STUDY OF EFFECT OF RAINFALL AND RIVER STAGE ON GROUNDWATER TABLE BY REGRESSION MODEL

Submitted by

ISLAM MOHAMMAD FAISAL

In partial fulfilment of the requirements for the Degree of Master of Science in Engineering (Water Resources)

Department of Water Resources Engineering

Bangladesh University of Engineering and Technology

Dhaka.

October, 1988.

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF WATER RESOURCES ENGINEERING

October 18, 1988

WE HEREBY RECOMMEND THAT THE THESIS PREPARED BY Mr. Islam Mohammad Faisal ENTITLED Study of Effect of Rainfall and River Stage on Groundwater Table by Regression Model BE ACCEPTED AS FULFILLING THIS PART OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN ENGINEERING (WATER RESOURCES).

Chairman of the Committee (Dr. A.F.M. Saleh)

Member (Dr. A. Hannan)

Member (Dr. A. Halim)

Member (Mr. M.A. Karim)

Head of the Department (Dr. A. Halim)

ABSTRACT

A linear multiple regression model has been developed to study the effect of total monthly rainfalls (TMR) on the yearly highest/lowest levels (HWL and LWL respectively) of groundwater table. Twelve different applications of the model has been made to simulate the annual highest/lowest groundwater levels in the Ganges-Kobadak (G-K) Project area. Groundwater developments in this area were kept to a minimum in the past, and this served the prerequisite of a relatively undisturbed subsurface regime for regression study. All the simulations by the model have been found to be statistically satisfactory. Particular emphasis has been put on the techniques of selecting the 'Best Subset' of independent variables (TMRs). It is found that a combination of efficient

algorithm and instantaneous judgement is necessary to obtain the optimum subset of variables which will generate the near largest coefficient of determination, R^2 and the smallest error of estimate, s at the same time.

Possibility of forecasting HWL/LWL with the help of resulting regression equations has also been discussed. It is found that, nine out of twelve of the regression equations are capable of making quite dependable forecast. Frequency analyses have been done for TMRs of appropriate rainfall stations so that forecast of highest or lowest water table due to TMRs having numerous combinations of return periods could be made.

The effect of river stage on groundwater table has also been studied using simple linear regression technique. It is found that, beyond a distance of about 2500 m from the bank, the water table remains virtually insensitive to the river stage.

ACKNOWLEDGEMENT

The author expresses sinceremost gratitude to Dr. A.F.M. Saleh, Chairman of the Board of Examiners for his constant moral support, advice, guidance and understanding throughout this study. Dr. Saleh took personal interest in all the major and minor details of this research work and offered a helping hand whenever needed.

The author also likes to thank the other members of the Board of Examiners : Dr. A. Hannan, Dr. A. Halim and Mr. M.A. Karim. All of them made painstaking review of the thesis and forwarded valuable suggestions.

Grateful acknowledgements are due to the organizations of Institute of Flood Control and Drainage Research (IFCDR), Bangladesh Water Development Board (BWDB), Master Plan Organization (MPO) and Bangladesh Meteorological Department (BMD) for providing all the necessary data.

The author wishes to express special thanks to Mr. Nazmul Haque, Programmer, Computer Center of Bangladesh University of Engineering and Technology (BUET) for his help in using the SPSS and other softwares. Finally, the author is profoundly indebted to his friends and colleagues Manas, Anis and Ferdous for their forebearance in dealing with all the impudent demands by the author in relation to this study which they satisfied promptly and even enthusiastically !

Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.0. Introduction

Modeling of groundwater system is at a preliminary stage in Bangladesh. Initial activities in this field began in around 1976 and they were mostly concerned with the determination of important groundwater related parameters like storage coefficient, transmissivity etc. The simple analytical techniques were used for these purposes. Later on, a number of numerical exercises were also carried on for regional assessment of the groundwater system (Ahmed,1986). But use of statistics, which is an equally powerful tool, is still absent in the studies of groundwater related phenomena.

The annual cycle of groundwater table fluctuation in Bangladesh is predominantly governed by the total monthly rainfalls in different months, specially in areas where there is no artificial interference on the groundwater regime. Also, presence of near by stream affects the level of groundwater table. So far, no attempt was made by any researcher in Bangladesh to statistically correlate the two above mentioned factors with the fluctuation of groundwater table. This encouraged the author to take up the present study.

The primary purpose of the proposed research is to develop a linear multiple regression model to study the effect of total monthly rainfalls (TMR) on the yearly maximum/minimum levels of groundwater table. For testing the goodness of fit, the model will be applied in the Ganges-Kobadak (G-K) Irrigation Project area. A location map of the study area is shown in Fig. 1.1. The G-K area is selected primarily because this area is under surface water irrigation scheme and groundwater developments were kept to a minimum in the past. This ensures a relatively undisturbed groundwater regime suitable for multiple regression study having constant coefficient of determination. Moreover, the effect of river stage on groundwater table at different distances will also be studied. Consequently, a characteristic distance, d_c will be identified for the study area beyond which the water table will supposedly remain insensitive to the fluctuation of river stage. Simple linear regression will be used for this part of the study which may be treated as a special case of multiple regression with single independent variable. Hence, the same methodology and techniques of inferences that will be discussed for the multiple regression study will also be applicable to the simple linear regression. The major incentives for the proposed undertaking based on statistical technique are :

a] No multiple regression model related to groundwater has so far been developed in Bangladesh. So, such a study will open

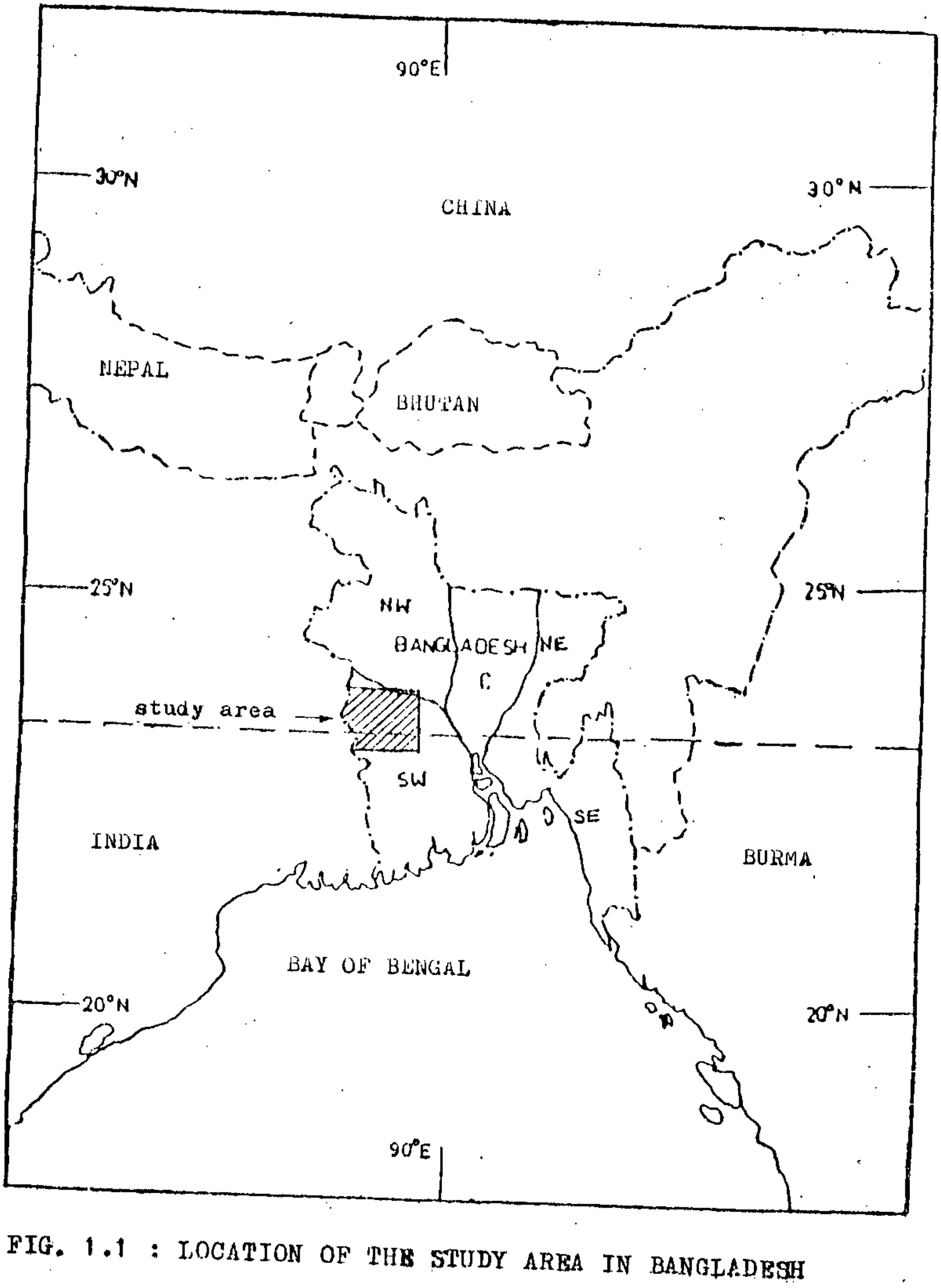


FIG. 1.1 : LOCATION OF THE STUDY AREA IN BANGLADESH

up a new arena of scientific inquisition in this field.

b] A regression model is conceptually simple. It does not require thorough understanding of the underlying principles of a natural process, nor does it require elaborate mathematical background.

c] Provided with adequate good quality data, a regression model is as good as a numerical model. Moreover, it does not require the tiresome 'calibration' phase; rather, a regression model gets automatically calibrated with the insertion of new set of inputs.

d] A regression model generates basic statistics of input and output with little additional effort which give valuable insight into the problem. It also generates variance-covariance and correlation matrix and thus, provides the degree of variability and interdependency of the variables. It produces weight factors from which relative contribution of each variable in the model can be estimated. Such features are absent in any other model.

e] A regression model may allow valuable supplementary analyses of distribution and trend. When adequate data are available, study of distribution and trend become easier by using input statistics generated through the process of model development.

f] A regression model is relatively cheap. Development and execution of other models are either cumbersome or costly. When such models are developed on a regional basis, computer facilities become a prerequisite to handle the huge amount of input and output data. But anyone having a pocket calculator may use the ultimate product of regression model - usually a simple linear equation.

g] The outcome of a multiple regression model is easily transferable to the field level. On the contrary, any other type requires special training for using the model and for interpreting the output.

1.1. Objectives of the Research

From the above made discussions, objectives of this study may be summarized as follows :

i] To develop a linear multiple regression model between monthly total rainfalls and yearly maximum and minimum levels of water table.

ii] To apply the model to six selected dug wells in the G-K Project area, each having water level records of about 20 years. Also to discuss the techniques of determining the 'Best Subset' of independent variables from statistical point of view with a suitable illustration.

iii] To discuss the possibility of using such model for forecasting of maximum and minimum groundwater levels.

iv] To study the effect of river stage on the groundwater table at different distances by simple linear regression and to get the characteristic distance, d_c beyond which the groundwater table is no more affected by the river stage.

Chapter 2

LITERATURE REVIEW

2.0. Models for Groundwater

A model is a tool to represent the simplified version of reality. A good model adequately depicts all the desired features of a physical or environmental process with certain degrees of approximation and idealization.

With the advent of sophisticated experimental and computational facilities, numerous models have already been developed in the field of groundwater which can be broadly classified into :

a) physical models

b) analog models

c) mathematical models

d) hybrid models

2.1. Background of Groundwater Modeling

The fundamental task in developing groundwater related model was done by the great French scientist Henry Darcy(1803-1858). His treatise of 1856 defined the relation, now known as the Darcy's law, governing groundwater flow in most alluvial and sedimentary formations. Later European contributors of the nineteenth century were J.Boussinesq, G.A.Daubree, J.Dupuit, P.Forchheimer and A.Thiem (Todd,1980). Their contributions were mainly of analytical nature and thereby served as the basic tools for developing analytical models.

In the twentieth century, tremendous advancements have been made in getting analytical solutions for the problems of radial flow into a well and time-variant flow through porous media. Among many distinguished contributors are C.V.Theis, C.E.Jacob, M.S.Hantush, R.E.Glover and C.W.Walton.

Consequence of such extensive analytical search was better understanding of the scope and limitations of this process. Researchers looked forward to getting alternative breakthroughs. And, a surge of physical and analog models evolved in the late fifties, which finally merged into the new wave of mathematical models (other than analytical) with the dawn of computer-age in late sixties.

Standard texts which contain good documentations of these developments include Walton(1970), Glover(1974), McWhorter and Sunada(1977), Bouwer(1978), Bear(1979), Freeze and Cherry(1979), Todd(1980) and Rethati(1983).

2.2. Types of Groundwater Models

2.2.1. Physical Model

A physical model is scaled down replica of the field conditions maintaining similarity from both physical and hydraulic points of view. Sand tank model is a typical example. This is one of the earliest type of models ever used to study groundwater flow. P. Forchheimer used one such model for study of well flow in Graz, Austria as early as 1898 (Todd,1980).

Although such a model is good enough for homogeneous and isotropic formation, elaborate treatment is necessary for simulation of features like nonhomogenity, anisotropy, capillary action etc. Another major disadvantage is its lack of flexibility to the changing geohydrological conditions.

2.2.2. Analog Model

Analog models are developed by noticing the similarity of governing equations for flow of fluid (laminar), heat and electricity. For example, Ohm's law is the electrical analog of Darcy's law for laminar flow of fluid. Viscous fluid model, membrane model, thermal model etc. are well known analog models. Relatively recent development in this discipline is the electrical analog model based on conductive solid/liquid or resistance-capacitance(RC) network. RC network is specially flexible and is capable of simulating time-varient flow in non-homogeneous aquifers. Todd(1980) includes extensive discussion on analog models.

2.2.3. Mathematical Model

Mathematical models can be arranged into four subgroups, namely:

i) analytical model
ii) numerical model
iii) Operation research (OR) model
iv) statistical model

2.2.3.1. Analytical Model

An analytical model is usually based on a number of rigorous mathematical equations which are subjected to specified initial and/or boundary conditions. Texts mentioned earlier contain numerous references of such models. Basic advantage of an analytical model is that the model is deterministic (always generates the same output for a specified set of inputs) and well-understood for relatively simple flow conditions. Underlying equations of an analytical model may also serve as the building blocks of equivalent numerical model.

However, analytical model has one serious drawback. As described by Thomas(1973), the equations of flow and continuity in the form of differential equations do not lend themselves easily to rigorous analytical solutions when boundaries are complex. So far, the only remedy is to switch over to alternative modeling techniques.

2.2.3.2. Numerical Model

A common feature of most numerical models is superposition of a regular or irregular grid system over the zone under study. Then the equation of flow is applied to each of the grid points, and using finite difference approximation, a system of linear equations is formed. Computer aided solution of such system usually yields groundwater levels at grid points. Sometimes the entire zone is sub-divided into a number of polygons and finite element technique is applied. Remson et al.(1971) have presented a very worthreading text detailing most of the numerical methods with their advantages and disadvantages. Other interesting titles are : Thomas(1973), Prickett(1975), Pinder and Gray(1977), Boonstra and Ridder(1981) and Wang and Anderson(1982). Since mid seventies, finite element technique has been modified into a more advanced and mathematically complicated form, called the boundary element technique. This method requires much smaller system of equations and hence, saves valuable computer storage. Brebbia(1978) covered the fundamentals of this technique.

Flexibility is the major advantage of a numerical model. The modeller can incorporate almost any pecularities he wants to in the model. However, it requires indepth knowledge and adequate experience to formulate a dependable model. In addition, it requires efficient algorithms, good programming skill and high speed computer for its execution. It is obvious that, developing a numerical model is very often a laborious, time-consuming and costly process. And the final run-time cost is also considerable.

2.2.3.3. Operation Research (OR) Model

Operation research is being used in different branches of engineering since late fifties. But its application in the field of groundwater is relatively new. Both the linear and the dynamic programming approaches have been used depending upon the nature of the problem. Models are developed to optimize different objective functions, such as, the net economic gain from conjunctive use of surface and groundwater subjected to a number of constraints. Among many distinguished contributors, to name a few, are: Dracap(1966), Domenico(1968), Cochran and Butcher(1970), Kleinecke(1971), Chaudhury et al.(1974), Heidari(1982), Gorelick et al.(1984), Willis(1985) and Jones et al.(1987).

OR models are mainly being used in the process of decision making which is also the intended purpose of all linear and dynamic models. Such help is of great use in water resources engineering from economic and environmental point of view.

2.2.3.4. Statistical Model

A statistical model generates output which is always associated with an element of chance or probability. Usually a statistical model has a number of parameters. As described by Haan(1977), they are to be determined in someway from the observed hydrological data. The validity and applicability of a statistical model depend directly on the characteristics of the data used to estimate the parameters. Statistical models can be classified into parametric and stochastic models. In a parametric model, once the parameters are known, the model becomes a deterministic one. A stochastic model produces different outputs even with the repeated use of a specified set of inputs. However, the generated outputs follow a statistical pattern. To exemplify - a multiple regression model is a parametric model and a model generating random events from a predefined distribution is a stochastic model.

It is interesting to note that parametric models are also inherently stochastic. Once the data determining the parameters get changed, so do the parameters.

A statistical model requires some minimum amount of data to produce a reliable output. But collection and monitoring of groundwater data usually have a very short history. Until late sixties, this was the major hindrance of using the powerful tools of statistics for the study of sub-surface flow. So far, majority of the statistical models related to groundwater dealt the problem of solute transport and groundwater contamination.

Pioneers in this field are: Chalky(1949), Scheideggar(1954), Reddell(1967) and de Jong(1969). From seventies and onwards, with the accumulation of workable amount of data, this discipline flourished rapidly. Among significant contributors are : Bibby(1971), Cooley(1973), Flores(1976), Bakr et al(1978), Dagan(1982), Gelhar and Axness(1983), and finally, Black and Freyberg(1987).

2.2.4. Hybrid Model

The term 'hybrid' indicates that the model is a composite one having certain combination of the categories already discussed. Vemuri and Karplus(1969) and Morris et al.(1972) used hybrid computer models which were combination of RC analog and computer based numerical model. They saved numerous iterations by explicitly solving the problems with the help of analog models. Solutions so obtained were then used as feedback to the numerical models.

Another typical example is a combination of statistical and numerical techniques. To avoid tedious trial and error process, linear regression model may be developed which will correlate the numerical model parameters with some basic input data. Reddell and Sunada(1967) and Eshett(1970) discussed such possibilities. Cooley(1973) discussed a hybrid of statistical and analytical techniques. As all the methods have some specific advantages over the others, more and more hybrid models are supposed to come out in future optimizing the overall simulation process.

2.3. Related Works to the Proposed Research Topic

As mentioned earlier, works that have been done so far in statistical modelling of groundwater are not voluminous. This is specially true for the study of rainfall-groundwater relationship by statistical technique. Among the beginners, Ubell worked on the effect of rainfall on groundwater storage in early fifties (Ubell,1953). Bogardi(1953) extended the findings of Ubell and studied the impact of precipitation as well as temperature on changes of water table. Further supporting works in this line were carried on by Csomane(1968), Rethati(1970) and Sing(1981) etc. Later on, Rethati(1983) discussed about a multiple linear regression model to simulate annual highest and lowest groundwater levels (HWL and LWL respectively) with the help of total monthly rainfalls (TMR) of previous water year. This particular work by Rethati will be the basic guide line for major part of this research work.

Study of the effect of river stage on groundwater level, which is also the secondary objective of this research work, is recently gaining increasing attention due to the introduction of the concept of conjunctive use of water resources. McWorther et al.(1972), Glover(1974), Labadie et al.(1983) and many others did excellent works in this field of stream-aquifer interaction.

In Bangladesh, the most relevant work to the proposed research topic was done by Haq and Sattar(1987). In their study, six years of data on annual groundwater table fluctuation was regressed with corresponding total annual rainfall. A simple single variable regression equation was developed which showed that the speculated relationship was very significant. They found the coefficient of determination to be 0.99 - almost a perfect linear fit! They also observed that the water level usually responded favorably after an accumulated rainfall depth of about 75 cm. This amount, as noted by the researchers, took about three monsoon months to accumulate.

Although this simple study precludes any possibility of using it as a forecasting model, it definitely encourages such effort. It will be discussed later on that, breaking up the total annual rainfall into its monthly components and using multiple regression technique may give in quite a dependable forecasting model.

Since 1976 - the beginning of groundwater related modeling in Bangladesh, nine other exercises have been undertaken by organizations like Bangladesh Water Development Board (BWDB), Bangladesh Agricultural Development Corporation (BADC) and Master Plan Organization (MPO). Most of the treatments were based on analytical techniques, except a few numerical ones. The North-West Bangladesh Groundwater Model was the most elaborate among them. Basic aim of all these activities was to achieve better planning and management policies for groundwater use. Some of the above mentioned models, given proper adjustments, may be used to predict the extreme groundwater levels based on previous rainfall pattern. For details, volume III of the second interim report by MPO(1984) may be consulted.

Apart from the models used for professional purposes, models are coming out in recent years from academicians and researchers. Khan(1982) has presented models for groundwater yield and recharge assessment based on combination of analytical techniques. Khan and Mawdsley(1984) also tried to assess the aquifer yield by linear programming. Ahmed(1986) developed another model using simple implicit finite difference scheme to study the groundwater system in the Mymensingh-Tangail area. However, none of these models were concerned with the linear or polynomial relationship between rainfall and groundwater table.

About the effect of river stage on groundwater level, Saleh(1985) developed an analytical model called, Watershed

Irrigation Potential Estimation (WIPE) model, to simulate groundwater movement in a watershed. For a small watershed in the North-West Bangladesh, he found that, beyond 2000 m from the river, the flux from the watertable to the river was negligible and the water table profile of the watershed was not affected by the water level in the river. Later on, Hoque(1986), Khan and Mawdsley(1986), and Michael(1986) dealt with the problem of stream-aquifer system, although their main emphasis was on the theme of conjunctive use of surface and groundwater.

It is quite noticeable that so far no physical, analog or statistical model related to groundwater is attempted by any modeller.

Chapter 3

PRINCIPLES OF LINEAR MULTIPLE REGRESSION

3.0. Definition

Multiple regression is a part of statistics which deals with the investigation of the relationship between three or more variables related in a probabilistic fashion (Devore, 1982).

3.1. The Linear Probabilistic Model

For the deterministic model $y = B_1 + \sum(B_i x_i)$ where $i=2,3,..,k$, the actual observed value of y is a linear function of variables $x_2, x_3, x_4, ..., x_k$. The generalization of this to a probabilistic model assumes that the expected value of y (dependent variable) is a linear function of x_i (independent variables); but for a particular set of x_i, the variable y differs from its expected value by a random amount. Mathematically,

$$y = B_1 + B_2x_2 + B_3x_3 + ... + B_kx_k + \varepsilon \qquad [3.1]$$

where ε is a random variable with $E(\varepsilon) = 0.0$ and $Var(\varepsilon) = \sigma^2$. To construct confidence and prediction intervals and to test hypotheses about the model parameters, it is also needed to assume that ε has a normal distribution.

Equation [3.1] is the most straight forward form of linear multiple regression. Sometimes the variables may have exponents over them making the equation to be of higher order. Also there may be terms formed by product of two or more variables, called interaction terms. The following equation implies a second order interaction model :

$$y = B_1+B_2x_2+B_3x_3+B_4x_2^2+B_5x_3^2+B_6x_2x_3+\varepsilon \qquad [3.2]$$

The presence of higher order terms indicate that the expected change in y depends on the change of values having higher order in such a way that the contours of regression function against those variables will be curved. And the presence of interaction terms imply that the expected change in y depends not only on the variables being increased or decreased but also on other variables forming the interaction terms.

Now, whether to include such higher order or interaction terms in the model solely depends on the nature of the problem being studied.

3.2. Estimating the Parameters

The multiple linear regression model with (k-1) variables and n observations y_1, y_2, ..., y_n has the form :

$$y_i = B_1+B_2x_{i2}+B_3x_{i3}+ \ldots\ldots +B_kx_{ik}+\varepsilon_i \qquad [3.3]$$

$$i = 1,\ 2,\ \ldots\ldots,\ n$$

For convenience of matrix notation, a dummy variable is introduced as x_{i1} associated to B_1 where $x_{i1}=1.0$ for i=1, 2,..,n. So equation [3.3] becomes :

$$y_i = B_1x_{i1}+B_2x_{i2}+ \ldots\ldots +B_kx_{ik}+\varepsilon_i \qquad [3.4]$$

Now the vector of observation Y, vector of random errors E, parameter vector B and design matrix X are defined as :

$$Y = \begin{matrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{matrix} \quad E = \begin{matrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{matrix} \quad B = \begin{matrix} B_1 \\ B_2 \\ . \\ . \\ B_k \end{matrix} \quad X = \begin{matrix} x_{11} & \ldots & x_{1k} \\ x_{21} & \ldots & x_{2k} \\ . & & . \\ . & . & . \\ x_{n1} & \ldots & x_{nk} \end{matrix}$$

Consequently, in matrix form, equation [3.4] turns out to be :

$$Y = XB + E \qquad [3.5]$$

If b be the sample estimate of the vector B using least square criteria, then, it can be shown with the help of matrix algebra that the normal equations corresponding to equation [3.5] takes the form :

$$(X^TX)b = X^TY \qquad [3.6]$$

where X^T is the transpose of the design matrix X. Multiplying both sides of equation [3.6] on the left by $(X^TX)^{-1}$, the solution matrix for sample estimates of parameters becomes :

$$b = (X^TX)^{-1}X^TY \qquad [3.7]$$

Equation [3.7] indicates that to get the vector **b**, it must be possible to invert the matrix X^TX, however, the transformation is not unconditional. If z_{ij} is defined to be $(x_{ij}-\bar{x}_j)/s_j$ and $Z=[z_{ij}]$, then $Z^TZ/(n-1)$ is the k*k correlation matrix $R=[r_{ij}]$, where r_{ij} is the correlation coefficient between the ith and the jth independent variables. By definition, $r_{ij}=1$ for $i=j$.

If $|r_{ij}|=1$ for some $i \neq j$, then the ith independent variable is a linear function of the jth independent variable and the rank of X^TX matrix will be less than k. But X^TX being a k*k matrix, its rank must be k to get it inverted. This means that an independent variable can not be a (perfect) linear function of any other independent variable. Moreover, an independent variable can not be linearly dependent on any linear function of the remaining independent variables, otherwise the rank of X^TX will drop down again. Even a near linear dependence in X may cause severe roundoff errors in $(X^TX)^{-1}$ and loss of significance leading to nonsensical estimates for B (Draper and Smith, 1981). This is why, very often the first step in regression analysis becomes the computation of correlation matrix.

3.3. Standardizing the Variables

When the values of variables in multiple regression analysis are large, it is advantageous to carry out a special coding for the variables. If $\overline{x}_j$ and s_j be the sample mean and standard deviation of x_{ij}'s ($i=1,...,n$), the coded form of x_j will be $x_j^*=(x_{ij}-\overline{x}_j)/s_j$. The coded value x_j^* simply represents any x_{ij} value in units of standard deviation above or below the mean. Careful observation easily reveals that the outcome of such transformation simply creates the Z matrix needed for correlation coefficients. So, standardizing the variables need no additional effort. But it has two important benefits :

a] it increases the numerical accuracy in all computations through less computer roundoff error.

b] it gives more accurate estimates than for the parameters of the uncoded model because the individual parameters of the coded model characterize the behavior of the regression function near the center of the data rather then near the origin.

3.4. Coefficients of Determination and the ANOVA Table

Recalling that the column vector **b** is the sample estimate of parameters and defining $x_i=(x_{i1},x_{i2},...,x_{ik})$, the ith estimate for the dependent variable becomes $\hat{y}_i=x_ib$. So, summing up for all the i's, the error sum of squares $SSE=\Sigma(y_i-\hat{y}_i)^2$ measures how much variability in the y_i's is not explained by the regression relationship. If SSE is quite small, all the observed points lie near the least square line, while if it is large, then there is much 'residual variability' even after taking into account the possibility of a linear relationship.

The total amount of variability in the y_i's can be measured by computing $SST=\Sigma(y_i-\bar{y})^2$ which is the total sum of squares of the y_i's about their mean. Hence, the coefficient of multiple regression R^2, indicating the proportion of variation in y_i's explained by linear regression is defined as :

$$R^2 = (SST - SSE)/SST$$
$$= 1 - SSE/SST \qquad [3.8]$$

The table given below, called the ANOVA (analysis of variance) table is quite helpful in calculating all the above terms and for other analyses using matrix notation.

ANOVA TABLE

Source	Degrees of freedom	Sum of squares
Mean	1	$n\bar{y}^2$
Regression	k-1	$b^TX^TY - n\bar{y}^2$
Residual	n-k	$Y^TY - b^TX^TY$
Total	n	Y^TY

Using the ANOVA table, SST, SSE and R^2 assume the following forms :

$$SST = Y^TY - n\bar{y}^2$$
$$SSE = Y^TY - b^TX^TY$$
$$R^2 = (b^TX^TY - n\bar{y}^2)/(Y^TY - n\bar{y}^2) \qquad [3.9]$$

Another important statistic, Var(ε) or σ^2 is represented by its sample estimate s^2 as :

$$s^2 = (Y^TY - b^TX^TY)/(n - k) \qquad [3.10]$$

which is also known as residual mean square.

3.5. Inferences on Regression Coefficients

To make inferences concerning **B**, the variance of **b** must be known. It is shown by Haan(1977) that the variance-covariance matrix of **b** is given as :

$$Cov(\mathbf{b}) = \sigma^2(X^TX)^{-1} \quad [3.11]$$

The variance of b_i is equal to the covariance of b_i with itself and is therefore σ^2 times the ith diagonal element of $(X^TX)^{-1}$. The covariance of b_i with b_j is σ^2 times the (i,j)th element of $(X^TX)^{-1}$. To get a confidence interval on each b_i, the underlying assumption would be that b_i/s_{bi} has a t distribution with (n-k) degrees of freedom where s_{bi} is the positive square root of covariance of b_i. Such assumption will be perfectly valid when the dependent variable is normally distributed. However, according to the Central Limit Theorem, if a hydrologic random variable is the sum of k independent effects and n stands for the number of observations, then, as n gets larger, the distribution of the variable tends to be normal. Usually, $n \geqslant 30$ gives good enough approximation. It is shown by different experimenter that n as small as 15 also works good if the underlying distribution of the dependent variable is not far from normal. Then the lower and the upper confidence intervals are given by :

$$L_{Bi} = b_i - t_{1-\alpha/2,n-k}s_{bi}$$

$$U_{Bi} = b_i + t_{1-\alpha/2,n-k}s_{bi} \quad [3.12]$$

To test the hypothesis that $H_0:B_i=0.0$ against $H_a:B_i\neq 0.0$, the test statistic is :

$$t = b_i/s_{bi} \qquad [3.13]$$

Here, H_0 is rejected if $|t| > t_{1-\alpha/2,n-k}$ which means that the ith independent variable is contributing significantly to explaining the variation in the dependent variable. Alternately, if the null hypothesis is accepted, then the corresponding independent variable is usually deleted from the model.

Conclusions resulting from the individual testing about regression coefficients may sometimes be misleading. For example, separate t tests may indicate that both B_i and B_j are statistically insignificant. This does not mean that both B_i and B_j should be eliminated from the model as B_j belonged to the model when B_i was tested and vice versa. This situation is likely to occur when the sample values of corresponding independent variables are highly correlated. However, B_i or B_j when used along may be quite significant.

In many circumstances, firstly the full size model involving the k carriers (including the $x_{i1}=1.0$) is made. Then further investigation is done to check whether a particular subset of l carriers provides almost as good a fit as the full k-carrier model. To serve this purpose, the required test statistic has an F distribution as :

$$F = \frac{(SSE_l - SSE_K)/(k - l)}{SSE_K/(n - k)} \qquad [3.14]$$

SSE_l = unexplained variation for the reduced model
SSE_K = unexplained variation for the full model

Rejection region : $F > F_{1-\alpha, k-1, n-k}$

A more detailed discussion about the selection of best subset will be made later on in section 3.11 of this chapter.

3.6. Test of Model Utility

Extending the above discussion to the extreme that the entire regression equation is not explaining a significant amount of the variation of the dependent variable, the null hypothesis will be $H_0: B_1 = B_2 = \ldots\ldots = B_k = 0.0$ versus H_a: at least one of these B's is not zero. Here use is made of the fact that the ratio of the mean square due to regression to the residual mean square has an F distribution with (k-1) and (n-k) degrees of freedom. The F statistic in matrix form may be given as :

$$F = \frac{(b^T X^T Y - n\bar{y}^2)/(k-1)}{(Y^T Y - b^T X^T Y)/(n-k)} \qquad [3.15]$$

Rejection region : $F > F_{1-\alpha, k-1, n-k}$

3.7. Inferences on the Regression Line and Individual Prediction

It is shown by Draper and Smith(1981) that the variance of the ith estimate of the dependent variable from the regression equation ($\hat{y}_i = x_i b$) can be given by :

$$\mathrm{Var}(\hat{y}_i) = \hat{\sigma}^2 x_i (X^T X)^{-1} x_i^T \qquad [3.16]$$

So the confidence intervals (CI) on $\hat{y}_i$ are expressed as :

$$L = x_i b - t_{1- /2,n-k} \; /(Var(\hat{y}_i))$$

$$U = x_i b + t_{1- /2,n-k} \; /(Var(\hat{y}_i)) \qquad [3.17]$$

The CIs on individual predicted value $\hat{y}_{pi}$ are also given by equation [3.17] with the following change :

$$Var(\hat{y}_{pi}) = \sigma^2(1 + x_i(X^TX)^{-1}x_i^T \qquad [3.18]$$

3.8. The Bonferroni Intervals

In some situations, CIs may be desired for two or more set of x_i's. For example if two such intervals are calculated at 95% level, then joint confidence coefficient would be (.95)*(.95)=.90 assuming intervals are independent to each other. But intervals are actually not independent, because same b and s^2 are used in each. The treatment of such joint confidence intervals rests on a mathematical result called the 'Bonferroni Inequality' and so, the joint CIs are often called the Bonferroni intervals. In general, if the $100(1-\alpha)$% confidence interval is computed for m different sets of x_i's, then the joint confidence coefficient on the resulting set of interval is at least $100(1-m\alpha)$% .

3.9. Additional Analyses

3.9.1. Identifying the Outliners by HAT Matrix Elements

In simple linear regression, diagonistic plots can be used to identify both points of large residuals and wild points well off from most of the sample x_i's. In multiple regression, due to presence of two or more independent variables, such plottings are virtually impossible. Hence, a new tool has been proposed, called the 'HAT' matrix which is defined as (Devore,1982) :

$$H = X(X^T X)^{-1} X^T \quad [3.19]$$

Obviously, the ith fitted value y_i is the product of the ith row of H with Y :

$$\hat{y}_i = h_{i1}y_1 + h_{i2}y_2 + \ldots\ldots\ldots \; h_{in}y_n \quad [3.20]$$

Thus the element h_{ij} gives the weight associated with jth observation in computing the ith predicted value. In particular, h_{ii} measures the influence of y_i on its own predicted value $\hat{y}_i$. It is therefore of great interest to know whether a particular h_{ii} is relatively large or small. Large h_{ii} will indicate y_i with large influence on the overall fit and such points may be excluded from the final model preparation. The rule of thumb is that any i for which $h_{ii} >$ 2k/n indicates that y_i is a point with large influence.

Another means for deciding whether or not the ith point has large influence is to consider the changes in parameter estimates when the ith data point is deleted from the sample. However, for this study, the HAT matrix approach has been resorted to.

3.9.2. Aptness of the Model

An effective approach to assessment of model adequacy is to compute the fitted or predicted values $\hat{y}_i$ and the residuals $e_i = y_i - \hat{y}_i$, and then plot various functions of these computed quantities. It can be shown that each residual is normally distributed with expected value of zero and when n gets reasonably larger, the standard deviation may be simply given by s - the standard error of estimate. Now, defining the standardized residual as $e_i^* = e_i/s$, a plot of e_i^* versus $\hat{y}_i$ is always recommended for multiple regression analysis. If the model is acceptable, then this residual plot should not exhibit any distinct pattern. Also the residuals should be randomly distributed about zero according to a normal distribution, so all but a very few e_i^*'s should lie between -2 and +2 (that is within two standard deviations of their expected value zero). Detailed discussion about identifying the difficulties and probable remedies are made by Daniel and Wood(1980) and Chatterjee and Price(1977).

In the previous section, discussion is made about using HAT matrix to identify the outliners. If such outliners result from errors in recording data values or faulty experiment, they are omitted from the final model. But if no assignable cause can be found for the outliners, it is still desirable to report the estimated equations both with and without outliners. An alternative procedure is to keep the outliners in the model but to put relatively less weight on them. One such method is MAD (minimize absolute deviation) and here values of parameters are to be found by iterative computational procedure. More information about alternative fitting techniques can be found in Mosteller and Tukey(1977).

