CS(AR)-4/97-98

Salinity Modelling of Ground Water in Saharanpur and Hardwar District

आपो हि ष्ठा मयोभुवः

NATIONAL INSTITUTE OF HYDROLOGY
JAL VIGYAN BHAWAN
ROORKEE - 247 667 (U.P.) INDIA

1997-98

CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The effect of salinity is one of the most important water quality consideration for agricultural waters. Generally, salinity is measured in terms of conductivity concentration. In the present study an effort is made to develop statistical models for the estimation of conductivity for pre monsoon and post monsoon seasons using routinely monitored water quality parameters of ground water wells in Saharanpur District (UP). Best subset procedure based on R and F values is used in model dissemination. It was found that alkalinity, Sulphate, Nitrate, Sodium and Calcium could be used as surrogate parameters for the prediction of conductivity. The predicted values of conductivity were compared with observed(actual) values and reasonably, good matchings were obtained.

It is noticed that there is not a single model which can be used to predict the conductivity levels. The variation in conductivity not only varies from site to site but also it varies from season to season. However, it is observed that alkalinity and hardness are the parameters which can be used to predict conductivity to a great extent. Because both the alkalinity and hardness are commonly measured water quality parameters, it is suggested both of them should be used to estimate the suitability of ground water with respect to irrigational and other beneficial water uses.

From the statistics arrived at and results/interpretation, it can be concluded that it would be necessary to have not one but two separate models (one for pre and another for post monsoon) so that the conductivity vis-a-vis salinity could be obtained (predicted) knowing other surrogate parameters like sulphate, nitrate, sodium etc.

1.0 INTRODUCTION:

Extensive areas of land in the arid and semi-arid regions of India have gone out of cultivation due to the rise of water table and accumulation of salt. Excessive irrigation and poor water management are the chief causes of water logging and salt build up. Development of soil salinity is a challenge to the permanence of irrigated agriculture. An accumulation of salts in soils leads to unfavourable soil-water-air relationships and decreases crop production. Gradually, the land goes out of cultivation, unless remedial measures are undertaken.

In India, salt affected soils are designated by different names according to their morphological, chemical and physical characteristics and the language of the region. Most commonly, the term usar which is derived from the ancient Sanskrit word ushtra, meaning sterile or barren is collectively used for all kinds of saline and alkali soils. The other words used in different regions of the country are reh, thur, kallar, rakkar and kulrati in Rajasthan, UP and Punjab; karl, chopan and lona in Maharashtra; karu, urppu, choudu and palachoudu in Tamil Nadu; phodus in Andhra Pradesh and kari, pokhali or kaipad in Kerala. However, these terms are not associated with any particular limit of the indices measuring salinity or alkalinity of the soils.

Ground water is one of the most important source of water. Therefore one has to pay attention for its quality. The quality of ground water is the resultant of all processes and reactions that have to be acted on the water from the moment it condensed in the atmosphere to the time it is discharged out by a well or spring. Besides atmospheric pollutants added to water, there are many sources that contribute contaminants to the ground water zone. The major sources contributing to the pollution problems are land disposal of solid wastes, sewage disposal on land, agricultural activities, leakages and spills, deep well disposal of liquid wastes and urban runoff and polluted surface water. In simple terms, the quality of water held with in the groundwater reservoir reflect the soil water percolating to the water table and further water/rock interactions occurring with in the groundwater body. In more detail, the primary controls on the solute content of groundwater as the original chemical quality of the water entering the zone of saturation; the distribution, solubility, exchange capacity and exchange selectivity of the minerals in the strata; the porosity and permeability of the rocks and the flow path of the water.

Contaminant of ground water by anthropogenic non-point sources attributed to development is well documented (Boumer 1978; Canter and Knox 1985; Novotny and Chesters 1981; Sandhu et al. 1977). Trilinear diagrams developed from the work of Piper (Grower, 1980) are frequently used to characterize the chemical quality of ground water. The components of natural water can be divided into five classes, namely dissolved inorganic ions and compounds, particulate inorganic compounds, dissolved organic compounds and particulate, organic materials and dissolved gases. The dissolved inorganic constituents are responsible for the conductivity of the water. In 1972, Grower (1980) found a correlation between dissolved solids and conductance in rivers in West Africa, the regression equations being:

$$(Ca+Mg+K+Na+Cl+HCO3) = 1.4 + 0.71 \text{ Conductivity}$$

and

$$(Ca+Mg+K+Na+Cl+HCO3+SiO2) = 1.1 + 0.95 \text{ Conductivity}$$

This study's objectives were to estimate the relative importance of each sources of conductivity and to develop the suitable regression models for the prediction of conductivity with the help of other commonly measured water quality parameters.

