at $t = 0$. If, however, the actual penalty price is less than $\hat{p}$, construction should be delayed; the optimal amount of delay y^* can be calculated from the simultaneous solution of Eqs. 13 and 14 as shown in Appendix I. Alternatively, an estimate of y^* can be obtained from

$$y^* = \alpha F^{\beta} - x_o \quad \ldots \quad (18)$$

in which $$F = \frac{kD^a}{pD} \quad \ldots \quad (19)$$

$$\alpha = 0.012\, a^{-4.7}\, r^{0.4} \quad \ldots \quad (20)$$

$$\beta = 5.58\, a^{1.3}\, r^{0.18} \quad \ldots \quad (21)$$

Note that one of the independent variables of Eq. 18 is F, defined in Eq. 19; F is identical to the "penalty factor" employed by Erlenkotter (2). Note also that the numerator and denominator of the right-hand side of Eq. 17 can be multiplied by x^a and x, respectively, without changing the value of F if x = 1 yr. The numerator of F is thus seen to be the cost of a system with capacity to meet 1 yr of demand, and the denominator is the rate of penalty

TABLE 3.—Comparison of ($x_o + y^*$) Values from Eqs. 18 and 24

Discount rate, r (1)	Economy-of-scale factor, a (2)	Penalty factor, F (3)	$x_o + y^*$ from exact Eq. 24 (4)	$x_o + y^*$ from approximate Eq. 18 (5)	Error, as a percentage (6)
0.05	0.5	55.207	20.000	18.274	8.63
		91.630	40.000	35.689	10.78
		120.713	60.000	51.369	14.38
		145.555	80.000	65.777	17.78
		167.552	100.000	79.219	20.78
0.05	0.6	40.188	20.000	18.807	5.97
		62.886	40.000	39.805	0.49
		79.833	60.000	59.356	1.07
		93.690	80.000	77.600	3.00
		105.568	100.000	94.770	5.23
0.05	0.7	30.228	20.000	20.072	−0.36
		44.146	40.000	43.565	−8.91
		53.745	60.000	65.159	−8.60
		61.207	80.000	85.016	−6.27
		67.372	100.000	103.461	−3.46
0.05	0.8	23.748	20.000	22.347	−11.73
		31.904	40.000	45.845	−14.61
		36.999	60.000	65.750	−9.58
		40.726	80.000	83.052	−3.81
		43.675	100.000	98.457	1.54
0.10	0.5	32.396	20.000	21.957	−9.78
		51.461	40.000	43.893	−9.73
		66.283	60.000	64.109	−6.85
		78.792	80.000	83.040	−3.80
		89.798	100.000	100.990	−0.99
0.10	0.6	23.829	20.000	20.944	−4.72
		35.501	40.000	44.617	−11.54
		43.976	60.000	66.969	−11.62
		50.826	80.000	88.134	−10.17
		56.666	100.000	108.331	−8.33
0.10	0.7	17.929	20.000	19.938	0.31
		24.857	40.000	42.519	−6.30
		29.509	60.000	63.283	−5.47
		33.092	80.000	82.537	−3.17
		36.042	100.000	100.603	−0.60
0.10	0.8	13.887	20.000	18.708	6.46
		17.727	40.000	36.676	8.31
		20.075	60.000	51.681	13.86
		21.785	80.000	64.748	19.06
		23.139	100.000	76.460	23.54

cost accrual, in dollars/year after 1 yr of allowing demand to go unsatisfied. Therefore, F is the ratio of capital to penalty costs and has units of time. In using Eq. 18, F must be in years if y^* and x_o are in years. It is important to note that F is inversely related to p; high importing or social costs imply low values of F.

Eq. 18 was developed from two-stage regression analysis of exact values from the solution of Eq. 24 in Appendix I. In solving Eq. 24, it was easier to treat F as the unknown than y^*. The economy of scale factor, a, was successively set equal to 0.5, 0.6, 0.7, and 0.8, and the discount rate, r, was set at 0.05 and 0.10. The sum, $x_o + y^*$, appearing as a single variable in Eq. 24, was assigned 50 values in the range of 2 yr–100 yr. A total of 400 exact values of F were obtained using Newton's method, 40 of which are shown

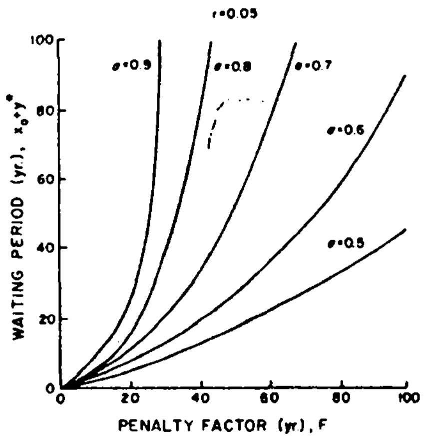


FIG. 6.—Optimal Waiting Period Versus Penalty Factor for Alternative Economy of Scale Factors

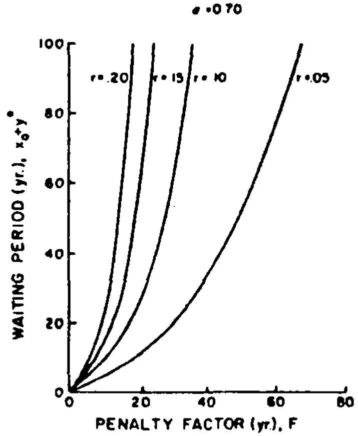


FIG. 7.—Optimal Waiting Period Versus Penalty Factor for Alternative Discount Rates

in Table 3. Graphical analysis of these values suggested the form of the model proposed for regression analysis.

Table 3 includes exact and approximate values of $x_o + y^*$ from Eqs. 24 and 18, respectively. The standard error of estimate based on 400 residuals obtained from the difference between $x_o + y^*$ values from the two equations is 6.2 yr, and the correlation coefficient is 0.977. Fig. 6 shows the variation of $x_o + y^*$ with F for alternative values of a and with r = 0.05. The optimal waiting period is seen to increase sharply as F increases (i.e., as the penalty cost, p, decreases), especially when economies of scale are small. For systems with greater economies (i.e., lower a-values), the increase in waiting period with F is less severe. Fig. 7 shows that for a given economy of scale factor, the optimal waiting period increases as F increases and has the sharpest rise when the discount rate is high.

Examples

In previous sections, Eq. 6 is presented for calculating the approximate optimal design period for expansions, Eq. 10 for the optimal design period of the initial projects, and Eqs. 18–21 for the optimal waiting period. Use of these equations is illustrated in this section. In general, the examples apply more to water supply planning in developing countries than in the United States.

It is assumed that the cost of water supply and distribution facilities for small communities abroad is \$300,000 for a system with 10^6-U.S. gal/day (3,785-m^3/day) capacity; this is the value of k in Eq. 3. Additionally, the economy of scale factor, a, is assumed to be 0.7. These values roughly correspond to those obtained by the first writer (3) from analysis of 65 new gravity systems, which provide disinfection but no other treatment constructed in Central America between 1965 and 1969.

Consider a community of 5,000 persons with average per capita water demand of 30 U.S. gal/day (0.114 m^3/day). The existing rate of demand is 150,000 U.S. gal/day (568 m^3/day) which is assumed to increase linearly at the rate D of 5,000 U.S. gal/day/yr (18.93 m^3/day/yr). Assume that the rate of demand has grown equal to the capacity of existing facilities in which case now is the time for an expansion if capacity is to equal or exceed demand. From Eq. 6, the design period, x^*, with $a = 0.7$ and $r = 0.05$ is found to be 13.5 yr. Multiplying by D, the required capacity x^*D is 67,500 U.S. gal/day (255 m^3/day) which, from Eq. 3 has approximate construction cost of \$45,000.

Now assume that the same community has no existing water system in which case the current demand for supply from local facilities goes unsatisfied; $D_o = 150{,}000$ U.S. gal/day (568 m^3/day). The number of elapsed years to the previous point of zero excess capacity x_o is D_o/D, or 30 yr. If the decision is made to build a supply system immediately, its design period x_i^* can be calculated from Eq. 10; the result with $x^* = 13.5$, $a = 0.7$, $r = 0.05$, and $x_o = 30$ is $x_i^* = 21.3$ yr. Thus, the initial project should have capacity ($D_o + x_i^* D$) equal to 256,400 U.S. gal/day (971 m^3/day); its cost from Eq. 3 is about \$116,000.

Instead of building the system immediately, it may be desirable to delay construction, in which case the waiting period model applies. It has been shown that if the project is implemented at the beginning of the planning period, its optimal cost is about \$116,000. From Eq. 17, the penalty price implicitly assigned to unsatisfied demand by building now is $\hat{p} = 0.05 \times 116{,}000/5{,}000 \times 30 \times 365 = 1.06 \times 10^{-4}$ \$/U.S. gal (2.8×10^{-2} \$/m^3). Equivalently, $\hat{p}$ is 10.6 ¢/10^3 U.S. gal (2.8 ¢/m^3). The question now is whether the benefit of publicly supplied water p from a local system differs from this value of $\hat{p}$. The equivalent question for the United States is whether the price of importing water from a neighboring town p is different from $\hat{p}$. If $p \geq 10.6$ ¢/10^3 U.S. gal (2.8 ¢/m^3), construction should proceed immediately. If, however, p is less than this amount, construction should be delayed. There is, of course, considerable difficulty in obtaining a value for p; alternative methods are considered in Ref. 4. For purposes herein, assume that $p = 9$ ¢/10^3 U.S. gal (2.38 ¢/m^3).

In order to determine the optimal waiting period, it is first necessary to calculate the penalty factor, F, from Eq. 19. For this purpose, the parameters upon which F depends are expressed in terms of dollars, gallons, and years. The value of k in cost Eq. 3 is \$300,000 when xD has units 10^6 U.S. gal/day. Equivalently $k = \$0.304$ (\$15.10) when xD is measured in U.S. gallons per year (cubic meters per year). The penalty price p is 9×10^{-5} \$/U.S. gal (2.38×10^{-2} \$/m^3), and the rate of demand increase D corresponding to 5,000 U.S. gal/day/yr is 1.825×10^6 U.S. gal/yr/yr (6,908 m^3/yr/yr). Substituting these values into Eq. 19 results in a value of $F = 44.8$ yr. From Eqs. 20 and 21, $\alpha = 0.01936$ and $\beta = 2.0468$ for $a = 0.7$ and $r = 0.05$. The resulting value of y^* from Eq. 18 with $x_o = 30$ is 16 yr. Construction should thus be delayed about 16 yr, at which time the rate of unsatisfied demand is expected to be 230,000 U.S. gal/day (871 m^3/day). This then will be the value of D_o for the Initial Deficit Model; the corresponding value of x_o is 46 yr. From Eq. 10, the optimal excess capacity period should be 23 yr implying the need for an initial project with capacity 0.345×10^6 U.S. gal/day (1,306 m^3/day), which from Eq. 3 is estimated to cost \$143,000.

Summary and Conclusions

Many of the underlying assumptions of the models of this paper are identical. To a large extent, they have not been explicitly stated, since this work has focused primarily on optimality conditions rather than model development. These assumptions are considered at length in Ref. 7 and are briefly summarized and examined as follows.

Among the more obvious assumptions are linearly increasing demand and instantaneous project implementation. It is also assumed that facilities are unable to produce beyond maximum capacity, that target demands are fixed (i.e., infinitely price inelastic), that cost functions for both initial construction and expansions are identical and constant over time, and that the planning horizon is infinite.

The assumption of linearly increasing demand is of doubtful accuracy except possibly for large municipalities. Medium and small cities in the United States and abroad are more likely to have geometrically increasing demand. The inaccuracies resulting from this assumption, however, are not serious. Muhich (8) compared optimal design periods for both geometric and linear rates of growth under several different conditions and found that although the periods associated with linear growth are somewhat longer, differences in total present value costs are negligible.

The assumptions of instantaneous project implementation and inflexible capacity are unrealistic but do not appear to be serious. Projects may require a year or two for implementation, and steps may be taken to extend excess capacities for a few years once design limits have been reached. These differences, however, are generally small compared to project design periods. Regarding the assumption of fixed target demand, evidence exists that municipal water use in the United States is indeed price inelastic except in rare cases. It appears that village demands in developing countries are similarly price inelastic since public water systems abroad are used primarily for supplying basic human requirements.

The assumption that the same cost function applies to initial projects as well as expansions is not well founded. Unfortunately, separate functions have not been reported in the technical literature. While they may indeed be different, it is likely that their economy of scale factors would be similar. It is unrealistic to assume that costs remain constant over time, but it is, after all, opportunity costs that are of concern rather than raw construction costs. Consequently, fluctuations in the economy that affect all costs more or less equally have little bearing on these models. The assumption of an infinite planning horizon can be easily relaxed by restricting n to finite value in the present worth factor of Eq. 2. For discount rates of about 10% and larger, a horizon limited to only a few rather than an infinite number of expansions has little effect on optimal design periods.

The economic life of facilities is assumed to be infinite. The more realistic case of finite life can be easily accommodated by including a multiplier in the cost function as shown by Manne (7); optimality conditions are unaffected. The cost function itself refers to entire undifferentiated systems which in reality consist of separate components. Water treatment plants, for example, include settling tanks, filters, and numerous additional facilities, and complete municipal systems include supply, treatment, and distribution works. Strictly speaking, model results are valid for integrated systems only if the economy of scale factor a is identical for all components. Regarding a, its value must be in the range 0-1.

Operating and maintenance costs are ostensibly ignored herein. Actually, they are assumed to be proportional to the amount of water supplied in the case of the Capacity Expansion and Initial Deficit Models. Since demand must be exactly satisfied in these first two models, it follows that such costs are predetermined and do not affect optimality conditions. This is not true, however, for the Waiting Period Model. Since part of the demand goes unsatisfied from local facilties, explicit account must be taken of operating costs, even under the assumption that they are proportional to output. Basically, such account can be taken by subtracting the unit cost of operation from p in the first term of Eq. 12. This is equivalent to stating that p does not represent social or importing cots, but rather the difference between such costs and the cost per gallon of operation. In reality, economies of scale are associated with operation, but they do not seem to be of such magnitude as to grossly invalidate this interpretation.

While a few additional assumptions might be cited, they are of minor consequence. More important are the conclusions that can be drawn from the work presented herein. These are primarily associated with the Waiting Period Model. Perhaps most noteworthy is the fact that the optimal waiting period prior to construction is a function of the tension between system costs and the losses associated with not satisfying demand from local facilities; such tension is reflected by F, the penalty factor. As shown in Figs. 6 and 7, the optimal waiting period is extremely sensitive to F.

In order to decide how long to wait before construction, a numerical value is needed for p, i.e., information is required either on the costs of importing (if demands are to be satisfied from a neighboring system) or the social benefits of publicly supplied water (if demands are not met). In the latter case, such values are extremely difficult to obtain, which generally requires investment decisions to be made in their absence. The mere act of deciding, however, implicitly assigns a value to such benefits; it can be estimated from Eq. 17. Assuming previous investment decisions were intended to be optimal, even if in fact they were not, their imputed benefits can be easily calculated by dividing the product of discount rate and construction cost by the rate of unsatisfied demand at the time of construction. The water supply benefits associated with pending current decisions under consideration can be similarly estimated. While

such calculations can result in extensive tables of imputed p-values, their importance for guiding future investment is questionable. The ability to impute benefits may be useful, but it does not relieve the planner of the need to more accurately assess them by willingness-to-pay or other measures.

Total present value costs associated with the Waiting Period Model are relatively insensitive to nonoptimal values of the decision variables. This was determined from Eq. 12. Results for the case of $r = 0.05$, $a = 0.7$, $x_o = 30$, and $F = 50$ (for which $y^* = 21.49$, $x_i^* = 21.29$, and $x^* = 13.51$) are fairly representative. With any two of the decision variables at their optimal values and the third at half its optimal value, the increase in total present value cost did not exceed 2%. Similarly, with the third at twice its optimal value, the increase did not exceed 2.5%. Costs are most sensitive to x_i and y.

An important conclusion from the Initial Deficit Model by Thomas (12) is that the optimal design period for systems with initially unsatisfied demand is always greater than that for expansions. In developing countries, it is not uncommon to find community water systems designed for 20 yr or more. Instinctively, such design criteria seem excessive in light of higher discount rates abroad than in the United States. However, 20-yr design periods may not be far from optimal due to the fact that most construction is for new systems with existing unsatisfied demand. Correspondingly, 20-yr design periods are probably far too long for expansions.

Appendix I.—Determination of Optimal Waiting Period

Mathematical expressions for optimal values of the decision variables in the Waiting Period Model include Eqs. 5, 13, and 14; furthermore, Eq. 15 is obtained from the simultaneous solution of Eqs. 13 and 14. Replacing D_o in Eq. 15 by Dx_o, the expression can be rewritten as

$$x_o + y^* = \left(\frac{kD^a}{pD}\right) r(x_o + y^* + x_i^*)^a \qquad (22)$$

substituting F for kD^a/pD and solving for x_i^* yields

$$x_i^* = \left(\frac{x_o + y^*}{rF}\right)^{1/a} - (x_o + y^*) \qquad (23)$$

This expression for x_i^* in terms of y^* and parameters can now be substituted into Eq. 13 to obtain the following expression, which contains only y^* as the unknown:

$$(x_o + y^*)^{(a-1)/a} = \frac{rx^a(rF)^{(a-1)/a}}{a[1 - \exp(-rx)]} \exp\left[-r\left(\left(\frac{x_o + y^*}{rF}\right)^{1/a} - (x_o + y^*)\right)\right] \qquad (24)$$

Note that y^* is always summed with x_o, which makes it possible to consider $x_o + y^*$ as a single unknown. It is also important to note that $(x_o + y^*) = 0$ when $F = 0$, and from Eq. 19 it is seen that $F \rightarrow 0$ as $p \rightarrow \infty$.

To determine $x_o + y^*$ from Eq. 24, values are needed for x, r, a, and F; x (actually x^*) can be obtained from Eq. 5, given values for r and a, and F can be evaluated from Eq. 19, given values for k, D, a, and p. Even with such data, Eq. 24 cannot be solved explicitly for $x_o + y^*$; in this study, Newton's method was used for solution. Once y^* is evaluated, x_1^* can be calculated from Eq. 23.

Appendix II.—References

1. Chenery, H. B., "Overcapacity and the Acceleration Principle," *Econometrica*, Vol. 20, No. 1, Jan., 1952, pp. 1-28.
2. Erlenkotter, D., "Optimal Plant Size With Time-Phased Imports," *Investments for Capacity Expansion*, A. S. Manne, ed., MIT Press, Cambridge, Mass., 1967.
3. Lauria, D. T., "Planning Small Water Supplies in Developing Countries," *Report*, U.S. Agency of International Development, Washington, D.C., 1972.
4. Lauria, D. T., "Water Supply Planning in Developing Countries," *Journal of the American Water Works Association*, Vol. 65, No. 9, Sept., 1973, pp. 583-587.
5. Lynn, W. R., "Stage Development of Wastewater Treatment Works," *Journal of the Water Pollution Control Federation*, Vol. 36, No. 6, June, 1964, pp. 722-751.
6. Manne, A. S., "Capacity Expansion and Probabilistic Growth," *Econometrica*, Vol. 29, No. 4, Oct., 1961, pp. 632-649.
7. Manne, A. S., ed., *Investments for Capacity Expansion: Size, Location and Time Phasing*, MIT Press, Cambridge, Mass., 1967.
8. Muhlich, A. J., "Capacity Expansion of Water Treatment Facilities," thesis presented to the Harvard University, at Cambridge, Mass., in 1966, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
9. Rachford, T. M., Scarato, R. F., and Tchobanoglous, G., "Time-Capacity Expansion of Waste Treatment Systems," *Journal of the Sanitary Engineering Division*, ASCE, Vol. 95, No. SA6, Proc. Paper 6957, Dec., 1969, pp. 1063-1077.
10. Rich, L. G., *Environmental Systems Engineering*, McGraw-Hill Book Co., Inc., New York, N.Y., 1973.
11. Rudd, D. F., and Watson, C. C., *Strategy of Process Engineering*, John Wiley and Sons, Inc., New York, N.Y., 1968.
12. Scarato, R. F., "Time-Capacity Expansion of Urban Water Systems," *Water Resources Research*, Vol. 5, No. 5, Oct., 1969, pp. 929-936.
13. Thomas, R. H., discussion of "Time Capacity Expansion of Urban Water Systems," by Thomas M. Rachford, Russell F. Scarato, and George Tchobanoglous, *Journal of the Sanitary Engineering Division*, ASCE, Vol. 96, No. SA4, Proc. Paper 7441, Aug., 1970, pp. 1017-1023.
14. Zimmerman, H., and Sovereign, M. G., *Quantitative Models for Production Management*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1974.

Appendix III.—Notation

The following symbols are used in this paper:

a = economy of scale factor;
$C(xD)$ = construction cost of system with capacity xD;
D = annual rate of demand increase;
D_o = initial unsatisfied rate of demand;
F = penalty factor;
k = cost of system with unit capacity;
n = number of expansions;
p = penalty cost of unsatisfied demand;
$\hat{p}$ = implicit penalty cost associated with no waiting period;

r = annual discount rate;
t = time;
x = design period of expansions;
x^* = optimal design period of expansions;
x_o = elapsed period;
x_1 = design period of initial project;
x_1^* = optimal design period of initial project;
y = waiting period; and
y^* = optimal waiting period.

APPENDIX F

SPSS
USER
INSTRUCTIONS

August 1981

Direct inquiries to:

Professor Donald T. Lauria
Department of Environmental Sciences and Engineering
School of Public Health
University of North Carolina
Chapel Hill, North Carolina 27514
USA

ABSTRACT

SPSS is a computer package which provides a wide variety of statistical services. Of greatest interest to us is the "regression" feature, which allows us to develop and compare different equations describing dependent and independent variables from input of raw statistical data. In the following example, statistics from 16 network designs are analyzed to develop relations for

1) Cost as a function of length and average diameter,
2) Cost as a function of tap spacing and per capita flow,
3) Average diameter as a function of per capita flow and network length.