3.10. Selection of the 'Best Subset'

Often an experimenter will have large number of independent variables or carriers and then wish to build a regression model involving a subset of those carriers. The use of the subset will make the resulting model more manageable, especially if more data is to be subsequently collected. This also provides a model which is easier to interpret than one with many more carriers (Devore,1982). Two basic questions in connection with the variable selection are :

i] If it is possible to examine all possible subsets of independent variables, which criteria should be used to select a model?

ii] If the number of variables is too large to check all possible combinations, what alternative techniques are available?

To answer such questions, statisticians developed both criteria for variable selection and alternative ways of getting a sufficiently good subset (it may also be the best) as discussed below.

3.10.1. Criterias for Variable Selection

As before, SSE_k will be used to indicate the error sum of squares with k carriers (including the $x_{i1}=1.0$ terms). For a fixed value of k, it is reasonable to identify the best model as the one having minimum SSE_k. The more difficult issue concerns comparison of SSE_k's for different values of k. Two different criteria, each one a simple function of SSE_k, are widely in use.

i] R_k^2, the coefficient of multiple determination for a k carrier model. Because R_k^2 will virtually always increase as k does (and can never decrease), it is not the k which maximizes the R_k^2 is of interest. Instead, a small k is needed to be identified for which R_k^2 is nearly as large as R^2 for all carriers in the model.

ii] The standard error of estimate s_k for k carrier model. The confidence intervals on the regression line are function of s_k, the line with the smallest standard error will have the narrowest confidence intervals and hence, will represent the most dependable prediction.

In this model study, both the criteria will be used to check the model performance. Many times the two criteria of the near-largest R_k^2 and smallest s_k give the same subset of carriers. Generally, with the increased number of variables, R_k^2 will always increase and s_k will hopefully decrease. But after some specific value of k, s_k or the standard error of estimate will tend to increase. This is a tip-off that added variables are not contributing significantly to the regression and can just as well be left out.

3.10.2. Techniques of Variable Selection

Three different methods are commonly in practice, namely, the Forward Selection (FS), the Backward Selection (BS) and the Stepwise Selection (SS).

3.10.2.1. Forward Selection (FS)

FS starts with no carriers in the model and considers fitting in turn the model with only x_2 (x_1 being always present as the constant 1), only $x_3, \ldots,$ and finally only x_m. The variables which, when fit, yield the largest absolute t ratio (which is $|b_j/s_{bj}|$) enters the model provided that the ratio exceeds the specified constant t_{in}. The process continues until at some step no absolute t ratio exceeds t_{in}. At 95% level of confidence, most t values are near 2 and so, $t_{in}=2.0$ is often used in FS technique.

3.10.2.2. Backward Selection (BS)

This method starts with the model in which all carriers under consideration are present. Let the set of all carriers be x_2, $x_3,..,x_k$. Then each absolute t ratio is examined and the smallest one is detected. If the smallest absolute t ratio is less than a specified constant t_{out}, then the corresponding carrier is eliminated from the model. The process is continued again for the reduced model until at some stage, all absolute ratios are at least equal to t_{out}. The model used is the one containing carriers which were not eliminated. For the same reason as mentioned above, t_{out} is usually taken to be 2.0.

3.10.2.3. Stepwise Selection (SS)

The stepwise procedure most widely used is a combination of FS and BS, denoted as SS. This procedure starts off as does the FS, by adding variables to the model; but after each addition examines those variables previously entered to see if any is a candidate for elimination. For example, if there are eight carriers under consideration and current set consists of x_2, x_3, x_5 and x_6 with x_6 having just been added, the t ratios t_2, t_3 and t_5 are examined. If the smallest absolute ratio is less than t_{out}, then the corresponding variable is eliminated from the model. The idea behind SS is that a variable may individually contribute little towards the increment of R^2 or decrement of s when other variable(s) with which it has got strong correlation is already present in the model. Such variables can be identified easily from the correlation matrix (CM) of the model being studied and this is another strong point in favor of working out the CM in the first place.

For SS process, to prevent the same variable from being repeatedly entered and removed, it is essential that $t_{in}>t_{out}$. For this study, $t_{in}=2.0$ and $t_{out}=1.975$ will be used as done in most of the standard packages available for stepwise regression (Devore, 1982). Currently, a number of efficient computer packages are available to take care of this very elaborate process of trial and error. A discussion on the packages will be made in the next chapter.

3.10.3. Some Final Comments on the Selection of Variables

The three automatic selection procedures FS, BS and SS will generally identify a very good model. But there is no guarantee that the model will be the best which could have been resulted from all possible combination of carriers. Above all, no matter which technique is used, care must be exercised to see that the resulting equation is rational.

In general, all the variables retained in a regression equation should make a significant contribution to the regression unless there is an overriding reason (theoretical or intuitive) for retaining a non-significant variable. The variables retained should have physical meaning. If two variables are equally significant when used alone but are not both needed, the one that is easiest to obtain should be used (Haan,1977). Finally, if there appears to be strong relationship between some of the potential carriers in a given data set, alternative method, say, the 'Ridge Regression' technique should be employed.

Chapter 4

MODEL DEVELOPMENT, DATA PROCESSING AND SOFTWARE REQUIREMENTS

4.0. Introduction

It was stated in Chapter 1 that, the basic tasks of this study were to develop a linear multiple regression model that could be used to forecast future extreme groundwater levels based on monthly total rainfall data and to study the effect of river stage on groundwater level at different distances from the bank line using simple linear regression technique.

In doing so, first of all, a brief description of the study area (where the developed model will be applied) will be given in the next section. Thereafter, each of the steps necessary to attain the above mentioned objectives will be discussed in detail.

In this regard, the first step will be to develop the conceptual models for both rainfall-groundwater level and river stage-groundwater level relationships. At this stage, model variables will be defined with due care, which will eventually determine the types and amount of data required for running the model, subsequent analyses and testing the forecasting potential when applicable.

The second step, then, will be to collect and process all the data and to point out their salient features that will be helpful in preparation and interpretation of model input and output. And the third and final step will be selection and/or development of appropriate softwares which will be able to carry on all the necessary operations.

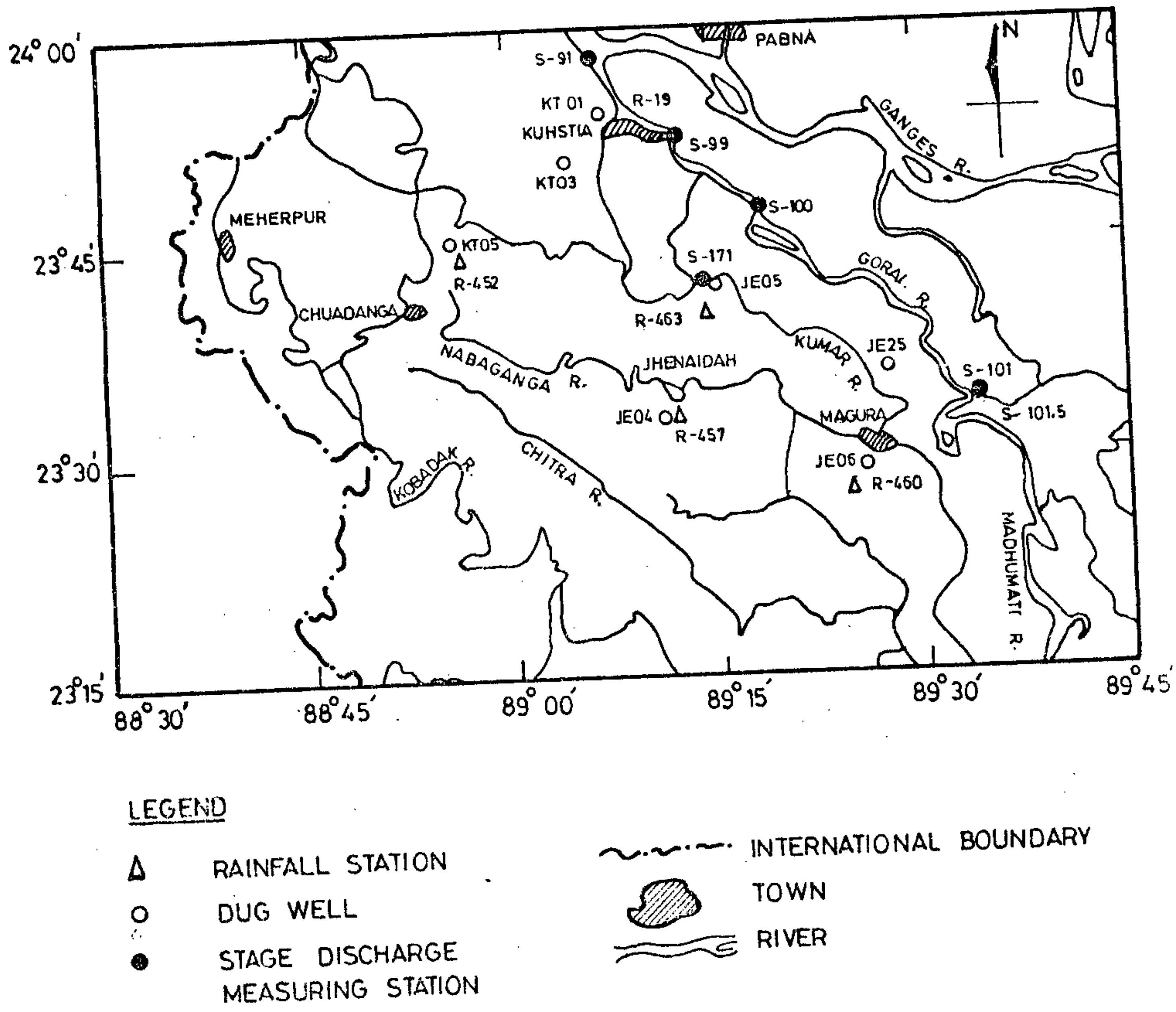
4.1 Brief Description of the Study Area

The Ganges-Kobadak (G-K) Irrigation Project covers areas from the districts of Kushtia, Jessore and Khulna. The area under present study lies between latitudes of 23°N to 24°N and longitudes of 88°E to 90°E. Major streams in and around the study area are the Ganges, the Gorai, the Kumar and the Nabaganga. The mean annual rainfall in the area is about 1550 mm and the mean evapotranspiration rate is about 1350 mm. Geologic formations of the upper layers of the area are composed of silt, sand and clay. Minimum annual recharge in the project are is estimated to be about 0.02 m. The average transmissivity and specific yield of the area are approximately 2000 m^2/day and 0.10 respectively (IECO,1980).

A map of this study area depicting all the salient features is given in the next page (Fig. 4.1). This figure shows all the locations of the wells under study along with the associated rainfall, river stage and discharge measuring stations.

FIG. 4.1 : THE STUDY AREA UNDER G-K PROJECT

(SCALE 1:750,000)



4.2. Development of the Linear Multiple Regression Model

This study is supposed to formulate a model showing the linear relationship between total monthly rainfalls and annual culminations of water table in an undisturbed groundwater regime. In reality, the phenomenon of water table fluctuation is affected by numerous factors other than rainfall. Because, the extent of rainfall which ultimately reaches the groundwater table depends on the land cover, slope of the land surface, soil moisture content, depth of root zone of the existing crop or plantation, rate of evapotranspiration, soil and air temperature, humidity, wind speed etc., and the list is not exhaustive. But from practical point of view, it is not feasible to include all the variables into a hypothetical relationship. Firstly because, many of them will contribute too little; secondly, adequate and reliable data are seldom available for all the affecting variables. Moreover, regression analysis requires that number of independent variables should be less than the number of samples or data sets; preferably, less than one third of the later. So, only the most important factors are usually included in a typical model study.

As stated earlier, the study area of G-K project has got an undisturbed groundwater regime ideal for regression analysis. The oldest wells in the project area were installed in 1961-62 period. Associated rainfall stations were also installed in the same time. But there was a discontinuity of records in the period of 1964-65. As the publications by BWDB documented groundwater levels from the year of 1963-64, it was not possible even to get an initial estimate of the missing values of 1964-65 by Forgo's method (Forgo,1968) which requires data of at least four previous years. So, data from 1966 and onwards were used for this study. This means that groundwater level data for different wells had the longest record length of about 20 years. So, to simulate extreme groundwater levels as dependent variables, number of independent variables should ideally be kept to 7 or 8. This is why, only the most significant contributing variables - the total monthly rainfalls were taken as independent variables in the proposed regression model. Breaking up the total annual rainfall into its monthly components greatly increases the flexibility of the model and also helps grasp the role of individual months in determining the annual culminations. Splitting the rainfalls into further smaller intervals will cause too many variables and it was revealed by preliminary model runs that, adding several month's rainfall together to reduce the number of variables also greatly reduce the prediction capability of the model. Hence, monthly total rainfalls as independent variables came out to be the best choice.

4.2.1. Variables for HWL Simulation

Apparently it may seem alright to think that the monthly rainfalls after the occurrence of LWL will contribute to the rise of water table. But due to the very slow rate of seepage through subsurface media, it is quite possible that some portion of rainfall occurring before the LWL may eventually enter the groundwater table after recording of LWL. This will definitely contribute to the water table rise and will not be encountered by the truncated series of monthly rainfalls (after LWL). So it is a better choice to take all the monthly rainfalls from previous HWL as independent variables to simulate the next HWL as dependent variable.

It will be shown later in this chapter that, in G-K project area, the groundwater level on average becomes maximum in the month of September and minimum in the month of May. So, monthly total rainfalls from September to August will be considered as independent variables to simulate the next HWL. Mathematically (equation 3.4) :

$$y_i = \sum B_j x_{ij} + \varepsilon_i$$
$$i = 1, 2, \ldots, n$$
$$j = 1, 2, \ldots, k$$

where y_i is the ith HWL; $x_{i,2}$ to $x_{i,13}$ are the total monthly rainfalls of September to August of the ith year; B_2 to B_{13} are the associated weight or contribution factors; B_1 is the constant or intercept term which takes care of the fixed component of the dependent variables and $x_{i1}=1$ for all i's, used for convenience of matrix notation. The upper limit of j, i.e., k will be equal to the number of dependent variables plus one. The upper limit of i denoted as 'n' is the number of samples or data sets available or being used for the model. Finally, ε_i is the ith random error term having expected value of zero and a constant variance for all observations as specified in section 3.1 of Chapter 3.

The LWL may also be included to simulate the next HWL. However, such inclusion is only encouraged when significant increase in coefficient of determination is noted. Groundwater levels at locations close to a nearby river will also be affected by river stage. But none of the six wells except JE05 are that close to the river Gorai or Kumar in the G-K area (Fig. 4.1). And, river stage of the Kumar will not be included in the simulations of HWL and LWL for JE05 due to the restriction on allowable number of carriers in the model.

Hence, initial number of independent variables (K-1) becomes 13 or 12 (with or without LWL) which is greater than one third of the number of samples (20 in our case). It will be shown later that, a number of months contribute too little into the model and therefore, may be dropped out. Consequently, with dropping out of insignificant variables and accumulation of more data, the model will attain the required stability.

4.2.2. Variables for LWL Simulation

Selection of independent variables for LWL simulation is rather straight forward. As the LWL occurs mostly on May, monthly rainfalls from September to April plus the previous HWL will make the nine independent variables. Hence, the starting point of the model formation will be (equation 3.4):

$$y_i = \sum B_j x_{ij} + \varepsilon_i$$
$$i = 1,2,\ldots,n$$
$$j = 1,2,\ldots,10$$

where y_i is the ith LWL; $x_{i,2}$ is the previous HWL; $x_{i,3}$ to $x_{i,10}$ are monthly total rainfalls of September to April; B_2 to B_{10} are associated contribution factors, B_1 is the intercept term; $x_{i1}=1$ for all i's; n is the number of data sets and upper limit of j, i.e., k is fixed as 10. Again, ε_i is the ith random error term.

This time, after dropping out insignificant variables, ultimate number of independent variables will be quite appropriate as demanded by multiple regression methodology.

4.2.3. Forecasting and the Frequency Analysis

One of the goals of multiple regression model development in this study is to use it for forecasting purposes. Once the model passes the goodness of fit test with significantly large coefficient of determination and relatively small standard error of estimate, it can be used for forecasting future extreme water levels based on 'design rainfalls' of previous contributing months. Here the term 'design rainfall' means the total monthly rainfall of a month corresponding to some predefined return period. Magnitude of such rainfall may be found from the probability plots based on an appropriate probability distribution for the rainfall data. So, frequency analysis of total monthly rainfall data is prerequisite to forecasting and will be done later in this chapter.

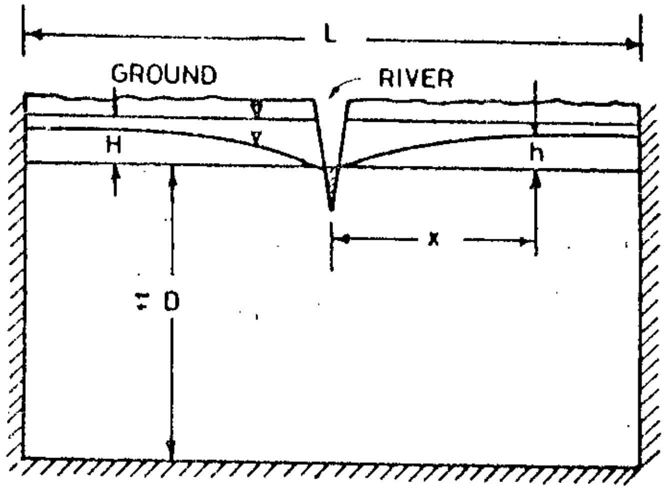
4.3. Study of Stream-Aquifer Interaction

Before starting with the nature of study to be carried on, it will be helpful to have a brief discussion about the simplest type of stream-aquifer interaction.

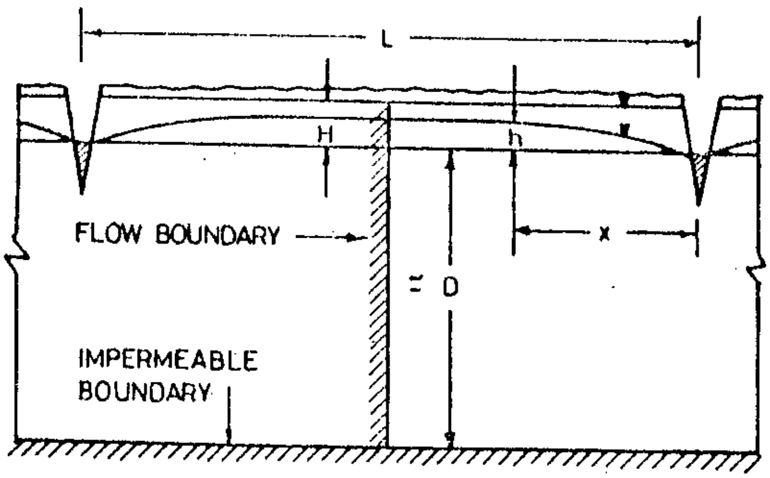
Pioneering work in this field was done by Glover (1974). For classical problem of river valley drainage or drainage by parallel streams as shown in Fig. 4.2, he assumed that:

i] Dupuit Forchheimer idealization is valid, i.e. horizontal gradient dh/dx is approximately equal to surface gradient dh/ds and it applies to the entire depth of the aquifer.

ii] The saturated thickness remains the same and may be approximated by the depth of impermeable layer from the drain or stream water level (physically it means that the flow over the drain level is negligible).



a. river valley drainage



b. parallel drainage

FIG. 4.4 STREAM-AQUIFER INTERACTION

Using continuity principle, the flow equation for transient state then becomes :

$$\frac{\partial^2 h}{\partial x^2} = \frac{S}{KD}\frac{\partial h}{\partial t} \qquad [4.1]$$

where,

K = permeability of the aquifer
D = thickness of the aquifer contributing to the stream as shown in Fig. 4.2
S = storage coefficient
x = distance measured along horizontal direction or along path of flow
t = time
h = height of water table above the drain level

Solution of equation 4.1 subject to the conditions

h = H for $0<x<L$ when $t = 0$
h = 0 for $x=0$ when $t > 0$
h = 0 for $x=L$ when $t > 0$

is :

$$h = H\frac{4}{\pi}\sum_{n=1,3,5}^{\infty}\frac{e^{-m}}{n}\,\mathrm{SIN}(n\pi x/L) \qquad [4.2]$$

where $m = n^2\pi^2\alpha t/L^2$, $\alpha = KD/S$
and L = width of the river valley or distance between the parallel drains

Now, water table profile on a vertical plane through the stream-aquifer system can be found by using equation [4.2] for any arbitrary time t. The profile will be similar to those shown in Fig.4.2. It is evident from this figure that, the smaller the distance x, the greater is the drawdown or lowering of water table. Now, if the stream level fluctuates, the water table will also try to react; however, sensitivity of the water table to fluctuating stream level will be declining with increasing distance from the stream. Accordingly, objectives of this research work related to stream-aquifer system are :

i] to show that groundwater level will be less sensitive to the fluctuation of river stage with increasing distance from the bank-line.

ii] to roughly define a characteristic distance d_c beyond which water table will remain virtually insensitive to the river stage fluctuation for the basin under study.

Simple linear regression will be used to attain both the objectives. Firstly, river stages at different locations will be correlated with corresponding groundwater levels at different distances from the bank. Secondly, the correlation coefficients so obtained will be plotted against the associated distances at which the groundwater levels will be measured.

For the 1st case, following simple linear regression will do:

$$y_i = b_1+b_2 s_i+\varepsilon_i \qquad [4.3]$$

where,

y_i = ith groundwater level at a location

b_1, b_2 = constants

s_i = ith value of river stage

To eliminate the possible effect of rainfall, river stage of 31st December and groundwater level of the next week will be used as independent and dependent variables respectively. River flow data should also be collected on the same date to confirm that recorded stages do not merely represent stagnant water level but correspond to the level of flowing stream augmented by the aquifer.

The nature of the plot of correlation coefficient R versus distance x is unknown. But, it may be expected that the value of R will decrease rapidly with increasing distance from the bank. As R = 0.75 is generally the lowest accepted value which is supposed to indicate moderately linear relationship, this value of R will be used to get the a characteristic distance d_c, as defined earlier.

4.4. Data Processing

From the elaborate discussions about model development, it is evident that four different categories of data were necessary to carry on the proposed study; namely :

i) groundwater level

ii) total monthly rainfall

iii) river stage

iv) river discharge

The source of groundwater level data were BWDB, the Institute of Flood Control and Drainage Research (IFCDR) and Bangladesh University of Engineering and Technology (BUET). The rainfall data were gathered from IFCDR, MPO and Bangladesh Meteorological Department (BMD), Dhaka. Finally, the river stage and discharge data were collected from both IFCDR and BWDB. The sections to follow cover each of the types in detail.

4.4.1. Groundwater Level

4.4.1.1. Type

Weekly depths of water table from the fixed measuring points were collected for a period of about 20 years for six selected wells in the G-K project area. Table 4.1 shows details of these wells. All the wells were dugwells and only these six selected wells had long enough records necessary for the development of a multiple regression forecasting model. However, a seventh well (JE25) of relatively recent installation was also used for the purpose of studying stream-aquifer interaction in the dry season along with the other wells.

Three different types of data were extracted from the collected weekly groundwater levels:

a) MML - mean monthly water level

b) HWL/LWL - highest/lowest water level (annual)

c) Water level of the 1st week of January of each year.

TABLE 4.1

LOCATION OF WELLS

NO.	WELL NO.	LOCATION	LATITUDE	LONGITUDE	INSTALLATION DATE
1	JE04	JHENAIDAH	$23^{0}32'30''$	$89^{0}10'45''$	JAN,1961
2	JE05	GORAGANJ	$23^{0}41'35''$	$89^{0}15'00''$	JAN,1961
3	JE06	MAGURA	$23^{0}29'05''$	$89^{0}25'20''$	JAN,1961
4	JE25	NOHATA	$23^{0}33'45''$	$89^{0}27'30''$	JAN,1977
5	KT01	KUSHTIA	$23^{0}45'00''$	$89^{0}07'00''$	FEB,1961
6	KT03	ELANGI, KUMARKHALI	$23^{0}51'00''$	$89^{0}04'00''$	SEPT,1962
7	KT05	ALAMDANGA	$23^{0}45'00''$	$88^{0}56'00''$	JAN,1961

(ALL WELLS ARE DUGWELLS)

Type (a) was used to study the typical yearly hydrograph as shown in Fig. 4.3. It was found that, the mean monthly depths of water table in all the wells become minimum in the month of September and maximum in the month of May. Further checking with type (b) revealed that almost all the HWLs and LWls also occurred in the same months. Hence, months of September and May were treated to be the months of annual culminations of water table.

Type (b) was used as model inputs. It was also used to check the goodness of fit of the predicted water level values.

Type (c), as stated earlier, was used to study the degree of influence of nearby stream on the contributing aquifer, obviously in the period of no rainfall.

4.4.1.2 Data Preparation

All the groundwater levels were converted into reduced levels (RL) in meter with respect to the mean sea level (MSL). To convert data from PWD to MSL, the following relationship was used (IECO,1964):

$$RL(MSL)_m = RL(PWD)_m + 1.507 \qquad [4.4]$$

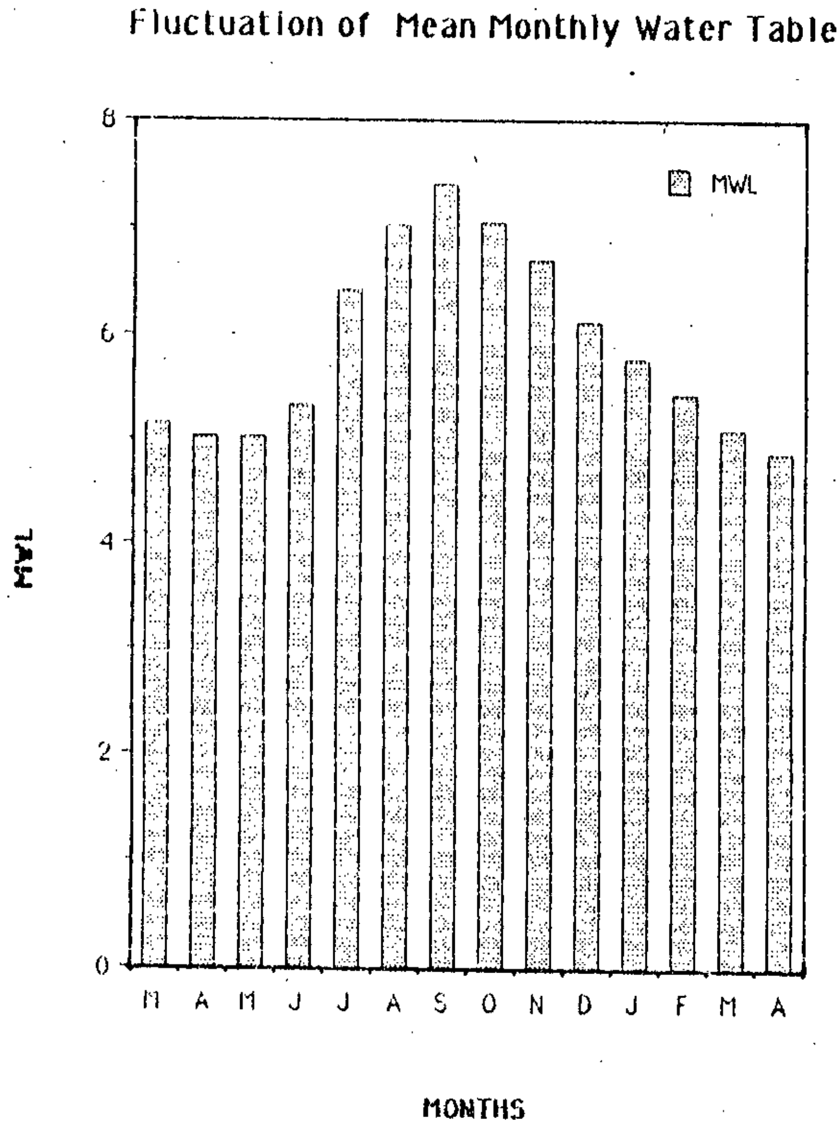


FIG. 4.3 : TYPICAL PLOT OF FLUCTUATION OF MEAN MONTHLY WATER TABLE

4.4.1.3. Additional Features

a) Missing/erratic water levels: It was noted during the preparation of model input that, certain extreme water levels were either missing or too erratic in comparison to the rest of the data. As groundwater levels have some sort of memory of their own due to slow rate of subsurface flow, deletion of such observations would cause discontinuity in a series where each value was dependent on the previous one. To overcome this difficulty, Forgo's method (Forgo,1968) of missing value estimation was used as a preliminary measure. This method is based on the trends of extremes of previous years. Fig. 4.4 illustrates this quite simple but reasonable procedure.

To get the LWL of 1971, firstly four previous differences of HWL and LWL were calculated. For example, HWL of 1966 and LWL of 1967 gave the first such difference. The process was repeated upto the difference of HWL of 1969 and LWL of 1970. Then, average of these four differences was substracted from the HWL of 1970 to get the LWL of 1971.

Although the method usually gives the first approximation to the desired extreme, it has one major drawback. It does not consider the potential factors affecting the extreme water levels, for example, rainfall pattern of previous water year. So, in this study, an iterative approach was used to improve the approximation. To start with, Forgo's estimates for

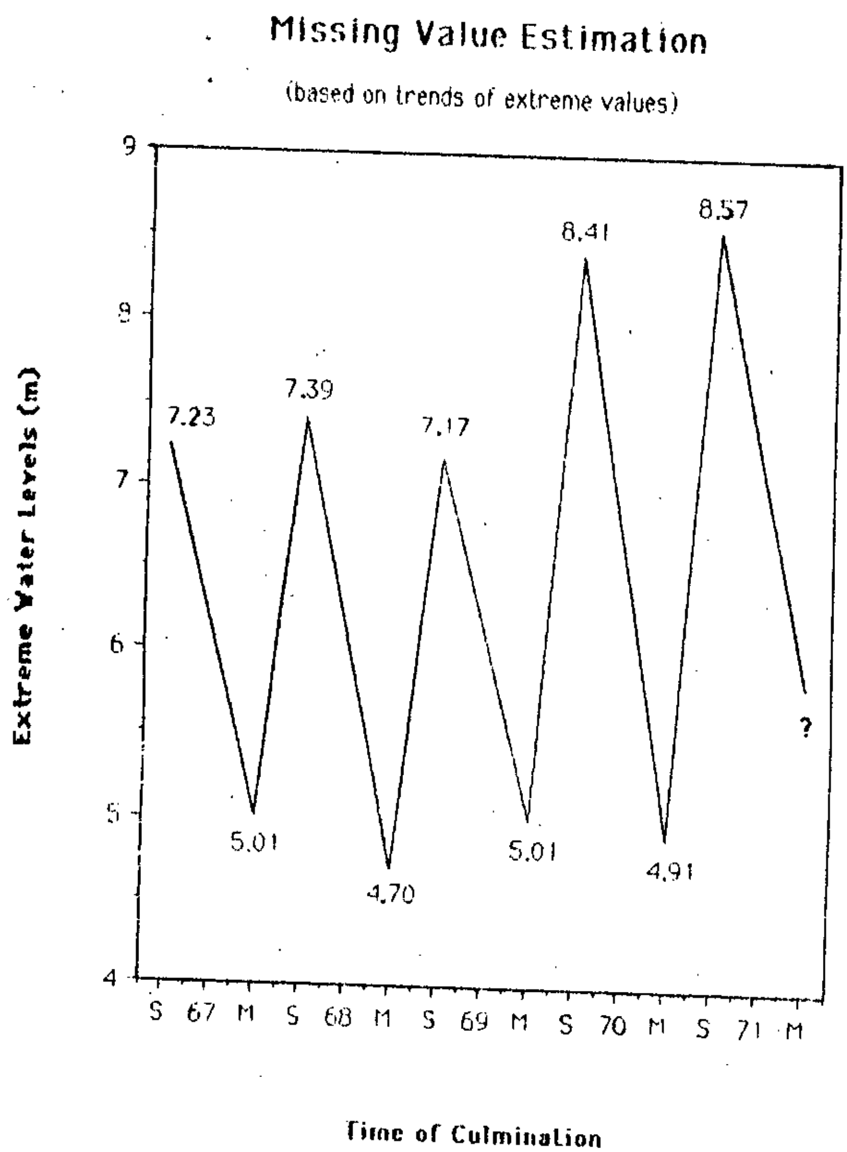
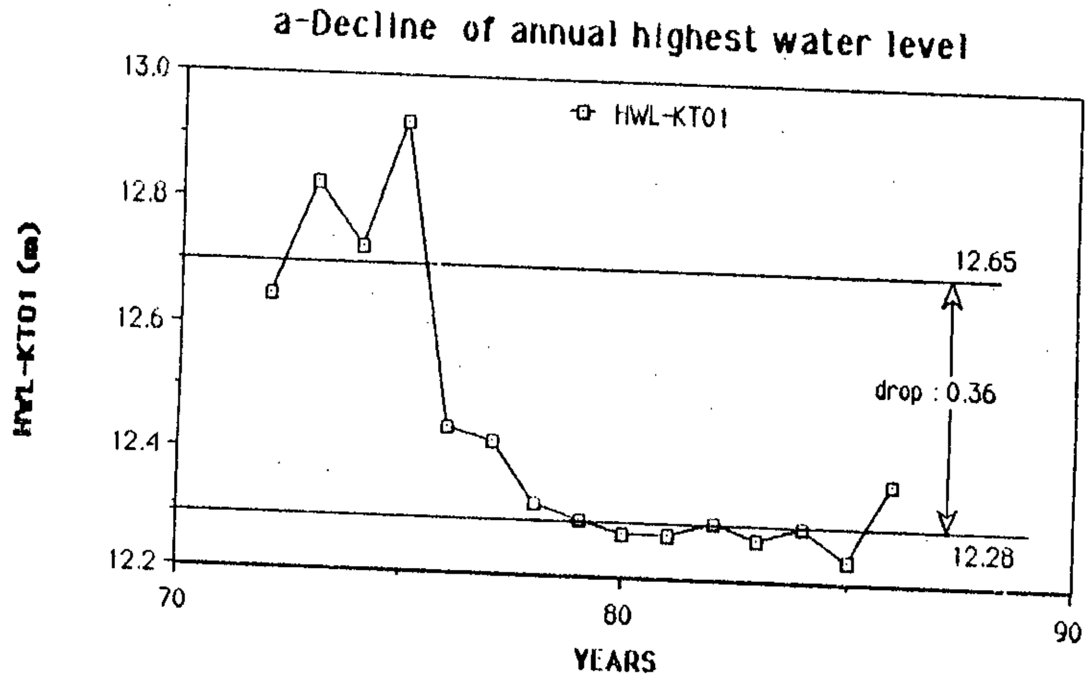


FIG. 4.4 : FORGO'S METHOD OF MISSING VALUE ESTIMATION

missing/erratic values were used in the regression model to get the 1st set of monthly contribution factors. Using these factors, total effective rainfall of previous water year (from September to August for HWL simulation) was calculated and compared with the same for other years. Such comparison indicates whether the Forgo's estimates should be increased or decreased. The process continued until change in R^2 due to latest modifications became negligible. It may be mentioned here that, out of about 240 extreme groundwater levels used in the model study, 7 were found to be erratic and 14 others missing.

b) Change in trends of HWL/LWL: It was noted that the wells KT01 and JE05 showed distinct sign of decline of annual HWL and LWL respectively (Fig.4.5). In well no. KT01, the average yearly HWL before 1975 was about 12.65 m, while the same from 1978 to 1985 became 12.288 m, indicating a permanent lowering of 0.36 m. However, as no other wells in the vicinity showed such sign of permanent lowering of HWL, this particular drop in KT01 should be treated with care. So far, no definite cause could be identified for this lowering.