The total salinity (specific conductivity is a measure of the concentration of salts in water and as such is related to the usability of water for irrigation of crops. The significance and interpretation of recommended salinity classifications (SC) are as follows;

1. Low salinity water (SC) can be used for irrigation with most crops on most soils with little likelihood that soil salinity will develop. Some leaching is required, but this occurs under normal irrigation practices except in soils of extremely low permeability.

2. Medium salinity water (SC 250-750) can be used if a moderate amount of leaching occurs. Plants with moderate salt tolerance can be grown in most cases.

3. High salinity water (SC-750-2000) can not be used on soils with restricted drainage. On soils with adequate drainage special management practices are required,

4. Very high salinity water (SC-2000) is not suitable for irrigation under ordinary conditions. It may be used under special circumstances where soils are permeable, drainage adequate excess leaching water is applied in profusion and extremely salt tolerant crops are grown.

2.0 NATURE AND EXTENT OF SALT PROBLEMS IN IRRIGATED AREAS OF INDIA:

The properties of soil are profoundly influenced by the quality of irrigation water, hydrological conditions and cultural practices. Broadly speaking, the salt problems in the irrigated areas of India are related to: i) the quality of water used for irrigation, ii) rainfall, and iii) depth of water table. Fluctuations in water table, leading to temporary water logging and seepage from canals, also modify the above factors. The salt affected areas of India can be classified into three categories a) arid and semi-arid regions with low rainfall and irrigated with poor quality ground water, b) regions of high water table and areas influenced by temporary waterlogging and seepage from canals and other sources and c) coastal regions influenced by sea water intrusion.

Salt problems in the arid and semi-arid regions having rainfall below 45 cm per annum are spreaded over the states of Gujrat and Rajasthan and in the districts of Agra and Mathura in Uttar Pradesh, Ferozepur and Sangrur in Punjab and Rohtak, Gurgaon and Mohindergarh in Haryana. In these areas soil salinity is primarily due to the use of poor quality irrigation water as other factors including rainfall, high water table and seepage are not operative. The severity of the salinity problem in these areas is further increased by the arid climate.

In Rajasthan, out of a total area of 1.1 million hectares under well irrigation about 57 percent of the area is affected by the problem of salinity and alkalinity (Paliwal, 1972). The salt affected area is about 70 percent of the irrigated area in the districts of Bikaner, Jaiselmer, Pali, Jodhpur, Bharatpur, Barmer, Nagpur, Jaipur and Bhilwara. On this basis a total area of more than 100,000 hectares of land is salt affected in the districts of Jaipur, Bhilwara and Bharatpur.

The mean chemical composition of well waters of some of the districts of Rajasthan shows that due to low rainfall, the ground waters of western region are more saline than that of eastern region. However, the adverse effect of poor quality water is not much visible due to sandy texture of most of the irrigated soils. Similarity, the salt problem is severe in the arid regions of western Gujrat where the quality of well waters is poor.

The salt problems in some of the areas of Punjab, Haryana, Uttar Pradesh and Delhi are due to the combined effect of poor quality irrigation water, high water table and inadequate drainage. In Punjab, the quality of irrigation water does not seem to pose a serious problem in the districts of Gurdaspur, Kapurthala, Hoshiarpur and Ludhiana while in Amritsar, Patiala and Sangrur districts it is of moderate level and the serious problem is in the district of Ferozepur. In Haryana, the problem is serious in the districts of Rohtak and Gurgaon. However, the salt problems in some of these areas are due to low permeability of the soils rather than high water table or poor quality water. In the Union Territory of Delhi, all the 5 blocks, viz; Alipur, Shahdara, Nangloi, Najafgarh and Mehrauli are influenced by the problem of soil salinity and poor quality of well waters. The irrigated areas have high water table and poor drainage.