RAW DATA

R Service Radius (m)	Q Per Capita Flow (LPCD)	L Network Length (m)	$\bar{D}$ Average Diameter (mm)	C Cost in $1,000's
100	20	1450	59.0	26.7
100	50	1450	88.0	37.4
100	100	1450	110.5	46.4
50	20	3007	42.6	44.4
50	50	3007	61.2	58.1
50	100	3007	79.8	72.9
8	50	10788	35.6	142.1
8	100	10788	43.0	160.6

100	20	2080	45.8	31.9
100	50	2080	66.8	42.2
100	100	2080	85.8	52.8
50	20	3780	35.4	49.8
50	50	3780	50.8	63.0
50	100	3780	65.2	76.8
13	50	12580	34.4	159.5
13	100	12580	42.7	184.9

MODELS TO BE TRIED

A. Cost as a function of length and diameter

1) $C = K + aL + a\bar{D}$

2) $C = K\ L^a\ \bar{D}^b$ [for SPSS: $\ln(C) = \ln(K) + a\ \ln(L) + b\ \ln(\bar{D})$

3) $C = K\ e^{aL}\ e^{b\bar{D}}$ [for SPSS: $\ln(C) = \ln(K) + aL + b\bar{D}$]

B. Cost as a function of service radius and per capita flow

4) $C = K\ R^a\ Q^b$ (for SPSS: $\ln(C) = \ln(K) + a\ \ln(R) + b\ \ln(Q)$]

5) $C = K\ e^{aR}\ e^{bQ}$ [for SPSS: $\ln(C) = \ln(K) + aR + bQ$

C. Average diameter as a function of per capita flow and network length

6) $\bar{D} = K\ Q^2\ L^b$ [for SPSS: $\ln(\bar{D}) = \ln(K) + a\ \ln(Q) + b\ \ln(L)$]

$\bar{D} = K\ e^{aQ}\ e^{bL}$ [for SPSS: $\ln(\bar{D}) = \ln(K) + aQ + bL$

SIMPLIFIED REGRESSION INPUT INSTRUCTIONS*

"CARD NUMBER" Represents the order in which the data is to be read in, and does not itself appear on the card.

"CARD NAME" (where shown), indicates the name of the SPSS function, which must appear in the first 15 columns of the card.

"CARD INPUT" Represents the instructions which are supplied by the user to the program, in columns 16-80 of each card.

Card Number	Card Name	Card Input
1	RUN NAME	The title the user wishes to appear on the output for identification of the run.
2	VARIABLE LIST	The names (up to eight characters in length) of the variables in the raw data to be supplied. These must appear in the same sequence as they appear in the data. Variable names are separated by commas.
3	INPUT MEDIUM	The form in which data will be read in (CARD or TAPE or DECK or OTHER).
4	N OF CASES	The number of complete observations input as raw data. A "complete observation" consists of a set of dependent and independent variables.
5	INPUT FORMAT	The format of input data. (FIXED or FREEFIELD or BINARY). FIXED format indicates data will appear in fixed fields, which must then be specified (e.g. FIXED (F10.5, I5, F8.2)). FREEFIELD implies that data will be separated only by commas.

*More complete documentation is available from Statistical Package for the Social Sciences by Norman H. Nie, C. Hadlai Hull, Jean Jenkins, Karin Steinbrenner and Dale H. Bent. (McGraw Hill, 1975).

Card Number	Card Name	Card Input
6 et seq.	COMPUTE	(optional). If variable transformations are desired, they must be specified here. This is done by writing an equation in which a new variable name is on the left side of an = sign, and the transformation of the input variable is on the right side. Examples and specifications of these transformation functions are attached. A new COMPUTE card must appear for each transformation.
7 et seq.	REGRESSION	The variables (both those input and those developed in COMPUTE statements) which will be included in regression equations described in cards 8 et seq. These are listed to the right of the expression VARIABLES =, with variable names separated by commas. If these cannot all fit on one card, then the first line should end after a complete variable name and comma, and the remaining varaibles may be listed on the following line starting in column 16. A SLASH MUST FOLLOW THE LAST VARIABLE NAME.
8 et seq.	(regression design card. No name is entered in the first 15 spaces, however.)	The proposed regression models to be examined are typed in columns 16-80. These are specified by writing REGRESSION = dependent variable name WITH independent variable 1, independent variable 2, etc. (e.g. REGRESSION = Z WITH X,Y). These cards must follow immediately after the REGRESSION cards (card number 7 et seq.). A regression design card is required for each model. A slash follows all regression models except the last. The mode of the regression must also be specified.... see Note 1.
9 et seq.	OPTIONS	(optional.) Additional manipulations of data or output which may be performed. These are identified by number. A list of these options is attached.
10 et seq.	STATISTICS	(optional.) The additional statistics besides basic regression data (regression terms, r^2, F-tests, standard errors, etc.) which the user wants. A list of these statistics follows. These optional statistics are listed by number.
11	READ INPUT DATA	No card input required. This card simply signifies that the data follows.

Card Number	Card Name	Card Input
12 et seq.	(no name)	The raw data being input for analysis, in the sequence and format specified in the VARIABLE LIST and INPUT FORMAT cards.
13 et seq.	FINISH	No card input. This signifies the end of the program.

Note 1: There are three ways in which SPSS can perform regressions: simple regression, hierarchical regression, and stepwise inclusion subject to statistical significance.

In simple regression, all variables are introduced into the equation simultaneously. This is signified by following all the variables on the right hand side with an even number in parentheses. (e.g. REGRESSION = Z WITH X,Y (2)/).

In hierarchical regression variables are introduced stepwise so that models using less than all of the variables are developed. Each variable is followed by even numbers in parentheses, with the higher numbered variables introduced first. Thus

REGRESSION = Z WITH X(4),Y(2)/

will develop two regression equations: Z = f(x) and Z = f(X,Y).

In stepwise inclusion subject to statistical significance, variables are introduced sequentially at each level of inclusion according to the fraction of variance they explain, and subject to the satisfaction of statistical significance tests specified by the user. These tests are indicated immediately after the dependent variable and consists of the maximum number of dependent variables, the minimum F-test acceptable, and the minimum fraction of a variable's variance unexplained by previously entered variables.

For example,

REGRESSION = Z(4,5.2,.2) WITH A(5), B(5), C(3), D(1), E(1), F(1)/

indicates that

(1) variables A & B will be examined first, then C, then D, E, and F.
(2) a maximum number of 4 independent variables will be considered in a regression equation.
(3) Independent variables will only be entered which have F-test values greater than 5.2, and
(4) That an independent variable will only be entered if at least 20% of its variance is unexplained by previously entered variables.

Graphic	Meaning	Example
/	Division	VARX=VARA/VARB
*	Multiplication	VARX=VARA*VARB
+	Addition	VARX=VARA+VARB
-	Subtraction	VARX=VARA-VARB
**	Exponentiation	VARX=VARA**2

In addition to these standard arithmetic operators, any of the variables or constants used in the expression may also be acted upon by one of the following prepared or packaged functions.

Mnemonic	Meaning	Example
SQRT	Square root	VARX=SQRT(VARA)
LN	Natural or Naperian logarithm	VARX=LN(VARA)
LG10	Base 10 logarithm	VARX=LG10(VARA)
EXP	Exponential (e^{arg})	VARX=EXP(VARA+VARC)
SIN	Sine†	VARX=SIN(VARA+VARB)
COS	Cosine†	VARX=COS(VARA)
ATAN	Arctangent†	VARX=ATAN(VARA)
RND	Round result to whole number	VARX=RND(VARA+VARC/6)
ABS	Absolute value (ignores sign)	VARX=ABS(VARA)
TRUNC	Truncate value (whole number without rounding)	VARX=TRUNC(VARA)
MOD10	Result is remainder of division by 10	VARX=MOD10(VARA)

†Argument is in radians.

In order to make use of the above functions, it is necessary to follow the mnemonic of the function with an expression entirely enclosed in parentheses. The parenthesized expression may be the name of a single variable, or it may be a more complex expression containing one or more variable names and/or constants.

The COMPUTE card, like most other SPSS cards, may be continued on successive cards if the entire statement cannot be completed on one physical card. When this is the case, columns 1 to 15 of succeeding cards are left blank, and the rest of the statement is completed in columns 16 to 80 of as many cards as needed.

The COMPUTE control card, unlike many other SPSS cards, may contain no more than one transformation though a transformation may take more than one physical card to complete. Each new statement must begin with the word COMPUTE starting in column 1 of the control field. For these reasons *it is incorrect to use a card like the following.*[1]

```
1              16
COMPUTE        NEWVAR=VARA*VARB    FIRST VAR=NEWVAP/VARC    VAR038=FIRSTVAR
               /(VAR029-1)
```

The use of this card would cause the run to be terminated and an error message to be reported.

When generating variable transformations by means of the COMPUTE card, the user need not be concerned with the amount of space (i.e., the number of digits) taken up by the results of the transformation since space is automatically provided by the system. The user should remember that if the calculated variables are intended for crosstabulations and other such procedures, there should be a reasonable number of categories for convenience. The user has at hand the RND and TRUNC functions, which convert mixed, i.e., numbers including decimal fractions, to whole numbers before the values are actually output onto the cases.

[1]Note that this is an incorrect control card (for demonstration only).

Along with the Statistic 4 plot, Y scores, predicted Y' values, and residuals are listed in raw-score (unstandardized) form. Also listed is the SEQNUM of each case, which is the sequence number of a case as it occurs in the file. The SEQNUM is generated by SPSS and is used as the vertical dimension of the Statistic 4 plot. To obtain the greatest utility from the plot, the user should attempt to sort cases along some meaningful dimension in a previous run. Sorting cases into a time sequence, for example, would allow the user to explore for time dependence or autocorrelation. As an aid in interpreting the plot, the user may obtain the so-called Durbin-Watson statistic by specifying Statistic 5 on the STATISTICS card. A tabled sampling distribution for the Durbin-Watson statistic, along with a discussion of its use in testing for autocorrelation, is provided by J. Johnston (1972).

Following the scatterplots requested by the user, standardized residuals and predicted Y' values may be output on the raw-output-data file for future use. The writing of standardized residuals and predicted Y' values is controlled by various combinations of Options 11 and 12 on the OPTIONS card. If neither Option 11 nor 12 is specified, the raw-output-data file will *not* be produced. Option 11 used alone causes output of standardized residuals; Option 12 alone causes output of standardized Y' values. Finally, the use of Options 11 and 12 together causes output of both standardized residuals and predicted Y' values.[1]

The default output format for the residual and/or predicted Y' values written on the raw-output-data file is 8F10.6. A maximum of eight residuals or Y' values, or a maximum of four pairs of residuals and Y' values, are written on *each* record. Output records of the residuals can be sequenced by specifying Option 10. In this case, the first 20 columns in the record are used for sequencing information, and only six residuals and/or Y' values are written (in 6F10.6 format) in columns 21 to 80. The sequencing information includes the SEQNUM in columns 1 to 8, record number per case in columns 9 and 10, and the first four characters of the file name (or subfile name if relevant) in columns 12 to 15.

Figure 20.8 in SEC. 20.10 shows output generated with Options 11 and 12. When both residuals and predicted Y' values are output, they appear as residual for the first equation, then Y' for the second, and so forth. The following information is also output.

1 A message indicating the number of residuals and/or predictors output, the number of cases, and the number of records written for each case.
2 The value assigned to residuals and/or Y' values if they are missing (99.0).
3 The possible range of valid values for residuals and/or Y'. Possible range is always +99.0 to −99.0 and extreme values beyond these limits are truncated. Since the residuals and predicted Y' values are output in standardized form, this range is quite generous.
4 A summary table showing the VARIABLE list and REGRESSION design statements for which residuals were output. The summary table also indicates the output record number, record columns, and the number of missing cases.

20.6 OPTIONS AVAILABLE FOR SUBPROGRAM REGRESSION

There are 15 options available with subprogram REGRESSION. Options are specified by the user on an OPTIONS card placed immediately after the REGRESSION procedure card. The OPTIONS card contains the control word OPTIONS beginning in column 1, and the list of desired options beginning in column 16. When more than one option is specified, option numbers are listed in order of increasing size, separated by commas. The general format of the OPTIONS card is

[1]If Options 11 and/or 12 are used, an operating system control card defining the raw-output-data file must be included (see Appendix E, F, G, or H). A RAW OUTPUT UNIT card may be needed to separate the residuals and/or predictors from raw output data written by other tasks in the run. When residuals and/or predictors are output, correlation matrices, etc. should probably not be output from the same REGRESSION task since they will all be written on the same raw-output-data file.

1	16
OPTIONS	number list

If the user wishes to specify, say, Options 4 and 9, the card would appear as

1	16
OPTIONS	4, 9

It should be noted that some options may be used only when one or more other options are also specified. For example, Options 5 and 9 presuppose that Option 4 is also specified. On the other hand, there are some options that are incompatible and may not be used on the same OPTIONS card; for example, Options 1 and 2 or 4 and 15 may not be specified for the same run. Options specified on the OPTIONS card are in effect for all VARIABLES lists and/or REGRESSION design statements contained on the procedure card.

OPTION 1 *Inclusion of missing data.* This option causes the subprogram to include all cases in the calculation of correlation coefficients regardless of any missing-data values which may be defined.

OPTION 2 *Pairwise deletion of missing data.* This option causes pairwise deletion of cases which contain missing-data values. With this option, a missing value for a particular variable causes that case to be eliminated from calculations involving that variable only. Pairwise deletion should be used when a researcher has many variables each with just a few missing values, and when listwise deletion (the default option) would reduce the number of cases farther than desired. The number of cases from which the degrees of freedom are calculated is, under this pairwise deletion option, the minimum number of cases that any correlation coefficient required by the particular REGRESSION design statement is based upon.

The user should be aware that serious problems may result from using pairwise deletion. As a result of computational inaccuracies, little confidence can be placed in multiple regression statistics when pairwise deletion is used. Occasionally, such anomalies as multiple correlation coefficients greater than 1.0, or negative sums of squares and F ratios, are obtained with pairwise deletion. Consequently, Option 2 is often not justified and should be used with extreme caution.

Default Option—Listwise Deletion of Missing Data. When neither Option 1 nor Option 2 is specified, cases with missing values are automatically eliminated from all calculations through listwise deletion. Thus, all means, standard deviations, and correlations are based on the same universe of data. While sample size may be decreased markedly, there are sound statistical reasons for preferring listwise deletion, as can be seen in the discussion of pairwise deletion. For further discussion of the various treatments of missing data the reader may refer to Sec. 19.4.

OPTION 3 *Suppression of variable labels.* Selection of this option causes suppression of the variable labels on the printed output. While resulting in a slight increase in processing speed, the absence of variable labels makes the output less convenient to use, especially when it is to be read by persons unfamiliar with the user's data.

OPTION 4 *Matrix input.* This option specifies that a matrix of correlation coefficients will be input by the user. Detailed specifications for the input of a correlation matrix are given in Sec. 20.4.

OPTION 5 *Input of means and standard deviations.* This option indicates that means and standard deviations are to be read-in preceding the input correlation matrix (see Sec. 20.4.2). *Option 5 may only be used when Option 4 is also used.*

OPTION 6 *Suppression of step-by-step output.* When this option is specified, only the summary table portion of the REGRESSION output will be printed.

OPTION 7 *Suppression of the summary table.* When this option is specified, only the step-by-step portion of the REGRESSION output will be printed.

OPTION 8 *Matrix output.* This option causes the correlation matrix or matrices used in the calculations to be output on a unit of the user's choice. In this case, an operating system control card defining the raw-output-data file must be prepared (see Appendix E, F, G, or H). A RAW OUTPUT UNIT card may also be needed to separate the matrix from the raw output data produced by other tasks in the run. The format of the output matrix is compatible with that required for input to subprogram REGRESSION and thus may be used for matrix input on subsequent runs. Means and standard deviations may also be output (see Option 15).

OPTION 9 *Input correlation matrix is indexed by the* VARIABLE LIST *card.* As described in Sec. 20.4, Option 9 indicates that the user will input only one large correlation matrix and that subsets of variables from the matrix will be used in various regression calculations. The use of this option is convenient when a large number of variables is to be read-in, and when several VARIABLES lists on the procedure card contain many variables in common. *Option 9 cannot be used without Option 4.*

Options 10 through 14 pertain to analysis of residuals.

OPTION 10 *Causes sequencing information to be entered in columns 1 through 20 of each record on the raw-output-data file.* SEQNUM is placed in columns 1 to 8, the record number in columns 9 and 10, and the first four characters of file or subfile name in columns 12 to 15. Six residuals and/or Y' values are written on the record starting in column 21 with format 6F10.6. If Option 10 is not used, output format is 8F10.6 for residuals and Y' values and no sequencing information is output.

OPTIONS 11 and 12 If Option 11 is used alone, standardized residuals are output on the raw-output-data file. Option 12 alone will cause output of standardized Y' values. When Options 11 and 12 are used together, both residuals and Y' values are output. If only plots are desired, neither Option 11 nor 12 should be used.

OPTION 13 This option is in effect only when pairwise deletion (Option 2) is being used and when data replacement is requested (RESID = mdrp, where mdrp > 0).

Option 13 *will create standardized predictors which are a weighted product of the existing data:*

$$\text{Weighted standardized predictor} = \frac{\text{number of independent variables in regression equation}}{\text{number of nonmissing independent variables}}\left(\Sigma B_i Z_i\right)$$

where B_i is the standardized regression coefficient β, Z_i is the standardized independent variable, and the summation is over all nonmissing variables entered in the regression equation. The standardized residual is then calculated from the weighted predictor.

If the proportion of independent variables which are missing exceeds the mdrp, the output residuals and/or predictors will not be weighted, but will have the value of 999.0.

If Option 13 is not specified no weighting is done.

OPTION 14 *Suppresses the printing of axes on the plots of standardized predictor versus standardized residual.* (These plots are obtained by specifying Statistic 6.)

OPTION 15 *Output of means and standard deviations.* This option causes means and standard deviations to be output on the raw-output-data file in 8F10.4 format. Means and standard deviations are output in separate sets, corresponding to separate VARIABLES lists, when more than one such list appears on the procedure card.

Cards obtained with Options 8 and 15 can be used for input on subsequent REGRESSION runs using Options 4 and 5. The user must note that the format is unalterable, and is not suitable for mean and standard deviation values greater than or equal to 1,000,000 and less than or equal to −100,000. *Option 15 cannot be used with matrix input.*

20.7 STATISTICS AVAILABLE WITH SUBPROGRAM REGRESSION

The optional statistics available with a REGRESSION run are specified by the user on a separate STATISTICS card. If present, the STATISTICS card is placed immediately after the OPTIONS card. If there is no OPTIONS card, the STATISTICS card is placed directly after the REGRESSION procedure card.

There are seven optional statistics available. The STATISTICS card has the usual format: the control word STATISTICS beginning in column 1 and a list of desired statistics beginning in column 16. In place of the list of statistics the user may put the keyword ALL, in which case all statistics are called for. However, only one of Statistics 1, 3, and 7 will be printed, and Statistics 4, 5, and 6 are in effect *only* when the user has specified on the REGRESSION procedure card that an analysis of residuals is to be performed. (The manner in which an analysis of residuals is called for is discussed in Sec. 20.2.2.4.) If none of the following statistics are desired, no STATISTICS card is placed in the deck.

STATISTIC 1 *Printout of the correlation matrix (matrices).* If this statistic is called for, a correlation matrix is printed for each VARIABLES= list appearing on the REGRESSION procedure card.

STATISTIC 2 *Means, standard deviations, and number of valid cases.* This statistic causes means and standard deviations to be printed for each VARIABLES= list appearing on the REGRESSION procedure card. In addition, the number of *valid cases* on which means and standard deviations have been computed are printed. For pairwise deletion, the number of valid cases is the number of cases not having missing values for a given variable. For listwise deletion, the number of valid cases is the number of cases not having missing values on *any* of the variables on the VARIABLES= list. Note that missing-data values are all counted as valid when Option 1 is specified by the user.

STATISTIC 3 *Forced printing of the correlation matrix and warning of bad elements.* Selection of this statistic forces the printing of the correlation matrix in the event that one or more correlation coefficients cannot be calculated. Correlation coefficients that cannot be calculated are represented in the matrix by the value of 99. If Statistic 3 is used without Statistic 1 or 7, the matrix will be printed only if one or more correlation coefficients are incalculable. If Statistic 3 is used *with* Statistics 1 or 7 (as when the keyword ALL is used), the correlation matrix will always be printed.

Statistic 3 is useful as a warning when the user is performing REGRESSION analysis on variables whose characteristics are somewhat unfamiliar. The appearance of the correlation matrix will alert the user to bad variables which should be dropped from the analysis the next time around.

Statistics 4 through 6 are used in connection with analysis of residuals. Both Statistics 4 and 5 are meaningful only if the file has been sorted in some relevant fashion (see Sec. 20.5).

STATISTIC 4 *Causes output of a plot of standardized residuals against the sequence of cases in a file.* This plot is obtained for only the *last* regression equation designated by the RESID=0 keyword. The plot is accompanied by a listing of unstandardized Y, Y', and residuals. Only 500 cases will be plotted.

STATISTIC 5 *Computes the Durbin-Watson statistic for residuals.* This statistic is based on the differences between the residuals of adjacent cases in a *sequenced* file and is used in a test for autocorrelation.

$$\text{Durbin-Watson statistic} = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

where e_i is the residual for case i and n is the number of cases.

STATISTIC 6 *Requests a plot of standardized residuals against standardized Y' values with residuals on the vertical axis.* Two plots are printed per page. These plots can be examined for abnormalities as described in Sec. 20.1.2.6.

STATISTIC 7 *Printout of correlation matrix and number of cases.* This statistic may be requested when pairwise deletion is specified (Option 2). When listwise deletion of missing data (default option) or inclusion of missing data (Option 1) is used, requesting Statistic 7 will cause Statistic 1 to be printed instead.

Statistic 7 causes a matrix to be printed in which the lower triangle contains the correlation coefficients, the upper triangle contains the number of cases used in building each correlation coefficient, and the diagonal contains the number of nonmissing cases for each variable. If both Statistics 7 and 1 are requested, only Statistic 7 will be printed.

20.8 PROGRAM LIMITATIONS OF SUBPROGRAM REGRESSION

LIMITATION 1 A maximum of 10 VARIABLES lists is allowed on a REGRESSION procedure card. Stated another way, a maximum of 10 correlation matrices will be constructed from raw data (or read with matrix input) on a single REGRESSION run.

LIMITATION 2 A maximum of 50 REGRESSION design statements is allowed per procedure card, irrespective of the number of VARIABLES lists appearing on the card.

LIMITATION 3 A maximum of 100 variables is allowed on any VARIABLES list, and a maximum of 200 variable names is allowed in the combined VARIABLES lists of the procedure card. Variables occurring in more than one VARIABLES list are counted once for each list.