In well no. JE05, sign of lowering is much more distinct. Here, although the HWL remained the same over the years, LWL declined quite sharply. From 1982 to 1983, the drop was about 4.068 m. Mean LWL before 1983 was found to be 5.86 m while the same 1983 to 1986 came out as 2.976 m, hence, the average drop



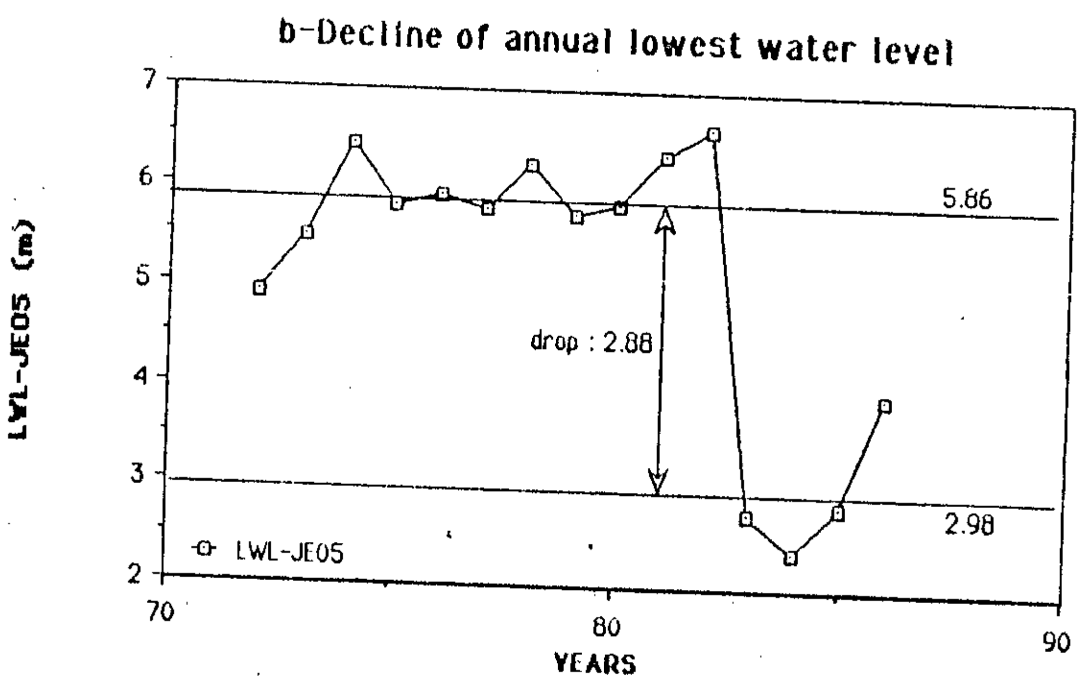


FIG. 4.5 : DECLINE OF ANNUAL EXTREME WATER LEVELS

becomes 2.884 m. The most probable cause of such significant lowering could be sudden and extensive withdrawal of groundwater. Although ground water extraction is reportedly not permitted in the command area of G-K project, a number of shallow tubewells (STW) which were existing earlier are still under operation (Michael, 1986). It is possible that, since 1983, extraction from such STWs near JE05 has been increased greatly for irrigation and domestic purposes. Even some new STWs could have been installed which were not reported duly to the authority.

An interesting feature was shown by the comparative plot of LWL and corresponding river stage of the nearby river Kumar (Fig. 4.6). It was noted that trends of LWL and river stage were quite similar upto 1982. But strangely in 1983, LWL suffered a huge drop and even went below the river stage. It remained so thereafter and never managed to come back to the original trend. This unnatural behavior of LWL in dry season strongly indicates the possibility of artificial interference to the aquifer. In other words, possibility of significant groundwater withdrawal cannot be ruled out.

4.4.2. Total Monthly Rainfall (TMR)

4.4.2.1. Station Selection

Six rainfall stations were selected carefully so that they remain close to the six selected wells of long duration record. This matching was necessary to avoid adjustments in

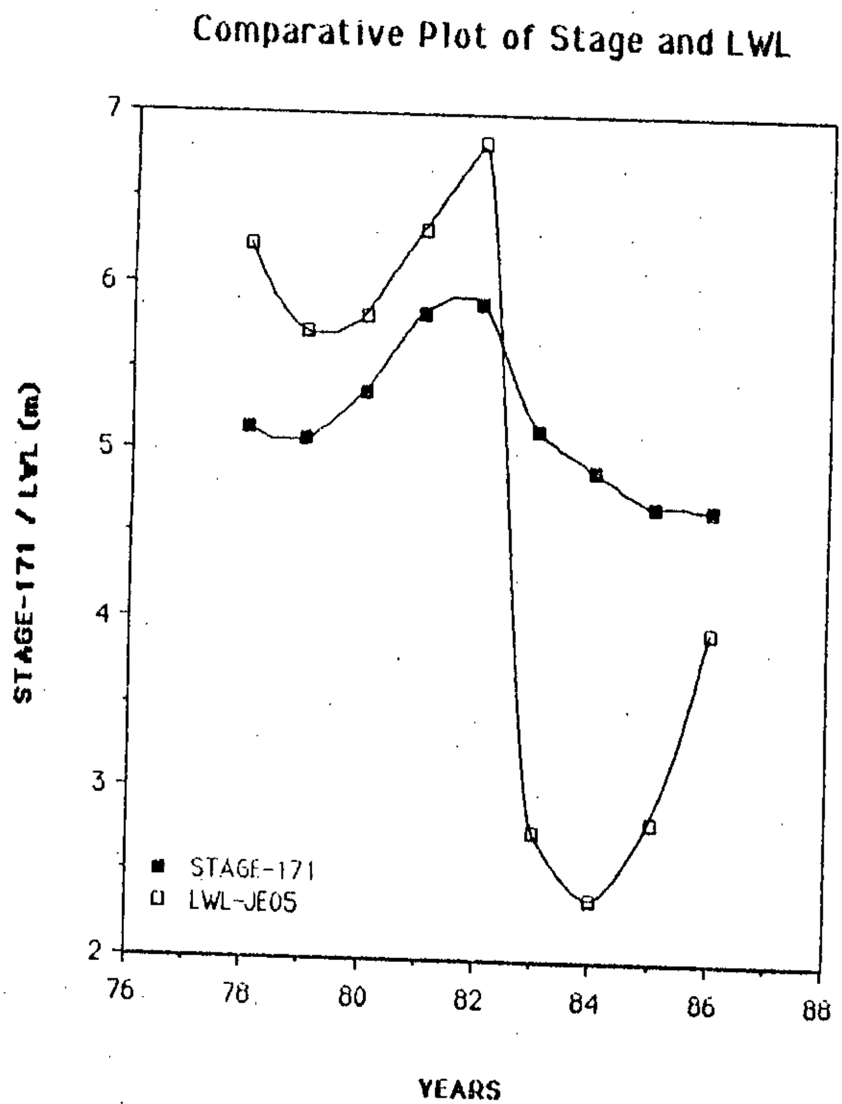


FIG. 4.6 : DECLINE OF RIVER STAGE AND LWL AFTER 1983

rainfall data used as direct input into the model. List of the selected rainfall stations along with the matching wells is presented in Table 4.2 and shown in Fig. 4.1.

4.4.2.2. Data Type

Data were collected on daily basis. But the processed data sheet also provided 10-day average, mean and total monthly rainfalls etc. Fig. 4.7 shows a typical plot of monthly total rainfalls over the years 1982 to 1985 for the station R-19. The plot shows clearly that, on average, the water year can be partitioned into two distinct periods - wet and dry. The wet period comprises the months of May to October and the dry period - the months of November to April. It is also noticeable that most of the rainfall in the wet season occurs during the months of June, July and August. And, the driest months over the year are November to February. These information were used later during the analyses of model outputs.

Throughout the study, rainfall values were reported in centimeter (cm).

4.4.2.3. Frequency Analysis

Frequency analysis was done for the rainfall stations R-463, R-460 and R-452 (given in Appendix-A), the corresponding matching wells are JE05, JE06 and KT05 respectively. For these wells, the proposed multiple regression model explained (as

TABLE 4.2

RAINFALL STATIONS

NO.	STATION NO.	LOCATION	MATCHING WELL	INSTALLATION
1	R-457	JHENAIDAH	JE04	13/03/61
2	R-463	SAILAKUPA	JE05	30/06/62
3	R-460	MAGURA	JE06	01/12/60
4	R- 19	KUSHTIA	KT01	01/01/61
5	R- 19	KUSHTIA	KT03	01/01/61
6	R-452	ALAMDANGA	KT05	18/09/62

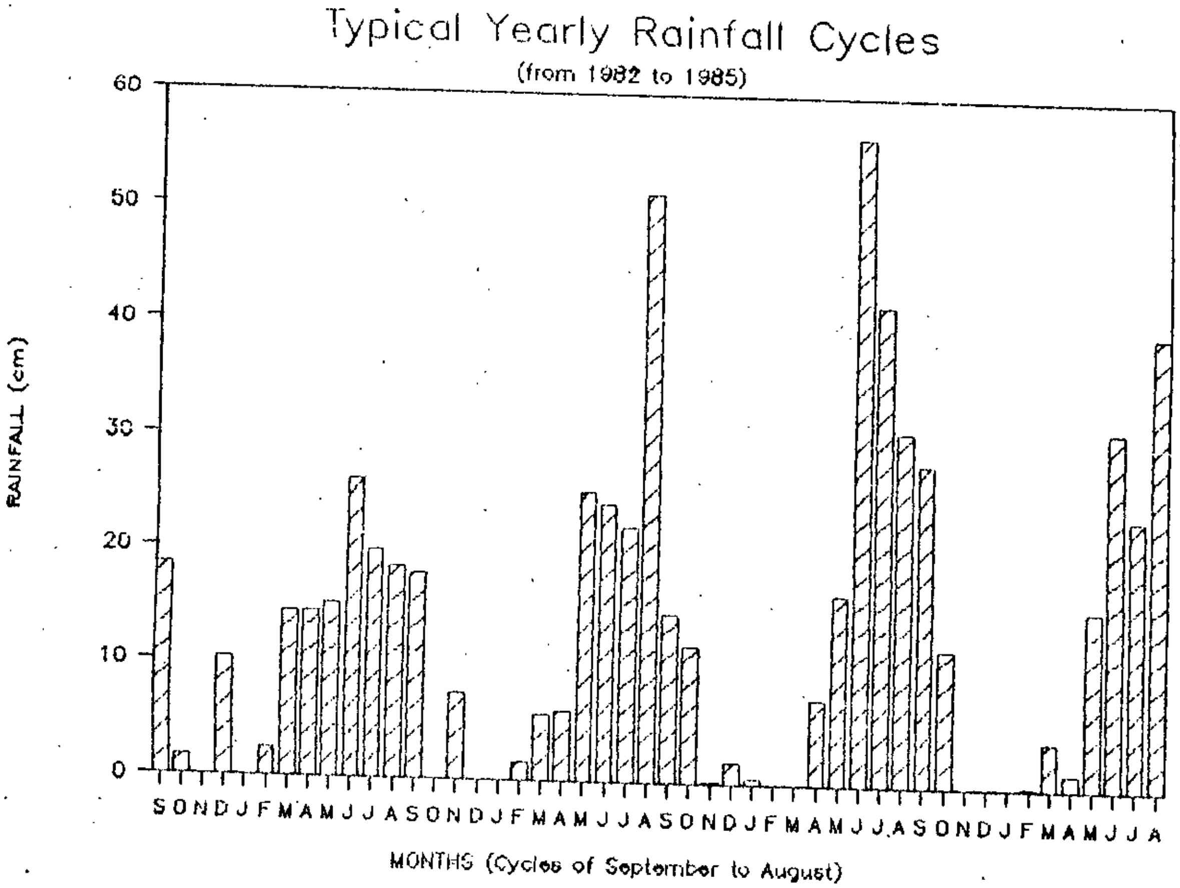


FIG. 4.7 : PLOT OF MONTHLY TOTAL RAINFALLS FROM SEPTEMBER-'82 TO AUGUST-'85 / R-19

will be shown in the next chapter) 90% or more of the variations in either HWL, LWL or both. Log-normal distribution was taken for frequency analysis as a first choice. The distribution was found to be good enough for the months of wet period. These months also passed the Kolmogorov-Smirnov (K-S) test of acceptability. Fig. 4.8 shows one such log-normal plot and K-S bounds at 95% confidence level for the month of August belonging to the station R-463.

On the other hand, months of dry period needed some special treatments due to presence of 'zero' rainfalls which could not be plotted on a log-normal paper. There are three methods in practice to tackle such specialty (Haan,1977).

i] To add a small constant to all the observations and then, to follow the usual procedure.

ii] To use partial series, excluding zero values or values below certain lower limit.

iii] To use the theorem of total probability to get a mixed distribution with a finite probability that $x = 0$ and a continuous distribution of probability for $x > 0$.

The third method seems to be theoretically more sound than the other two. It does not distort the data in any way - either by adding constant to 'zero's or by deleting zero and/or near zero values in the series. However, it requires elaborate

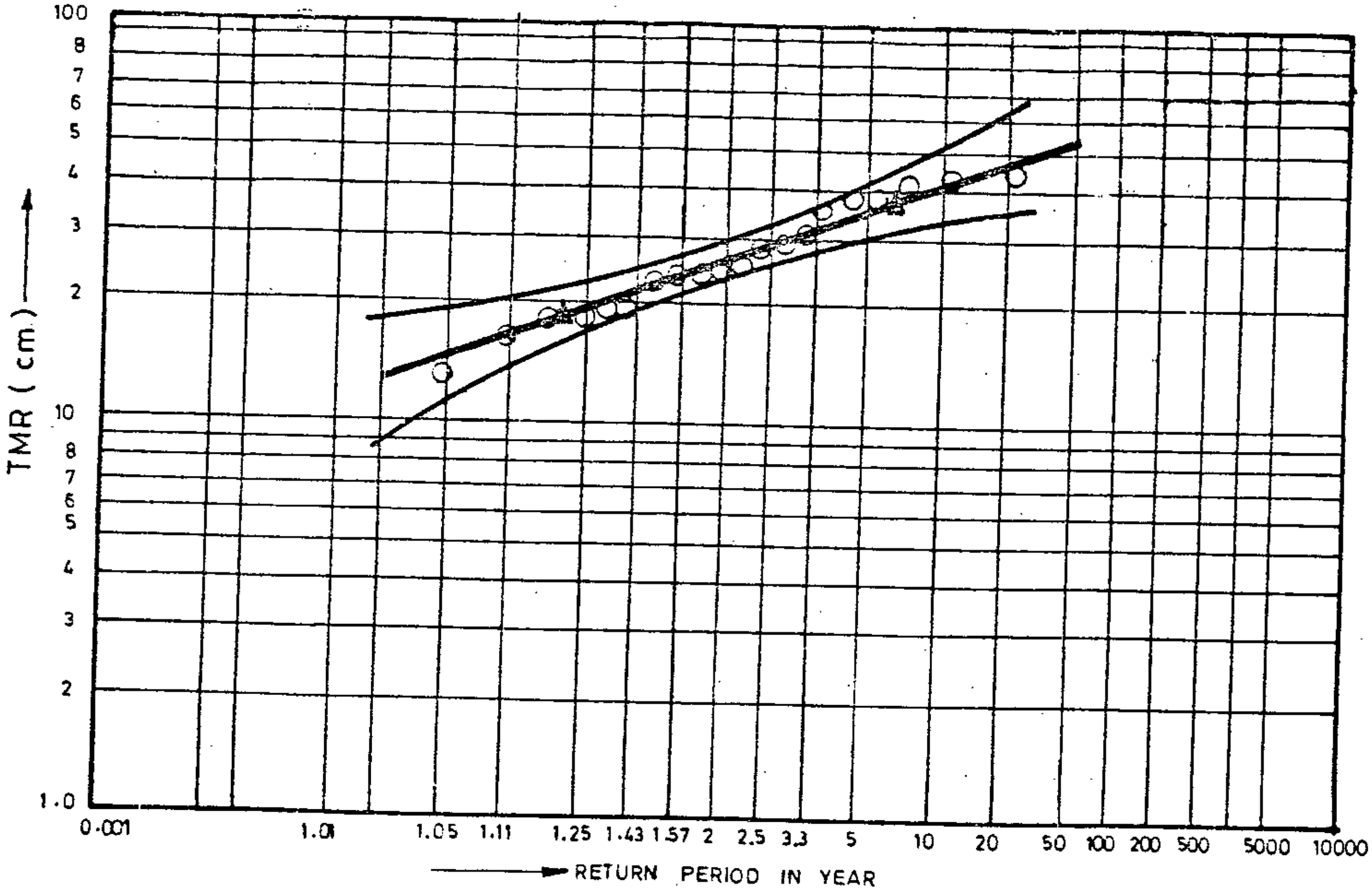


FIG. 4.8 PROBABILITY PLOT OF AUGUST/ R-463

calculation and its real merit from practical point of view is not yet well established. So, methods (i) and (ii) were resorted to in this study.

In method (i) a rainfall of 1 mm (0.1 cm) was assigned to all the 'zero' rainfalls. The plotting positions were then calculated and plotted accordingly. Fig. 4.9 shows one such plot for the month of February of station R-452. It is clearly revealed in the figure that, rains of >1.0 cm arranged themselves into a hypothetical straight line (line A in Fig. 4.9). But rains of lesser magnitude scattered well off that line. And the theoretical log-normal line (line B in Fig. 4.9) failed to be a good fit to the plotting positions. Similar characteristics were noted for all other months of dry season and for all the stations under study. As the purpose of analysis was to predict design rains of longer return periods, it was concluded that giving equal weights to all the data points would produce highly erroneous results. At this point, method (ii), although biased to some extent, seemed to be an intuitively better choice.

Method (ii) turned out to be quite befitting for the months of dry season. Here total monthly rainfalls of <1.0 cm were deleted from calculation. So, the resulting series became curtailed at the lower end. The new plotting positions, however, showed excellent agreement with the new theoretical lines. Again, referring to Fig. 4.9, line C is the new

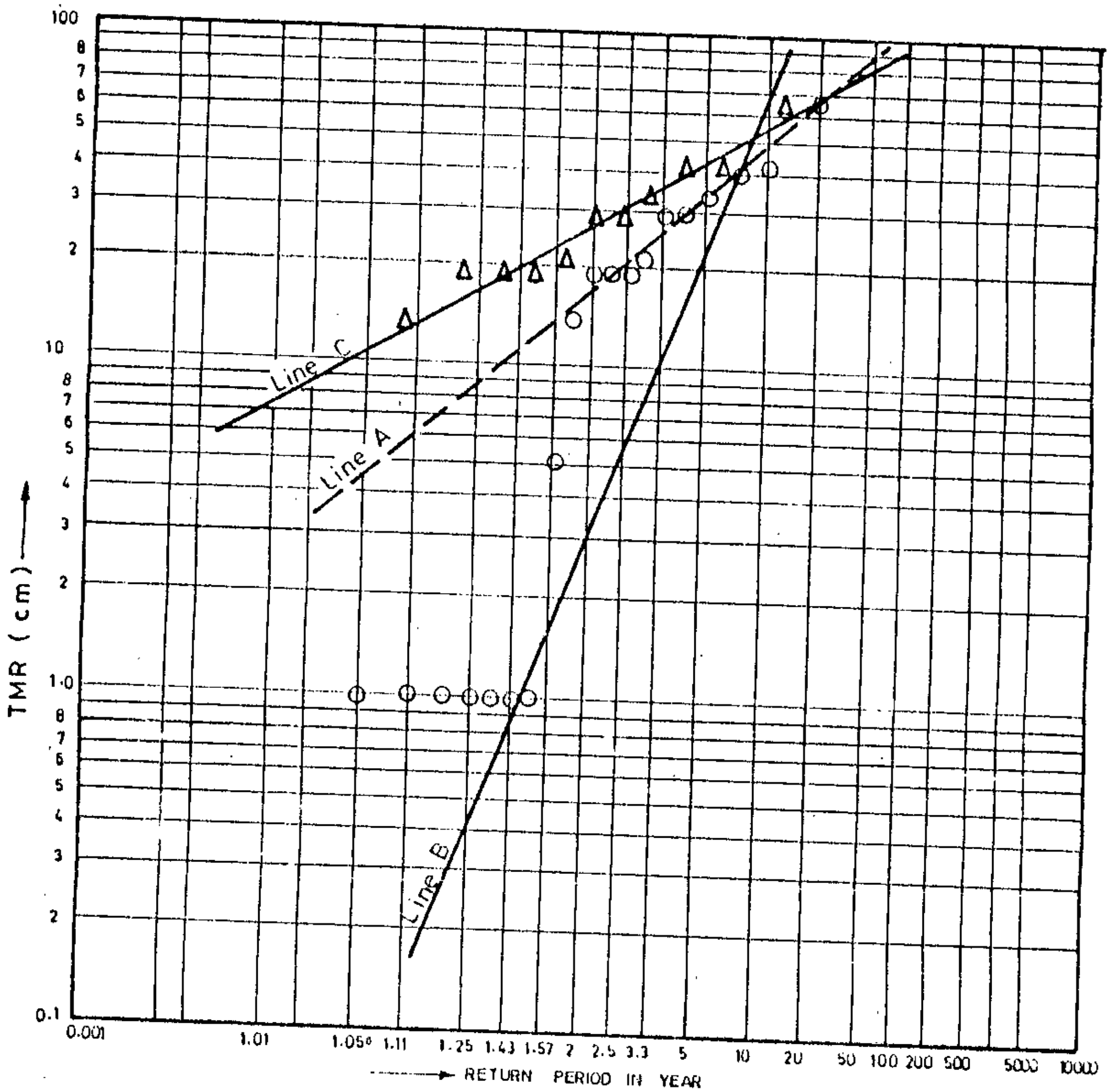


FIG. 4.9 PROBABILITY PLOTS OF FEBRUARY / R - 452

theoretical line showing a very good fit indeed. Finally, recapitulating that the interest of investigation lies towards the rainfalls of longer return period, it was concluded that method (ii) was reasonable and better choice compared to method (i). However, the months of December and January were excluded from the frequency analysis due to their insignificant contributions into the model, as will be shown later in Chapter 5.

4.4.3. River Stage and Discharge

4.4.3.1. Type

Water level (WL) and discharge (Q) were available at IFCDR on a daily basis from the water year of 1969-70 to water year of 1986-87 for the six selected locations (adequate for the proposed nature of study). Table 4.3 shows salient features of these stations. Relative locations of the river stations were already shown in Fig. 4.6.

4.4.3.2. Linear Interpolation

To carry on the proposed stream-aquifer interaction study, linear interpolation was used to get river stages at intermediate locations between pairs of stations. For this part of the study, the first location was at half way between stations S-91 and S-99; the second location was at 33 km downstream from station S-100 (and 42.375 km upstream from station S-101), both on the river Gorai (Fig. 4.1). The

TABLE 4.3

STAGE / DISCHARGE STATIONS

NO.	STATION NO.	LOCATION	DATA TYPE	RIVER
1	S-91	TALBARIA	WL	39,GANGES
2	S-99	GORAI RLY. BRIDGE	WL,Q	42,GORAI-MADHUMOTI
3	S-100	JANIPUR	WL	42,GORAI-MADHUMOTI
4	S-101	KAMARKHALI	WL,Q	42,GORAI-MADHUMOTI
5	S-101.5	KAMARKHALI	WL	42,GORAI
6	S-171	GORAGANJ	WL,Q	65,KUMAR

WL : WATER LEVEL,   Q : DISCHARGE

underlying assumption was that the river has a considerably flat bottom slope, and hence water surface slope, which favored the direct linear interpolation.

To check the validity of the assumption, WLs of 31st December of S-99 were plotted against WLs of S-91. The correlation coefficient R by simple linear regression was found to be 0.79 (Fig. 4.10.a). It suggests that the linear interpolation for the location halfway between stations S-91 and S-99 was acceptable; but, in-situ water level should have been used to get exact water levels. Interestingly enough, plotting of WLs of S-101 versus that of S-100 gave R=0.95 which may be considered as excellent (Fig. 4.10.b).

The probable cause of slight distortion from linearity for the first case may be attributed to the presence of the district town Kushtia along the south bank between stations S-91 and S-99. It is quite possible that certain amount of water is being withdrawn from the river for municipal or other purposes. Also, there may be some return flow from the town into the river Gorai, the ultimate result being local change in slope of the water surface profile.

So, for the stage of the location halfway between S-91 and S-99, some unavoidable error was initially introduced. Interpolated data at this location were correlated to groundwater levels of KT01 and KT03. On the other hand, linear

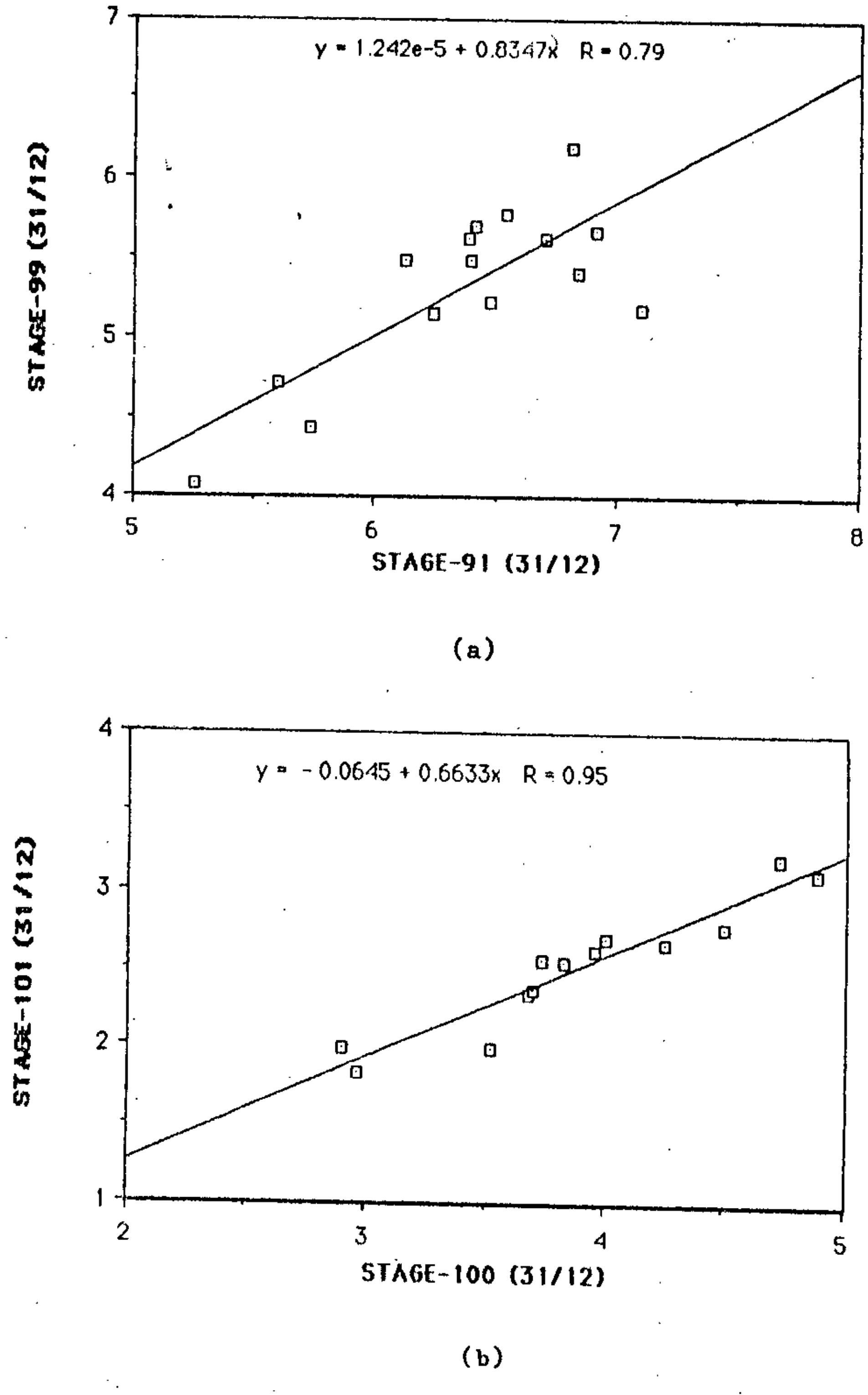


FIG. 4.10 : CORRELATION BETWEEN UPSTREAM AND DOWNSTREAM STAGES

interpolation to get stage at 33 km downstream from S-100 was good enough approximation. Data so obtained were used to study the correlation with groundwater levels of JE25. But, stages of station S-171 on Kumar required no such interpolation and were used along with groundwater levels of JE05 for similar study.

4.4.3.3. River Flow

It was assumed in the theoretical part of stream-aquifer interaction that the stream was flowing continuously with fluctuating water levels which affected the level of water table of the surrounding aquifer. To supplement this assumption, discharge data at two selected stations were collected. The first station was S-101 on river 42, Gorai-Madhumati which was the last of the stations 91-99-100 from upstream to downstream being used for the study. The other station was S-171 on river 65, Kumar. Data from both the locations confirmed that there was continuous flow in the rivers during December-January period, specifically on 31st of December.

4.5. The Software Requirements

A number of powerful softwares are available for doing multiple regression analysis like SPSS, SAS, BMDP, MINITAB, STATWORKS, SELECT, etc. However, not all of them are equally suitable to carry on all the proposed analyses of this study.

Among the packages mentioned above, STATWORKS (Apple Macintosh compatible) is the most user friendly and it has got excellent graphics display. For major statistical jobs, SPSS and SAS are being universally used, although their memory requirements are very high and they are easy to handle when installed in the Mainframe system. Graphics of SPSS and SAS are not as vivid as that of STATWORKS . For this study, both SPSS and STATWORKS were used as per job requirement.

The other packages are currently not available at BUET Computer Center. But they have got some powerful features. For example, MINITAB uses the HAT matrix technique to identify the outliners - a unique feature not available in other packages (the author developed a program of his own which includes this specialty). BMDP and SELECT are very versatile in generating the so called 'best subsets' comprising different number of carriers. For this study, the 'STEPWISE' option of SPSS was used along with the special option of 'FORCED ENTRY'.

It is obvious that, all the packages mentioned above are good enough for simple linear regression used in the stream-aquifer interaction study.

Chapter 5

RESULTS AND DISCUSSIONS

5.0. Introduction

In this chapter, all the steps of analyses will be followed in detail and pertinent comments will be made. Firstly, the multiple regression model will be covered along with the selection of best subset and then, the study of stream-aquifer interaction will be taken up.

As all the wells studied by multiple regression model were subjected to the same procedure, the case HWL simulation of well no. JE06 will be picked up for detailed analysis. The reason for choosing JE06 is that it has exhibited excellent match between actual and simulated groundwater levels and thus allows further fine tunning of the original 12-carrier (x_{i1}, reported as X(1) in the subsequent discussion is actually a dummy variable, and hence, will not be referred to as a carrier) model into a smaller sized 'best subset'. To avoid repetition, findings for other wells regarding simulation of HWL and LWL will be given in tabular form in Appendix-B. However, brief comments will be made about salient features of inputs and outputs for these wells when necessary.

5.1. Sample Analysis for HWL of Well no. JE06

5.1.1 Basic Input Statistics

To provide an insight into the quality of input data, basic statistics of all the variables were calculated first. Table 5.1 gives such statistics for simulation of HWL of well no. JE06. Here X(1) represents the variable associated to the constant (hence, X(1) always equals to 1.0), X(2) through X(13) stand for TMRs of September to August and finally, YF represents the field or actual HWL.

The statistics given include mean(MEAN), minimum (MN), maximum (MX), standard deviation (STD), skewness (SKEW) and kurtosis (CUR). The last three properties for X(1) were reported as 99.99 which simply means 'Not Applicable'.

As the underlying assumption of the multiple regression analysis was that the variables had normal (or at least near normal) distribution, these basic statistics point out to what extent the assumption was satisfied. For a theoretical normal distribution, the coefficients of skewness and kurtosis are 0.0 and 3.0 respectively. However, Table 5.1 makes it pretty clear that the carriers X(4), X(5) and X(6) (stand for the TMRs of November, December and January respectively) had properties greatly different from than those of normal. Also they had STDs greater than the respective MEANs.

TABLE 5.1

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	24.92	10.97	41.78	9.18	.11	2.71
X(3)	13.02	.51	33.27	9.21	.57	3.38
X(4)	4.62	.00	30.61	7.46	2.72	11.82
X(5)	2.44	.00	20.61	5.23	2.88	11.88
X(6)	1.53	.00	13.13	3.12	3.37	15.32
X(7)	2.12	.00	9.19	2.43	1.97	6.99
X(8)	3.52	.00	8.81	3.32	.28	1.89
X(9)	10.05	.00	27.55	6.44	.96	5.57
X(10)	20.48	2.06	49.66	12.12	.63	3.92
X(11)	28.24	.00	76.76	17.45	1.15	5.73
X(12)	32.91	6.39	65.41	14.74	.50	3.64
X(13)	29.08	14.40	57.20	11.94	1.12	4.04
YF	5.004	3.426	5.913	.637	-.842	4.214

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	3.7606	***	***	.0000	.0000
X(2)	-.0251	.626	15.476	.2307	.0993
X(3)	.0322	.419	10.348	.2962	.1275
X(4)	.0185	.086	2.115	.1382	.0595
X(5)	-.0105	.026	.635	.0551	.0237
X(6)	-.0089	.014	.336	.0277	.0119
X(7)	-.0673	.142	3.522	.1636	.0705
X(8)	-.0443	.156	3.851	.1472	.0634
X(9)	-.0367	.369	9.126	.2365	.1018
X(10)	.0233	.478	11.811	.2827	.1217
X(11)	-.0024	.068	1.684	.0421	.0181
X(12)	.0160	.528	13.045	.2362	.1017
X(13)	.0390	1.134	28.049	.4660	.2007

So, it could be guessed even before getting the regression coefficients or contribution factors that such carriers would contribute little towards the enhancement of the model; on the contrary, they would increase the standard error of estimate. It is no surprise that subsequent analyses confirmed this prior suspicion.

5.1.2. The Correlation Matrix (CM)

Technique of generating the correlation matrix or CM has already been discussed in Chapter 3. Its main utility is that it helps to identify the degree of 'statistical dependency' or correlation between the predefined independent variables. The CM for simulation of HWL of well no. JE06 is given in Table 5.2 (for the 12-carrier original model) which is a 13*13 symmetric matrix. The diagonal elements of this matrix give the correlation of a carrier with itself which is always 1.0. The off diagonal elements give the cross-correlations between the carriers corresponding to the rows and columns of the elements. It is clearly noticeable from the three inner boxes within the CM that all the high cross-correlation terms are related either to the months of November, December or January. Hence, these three months may be considered as the most problematic ones. Interestingly enough, basic input statistics also pointed out to the same three months as major deviants. Hence both input statistics and correlation matrix jointly suggest that performance of a 10-carrier model excluding the

TABLE 5.2

CORRELATION MATRIX FOR SIMULATION OF HWL OF WELL NO. JE06

Data File, RESI1

HWL	1.000	-0.109	0.441	0.166	-0.125	0.170	-0.599	-0.099	0.139	0.105	0.469	0.172	0.674
SEPT	-0.109	1.000	-0.118	-0.277	-0.339	-0.401	0.078	-0.232	-0.223	-0.225	-0.091	0.252	0.281
OCT	0.441	-0.118	1.000	0.065	0.024	0.383	-0.021	-0.277	0.217	-0.283	0.093	-0.241	0.233
NOV	0.166	-0.277	0.065	1.000	0.793	0.756	-0.208	-0.133	0.644	0.063	0.041	0.082	0.026
DEC	-0.125	-0.339	0.024	0.793	1.000	0.763	-0.155	0.007	0.610	-0.000	0.137	0.003	-0.157
JAN	0.170	-0.401	0.383	0.756	0.763	1.000	-0.112	-0.264	0.669	0.011	0.324	0.060	-0.074
FEB	-0.599	0.078	-0.021	-0.208	-0.155	-0.112	1.000	-0.093	0.009	0.184	-0.436	-0.154	-0.477
MAR	-0.099	-0.232	-0.277	-0.133	0.007	-0.264	-0.093	1.000	-0.176	0.276	-0.099	0.009	-0.037
APL	0.139	-0.223	0.217	0.644	0.610	0.669	0.009	-0.176	1.000	0.133	0.239	0.048	0.233
MAY	0.105	-0.225	-0.283	0.063	-0.000	0.011	0.184	0.276	0.133	1.000	0.367	0.056	-0.189
JUNE	0.469	-0.091	0.093	0.041	0.137	0.324	-0.436	-0.099	0.239	0.367	1.000	0.059	0.335
JULY	0.172	0.252	-0.241	0.082	0.003	0.060	-0.154	0.009	0.048	0.056	0.059	1.000	-0.069
AUG	0.674	0.281	0.233	0.026	-0.157	-0.074	-0.477	-0.037	0.233	-0.189	0.335	-0.069	1.000
	H	S	O	N	D	J	F	M	A	M	J	J	A

two most disturbing months of December and January (also statistically insignificant, as will be shown in the coming sections) should be checked against the initial 12-carrier model.

5.1.3. Estimating the Parameters and the ANOVA Table

5.1.3.1. Inferences on the Regression Coefficients

After getting the basic input statistics and the correlation matrix, the next step becomes determining the regression coefficients. Table 5.3 contains a complete list of such coefficients for simulation of HWL of well no. JE06. The first column gives the values obtained using equation 3.7. As the process involves thousands of operations which may cause accumulation of roundoff errors, a check column CB(I) is also reported side by side. Here the coefficients were determined by direct solution of equation 3.6 using the Cholesky's algorithm (Rice,1983). The standard deviation of the coefficients are reported in the next column as STD(I). Finally the column of T(I) gives the t statistics for each of the coefficients.