The salt affected lands in the central and southern states are mostly confined in the black cotton soil region in the semi-arid regions comprising of parts of Maharashtra, Karnataka and Andhra Pradesh. Saline, alkali and saline-alkali soils in varying gradations are observed in these areas. Nearly all the deep black soils and soils in the low lying areas are potentially saline or alkali, depending upon the depth and nature of alkalization. Improper use of irrigation water has decreased the subsoil water table and ultimately by secondary salinization salt affected soils have developed. In some of these areas gypsum is found in the B horizon which helps in keeping the soil in good condition. Almost all the salt affected black cotton soils of the south have a good reserve of lime throughout the profile. This may act as a source of calcium and help in their better utilization.

The saline-alkali soils of the Kaveri delta have been classified into Old delta and New delta. The old delta soils are dark grey, clayey, low-lying and poorly drained, with a water table at about 1.5 meters depth. The New delta soils are light textured, more dispersed and have low permeability.

Salt affected soils of coastal lands. Along the coastal line of India, thousands of hectares of land is salt affected by sea water intrusion . Large areas of land along the mouth of rivers and creeks in West Bengal, Orissa, Andhra Pradesh, Tamil Nadu, Maharashtra and Gujrat are subjected to periodic tidal waves and are inundated with sea water. Such areas are devoid of all types of vegetation.

2.1 WATERLOGGING AND SOIL SALINITY:

Studies at the Land Reclamation, Irrigation and Power Research Institute, Amritsar, showed a rise of water table from 2.6 to 6.6 meters during the period 1940-60(Uppal, 1962). It was observed that the salt problem in high water table areas is mostly governed by the amount and intensity of rainfall. In the monsoon season when water table rises and comes near the surface, the land is temporarily waterlogged due to which salinity is reduced. But after rains and with onset of winter, water is lost by evaporation and salts are deposited on the surface. During the cropping period of winter crop(rabi) such lands are irrigated and the salts present in the soil profile move up and down. Consequently, total soil salinity is controlled by the period of waterlogging, water transmission, characteristics of soil, the salt content of ground water and the evaporative condition of the region. Topography of the land also modifies the ultimate soil salinity.

2.2 Chemistry of Salt affected soils :

The main processes occurring in soils while irrigating with poor quality waters are : i) ion exchange between cations in irrigation water and those present on soil exchange complex, ii) dissolution and precipitation of $CaCO_3$, iii) weathering of minerals, iv) hydration and dehydration of soil as a result of fluctuation in soil moisture, v) leaching down of ions, vi) upward movement of ions through capillary activity, and vii) mineral nutritional characteristics of the crop grown. Amongst these processes, cation exchange is the most important process which governs the accumulation of excessive sodium during irrigation with saline water and reclamation of alkali soils during treatment with gypsum.

The finest fractions of soil particles called clay, have a certain charge deficit. This charge is balanced by the charge of the adsorbed cation taken from the soil solution. These adsorbed ions on the clay phase are replaced by any other cation. However, the rate of reaction depends upon its chemical affinity to the soil, its valency and size. These replaceable ions are called exchangeable cations. The maximum adsorption capacity of the soil for cations is called cation exchange capacity which is generally expressed as meq/100 gm of soil. An order of replaceability of any ion present on the soil can be expressed as

$$NH_4^+ > Mg^{++} > Ca^{++} > K^+ > Na^+$$

This means that Ca^{++} will replace Na^+ from the exchange complex more easily while the replacement of Ca^{++} by Na^+ is more difficult.

In normals soils, generally Ca, Mg, K and NH_4 constitute 70, 25, 3 and 2 percent respectively of the exchangeable cations. such soils are dominated by Ca followed by Mg, K and NH_4. But in case of alkali soils the proportion of exchangeable Na is more and increases with the pH of soil. The pH of alkali soils generally increases with the degree of Na saturation . Exchangeable sodium percent (ESP) value of the soil is further influenced by the nature of the anion associated with the sodium salt in the solution phase.

A clay particle with sodium and calcium ions attached tends to hydrolyse. When a sodium ion is exchanged for a hydrogen ion and the sodium ion combines with a molecule of water, sodium hydroxide (NaOH) is formed.

Na Na
Ca Clay Particle Na + H_2O ----> NaOH + H Clay Particle Ca
Na Na

When carbon dioxide (CO_2) is present in the soil air, it readily reacts with the water to form hydrogen carbonate (H_2CO_3). However, the sodium hydroxide reacts readily with the hydrogen carbonate to form sodium carbonate.

$$H_2CO_3 + 2NaOH = Na_2CO_3 + 2H_2O$$

The Na_2CO_3 is gradually removed with extensive leaching and the soil is left with hydrogen ions having replaced sodium ions. This increase of hydrogen ions is reflected in a lower pH.