LIMITATION 4 A maximum of 400 variable names may be used in the combined REGRESSION design statements on any procedure card. Each occurrence of a variable as either a dependent or independent variable counts as one in this total. A maximum of 100 *different* variables is allowed for a single REGRESSION design statement.

LIMITATION 5 This limitation applies to the IBM 360-370 version of SPSS. Other users should consult Appendix F, G, or H, or their local computation center personnel.

```
//PW=[redacted] ...REGION=200K
// EXEC SPSS
//SYSIN DD *
RUN NAME          BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA*A, YEMEN
VARIABLE LIST     SERVRAD,LPCD,NLENGTH,DBAR,COST
INPUT MEDIUM      CARD
N OF CASES        16
INPUT FORMAT      FREEFIELD
COMPUTE           LSERVRAD=LN(SERVRAD)
COMPUTE           LFLOW=LN(LPCD)
COMPUTE           LNLENGTH=LN(NLENGTH)
COMPUTE           LDBAR=LN(DBAR)
COMPUTE           LCOST=LN(COST)
REGRESSION        VARIABLES=SERVRAD,LPCD,NLENGTH,DBAR,COST,LCOST,LSERVRAD,LNLENGTH,
                  LFLOW,LDBAR/
                  REGRESSION=COST WITH NLENGTH,DBAR(2)/
                  REGRESSION=LCOST WITH LNLENGTH,LDBAR(2)/
                  REGRESSION=LCOST WITH NLENGTH,DBAR(2)/
                  REGRESSION=LCOST WITH LSERVRAD(4),LFLOW(2)/
                  REGRESSION=LCOST WITH SERVRAD,LPCD(2)/
                  REGRESSION=LDBAR WITH LFLOW(4),LNLENGTH(2)/
                  REGRESSION=LDBAR WITH LPCD,NLENGTH(2)
STATISTICS        1
READ INPUT DATA
100,20,1450,59.0,26.7
100,50,1450,88.0,37.4
100,100,1450,110.5,46.4
50,20,3007,42.6,44.4
50,50,3007,61.2,58.1
50,100,3007,79.8,72.9
8,50,10788,35.6,142.1
8,100,10788,43.0,160.6
100,20,2080,45.8,31.9
100,50,2080,66.8,42.2
100,100,2080,85.8,52.8
50,20,3780,35.4,49.8
50,50,3780,50.8,63.0
50,100,3780,65.2,76.8
13,50,12580,34.4,159.5
13,100,12580,42.7,184.9
FINISH
//SYSPRINT DD SYSOUT=A
//
CARD COUNT= 00044
```

-43-

SPSS BATCH SYSTEM

SPSS FOR OS/360, VERSION H, RELEASE ...   08/14/81   PAGE 1

```
SPSS FOR OS/360, VERSION H, RELEASE 8.0, MAY 15, 1979

DEFAULT SPACE ALLOCATION..      ALLOWS FOR..     51 TRANSFORMATIONS
WORKSPACE      35840 BYTES                      204 RECODE VALUES + LAG VARIABLES
TRANSPACE       5120 BYTES                      822 IF/COMPUTE OPERATIONS

          1 RUN NAME        BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN
          2 VARIABLE LIST   SERVRAD,LPCD,NLENGTH,DBAR,COST
          3 INPUT MEDIUM    CARD
          4 N OF CASES      16
          5 INPUT FORMAT    FREEFIELD
          6 COMPUTE         LSERVRAD=LN(SERVRAD)
          7 COMPUTE         LFLOW=LN(LPCD)
          8 COMPUTE         LNLENGTH=LN(NLENGTH)
          9 COMPUTE         LDBAR=LN(DBAR)
         10 COMPUTE         LCOST=LN(COST)
         11 REGRESSION      VARIABLES=SERVRAD,LPCD,NLENGTH,DBAR,COST,LCOST,LSERVRAD,LNLENGTH,
         12                 LFLOW,LDBAR/
         13                 REGRESSION=COST WITH NLENGTH,DBAR(2)/
         14                 REGRESSION=LCOST WITH LNLENGTH,LDBAR(2)/
         15                 REGRESSION=LCOST WITH NLENGTH,DBAR(2)/
         16                 REGRESSION=LCOST WITH LSERVRAD(4),LFLOW(2)/
         17                 REGRESSION=LCOST WITH SERVRAD,LPCD(2)/
         18                 REGRESSION=LDBAR WITH LFLOW(4),LNLENGTH(2)/
         19                 REGRESSION=LDBAR WITH LPCD,NLENGTH(2)
         20 STATISTICS      1

***** REGRESSION PROBLEM REQUIRES   1680 BYTES WORKSPACE, NOT INCLUDING RESIDUALS *****

         21 READ INPUT DATA
```

44

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

FILE NONAME (CREATION DATE = 08/14/81)

CORRELATION COEFFICIENTS

A VALUE OF 99.00000 IS PRINTED
IF A COEFFICIENT CANNOT BE COMPUTED.

	SERVRAD	LPCD	NLENGTH	DBAR	COST	LCOST	LSERVRAD	LNLENGTH	LFLOW	LDBAR
SERVRAD	1.00000	-0.19481	-0.87339	0.67316	-0.86269	-0.89818	0.94606	-0.94793	-0.23080	0.70006
LPCD	-0.19481	1.00000	0.24060	0.46821	0.42555	0.50104	-0.23009	0.22346	0.97232	0.45341
NLENGTH	-0.87339	0.24060	1.00000	-0.62032	0.97868	0.93682	-0.95330	0.96865	0.28505	-0.66913
DBAR	0.67316	0.46821	-0.62032	1.00000	-0.48785	-0.43599	0.63729	-0.69910	0.45970	0.98617
COST	-0.86269	0.42555	0.97868	-0.48785	1.00000	0.97479	-0.93840	0.94827	0.46264	-0.53007
LCOST	-0.89818	0.50104	0.93682	-0.43599	0.97479	1.00000	-0.93067	0.94736	0.54967	-0.47811
LSERVRAD	0.94606	-0.23009	-0.95330	0.63729	-0.93840	-0.93067	1.00000	-0.96597	-0.27260	0.67930
LNLENGTH	-0.94793	0.22346	0.96865	-0.69910	0.94827	0.94736	-0.96597	1.00000	0.26474	-0.73338
LFLOW	-0.23080	0.97232	0.28505	0.45970	0.46264	0.54967	-0.27260	0.26474	1.00000	0.44611
LDBAR	0.70006	0.45341	-0.66913	0.98617	-0.53007	-0.47811	0.67930	-0.73338	0.44611	1.00000

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

LNLENGTH	-0.94793	0.22346	0.96865	-0.69910	0.94827	0.94736	1.00000	-0.96597	-0.27260	0.67930
LFLOW	-0.23080	0.97232	0.28505	0.45970	0.46264	0.54967	-0.96597	1.00000	0.26474	-0.73338
LDBAR	0.70006	0.45341	-0.66913	0.98617	-0.53007	-0.47811	0.67930	-0.73338	0.44611	1.00000

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN 08/14/81 PAGE 3

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 1

DEPENDENT VARIABLE.. COST

VARIABLE(S) ENTERED ON STEP NUMBER 1.. NLENGTH
DBAR

MULTIPLE R	0.99042		ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.98093		REGRESSION	2.	40114.59584	20057.29792	334.35181
ADJUSTED R SQUARE	0.97800		RESIDUAL	13.	779.85371	59.98875	
STANDARD ERROR	7.74524						

---------------- VARIABLES IN THE EQUATION -----------------

VARIABLE	B	BETA	STD ERROR B	F
NLENGTH	0.1374646D-01	1.09893	0.00061	506.462
DBAR	0.4529557	0.19384	0.11410	15.759
(CONSTANT)	-15.44930			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

ALL VARIABLES ARE IN THE EQUATION

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN 08/14/81 PAGE 4

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 1

DEPENDENT VARIABLE.. COST

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
NLENGTH	0.97868	0.95781	0.95781	0.97868	0.1374646D-01	1.09893
DBAR	0.99042	0.98093	0.02312	-0.48785	0.4529557	0.19384
(CONSTANT)					-15.44930	

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN 08/14/81 PAGE 5

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 2

DEPENDENT VARIABLE.. LCOST

VARIABLE(S) ENTERED ON STEP NUMBER 1.. LDBAR
LNLENGTH

-45-

STANDARD ERROR 0.01984

---------------- VARIABLES IN THE EQUATION ----------------

VARIABLE	B	BETA	STD ERROR B	F
LDBAR	0.7889383	0.46882	0.02086	1431.075
LNLENGTH	1.012085	1.29119	0.00971	10854.878
(CONSTANT)	-7.279131			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

ALL VARIABLES ARE IN THE EQUATION

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN 08/14/81 PAGE 6

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 2

DEPENDENT VARIABLE.. LCOST

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
LDBAR	0.47811	0.22859	0.22859	-0.47811	0.7889383	0.46882
LNLENGTH	0.99954	0.99908	0.77049	0.94736	1.012085	1.29119
(CONSTANT)					-7.279131	

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

-46-

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN 08/14/81 PAGE 7

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 3

DEPENDENT VARIABLE.. LCOST

VARIABLE(S) ENTERED ON STEP NUMBER 1.. NLENGTH
DBAR

			ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
MULTIPLE R	0.95492		REGRESSION	2.	5.05773	2.52886	67.25010
R SQUARE	0.91166		RESIDUAL	13.	0.48885	0.03760	
ADJUSTED R SQUARE	0.89831						
STANDARD ERROR	0.19392						

---------------- VARIABLES IN THE EQUATION ----------------

VARIABLE	B	BETA	STD ERROR B	F
NLENGTH	0.1577962D-03	1.08316	0.00002	106.462
DBAR	0.6420215D-02	0.23592	0.00286	5.051
(CONSTANT)	3.027702			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

ALL VARIABLES ARE IN THE EQUATION

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

--------TOP-OF-FORM--------

FILE NONAME (CREATION DATE = 08/14/81)

* * * * * * * * * * * * MULTIPLE REGRESSION * * * * * * * * * * * * VARIABLE LIST 1

DEPENDENT VARIABLE.. LCOST REGRESSION LIST 3

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
NLENGTH	0.93682	0.87762	0.87762	0.93682	0.1577962D-03	1.08316
DBAR	0.95492	0.91186	0.03424	-0.43599	0.6420215D-02	0.23592
(CONSTANT)					3.027702	

--------TOP-OF-FORM--------

FILE NONAME (CREATION DATE = 08/14/81)

* * * * * * * * * * * * MULTIPLE REGRESSION * * * * * * * * * * * * VARIABLE LIST 1

DEPENDENT VARIABLE.. LCOST REGRESSION LIST 4

VARIABLE(S) ENTERED ON STEP NUMBER 1.. LSERVRAD

MULTIPLE R	0.93067	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.86615	REGRESSION	1.	4.80418	4.80418	90.59643
ADJUSTED R SQUARE	0.85659	RESIDUAL	14.	0.74240	0.05303	
STANDARD ERROR	0.23028					

------------ VARIABLES IN THE EQUATION ------------

VARIABLE	B	BETA	STD ERROR B	F
LSERVRAD	-0.6096454	-0.93067	0.06405	90.596
(CONSTANT)	6.474811			

------------ VARIABLES NOT IN THE EQUATION ------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
LFLOW	0.31973	0.84084	0.92569	31.371

* *

VARIABLE(S) ENTERED ON STEP NUMBER 2.. LFLOW

MULTIPLE R	0.98020	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.96079	REGRESSION	2.	5.32907	2.66453	159.25303
ADJUSTED R SQUARE	0.95475	RESIDUAL	13.	0.21751	0.01673	
STANDARD ERROR	0.12935					

------------ VARIABLES IN THE EQUATION ------------

VARIABLE	B	BETA	STD ERROR B	F
LSERVRAD	-0.5525516	-0.84351	0.03739	218.345
LFLOW	0.3017877	0.31973	0.05388	31.371
(CONSTANT)	5.069398			

------------ VARIABLES NOT IN THE EQUATION ------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

MAXIMUM STEP REACHED

(CONSTANT) 5.069398

MAXIMUM STEP REACHED

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 4

DEPENDENT VARIABLE.. LCOST

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
LSERVRAD	0.93067	0.86615	0.86615	-0.93067	-0.5525516	-0.84351
LFLOW	0.98020	0.96079	0.09463	0.54967	0.3017877	0.31973
(CONSTANT)					5.069398	

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 5

DEPENDENT VARIABLE.. LCOST

VARIABLE(S) ENTERED ON STEP NUMBER 1.. SERVRAD
LPCD

MULTIPLE R	0.95773	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.91724	REGRESSION	2.	5.08757	2.54378	72.04427
ADJUSTED R SQUARE	0.90451	RESIDUAL	13.	0.45901	0.03531	
STANDARD ERROR	0.18791					

------------------ VARIABLES IN THE EQUATION ------------------

VARIABLE	B	BETA	STD ERROR B	F
SERVRAD	-0.1385608D-01	-0.83216	0.00135	104.653
LPCD	0.6199952D-02	0.33892	0.00149	17.360
(CONSTANT)	4.609737			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

ALL VARIABLES ARE IN THE EQUATION

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 5

DEPENDENT VARIABLE.. LCOST

SUMMARY TABLE

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
SERVRAD	0.89818	0.80673	0.80673	-0.89818	-0.1385608D-01	-0.83216
LPCD	0.95773	0.91724	0.11051	0.50104	0.6199952D-02	0.33892
(CONSTANT)					4.609737	

- - - - - - - - - T O P - O F - F O R M - - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN 08/14/81 PAGE 13

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 6

DEPENDENT VARIABLE.. LOBAR

VARIABLE(S) ENTERED ON STEP NUMBER 1.. LFLOW

MULTIPLE R	0.44611
R SQUARE	0.19902
ADJUSTED R SQUARE	0.14180
STANDARD ERROR	0.33475

ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
REGRESSION	1.	0.38980	0.38980	3.47854
RESIDUAL	14.	1.56884	0.11206	

---------------- VARIABLES IN THE EQUATION -----------------

VARIABLE	B	BETA	STD ERROR B	F
LFLOW	0.2502215	0.44611	0.13416	3.479
(CONSTANT)	3.030987			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
LNLENGTH	-0.91566	-0.98661	0.92991	475.531

49

* *

VARIABLE(S) ENTERED ON STEP NUMBER 2.. LNLENGTH

MULTIPLE R	0.98929
R SQUARE	0.97869
ADJUSTED R SQUARE	0.97541
STANDARD ERROR	0.05667

ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
REGRESSION	2.	1.91689	0.95845	298.45738
RESIDUAL	13.	0.04175	0.00321	

---------------- VARIABLES IN THE EQUATION -----------------

VARIABLE	B	BETA	STD ERROR B	F
LFLOW	0.3861885	0.68853	0.02355	268.876
LNLENGTH	-0.4265072	-0.91566	0.01956	475.531
(CONSTANT)	5.985553			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

MAXIMUM STEP REACHED

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

- - - - - - - - - T O P - O F - F O R M - - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN 08/14/81 PAGE 14

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 6

DEPENDENT VARIABLE.. LOBAR

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN   08/14/81   PAGE 14

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * * *   VARIABLE LIST 1
REGRESSION LIST 6

DEPENDENT VARIABLE.. LOBAR

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
LFLOW	0.44611	0.19902	0.19902	0.44611	0.3861885	0.68853
LNLENGTH	0.98929	0.97869	0.77967	-0.73338	-0.4265072	-0.91566
(CONSTANT)					5.985553	

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN   08/14/81   PAGE 15

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * * *   VARIABLE LIST 1
REGRESSION LIST 7

DEPENDENT VARIABLE.. LOBAR

VARIABLE(S) ENTERED ON STEP NUMBER 1.. LPCD
NLENGTH

MULTIPLE R	0.92110	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.84842	REGRESSION	2.	1.66176	0.83088	36.38284
ADJUSTED R SQUARE	0.82510	RESIDUAL	13.	0.29688	0.02284	
STANDARD ERROR	0.15112					

---------------- VARIABLES IN THE EQUATION ----------------

VARIABLE	B	BETA	STD ERROR B	F
LPCD	0.7089254D-02	0.65215	0.00121	34.365
NLENGTH	-0.7151045D-04	-0.82604	0.00001	55.134
(CONSTANT)	3.930578			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

ALL VARIABLES ARE IN THE EQUATION

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN   08/14/81   PAGE 16

FILE NONAME (CREATION DATE = 08/14/81)

* M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * * *   VARIABLE LIST 1
REGRESSION LIST 7

DEPENDENT VARIABLE.. LOBAR

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
LPCD	0.45341	0.20558	0.20558	0.45341	0.7089254D-02	0.65215
NLENGTH	0.92110	0.84842	0.64284	-0.66913	-0.7151045D-04	-0.82604
(CONSTANT)					3.930578	

- - - - - - - - - T O P - O F - F O R M - - - - - - - - -

BRANCH NETWORK STATISTICS FOR ZONE 1 & 2 , SANA'A, YEMEN   08/14/81   PAGE 17

```
TRANSPACE REQUIRED..      500 BYTES
      5 TRANSFORMATIONS
      0 RECODE VALUES + LAG VARIABLES
     15 IF/COMPUTE OPERATIONS

CPU TIME REQUIRED..    0.66 SECONDS


                 22 FINISH

         NORMAL END OF JOB.
            22 CONTROL CARDS WERE PROCESSED.
             0 ERRORS WERE DETECTED.
```


KOLSKY SPSSTEST 18-PAGES 394-LINES 0-PLOTS # UNC.ES.F219K

KOLSKY SPSSTEST 18-PAGES 394-LINES 0-PLOTS # UNC.ES.F219K
KOLSKY SPSSTEST 18-PAGES 394-LINES 0-PLOTS # UNC.ES.F219K
KOLSKY SPSSTEST 18-PAGES 394-LINES 0-PLOTS # UNC.ES.F219K

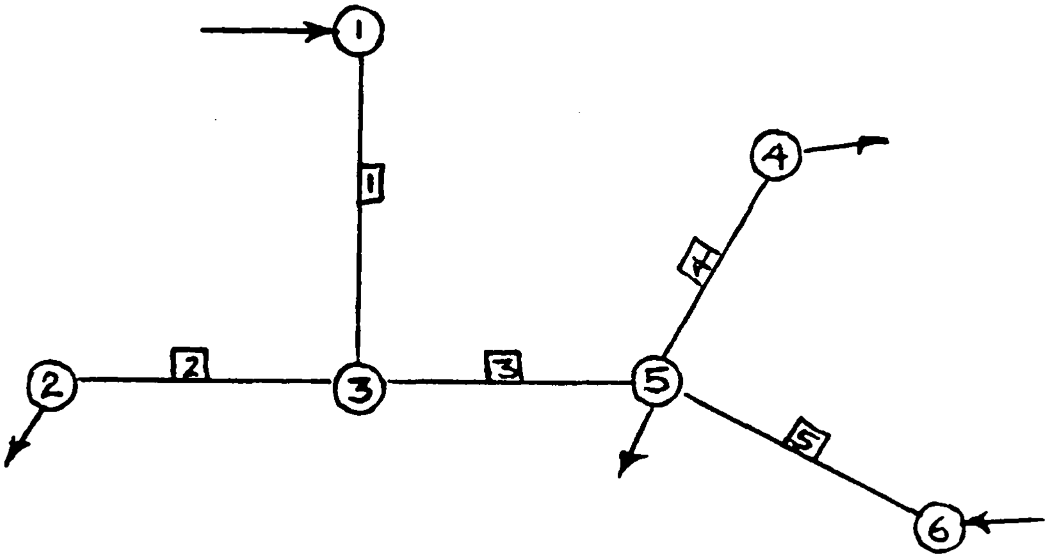
APPENDIX G

LINEAR PROGRAMMING MODEL
FOR LEAST COST DESIGN OF BRANCHED (NON-LOOPED)
WATER DISTRIBUTION SYSTEMS AND USER INSTRUCTIONS
FOR BASIC MICROCOMPUTER PROGRAMS
"NODELINK" AND "BRANCH"*

1982

*Developed by Keith Little, Department of Environmental Sciences and Engineering, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514.

BRANCHED NETWORK LEAST COST DESIGN EXAMPLE:



NETWORK CHARACTERISTICS

Node #	Elevation	Input	Demand	HGL of Input	Link #	Length
1	1m	0.5 lps	0 lps	20.0 m	1	500 m
2	3	0	0.4	---	2	1200
3	1	0	0	---	3	1000
4	5	0	0.15	---	4	1500
5	2	0	0.2	---	5	1500
6	2	0.25	0	12.0		

Link #1 is an existing 50 cm with friction coefficient of 100.

DESIGN CRITERIA

Available pipe dia's are 2 cm @ $100/m and 50 cm @ $150/m
Friction coeff. in Hazen-Williams eq'n for new pipe is 140
Minimum residual head at all terminal nodes is 1.0 m
Peaking factor = 5.0

OBJECTIVE: Minimize construction costs while meeting design criteria.

1. <u>Linear Programming Formulation of Example Design Problem</u>

Let's express the design criteria as linear, mathematical equations/inequalities.