As suspected earlier, individual t statistics for the months of November, December and January are all found to be less than $t_{1-\alpha/2,n-k}$ (at $\alpha=0.05$ in this case) which is equal to 2.57 (Table 5.3). Now to check whether deletion of the two worst carriers (i.e., TMRs of December and January) makes any

TABLE 5.3

REGRESSION COEFFICIENTS AND STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	3.7606	3.7606	.2526	14.8849
2	-.0251	-.0251	.0064	-3.9440
3	.0322	.0322	.0064	5.0545
4	.0185	.0185	.0153	1.2121
5	-.0105	-.0105	.0188	-.5604
6	-.0089	-.0089	.0406	-.2189
7	-.0673	-.0673	.0304	-2.2155
8	-.0443	-.0443	.0168	-2.6414
9	-.0367	-.0367	.0117	-3.1272
10	.0233	.0233	.0062	3.7930
11	-.0024	-.0024	.0053	-.4570
12	.0160	.0160	.0034	4.7110
13	.0390	.0390	.0067	5.8213

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	450.6903
REGRESSION	12.0000	6.7680
RESIDUAL	5.0000	.1307
TOTAL	18.0000	457.5889

MSSQ= .0261
MS = .1616
MRSQ= .9811
MR = .9905
F = 21.5838

significant difference, equation 3.14 was used. It was found in a separate model run using ten carriers (so, l=11) that SSE_l=0.1417 for the reduced model. For the full model, SSE_k was found to be 0.1307 which was reported as RESIDUAL in Table 5.3. Now, from equation 3.14, the value of F was found to be 0.21, whereas, the critical value of F or $F_{1-\alpha,k-l,n-k}$ (α =0.05) is 5.79. Hence the null hypothesis becomes accepted indicating that TMRs of December and January failed to explain significant amount of variation of the dependent variable HWL. So, once again the analysis suggests that TMRs of December and January may be excluded from the final model.

Scanning through Table 5.3, a further detection of weak carrier is possible which is X(11) or TMR of June. It is rather unexpected that a month like June having very high rainfall intensity fails to be statistically significant. The probable explanation of such outcome is that the carrier badly suffered from the problem of multicollinearity (Devore,1982). Literally it means that TMR of June could be expressed by a linear function of a number of the remaining independent variables. Unfortunately, the correlation matrix does not give any direct indication to this kind of interdependency. The only facial symptoms of such flaw are that the respective regression coefficient fails the t-test and bears the sign opposite to what is expected. Evidently, statistic from Table 5.3 that B(11) (or equivalently b_{11}, the contribution factor of June) came out to be -0.0024 with t-ratio of -0.457 makes the diagonosis almost conclusive. Possible remedy may be to use the 'Ridge Regression' technique which is designed to handle such problems or to use some kind of multivariate analysis like the method of 'Principal Components'.

The carrier X(11), however, was retained in the proposed 10-carrier model; firstly because it has got strong hydrologic significance and secondly due to its role in optimizing the process of variable selection to get the 'Best Subset' from statistical point of view.

5.1.3.2: ANOVA Table and the Test of Model Utility

The 'Analysis of Variance' or ANOVA table is presented in the lower part of Table 5.3. Meaning of different terms will be obvious when compared with the same of ANOVA table shown in section 3.4 of Chapter 3. Additional terms reported have the following meaning :

MSSQ = sample estimate of variance, s^2

MS = standard error of estimate, s

MRSQ = multiple coefficient of determination, R^2

MR = multiple correlation coefficient, R

F = F statistic for the full model

As given by the table, R^2 is 0.9811 in this case which means that the model explained 98.11% of the total variation of the dependent variable and so, the simulation may be treated as excellent. A further test of model utility may be done by F-test. The F value of 21.58 reported in this table was estimated from equation 3.15. The corresponding critical value $F_{1-\alpha, k-1, n-k}$ (at α=0.05) is 4.68. Clearly, the alternate hypothesis becomes accepted which means that the model is significantly explaining the variation of HWL.

In short, ANOVA reflects the overall performance of the model by R^2, s and F. This table also greatly helps in the process of variable selection where both R^2 and s are needed.

5.1.4. Inferences on Prediction

One way of checking the goodness of fit of prediction by the model is to calculate the confidence limits (at some predefined level, say, at 95%) on each of the predictions and then, to superimpose those limits on the plot of actual values of dependent variable. Table 5.4 gives such limits (calculated from equation 3.17) along with the actual HWLs for well no. JE06. The corresponding plot is shown in Fig. 5.1. As revealed by the figure, all the actual HWLs remained within the spectrum of prediction outlined by the upper and the lower confidence limits. It visually confirms the excellent nature of prediction by the model. A further complementary plot of actual and simulated HWLs is shown in Fig.5.2.

It should be noted at this point that, due to lack of adequate data, all the available observations were used for model development. So, goodness of individual prediction based on

TABLE 5.4

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	5.507	5.512	-.005	-.084	5.220	5.803	.0208	.7956
2	4.923	5.122	-.199	-4.039	4.914	5.330	.0106	.4059
3	5.608	5.683	-.075	-1.333	5.392	5.973	.0206	.7903
4	4.700	4.547	.153	3.246	4.295	4.800	.0156	.5984
5	5.913	5.908	.005	.083	5.606	6.211	.0224	.8584
6	4.450	4.432	.018	.407	4.213	4.650	.0117	.4478
7	5.404	5.408	-.004	-.069	5.130	5.685	.0189	.7215
8	5.660	5.499	.161	2.846	5.220	5.777	.0190	.7277
9	5.026	5.074	-.048	-.964	4.792	5.357	.0196	.7493
10	4.750	4.761	-.011	-.225	4.487	5.035	.0184	.7041
11	5.150	5.186	-.036	-.692	4.863	5.509	.0256	.9785
12	5.026	5.139	-.113	-2.249	4.886	5.392	.0157	.5994
13	5.150	5.089	.061	1.190	4.849	5.329	.0141	.5406
14	4.036	4.112	-.076	-1.871	3.803	4.421	.0234	.8955
15	4.452	4.419	.033	.738	4.106	4.733	.0241	.9211
16	3.426	3.424	.002	.050	3.110	3.739	.0242	.9280
17	5.104	5.032	.072	1.410	4.790	5.274	.0144	.5507
18	5.784	5.723	.061	1.051	5.433	6.013	.0206	.7873

The critical HAT(I) value is : 1.4444

YF=ACTUAL HWL, P=PREDICTED HWL, YPVAR=VARIANCE OF PREDICTION

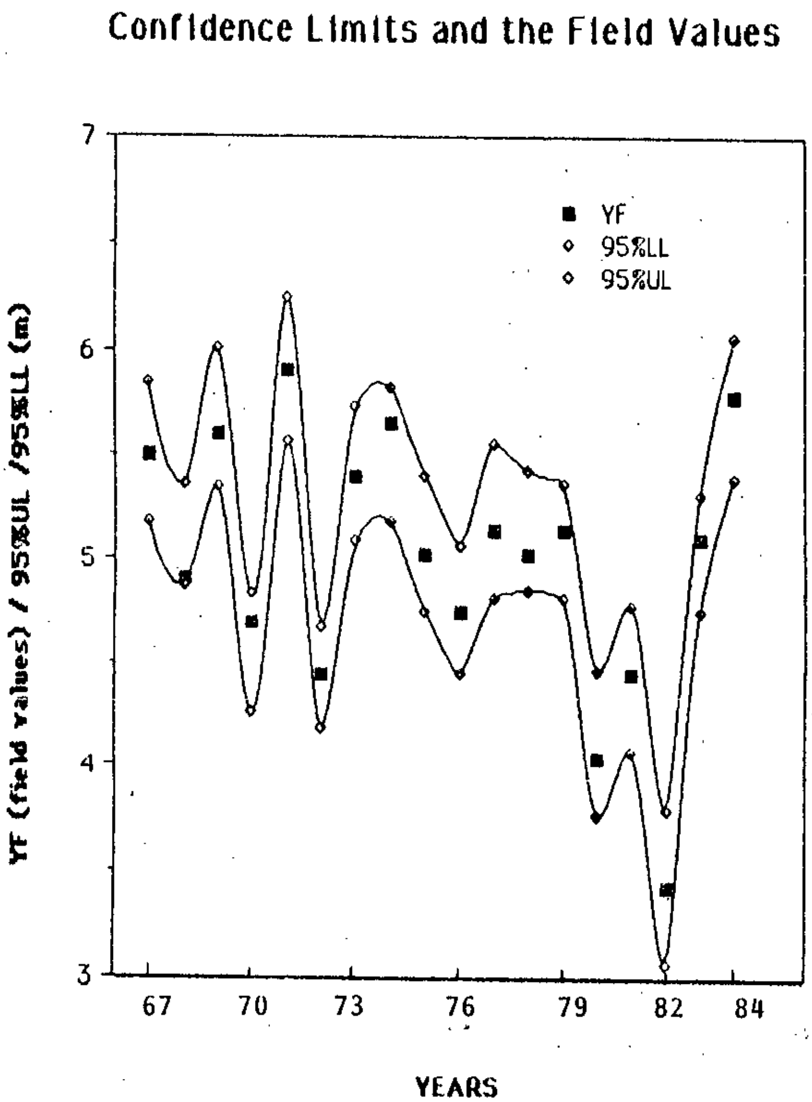


FIG. 5.1 : CONFIDENCE LIMITS OF PREDICTIONS AND THE FIELD VALUES

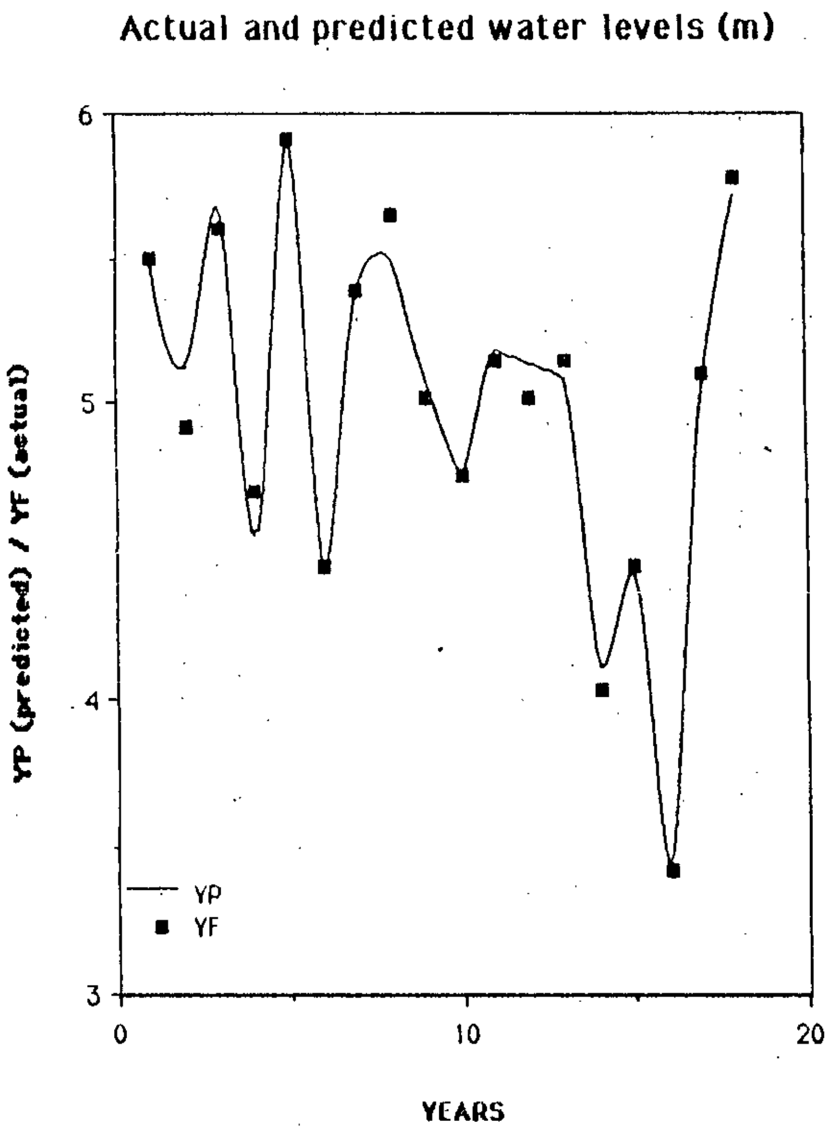


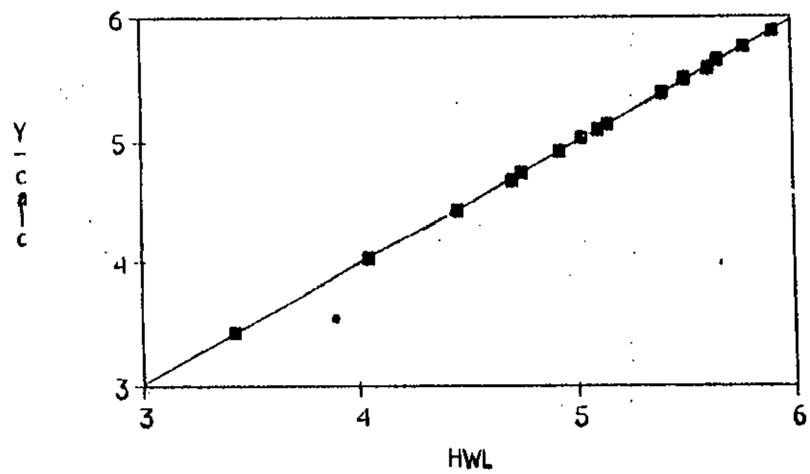
FIG. 5.2 : ACTUAL AND SIMULATED HWLs FOR WELL NO. JE06

the observation which was not used for the model development could not be tested. However, with the accumulation of more data in future, the model can be put in a real test to judge its prediction capability.

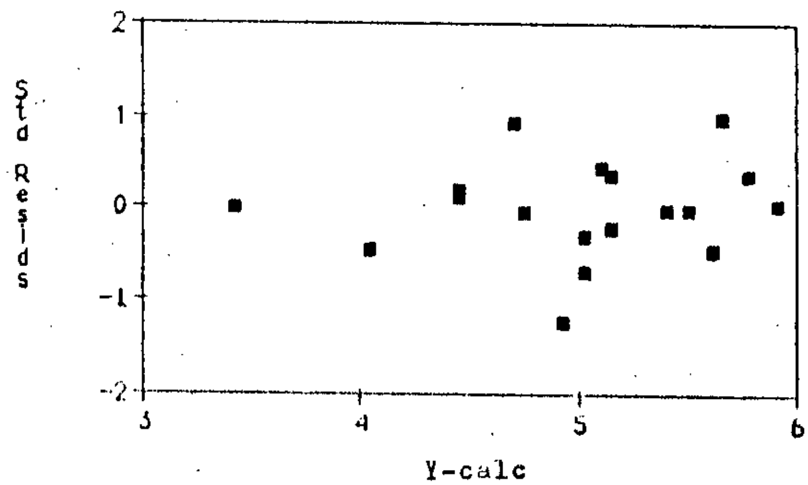
5.1.5 Study of HAT Elements and Residuals

As discussed in Chapter 3, the diagonal elements of HAT matrix (given by equation 3.19) measures the influence of y_i on its own predicted value $\hat{y}_i$. When an h_{ii} gets bigger than the critical value of 2k/n, the ith observation of the dependent variable y_i may be considered to be a point of large influence. Both such diagonal h_{ii}s and the critical value are given in Table 5.4. It is noted that none of the HAT values exceeds the critical value given at the end of the table. Recapitulating that missing and erratic y_is were smoothened out during the phase of data processing, such finding is not unexpected. Rather, it indicates the effectiveness of the previous treatments.

The final assessment of a model is usually done by examining the plot of standardized residual versus the predicted variable $\hat{y}_i$. Such a plot for well no. JE06 is given in Fig. 5.3. The nature of the scatter plot of residuals seems to be an ideal one. All the residuals are randomly distributed about zero, but none of them exceeds the limit of -2 to +2. The figure also exhibits another plot of $\hat{y}_i$ versus y_i in ascending



(a)



(b)

FIG. 5.3 : (a) PLOT OF $\hat{y}_i$ VERSUS y_i

(b) PLOT OF e_i^* VERSUS $\hat{y}_i$

order, again the success of linear simulation is clearly displayed. Hence, both the HAT elements and the residual plot promptly disclose the sound performance of the model.

5.1.6 Physical Interpretation: Contribution and Capriciousness of the Independent Variables

Before going for selection of best subset, discussion will be made in this section about the physical interpretation of regression coefficients and associated terms. It is worth recalling that in article 5.1.3.1, significance of a carrier was judged by its t ratio and not by the associated regression coefficient or contribution factor. A contribution factor b_j near zero does not necessarily imply weak relationship between associated carrier x_j and the dependent variable. In fact, b_j can be made very near to zero by multiplying each y_i by a small number c. In reality, this may happen due to change in units of measurement. The effect will be that the new b_js will be c times the old one, but the new s will also be c times the old one, so that t statistic will have the same value. So, a near zero b_j may simply be converted into two or three digit figures which will apparently look very strong contributor, but its real contribution into the model will remain unchanged.

The next question that follows the identification of potential carriers is how to judge their relative contributions into the model. According to Rethati(1983), the role of a month in the development of groundwater can be evaluated by the product of appropriate contribution factor and mean TMR of that month. Likewise, the contribution factor of a month and the standard deviation of corresponding TMR multiplied together gives the 'capriciousness' of the month in contributing to storage.

In Table 5.1, the column of CONT gives such absolute contribution of each month and %CONT gives the same in percentage expressed with respect to the sum of CONTs for all the carriers ('***' for X(1) means 'Not Applicable'). The 'capriciousness' as mentioned above, is given in column of SENS and relative value of the same is given in the next column of RSENS. Now looking at these statistics, comparisons of the real contribution and capriciousness of different carriers become much easier. For example, the largest contributor in simulating the HWL of well no. JE06 is identified as X(13) or TMR of August, which accounts for 28.05% of the total contribution by all the carriers. The model is most sensitive to this particular carrier as revealed by the associated RSENS of 0.2 (equivalent to 20%). On the other hand, contribution of X(16) or TMR of January is virtually nil (only 0.336%) and the model is least sensitive to its value as indicated by corresponding RSENS of 0.0119.

A word of caution, however, should be mentioned at this point. Statistical models are often called the 'Black Box' models because they are not meticulous about the physical relationship between dependent and independent variables. The mere 1.68% relative contribution of X(11) does not mark it as a poor contributor from hydrologic point of view and probable presence of multicollinearily has already been discussed. Hence, statistical inferences do not necessarily imply cause and effect relationship.

5.1.7. Selection of the 'Best Subset'

As discussed in section 3.10.2.3 of Chapter 3, the Stepwise Selection (SS) is so far that best available technique to get the best subset of carriers. Hence, SS is the method resorted to in this study. Different steps in variable selection for simulation of HWL of well no. JE06 are shown in Table 5.5 which reveals some very interesting features as discussed below:

i] The process terminated abruptly just after the inclusion of the most important carrier X(13) or TMR of August. This happened because the highest absolute t-ratio for the next entry was 1.83 which was less than t_{in} of 2.0. This undue termination demonstrates the major drawback of any mechanical process. At this point, mere common sense dictates that more variables should be entered to improve both R^2 and s of the model. So, the second carrier X(7) was forcibly entered using a special option of SPSS.

TABLE 5.5

SEQUENCES OF STEPWISE SELECTION

STEP NO.	CARRIER TO ENTER	ABSOLUTE t-RATIO	CARRIER TO OUT	R^2	s
1	X(13)	3.60	NONE	0.45	0.49
2	X(7)	1.83*	NONE	0.55	0.45
3	X(3)	2.24	NONE	0.66	0.41
4	X(10)	2.99	NONE	0.80	0.33
5	X(12)	2.03	NONE	0.85	0.29
6	X(9)	1.36*	NONE	0.87	0.28
7	X(2)	2.35	NONE	0.92	0.23
8	X(8)	3.03	NONE	0.96	0.18
9	X(4)	2.10	NONE	0.97	0.15
10	X(11)	1.40*	X(4)**	0.98	0.14
11	X(5)	0.67*	IGNORED	0.98	0.15
12	X(6)	0.22*	IGNORED	0.98	0.16

* : FORCED ENTRY

** : ELIMINATION OF X(4) IS INVALID DUE TO FORCED ENTRY

IGNORED : STEPS CARRIED ON ONLY TO MONITOR R^2 AND s .

ii] After forced entry of the second carrier, the next three carriers X(3), X(10) and X(12) entered directly into the model, each time making considerable improvement of the model. Moreover, just after entering the 3rd carrier, t-statistic of the 2nd one or X(7) became significant, thus validated its presence in the model.

iii] The process terminated again at the sixth entry as the largest absolute t (this time for X(9)) was only 1.36 against the required value of greater than t_{in} or 2.0. But recalling that the allowable number of carriers for 20 years' of observation may be 7 or 8, X(9) was included in the model. This inclusion further increased the R^2 and decreased the s.

iv] The next three carriers after X(9), which were X(2), X(8) and X(4), entered directly into the model. The first two entry of X(2) and X(8) considerably improved the model performance satisfying both the criteria for R^2 and s. The entry of X(4), however, stirred the model only slightly.

v] Despite the forced entry of X(7) and X(9), carriers in each step which were already present in the model maintained t-ratios greater than t_{out} or 1.9748. So, no case of dropping out occurred upto step no. 9.

vi] The tenth entry of X(11) was again a forced one, done only to monitor the behavior of R^2 and s. This also caused dropping out of X(4). But replacement of X(4) by X(11) introduced no change in the model performance. However, X(11) is more preferable due to its strong hydrologic significance.

vii] The last two carriers entered by force were X(5) and X(6) by sequence. Both the entry failed to increase the R^2 significantly, rather, they caused successive increase in s, thus deteriorated the overall model performance.

Interpretation of model performance during this entire selection process becomes much easier by simply having a glimpse of Fig. 5.4. It depicts the nature of variation of R^2 and s with increasing N - the number of carriers in the model. A bit of scanning through the figure shows that, after the eighth entry, model performance became almost optimum. However, after the tenth entry, R^2 reached very close to the possible maximum and at the same time, s became the minimum. This is the so called 'optimum configuration' of a model having the near-largest R^2 and smallest s (statistics of this 10-carrier model is given in Appendix-B). Considering the fact that the tenth entry was X(11) or TMR of June which is supposed to be an important hydrologic contributor, the 10-carrier model excluding TMRs of December and January may be declared as the 'Best Subset' for HWL of well no. JE06.

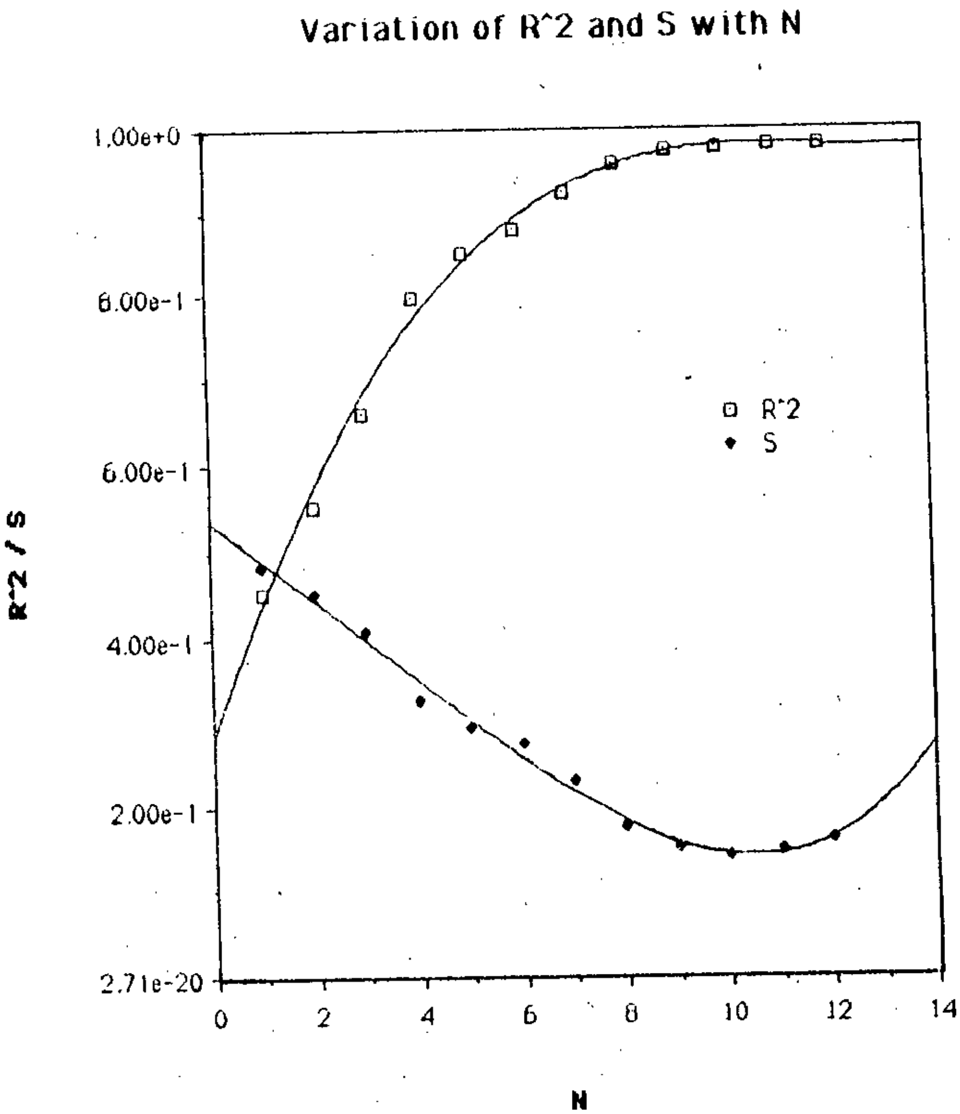


FIG. 5.4 : MODEL PERFORMANCE IN TERMS OF R^2, s AND N

5.1.8. Forecasting for the Future HWL/LWL

One of the objectives of this study was to discuss the possibility of using the multiple regression model for forecasting of annual culminations of the water table. To use a model for forecasting, its quality of prediction should be good enough, or in other words, its R^2 should be 0.8 or higher. It is seen from Table 5.6 that, nine out of twelve of the simulations passed this criteria, and hence, corresponding regression equations can be used for predicting the future HWL or LWL. It is advisable that, instead of using the full twelve or nine carrier model (for HWL and LWL respectively), the optimum subset of carriers should be identified first. And forecasting should be attempted by using such 'Best Subset' which will ensure the narrowest spectrum of prediction.

Input for forecasting may be extracted from the probability plots of potential carriers as given in Appendix-A. A number of forecasting exercises are shown in Table 5.7 for simulation of HWL of well no. JE06. The 10-carrier optimum subset was used for all these predictions.

The mean values of the potential carriers in the table were taken from Table 5.1. The corresponding simulated value of HWL of 5.02 m is close to the actual mean HWL of 5.00 (the is actually due to roundoff error during computations).

TABLE 5.6

MODEL PERFORMANCE SUMMARY

WELL NO.	SIMULATION	R^2	s
JE04	HWL	0.80	0.51
	LWL	0.75	0.34
JE05	HWL	0.96	0.28
	LWL	0.93	0.19
JE06	HWL	0.98	0.16
	LWL	0.78	0.27
KT01	HWL	0.86	0.15
	LWL	0.83	0.29
KT03	HWL	0.84	0.31
	LWL	0.80	0.43
KT05	HWL	0.90	0.44
	LWL	0.74	0.27

TABLE 5.7

FORECASTING FOR HWL OF WELL NO. JE06

Months	Contribution Factors	Total Monthly Rainfalls (cm)				
		Mean	10-Year	20-Year	50-Year	100-Year
SEPT	-0.0253	24.92	40.00	45.00	52.50	58.00
OCT	0.0318	13.02	29.90	40.00	57.50	73.00
NOV	0.0112	4.62	11.40	15.00	18.60	22.00
FEB	-0.0691	2.12	8.50	12.70	18.00	22.50
MAR	-0.0492	3.52	9.00	10.75	13.33	15.00
APL.	-0.0403	10.05	21.25	28.75	39.00	48.00
MAY	0.0257	20.48	40.00	50.00	65.00	75.00
JUNE	-0.0044	28.24	60.00	76.00	100.00	133.33
JULY	0.0164	32.91	55.00	67.00	78.00	87.50
AUG	0.0418	29.08	47.50	56.00	67.50	74.00
SIMULATED HWL (m) :		5.02	5.58	5.88	6.32	6.49

[Constant or the Intercept Term of Regression Euqation = 3.7533]

The table also gives the probable HWLs corresponding to TMRs having return periods of 10, 20, 50 and 100 years. It is noted from the table that TMRs of 100 year return period bring the groundwater table to a level of 6.495 m which is very close to the ground level of 6.805 m in the location of the well. However, it is very unlikely that in reality, all the potential carriers will have a magnitude of 100 year return period in the same year. The highest recorded value of HWL from 1966 to 1986 is 5.913 m which is less than the magnitude of 6.495 m resulting from carriers of 100 year return period. It is obvious that numerous combinations of return periods to different carriers may be tried to predict a HWL or LWL.

The forecasted value of HWL or LWL will be helpful in the process of decision making. For example, a very high HWL indicates a possibility of water logging in the area and a very low LWL indicates the possibility of drying out of surface water sources which were being being augmented by the aquifer. Moreover, the difference of HWL and LWL multiplied by the specific yield of the aquifer material gives the amount of annual recharge into the aquifer.

An interesting exposition of the model performance may be reported here. As given by Table 5.7, the forecasted HWL of JE06 for the mean values of associated TMRs is 5.02 m. The same for LWL (using the full model) is 0.84 m. The difference of HWL and LWL becomes 4.17 m or 4170 mm. The well JE06 is located in the upazila of Magura and the average specific yield of this area is given to be 0.06 by Karim(1984). Hence, the annual recharge on a year of mean rainfalls in each of the months becomes 4170*0.06=250.2 mm. And Karim(1984) has reported the mean annual recharge for the upazila Magura as 262 mm - a reasonably close figure to the predicted 250.2 mm by the model.

Such forecasting exercises may be carried on for other wells and the potential of forecasting is obviously enormous. In fact, the use of forecasting by a model is only limited by the scope and limitations of the model itself.

5.2 Brief Discussion about the Simulation of HWL and LWL of Other Wells

All the steps followed above for analysis of HWL simulation for well no. JE06 may be repeated for the rest of the wells. So, instead of going through the same process many more times, relevant findings about HWL and LWL simulation for these wells are given in Appendix-B. However, the discussion to follow will outline salient features of the simulations for the wells other than JE06.

For simulation of HWL of well no. JE04, 12-carrier model like that of JE06 was used. Again the carrier X(13) came out to be the strongest contributor. But this time, the contribution of X(11) increased to 11.67%, from that of just 1.68% for JE06. For simulation of HWL for other wells (except JE06 and JE04), previous LWL was introduced as an additional carrier. This was done to improve the model performance. So, in associated tables for these wells, number of carriers became 13 where the LWL had the status of X(11). The TMRs had the same sequence with X(10) representing TMR of May and X(12) that of June etc. It should be mentioned here that LWL was not used in the case of JE04, because it failed to improve the model performance significantly. And the 12-carrier model of JE06 worked excellent without the inclusion of LWL. As the number of carriers should always be kept to the possible minimum, LWL was not added to this model.

It was found from examining the %CONTS of different carriers that, for all these wells LWL played a very significant role. The secondmost important carrier was identified to be X(14) or TMR of August which was the strongest contributor when LWL was not included. One exception was noted for X(14) of JE05 which could be attributed to the same problem of multicollinearity which happened to X(11) of JE06.

For simulation of LWL, 9 carrier models were used. The first carrier X(2) stood for previous HWL, the successive carriers were TMRs of September to April to simulate the LWL occurring on May. For all the wells, X(2) or the HWL played a very significant role. Among rest of the carriers, x(3) and x(10) or TMRs of September and April were found to be dominating. Again it seemed pretty difficult to interpret the contributions physically. For example, contribution of X(3) varried from 0.5% for KT03 to 27.25% for JE05. However, from statistical point of view, such variation is immaterial unless it affects the quality of prediction by the model. And as revealed by Table 5.6 which shows the summary of model performances in terms of R^2 and s, all of the models seem to possess the characteristics required for dependable prediction or forecasting.

As pointed out in article 4.4.1.3 of Chapter 4, LWL of JE05 suffered an average lowering of about 2.884 m since 1983. To avoid this new trend in the data set, 16 years of data from 1967 to 1982 were used for simulation. Another trial run of the model was done including all the data set upto 1986. It was found that, in the former case the model performance was excellent, but for the later, quality of prediction severely deteriorated. It simply indicates that the model will work good only if the groundwater regime remains undisturbed. Hence, regression equation developed for simulation of LWL of JE05 should not be used for forecasting due to recent change in the regime condition.

About the selection of 'Best Subset', procedure described in section 5.1.7 for HWL of JE06 may be repeated for all the remaining cases under study. Required statistics generated by the full models (without deletion of any TMR, except for LWL of KT05) are given in Appendix-B. Careful examination of these statistics discloses that, in general, TMRs of December and January were the weakest contributors and hence, may be deleted from the model. Another sample analysis for variable selection by SS technique, this time for LWL of KT05, supported this postulation. For other wells, however, detailed study is recommended to exactly identify the optimum set of variables.

The model generated regression equations which have R^2 equal to or greater than 0.8 can be used for forecasting. The same procedure outlined in section 5.1.8 may be followed. As stated in the same section, the optimum subset of variables should be used for this purpose. Due to recent change in trend of the LWL of JE05, corresponding regression equation is not usable for forecasting.

Finally, examining the Table 5.6, it becomes evident that simulations of HWL were always of better quality than those of LWL. Hence, a probable conclusion may be that groundwater regime suffers some interference during the dry season.

5.3. Results and Discussion on the Study of Stream-Aquifer Interaction

The procedure to be followed to attain the objectives related to the stream-aquifer interaction study was outlined in section 4.3 of Chapter 4. To be brief, the groundwater levels at different distances from the bank line were regressed with corresponding river stages and a set of correlation coefficients (R) were derived. One such plot depicting the linear relationship between groundwater level of JE05 and stage of river at station S-171 is shown in Fig. 5.5. The correlation coefficients so obtained were indicative of the degree of sensitivity of the water table to the river stage. A plot of R versus x (distance) was then made which showed the declining trend of R with increasing x (Fig. 5.6). From this figure, the characteristic distance d_c was found to be about 2500 m. Hence this preliminary study suggests that in the G-K Project area, the groundwater table beyond a distance of about 2500 m from the bank will probably remain insensitive to the fluctuation of river stage.