When soils contain calcium carbonate ($CaCO_3$)or gypsum, calcium is dissolved into the soil solution. This available calcium is exchanged for sodium during the leaching process to obtain a normal soil.

Na Na Na
Ca Clay Particle +$CaCO_3$ ----> Na_2CO_3 + Ca ClayParticle Ca
Na Na Na

2.3 Effect of Salts on Soils :

The physical and chemical properties of irrigated soils depend largely on the chemical composition of irrigation, soil type, drainage characteristics and climatic conditions. the main factors are :

i) Concentration of salts in the soil solution and
ii) Nature and degree of exchangeable cations.

Under field conditions, it is difficult to separate the two effects completely because the process of ion exchange begins simultaneously, as the concentration of soil solution increases. However, the physical properties of the soil are more deteriorated in the presence of excessive exchangeable sodium.

2.4 Chemical Properties :

Salinity Hazard : When the salt distribution in the soil profile becomes excessive, crop growth is deteriorated. Such soils are often recognised by the presence of a white surface crust. In saline water irrigated areas , particularly in the arid regions of Rajasthan, the EC of the saturation extract of the soils is usually about three times and even more than that of irrigation water. In general , salt concentration is maximum on the surface and decreases with increase in depth (Maliwal, 1968; Mehta, 1970). In irrigated soils, the extent of salinity is more related to the process of salt water movement with in the soil column than whether the salts are moving downwards with the leaching effect of irrigation water or are moving up by capillarity under the thermal gradient. In a well drained light textured soil immediately after irrigation, the soil salify, as measured on the basis of saturation extract of the soil sample, would be less at surface layers and it would be more at deeper layers. On the contrary, in heavy soils accumulation of salts would always be more on the surface due to impeded drainage. Quantitatively, under similar soil water conditions, more salinity would be observed in arid than in humid regions and in heavy textured soils than in light soils.

Sodium Hazard : In irrigated soils , the cations present in irrigation water as soluble salt, take part in an exchange reaction with the soil. The main cation exchange which take place is between sodium ion in irrigation water and calcium ion in the exchange complex of the soil.

$$\text{Ca Clay Particle Ca} + 2Na^{+} = \text{Ca Clay Particle} \begin{matrix} Na \\ Na \end{matrix} + Ca^{++}$$

According to this process a soil may or may not be changed into an alkali soil depending upon the relative proportion of sodium and calcium ions in the equilibrium solution. However, the process of alkalisation could be very quick and can come to an equilibrium condition if the relative proportion of Na : Ca+Mg, as expressed by the sodium adsorption ratio, is high. The exchangeable sodium percentage increases with the total salt concentration and /or SAR of irrigation water or soil solution.

In arid and semi-arid region , as the soil solution becomes concentrated through evaporation or transpiration, the solubility limits of calcium carbonate and magnesium carbonate are exceeded. As a result of this they are precipitated with a corresponding increase in the relative proportion of sodium in the soil solution. Hence, a part of exchangeable calcium and magnesium is replaced by sodium , increasing the sodium content. This leads to the conclusion that any chemical process in the soil, causing a decrease in calcium, either in solution or exchange phase, will lead to enhancement of alkalinity.

In addition to the above process, the extent of degree of sodium saturation will be governed by the chemical affinity of t ion to the soil, its replacing power in comparison to the other ions and its own replaceability by other ions. Fortunately amongst sodium, calcium and magnesium, the replacing power both of calcium and magnesium and their chemical affinity to soil are more than sodium. Hence, a higher proportion of sodium is necessary in the soil solution to replace calcium. The proportion of sodium to calcium and magnesium should be above 50 percent in the soil solution before it is adsorbed in significant amounts. Moreover, more Na is adsorbed at the same salt concentration and the same Na content when Mg is present rather than Ca, as the replaceability of Mg is more that that of Ca. It may be summarised that the process of alkalisation and its extent is influenced by the total salt concentration, ionic composition, relative proportion of sodium to other cations, nature and total amount of clay and their ion exchange characteristics.

Magnesium Hazard : Sometimes more magnesium is present in irrigation water that calcium. This increases the degree of magnesium saturation which also deteriorates the soil structure. Irrigation waters having relatively more magnesium than calcium are likely to decrease soil productivity. About 73 percent of the 4162 samples of poor quality well waters of Rajasthan were observed to have more Mg than Ca (Paliwal, 1972).