Let

x_1^* = the length of 2 cm pipe in link 1

x_2 = the length of 50 cm pipe in link 1

x_3 = the length of 2 cm pipe in link 2

x_4 = the length of 50 cm pipe in link 2

x_5 = the length of 2 cm pipe in link 3

x_6 = the length of 50 cm pipe in link 3

x_7 = the length of 2 cm pipe in link 4

x_8 = the length of 50 cm pipe in link 4

x_9 = the length of 2 cm pipe in link 5

x_{10} = the length of 50 cm pipe in link 5

(* Note that link 1 exists as 50 cm but we'll define x_1 anyway)

Let

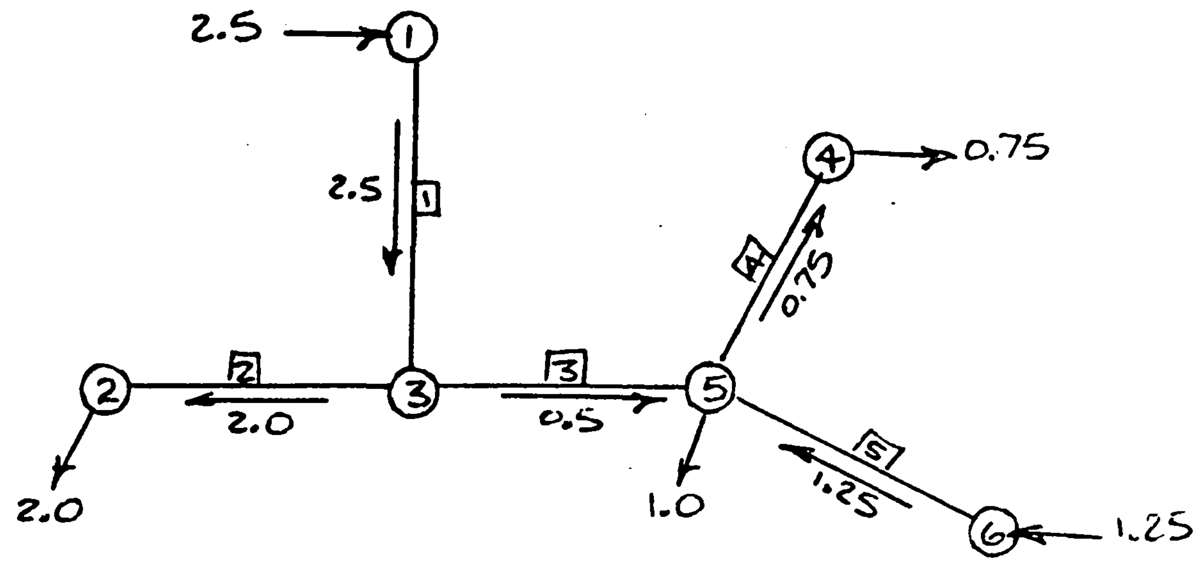
hij = the slope of the hydraulic gradient in pipe j of link i for the link flowrate

From the Hazen-Williams relationship (for Q in lps, L in m, Dia in cm, and h in m/m)

$$hij = 1.62 \times 10^5 \left(\frac{Q_j}{F_j}\right)^{1.85} (\text{Dia of pipe } i)^{-4.87}$$

F_j is the friction coefficient for pipe j.

The design link flows are found to be (using the peaking factor)...



The hij's can then be determined

$h_{11} = 6.02$	$h_{32} = 2.56 \times 10^{-8}$
$h_{12} = 9.37 \times 10^{-7}$	$h_{41} = 0.35$
$h_{21} = 2.14$	$h_{42} = 5.42 \times 10^{-8}$
$h_{22} = 3.33 \times 10^{-7}$	$h_{51} = 0.90$
$h_{31} = 0.17$	$h_{52} = 1.40 \times 10^{-7}$

Let's now write the headloss constraints. We need to define a reference input node so that headloss constraints can be written from this reference node to any other nodes that may be necessary. Let the reference input node be node #1.

There are two types of headloss constraints. The first type specifies that the headloss between the reference input node and each of the non-input nodes in the system is $\leq$ that headloss which will just satisfy the minimum residual head requirement at the node. (It is generally sufficient to write these constraints only for "terminal" nodes, i.e. those nodes at the extremities of the system. If an interior node has a high elevation relative to the rest of the system, a constraint of this type should also be written for it.)

The headloss constraint for non-input, terminal node 2 is

$$\underbrace{h_{11}x_1 + h_{12}x_2}_{\text{link 1}} + \underbrace{h_{21}x_3 + h_{22}x_4}_{\text{link 2}} \leq \underbrace{20.0 - (3.0 + 1.0)}_{\text{Maximum headloss}}$$

i.e., the headloss in link 1 + the headloss in link 2 must be $\leq$ the maximum allowable head loss.

Similarly, for non-input terminal node 4, the headloss constraint is

$$h_{11}x_1 + h_{12}x_2 + h_{31}x_5 + h_{32}x_6 + h_{41}x_7 + h_{42}x_8 \leq 20.0 - (1.0 + 5.0)$$

The second type of headloss constraint is for input nodes. It can be shown that, for every input node (a) the HGL of the reference input minus the headloss between the reference and some point on the path between the reference and the other input, must be equal to (b) the HGL of the (non-reference) input minus the headlosses to the same intermediate point along the same path. This is simply an awkward way to say that the pressure at any point inthe system is the same regardless of how the water got there!

For the example network, such a headloss constraint must be written between the reference node 1 and the other input node 6. Letting node 3 be the point along the path from 1 to 6 at which the headlosses must be equal, we can write

$$20.0 - (h_{11}x_1 + h_{12}x_2) = 12.0 - (h_{51}x_9 + h_{52}x_{10}) + (h_{31}x_5 + h_{32}x_6)$$

or

$$-h_{11}x_1 - h_{12}x_2 - h_{31}x_5 - h_{32}x_6 + h_{51}x_9 + h_{52}x_{10} = 12.0 - 20.0$$

The remaining design constraints simply state that the sum of the length's of the pipes selected for any link will be equal to the length of that link. There is one such equality constraint for every link (existing also) in the network. For the example, we can write...

$$x_1 + x_2 = 500$$
$$x_3 + x_4 = 1200$$
$$x_5 + x_6 = 1000$$
$$x_7 + x_8 = 1500$$
$$x_9 + x_{10} = 1500$$

Linear programming requires (as does our design) that the decision variables (the x's) be non-negative, or

$$x_1, x_2, \ldots, x_{10} \geq 0$$

These non-negativity constraints do not explicitly appear in the LP formulation but are understood to exist.

We would now be ready to solve the LP except for two problems with our formulation. LP does not know how to solve inequality constraints (which do result from the Type 1 headloss constraints) now does it know how to solve equalities with negative right-hand-sides (which may result from the Type 2 headloss constraints).

The inequalities are easily made equalities by defining new variables that represent the "slack" headloss available on any path between the reference and the non-input nodes for which headloss constraints have been written. If any slack variable has a non-zero value in the solution, it simply implies that the constraint was not binding (it could have been left out without affecting the solution). Defining slack variables x_{11} and x_{12} for the two Type 1 headloss constraints, we can rewrite them as equalities...

$$h_{11}x_1 + h_{12}x_2 + h_{21}x_3 + h_{22}x_4 + x_{11} = 20.0 - (1.0 + 1.0)$$
$$h_{11}x_1 + h_{12}x_2 + h_{31}x_5 + x_{32}x_6 + h_{41}x_7 + h_{42}x_8 + x_{12} = 20.0 - (1.0 + 5.0)$$

Every decision variable (x's) must have an associated cost coefficient so that they can appear in the objective function. Since we don't want to prevent the slack variables x_{11} and x_{12} from appearing in the solution, we let $c_{11}, c_{12} = 0$.

The negative right hand side resulting from the Type 2 headloss constraint is resolved by multiplying both sides by -1. The constraint for input node 6 becomes

$$h_{11}x_1 + h_{12}x_2 + h_{31}x_5 + h_{32}x_6 - h_{51}x_9 - h_{52}x_{10} = 20.0 - 12.0$$

For reasons that won't be explored here, we must add an "artificial" variable to the left-hand-side of every length and Type 2 headloss constraint. There variables are "artificial" because all constraint are already legitimate equalities, and adding anything to only one side of an equality is illegitimate, hence they are articial. To prevent these artificial variables from appearing in the solution we will penalize them heavily in the objective function. Let the cost coefficients of the artifical variables x_{13}, $x_{14} \ldots, x_{18}$ be much larger than the largest legitimate cost coefficient, say 10 x \$150 = \$1500. If any of the artificial variables appear in the solution, it will mean that the problem is infeasible.

While we're on the subject of manipulating cost coefficients, we must fix the cost coefficients of the candidate pipes in the link which already exists to ensure that no non-existing diameter appears in the solution. For the non-existing candidate diameters, we'll let their cost coefficients be the same as the artifical variables, or \$1500. Since the cost of the existing diameter is \$0, that's what it will be. We have then c_1 = \$1500 and c_2 = \$0. Finally, the objective function that LP will seek to minimize subject to the constraints is the mathematical expression of the construction costs. The cost of any pipe j is its length (x_j) times its unit cost, c_j. The objective function, including slack and artificial variables, is then

$$\min z = \underbrace{c_1x_1 + \dots + c_{10}x_{10}}_{\text{original variables}} + \underbrace{c_{11}x_{11} + c_{12}x_{12}}_{\text{slack variables}} + \underbrace{c_{13}x_{13} + \dots + c_{18}x_{18}}_{\text{artificial variables}}$$

LP codes are written to either maximize or minimize. The code used in "BRANCH" is a maximization. Since maximizing z is equivalent to minimizing -z, we write

$$\min -z = -c_1x_1 - c_2x_2 - \dots - c_{18}x_{18}$$

Writing all the constraints and the objective function in a matrix of coefficients of the decision variables where each row is a constraint (the last row is the objective function) and each column is a decision variable, the LP formulation is complete and expressed in a form suitable to LP's simplex algorithm.

		PIPE LENGTH VARIABLES										SLACK VAR'S		ARTIFICIAL VARIABLES						RHS
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	
CONSTRAINTS	TYPE–NODE	6.02	9.37×10^{-7}	2.14	3.33×10^{-7}	0	0	0	0	0	0	1	0	0	0	0	0	0	0	116
	TYPE–NODE	6.02	9.37×10^{-7}	0	0	0.17	2.56×10^{-8}	0.35	5.42×10^{-8}	0	0	0	1	0	0	0	0	0	0	14
	NODE–NODE	6.02	9.37×10^{-7}	0	C	0.17	2.56×10^{-8}	0	0	-.90	-1.40×10^{-7}	0	0	1	0	0	0	0	0	8
	LENGTH	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	500
		0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1200
		0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1000
		0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1500
		0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1500
OBJ FN		-$1500	$0	-$100	-$150	-$100	-$150	-$100	-$150	-$100	-$150	$0	$0	-$1500	-$1500	-$1500	-$1500	-$1500	-$1500	-Z
		LINK 1		LINK 2		LINK 3		LINK 4		LINK 5										

2. User Instructions for "Nodelink" and "Branch"

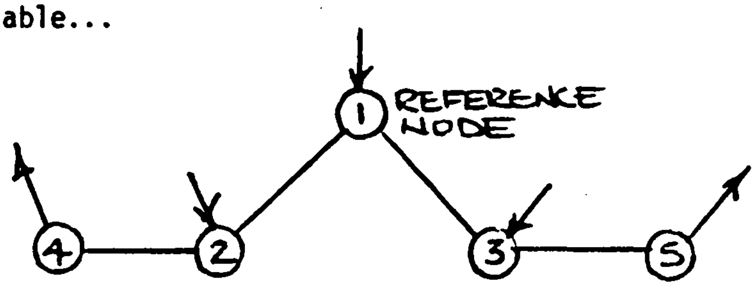
The least cost design of branched water distribution systems is accomplished by means of the computer programs, "NODELINK" and "BRANCH." Both NODELINK and BRANCH are written in the Basic programming language, a language for which some dialect is supported by virtually all microcomputers. The major difference in the language among various types of microcomputers is input/output commands,i.e. READ, INPUT, PRINT, LPRINT commands. This is especially true when writing to disk files. Source listings for NODELINK and BRANCH are in the Appendix. The listings are for the CP/M based Osborne 1 microcomputer.

NODELINK and BRANCH function sequentially. First, NODELINK reads the original data that describe network characteristics and design criteria. NODELINK transforms this data into a format suitable for the linear programming algorithm and writes this transformed data to a sequential data file on a diskette. BRANCH reads the transformed data and activates the linear programming algorithm which iterates until a least cost design is found or the problem is determined to be infeasible. NODELINK and BRANCH were designed as separate programs in order to conserve computer memory making more memory available for data manipulation. With minor modifications, NODELINK and BRANCH could be merged into a single program. This would be necessary if a disk drive were not available.

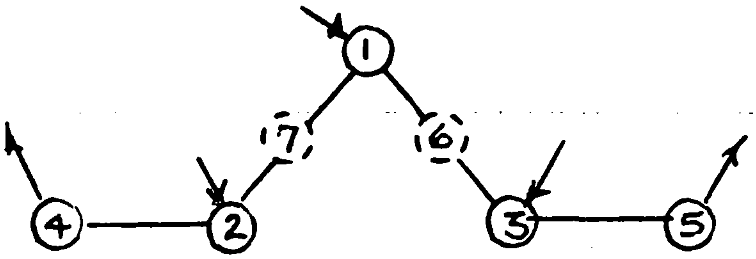
Before illustrating the data input format for the example design problem a few comments on the node and link numbering system and multiple sources are appropriate.

- The nodes and links may be numbered arbitrarily, but the sequence of node numbers and link numbers must be an integer sequence beginning with 1. In other words, for a 2-link, 3-node (there are always n + 1 nodes in an n-link branched network) network, the links must be numbered 1, 2 and the nodes 1, 2, 3 (not 1.2, 3, 5 for example). Within this framework, the actual numbering scheme can be arbitrary.

- If the network has more than one input node, the first node on the path(s) from the reference node to the other input node(s) must be a non-input node. This is so that Type 2 headloss constraints can be written to these interior nodes. For example, the following network is not acceptable...



The pseudo-nodes (6) and (7) must be introduced...



The pseudo-nodes and the resulting new links are then treated just as if they were original nodes and links.

Let's input the data for the example design problem. The data are entered in DATA statements beginning with data statement number 2000 and continuing in sequential order (using any desired increment).

We'll use an increment of 10. Each item of data is separated from others by a comma.

DATA Statement 2000:

This statement is for project identification. Any sequence of alphanumeric characters is acceptable if the sequence begins with a letter. We enter ...

2000 DATA Project Example Network

DATA Statement 2010:

The first entry on 2010 is the number of links, NL. The second entry, NT, is the number of headloss constraints. For single source networks with no Type 1 headloss constraints written for interior (non-terminal) nodes, NT equals the number of terminal nodes -1 (if the input is at a terminal node) or the number of terminal nodes (if the input is at an interior node). For multiple source networks with no interior node Type 1 constraints and the reference node at a terminal node, NT equals the number of terminal nodes -1 plus the number of non-reference node inputs at interior nodes. If the reference node is an interior node and no Type 1 interior node constraints are to be written, NT equals the number of terminal nodes plus the number of non-reference interior node inputs. If Type 1 headloss constraints are to be written for interior nodes, these nodes are treated as terminal nodes subject to Type 1 constraints in determining NT. (The total number of constraints will be NT + NL). For our example, there are 2 Type 1 terminal nodes and 1 Type 2 terminal node; therefore NT = 3. The last entry on 2010 is the number of candidate pipe diameters, ND.

2010 DATA 5, 3, 2

DATA Statement 2020:

The first entry is the flow peaking factor, PF. Next is the minimum residual head at the nodes, MR. Finally the hydraulic grade line elevation at the reference node is entered, RH.

2020 DATA 5.0, 1.0, 20.0

DATA Statement 2030:

The first entry is the total number of links already existing, NE. The next entry, NU, specifies the set of units that the data are in and that the solution will use.

For NU = 1: Flows are in million gallons per day and lengths and pipe diameters are in feet.

For NU = 2: Flows are in liters per second. Lengths are in meters, and diameters are in centimeters.

For NU = 3: Flows are in liters per second, lengths are in meters, and diameters are in inches.

The last entry, NO, is a print option that displays intermediate results if NO = 1 or no intermediate results if NO = 0. NO = 1 is much more entertaining.

2030 DATA 1, 2, 1

DATA Statement 2040:

The ND candidate pipe diameters are entered here. They should be entered in a logical order, for instance first entry smallest diameter and last entry largest diameter, because the solution does not give the diameters but rather a number that corresponds to the entry number in this DATA statement.

2040 DATA 2,50

DATA Statement

The ND unit pipe costs are entered here in the same sequence as the corresponding pipe diameters in 2040.

2050 DATA 100.00, 150.00

DATA Statements 2060 - 3100:

There is one DATA statement for each of the NL links in the network.

These statements contain the network geometry and hydraulic characteristics. It is essential that these statements be entered in consecutive order of the link numbers, i.e. 2060 DATA (Link 1), 1500 DATA (Link 2), ..., DATA (Link NL). The first entry is the link number. The second entry is "1" if the link exists, "0" otherwise. If the link exists, the third entry is the diameter number (not the diameter itself) of the existing pipe corresponding to the candidate diameters in statement 2040 (Note the existing diameter must be included as one of the candidate diameters). If the link does not exist, the third entry is "0".

The fourth link entry is the link's roughness coefficient.

The fifth link entry is its length.

The next three entries correspond to the link's "near node" with respect to the reference node. "Near" means if a path were traced from the reference node to the link, the link's "near node" would be reached first. The link's "far node" would be reached last. The first "near node" entry is the node number, the next entry is the node's input or demand. If the node has an input, enter the input flow. If the node has a demand, enter the flow demanded preceded by a "-". If the node is merely a junction or pseudo-node, enter "0." The last "near node" entry is the ground elevation if the node is a non-input node, or the elevation of the hydraulic grade line if it is an input node.

The next three entries are for the link's "far node." They are identical in format to those described for the "near node." (It is crucial that the "near node" and "far node" entries are in the proper sequence - this is part of the system by which the computer can understand the network geometry.)

The last link entry is "1" if the link's far node is a terminal node. It is "0" if the links far node is not a terminal node (but is an interior node). Finally, it is "2" if the link's <u>near</u> node is the reference node, regardless of whether the far node is terminal or interior. (If the reference node is interior, more than one link will have a "2" entry here.).

The link DATA statements for the example are...

```
2060 DATA  1, 1, 2, 100, 50, 1, 0.5, 20.0, 3, 0, 1.0, 2
2070 DATA  2, 0, 0, 140, 120, 3, 0, 1.0, 2, -0.4, 3.0, 1
2080 DATA  3, 0, 0, 140, 100, 3, 0, 1.0, 5, -0.2, 2.0, 0
2090 DATA  4, 0, 0, 140, 150, 5, -0.2, 2.0, 4, -0.15, 5.0, 1
3100 DATA  5, 0, 0, 140, 150, 5, -0.2, 2.0, 6, 0.25, 12.0, 1
```

After the last DATA statement, there is an END statement...

```
3120 END
```

The problem is now ready to run.

The example problem was run on an CP/M based Osborne 1 microcomputer which uses a Z80A microprocessor. The problem took approximately three minutes to solve. The solution is:

Link #	Length of 2 cm	Length of 50 cm
1	0 m	500 m
2	7.48	1192.52
3	85.09	914.92
4	0	1500
5	6.69	1493.31

The construction costs are $775,038.

APPENDIX

*SOURCE LISTING FOR NODELINK***

```
READ T$,NL,NT,ND,PF,MR,RH,NE,NU,NO
LL=3*NL+2
DIM DA(NL,14),LP(NT,LL),P(ND),DI(ND)
FOR I=1 TO ND:READ DI(I):NEXT I
FOR I=1 TO ND:READ P(I):NEXT I
FOR I=1 TO NL:FOR J=1 TO 10:READ DA(I,J):NEXT J
FOR J=12 TO 13:READ DA(I,J):NEXT J:NEXT I
PRINT "THE ORIGINAL DATA FOLLOWS":PRINT
FOR I=1 TO NL:FOR J=1 TO 14
PRINT DA(I,J);
NEXT J:PRINT:NEXT I
S=0
FOR I=1 TO NL:S=S+DA(I,10):NEXT I
FOR I=1 TO NL
IF DA(I,13) <> 2 GOTO 180
S=S+DA(I,7)
GOTO 190
NEXT I
IF ABS(S) > .01 THEN PRINT "WARNING--THE NETWORK INPUTS AND DEMANDS ARE OUT OF
IF ABS(S)>.01 THEN PRINT "BALANCE BY";ABS(S)
REM THE ITERATIONS TO DETERMINE LINK FLOWS BY CUMULATING FLOWS AT DOWNSTREAM
REM NODES FOLLOW
FOR I=1 TO NL
IF DA(I,13) <> 1 GOTO 340
IF DA(I,10) > 0 GOTO 280
REM DA(I,10)<0 IMPLIES A DEMAND NODE
DA(I,11)=DA(I,10)
DA(I,14)=-DA(I,10)
GOTO 340
IF DA(I,13)=2 GOTO 320
DA(I,11)=DA(I,10)
DA(I,14)=-DA(I,10)
GOTO 340
DA(I,11)=DA(I,10)
DA(I,14)=DA(I,10)
NEXT I
REM WE NOW HAVE LINK FLOWS IN ALL TERMINAL LINKS
REM FLOW IS + IF DIRECTED AWAY FROM THE REFERENCE INPUT
FOR I=1 TO NL
IF DA(I,14) <> 0 GOTO 510
K=0
FOR J=1 TO NL
IF DA(I,9) <> DA(J,6) GOTO 460
K=K+1
IF DA(J,14)=0 GOTO 500
DA(I,11)=DA(I,11)-DA(J,14)
REM THE NEGATIVE IN ABOVE STATEMENT IS BECAUSE FLOWS ENTERING NODES ARE NEG.
NEXT J
DA(I,11)=DA(I,11)+DA(I,10)
DA(I,14)=-DA(I,11)
GOTO 510
IF K>0 THEN DA(I,11)=0
NEXT I
FOR I=1 TO NL:IF DA(I,11)=0 GOTO 370
NEXT I
PRINT:PRINT "THE LINK FLOWS HAVE BEEN DETERMINED":PRINT
PRINT "THE LINK FLOWS ARE (+ FLOWS ARE AWAY FROM REF NODE)...":PRINT
FOR I=1 TO NL:PRINT "LINK ";I,"FLOW = ";DA(I,14):PRINT
NEXT I
```

```
580 FOR I=1 TO NT:FOR J=1 TO LL:LP(I,J)=0:NEXT J:NEXT I
590 R=0
600 RC=0
610 RC=RC+1
620 REM RC=1 IMPLIES TERMINAL NODE CONSTRAINTS(BOTH SOURCE AND DEMAND TYPE)
630 REM RC=2 IMPLIES CONSTRAINTS FOR SOURCES AT INTERIOR NODES
640 FOR I=1 TO NL
650 IF RC=1 GOTO 720
660 IF DA(I,13) >0 GOTO 970
670 IF DA(I,10) <= 0 GOTO 970
680 R=R+1
690 IF R>NT GOTO 1010
700 LP(R,1)=1
710 GOTO 760
720 IF DA(I,13) <> 1 GOTO 970
730 R=R+1
740 IF DA(I,10) <0 THEN LP(R,1)=0
750 IF DA(I,10) >0 THEN LP(R,1)=1
760 LP(R,LL)=DA(I,12)
770 C=3*(I-1)+2
780 LP(R,C)=DA(I,1)
790 C=C+1
800 LP(R,C)=DA(I,4)
810 C=C+1
820 LP(R,C)=DA(I,14)
830 K=I
840 GOTO 860
850 K=J
860 FOR J=1 TO NL
870 IF DA(K,13)=2 GOTO 970
880 IF DA(J,9) <> DA(K,6) GOTO 960
890 C=3*(J-1)+2
900 LP(R,C)=DA(J,1)
910 C=C+1
920 LP(R,C)=DA(J,4)
930 C=C+1
940 LP(R,C)=DA(J,14)
950 GOTO 850
960 NEXT J
970 NEXT I
980 IF RC=2 GOTO 1010
990 REM NOW WRITE LP ROWS FOR SOURCES AT INTERIOR NODES
1000 GOTO 610
1010 PRINT
1020 PRINT "MATRIX LP COMING UP...":PRINT
1030 FOR I=1 TO NT:FOR J=1 TO LL:PRINT LP(I,J);:NEXT J:PRINT:NEXT I
1040 PRINT "THE DATA WILL NOW BE WRITTEN TO A SEQUENTIAL DATA FILE NAMED LPDATA
1050 OPEN "O",#1,"B:LPDATA.DAT"
1060 CLOSE #1
1070 KILL "B:LPDATA.DAT"
1080 OPEN "O",#1,"B:LPDATA.DAT"
1090 PRINT #1,T$
1100 PRINT #1,NT,NL,ND,PF,MR,RH,NE,NU,NO
1110 FOR I=1 TO ND
1120 PRINT #1,P(I)
1130 NEXT I
1140 REM TRANSFER THE EXISTING LINK NUMBERS AND THEIR DIA NUMBERS
1150 FOR I=1 TO NL
1160 IF DA(I,2)=0 GOTO 1180
1170 PRINT #1,DA(I,1),DA(I,3)
1180 NEXT I
```

```
1190 FOR I=1 TO ND
1200 PRINT #1,DI(I)
1210 NEXT I
1220 FOR I=1 TO NT
1230 PRINT #1,LP(I,1)
1240 FOR J=1 TO NL
1250 C=0
1260 FOR K=1 TO 3
1270 R=3*(J-1)+K+1
1280 N=LP(I,R)
1290 IF N <> 0 GOTO 1330
1300 C=C+1
1310 IF C=1 THEN PRINT #1,0
1320 GOTO 1340
1330 PRINT #1,N
1340 NEXT K
1350 NEXT J
1360 PRINT #1,LP(I,LL)
1370 NEXT I
1380 FOR I=1 TO NL
1390 PRINT #1,DA(I,5)
1400 NEXT I
1410 CLOSE #1
1420 PRINT "TYPE 'RUN BRANCH'"
2000 DATA PROJECT EXAMPLE NETWORK
2010 DATA 5,3,2
2020 DATA 5.0,1.0,20.0
2030 DATA 1,2,1
2040 DATA 2,50
2050 DATA 100.00,150.00
2060 DATA 1,1,2,100,500,1,0.5,20.0,3,0,1.0,2
2070 DATA 2,0,0,140,1200,3,0,1.0,2,-0.4,3.0,1
2080 DATA 3,0,0,140,1000,3,0,1.0,5,-0.2,2.0,0
2090 DATA 4,0,0,140,1500,5,-0.2,2.0,4,-0.15,5.0,1
2100 DATA 5,0,0,140,1500,5,-0.2,2.0,6,0.25,12.0,1
2110 END
```