The salient features and results of this study are summarized in Table 5.8

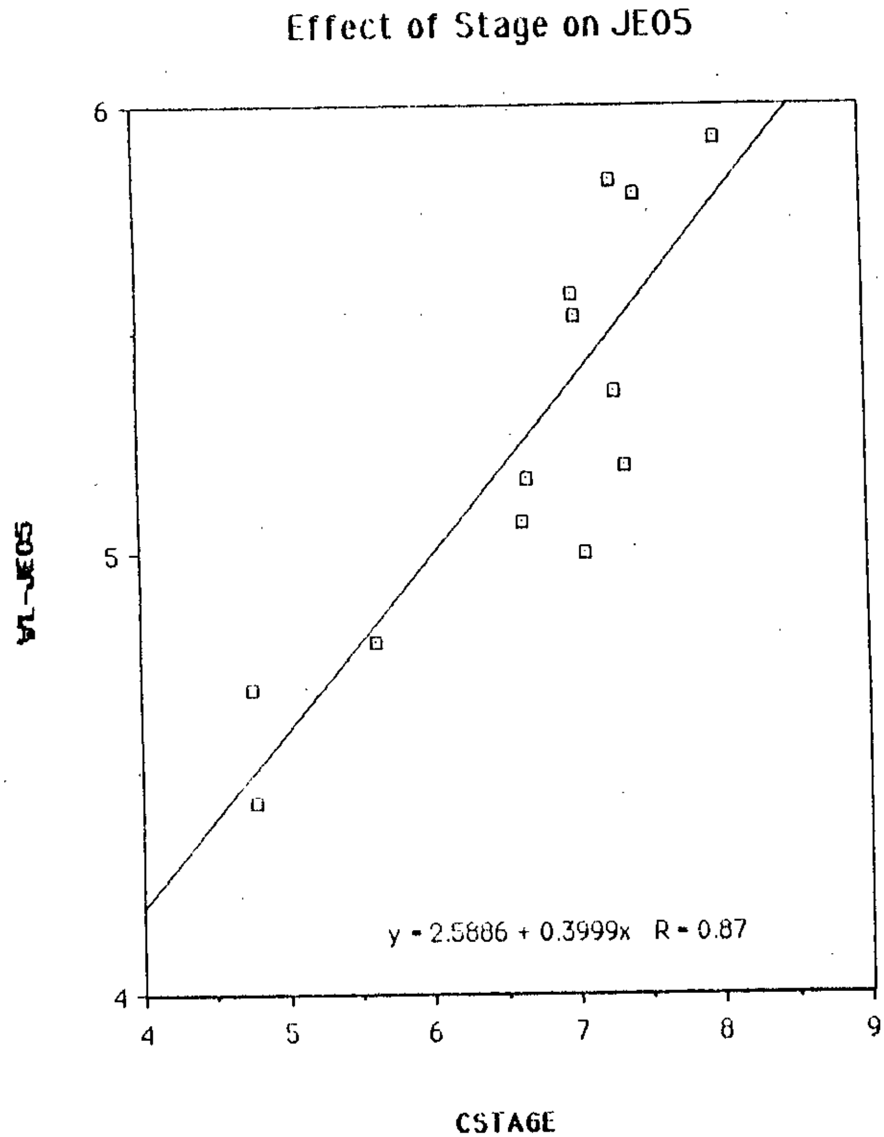


FIG. 5.5 : LINEAR RELATIONSHIP BETWEEN RIVER STAGE AT S-171 AND GROUNDWATER LEVEL OF WELL NO. JE05

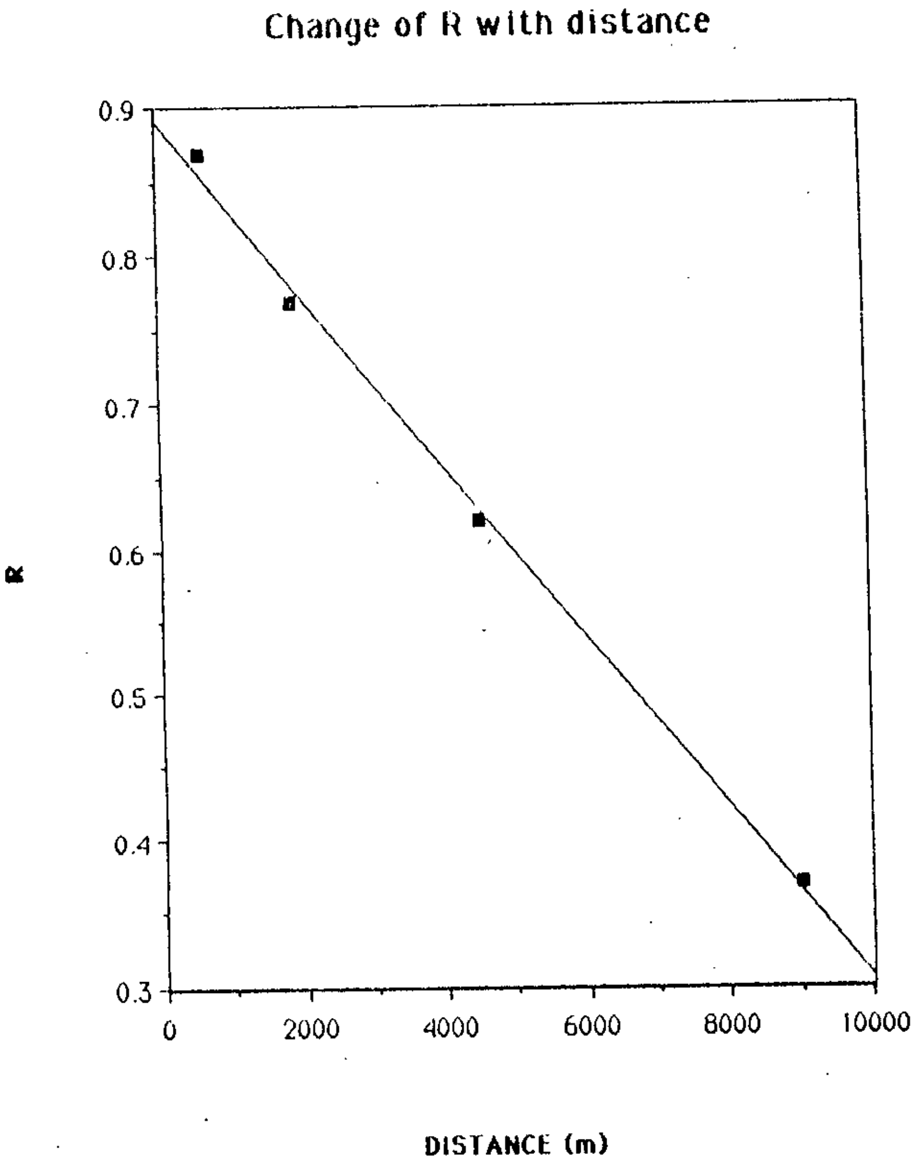


FIG. 5.6 : DECLINE OF R WITH INCREASING DISTANCE FROM THE RIVER BANK

TABLE 5.8

SUMMARY TABLE OF STREAM-AQUIFER INTERACTION STUDY

LOCATION OF RIVER STAGE MEASUREMENT	MATCHING WELL FOR GROUNDWATER LEVEL	DISTANCE OF WELL FROM THE BANK(m)	CORRELATION COEFFICIENT
S-171, on river 65, Kumar.	JE05	750	0.87
At 7.5 km downstream from S-91, on river 39, Ganges.	KT01	1875	0.77
At 33 km downstream from S-100 , on river 42, Gorai-Madhumoti.	JE25	4500	0.62
At 7.5 km downstream from S-91, on river 39, Ganges.	KT03	9000	0.37

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

6.0. Conclusions from the Multiple Regression Model and Stream-Aquifer Interaction Study

In accordance with the 'Objectives of the Research', the multiple regression model was developed and applied successfully to the six selected dugwells in the G-K Project area. The model was used to simulate both HWL and LWL. As revealed by Table 5.6, all the simulations explained significant amount of variation in the dependent variable (HWL/LWL). In general, simulations of HWLs were found to be more successful than those of LWLs. This minor weakness of the later may be attributed to the probable interference on the groundwater regime in the dry season.

The model has generated percent contributions (%CONTS) for all the carriers or independent variables for simulation of HWLs and LWLs. Such factors should be interpreted statistically, and not physically. Statistical inferences do not necessarily imply cause and effect relationship. However, contribution of a carrier which strongly contradicts the intuition or physical findings should be treated with care. Such problem may arise due to presence of multicollinearity among the carriers.

The technique of variable selection to determine the statistically 'Best Subset' was discussed in detail with a case study of HWL simulation for well no. JE06. It was pointed out that a combination of efficient algorithm and instantaneous judgement is necessary to get the optimum subset of carriers having near largest R^2 and smallest s.

The multiple regression model has generated a set of multiple regression equations, each for a particular case of HWL or LWL simulation. Possibility of using such regression equation for forecasting was discussed with several exercises using TMRs of different return periods for well no. JE06. The forecasted values of HWL and annual recharge (calculated from the difference of forecasted value of HWL and LWL) were found to be quite satisfactory when compared with the actual values. It was concluded that nine out of twelve of the regression equations could be used for forecasting. However, regression equation for simulation of LWL of JE05 should not be used for forecasting due to recent change in trend of LWL in this well.

The study of stream-aquifer interaction showed that the effect of river stage on groundwater table decayed with the increasing distance from the bank. It was found that, beyond a distance of about 2500 m from the bank, groundwater level remained statistically insensitive to the river stage.

6.1. Recommendations

The following recommendations are made in respect to the future study.

i] The model generated contribution factors should be updated with accumulation of more data. Also the best or optimum subset for all the simulations should be identified.

ii] The technique of 'Ridge Regression' or the method of 'Principal Components' should be tried to eliminate the effect of multicollinearity among the carriers and the outcomes should be compared with those of the current study.

iii] The exercises may be extended to the piezometric wells in future to evaluate the model performance.

REFERENCES

Ahmed, S., Study of groundwater system in the Mymmensingh-Tangail area using numerical methods, M.Sc. Engg.(Water Resources) thesis , BUET, Dhaka.
1986

Bakr, A.A., et al., Stochastic analysis of spatial variability in subsurface flows,1, comparison of one and three dimensional flows, Water Res. Research, 14(2):263-271.
1978

Bear, J., Hydraulics of groundwater, McGraw-Hill Book Company, New York, 557 pp.
1979

Bibby, R.,Statistical error analysis of groundwater systems, dissertation, Colorado State Univ.,Fort Collins.
1971

Black, T.C. and D.L.Freyberg, Stochastic modeling of vertically averaged concentration uncertainity in a perfectly stratified aquifer, Water Res. Research, 23(6):997-1004.
1987

Bogardi, J., Impact of precipitation and temperature on changes of water table, Hidrologai Kozlony, Nos 11-12.
1953

Boonstra,J., and N.A. De Ridder, Numerical modeling of groundwater basin, ILRI, Wageningen, The Netherlands.
1981

Bouwer, H., Groundwater hydrology, McGraw-Hill, 480 pp.
1978

Brebbia, C.A., The boundary element method for engineers, Halsted Press, New York, 189 pp.
1978

Chalky, J.N., et al., A method for estimating volume-surface ratios, Science 110:295p.
1949

Chatterjee, S., and B. Price, Regression analysis by example, John Wiley, New York.
1977

Chaudhury, M.T., et al. Optimal conjunctive use model for Indus basin, J. of Hydraulics Divn.,ASCE,100(HY5):667-687.
1974

Cochran, G.F., and W.S.Butcher, Dynamic programming for optimal conjunctive use, Water Res. Research, 6(3).
1970

Cooley, R.L., et al., Some application of statistical methods to groundwater flow system analysis, Tech. Rept. Series H-W, Hydrology and Water Resources Pub. No.14, Nevada.
1973

Csomane, S.K., Determination of graundwater levels of different
1968 probability, Vizugyi Kozlem, No. 2.

Dagan, G., Stochastic modeling of groundwater by unconditional
1982 and conditional probablities, Water Res. Research, 18(4):835-848.

Daniel, C., and F. Wood, Fitting equations to data (2nd ed.),
1980 John Wiley, New York.

Davis, G.B., and T.R.Hoffmann, Fortran'77: A structured,
1985 disciplined style...., McGraw-Hill Book Company.

de Jong de Josselin, G., The tensor characteristics of the dis-
1969 persion coefficient in anisotropic porous media, Paper no.412, Symposium on the fundamentals of transport phenomena in porous media, p.7, Technion.

Devore, J.L., Probablity and statistics for engineering and
1982 sciences, Brooks/Cole Pub. Co., California, 640 pp.

Domenico, P.A., et al., Optimal groundwater mining, Water Res.
1968 Research, 4(2):247-255.

Dracap, J.A., The optimal use of groundwater and surface water
1966 system : a parametric linear programming approach, Thec. Rept. 6-24, Water Res. Center Contribution no.107, Univ. of California, Barkely.

Draper, N.R., and H.Smith, Applied regression analysis(2nd ed.),
1981 John Wiley & Sons, New York.

Eshett, A., Groundwater system analysis, dissertation,
1970 Colorado State University, Fort Collins, USA.

Flores, W.E.Z., Stochastic management of stream aquifer system,
1976 Advances in Groundwater Hydrology, Ed. Z.A.Saleem, AWWA, 309-328 pp.

Forgo, L., Predicting the trend of groundwater level, Vizugyi
1968 Kozlem, NO. 4.

Freeze, R.A., and J.A.Cherry, Groundwater, Prentice Hall, New
1979 Jersy.

Gelhar, L.W., and C.L.Axness, Three dimensional stochastic
1983 analysis of macrodispersion in aquifers, Water Res. Research, 19(1):161-180.

Glover, 1974 R.E., Transient groundwater hydraulics, Dept. of Civil Engg., Colorado State Univ., Fort Collins.

Gorelick, 1984 S.M., et al., Aquifer reclammation design : The use of cotaminant transport simulation combined with nonlinear programming, Water Res. Research, 20(4):415-427.

Haan, 1977 C.T., Statistical methods in hydrology, Iowa State Univ., Ames, 378 pp.

Haq, 1987 K.A., and M.A.Sattar, Groundwater table fluctuation and response to rainfall on a microlevel basin, BRRI research paper, Joydevpur, Dhaka.

Heidari, 1982 M., Application of linear systems theory and linear programming to groundwater managemant in Kansas, Water Res. Research, 18(6):1003-1012.

Hoque, 1986 M.M., Conjunctive use of groundwater and surface water in Ganges-Kobadak Project, 1st interim report, IFCDR, BUET.

IECO, 1964 EPWAPDA Master Plan, Dahka.

IECO, 1980 South-West regional plan : a consultancy report for BWDB, Dhaka, Bangladesh.

Jones, 1987 LaDon, et al., Optimal control of nonlinear groundwater hydraulics using differencial dynamic programming, Water Res. Research, 23(11):2091-2096.

Karim, 1984 M.A., Upzilla-wise groundwater recharge condition of Bangladesh, Groundwater Circle, BWDB, Dhaka.

Khan, 1982 L.R., Modeling of groundwater recharge and yield, Ph.D thesis, Univ. of Newcastle upon Tyne, U.K.

Khan, 1984 L.R., and J.A.Mawdsley, Assessment of aquifer yield using linear programming technique, J. of IEB,2(3):9-16.

Khan, 1986 L.R., and J.A.Mawdsley, Groundwater yield assessment of a stream-aquifer system using a lumped model, Proceedings of groundwater workshop, Roorkee, India.

Kleinecke,D., Use of linear programming for estimating
1971 geohydrologic parameters of groundwater basins, Water Res. Resources, vol.7, 367-374 pp.

Labadie, J.W., et al., A river basin network model for conjunc-
1983 tive use of surface and ground water, Report NO. 125, WRRI, CSU, Fort Collins.

McWorther,D.B., et al., Analytical model of stream-aquifer
1972 systems with pumping and recharge, Completion Report, CSU, Fort Collins.

McWhorter,D.B., and D.K.Sunada, Groundwater hydrology and
1977 hydraulics, Water Res. Pub., Fort Collins, Colo.,290pp.

Michael, A.M., Conjunctive use of ground and surface water for
1986 irrigation in Bangladesh, BARC, Dhaka.

Morris, W.J., and et al., Combined surface water groundwater
1972 analysis of hydrological systems with the aid of hybrid computer, Water Res. Billetin, vol.8, 63-76 pp.

Mosteller,F., and J. Tukey, Data analysis and regression,
1977 Addision-Wesley, Reading, Mass.

MPO, National Water Plan Project, second interim report,
1984 vol.III, Master Plan Organization, Dhaka.

Pinder, T.A., and W.G.Gray, Finite element simulation in surface
1977 and subsurface hydrology, Academic Press, NY, 295pp.

Prickett,T.A., Modeling techniques for groundwater evaluation,
1975 Advances in Hydroscience, Academic Press, New York.

Reddell, D.L., and D.K.Sunada, Numerical simulation of dispersion
1967 in groundwater aquifers, CSU, Hyd. Paper No. 41, 79p.

Remson, I., et al., Numerical methods in subsurface hydrology,
1971 John Wiley & Sons, New York, 389 pp.

Rethati, L., Determination of design groundwater level,
1970 Elotervezes-Melyepites, Budapest.

Rethati, L., Groundwater in civil engineering, Elsevier
1983 Scientific Pub. Co., Amsterdam, 478 pp.

Rice J.R., Numerical methods, software, and analysis, McGraw-
1983 Hill Int. Book Copany.

Saleh, 1985 A.F.M., Hydrological evaluation of irrigation from groundwater : a case study in Bangladesh, Ph.D. Thesis, Dept. of Agricultural Engg., Cornell Univ., Ithaca.

Scheideggar, 1954 A.E., Statistical Hydrodynamics in porous media, Journal of Applied Physics, vol.25, no.8.

Scheideggar, 1974 A.E., The physics of flow through porous media, (3rd ed.), Univ. of Toronto, 353 pp.

Sing, 1981 J., A multiple regression model for prediction of average water table depth in lower part of Ghaggar Basin, ICAE, Paper No. 225, Bangkok.

Thomas, 1973 R.G., Groundwater models, FAO Pub. No.21, The United Nations, Rome.

Todd, 1980 D.K., Groundwater hydrology(2nd ed.), John Wiley and Sons, New York, 535 pp.

Ubell, 1953 K., Groundwater storage induced by precipitation, Vizgyi Kozlem, No. 11.

Vemuri, 1969 V., and W.J.Karplus, Identification of nonlinear parameters of groundwater basins by hybrid computation, Water Res. Research, vol.5, 172-185 pp.

Walton, 1970 W.C., Groundwater resources evaluation, McGraw-Hill, New York, 664 pp.

Wang 1982 H.F., and M.P.Anderson, Introduction to groundwater modeling, W.H.Freeman and Co., San Fransisco, 237 pp.

Willis, 1985 R.A., A planning model for the management of groundwater hydraulics : Theoritical developments and numerical experiments, Water Res. Research, 21(10):1476-1482.

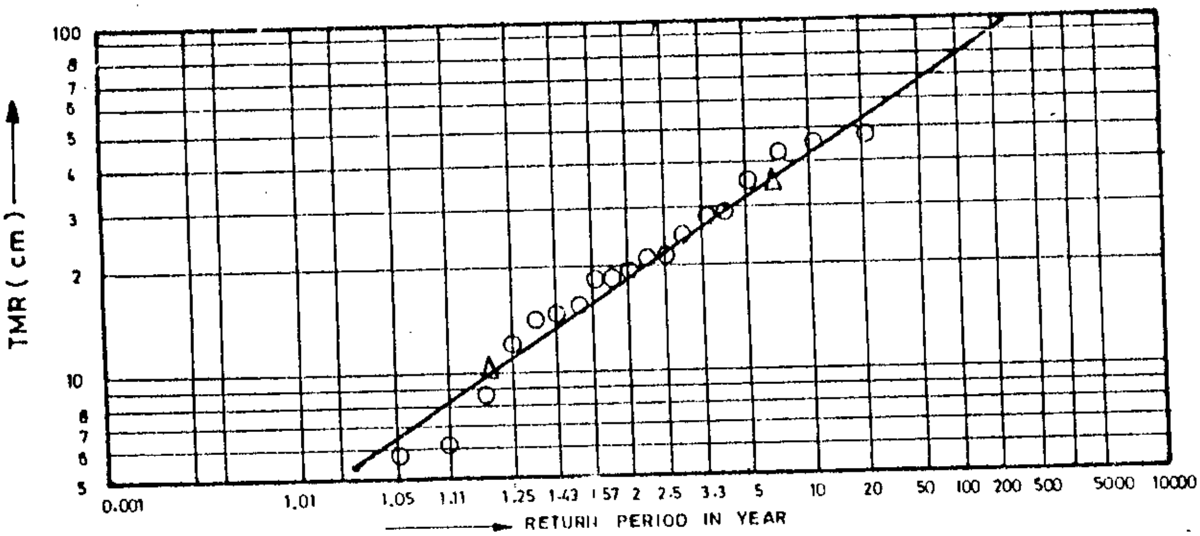
Yevjevich, 1972 V.M., Probablity and statistics in hydrology, Water Res. Pub., Fort Collins, Colorado.

APPENDIX - A

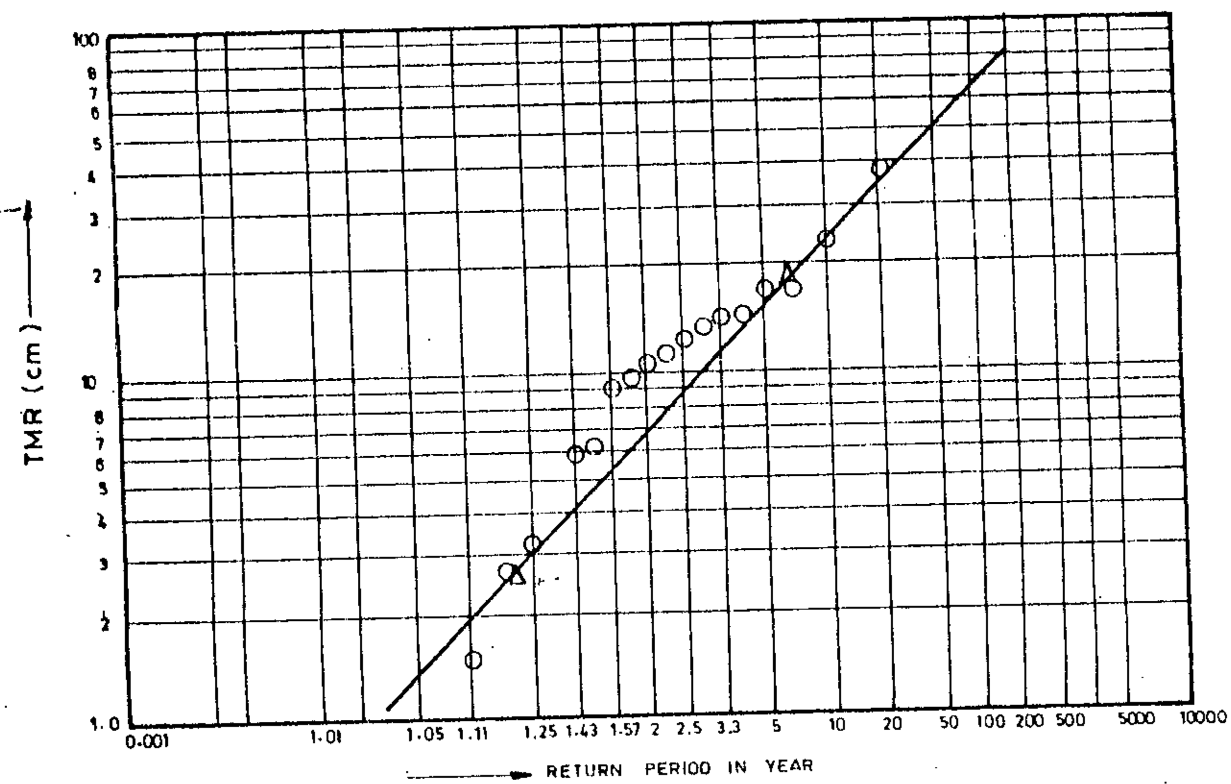
Probability Plots of Total Monthly Rainfalls for