3.0 DESCRIPTION OF STUDY AREA:

3.1 Saharanpur District:

The study was conducted for Saharanpur district of Uttar Pradesh (India). The are is part of the Indogangetic plains and lies between latitude 30°26′ N - 29°34′ N and longitude 78°11′ E - 77°6′ E. Saharanpur is one of the important towns of Uttar Pradesh. In Western UP rapid industrial and agricultural growth has taken place during last three decades. This is likely to become manifold in near future particularly in areas like Saharanpur where the necessary industrial nucleus already exists. A variety of industries already have been set up in the district such as paper, textile, sugar, food processing, small scale steel industries and cottage industries etc.

Physiographically the area is generally flat except Siwalik hills in the north and north east. The area is devoid of relief features of any prominence except from deep gorges cut by nallas and rivers flowing through the area. The district is bounded by river Yamuna in the West and river Ganga in the east.

Regarding the drainage of the area, the rivers generally flow from north to south. These rivers during most of the non-monsoon season carry water drained into them from ground water storage. Some of the important rivers of the district are the Ganga, Yamuna, Hindon, Krishna and the Kali (West). Apart from these rivers, the western Ganga canal and Eastern Yamuna canal also drain the area.

The climate of the district Saharanpur as that of the greater part of Indian subcontinent is characterised by moderate type of subtropical monsoonic climate. In general, the average normal monsoon rainfall in the district is about 485.6 mm. The temperature ranges from 8°C in winter to 40°C in summer. Major part of the rainfall (about 75%) is received during the monsoon period. It has been observed that the rainfall is heaviest in the northern region of the district, close to foot hills of Himalayas and becomes lesser southward.

The area under study is a part of west Indogangetic plain which is mainly composed of pleistocene and subrecent alluvium brought down by river action from the Himalayan region. In other words alluvium is made up of recent unconsolidated fluviatile formations comprising of sand, silt, clay and kankar with occasional beds of gravel. The deposits of sand beds of varying thickness are the main source of groundwater in the area.

The groundwater conditions in all alluvial parts are considerably influenced by the varying lithology of the subsurface formations. As the general fluviatile nature of the deposits of indogangetic plains it has been observed that the strata exhibit great variation both laterally and vertically. The main source of water which sustains groundwater body in fine to coarse grained sands is rainfall. Other sources of groundwater replenishment are infiltration from rivers, canals and return flow from irrigation and inflow from the neighbouring areas.

The most common groundwater structures in the area are shallow and deep tubewells. Dugwells are also used as source for drinking water as well as irrigation, but to a lower extent.

Based on the lithological logs and water table fluctuation data, two types of aquifers have been delineated in area (Singh et al., 1979). The upper one is the shallow unconfined aquifer which generally extents to depths around 25m. The deeper aquifers are confined to semi-confined in nature and located at depths around 30 - 140 m, below ground level separated by three to four aquifers at average depths of 30 m to 55 m, 65 m to 90 m and 120 southward trend of groundwater flow both in unconfined aquifers and confined aquifers.

3.2 Hardwar District:

The area under study is a part of the Indogangetic plains and lies between latitude 29°30′ N - 30°15′ N and longitude 77°45′ E - 78°20′ E in the western Uttar Pradesh (Fig.2).

Physiographically the area is generally flat except for the Siwalik hills in the north and north east. The area is devoid of relief features of any prominence except for deep gorges cut by nallas and rivers flowing through the area. The area is bounded by the river Yamuna in the west and river Ganga in the east.

The climate of the area as that of the greater part of Indian subcontinent is characterised by moderate type of subtropical monsoonic climate. The average annual rainfall in the area is about 1000 mm, major part of which is received during monsoon period. The major landuse is agriculture and there is no effective forest cover. The soils of the area are loam to silty loam and are free from carbonates.

The most common ground water structure in the area are shallow and deep tubewells . Dugwell are also used as source for drinking water as well as irrigation,but to a lower extent. The ground water body is contained in fine to coarse-grained sands recharged by rainfall. Other sources of ground water replenishment are infiltration from rivers, canals and return flow from irrigation and inflow from the neighbouring areas.