BRANCH SOURCE LISTING

```
10 OPEN "I",#1,"B:LPDATA.DAT"
20 INPUT #1,T$,NT,NL,ND,PF,MR,RH,NE,NU,NO
30 IY=NL*ND+1
40 IP=IY-1
50 IZ=IY+NL+NT
60 IW=NT+NL
70 IX=IZ-1
80 DIM DI(ND),D(IW,IZ),P(IX),IB(IW),SC(IX)
90 MP=0
100 FOR I=1 TO ND
110 INPUT #1,P(I)
120 IF P(I) > MP THEN MP=P(I)
130 P(I)=-P(I)
140 NEXT I
150 MP=10*MP
160 FOR I=1 TO NL-1
170 L=ND*I
180 FOR J=1 TO ND
190 L=L+1
200 P(L)=P(J)
210 NEXT J
220 NEXT I
230 FOR I=IY+NT TO IX
240 P(I)=-MP
250 NEXT I
260 IF NE=0 GOTO 390
270 FOR II=1 TO NE
280 INPUT #1,JJ
290 INPUT #1,KK
300 FOR I=1 TO NL
310 IF I<>JJ GOTO 370
320 FOR J=1 TO ND
330 L=(I-1)*ND+J
340 P(L)=-MP
350 IF J=KK THEN P(L)=0
360 NEXT J
370 NEXT I
380 NEXT II
390 FOR I=1 TO ND
400 INPUT #1,DI(I)
410 NEXT I
420 IF NO=1 GOTO 450
430 PRINT "THE DATA ARE BEING READ...":PRINT
440 GOTO 490
450 PRINT "THE DIAMETERS ARE..."
460 FOR I=1 TO ND
470 PRINT DI(I);
480 NEXT I:PRINT
490 FOR I=1 TO IW:FOR J=1 TO IZ:D(I,J)=0:NEXT J:NEXT I
500 REM THE ORDER OF CONSTRAINTS IS ...TYPE 1 HL CONSTRAINTS ARE FIRST IN NODE
510 REM ORDER. THEN ARE TYPE 2 HL CONSTRAINTS IN NODE ORDER. LAST ARE LINK
520 REM LENGTH CONSTRAINTS IN LINK ORDER.
530 REM LOAD THE HEADLOSS CONSTRAINTS' COEFFICIENTS
540 FOR I=1 TO NT
550 INPUT #1,IS
560 REM IS=0 IMPLIES THAT THE CONSTRAINT IS A DEMAND TYPE
570 REM IS=1 IMPLIES THAT THE CONSTRAINT IS A SOURCE TYPE
580 REM INTERIOR SOURCES AND TERMINAL SOURCES (IS=1) ARE TREATED IDENTICALLY
590 IF IS=1 THEN P(IP+I)=-MP
```

```
600 IF IS=0 THEN P(IP+I)=0
610 L=0
620 FOR J=1 TO NL
630 INPUT #1,JY
640 REM JY=0 IMPLIES THAT LINK J IS NOT ON THE CONSTRAINT'S PATH
650 IF JY=0 THEN L=L+ND
660 IF JY=0 GOTO 840
670 INPUT #1,C
680 INPUT #1,Q
690 IF NU=1 THEN CO=10.55
700 IF NU=2 THEN CO=162000!
710 IF NU=3 THEN CO=1730
720 FOR K=1 TO ND
730 L=L+1
740 D(I,L)=CO*(C^(-1.85))*((ABS(Q)*PF)^1.85)*(DI(K)^(-4.87))
750 IF IS=0 THEN GOTO 800
760 REM THIS IS FOR TYPE 2 (SOURCE) HEADLOSS CONSTRAINTS
770 IF Q>0 THEN D(I,L)=-D(I,L)
780 REM WHEN Q>0 THEN HEAD IS GAINED IN THAT LINK ON PATH FROM SOURCE NODE
790 GOTO 830
800 REM THIS IS FOR TYPE 1 (DEMAND) HEADLOSS CONSTRAINTS
810 IF Q<0 THEN D(I,L)=-D(I,L)
820 REM WHEN Q<0 HEAD IS GAINED IN THAT LINK ON PATH FROM REF NODE
830 NEXT K
840 NEXT J
850 INPUT #1,EL
860 IF IS=0 GOTO 940
870 D(I,IZ)=EL-RH
880 IF D(I,IZ)>0 GOTO 950
890 FOR J=1 TO IP
900 D(I,J)=-D(I,J)
910 NEXT J
920 D(I,IZ)=-D(I,IZ)
930 GOTO 950
940 D(I,IZ)=RH-(EL+MR)
950 NEXT I
960 IF NO=0 GOTO 1000
970 PRINT "THE COST COEFFICIENTS FOLLOW..."
980 FOR I=1 TO IX:PRINT P(I);:NEXT I
990 REM LOAD THE LENGTH CONSTRAINTS' COEFFICIENTS
1000 FOR I=NT+1 TO IW
1010 L=ND*(I-(NT+1))
1020 FOR K=1 TO ND
1030 J=L+K
1040 D(I,J)=1
1050 NEXT K
1060 INPUT #1,LE
1070 D(I,IZ)=LE
1080 NEXT I
1090 FOR I=1 TO IW
1100 J=IP+I
1110 D(I,J)=1
1120 NEXT I
1130 CLOSE #1
1140 IF NO=1 GOTO 1170
1150 PRINT "RELAX WHILE I WORK ON THIS PROBLEM..."
1160 GOTO 1210
1170 PRINT:PRINT:PRINT "MATRIX D FOLLOWS...":PRINT
1180 FOR I=1 TO IW
1190 FOR J=1 TO IZ
```

```
1200 PRINT D(I,J);:NEXT J:PRINT:PRINT:NEXT I
1210 FOR N=IY TO IX
1220 FOR L=1 TO IW
1230 IF D(L,N)=1 GOTO 1260
1240 NEXT L
1250 GOTO 1270
1260 IB(L)=N
1270 NEXT N
1280 Z=0
1290 FOR N=1 TO IW
1300 IB=IB(N)
1310 Z=Z+D(N,IZ)*P(IB)
1320 NEXT N
1330 NP=0
1340 SM=0
1350 FOR N=1 TO IX
1360 FOR I=1 TO IW
1370 IF N=IB(I) GOTO 1480
1380 NEXT I
1390 SU=0
1400 FOR I=1 TO IW
1410 J=IB(I)
1420 SU=SU+P(J)*D(I,N)
1430 NEXT I
1440 SC(N)=P(N)-SU
1450 IF SC(N) <= SM GOTO 1480
1460 SM=SC(N)
1470 PC=N
1480 NEXT N
1490 FOR M=1 TO IW
1500 IB=IB(M)
1510 SC(IB)=0
1520 NEXT M
1530 IF SM <=0 GOTO 1990:REM WE HAVE OPTIMALITY
1540 NP=NP+1
1550 SL=1E+30
1560 FOR M=1 TO IW
1570 IF D(M,PC) > 0 GOTO 1590
1580 GOTO 1640
1590 Q=D(M,IZ)/D(M,PC)
1600 IF (Q-SL) < 0 GOTO 1620
1610 GOTO 1640
1620 PR=M
1630 SL=Q
1640 NEXT M
1650 IB(PR)=PC
1660 DV=D(PR,PC)
1670 FOR N=1 TO IZ
1680 CR=D(PR,N)
1690 D(PR,N)=CR/DV
1700 NEXT N
1710 IF NO=1 GOTO 1740
1720 PRINT:PRINT "IF YOU THINK THIS IS SLOW, TRY IT BY HAND!"
1730 GOTO 1780
1740 PRINT "SIMPLEX CRITERIA"
1750 FOR N=1 TO IX
1760 PRINT SC(N);
1770 NEXT N
1780 N=NP+1
1790 IF NO=0 GOTO 1820
```

```
1800 PRINT:PRINT "THE OBJECTIVE FUNCTION VALUE IS ";-Z:PRINT
1810 PRINT "TABLEAU # ";N
1820 FOR M=1 TO IW
1830 IF (M-PR)=0 GOTO 1900
1840 CM=-D(M,PC)
1850 FOR N=1 TO IZ
1860 IM=D(PR,N)*CM
1870 SK=D(M,N)
1880 D(M,N)=SK+IM
1890 NEXT N
1900 IF NO=0 GOTO 1960
1910 PRINT "X(";IB(M);")";:PRINT
1920 FOR N=1 TO IZ
1930 PRINT D(M,N);
1940 NEXT N
1950 PRINT:PRINT
1960 NEXT M
1970 Z=Z+SL*SM
1980 GOTO 1340
1990 PRINT:PRINT "ENOUGH OF THIS ITERATING...":PRINT:PRINT
2000 PRINT "THE LEAST COST DESIGN FOLLOWS...":PRINT
2010 PRINT "THE COST OF THE DESIGN IS ";-Z
2020 PRINT
2030 L=0
2040 FOR I=1 TO NL
2050 FOR J=1 TO ND
2060 L=L+1
2070 FOR K=1 TO IW
2080 IF IB(K) <> L GOTO 2100
2090 PRINT "THE LENGTH OF DIAMETER  ";J;"IN LINK ";I;" IS ";D(K,IZ)
2100 NEXT K
2110 NEXT J
2120 NEXT I
2130 PRINT
2140 L=0
2150 FOR I=IY TO IY+NT-1
2160 L=L+1
2170 FOR K=1 TO IW
2180 IF IB(K) <> I GOTO 2200
2190 PRINT "THE SLACK IN TERMINAL NODE CONSTRAINT ";L;"IS ";D(K,IZ)
2200 NEXT K
2210 NEXT I
2220 FOR I=IY+NT TO IX
2230 FOR K=1 TO IW
2240 IF IB(K) <> I GOTO 2260
2250 PRINT "THIS PROBLEM IS INFEASIBLE--THERE IS EXCESSIVE HEADLOSS"
2260 NEXT K
2270 NEXT I
2280 FOR I=IY TO IP+NT
2290 FOR K=1 TO IW
2300 IF IB(K) <> I GOTO 2340
2310 IF P(I)=0 GOTO 2340
2320 PRINT "THIS PROBLEM IS INFEASIBLE"
2330 PRINT "THERE IS AN UNSATISFIED MULTIPLE SOURCE CONSTRAINT"
2340 NEXT K
2350 NEXT I
2360 END
```

APPENDIX H

USER INSTRUCTIONS FOR THE CLOSED-CIRCUIT WATER DISTRIBUTION SYSTEM SIMULATOR PROGRAM "LOOP" IN THE BASIC LANGUAGE FOR MICROCOMPUTERS*

1982

* Developed by Keith Little, Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC 27514

General

"LOOP" is a program written in the BASIC language that simulates flows and pressures in a looped (closed circuit) water distribution system. Some dialect of the BASIC language is supported by virtually all microcomputers. LOOP accomplishes three algorithmic tasks. The first is, from user-specified nodal inputs and demands and system geometry, to determine an initial flow-balanced network. From this flow-balanced network, the second algorithm uses the hardy-Cross technique to systematically change the link flows in such a way that the headlosses around each loop cancel to within a user-specified tolerance. The third algorithm calculates final link headlosses and nodal pressures based on the flow distribution determined from Hardy-Cross, LOOP's source listing is in the Appendix.

The Hardy-Cross method is well suited for microcomputers. The network is described mathematically by a system of simultaneous, nonlinear equations. The network simulation is the solution to this system. Network simulation algorithms designed for mainframe computers operate on all equations in this system at the same time using numerical techniques such as Newton's method. This strategy requires a considerable internal memory even for moderately-sized networks. Hardy-Cross is essentially Newton's method applied to a single equation (of the system) at a time, thereby greatly reducing internal memory requirements. The cost is that convergence to the solution is slower and, in some cases, may not occur at all.

Limitations

LOOP is designed to simulate networks consisting of a single input, multiple demands, and a geometry in which the number of links (pipes between input/demand points or nodes) in each loop is either 2, 3, or 4. LOOP is, in fact, written for links per loop (the standard urban layout); however, "pseudo-nodes" can be introduced in loops consisting of 2 or 3 links to effect a four-link loop configuration. A pseudo-node is an artificial node with no input or demand.

The network is assumed to have a single input and a single known hydraulic grade line (HGL) elevation either at the input node (the common situation) or some other node. LOOP will find a "solution" for multiple inputs and a single specified HGL elevation, but there is no guarantee that such a system is physically realistic, i.e., rarely can one specify an input at a node without also knowing the pressure at that node.

LOOP does not accomodate in-line hydraulic elements such as booster pumps, pressure reducing valves, etc.

The units for the network data and solutions are english. Lengths, headlosses, HGL and ground elevations are in feet. Pipe diameters are in inches. Input(s) and demands are in cubic feet per second. Pressures are in pounds per square inch.

The data for the modified network are entered as follows:

Data statement 4000 contains any descriptive name (alpha-numeric) that begins with a letter.

```
4000 DATA EXAMPLE NETWORK
```

Data statement 4010 contains the stopping criterion for the Hardy-Cross headloss-balancing algorithm. A large value of the criterion results in rapid convergence at the cost of lesser accuracy and vice versa.

```
4010 DATA 1E-04
```

Data statement 4020 contains, respectively, the numbers of loops, links, and nodes (including pseudo-nodes).

```
4020 DATA 2, 7, 6
```

Data statement 4030 contains the link lengths in sequential order of link numbers, i.e., 1, 2, ...

```
4030 DATA 1200, 1000, 1200, 1000, 1000, 500, 500
```

Data statement 4040 contains the link diameters also in sequential order of link numbers.

```
4040 DATA 12, 12, 15, 8, 12, 20, 20
```

Data statement 4050 contains the link's C value (Hazen-Williams coefficient) in sequential order of link numbers.

```
4050 DATA 100, 100, 100, 100, 100, 100, 100
```

Data statement 4060 contains the ground elevations of the nodes in sequential order of node numbers. Note that ground elevations for any pseudo-nodes must be known.

```
4060 DATA 2, 1, 3, 2, 5, 3
```

Data statement 4070 contains, respectively, the node number at which the hydraulic grade line elevation is known and the elevation. This will typically be at the input node.

```
4070 DATA 3, 50
```

Data statement 4080 contains the input and demands of the nodes in sequential order of node numbers. The demands are input as negative quantities. Pseudo-nodes have 0 demand.

```
4080 DATA -1.5, -2.5, 10, -2, -4, 0
```

The remaining data statements specify the network geometry. There are four statements (one for each link) for every loop and three entries per statement. The (sets of) loop statements are entered in sequential order of loop numbers. The loops are not explicitly numbered in the statements but it is necessary (for the user to understand the solution) that this ordering scheme be followed. In other words, the first four statement will be designated by the program as for loop #1, the second four for loop #2, and so on. Each of the four link data statements for a given loop contains, respectively, the link number, the "counter clockwise" node's number, and the clockwise" node's number. The counter clockwise node for a given link in a given loop is the node (attached to that link) that is first encountered when traveling in a clockwise direction around that loop. The clockwise node is the last encountered. For example, the data statement describing link 4 relative to loop 1 is: Data 4, 4, 3. The statement for link 4 relative to loop 2 is: Data 4, 3, 4. Within a loop, the link statements can be arbitrarily ordered. For loop 1, we have:

```
4090 DATA 1, 3, 1
4100 DATA 2, 1, 2
4110 DATA 3, 2, 4
4120 DATA 4, 4, 3
```

Similarly, for Loop 2:

```
4130 DATA 6, 6, 5
4140 DATA 4, 3, 4
4150 DATA 7, 4, 6
4160 DATA 5, 5, 3
```

After the last data statement, there is an end statement,

```
4170 END
```

The example network was run on an Osborne 1 Microcomputer using the Z-80A microprocessor. The solution is given below.

```
THE LINK FLOWS AND HEADLOSSES FOLLOW...
FOR LINK #  1  THE FLOW IS  3.68067 AND THE HEADLOSS IS  12.5434
FOR LINK #  2  THE FLOW IS  2.18067 AND THE HEADLOSS IS  3.96883
FOR LINK #  3  THE FLOW IS  .31933 AND THE HEADLOSS IS  .0459563
FOR LINK #  4  THE FLOW IS  1.61993 AND THE HEADLOSS IS  16.497
FOR LINK #  5  THE FLOW IS  4.69939 AND THE HEADLOSS IS  16.4266
FOR LINK #  6  THE FLOW IS  .699396 AND THE HEADLOSS IS  .0201175
FOR LINK #  7  THE FLOW IS  .699396 AND THE HEADLOSS IS  .0201175
THE FLOW DIRECTIONS RELATIVE TO THE LOOPS ARE...

THE FLOW DIRECTIONS RELATIVE TO LOOP 1 ARE...
THE FLOW IN LINK # 1 IS CLOCKWISE
THE FLOW IN LINK # 2 IS CLOCKWISE
THE FLOW IN LINK # 3 IS COUNTER-CLOCKWISE
THE FLOW IN LINK # 4 IS COUNTER-CLOCKWISE

THE FLOW DIRECTIONS RELATIVE TO LOOP 2 ARE...
THE FLOW IN LINK # 6 IS COUNTER-CLOCKWISE
THE FLOW IN LINK # 4 IS CLOCKWISE
THE FLOW IN LINK # 7 IS COUNTER-CLOCKWISE
THE FLOW IN LINK # 5 IS COUNTER-CLOCKWISE
THE NODE EL'S,HGL'S, AND PRESSURES ARE...
FOR NODE # 1 THE GROUND EL IS 2 ,THE HGL IS 37.4566 AND THE PRESSURE IS 15.3448
FOR NODE # 2 THE GROUND EL IS 1 ,THE HGL IS 33.4877 AND THE PRESSURE IS 14.06
FOR NODE # 3 THE GROUND EL IS 3 ,THE HGL IS 50 AND THE PRESSURE IS 20.3406
FOR NODE # 4 THE GROUND EL IS 2 ,THE HGL IS 33.5337 AND THE PRESSURE IS 13.6471
FOR NODE # 5 THE GROUND EL IS 5 ,THE HGL IS 33.5739 AND THE PRESSURE IS 12.3662
FOR NODE # 6 THE GROUND EL IS 3 ,THE HGL IS 33.5538 AND THE PRESSURE IS 13.223
```