Stations R-463, R-460 and R-452

LOG NORMAL PROBABILITY PAPER

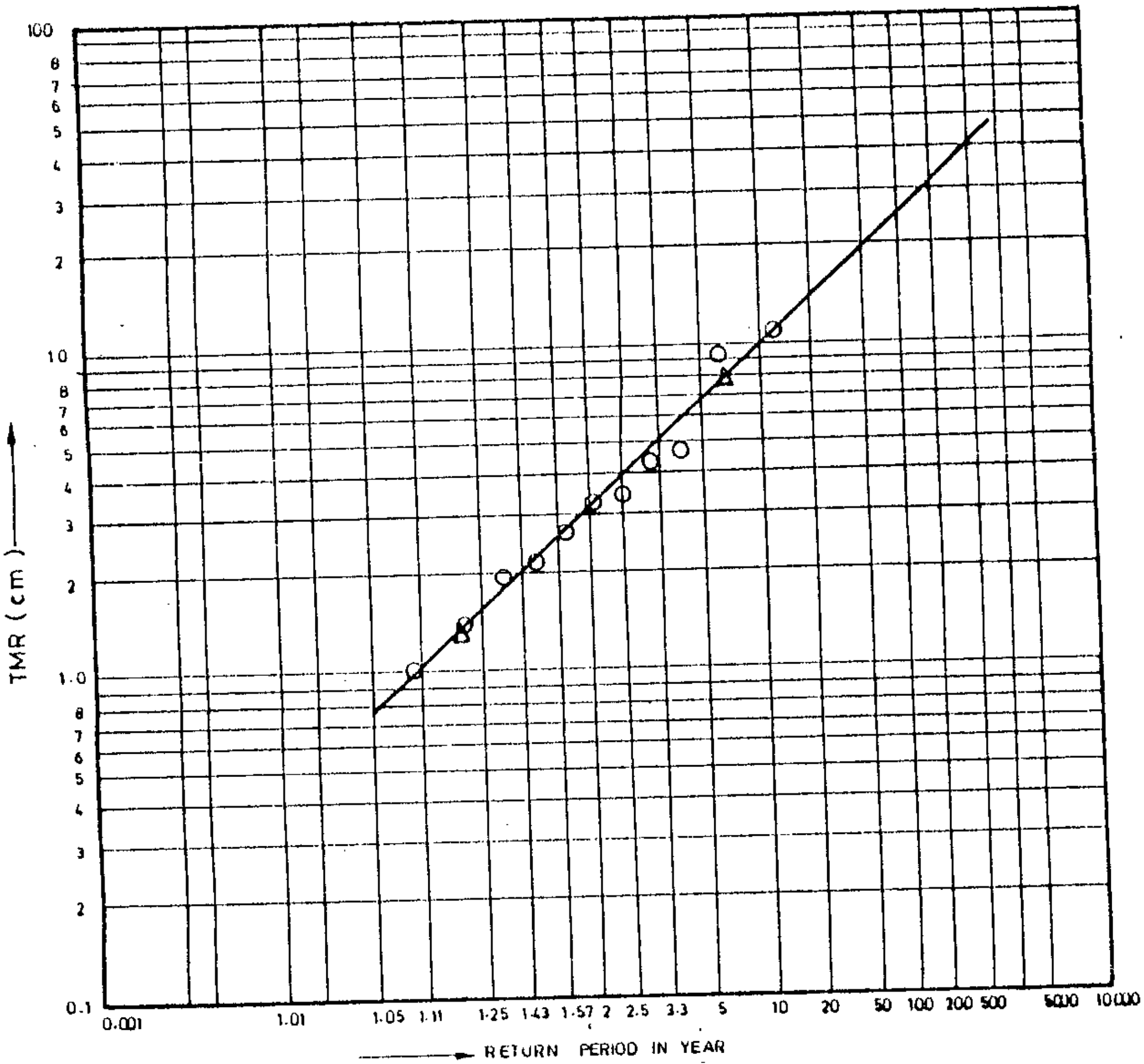


PROBABILITY PLOT OF SEPTEMBER / R-452



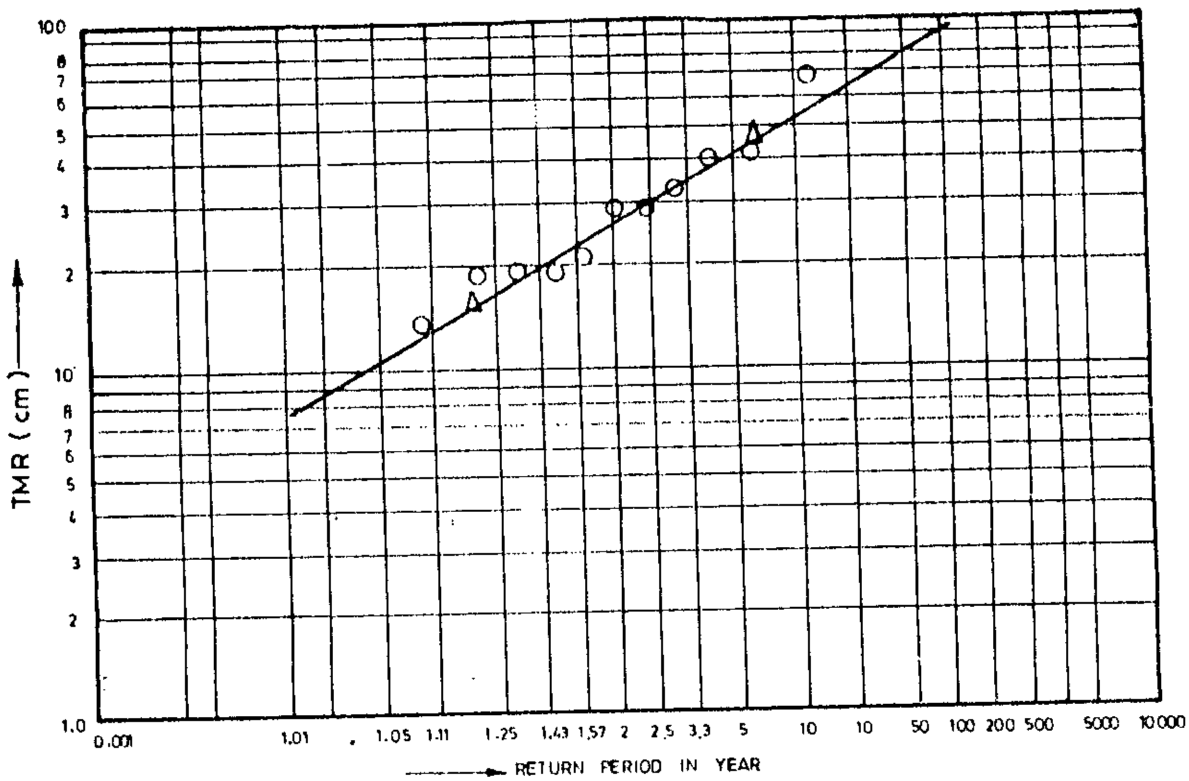
PROBABILITY PLOT OF OCTOBER / R-452

LOG NORMAL PROBABILITY PAPER

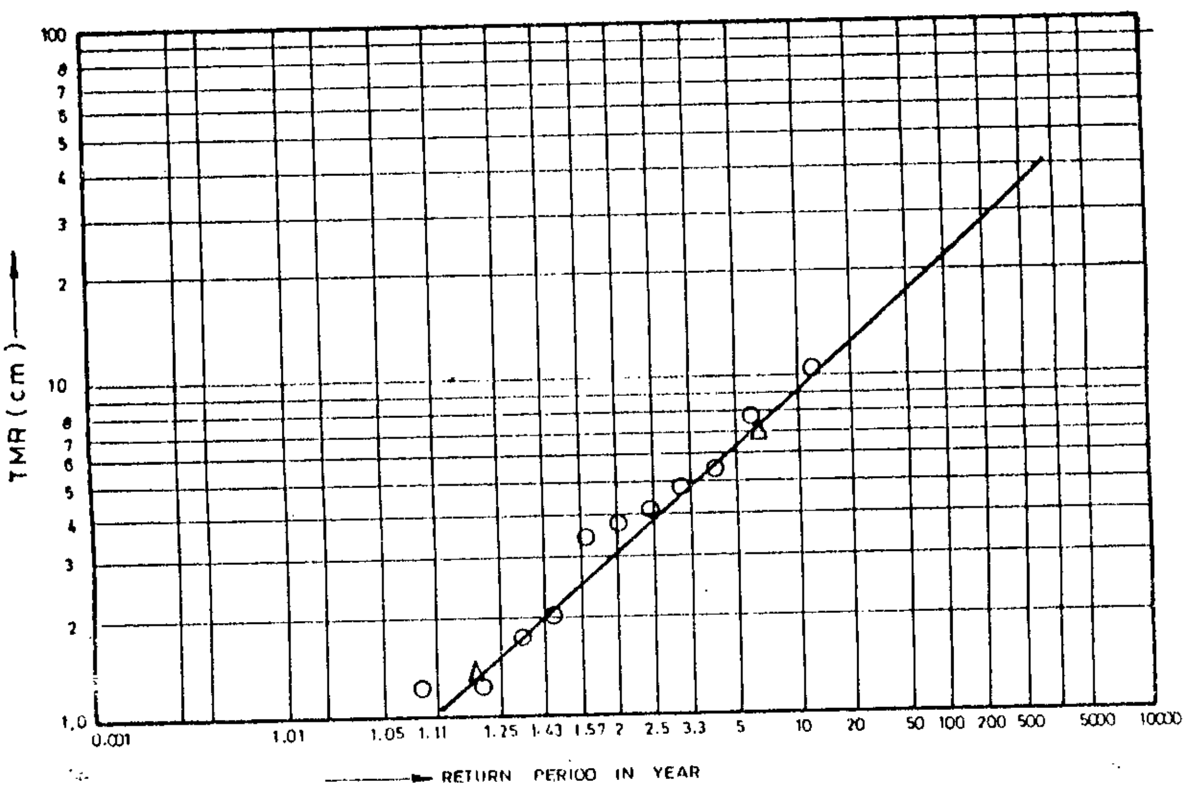


PROBABILITY PLOT OF NOVEMBER / R-452

LOG NORMAL PROBABILITY PAPER

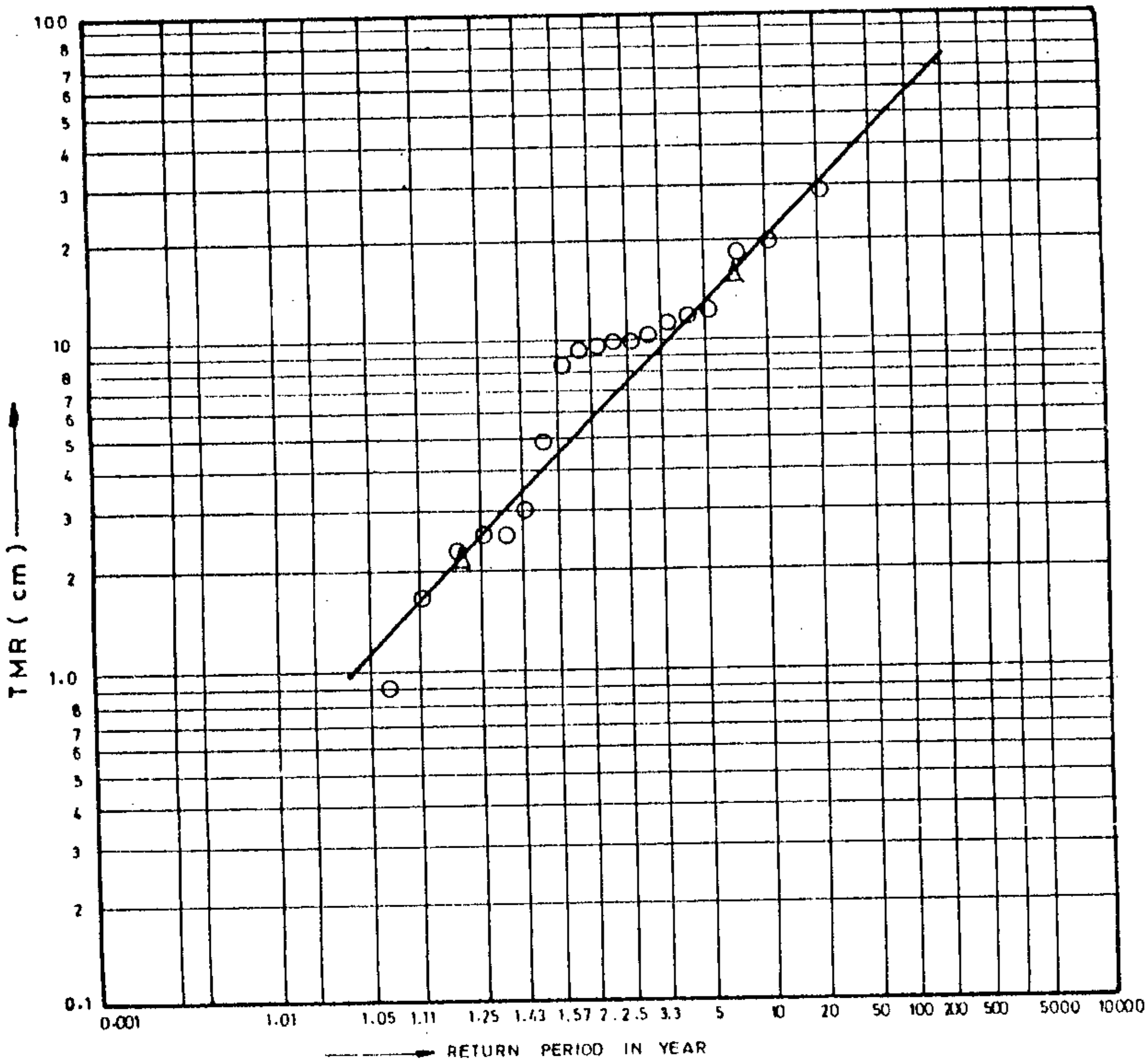


PROBABILITY PLOT OF FEBRUARY / R - 452



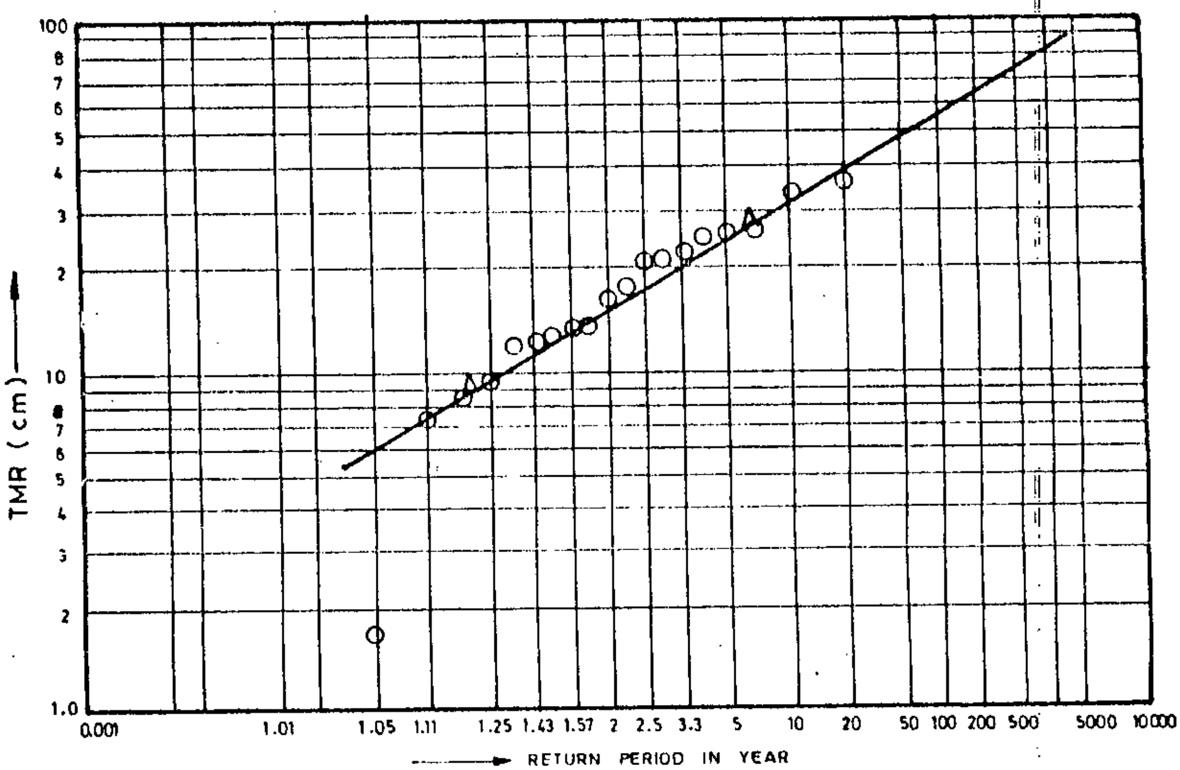
PROBABILITY PLOT OF MARCH / R - 452

LOG NORMAL PROBABILITY PAPER

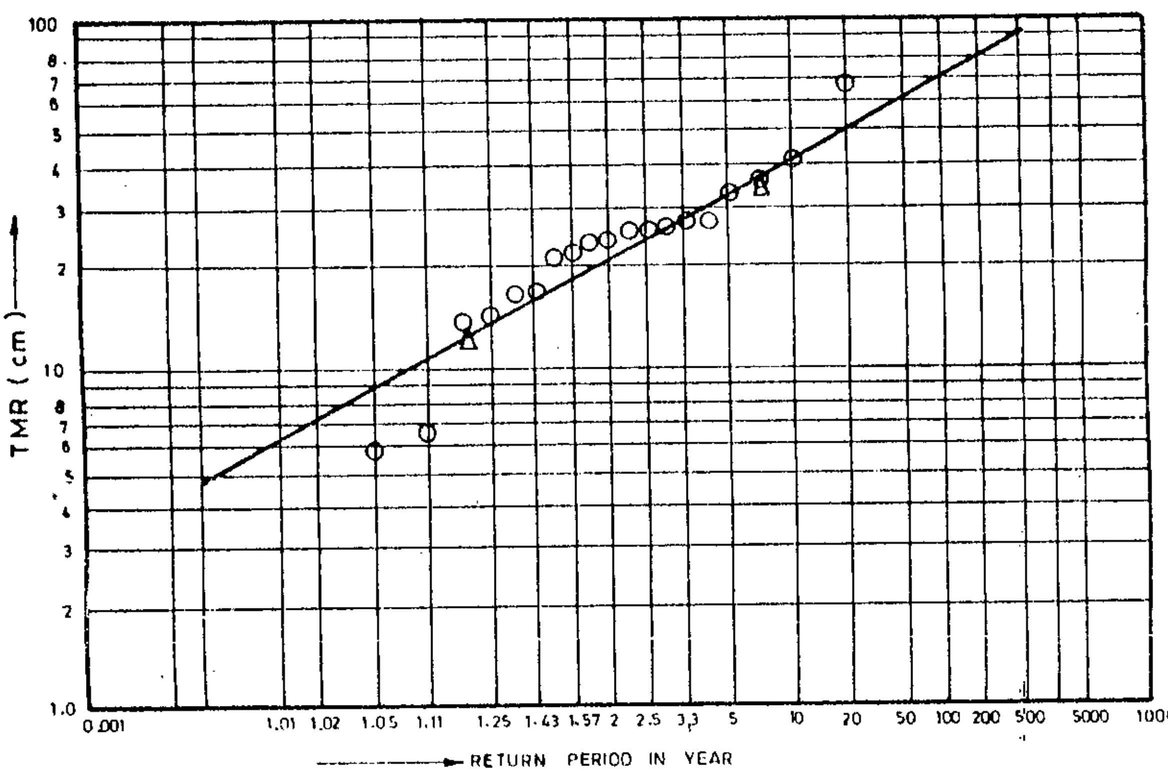


PROBABILITY PLOT OF APRIL / R - 452

LOG NORMAL PROBABILITY PAPER

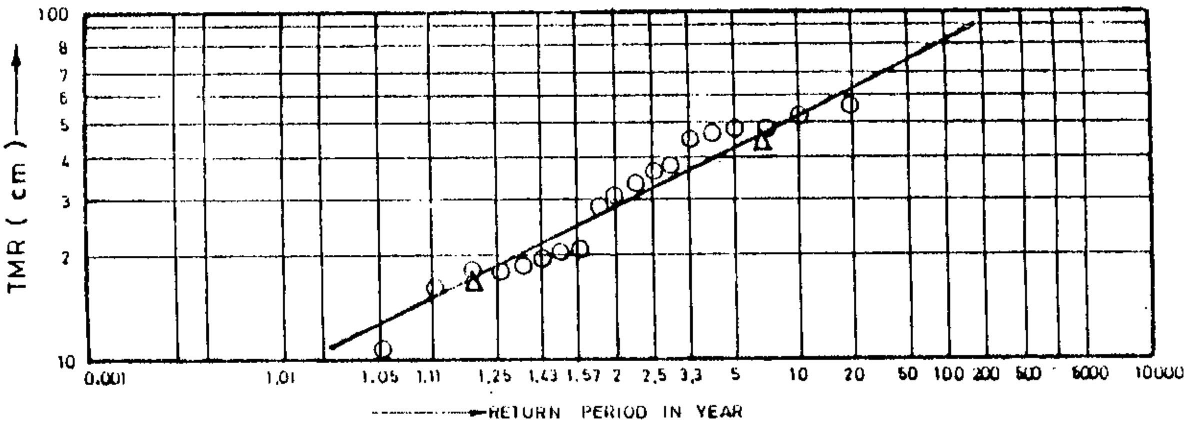


PROBABILITY PLOT OF MAY / R-452

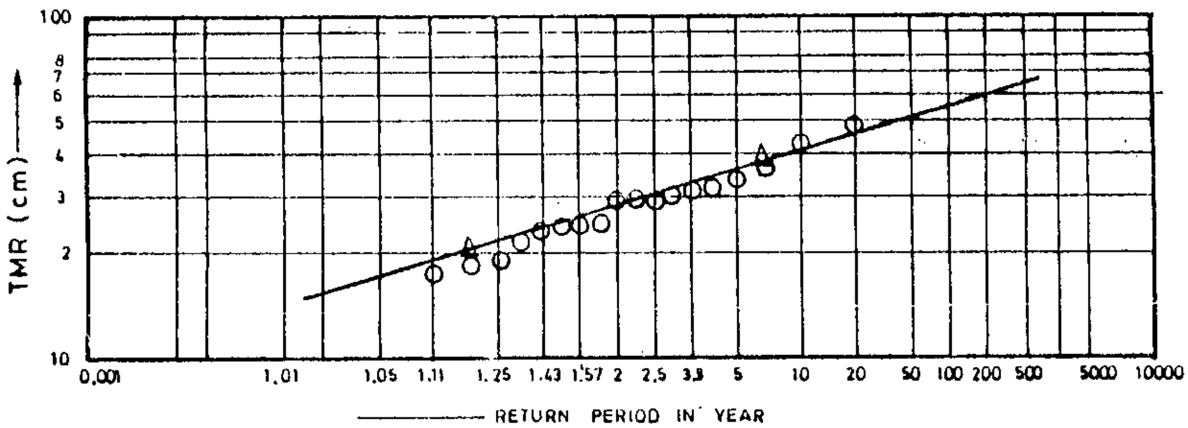


PROBABILITY PLOT OF JUNE / R-452

LOG NORMAL PROBABILITY PAPER

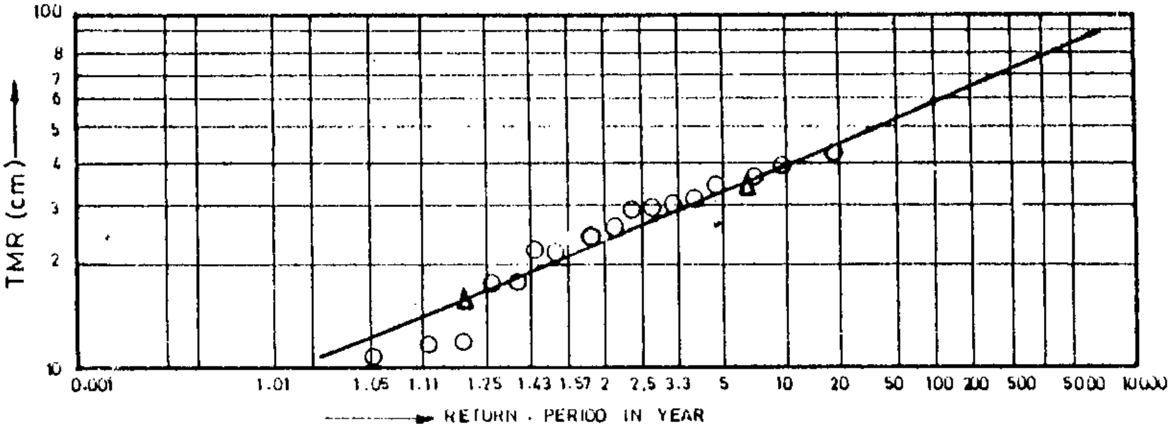


PROBABILITY PLOT OF JULY / R - 452

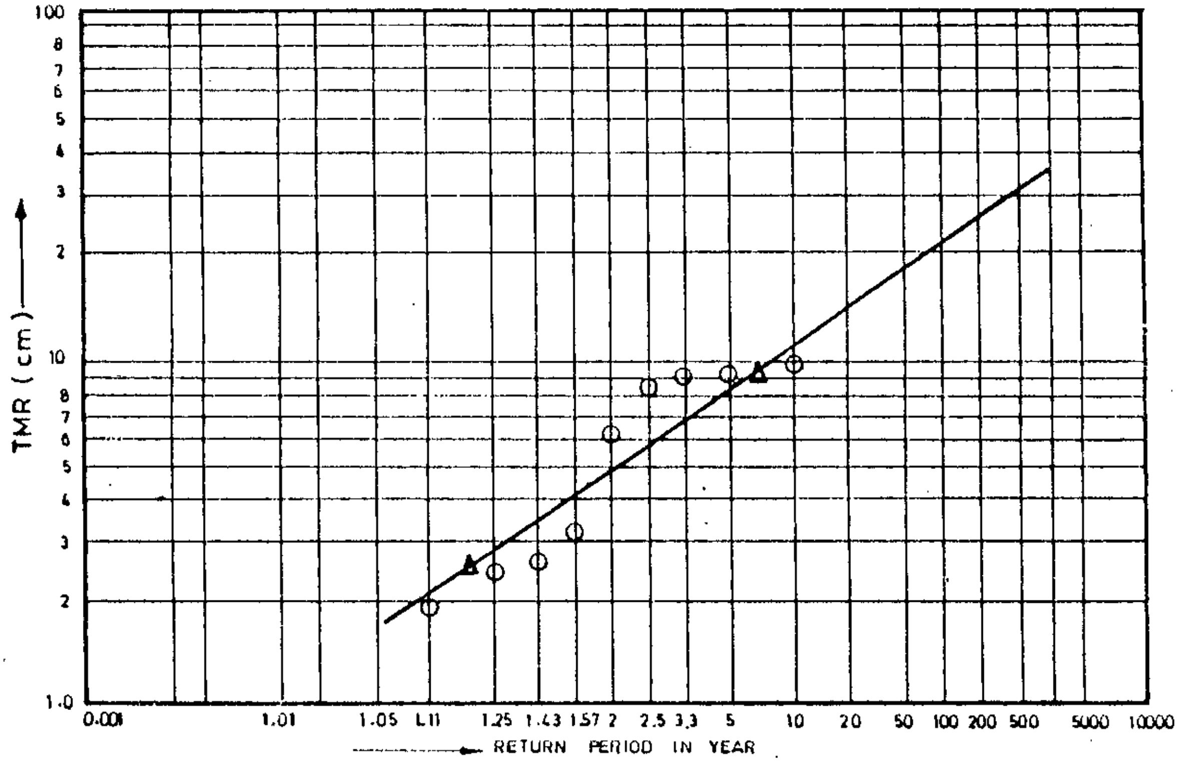


PROBABILITY PLOT OF AUGUST / R-452

LOG NORMAL PROBABILITY PAPER

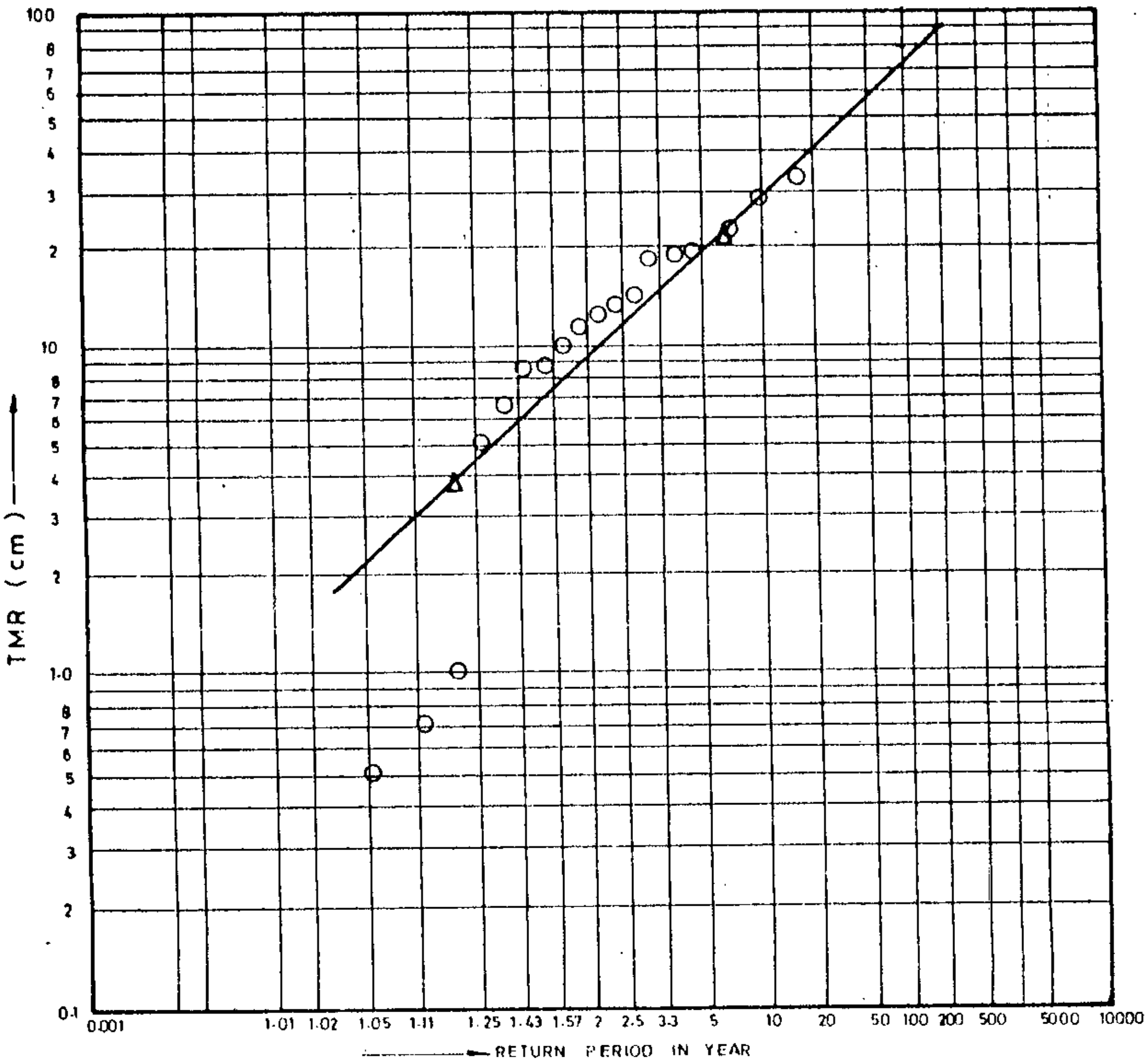


PROBABILITY PLOT OF SEPTEMBER / R-460



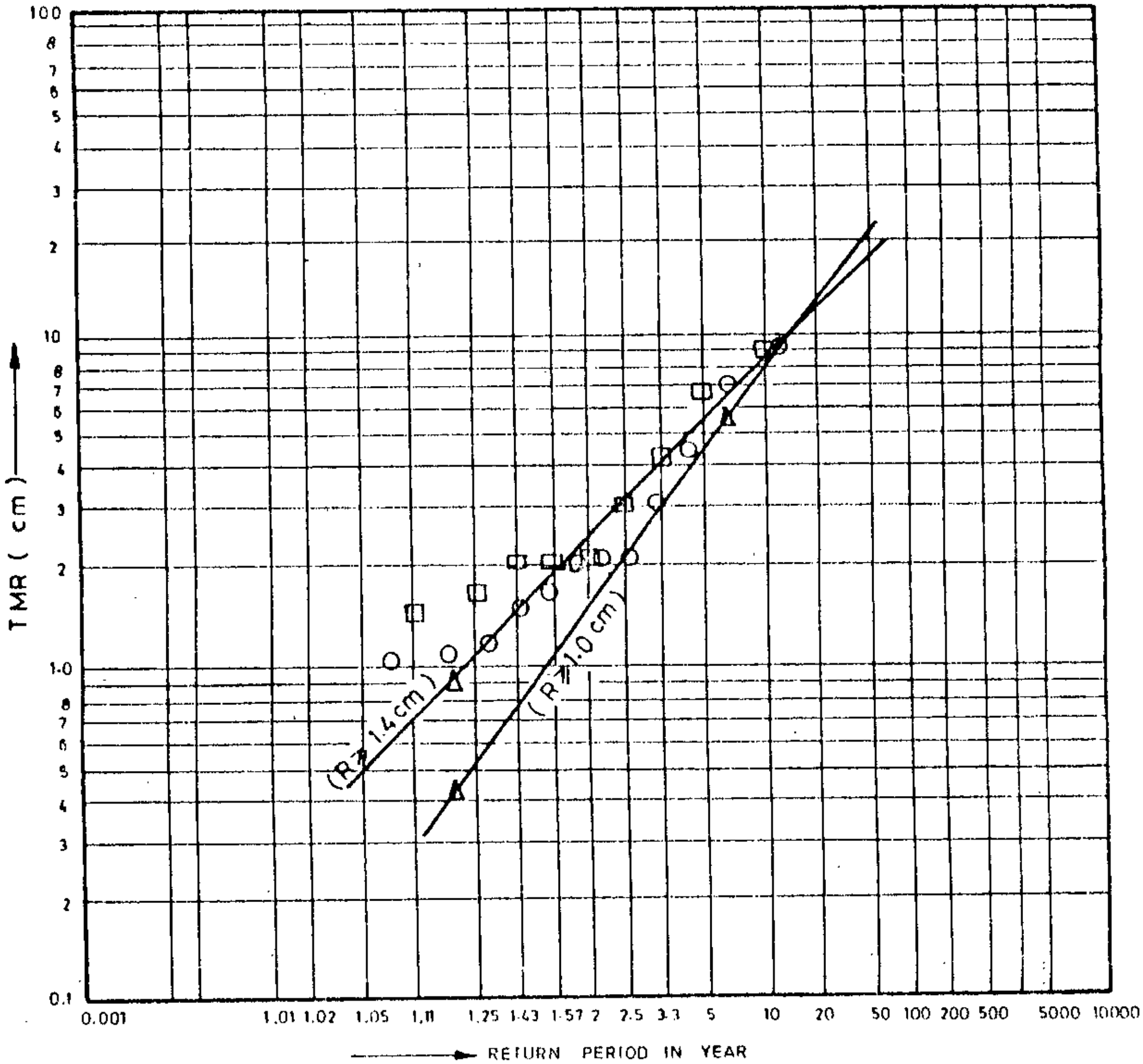
PROBABILITY PLOT OF NOVEMBER / R-460

LOG NORMAL PROBABILITY PAPER



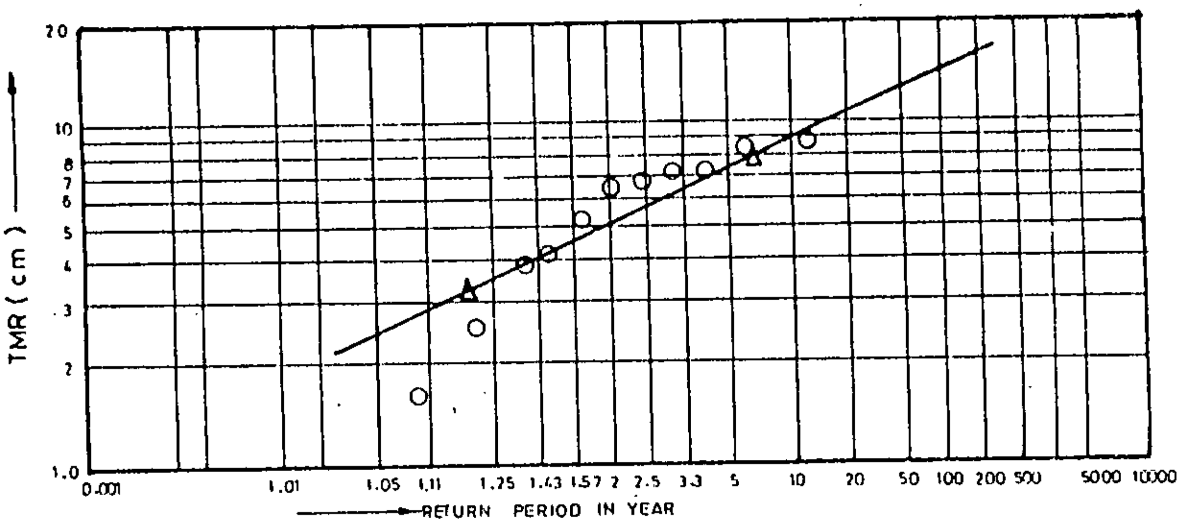
PROBABILITY PLOT OF OCTOBER / R-460

LOG NORMAL PROBABILITY PAPER

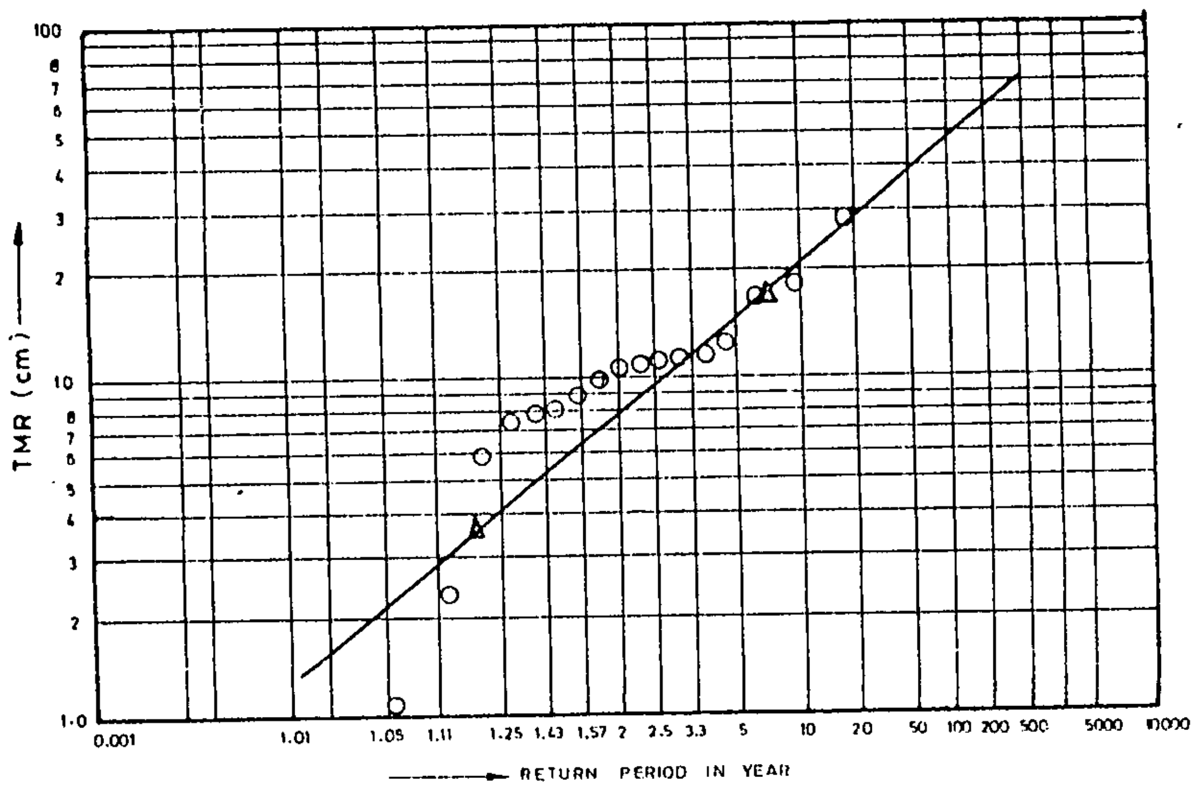


PROBABILITY PLOT OF FEBRUARY / R-460

LOG NORMAL PROBABILITY PAPER

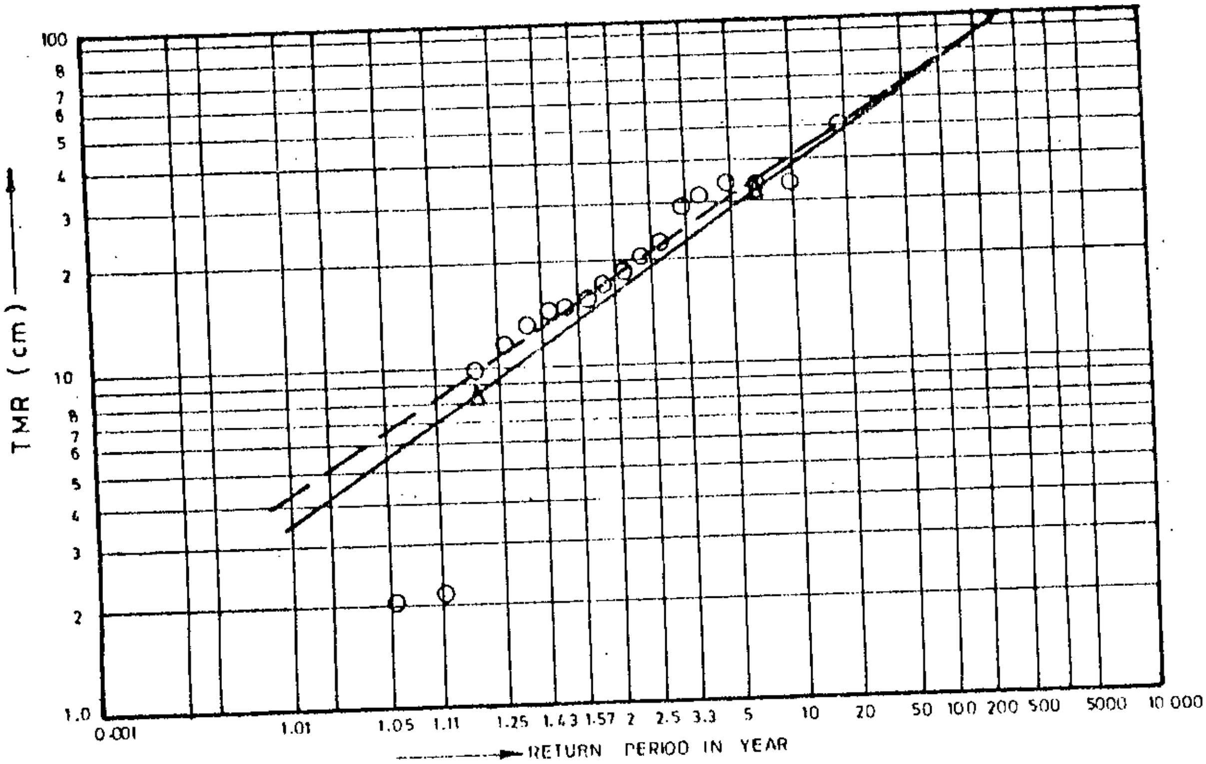


PROBABILITY PLOT OF MARCH / R-460

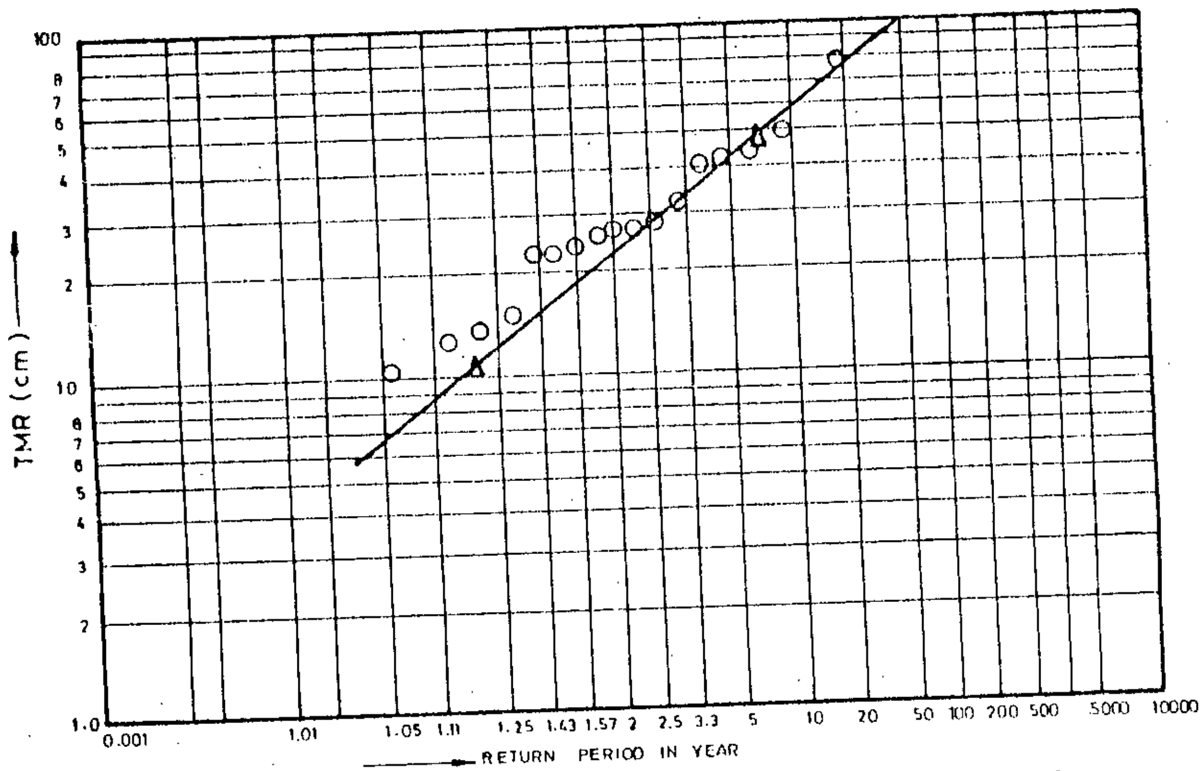


PROBABILITY PLOT OF APRIL / R-460

LOG NORMAL PROBABILITY PAPER

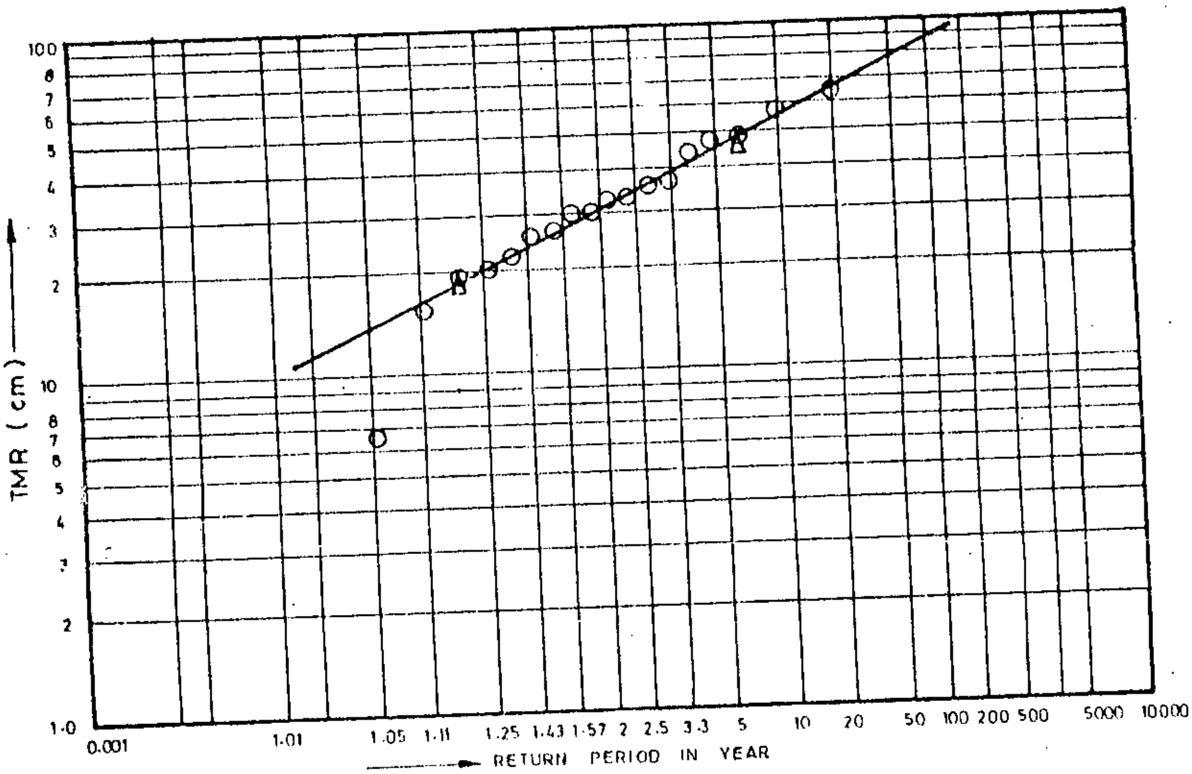


PROBABILITY PLOT OF MAY / R-460

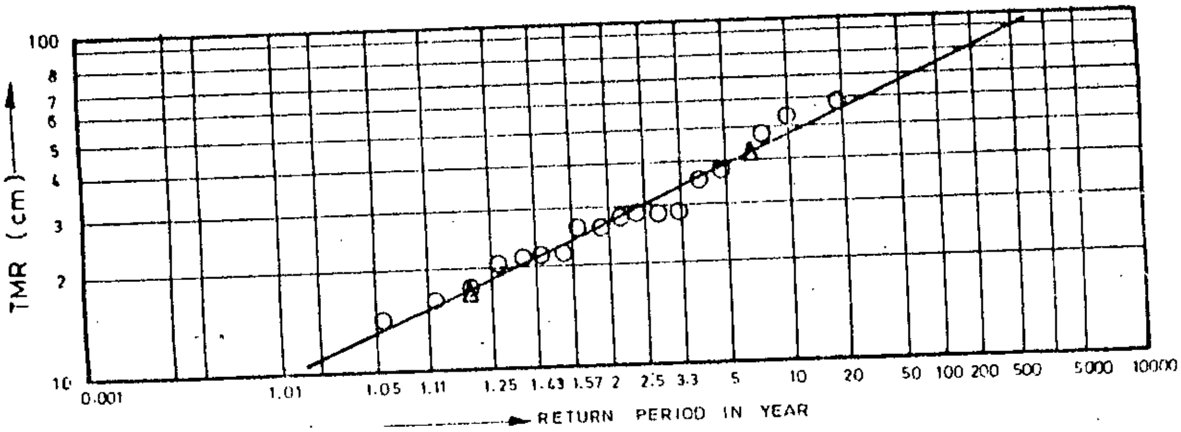


PROBABILITY PLOT OF JUNE / R-460

LOG NORMAL PROBABILITY PAPER

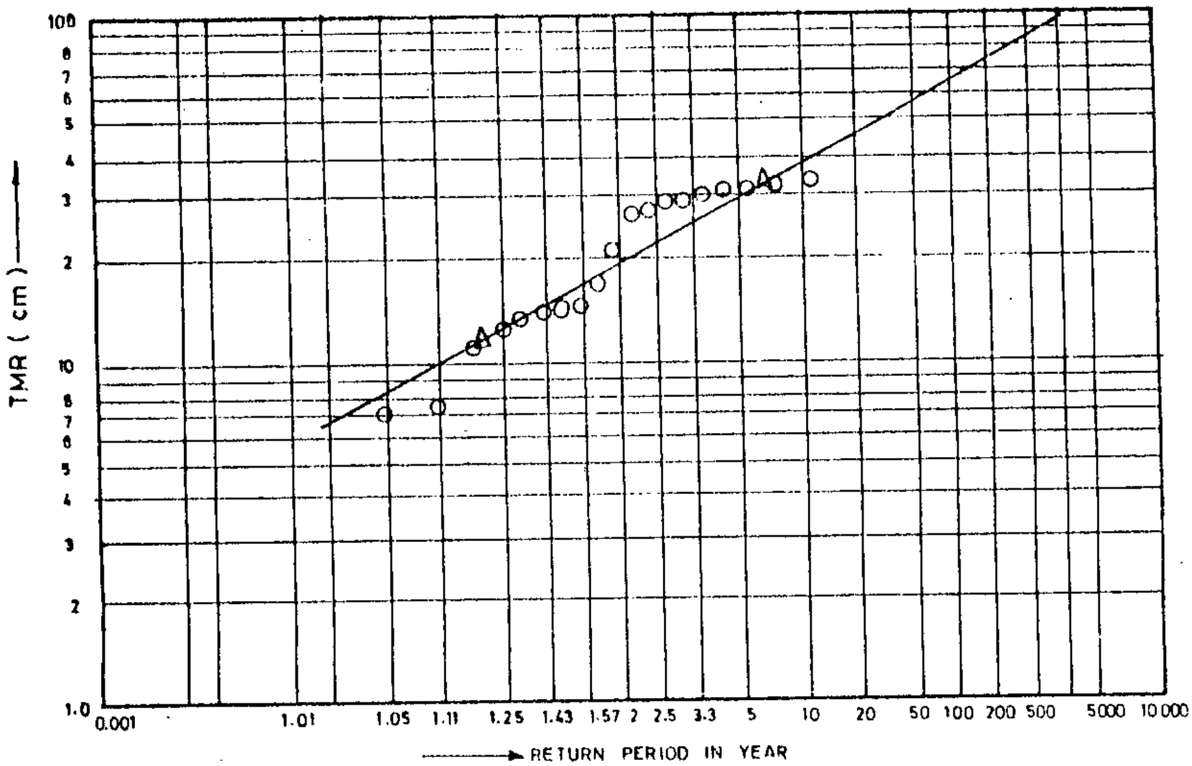


PROBABILITY PLOT OF JULY / R - 460

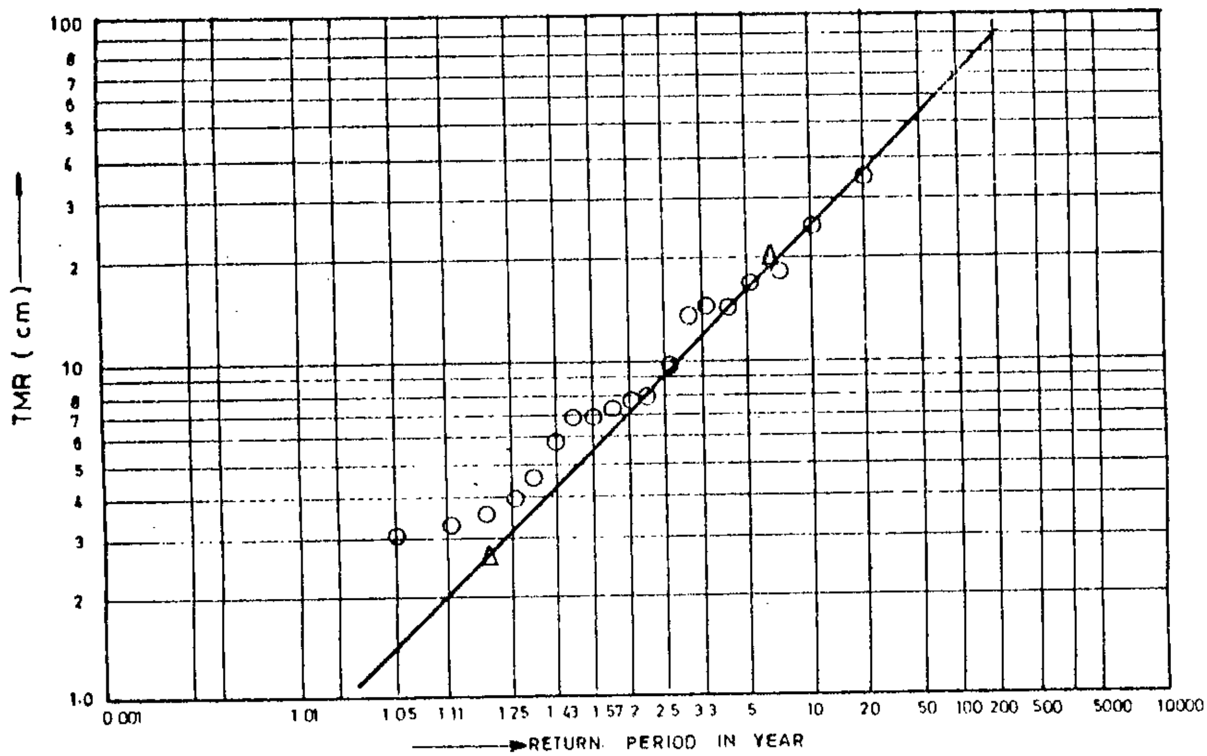


PROBABILITY PLOT OF AUGUST / R-460

LOG NORMAL PROBABILITY PAPER

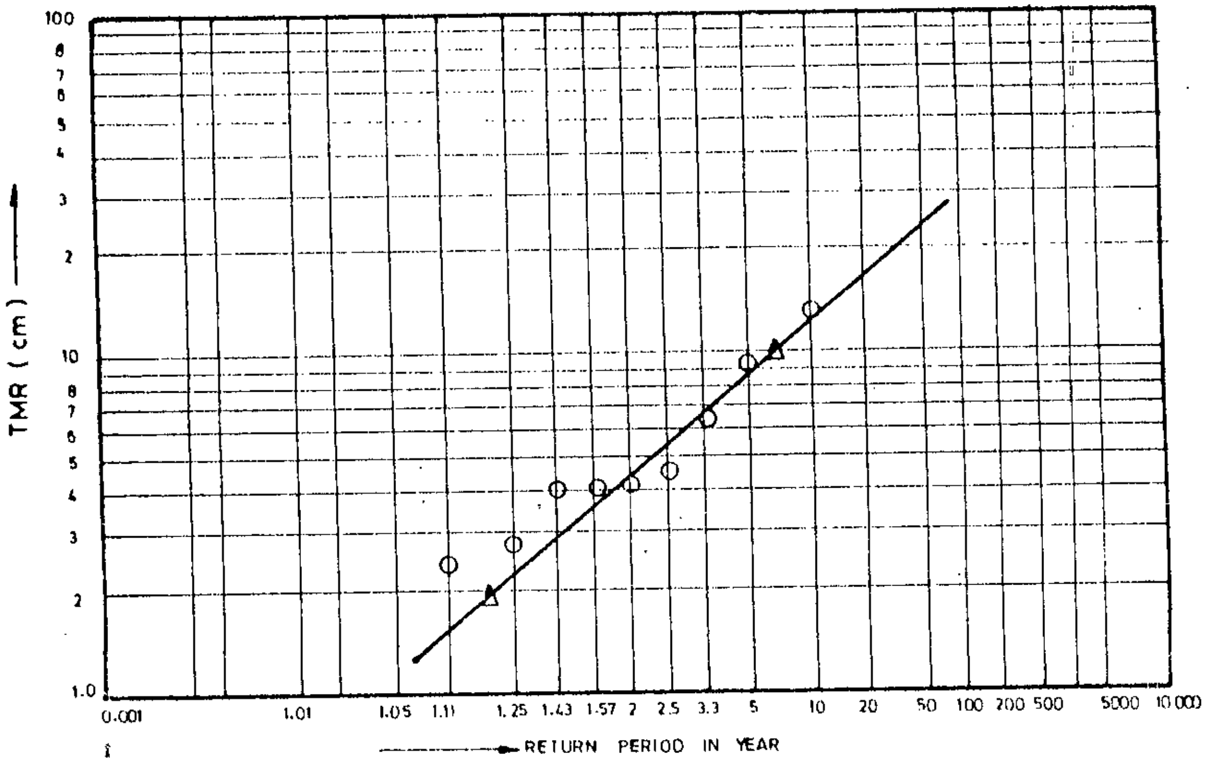


PROBABILITY PLOT OF SEPT. / R-463.

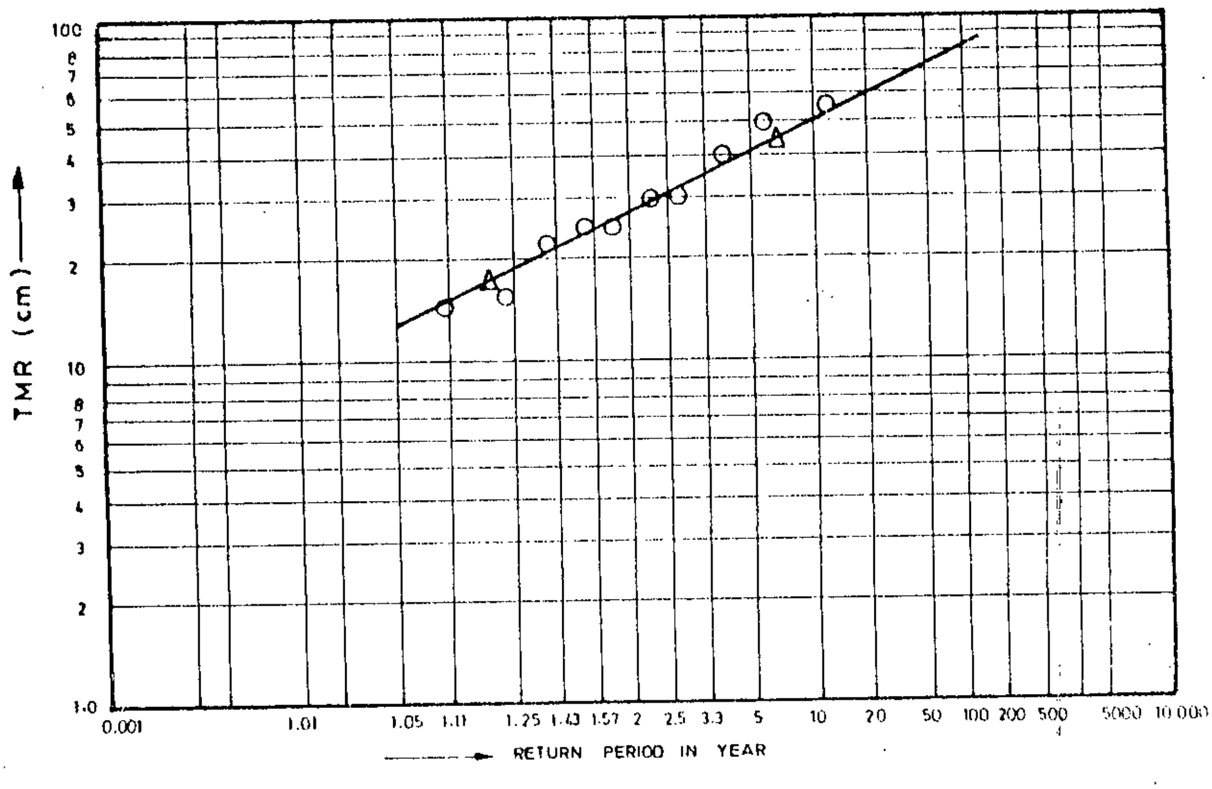


PROBABILITY PLOT OF OCT. / R-463

LOG NORMAL PROBABILITY PAPER

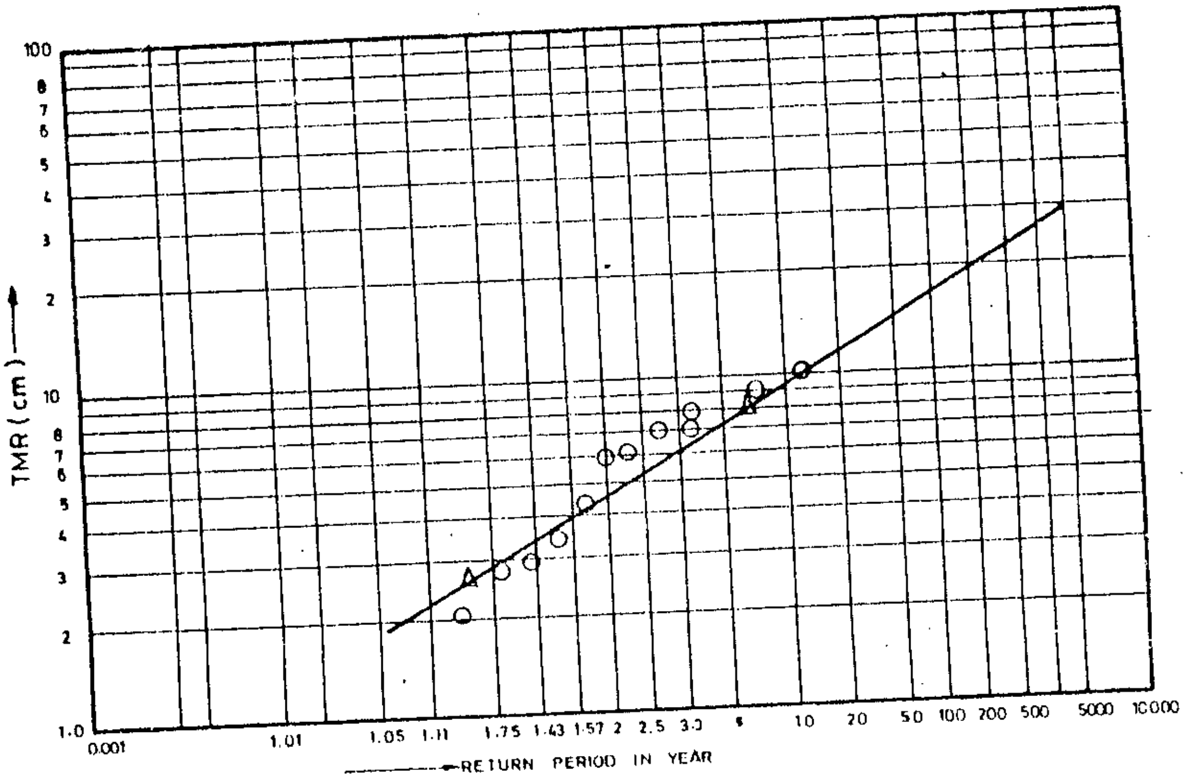


PROBABILITY PLOT OF NOV./R-463

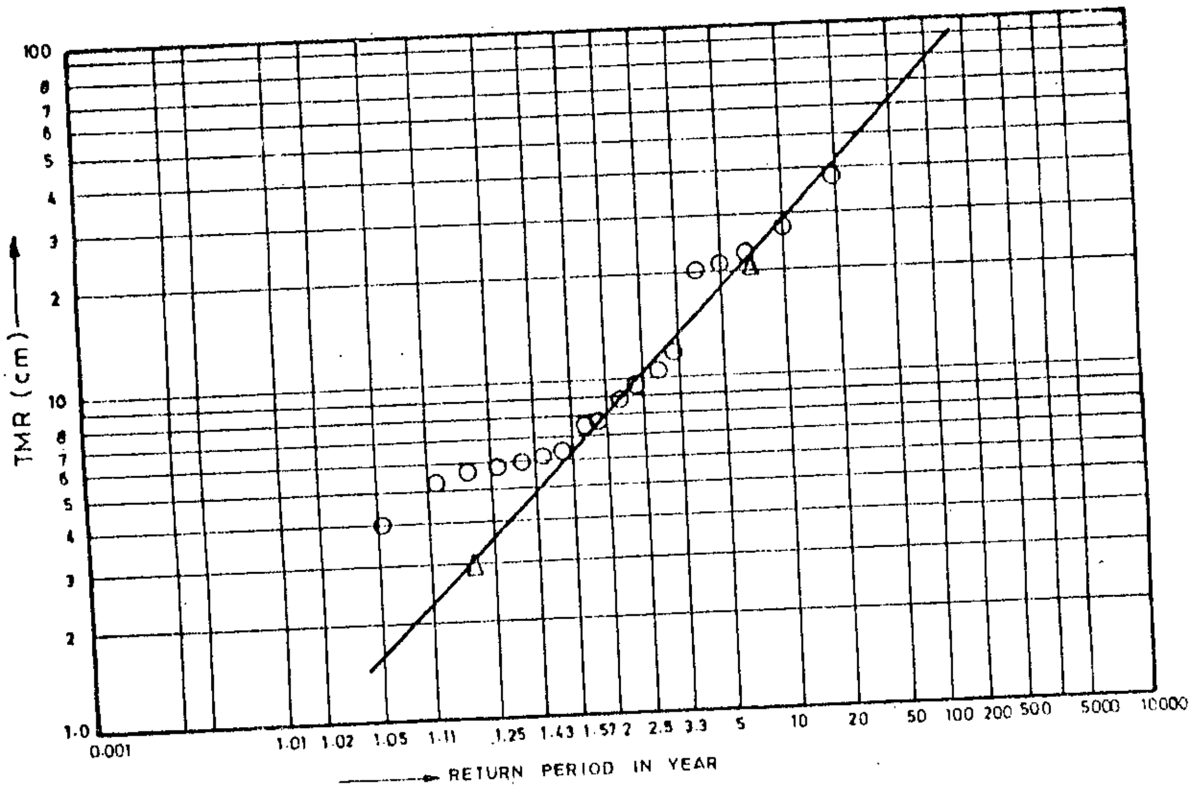


PROBABILITY PLOT OF FEB. /R- 463

LOG NORMAL PROBABILITY PAPER

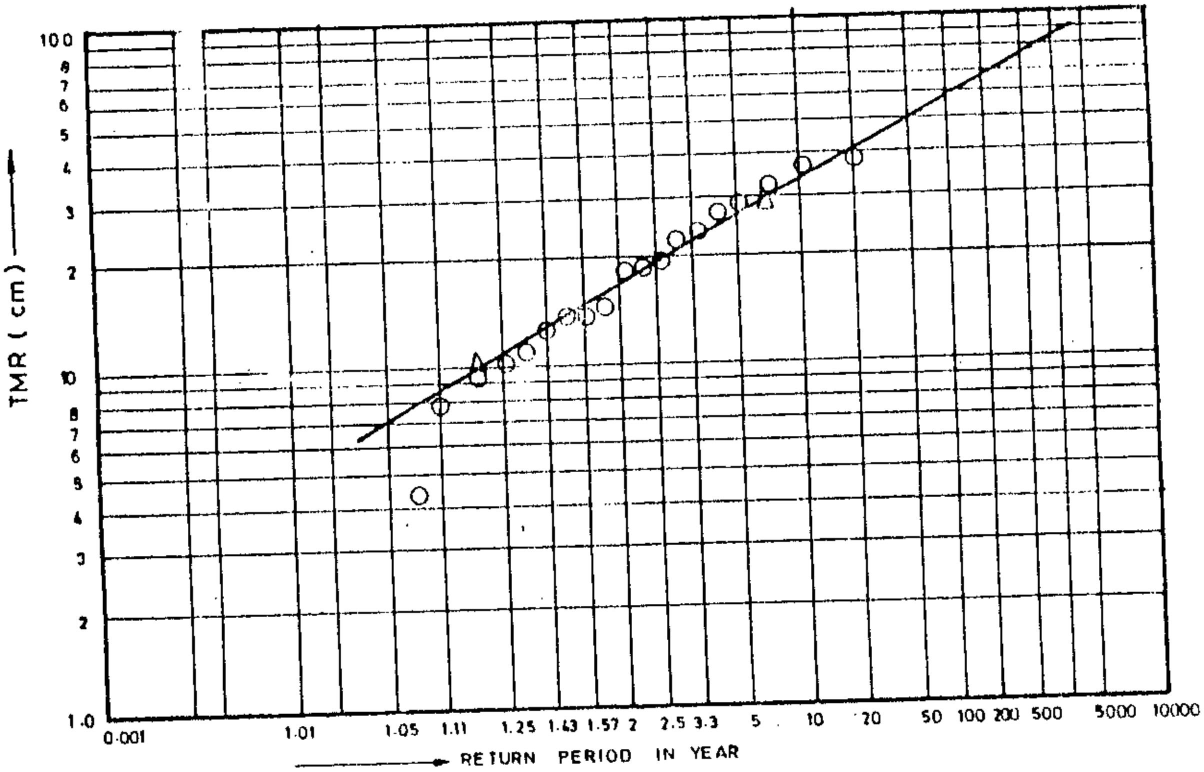


PROBABILITY PLOT OF MAR / R-463

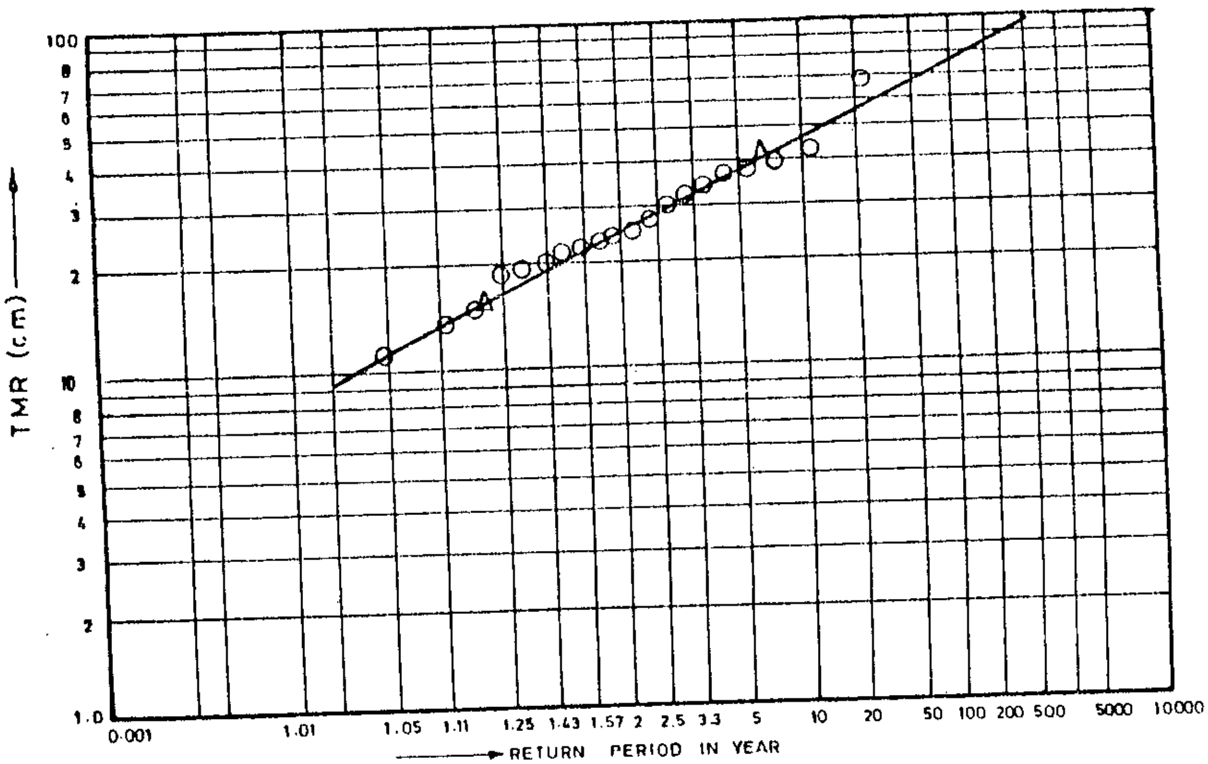


PROBABILITY PLOT OF APRIL / R- 463

LOG NORMAL PROBABILITY PAPER

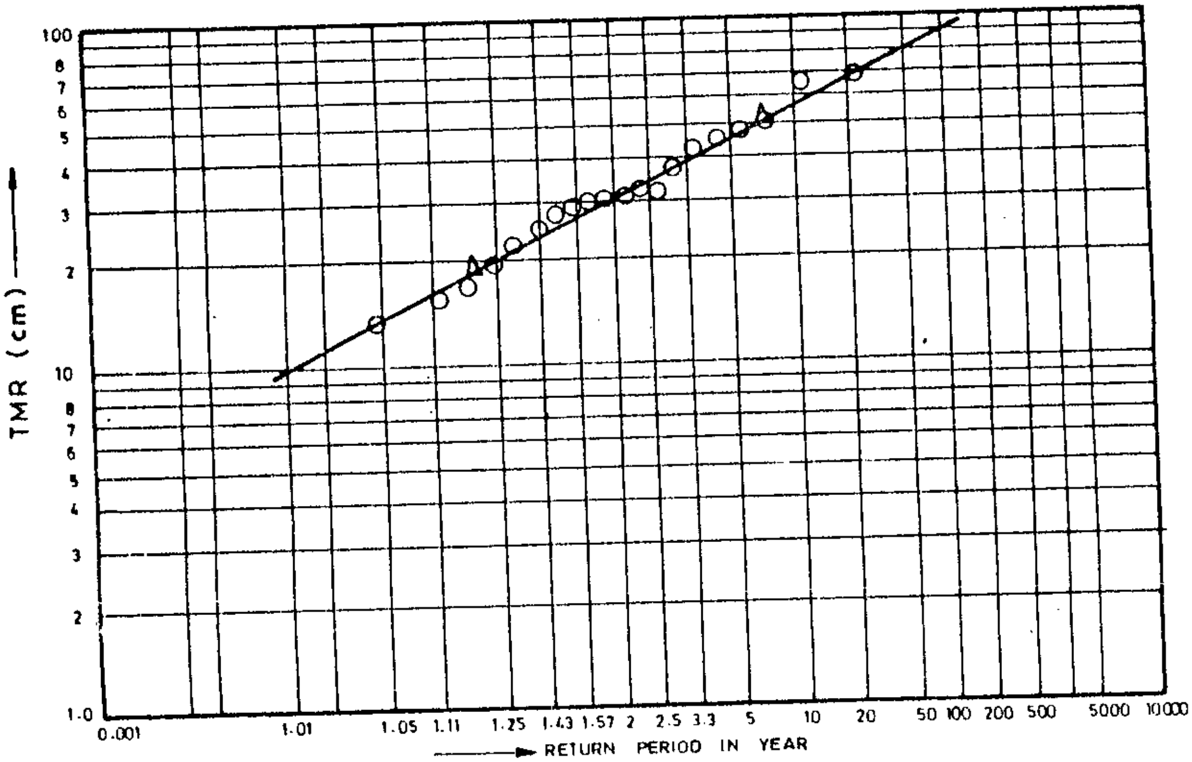


PROBABILITY PLOT OF MAY / R - 463

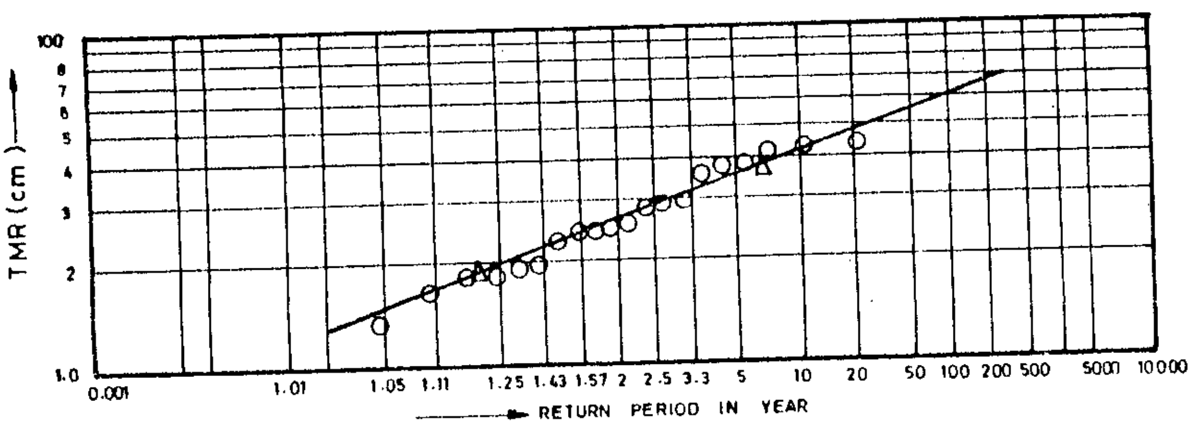


PROBABILITY PLOT OF JUNE / R - 463

LOG NORMAL PROBABILITY PAPER



PROBABILITY PLOT OF JULY / R-463



PROBABILITY PLOT OF AUGUST / R-463

APPENDIX - B

Simulation Tables for HWLs and LWLs of the Six Selected Wells in the G-K Project Area

SIMULATION OF HWL OF JE06 USING THE BEST SUBSET - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	24.92	10.97	41.78	9.18	.11	2.71
X(3)	13.02	.51	33.27	9.21	.57	3.38
X(4)	4.62	.00	30.61	7.46	2.72	11.82
X(5)	2.12	.00	9.19	2.43	1.97	6.99
X(6)	3.52	.00	8.81	3.32	.28	1.89
X(7)	10.05	.00	27.55	6.44	.96	5.57
X(8)	20.48	2.06	49.66	12.12	.63	3.92
X(9)	28.24	.00	76.76	17.45	1.15	5.73
X(10)	32.91	6.39	65.41	14.74	.50	3.64
X(11)	29.08	14.40	57.20	11.94	1.12	4.04
YF	5.004	3.426	5.913	.637	-.842	4.214

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	3.7353	***	***	.0000	.0000
X(2)	-.0253	.631	14.932	.2326	.0999
X(3)	.0318	.414	9.787	.2928	.1257
X(4)	.0112	.052	1.220	.0833	.0358
X(5)	-.0691	.146	3.461	.1681	.0722
X(6)	-.0492	.173	4.095	.1636	.0702
X(7)	-.0403	.405	9.583	.2595	.1114
X(8)	.0257	.526	12.452	.3114	.1337
X(9)	-.0044	.124	2.925	.0764	.0328
X(10)	.0164	.540	12.778	.2419	.1038
X(11)	.0418	1.216	28.768	.4995	.2145

SIMULATION OF HWL OF JE06 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	3.7353	3.7353	.2139	17.4633
2	-.0253	-.0253	.0052	-4.8506
3	.0318	.0318	.0045	7.0990
4	.0112	.0112	.0074	1.5118
5	-.0691	-.0691	.0250	-2.7701
6	-.0492	-.0492	.0129	-3.8227
7	-.0403	-.0403	.0091	-4.4096
8	.0257	.0257	.0042	6.1271
9	-.0044	-.0044	.0031	-1.3952