APPENDIX

```
***SOURCE LISTING FOR LOOP***

0 READ T$
0 READ SC
0 REM T$ IS THE PROJECT TITLE
0 READ N,L,NN
  REM N IS THE # OF LOOPS AND L IS THE # OF LINKS
0 REM NN IS THE # OF NODES
0 M=4*N
0 DIM L1(L),S(L),Q(L),QC(L),K(L),C(L),HL(L),LC(N),SQ(2)
0 DIM P(NN,4),D(M,4),I2(4),F(4),B(4),S1(4)
00 FOR I=1 TO L
0 READ L1(I)
20 NEXT I
30 FOR I=1 TO L
40 READ S(I)
50 NEXT I
60 FOR I=1 TO L
0 READ C(I)
80 NEXT I
90 REM P(I,1) ARE THE NODAL GROUND ELEVATIONS
00 FOR I=1 TO NN
10 READ P(I,1)
20 NEXT I
0 REM P(I,2) WILL BE THE NODAL HGL'S--ONLY ONE WILL BE KNOWN INITIALLY
40 FOR I=1 TO NN:P(I,2)=0:NEXT I
50 READ I
60 READ P(I,2)
70 REM P(I,2) IS THE KNOWN NODAL HGL FROM WHICH ALL OTHER HGL'S ARE DETERMINED
80 REM P(I,4) IS THE VECTOR OF NODAL INPUT/DEMANDS
0 C=0
00 FOR I=1 TO NN
10 READ P(I,4)
20 C=C+P(I,4)
30 NEXT I
40 FOR I=1 TO NN
0 NEXT I
60 IF ABS(C)<.01 GOTO 390
70 PRINT "THE NODAL INPUTS AND DEMANDS ARE NOT IN AGREEMENT"
80 PRINT "CHECK YOUR DATA AND TRY AGAIN!":END
90 FOR I=1 TO M
00 READ D(I,1):READ D(I,3):READ D(I,4)
0 NEXT I
20 PRINT "THE FLOW-BALANCING ALGORITHM WILL BEGIN..."
30 FOR I=1 TO L
40 Q(I)=0
50 QC(I)=0
60 NEXT I
  REM FIND ALL "CORNER" NODES (THOSE WITH 2 ATTACHED,UNASSIGNED  LINKS)
80 REM ASSIGN Q TO THOSE 2 LINKS TO MEET DEMAND AT NODE I
90 PRINT "WE'LL NOW CONSIDER NODES WITH EXACTLY 2 Q-UNSPECIFIED LINKS"
00 FOR I=1 TO NN
10 PRINT "THE NODE # IS";I
20 FOR K=1 TO 4:S1(K)=0:NEXT K
0 REM VECTOR S1 WILL CONTAIN THE #'S OF THE LINKS ATTACHED TO NODE I
40 S=0
50 FOR J=1 TO M
60 IF D(J,3)=I GOTO 590
70 IF D(J,4)=I GOTO 590
80 GOTO 650
```

```
90 PRINT "LINK #";D(J,1);"BEING CONSIDERED"
00 FOR K=1 TO 4
10 IF D(J,1)=S1(K) GOTO 650
20 NEXT K
30 S=S+1
40 S1(S)=D(J,1)
50 NEXT J
60 PRINT "THE LINKS ATTACHED TO NODE";I;"ARE..."
70 FOR K=1 TO 4:PRINT S1(K):NEXT K
80 REM NOW COUNT THE UNASSIGNED LINKS IN VECTOR S1
90 C=0
00 FOR K=1 TO 4
10 IF S1(K)=0 GOTO 730
20 IF QC(S1(K))=0 THEN C=C+1
30 NEXT K
40 IF C<>2 GOTO 1220
50 PRINT "NODE";I;"HAS EXACTLY 2 Q-UNASSIGNED LINKS"
60 REM NOW ASSIGN DIRECTION INDICES TO THE 2 LINKS AND
70 REM UPDATE THE NODAL INPUT/OUTPUT VECTOR,P(*,4)
80 FOR K=1 TO 4
90 IF S1(K)=0 GOTO 1090
00 IF QC(S1(K))<>0 GOTO 1090
10 FOR J=1 TO M
20 IF D(J,1)<>S1(K) GOTO 1080
30 IF D(J,4)<>I GOTO 960
40 CL=D(J,1):REM CL IS A CLOCKWISE LINK FROM NODE I
50 IF P(I,4)<0 THEN D(J,2)=1
60 IF P(I,4)>0 THEN D(J,2)=-1
70 REM CHECK IF LINK CL EXISTS IN ANOTHER LOOP AS A CC LINK
80 REM IF SO, UPDATE ITS DIRECTION INDEX ALSO
90 FOR JJ=1 TO M
00 IF D(JJ,1)<>CL GOTO 930
10 IF JJ=J GOTO 930
20 D(JJ,2)=-D(J,2)
30 NEXT JJ
40 P(D(J,3),4)=P(D(J,3),4)+P(I,4)/2
50 GOTO 1090
60 CC=D(J,1):REM CC IS A COUNTER-CLOCKWISE LINK FROM NODE I
70 IF P(I,4)<0 THEN D(J,2)=-1
80 IF P(I,4)>0 THEN D(J,2)=1
90 REM CHECK IF LINK CC EXISTS IN ANOTHER LOOP AS A CL LINK
000 REM IF SO, UPDATE ITS DIRECTION INDEX ALSO
010 FOR JJ=1 TO M
020 IF D(JJ,1)<>CC GOTO 1050
030 IF JJ=J GOTO 1050
040 D(JJ,2)=-D(J,2)
050 NEXT JJ
060 P(D(J,4),4)=P(D(J,4),4)+P(I,4)/2
070 GOTO 1090
080 NEXT J
090 NEXT K
100 REM NOW ASSIGN Q'S TO THE 2 LINKS
110 FOR K=1 TO 4
120 IF QC(S1(K))=0 THEN Q(S1(K))=ABS(P(I,4)/2)
130 IF QC(S1(K))=0 THEN QC(S1(K))=1
140 NEXT K
150 REM NOW UPDATE THE INPUT/OUTPUT VECTOR FOR NODE I
160 P(I,4)=0
170 PRINT "THE FLOW VECTOR IS..."
180 FOR J=1 TO L:PRINT J,Q(J):NEXT J
190 PRINT "THE NODAL I/O VECTOR IS..."
200 FOR J=1 TO NN:PRINT J,P(J,4):NEXT J
```

```
1210 GOTO 1230
1220 NEXT I
1230 FOR I=1 TO L
1240 IF QC(I)=0 GOTO 1270
1250 NEXT I
1260 GOTO 2100
1270 PRINT "WE'LL NOW LOOK AT NODES WITH 1 Q-UNSPECIFIED LINK"
1280 REM NOW FIND THE NODES WITH A SINGLE UNSPECIFIED LINK
1290 REM THEN SATISFY THEIR DEMAND
1300 FOR I=1 TO NN
1310 PRINT "THE NODE # IS ";I
1320 FOR K=1 TC 4:S1(K)=0:NEXT K
1330 REM VECTOR S1 WILL CONTAIN THE #'S OF LINKS ATTACHED TO NODE I
1340 S=0
1350 FOR J=1 TO M
1360 IF D(J,3)=I GOTO 1390
1370 IF D(J,4)=I GOTO 1390
1380 GOTO 1450
1390 PRINT "LINK #";D(J,1);"BEING CONSIDERED"
1400 FOR K=1 TO 4
1410 IF D(J,1)=S1(K) GOTO 1450
1420 NEXT K
1430 S=S+1
1440 S1(S)=D(J,1)
1450 NEXT J
1460 PRINT "THE LINKS ATTACHED TO NODE I ARE..."
1470 FOR K=1 TO 4:PRINT S1(K):NEXT K
1480 REM NOW COUNT THE Q-UNASSIGNED LINKS IN VECTOR S1
1490 C=0
1500 FOR K=1 TO 4
1510 IF S1(K)=0 GOTO 1530
1520 IF QC(S1(K))=0 THEN C=C+1
1530 NEXT K
1540 IF C<>1 GOTO 1900
1550 PRINT "NODE";I;"HAS EXACTLY 1 Q-UNSPECIFIED LINK"
1560 REM NOW ASSIGN THE DIRECTION INDEX TO THE LINK AND
1570 REM UPDATE THE NODAL I/O VECTOR
1580 FOR K=1 TO 4
1590 IF S1(K)=0 GOTO 1620
1600 IF QC(S1(K))<>0 GOTO 1620
1610 GOTO 1630:REM S1(K) WILL BE THE UNSPECIFIED LINK #
1620 NEXT K
1630 CC=0:REM CC WILL COUNT THE # OF TIMES THAT THE I/O FOR
1640 REM FOR THE NODE ATTACHED TO LINK S1(K) IS UPDATED
1650 FOR J=1 TO M
1660 IF D(J,1)<>S1(K) GOTO 1790
1670 IF D(J,4)<>I GOTO 1740
1680 IF P(I,4)<0 THEN D(J,2)=1
1690 IF P(I,4)>0 THEN D(J,2)=-1
1700 IF CC>=1 GOTO 1790
1710 P(D(J,3),4)=P(D(J,3),4)+P(I,4)
1720 CC=CC+1
1730 GOTO 1790
1740 IF P(I,4)<0 THEN D(J,2)=-1
1750 IF P(I,4)>0 THEN D(J,2)=1
1760 IF CC>=1 GOTO 1790
1770 P(D(J,4),4)=P(D(J,4),4)+P(I,4)
1780 CC=CC+1
1790 NEXT J
1800 REM NOW ASSIGN Q TO LINK S1(K)
```

```
1810 Q(S1(K))=ABS(P(I,4))
1820 QC(S1(K))=1
1830 REM NOW UPDATE I/O VECTOR FOR NODE I
1840 P(I,4)=0
1850 PRINT "THE FLOW VECTOR IS..."
1860 FOR J=1 TO L:PRINT J,Q(J):NEXT J
1870 PRINT "THE NODAL I/O VECTOR IS..."
1880 FOR J=1 TO NN:PRINT J,P(J,4):NEXT J
1890 GOTO 1910
1900 NEXT I
1910 REM NOW CHECK IF ALL LINKS HAVE ASSIGNED Q'S--IF NOT BO BACK THRU NETWORK
1920 FOR I=1 TO L
1930 IF QC(I)=0 GOTO 490
1940 NEXT I
1950 REM CHECK IF ALL ASSIGNED Q'S ARE NON-ZERO--THE HARDY CROSS
1960 REM ALGORITHM CAN ONLY HANDLE NON-ZERO LINK Q'S
1970 FOR I=1 TO L
1980 IF Q(I)<>0 THEN GOTO 2040
1990 PRINT "LINK #";I;"HAS BEEN ASSIGNED A ZERO FLOW. THE HARDY-"
2000 PRINT "CROSS ALGORITHM CANNOT HANDLE THIS LINK FLOW. I SUGGEST"
2010 PRINT "THAT YOU MAKE A VERY SMALL CHANGE TO THE NETWORK INPUT AND"
2020 PRINT "A COMPENSATING CHANGE TO ONE OF THE NODES CONNECTED TO LINK #";I;
2030 PRINT "AND TRY AGAIN":END
2040 NEXT I
2050 PRINT "THE D MATRIX AFTER THE FLOW BALANCING ALGORITHM IS..."
2060 FOR I=1 TO M
2070 FOR J=1 TO 4
2080 PRINT D(I,J);
2090 NEXT J:PRINT:NEXT I
2100 PRINT:PRINT
2110 PRINT "THE NETWORK NOW HAS A STARTING,BALANCED FLOW DISTRIBUTION"
2120 PRINT:PRINT "THE HARDY-CROSS ALGORITHM WILL NOW BEGIN..."
2130 REM COMPUTE K VALUES FOR HAZEN-WILLIAMS FORMULA
2140 FOR I=1 TO L
2150 K(I)=(4720/(C(I)*(S(I)/12)^2.63))^1.85
2160 NEXT I
2170 IT=0
2180 IT=IT+1
2190 T=0
2200 W=1
2210 FOR I=1 TO 4
2220 I2(I)=D(W,1)
2230 S1(I)=D(W,2)
2240 W=W+1
2250 NEXT I
2260 T9=0
2270 B1=0
2280 REM USING THE Q-BALANCED NETWORK APPLY NEWTON'S INTERACTION EQUATION
2290 FOR I=1 TO 4
2300 J=I2(I)
2310 F(I)=K(J)*Q(J)*S1(I)*ABS(Q(J))^.85*L1(J)
2320 B(I)=1.85*K(J)*ABS(Q(J))^.85*L1(J)
2330 T9=T9+F(I)
2340 B1=B1+B(I)
2350 NEXT I
2360 C1=T9/B1
2370 T=T+ABS(C1)
2380 FOR I=1 TO 4
2390 J=I2(I)
2400 REM CORRECTION TO Q APPLIED
```

```
2410 Q(J)=(Q(J)*S1(I)-C1)*S1(I)
2420 NEXT I
2430 REM CHECK IF ALL LOOPS HAVE BEEN CORRECTED
2440 IF W<M GOTO 2210
2450 REM CHECK IF CONVERGENCE IS WITHIN LIMIT SPECIFICATIONS
2460 IF T<SC GOTO 2480
2470 PRINT "HARDY-CROSS ITERATION # ";IT;" COMPLETED":GOTO 2180
2480 FOR I=1 TO L
2490 HL(I)=847000!*((ABS(Q(I))/C(I))^1.85)*(S(I)^(-4.87))*L1(I)
2500 NEXT I
2510 REM NOW UPDATE D(I,2) FOR REVERSED FLOWS
2520 W=0
2530 FOR I=1 TO 4
2540 W=W+1
2550 J=D(W,1)
2560 IF Q(J)<0 THEN D(W,2)=-D(W,2)
2570 NEXT I
2580 IF W<M GOTO 2530
2590 PRINT:PRINT "THE FLOWS AND HEADLOSSES IN LINKS HAVE BEEN DETERMINED.
2595 PRINT "TYPE 'RETURN' FOR THE SIMULATION RESULTS"
2600 PRINT "FOR ";T$
2610 INPUT F$
2620 PRINT "THE LINK FLOWS AND HEADLOSSES FOLLOW...":PRINT
2630 FOR I=1 TO L
2640 PRINT "FOR LINK # ";I;" THE FLOW IS ";ABS(Q(I));"AND THE HEADLOSS IS ";HL(I)
2650 NEXT I
2660 INPUT "TYPE 'RETURN' FOR THE FLOW DIRECTIONS RELATIVE TO THE LOOPS";F$
2670 PRINT "THE FLOW DIRECTIONS RELATIVE TO THE LOOPS ARE..."
2680 W=0
2690 FOR J=1 TO N
2700 PRINT:PRINT "THE FLOW DIRECTIONS RELATIVE TO LOOP";J;"ARE..."
2710 FOR I=1 TO 4
2720 W=W+1
2730 IF D(W,2)=1 THEN PRINT "THE FLOW IN LINK #";D(W,1);"IS CLOCKWISE"
2740 IF D(W,2)=-1 THEN PRINT "THE FLOW IN LINK #";D(W,1);"IS COUNTER-CLOCKWISE"
2750 NEXT I
2760 NEXT J
2770 REM NOW COMPUTE HGL'S AT ALL NODES
2780  FOR I=1 TO N
2790 FOR J=1 TO 4
2800 C=4*(I-1)+J:REM C IS THE ROW # OF MATRIX D
2810 K=D(C,3):REM K IS THE NODE # BEING CONSIDERED
2820 IF P(K,2)=0 GOTO 2850
2830 IF P(D(C,4),2)>0 GOTO 2850
2840 P(D(C,4),2)=P(K,2)-HL(D(C,1))*D(C,2)
2850 NEXT J
2860 NEXT I
2870 REM NOW CHECK IF ALL NODES HAVE BEEN ASSIGNED HGL
2880 FOR I=1 TO NN
2890 IF P(I,2)=0 THEN GOTO 2780
2900 NEXT I
2910 REM NOW COMPUTE THE PRESSURE AT EACH NODE
2920 F2=62.32/144
2930 FOR I=1 TO NN
2940 P(I,3)=F2*(P(I,2)-P(I,1))
2950 NEXT I
2960 PRINT:PRINT:INPUT "TYPE 'RETURN' FOR NODE ELEVATIONS,HGL'S,AND PRESSURES";F$
2970 PRINT "THE NODE EL'S,HGL'S, AND PRESSURES ARE...":PRINT
2980 FOR I=1 TO NN
2990 PRINT "FOR NODE #";I;"THE GROUND EL IS";P(I,1);",THE HGL IS";P(I,2);
2995 PRINT "AND THE PRESSURE IS ";P(I,3)
3000 NEXT I
```

```
4000 DATA EXAMPLE NETWORK
4010 DATA 1E-04
4020 DATA 2,7,6
4030 DATA 1200,1000,1200,1000,1000,500,500
4040 DATA 12,12,15,8,12,20,20
4050 DATA 100,100,100,100,100,100,100
4060 DATA 2,1,3,2,5,3
4070 DATA 3,50
4080 DATA -1.5,-2.5,10,-2,-4,0
4090 DATA 1,3,1
4100 DATA 2,1,2
4110 DATA 3,2,4
4120 DATA 4,4,3
4130 DATA 6,6,5
4140 DATA 4,3,4
4150 DATA 7,4,6
4160 DATA 5,5,3
4170 END
```

WET-WELL VOLUME FOR FIXED-SPEED PUMPS

Albert B. Pincince

A criterion sometimes used for determining wet-well capacity is prevention of too frequent starting and stopping of pumps. A considerable amount of heat is generated during the starting of a pump motor. This heat should be dissipated and the motor allowed to cool before it is restarted. For small motors that generate little heat, there seems to be almost no limit as to how many cycles are permissible, but operating cycles of perhaps 30 min or more are desirable for large pumps.

The objective of this paper is to present equations to determine the wet-well volume required to maintain the cycle greater than a given time. This volume can be obtained by differentiating the equation for cycle time with respect to inflow and setting the derivative equal to zero.

The cycle time, T, equals the off-time plus on-time:

$$T = t_{off} + t_{on} \quad \text{.......... 1}$$

$$t_{off} = V/Q_{in} \quad \text{.......... 2}$$

where

V = volume of suction well between pump start and stop, and

Q_{in} = rate of inflow to station.

The on-time is:

$$t_{on} = V/(Q_{out} - Q_{in}) \quad \text{....... 3}$$

in which Q_{out} = rate of pump discharge. Equations 2 and 3 can be substituted into Equation 1 to yield:

$$T = V\left(\frac{1}{Q_{in}} + \frac{1}{Q_{out} - Q_{in}}\right) \quad \text{... 4}$$

Captain Albert B. Pincince, MSC, is an instructor in the Department of Preventive Medicine, U. S. Army Medical Field Service School, Brooke Army Medical Center, Fort Sam Houston, Texas.

Setting the derivative dT/dQ_{in} of Equation 4 to zero yields:

$$Q_{in} = \frac{Q_{out}}{2} \quad \text{.......... 5}$$

as the flow at which the minimum pump cycle occurs. The minimum cycle time is obtained by substituting Equation 5 into Equation 4:

$$T_{min} = \frac{2V}{Q_{in}} = \frac{4V}{Q_{out}} \quad \text{....... 6}$$

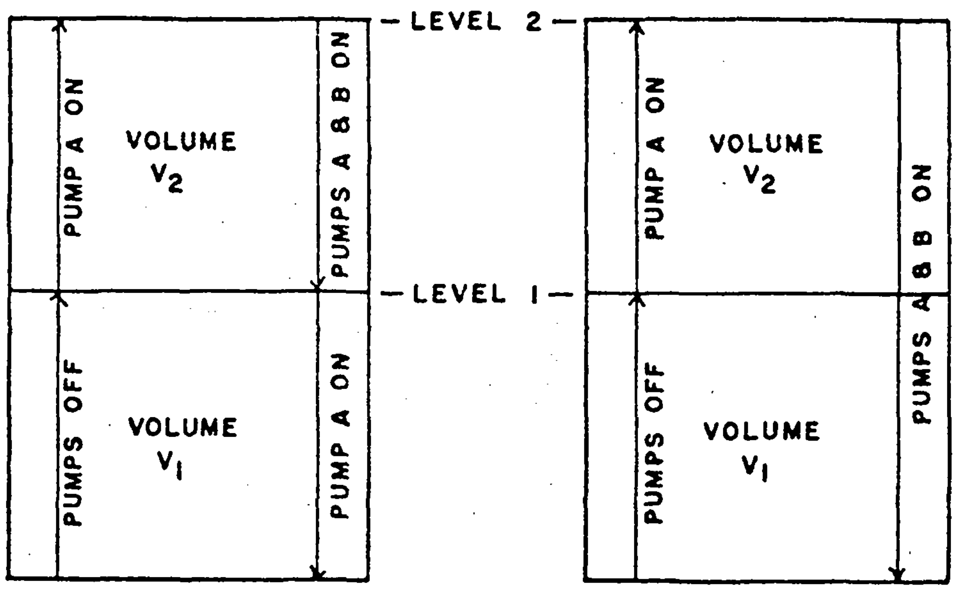
where T_{min} = minimum cycle time. Equation 6 can be rearranged (1) (2) to obtain the volume required for a single constant-speed pump to maintain a given minimum cycle time:

$$V = \frac{T_{min} Q_{out}}{4} \quad \text{.......... 7}$$

Multiple Pumps

Equation 7 also applies for the first pump of a station with several pumps. A similar analysis can be used to determine the drawdown volume for subsequent pump combinations. In the latter case, let the first pump have a capacity of A. Let the second pump have a capacity of B. Whenever flow is less than A, only the first pump is used. When flow is greater than A, say $A + b$, pump capacity equal to $A + B$ is provided.

Two schemes for the operating schedules of these pumps can be used. In the first scheme (Figure 1a) with flow greater than A, the first pump goes on when the liquid level reaches "1." Additional capacity B is provided after the liquid level reaches "2," and capacity $A + B$ is provided until the liquid level is down to "1." At this point



a) SCHEME I b) SCHEME II

FIGURE 1.—*This schematic shows two operating schemes for pumping from the wet well.*

(with flow greater than A), the liquid level begins to rise again and discharge capacity is A.

In the second scheme (Figure 1b), the first pump goes on when the liquid level reaches "1." After the liquid level reaches "2," capacity $A + B$ is provided until the entire wet well is emptied. The wet well again begins to fill, but there is no discharge until liquid level reaches "1," when the first pump goes on.

The same results would be obtained for both cases, of course, if additional capacity were supplied by stopping the first pump and starting a pump with capacity $A + B$ instead of keeping the first pump in operation and adding a second pump.

Wet-Well Volume for Scheme I

In Scheme I, when inflow is greater than A, the second pump is off while the water level rises from "1" to "2." The water level does not go below "1" because only capacity A is provided when the liquid reaches "1." The off-time, then, is:

$$t_{off} = \frac{V_2}{(A + b) - A} = \frac{V_2}{b} \quad \ldots.8$$

and on-time is the time required to drain the volume between levels "1" and "2":

$$t_{on} = \frac{V_2}{(A + B) - (A + b)} = \frac{V_2}{B - b} \quad ..9$$

The cycle is:

$$T_{2nd\ pump} = \frac{V_2}{b} + \frac{V_2}{B - b} \quad \ldots.10$$

Setting $dT_{2nd\ pump}/db$ to zero one obtains:

$$b = B/2 \quad \ldots\ldots\ldots.11$$

at the minimum pump cycle and:

$$V_2 = \frac{T_{min2}B}{4} \quad \ldots\ldots\ldots.12$$

as the drawdown volume necessary to keep the cycle time of the second pump greater than T_{min2}.

Wet-Well Volume for Scheme II

In the Scheme II when flow is greater than A, the pumps providing additional capacity B are on to empty both volumes "1" and "2," and the on time is:

$$t_{on} = \frac{V_1 + V_2}{(A + B) - (A + b)}$$

$$= \frac{V_1 + V_2}{B - b} \quad ..13$$

The first pump starts when the wet-well level reaches level "1," but the additional pumps are off while the wet-well level reaches level "2." The off-time is:

$$t_{off} = \frac{V_1}{A + b} + \frac{V_2}{(A + b) - A}$$

$$= \frac{V_1}{A + b} + \frac{V_2}{b} \quad ..14$$

The cycle time is Equations 13 plus 14 and is:

$$T_{2nd\ pump} = \frac{V_1 + V_2}{B - b} + \frac{V_1}{A + b} + \frac{V_2}{b} \quad ..15$$

To find the minimum cycle time, the derivative of Equation 15 is obtained and set to zero:

$$\frac{dT_{2nd\ pump}}{db} = \frac{V_1 + V_2}{(B - b)^2} - \frac{V_1}{(A + b)^2} - \frac{V_2}{b^2} = 0 \quad ..16$$

It is convenient to relate influent flow and pump discharge to the capacity of the first pump and to express the drawdown volume, V_2, as a multiple of V_1 by writing these terms as dimensionless ratios. Let $\beta = b/A$ be a dimensionless flow and $\gamma = B/A$ be a dimensionless pump discharge. Further, let $V' = V_2/V_1$ be a dimensionless drawdown volume. Then Equation 16 becomes:

$$\frac{1 + V'}{(\gamma - \beta)^2} - \frac{1}{(1 + \beta)^2} - \frac{V'}{\beta^2} = 0 \quad ..17$$

Equation 15 also can be expressed in the same dimensionless terms as:

$$\frac{1 + V'}{\gamma + \beta} + \frac{1}{1 + \beta} + \frac{V'}{\beta} = \frac{T_{2nd\ pump} A}{V_1} = \tau \quad ..18$$

in which τ is a dimensionless time. From Equation 6:

$$\frac{A}{V_1} = \frac{4}{T_{min\ 1st\ pump}} \quad ......19$$

After substitution, the dimensionless time can be expressed as:

$$\mathrm{T} = 4\left(\frac{T_{2nd\ pump}}{T_{min\ 1st\ pump}}\right) \quad ....20$$

It is reasonable to let $T_{min\ 2nd\ pump}$ equal $T_{min\ 1st\ pump}$, although they need not be equal. With this condition, T equals 4 at the minimum cycle, and Equation 18 becomes:

$$\frac{1 + V'}{\gamma - \beta} + \frac{1}{1 + \beta} + \frac{V'}{\beta} = 4 \quad ...21$$

A solution for V' as a function of γ is sought. Equations 17 and 21 can be solved for V' and then equated, and the resulting equation solved for $\gamma - \beta$:

$$\gamma - \beta = \frac{4\beta^3 + 8\beta + 5\beta + 1}{4\beta^2 + 8\beta + 3} \quad ...22$$

The dimensionless volume V' then can be obtained after rearranging Equation 21:

$$V' = \frac{3\chi\beta + 4\chi\beta^2 - \beta - \beta^2}{\beta + \beta^2 + \chi + \chi\beta} \quad ...23$$

in which $\chi = \gamma - \beta$.

Equation 22 can be solved for values of β and the result substituted into Equation 23. The resulting values of V' can, in turn, be related to γ because $\gamma = \beta + \chi$.

The results of these calculations are shown in Figure 2, in which V', the ratio of additional drawdown volume to volume for one pump, is plotted vs. γ, the ratio of additional pump discharge to discharge for one pump. To calculate additional drawdown volume

V_2, volume V_1 is calculated from Equation 7. Next, γ is calculated and V' picked off from Figure 2. V_2 is the product V_1V'.

The wet-well reduction that can be achieved by using Scheme II instead of Scheme I for multiple pumps is illustrated in Figure 3. In this figure, the ratio of the additional volume for Scheme II (from Figure 2) to that for Scheme I is plotted vs. γ. (The dimensionless volume, taken from Equation 12, for Scheme I equals γ.) Figure 3 shows, for example, that if additional pump capacity is equal to the base capacity, the additional wet-well volume required for Scheme II is about four-tenths that for Scheme I. Thus, substantial savings in wet-well volume sometimes can be attained.

Figures 2 and 3 indicate that the equation developed for Scheme II cannot be used for γ (the ratio of additional pump capacity to base pump capacity) of less than one-third because it yields negative volumes in this range. This limitation should not detract greatly from the usefulness of this approach because increases in pump capacity are generally greater than one-third.

Example

Let the capacity of the first pump be 1,000 gpm (3.8 cu m/min) and the dis-

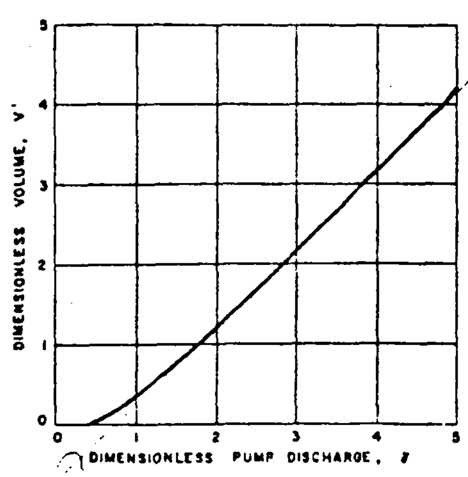


FIGURE 2.—The dimensionless volume required depends on pump discharge (Scheme II).

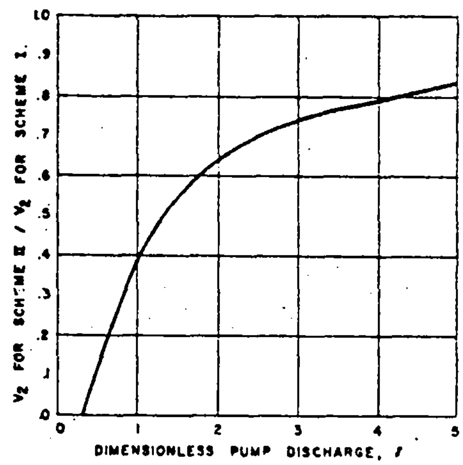


FIGURE 3.—The ratio of volume increment for Scheme II to volume increment for Scheme I varies with values for pump discharge.

charge with the second pump operating be 3,000 gpm (11.4 cu m/min). Calculate the required drawdown volumes for a minimum cycle time of 10 min.

From Equation 7, the drawdown volume for the first pump is:

$$V_1 = \frac{(10 \text{ min})\ (1{,}000 \text{ gpm})}{4}$$

$$= 2{,}500 \text{ gal } (9.5 \text{ cu m})$$

For Scheme I, the additional drawdown volume for the second pump (from Equation 12) is:

$$V_2 = \frac{(10 \text{ min})\ (2{,}000 \text{ gpm})}{4}$$

$$= 5{,}000 \text{ gal } (19 \text{ cu m})$$

For Scheme II, γ, the ratio of the discharge increment to the discharge of the first pump, is 2,000/1,000, or 2. From Figure 2, V' is 1.28 and V_2 is, therefore, 1.28 times 2,500, or 3,200 gal (12.1 cu m), a reduction of 1,800 gal (6.8 cu m) from Scheme I.

Subsequent Pumps

A similar analysis can be made for the third or subsequent combinations. The equations become longer, however, and unwieldy. It is better, in these cases, to use a trial-and-error method.

$b_1 = (3.2 - 1.2)/3 = 1.0$

$V' = b_0 + b_1\gamma$ $b_0 = 4.2 - 5 = -0.8$

$V' = -0.8 + \gamma$, $\gamma > 0.8$

For example, the equation for the cycle time for the third pump is:

$$T_{3rd\ pump} = \frac{V_1 + V_2 + V_3}{C - c} + \frac{V_1}{A + B + c} + \frac{V_2}{B + c} + \frac{V_3}{c} \quad ..24$$

where C is increase in pump discharge by addition of the third pump and c is the inflow greater than the capacity of the first two pumps combined. The terms V_1, V_2, A, B, and C have been determined, and the others are required. One trial-and-error solution is to try a value of V_3 and to find the corresponding minimum cycle by calculating $T_{3rd\ pump}$ for various values of c. Another value of V_3 is selected and the process continued until the desired minimum cycle time is obtained.

Application

Economies resulting from use of variable-speed drive for pumping units have allowed these units to replace constant-speed pumps in many instances. Among the advantages in the former units is that wet-well size can be reduced greatly. In addition, flow surges, which occur when pumps start and stop, are eliminated because the discharge matches incoming flow. Constant-speed drives still appear to be more economical in some cases, however, and should continue to find use. Williams and Kubik (3) indicate that the constant-speed pumps may be more efficient than variable-speed pumps in cases where static lift predominates over pipeline friction.

Summary

Equations for determining wet-well capacities to prevent too frequent starting of fixed-speed wastewater pumps have been derived and presented. The equations show that substantial wet-well volume reductions can be obtained if adequate consideration is given to the operating scheme.

Appendix

Notation

A = pump discharge for first pump or pump combination,
B = additional pump discharge for second pump or pump combination,
b = difference between influent flow rate and A,
C = additional pump discharge for third pump or pump combination,
c = difference between influent flow rate and $A + B$,
Q_{in} = influent flow rate,
Q_{out} = pump discharge,
T = cycle time,
T_{min} = minimum cycle time,
t_{on} = on time,
t_{off} = off time,
V = volume of wet-well between pump start and stop,
V_1 = volume of wet-well drained by first pump,
V_2 = volume of wet-well between start of first pump and start of second pump,
V' = dimensionless volume = V_2/V_1,
β = dimensionless inflow = b/A,
γ = dimensionless pump discharge = B/A,
τ = dimensionless time, and
χ = dimensionless flow = $\gamma - \beta$.

References

1. Burlingame, R. S., Unpublished design computations (1951)
2. Burdoin, A. J., "Pumps, Measuring Devices, Hydraulic Controls." Seminar Papers on Waste Water Treatment and Disposal, Boston Society of Civil Engineers, 137 (1961).
3. Williams, S. W., Jr., and Kubik, A. W., "Variable Speed Drives for Sewage Pumps." *Jour. Water Poll. Control Fed.*, **35**, 1493 (1963).

APPENDIX J

The Linear Decision Rule in Reservoir Management and Design. 1, Development of the Stochastic Model

CHARLES REVELLE

Johns Hopkins University, Baltimore, Maryland 21218

ERHARD JOERES

Johns Hopkins University, Baltimore, Maryland 21218

WILLIAM KIRBY

U. S. Geological Survey, Washington, D.C. 20242

Abstract. With the aid of a linear decision rule, reservoir management and design problems often can be formulated as easily solved linear programing problems. The linear decision rule specifies the release during any period of reservoir operation as the difference between the storage at the beginning of the period and a decision parameter for the period. The decision parameters for the entire study horizon are determined by solving the linear programing problem. Problems may be formulated in either the deterministic or the stochastic environment.

INTRODUCTION

The problem of reservoir management is drawing increased attention in this decade primarily as a result of the introduction of systems analysis and operations research methodology to the field of water resource planning. The tools of these disciplines are being applied to integrate the many functions of a reservoir, including flow augmentation, flood protection, recreation, irrigation, and hydropower. In addition the techniques are aimed at making risk explicit, so that the possible consequences of certain patterns of reservoir management are clear. Finally the methods are seeking rational decision rules or policy functions to simplify the decision-making in reservoir regulation. The goal then is a set of decision rules which are easy to apply and which when applied meet the multiple objectives of the reservoir system with explicit statements of risk.

The recent literature on reservoir regulation has concentrated on optimization methods in seeking policy functions. *Thomas and Watermeyer* [1962] applied linear programing to determine the set of releases maximizing the expected value of benefits. *Loucks* [1968] proposed a stochastic linear programing model to determine a strategy for releases, given the current state of the system and previous inflow. The objective was to minimize the sum of the squared deviations from target flows. Input data consisted of an inflow transition probability matrix. Reservoir volumes, releases, and inflows were restricted to a small set of integers to prevent expansion of the problem to an unmanageable size.

Young [1967] applied dynamic programing to determine the set of releases minimizing a loss function over a long record in which the inflow in each year is given. Using the releases supplied by the program, he then performed several regressions to relate the optimal releases to storage and inflows. He concluded that linear rules provide as good or better fit to the data than more complicated rules, e.g., quadratic or cubic.

Hall et al. [1968] maximized the total return from the operation of a single reservoir, where returns accrue from the sale of water and energy. Inflows during the planning period followed a given sequence, and optimal decisions on releases and energy supply were derived by dynamic programing.

Schweig and Cole [1968] considered the control rules for two linked reservoirs allocating water to meet a common demand. Dynamic programing was utilized to select a set of release rules which are functions of reservoir contents. Their objective was to minimize the total of long-term costs of transmission and shortage. Inflows were treated as random variables.

A recent and thorough review of many of the approaches in the last decade is presented by *Roefs* [1968].

THE LINEAR RULE

A fundamental approach for optimizing reservoir regulation that has promise for development and application is the linear decision rule. The linear decision rule appears first to have been incorporated into an optimization method by *Charnes et al.* [1958]. They treated a problem of refining heating oil to meet stochastic weather-dependent demands in which the quantity refined in each period was chosen to be linear in the demand of the previous period. Unsolvable programing problems were converted to linear programing problems by use of this device. *Young* [1967] introduced the rule to water resources planning, utilizing the rule in a postoptimization analysis of a set of reservoir releases.

As applied to a reservoir, the simplest form of the linear decision rule is

$$x = s - b$$

where x is the release during a period of reservoir operation; s is the storage at the end of the previous period; and b is a decision parameter chosen to optimize some criterion function. This rule is to be interpreted as an aid to the reservoir operator's judgment in selecting a release commitment to be honored under normal conditions. In exceptional cases, however, the actual release during the time period might have to differ from the commitment x. For example the optimal value of the decision parameter b might be negative, so that commitments might be made to release more than was in storage at the beginning of the time period. Under normal circumstances this commitment might be perfectly feasible, but one then would have to be prepared to take the consequences of insufficient inflow during the period.

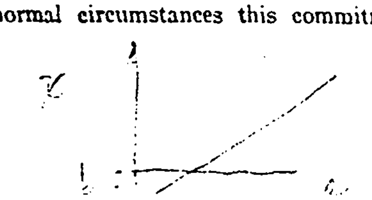
Such a rule would be easily applied in practice and is in addition intuitively appealing in its structure. Additional advantages are detailed after the problem formulations. It is necessary, however, to point out that a linear decision rule might not be the best rule for any given system. A power rule, a fractional rule, or some combination thereof with different rules for each period might yield a better value of the criterion function. But such rules frequently lead to unwieldy problems that are exceedingly difficult to solve. Formulations utilizing the linear decision rule on the other hand have been examined for mathematical tractability and have been found in many cases to lead to linear programing problems.

These linear decision rules can be applied in two frameworks: (1) the stochastic framework where the magnitudes of reservoir inputs are treated as random variables unknown in advance and (2) the deterministic framework where the magnitude of each input in a sequence is specified in advance, either from historic records or from a simulation or synthesis based on the statistical properties of the streamflow process.

A RESERVOIR DESIGN PROBLEM

A dam is to be built to provide a regulated outflow for waste dilution, water supply, and other uses and to provide pools for recreation and flood control. The intent of the dam builder is to provide a dependable supply for the downstream users. To this end he will issue at the beginning of each time period a commitment to release a total volume of exactly x_i during the ith time period insofar as reasonably possible. The downstream users will consider this release commitment in planning their activities for the time period; should the actual release either exceed or fall short of this commitment, their plans might go awry.

The projected requirements of the downstream users are expressed by minimum acceptable releases q_i to be supplied in period i. To prevent excessive channel erosion, waterlogging of fields near the stream, and other damage that would occur if the release were too large, the release during period i should not



$b_i = S_{i-1} - q_i$

exceed the volume f_i. Another set of requirements is imposed by the other uses of the reservoir. For recreational and esthetic (insect and odor control) purposes it is desirable to maintain the storage in the reservoir above a lower limit s_{min} during all time periods; this limit is not critical, and it will be satisfactory if the storage at the end of each period lies above this level. An additional requirement imposed by flood control considerations is that a freeboard of at least v_i be available at the end of each period for storing floods that might occur in the next period.

The engineering problem is to find an operating policy (a formula for the release commitments x_t) that causes the requirements to be satisfied while minimizing the size, and hence the cost, of the dam required.

DETERMINISTIC FORMULATION

The example is structured first in the deterministic environment. A 20-year sequence of monthly inputs is postulated. It is required to find twelve linear decision rule parameters, one for each month of the year, that minimize the reservoir capacity required to meet the specified performance characteristics with the postulated input sequence. The following symbols are defined:

- q_i, minimum release to be provided in the ith month of the year;
- f_i, maximum allowable release in the ith month of the year;
- v_i, flood storage capacity required at the end of the ith month of the year;
- b_i, linear decision rule parameter for the ith month of the year, to be determined;
- c, reservoir capacity, to be determined;
- s_{min}, minimum storage to be maintained, expressed as a fraction a_m of the reservoir capacity;
- s_0, initial storage in the reservoir, expressed as a fraction a_0 of the capacity;
- r_t, postulated reservoir input in the tth month of operation;
- x_t, release during the tth month of operation, to be determined by the linear decision rule;
- s_t, storage at the end of the tth month of operation, to be determined by the linear decision rule.

All variables are measured in volumetric units. The variables q, f, v, and b are indexed by a parameter $i = 1, \cdots, 12$ because their values in the ith month are the same from year to year. The variables r, x, and s, however, do not follow a regular cyclic pattern and therefore are indexed by the parameter $t = 1, \cdots, 240$. The correspondence between i and t therefore is

$$i = t(\text{mod } 12) = \text{remainder of } t/12$$

The linear decision rule then is

$$x_t = s_{t-1} - b_i$$

The equation of continuity for the reservoir is

$$s_t = s_{t-1} - x_t + r_t$$

Substitution of the linear decision rule into the continuity equation yields

$$s_t = b_i + r_t \tag{1}$$

Substituting a similar equation for s_{t-1} into the decision rule yields the following expression for the release during period t:

$$x_t = r_{t-1} + b_{i-1} - b_i \tag{2}$$

The engineering specifications on release commitments and storage utilization can be expressed mathematically by treating them as limitations on the range of decisions acceptable at each point in time at which decisions are to be made. The constraints take the following form:

1.1a The freeboard $c - s_t$ at the end of period t must be greater than v_i.

$$c - s_t \geq v_i \qquad (t = 1, \cdots, 240)$$

In the deterministic sense this is equivalent to saying that the decision at the beginning of time period t should not lead to insufficient freeboard at the end of t, given the extremes of the hydrologic record. The longer the record in general the more severe the observed extremes and the lower the probability of violating the constraints in practice.

1.2a The storage at the end of period t must be greater than the minimum storage required.

$$s_t \geq s_{min} \qquad (t = 1, \cdots, 240)$$

In terms of the decision to be made at the beginning of period t, the constraint limits the control function to those linear rules which lead to storages exceeding s_{min}, given the extremes of the hydrologic record. The longer the record the smaller the likelihood of observing a worse extreme and hence violating the constraint.

1.3a The release in period t must exceed q_t.

$$x_t \geq q_t \qquad (t = 1, \cdots, 240)$$

1.4a The release in period t must be less than f_t.

$$x_t \leq f_t \qquad (t = 1, \cdots, 240)$$

These last two constraints further limit the range of decision rules which can be considered.

Substitution of equations 1 and 2 into 1.1a to 1.4a yields

1.1b $$c - b_i \geq v_i + r_t$$

1.2b $$b_i \geq s_{min} - r_t$$

1.3b $$b_{i-1} - b_i \geq q_i - r_{t-1}$$

1.4b $$b_{i-1} - b_i \leq f_i - r_{t-1}$$

$$(t = 1, 2, \cdots, 240)$$

This small problem (small in the sense that only twenty years of records are considered) poses about 960 constraints at first glance. A remarkable property, however, is observed in the constraints, namely that each constraint appears in the same form twenty times, except for a different stipulation on the right-hand side. Of each constraint's twenty appearances then, one occurrence should be more restrictive than any other. Only this dominant constraint need be retained.

In writing the constraint set in its final form, the term s_{min}, the minimum storage, is set equal to some fraction of the total capacity and the term s_0, the initial storage, is some different and larger fraction of the capacity

$$s_{min} = a_m \cdot c$$

$$s_0 = a_0 \cdot c \qquad (a_0 \geq a_m)$$

The constraint set now becomes

1.1c $$c - b_i \geq \max_n (r_{i+12n}) + v_i$$

$$(i = 1, \cdots, 12)$$

1.2c $$a_m c - b_i \leq \min_n (r_{i+12n})$$

$$(i = 1, \cdots, 12)$$

1.3c $$b_{i-1} - b_i \geq q_i - \min_n (r_{i-1+12n})$$

$$(i = 2, \cdots, 12)$$

$$b_{12} - b_1 \geq q_1 - \min_n (r_{12+12n})$$

$$a_0 c - b_1 \geq q_1$$

1.4c $$b_{i-1} - b_i \leq f_i - \max_n (r_{i-1+12n})$$

$$(i = 2, \cdots, 12)$$

$$b_{12} - b_1 \leq f_1 - \max_n (r_{12+12n})$$

$$a_0 c - b_1 \leq f_1$$

The total number of constraints is now 50 rather than the 960 encountered earlier. The number of unknowns is 13. The objective is to minimize the size of the reservoir

Minimize c

The problem is finding the smallest reservoir that will deliver flows in the specified range over the entire record under the added constraint of a linear decision rule. The results of solution will be the required reservoir capacity and the twelve decision parameters constituting the decision rule for management of the reservoir.

Constraints 1.1c-1.4c ensure absolutely that the release and storage requirements would be met if this optimal linear decision rule were applied to the postulated input sequence. In practice, however, future reservoir inflows are not known with certainty, so there is no absolute assurance that this policy will yield the desired releases and storages in the future. On the contrary one may estimate that in each month i there is a probability of 2/21 that the input will lie outside the 20-year recorded range of inputs. Consequently in the absence of any information to the contrary one might expect that in each future month the probability of violating some of the constraints also would be 2/21.