10	.0164	.0164	.0027	6.0204
11	.0418	.0418	.0045	9.3353

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SCS
MEAN	1.0000	450.6903
REGRESSION	10.0000	6.7569
RESIDUAL	7.0000	.1418
TOTAL	18.0000	457.5889

MSSQ= .0203
MS = .1423
MRSQ= .9794
MR = .9897
F = 33.3630

SIMULATION OF HWL OF JE06 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	5.507	5.510	-.003	-.048	5.281	5.738	.0145	.7161
2	4.923	5.103	-.180	-3.648	4.938	5.267	.0075	.3714
3	5.608	5.641	-.033	-.594	5.428	5.854	.0126	.6208
4	4.700	4.494	.206	4.386	4.367	4.620	.0044	.2186
5	5.913	5.928	-.015	-.261	5.692	6.165	.0155	.7662
6	4.450	4.406	.044	.979	4.240	4.573	.0077	.3791
7	5.404	5.428	-.024	-.439	5.208	5.648	.0134	.6611
8	5.660	5.521	.139	2.452	5.345	5.698	.0086	.4252
9	5.026	5.099	-.073	-1.451	4.874	5.324	.0141	.6949
10	4.750	4.769	-.019	-.409	4.544	4.995	.0141	.6966
11	5.150	5.211	-.061	-1.188	4.955	5.467	.0182	.8984
12	5.026	5.114	-.088	-1.748	4.950	5.278	.0074	.3665
13	5.150	5.109	.041	.797	4.917	5.301	.0102	.5026
14	4.036	4.143	[illegible]	-2.644	3.906	4.379	.0155	.7632
15	4.452	4.405	[illegible]	1.046	4.153	4.658	.0177	.8717
16	3.426	3.417	[illegible]	.617	3.127	3.697	.0173	.8530
17	5.104	5.021	[illegible]	1.635	4.840	5.201	.0090	.4447
18	5.784	5.720	.064	1.112	5.486	5.954	.0152	.7495

The critical HAT(I) value is = .2222

SIMULATION OF HWL OF KT01 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	24.61	9.14	97.20	19.37	3.17	14.67
X(3)	12.00	2.39	34.45	8.50	1.44	5.31
X(4)	1.46	.00	6.68	2.02	1.32	4.23
X(5)	1.13	.00	5.92	2.01	1.63	4.53
X(6)	.62	.00	5.41	1.31	3.16	13.51
X(7)	1.75	.00	6.32	1.80	1.22	4.38
X(8)	2.72	.00	10.82	3.37	1.50	4.72
X(9)	8.48	1.40	25.08	7.01	1.32	4.46
X(10)	17.55	.63	39.29	10.53	.41	3.19
X(11)	10.77	10.21	11.65	.48	.65	2.43
X(12)	23.36	7.06	57.37	13.34	1.35	4.87
X(13)	29.71	9.82	65.61	15.34	1.03	4.39
X(14)	25.75	12.86	59.03	12.21	1.48	5.21
YF	12.476	12.236	12.927	.216	.677	2.658

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	5.1614	***	***	.0000	.0000
X(2)	.0039	.095	1.105	.0747	.0489
X(3)	-.0159	.191	2.218	.1351	.0885
X(4)	.0342	.050	.581	.0691	.0452
X(5)	-.0266	.030	.350	.0536	.0351
X(6)	-.0179	.011	.130	.0235	.0154
X(7)	.0341	.060	.695	.0615	.0402
X(8)	.0392	.107	1.241	.1321	.0865
X(9)	-.0483	.409	4.759	.3383	.2215
X(10)	.0012	.021	.247	.0127	.0083
X(11)	.6515	7.015	81.603	.3098	.2028
X(12)	.0120	.279	3.247	.1594	.1044
X(13)	.0014	.042	.488	.0217	.0142
X(14)	-.0111	.287	3.336	.1359	.0890

SIMULATION OF HWL KT01 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	5.1614	5.1614	2.1914	2.3553
2	.0039	.0039	.0029	1.3417
3	-.0159	-.0159	.0064	-2.4792
4	.0342	.0342	.0285	1.1983
5	-.0266	-.0266	.0335	-.7944
6	-.0179	-.0179	.0365	-.4918
7	.0341	.0341	.0380	.8979
8	.0392	.0392	.0214	1.8276
9	-.0483	-.0483	.0116	-4.1735
10	.0012	.0012	.0058	.2102
11	.6515	.6515	.2061	3.1618
12	.0120	.0120	.0047	2.5326
13	.0014	.0014	.0032	.4420
14	.0111	.0111	.0064	1.7296

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	2957.2861
REGRESSION	13.0000	.7213
RESIDUAL	5.0000	.1196
TOTAL	19.0000	2958.1270

MSSQ= .0239
MS = .1546
MRSQ= .8578
MR = .9262
F = 2.3200

SIMULATION OF HWL OF KTO1 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	12.725	12.698	.027	.210	12.425	12.972	.0183	.7658
2	12.550	12.516	.034	.271	12.300	12.732	.0114	.4765
3	12.623	12.563	.060	.476	12.313	12.813	.0153	.6409
4	12.546	12.552	-.006	-.049	12.243	12.861	.0234	.9767
5	12.300	12.282	.018	.150	12.011	12.552	.0179	.7487
6	12.649	12.514	.135	1.065	12.322	12.707	.0091	.3794
7	12.826	12.769	.057	.441	12.527	13.012	.0144	.6027
8	12.725	12.612	.113	.888	12.340	12.884	.0182	.7606
9	12.927	13.004	-.077	-.599	12.714	13.295	.0206	.8629
10	12.445	12.378	.067	.538	12.149	12.607	.0128	.5361
11	12.421	12.463	-.042	-.336	12.193	12.733	.0179	.7471
12	12.316	12.304	.012	.096	12.003	12.605	.0222	.9267
13	12.369	12.480	-.111	-.897	12.221	12.738	.0164	.6845
14	12.268	12.260	.008	.069	11.957	12.562	.0224	.9364
15	12.268	12.165	.103	.837	11.877	12.454	.0204	.8534
16	12.293	12.390	-.097	-.789	12.105	12.675	.0199	.8300
17	12.266	12.422	-.156	-1.274	12.170	12.674	.0156	.6514
18	12.288	12.351	-.063	-.513	12.058	12.644	.0211	.8873
19	12.236	12.317	-.081	-.662	12.049	12.585	.0177	.7381