These observations indicate two shortcomings of the deterministic formulation. First the deterministic formulation yields no explicit statement of the reliability with which the reservoir will meet the specified performance objectives in the future. Second the reservoir's reliability is fixed fortuitously by the specific postulated input sequence and is not under the direct control of the designer. Chance-constrained programing introduced by Charnes and others in the same series of papers that brought forth the notion of the linear decision rule can be used to eliminate these deficiencies in the deterministic formulation of the reservoir management problem.

CHANCE-CONSTRAINED FORMULATION

The example is now restructured in the stochastic environment. Flows in particular periods are not specified and are known only with some probability. That is, the total discharge in the ith month of the year is treated as a random variable R_i having the cumulative probability distribution function

$$F_{R_i}(r) = P[R_i \leq r]$$

In addition the constraints are now expressed as limitations on the allowable risk of violating the performance requirements.

To illustrate the construction and interpretation of the chance constraints, the flood freeboard requirement is treated in detail. The other constraints are formulated in the same general way and represent a similar point of view.

The freeboard requirement is that a volume of at least v_i be available at the end of month i for temporary storage of flood peaks. Thus honoring the release commitment for month i should lead to a storage volume no greater than $c - v_i$ at the end of the month. The decision parameter for month i therefore should be chosen so that the inequality

$$b_i + R_i \leq c - v_i$$

is true.

Now b_i, c, and v_i are constants by hypothesis and R_i is a random variable assuming different values in the ith months of different years. In general then unless $c - v_i - b_i$ exceeds the maximum possible value of R_i, the inequality occasionally will be false. This inequality moreover does not express the flood freeboard constraint in a form acceptable to mathematical programing algorithms.

The problem was sidestepped in the deterministic formulation by replacing the random variable R_i by the postulated input sequence. The price paid for this evasion was described above. It is perfectly possible, however, to come to grips with the problem by recognizing that the probability of the inequality's being true can be determined.

The inequality represents the event that honoring the release commitment given by the linear decision rule would lead to sufficient flood storage capacity at the end of the month. The probability of this event is given by the cumulative probability distribution function of the input R_i as follows:

$$P[b_i + R_i \leq c - v_i] = P[R_i \leq c - v_i - b_i] = F_{R_i}(c - v_i - b_i)$$

If in addition to v_i arbitrary values of c and b_i are proposed, the designer can get an estimate of this probability from the empiric frequency function of observed discharges in month i. If this probability is close to unity, sufficient flood freeboard will be available at the end of the ith month in most years, and this choice of c and b_i will be acceptable from the flood freeboard standpoint. If this probability is less than some value α_i, however, this choice c and b_i does not provide sufficiently dependable flood storage capacity. The choice of α_i of course is the designer's.

It is not necessary to use trial and error to delineate the collection of (c, b_i) pairs yielding sufficiently reliable flood storage capacity. This collection contains all (c, b_i) pairs for which the inequality

$$F_{R_i}(c - v_i - b_i) \geq \alpha_i$$

is true. This inequality may be called a chance constraint on b_i and c. For mathematical programing it is preferable to rewrite the chance constraint as its certainty equivalent

$$c - v_i - b_i \geq r_i(\alpha_i)$$

where $r_i(\alpha_i)$ is the 100 α_i percentile of the input R_i. (That is, $r_i(\alpha_i)$ is the solution for x of the equation $F_{R_i}(x) = \alpha_i$.) Thus the admissible values of c and b_i must satisfy the constraint

$$c - b_i \geq r_i(\alpha_1) + v_i$$

Although formally identical to the deterministic formulation, this probabilistic representation of the flood freeboard requirement has several advantages. Most important the chance-constrained formulation comes squarely to grips with the impossibility of absolutely ensuring the specific performance of a reservoir fed by random inputs. As one result this formulation attaches a statement of reliability to the mathematical representation of each performance requirement. The level of reliability at which each requirement is satisfied moreover is under the direct control of the designer.

A related advantage of the probabilistic formulation is that it clarifies the operational significance of the decision rule. In the deterministic formulation, one might interpret the linear decision rule as a specification of the actual reservoir outflow during the next month. In practice, however, this interpretation may lead to confusion when it is recognized that excessively large or small inflows during a month may make it physically impossible to release a specified volume. The probabilistic formulation on the other hand emphasizes that the linear decision rule is merely an aid to the operator's judgment in deciding how much to release during a month. If the rule is followed, the release commitment will be compatible with the reservoir performance requirements with a specified degree of reliability. When a conflict does arise, however, the operator has the ability to adjust the actual release in the light of the specific conditions of the case.

Finally the chance-constrained formulation of the performance requirements seems to permit more direct economic interpretation of the constraints than the deterministic formulation. It might be asked for example if there would be any advantage in changing the flood control performance requirement. The form of this requirement suggests that the specified freeboard v_i is based on hydrologic analysis of a standard design flood and that more detailed physical and economic data on the relation between flood damages and the flood freeboard are not readily available. Thus the designer cannot immediately interpret the marginal costs that the deterministic formulation would associate with changes in the freeboard specification v_i. In the probabilistic formulation on the other hand the marginal costs are associated with changes in the reliability with which the specified freeboard is made available and hence with changes in the reliability of protection against the design flood. The economic consequences of changes in this reliability appear clearer than those of changes in the freeboard specification.

The remaining performance requirements are now formulated as chance constraints using the same technique as in the flood freeboard constraint. For uniformity the major steps in the derivation of the flood freeboard constraint are repeated.

The results 1.1*b* through 1.4*b* are used to recast the original performance requirements as chance constraints.

2.1*a* The freeboard at the end of period i must exceed v_i with probability α_1:

$$P[c - b_i \geq R_i + v_i] \geq \alpha_1 \qquad (i = 1, \cdots, 12)$$

2.2*a* The storage at the end of period i must exceed s_{min} with probability α_2.

$$P[b_i \geq s_{min} - R_i] \geq \alpha_2 \qquad (i = 1, \cdots, 12)$$

This statement restricts the solution to decision rules which in at least a fraction α_2 of their applications lead to no conflict between the release commitment and the minimum storage requirement.

2.3*a* The release in period i is at least q_i with probability α_3.

$$P[b_{i-1} - b_i \geq q_i - R_{i-1}] \geq \alpha_3 \qquad (i = 1, 2, \cdots, 12)$$

2.4*a* The release in period i is less than f_i with probability α_4.

$$P[b_{i-1} - b_i \leq f_i - R_{i-1}] \geq \alpha_4 \qquad (i = 1, 2, \cdots, 12)$$

Constraints 2.3*a* and 2.4*a* restrict the choice of release commitments in period $i - 1$ to ensure that it will be possible with the specified reliability to make commitments in the desired range at the beginning of month i.

Constraints 2.1a to 2.4a do not imply guarantees about the storages or flows, nor do they say how excessive or how insufficient storages or flows may be. The constraints do imply, however, that most decisions made with the release rules will not lead to conflicts between release commitments and storage and release requirements.

These chance constraints open the way for a new kind of constraint, not possible in the deterministic formulation. A single example is offered, but the idea should be easy to extend. We have already presented a chance constraint on low flow (2.3a), and q_i must be achieved with probability α_3 (say .90). We may now buttress the formulation with a second level constraint on low flow, indicating that q_i' (some lower value) must be achieved with probability α_3' (some higher probability, say .95). For example the release in the ith period should exceed 500 million gallons with probability .90, but flows above 400 million gallons must be achieved at least 95% of the time. Mathematically this second level constraint has the same form as the first level constraint.

The chance constraints 2.1a to 2.4a are converted to a more convenient form with the aid of the cumulative distribution functions of the random monthly inputs R_i.

2.1b $$F_{R_i}(c - b_i - v_i) \geq \alpha_1$$

2.2b $$1 - F_{R_i}(s_{min} - b_i) \geq \alpha_2$$

2.3b $$1 - F_{R_{i-1}}(q_i - b_{i-1} + b_i) \geq \alpha_3$$

2.4b $$F_{R_{i-1}}(f_i - b_{i-1} + b_i) \geq \alpha_4$$

Let us suppose that $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = .90$. The cumulative distribution functions F_{R_i} can be estimated from monthly streamflow records. The following symbols are defined:

- $r_i^{.90}$, the flow which is exceeded in period i only 10% of the time (the value that the random variable has less than 90% of the time). $F_{R_i}(r_i^{.90}) = .90$;
- $r_i^{.10}$, the value which the flow in period i falls below only 10% of the time. $F_{R_i}(r_i^{.10}) = .10$.

The explicit statements of the chance constraints are

2.1c $$c - b_i \geq r_i^{.90} + v_i \quad (i = 1, \cdots, 12)$$

2.2c $$a_m c - b_i \leq r_i^{.10} \quad (i = 1, \cdots, 12)$$

2.3c $$b_{i-1} - b_i \geq q_i - r_{i-1}^{.10} \quad (i = 2, \cdots, 12)$$

$$b_{12} - b_i \geq q_1 - r_{12}^{.10}$$

$$a_0 c - b_1 \geq q_1$$

2.4c $$b_{i-1} - b_i \leq f_i - r_{i-1}^{.90} \quad (i = 2, \cdots, 12)$$

$$b_{12} - b_1 \leq f_1 - r_{12}^{.90}$$

$$a_0 c - b_1 \leq f_1$$

Of course the criterion is still

Minimize c

The results of solving such a problem are a reservoir size and twelve constants, each constant determining the release for a given period based on the storage at the end of the preceding period.

EXTENSIONS

There are numerous other problems which can be structured in the same general way. A brief list includes

1. Maximize the expected value of the average summer flow where a reservoir size is specified as well as the minimum storage and minimum and maximum releases to be reliably maintained. Such a criterion might be used in a situation in which waste dilution is a principal concern.
2. Maximize the expected value of the average storage or maximize the storage that can be maintained with 90% reliability where reservoir size is given in addition to certain operational requirements. These criteria might be applied if recreation were an important use of the reservoir.
3. Maximize the freeboard volume that can be maintained with 90% reliability or maximize the expected value of the freeboard volume where reservoir size and operational requirements are specified. These criteria would be applicable if flood control were a primary use of the reservoir.

4. Minimize the average of the absolute deviations of the expected values from the target flows in each month where reservoir size and operational requirements are specified.

Initial investigation shows that most of these problems can be treated in the same general way as the example problem.

Another problem whose study can be facilitated by use of the linear decision rule is an economic analysis of low flow augmentation. In the example we have shown how to determine the minimum size (and hence minimum cost) reservoir to deliver a given low flow. We can use another linear program to determine the least cost pattern of treatment efficiencies to meet dissolved oxygen standards downstream from the reservoir [*ReVelle et al.*, 1968]. For each low flow a minimum cost reservoir and minimum cost treatment pattern may be determined. If augmentation is economic, the curve of total cost (reservoir and treatment) as a function of low flow should show a minimum. Whether this occurs will undoubtedly vary with the specific circumstances, but each case can be explicitly investigated for the trade off between augmentation and treatment.

A second problem of interest is the consideration of systems of interconnected reservoirs, where the release from one or more reservoirs plus tributary flows constitutes the input to the next. A possible criterion is to minimize the total cost of building the reservoir system that meets certain maximum and minimum flow requirements with all releases following a linear rule. A prior drawback to such a problem was the fixed and concave cost functions which reservoirs display. A recent algorithm by *Walker and Lynn* [1968] has been shown to be efficient in locating optimal solutions to minimizing problems with fixed and concave cost objective functions and thus would be a promising method to apply here. In addition the criteria of maximizing or minimizing flow or minimizing the range of flows or deviations from targets could be utilized.

CONCLUSIONS

The use of the linear decision rule in either of the two frameworks presented yields a number of distinct advantages and several limitations.

First although the solutions obtained under the linear rule may not be optimal relative to the class of all possible decision procedures, this seems a small concession to obtain an optimization problem which can be solved with known techniques. *Young's* [1967] results moreover indicate that linear decision rules may be as good as more complicated rules. In addition the linear rule is intuitively appealing and simple to apply in practice. Furthermore the dominance relations in the deterministic framework and the chance-constrained relations in the stochastic framework both lead to optimization problems of small size, so that a computer solution is not burdensome.

Another significant advantage of the linear decision rule, especially when used in conjunction with chance constraints, is its clarification of the role of operating policies in optimal reservoir design. Computers and mathematical optimization procedures notwithstanding, we believe that reservoir design is a creative art and that the quality of the design depends in great part on the designer's ability to visualize the interaction of all components of the proposed system. Use of the linear decision rule in formulating the performance requirements either as deterministic constraints or as certainty equivalents of chance constraints gives the designer a precise representation of the interaction of operating policies and reservoir capacity. It shows in particular that the optimal reservoir capacity is a function of the operating policy and therefore suggests that more attention be given to selection of an operating policy before choosing the physical capacity of the reservoir.

It is necessary to point out that constraints on the range of flows, if there is no consideration of flows in the objective function, imply that the benefit function associated with flow is relatively flat. It is obvious however that a rough indication of the benefit function's shape is given by the placement of the release constraints. The device of multiple constraints satisfied with different probabilities can be used to improve the definitions of any such benefit functions in the chance-constrained formulation.

Finally the element seen as the most advantageous feature of such formulations is that risk is made explicit in stochastic problems. Reservoirs operated under linear rules based on chance-constrained formulations should perform at the levels of certainty that the designer or decision maker specifies.

APPENDIX: AN EXAMPLE PROBLEM IN THE STOCHASTIC ENVIRONMENT

The reservoir design problem is structured here in the chance-constrained programing formulation. The problem is modified by the additional definition that

$$b_i = h_i - g_i$$

This definition is included so that the decision parameters b_i may take on positive or negative values. Most linear programing codes restrict solutions to positive variables, making this substitution necessary.

The problem is to minimize the size of the reservoir which meets requirements on freeboard, minimum storage, low flow, and high flow with 90% reliability for each month. The mathematical formulation is

Minimize c

subject to

1. $$c + g_i - h_i \geq r_i^{.90} + v_i \qquad i = 1, 2, \cdots, 12$$

2. $$a_m c + g_i - h_i \leq r_i^{.10} \qquad i = 1, 2, \cdots, 12$$

3. $$g_i - h_i - g_{i-1} + h_{i-1} \geq q_i - r_{i-1}^{.10} \qquad i = 2, 3, \cdots, 12$$

$$g_1 - h_1 - g_{12} + h_{12} \geq q_1 - r_{12}^{.10}$$

$$a_0 c + g_1 - h_1 \geq q_1$$

4. $$g_i - h_i - g_{i-1} + h_{i-1} \leq f_i - r_{i-1}^{.90} \qquad i = 2, 3, \cdots, 12$$

$$g_1 - h_1 - g_{12} + h_{12} \leq f_1 - r_{12}^{.90}$$

$$a_0 c + g_1 - h_1 \leq f_1$$

Table 1 lists the data for the problem, including the tenth and ninetieth percentiles of the probability distribution functions of the monthly discharges of a stream in Maryland. The problem is solved twice: once under the assumption that the fractional storage to be maintained with 90% reliability is 0.40 and again with the fractional storage at 0.10. The decision rules and capacities derived from the solution of the two problems are listed in Table 2.

The reader is reminded that not every reservoir design problem has a feasible solution. It would not be possible for example to deliver outputs reliably and consistently in excess of inputs. In this connection it is instructive to examine the first twelve constraints of set 3. Since all these constraints are to be satisfied simultaneously, adding them yields the following necessary condition for the existence of a feasible solution:

$$\sum_{i=1}^{12} q_i \leq \sum_{i=1}^{12} r_i^{.10}$$

That is, there exists no reservoir meeting the specified performance objectives under a linear decision rule unless the sum of the desired releases is less than the sum of the monthly inputs occurring at the corresponding desired reliability level. Although most linear programing codes automatically check for the existence of feasible solutions, this necessary condition is a natural one for the designer himself to keep in mind.

To appreciate better the significance of this necessary condition, suppose the monthly reser-

TABLE 1. Data for the Example Problem

(a) The 10 and 90 Percentiles of the Probability Distributions of Monthly Flows

Month	Flow exceeded 90% of the time, $r_i^{.90}$, billion gallons	Flow exceeded 10% of the time, $r_i^{.10}$, billion gallons
$i = 1$	4.41	14.46
2	4.54	15.50
3	6.29	18.40
4	5.54	16.82
5	4.89	14.92
6	3.82	12.00
7	3.19	12.01
8	2.54	12.59
9	2.18	9.62
10	2.54	8.82
11	2.73	10.19
12	3.57	11.55
Sum	46.24	156.88

(b) Operational Requirements

$a_0 = 0.60$	
$q_i = q = 3.80$	Problem 1: a_m = 40% of capacity.
$f_i = f = 16.00$	Problem 2: a_m = 10% of capacity.
$v_i = v = 4.00$	

TABLE 2. Linear Decision Rules

Month	Problem 1 $a_m = .40$	Problem 2 $a_m = .10$
1	$X_1 = S_{12} - 9.68$	$X_1 = S_{12} + 2.16$
2	$X_2 = S_1 - 10.29$	$X_2 = S_1 + 1.55$
3	$X_3 = S_2 - 10.39$	$X_3 = S_2 + 0.81$
4	$X_4 = S_3 - 12.88$	$X_4 = S_3 - 1.65$
5	$X_5 = S_4 - 14.62$	$X_5 = S_4 - 3.39$
6	$X_6 = S_5 - 15.71$	$X_6 = S_5 - 4.48$
7	$X_7 = S_6 - 15.73$	$X_7 = S_6 - 4.50$
8	$X_8 = S_7 - 15.12$	$X_8 = S_7 - 3.89$
9	$X_9 = S_8 - 13.86$	$X_9 = S_8 - 2.63$
10	$X_{10} = S_9 - 12.24$	$X_{10} = S_9 - 1.01$
11	$X_{11} = S_{10} - 10.98$	$X_{11} = S_{10} + 0.25$
12	$X_{12} = S_{11} - 9.91$	$X_{12} = S_{11} + 1.32$
Required capacity	33.700	22.467

All units are billion gallons

voir inputs occur in a time-stationary Markovian normal autoregressive process with correlation coefficient ρ. If 90% reliability is desired, the necessary condition for feasibility is that the sum of the desired releases be no greater than $12\mu - 15.6\sigma$. When $\rho = .5$, however, the total annual input is normal with mean $\mu_A = 12\mu$ and standard deviation $\sigma_A = 5.65\sigma$. The maximum feasible annual release therefore is $\mu_A - 2.85\sigma_A$; the probability that the annual input will exceed this volume is about 0.998. Operation at 90% reliability in each month thus yields a maximum feasible annual release far smaller than the reliable annual input. On the other hand the system can deliver the maximum feasible annual release even during a long sequence of very low (tenth percentile, in this example) inputs. It is expected that use of longer time intervals would yield greater maximum feasible releases at the same reliability. (In this case, however, the release rule gives the operator less guidance.) Because of these trade offs among reliability, time interval, and feasible release the designer must give careful consideration to his actual requirements for reliability and decision point spacing.

A similar analysis of the fourth constraint set yields a further necessary condition for the existence of a feasible solution

$$\sum_{i=1}^{12} f_i \ge \sum_{i=1}^{12} r_i^{.90}$$

Finally for the high flow and low flow constraints to hold simultaneously in period i it is necessary that

$$f_i - q_i \ge r_{i-1}^{.90} - r_{i-1}^{.10}$$

An interesting feature of the result is that the differences between the capacity of the reservoir and the minimum storage to be maintained are the same for both problems

Problem 1: $a_m = 0.10$

$$c - 0.10c = 22.467 - 2.247 = 20.220$$

Problem 2: $a_m = 0.40$

$$c - 0.40c = 33.700 - 13.480 = 20.220$$

That these quantities are exactly the same is not surprising, since 20.220 billion gallons evidently represents the storage volume needed to maintain the desired control of releases. We tentatively call this volume the control volume K. For this problem, it seems to be the value the reservoir capacity would take on if the storage required at the end of each interval exceeded zero with 90% reliability. If this result were general, an immediate benefit would be considerable simplification of analysis of the solution's sensitivity to minimum storage.

Assume the control volume can be determined by a single run. Recalling that a_m is the fractional storage maintained with 90% reliability, we can write

$$K + a_m c = c$$

or

$$c = K/(1 - a_m)$$

Thus for any specification of $a_m (a_m \le a_0)$ different from the original specification, the reservoir capacity could be quickly calculated, assuming that other particulars of the problem remain the same.

Although constant control volumes have been observed in all of the several cases studied thus far, our experience is not sufficient to suggest that a constant control volume is a general result; we feel that further research is needed on this phenomenon.

Acknowledgment. Publication of this paper is authorized by the Director, U. S. Geological Survey. The authors wish to thank Mr. James Janis for his assistance in the preparation of the example problems.

Charnes, A., Cost hor approach oil, *Manag*
Hall, W. A., timization pose reser *Resour. R*
Loucks, D. Regulation *Civil Eng*
ReVelle, C., Linear pro management February
Roefs, T., the art, *Publ.*, 320-

REFERENCES

Charnes, A., W. W. Cooper, and G. H. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, *Management Sci.*, *4*(3), 235-263, April 1958.

Hall, W. A., W. S. Butcher, and A. Esogbue, Optimization of the operation of a multiple purpose reservoir by dynamic programming, *Water Resour. Res.*, 4(3), 471-470, June 1968.

Loucks, D. P., Computer Models for Reservoir Regulation, *J. Sanit. Eng. Div., Amer. Soc. Civil Engrs.*, *94*(SA4), 657-669, August 1968.

ReVelle, C., D. P. Loucks, and W. R. Lynn, Linear programming applied to water quality management, *Water Resour. Res.*, *4*(1), 1-10, February 1968.

Roefs, T., Reservoir management: the state of the art, *IBM Washington Scientific Center Publ.*, 320-3508, July 1968.

8- eig, Z., and J. A. Cole, Optimal control of aked reservoirs, *Water Resour. Res.*, *4*(3), 479-498, June 1968.

Thomas, H. A., and P. Watermeyer, Mathematical models: a stochastic sequential approach, in *Design of Water Resource Systems*, Maass et al., Harvard University Press, Cambridge, Massachusetts, 540-569, 1962.

Walker, W., and W. Lynn, Adjacent extreme point algorithms for the fixed charge problem, 23 pp., mimeographed, Cornell University, Ithaca, New York, 1968.

Young, G. K., Finding reservoir operating rules, *J. Hydraulics Div., Amer. Soc. Civil Engrs.*, *93*(HY6), 297-321, November 1967.

(Manuscript received February 3, 1969;
revised May 5, 1969.)