The critical HAT(I) value is : 1.4737

SIMULATION OF LWL OF KT01 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	12.49	12.27	12.73	.21	.62	2.62
X(3)	25.47	10.03	97.20	19.55	3.17	14.50
X(4)	12.39	2.39	34.45	8.57	1.38	5.18
X(5)	1.43	.00	6.68	2.07	1.36	4.22
X(6)	1.16	.00	5.92	2.07	1.56	4.29
X(7)	.63	.00	5.41	1.35	3.08	12.92
X(8)	1.84	.00	6.32	1.81	1.15	4.26
X(9)	2.27	.00	10.20	2.82	1.71	6.08
X(10)	8.61	1.40	25.08	7.18	1.24	4.23
YF	10.786	10.211	11.650	.482	.556	2.339

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	6.2022	***	***	.0000	.0000
X(2)	.3359	4.195	82.644	.0720	.0733
X(3)	-.0057	.145	2.848	.1110	.1131
X(4)	.0049	.061	1.195	.0420	.0428
X(5)	-.0500	.071	1.406	.1038	.1058
X(6)	.0658	.076	1.500	.1361	.1387
X(7)	-.0485	.030	.600	.0654	.0666
X(8)	.0870	.160	3.145	.1578	.1608
X(9)	.0123	.028	.549	.0346	.0353
X(10)	.0360	.310	6.113	.2588	.2637

SIMULATION OF LWL OF KT01 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CD(I)	STD(I)	T(I)
1	6.2022	6.2022	4.4132	1.4054
2	.3359	.3359	.3495	.9613
3	-.0057	-.0057	.0042	-1.3441
4	.0049	.0049	.0095	.5142
5	-.0500	-.0500	.0426	-1.1749
6	.0658	.0658	.0642	1.0260
7	-.0485	-.0485	.0606	-.8004
8	.0870	.0870	.0555	1.5674
9	.0123	.0123	.0421	.2912
10	.0360	.0360	.0138	2.6113

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	2093.9509
REGRESSION	9.0000	3.2954
RESIDUAL	8.0000	.6617
TOTAL	18.0000	2097.9080

MSSQ=	.0827
MS =	.2876
MRSQ=	.8328
MR =	.9126
F =	4.4269

SIMULATION OF LWL OF KT01 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	10.410	10.608	-.198	-1.898	10.223	10.992	.0427	.5157
2	10.710	10.476	.234	2.183	10.177	10.776	.0259	.3132
3	10.311	10.170	.141	1.369	9.676	10.664	.0705	.8528
4	10.300	10.437	-.137	-1.330	10.112	10.762	.0305	.3694
5	10.386	10.366	.020	.189	10.067	10.665	.0258	.3125
6	10.945	10.950	-.005	-.050	10.717	11.184	.0158	.1910
7	11.506	11.346	.160	1.394	10.937	11.754	.0481	.5820
8	11.403	11.004	.399	3.503	10.589	11.418	.0497	.6004
9	10.211	10.571	-.360	-3.523	10.215	10.926	.0366	.4419
10	11.450	11.396	.054	.471	10.920	11.872	.0656	.7932
11	10.740	10.779	-.039	-.363	10.260	11.298	.0780	.9426
12	10.516	10.808	-.292	-2.780	10.555	11.062	.0186	.2251
13	10.235	10.389	-.154	-1.505	9.971	10.807	.0506	.6114
14	11.650	11.776	-.126	-1.081	11.326	12.226	.0586	.7082
15	11.287	11.338	-.051	-.454	10.899	11.777	.0557	[illegible]
16	10.516	10.531	-.015	-.144	10.090	10.972	.0563	.6808
17	10.930	10.828	.102	.931	10.389	11.268	.0559	[illegible]
18	10.636	10.369	.267	2.514	9.986	10.751	.0422	.5107

The critical HAT(I) value is : 1.1111

SIMULATION OF HWL OF KTO3 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	24.61	9.14	97.20	19.37	3.17	14.67
X(3)	12.00	2.39	34.45	8.50	1.44	5.31
X(4)	1.46	.00	6.68	2.02	1.32	4.23
X(5)	1.13	.00	5.92	2.01	1.63	4.53
X(6)	.62	.00	5.41	1.31	3.16	13.51
X(7)	1.75	.00	6.32	1.80	1.22	4.38
X(8)	2.72	.00	10.82	3.37	1.50	4.72
X(9)	8.48	1.40	25.08	7.01	1.32	4.46
X(10)	17.55	.63	39.29	10.53	.41	3.19
X(11)	9.27	7.48	10.45	.65	-.90	5.81
X(12)	23.36	7.06	57.37	13.34	1.35	4.87
X(13)	29.71	9.82	65.61	15.34	1.03	4.39
X(14)	25.75	12.86	59.03	12.21	1.48	5.21
YF	12.596	11.800	13.800	.401	1.040	7.446

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	11.6365	***	***	.0000	.0000
X(2)	.0032	.079	2.788	.0622	.0374
X(3)	-.0017	.020	.707	.0142	.0085
X(4)	-.1053	.154	5.431	.2130	.1282
X(5)	-.0188	.021	.752	.0379	.0228
X(6)	.2035	.126	4.460	.2667	.1605
X(7)	-.0606	.106	3.739	.1090	.0656
X(8)	.0101	.027	.968	.0340	.0205
X(9)	.0142	.120	4.233	.0992	.0597
X(10)	.0129	.227	8.003	.1362	.0819
X(11)	-.0686	.637	22.449	.0445	.0268
X(12)	.0045	.106	3.728	.0604	.0363
X(13)	.0079	.236	8.331	.1219	.0734
X(14)	.0379	.976	34.410	.4625	.2783

SIMULATION OF HWL OF KTO3 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	11.6365	11.6365	2.6689	4.3600
2	.0032	.0032	.0053	.6046
3	-.0017	-.0017	.0148	-.1124
4	-.1053	-.1053	.0517	-2.0361
5	-.0188	-.0188	.0626	-.3010
6	.2035	.2035	.0776	2.6218
7	-.0606	-.0606	.0830	-.7296
8	.0101	.0101	.0490	.2059
9	.0142	.0142	.0286	.4949
10	.0129	.0129	.0122	1.0629
11	-.0686	-.0686	.2452	-.2799
12	.0045	.0045	.0103	.4408
13	.0079	.0079	.0065	1.2158
14	.0379	.0379	.0141	2.6946

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	3014.5755
REGRESSION	13.0000	2.4260
RESIDUAL	5.0000	.4663
TOTAL	19.0000	3017.4678

MSSQ= .0933
MS = .3054
MRSQ= .8388
MR = .9159
F = 2.0012

SIMULATION OF HWL OF KTO3 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	12.692	12.564	.128	1.012	12.010	13.118	.0752	.8066
2	12.512	12.338	.174	1.389	11.857	12.820	.0568	.6094
3	12.893	12.743	.150	1.167	12.298	13.187	.0485	.5200
4	12.817	12.844	-.027	-.214	12.255	13.434	.0852	.9133
5	13.800	13.679	.121	.878	13.145	14.213	.0699	.7497
6	12.357	12.103	.254	2.058	11.693	12.512	.0411	.4409
7	12.588	12.740	-.152	-1.211	12.287	13.194	.0504	.5404
8	12.716	12.663	.053	.414	12.122	13.205	.0718	.7698
9	12.588	12.541	.047	.372	11.967	13.115	.0807	.8658
10	12.235	12.263	-.028	-.226	11.786	12.740	.0558	.5979
11	12.616	12.582	.034	.266	12.050	13.115	.0696	.7460
12	12.488	12.446	.042	.337	11.852	13.039	.0863	.9259
13	12.616	12.939	-.323	-2.562	12.459	13.419	.0565	.6061
14	12.817	12.781	.036	.284	12.228	13.333	.0748	.8021
15	12.640	12.499	.141	1.115	11.938	13.060	.0771	.8265
16	11.800	11.922	-.122	-1.033	11.349	12.495	.0805	.8632
17	12.257	12.497	-.240	-1.960	11.964	13.030	.0696	.7466
18	12.787	12.780	.007	.052	12.187	13.373	.0862	.9244
19	12.107	12.401	-.294	-2.432	11.869	12.934	.0695	.7456

The critical HAT(I) value is : 1.4737

SIMULATION OF LWL OF KTO3 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	12.62	11.80	13.80	.39	1.06	8.01
X(3)	25.47	10.03	97.20	19.55	3.17	14.50
X(4)	12.39	2.39	34.45	8.57	1.38	5.18
X(5)	1.43	.00	6.48	2.07	1.36	4.22
X(6)	1.16	.00	5.92	2.07	1.56	4.29
X(7)	.63	.00	5.41	1.35	3.08	12.92
X(8)	1.84	.00	6.32	1.81	1.15	4.26
X(9)	2.27	.00	10.20	2.82	1.71	6.08
X(10)	8.61	1.40	25.08	7.18	1.24	4.23
YF	9.299	7.483	10.455	.657	-1.029	6.059

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	.2164	***	***	.0000	.0000
X(2)	.6495	8.198	81.413	.2559	.1218
X(3)	.0020	.050	.500	.0386	.0184
X(4)	.0175	.216	2.148	.1497	.0712
X(5)	.0279	.040	.395	.0579	.0276
X(6)	.0288	.033	.331	.0596	.0284
X(7)	.1794	.113	1.118	.2419	.1151
X(8)	-.1883	.345	3.431	.3414	.1625
X(9)	-.0653	.148	1.474	.1843	.0877
X(10)	.1074	.925	9.190	.7719	.3673

SIMULATION OF LWL OF KTO3 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	.2164	.2164	3.8485	.0562
2	.6495	.6495	.3001	2.1641
3	.0020	.0020	.0064	.3111
4	.0175	.0175	.0142	1.2328
5	.0279	.0279	.0644	.4334
6	.0288	.0288	.0957	.3012
7	.1794	.1794	.0895	2.0044
8	-.1883	-.1883	.0831	-2.2664
9	-.0653	-.0653	.0630	-1.0358
10	.1074	.1074	.0211	5.0870

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	1556.3736
REGRESSION	9.0000	5.8830
RESIDUAL	8.0000	1.4639
TOTAL	18.0000	1563.7205

MSSQ= .1830
MS = .4278
MRSQ= .8007
MR = .8948
F = 3.5722

<u>SIMULATION OF LWL OF KTO3 - (c)</u>

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	8.357	8.831	-.434	-5.174	8.287	9.376	.0858	.4687
2	9.339	9.368	-.029	-.313	8.920	9.817	.0581	.3177
3	9.287	9.445	-.158	-1.700	8.720	10.170	.1520	.83[illegible]9
4	9.200	9.262	-.062	-.673	8.779	9.745	.0677	[illegible]
5	9.363	9.490	-.127	-1.359	9.775	10.206	.1[illegible]79	.8[illegible]4
6	8.723	8.614	.109	1.244	8.228	9.001	.0432	.2[illegible]59
7	9.693	9.958	-.265	-2.732	9.477	10.438	.0668	.3640
8	7.487	7.982	-.499	-6.666	7.362	8.602	.1110	[illegible]
9	9.288	[illegible]	.570	6.069	8.418	9.218	.0467	[illegible]
10	10.224	[illegible]	[illegible]	-.843	9.610	11.014	[illegible]	[illegible]
11	9.660	[illegible]	.046	.472	8.844	10.400	.175[illegible]	[illegible]
12	9.211	8.777	.434	4.712	8.422	9.132	.0364	[illegible]
13	9.235	9.034	.201	2.177	8.437	9.631	.10[illegible]	.5[illegible]
14	10.455	10.272	.183	1.751	9.579	10.963	.1[illegible]	[illegible]
15	9.644	9.449	.195	2.020	8.830	10.068	.1[illegible]08	[illegible]
16	8.950	9.110	-.160	-1.792	8.411	9.810	.1416	.7[illegible]
17	9.497	9.798	-.301	-3.166	9.153	10.443	.1202	.5[illegible]
18	9.617	9.232	.385	4.002	8.696	9.768	.0831	.4541

The critical HAT(I) value is : 1.1111

SIMULATION OF HWL KTO5 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	22.53	5.77	47.04	12.61	.73	3.11
X(3)	10.75	.00	30.77	8.10	.75	4.03
X(4)	2.36	.00	10.97	3.16	1.72	5.92
X(5)	.82	.00	6.40	1.93	2.47	8.29
X(6)	.66	.00	4.72	1.22	2.41	9.39
X(7)	1.78	.00	6.58	1.88	.96	4.00
X(8)	2.60	.00	10.67	2.91	1.49	5.33
X(9)	9.49	.91	29.39	7.31	1.17	5.04
X(10)	17.67	1.65	35.81	8.89	.37	3.19
X(11)	9.19	7.44	10.10	.70	-.90	4.06
X(12)	24.44	5.82	58.00	12.16	1.01	5.50
X(13)	31.58	10.64	55.30	13.93	.24	2.08
X(14)	29.05	17.43	48.54	8.06	.73	3.96
YF	11.919	9.988	13.400	.749	-1.141	5.918

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	11.1712	***	***	.0000	.0000
X(2)	-.0476	1.071	9.284	.5998	.1648
X(3)	-.0234	.251	2.175	.1891	.0520
X(4)	.0311	.073	.636	.0982	.0270
X(5)	-.1461	.120	1.041	.2827	.0777
X(6)	.1616	.107	.927	.1975	.0543
X(7)	-.1472	.262	2.273	.2765	.0760
X(8)	-.0938	.244	2.114	.2733	.0751
X(9)	-.0070	.066	.572	.0509	.0140
X(10)	-.0433	.765	6.631	.3851	.1058
X(11)	.6233	5.727	49.621	.4344	.1194
X(12)	.0097	.237	2.057	.1181	.0324
X(13)	-.0016	.050	.436	.0222	.0061
X(14)	-.0883	2.566	22.234	.7119	.1956

SIMULATION OF HWL OF KT05 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	11.1712	11.1712	4.6085	2.4241
2	-.0476	-.0476	.0169	-2.8200
3	-.0234	-.0234	.0201	-1.1640
4	.0311	.0311	.0593	.5244
5	-.1461	-.1461	.1734	-.8427
6	.1616	.1616	.1653	.9778
7	-.1472	-.1472	.1061	-1.3871
8	-.0938	-.0938	.1123	-.8351
9	-.0070	-.0070	.0199	-.3495
10	-.0433	-.0433	.0389	-1.1123
11	.6233	.6233	.3304	1.8862
12	.0097	.0097	.0121	.8002
13	-.0016	-.0016	.0112	-.1424
14	-.0883	-.0883	.0382	-2.3134

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	2699.3079
REGRESSION	13.0000	9.1250
RESIDUAL	5.0000	.9691
TOTAL	19.0000	2709.4019

MSSQ= .1938
MS = .4402
MRSQ= .9040
MR = .9508
F = 3.6216

SIMULATION OF HWL KTO5 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	12.247	12.135	.112	.913	11.316	12.954	.1643	.8475
2	12.036	11.966	.070	.579	11.266	12.667	.1202	.6204
3	12.418	12.038	.380	3.064	11.509	12.566	.0685	.3536
4	11.881	11.766	.115	.970	10.905	12.626	.1814	.9360
5	10.350	10.439	-.089	-.863	9.611	11.268	.1683	.8684
6	9.988	10.418	-.430	-4.309	9.815	11.022	.0892	.4604
7	11.220	11.136	.084	.748	10.329	11.944	.1598	.8245
8	11.628	11.689	-.061	-.525	10.896	12.482	.1541	.7949
9	11.905	12.060	-.155	-1.300	11.262	12.858	.1561	.8052
10	12.110	11.608	.502	4.146	10.964	12.251	.1015	.5235
11	13.400	13.620	-.220	-1.642	12.882	14.358	.1336	.6891
12	12.311	12.209	.102	.826	11.373	13.046	.1715	.8847
13	12.210	11.946	.264	2.162	11.205	12.687	.1346	.6946
14	12.235	12.348	-.113	-.926	11.733	12.963	.0927	.4783
15	12.186	12.388	-.202	-1.656	11.558	13.218	.1689	.8714
16	11.881	12.029	-.148	-1.243	11.205	12.853	.1664	.8586
17	12.470	12.459	.011	.089	11.667	13.250	.1536	.7923
18	12.030	11.935	.095	.793	11.078	12.791	.1799	.9281
19	11.960	12.277	-.317	-2.651	11.498	13.057	.1489	.7685

The critical HAT(1) value is : 1.4737

SIMULATION OF LWL OF KT05 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	12.22	11.63	13.40	.46	1.81	8.25
X(3)	21.11	5.77	46.02	11.89	.88	4.36
X(4)	10.64	.00	30.77	8.64	1.15	5.72
X(5)	1.83	.00	10.97	3.28	2.53	10.27
X(6)	2.34	.00	10.67	3.29	1.86	7.16
X(7)	9.36	.91	29.39	8.47	1.43	5.94
X(8)	19.98	1.65	35.81	8.96	-.44	4.81
YF	9.649	9.379	10.100	.288	.919	2.808

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	4.0147	***	***	.0000	.0000
X(2)	.4660	5.692	92.455	.2124	.3582
X(3)	.0050	.106	1.720	.0596	.1006
X(4)	-.0150	.159	2.585	.1292	.2180
X(5)	-.0112	.020	.332	.0366	.0618
X(6)	.0257	.060	.978	.0846	.1427
X(7)	.0040	.037	.609	.0339	.0571
X(8)	-.0041	.081	1.321	.0365	.0615

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HA
1	9.379	9.308	.071	.762	8.829	9.786	.0414	.586
2	9.445	9.472	-.027	-.284	9.071	9.873	.0291	.412
3	10.025	9.700	.325	3.246	9.298	10.101	.0291	.4127
4	10.100	10.060	.040	.395	9.459	10.661	.0654	.9257
5	9.735	9.835	-.100	-1.022	9.310	10.359	.0498	.7057
6	9.491	9.669	-.178	-1.874	9.171	10.167	.0449	.6359
7	9.467	9.662	-.195	-2.056	9.105	10.218	.0561	.7944
8	10.100	10.031	.069	.683	9.419	10.643	.0679	.9614
9	9.490	9.485	.005	.052	8.866	10.104	.0693	.981
10	9.440	9.362	.078	.831	8.767	9.956	.0640	.906
11	9.467	9.557	-.090	-.954	9.043	10.072	.0479	.678

The critical HAT(I) value is : 1.4545

SIMULATION OF LWL OF KT05 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	4.0147	4.0147	2.5954	1.5468
2	.4660	.4660	.2222	2.0974
3	.0050	.0050	.0140	.3587
4	-.0150	-.0150	.0142	-1.0515
5	-.0112	-.0112	.0322	-.3475
6	.0257	.0257	.0331	.7767
7	.0040	.0040	.0113	.3536
8	-.0041	-.0041	.0169	-.2409

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	1024.1352
REGRESSION	7.0000	.6184
RESIDUAL	3.0000	.2119
TOTAL	11.0000	1024.9655

MSSQ=	.0706
MS =	.2658
MRSQ=	.7448
MR =	.8630
F =	1.2509

SIMULATION OF HWL OF JEO4 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	21.26	4.24	43.15	8.60	.86	4.82
X(3)	10.48	.00	27.10	6.99	.70	3.35
X(4)	2.42	.00	10.65	3.25	1.42	4.37
X(5)	1.47	.00	10.19	3.05	2.18	7.05
X(6)	.73	.00	3.28	1.16	1.58	4.30
X(7)	1.72	.00	8.41	1.94	2.19	9.48
X(8)	4.44	.00	14.50	4.69	.95	3.39
X(9)	7.72	.43	27.97	6.98	1.96	6.95
X(10)	17.83	1.40	36.12	7.69	.34	4.31
X(11)	26.30	6.15	56.42	11.78	.62	4.32
X(12)	28.86	5.93	47.32	10.05	.10	3.49
X(13)	27.82	12.47	51.30	10.20	.68	3.19
YF	7.517	6.000	8.570	.728	-.113	2.529

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	6.3226	***	***	.0000	.0000
X(2)	-.0178	.380	6.997	.1536	.0494
X(3)	-.0407	.426	7.859	.2843	.0915
X(4)	-.0580	.140	2.583	.1882	.0606
X(5)	.0115	.017	.313	.0351	.0113
X(6)	.1931	.141	2.606	.2237	.0720
X(7)	-.1679	.289	5.334	.3261	.1049
X(8)	-.0556	.247	4.547	.2607	.0839
X(9)	.0473	.365	6.735	.3301	.1062
X(10)	.0172	.306	5.643	.1320	.0425
X(11)	-.0241	.633	11.669	.2835	.0912
X(12)	.0356	1.028	18.953	.3579	.1152
X(13)	.0522	1.452	26.761	.5321	.1713

SIMULATION OF HWL OF JEO4 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	6.3226	6.3226	.8363	7.5602
2	-.0178	-.0178	.0247	-.7223
3	-.0407	-.0407	.0228	-1.7831
4	-.0580	-.0580	.0462	-1.2547
5	.0115	.0115	.0574	.2006
6	.1931	.1931	.1576	1.2255
7	-.1679	-.1679	.0926	-1.8144
8	-.0556	-.0556	.0397	-1.4019
9	.0473	.0473	.0231	2.0447
10	.0172	.0172	.0189	.9097
11	-.0241	-.0241	.0153	-1.5691
12	.0356	.0356	.0152	2.3444
13	.0522	.0522	.0159	3.2891

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	1186.6562
REGRESSION	12.0000	8.5231
RESIDUAL	8.0000	2.0741
TOTAL	21.0000	1197.2534

MSSQ= .2593
MS = .5092
MRSQ= .8043
MR = .8968
F = 2.7395

SIMULATION OF HWL OF JEO4 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	8.220	7.880	.340	4.132	7.210	8.551	.1299	.5011
2	7.230	7.495	-.265	-3.668	6.743	8.248	.1637	.6315
3	7.390	7.880	-.490	-6.635	7.086	8.674	.1823	.7032
4	7.170	7.623	-.453	-6.318	6.844	8.402	.1754	.6767
5	8.410	7.713	.697	8.283	7.000	8.427	.1470	.5671
6	8.570	8.203	.367	4.282	7.428	8.978	.1736	.6694
7	8.570	8.315	.255	2.981	7.592	9.037	.1510	.5822
8	6.680	6.957	-.277	-4.148	6.442	7.473	.0768	.2962
9	7.700	7.085	.615	7.986	6.269	7.901	.1926	.7430
10	8.260	8.137	.123	1.491	7.371	8.903	.1696	.6541
11	6.850	6.843	.007	.109	6.021	7.664	.1950	.7520
12	7.090	7.101	-.011	-.160	6.540	7.663	.0911	.3515
13	8.340	8.344	-.004	-.050	7.512	9.177	.2003	.7726
14	7.010	7.132	-.122	-1.744	6.349	7.915	.1773	.6838
15	6.000	6.006	-.006	-.104	5.235	6.777	.1718	.6625
16	6.700	6.837	-.137	-2.052	6.287	7.388	.0876	.3377
17	7.140	7.255	-.115	-1.605	6.387	8.123	.2178	.8400
18	6.910	6.823	.087	1.260	5.939	7.707	.2257	.8705
19	8.180	8.598	-.418	-5.111	7.823	9.373	.1735	.6691
20	7.820	8.057	-.237	-3.026	7.337	8.776	.1495	.5767
21	7.620	7.575	.045	.592	6.933	8.216	.1190	.4589

The critical HAT(I) value is : 1.2381

SIMULATION OF LWL OF JEO4 (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	7.50	6.00	8.57	.73	-.05	2.51
X(3)	21.26	4.24	43.15	8.60	.86	4.82
X(4)	10.48	.00	27.10	6.99	.70	3.35
X(5)	2.42	.00	10.65	3.25	1.42	4.37
X(6)	1.47	.00	10.19	3.05	2.18	7.05
X(7)	.73	.00	3.28	1.16	1.58	4.30
X(8)	1.72	.00	8.41	1.94	2.19	9.48
X(9)	4.44	.00	14.50	4.69	.95	3.39
X(10)	7.72	.43	27.97	6.98	1.96	6.95
YF	5.023	4.000	6.000	.503	.084	3.229

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	1.2303	***	***	.0000	.0000
X(2)	.4576	3.434	77.738	.3331	.2868
X(3)	-.0147	.312	7.062	.1262	.1087
X(4)	.0175	.183	4.145	.1221	.1051
X(5)	.0273	.066	1.494	.0886	.0763
X(6)	.0341	.050	1.140	.1041	.0897
X(7)	.0624	.046	1.034	.0723	.0622
X(8)	.0162	.028	.630	.0314	.0270
X(9)	.0205	.091	2.058	.0961	.0827
X(10)	.0269	.208	4.699	.1875	.1615

SIMULATION OF LWL OF JEO4 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	1.2303	1.2303	1.1990	1.0261
2	.4576	.4576	.1555	2.9424
3	-.0147	-.0147	.0129	-1.1349
4	.0175	.0175	.0139	1.2584
5	.0273	.0273	.0248	1.1029
6	.0341	.0341	.0347	.9841
7	.0624	.0624	.1033	.6042
8	.0162	.0162	.0508	.3178
9	.0205	.0205	.0237	.8662
10	.0269	.0269	.0141	1.9021

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	529.9114
REGRESSION	9.0000	3.8031
RESIDUAL	11.0000	1.2564
TOTAL	21.0000	534.9709

MSSQ=	.1142
MS =	.3300
MRSQ=	.7517
MR =	.8670
F =	3.6996

SIMULATION OF LWL OF JE04 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	5.400	4.734	.666	12.325	4.489	4.980	.0186	.1628
2	5.250	5.223	.027	.507	4.757	5.690	.0671	.5876
3	5.010	5.305	-.295	-5.897	4.819	5.792	.0730	.6392
4	4.700	4.556	.144	3.070	4.235	4.876	.0317	.2776
5	5.010	4.972	.038	.760	4.589	5.355	.0454	.3973
6	4.910	5.069	-.159	-3.244	4.612	5.527	.0645	.5649
7	5.850	5.610	.240	4.106	5.187	6.032	.0551	.4827
8	4.650	5.115	-.465	-9.992	4.716	5.514	.0491	.4301
9	4.470	4.584	-.114	-2.548	4.188	4.980	.0484	.4233
10	5.290	5.527	-.237	-4.472	5.102	5.951	.0557	.4874
11	5.010	4.945	.065	1.290	4.619	5.272	.0329	.2879
12	4.330	4.534	-.204	-4.716	4.207	4.861	.0330	.2892
13	5.050	5.323	-.273	-5.397	4.780	5.865	.0909	.7955
14	6.000	5.785	.215	3.581	5.303	6.268	.0719	.6291
15	4.500	4.583	-.083	-1.844	4.235	4.931	.0374	.3274
16	4.000	4.117	-.117	-2.931	3.640	4.594	.0702	.6144
17	5.690	5.537	.153	2.694	4.986	6.087	.0935	.8189
18	5.540	5.332	.208	3.746	4.801	5.864	.0871	.7629
19	4.770	4.645	.125	2.611	4.293	4.998	.0384	.3358
20	5.120	5.265	-.145	-2.837	4.934	5.596	.0338	.2960
21	4.940	4.728	.212	4.301	4.348	5.107	.0445	.3897

The critical HAT(I) value is : .9524

SIMULATION OF HWL OF JE05 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	22.70	7.10	48.14	10.96	.39	3.02
X(3)	10.39	1.32	34.85	8.31	1.60	6.07
X(4)	2.53	.00	13.21	3.61	1.77	6.34
X(5)	1.22	.00	7.49	2.32	1.88	5.72
X(6)	.42	.00	2.79	.74	2.30	8.38
X(7)	1.59	.00	5.18	1.71	.78	2.92
X(8)	3.56	.00	9.65	3.25	.50	2.25
X(9)	11.18	.41	37.08	9.57	1.40	4.79
X(10)	18.16	.48	38.20	10.49	.41	2.84
X(11)	5.39	2.37	6.83	1.25	-1.68	5.13
X(12)	27.86	11.10	66.77	12.67	1.52	6.78
X(13)	33.62	13.21	68.04	15.25	.97	3.90
X(14)	27.62	13.16	43.74	9.43	.38	2.35
YF	9.334	7.257	10.120	.751	-1.346	4.993

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	6.7051	***	***	.0000	.0000
X(2)	-.0501	1.137	16.683	.5490	.1734
X(3)	.0178	.185	2.709	.1476	.0466
X(4)	.0169	.043	.625	.0609	.0192
X(5)	-.0337	.041	.601	.0783	.0247
X(6)	-.3538	.150	2.196	.2627	.0830
X(7)	-.0712	.113	1.658	.1218	.0385
X(8)	.0436	.155	2.279	.1416	.0447
X(9)	-.0120	.134	1.972	.1151	.0364
X(10)	-.0262	.476	6.983	.2750	.0869
X(11)	.4766	2.568	37.683	.5951	.1880
X(12)	.0288	.803	11.779	.3648	.1153
X(13)	.0288	.969	14.215	.4393	.1388
X(14)	-.0015	.042	.617	.0144	.0045

SIMULATION OF HWL OF JE05 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	6.7051	6.7051	1.0888	6.1585
2	-.0501	-.0501	.0102	-4.9278
3	.0178	.0178	.0093	1.9152
4	.0169	.0169	.0445	.3791
5	-.0337	-.0337	.0453	-.7442
6	-.3538	-.3538	.2252	-1.5713
7	-.0712	-.0712	.0541	-1.3160
8	.0436	.0436	.0644	.6772
9	-.0120	-.0120	.0095	-1.2588
10	-.0262	-.0262	.0326	-.8044
11	.4766	.4766	.1102	4.3249
12	.0288	.0288	.0163	1.7724
13	.0288	.0288	.0083	3.4812
14	-.0015	-.0015	.0232	-.0657

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	1742.3404
REGRESSION	13.0000	10.2450
RESIDUAL	6.0000	.4754
TOTAL	20.0000	1753.0609

MSSQ=	.0792
MS =	.2815
MRSQ=	.9557
MR =	.9776
F =	9.9455

SIMULATION OF HWL OF JEO5 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	8.827	9.110	-.283	-3.207	8.840	9.380	.0194	.2449
2	9.741	9.848	-.107	-1.095	9.345	10.350	.0671	.8464
3	9.970	9.741	.229	2.297	9.369	10.113	.0368	.4648
4	9.970	9.984	-.014	-.141	9.567	10.401	.0463	.5839
5	8.900	8.788	.112	1.259	8.281	9.295	.0684	.8634
6	10.120	9.923	.197	1.947	9.448	10.398	.0600	.7577
7	8.952	8.861	.091	1.022	8.427	9.294	.0499	.6294
8	9.994	9.814	.180	1.805	9.380	10.247	.0499	.6303
9	9.766	9.854	-.088	-.902	9.471	10.237	.0390	.4916
10	9.842	9.735	.107	1.085	9.278	10.192	.0555	.7000
11	8.547	8.873	-.326	-3.817	8.513	9.233	.0344	.4340
12	9.790	9.652	.138	1.409	9.133	10.171	.0715	.9021
13	9.994	10.070	-.076	-.765	9.590	10.551	.0614	.7747
14	9.308	9.447	-.139	-1.489	8.932	9.961	.0703	.8871
15	8.979	8.915	.064	.708	8.551	9.280	.0354	.4463
16	9.613	9.811	-.198	-2.065	9.333	10.290	.0609	.7684
17	9.232	9.105	.127	1.377	8.612	9.598	.0645	.8142
18	8.056	8.010	.046	.570	7.498	8.522	.0697	.8799
19	9.815	9.899	-.084	-.855	9.366	10.432	.0755	.9533
20	7.257	7.233	.024	.337	6.707	7.759	.0735	.9276

The critical HAT(I) value is : 1.4000

SIMULATION OF LWL OF JE05 - (a)

BASIC INPUT STATISTICS

VARIABLES	MEAN	MN	MX	STD	SKEW	CUR
X(1)	1.00	1.00	1.00	99.99	99.99	99.99
X(2)	9.52	8.55	10.12	.52	-.64	2.40
X(3)	22.04	7.34	33.45	8.93	-.19	1.91
X(4)	10.25	1.32	34.85	8.72	1.83	6.84
X(5)	2.33	.00	9.22	2.75	1.17	4.46
X(6)	1.39	.00	7.49	2.54	1.66	4.82
X(7)	.43	.00	2.79	.81	2.27	7.87
X(8)	1.79	.00	5.18	1.74	.69	2.97
X(9)	3.89	.00	9.65	3.22	.40	2.43
X(10)	11.70	.71	37.08	9.77	1.52	5.24
YF	5.893	4.913	6.834	.444	.035	4.602

VARIABLES	COEFFS	CONT	%CONT	SENS	RSENS
X(1)	5.7728	***	***	.0000	.0000
X(2)	-.1761	1.676	46.082	.0907	.0616
X(3)	.0450	.992	27.256	.4016	.2726
X(4)	-.0059	.060	1.649	.0510	.0346
X(5)	.0665	.155	4.263	.1832	.1244
X(6)	.0593	.083	2.269	.1507	.1023
X(7)	-.0517	.022	.613	.0419	.0284
X(8)	.0525	.094	2.590	.0913	.0620
X(9)	.0710	.276	7.582	.2288	.1553
X(10)	.0239	.280	7.696	.2340	.1588

SIMULATION OF LWL OF JE05 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	5.7728	5.7728	1.2911	4.4711
2	-.1761	-.1761	.1479	-1.1908
3	.0450	.0450	.0114	3.9473
4	-.0059	-.0059	.0062	-.9498
5	.0665	.0665	.0261	2.5513
6	.0593	.0593	.0275	2.1567
7	-.0517	-.0517	.0789	-.6545
8	.0525	.0525	.0412	1.2744
9	.0710	.0710	.0257	2.7577
10	.0239	.0239	.0066	3.6388

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	555.6863
REGRESSION	9.0000	2.7393
RESIDUAL	6.0000	.2146
TOTAL	16.0000	558.6402

MSSQ=	.0358
MS =	.1891
MRSQ=	.9273
MR =	.9630
F =	8.5094

SIMULATION OF LWL OF JE05 - (c)

PREDICTION STATISTICS

I	YF	YP	ERR	%ERR	95%LL	95%UL	YPVAR	HAT
1	5.523	5.612	-.089	-1.613	5.306	5.918	.0248	.6946
2	5.651	5.821	-.170	-3.007	5.602	6.040	.0128	.3576
3	5.752	5.741	.011	.193	5.430	6.052	.0257	.7184
4	6.100	6.019	.081	1.325	5.683	6.355	.0300	.8378
5	5.995	5.969	.026	.442	5.636	6.301	.0293	.8195
6	4.913	5.069	-.156	-3.181	4.731	5.407	.0303	.8485
7	5.474	5.307	.167	3.059	5.049	5.564	.0177	.4939
8	6.437	6.640	-.203	-3.154	6.404	6.876	.0147	.4122
9	5.803	5.749	.054	.927	5.534	5.964	.0123	.3431
10	5.904	5.723	.181	3.072	5.522	5.924	.0107	.3002
11	5.779	5.896	-.117	-2.031	5.581	6.212	.0265	.7399
12	6.236	6.111	.125	2.002	5.832	6.391	.0208	.5808
13	5.727	5.728	-.001	-.016	5.403	6.053	.0281	.7854
14	5.827	5.833	-.006	-.107	5.569	6.097	.0185	.5102
15	6.337	6.347	-.010	-.161	6.012	6.683	.0299	.8361
16	6.834	6.727	.107	1.566	6.417	7.037	.0255	.7137

The critical HAT(I) value is : 1.2500

SIMIMULATION OF LWL OF JE06 - (b)

REGRESSION COEFFICIENTS AND ITS STATISTICS

I	B(I)	CB(I)	STD(I)	T(I)
1	.5765	.5765	.7874	.7321
2	.0632	.0632	.1529	.4137
3	-.0173	-.0173	.0102	-1.7089
4	.0384	.0384	.0120	3.2058
5	-.0530	-.0530	.0193	-2.7498
6	.0476	.0476	.0259	1.8361
7	.0064	.0064	.0586	.1086
8	-.0258	-.0258	.0314	-.8227
9	.0384	.0384	.0314	1.2245
10	.0041	.0041	.0189	.2161

ANALYSIS OF VARIENCE TABLE

SOURCE	DOF	SOS
MEAN	1.0000	11.5040
REGRESSION	9.0000	1.4906
RESIDUAL	6.0000	.4277
TOTAL	16.0000	13.4222

MSSQ= .0713
MS = .2670
MRSQ= .7770
MR = .8815
F = 2.3235