Based on the lithological logs and water table fluctuation data, two types of aquifers have been delineated in the area. The upper one is the shallow unconfined aquifer which generally extends to depths around 25 m. The deeper aquifers are confined to semi-confined in nature and located at depths around 30 to 140 m, below ground level separated by three to four aquifers at average depths of 30 to 55 m, 65 to 90 m, and 120 to 140 m. Water table contours in the area indicate the southward trend of ground water flow in unconfined and confined aquifers.

3.3 Water Quality Sampling:

In the present study twenty two wells covering the Saharanpur district and twenty five wells in Hardwar district were choosen. The wells were selected based on their proximity to source of pollution i.e. closeness to industry, agricultural fields (pesticides etc.), sewage outfalls etc. Selected wells were operating. Sampling was carried out in the months of June (pre-monsoon) and November (post-moonsoon), 1987 in Saharanpur District and during July and December, 1995 in Hardwar District. The integrated samples were collected by lowering the container in the open wells. The samples collected were stored in clean plastic bottles fitted with screw caps. About two litres of water sample was collected. The quality of water depends on a large number of individual hydrological, physical, chemical and biological factors. Some parameters are of special importance and deserve frequent attention and observation, other gives a rough picture of water body and its quality status. During present study, the chemical properties and the constituents of water analysed are pH, specific conductance, Hardness, Alkalinity (carbonates and bicarbonates), temperature and major cations and anions.

3.4 Ground Water Quality of the Study Area :

The quality of ground water of both district varies from place to place with the depth of water table. The water in shallow aquifers is rich in bicarbonates and alkaline earth metals.

In Saharanpur District, polluted Hindon river is interacting with ground water causing the deterioration of ground water quality. The chemical nature of shallow ground water in the area shows that the water is suitable for irrigation as well as for domestic purposes.

In Hardwar District, higher values of various constituents at many places make the water unfit for domestic applications. The ground water of area can be safely used for irrigation with most crops on most soils but may cause some problem if the soil permeability is very poor. More than 50% samples of the district falls under high salinity zone, such waters should not be used on soils with restricted drainage. Special management for salinity control may be required and plants with good salt tolerance should be selected.

4.0 GENERAL MODEL BUILDING:

The basic model that we consider is

$$Y = X\beta + e \tag{1}$$
$$var(e) = \sigma^2 I_n$$

in which, X is data matrix, Y is dependent variable vector, I_n is an identity matrix, β is unknown vector, and e is an error vector having zero mean and constant standard deviation (σ).

4.1 Parameter Estimation

The method of ordinary least squares is the most widely used method because of the simple concept and no assumption is necessary on the probability distribution of data. This method will be used for estimating parameters for phosphorus and is briefly described here. Let y_i (i = 1,......,n) be the response variables with mean μ_i. Consider a following linear model expressing the relationship between means μ_i and the explanatory variables x_{ij} (value of jth independent variable for observation i).

$$\mu_i = \sum_{j=0}^{k} \beta_j x_{ij} \tag{2}$$

where, x_{i0} = 1, β_j are the parameters to be estimated, and k+1 are the number of parameters in the model.

The method of least squares minimizes the sum of squares between the observed and model computed-values of dependent variable and in the process estimates the parameters that will provide the least sum of squares of error. The function to be minimized with respect to the model parameters (β_j) is

$$\eta = \sum_{i=1}^{n}(y_i - \sum_{j=0}^{j=k}\beta_j x_{ij})^2 \qquad (3)$$

where, y_i are the observed values of the dependent variable and η is the sum of squares of errors. The above function is minimized by partially differentiating the above equation with parameters and equating to zero. The equations so derived are solved for the parameters. The general solution of these equations in the matrix notation are as follows (Draper and Smith, 1981).

$$\hat{\beta} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y} \qquad (4)$$

where, ^{T} represents the transpose of the matrix, and $\hat{\beta}$ is the vector of estimated parameters, **y** is an n dimensional vector of dependent variables, and **x** is the nx(k+1) coefficient matrix of independent variables.

The vector of model-computed dependent variables $\hat{y}$ used for estimating the standard deviation and other statistical tests in matrix notation is defined as $\mathbf{x}\hat{\beta}$. The point estimates of the parameters obtained from Equation (4) and the regression process itself need to be examined before these could be used in the model application. In order to facilitate statistical diagnostics on parameters and regression, the errors $(y_i - \hat{y}_i)$ are assumed independent, normally distributed with zero mean and constant variance. Specifically, the following question must be answered

(1) Is the linear model suitable in explaining the variation in dependent variables (y_j) at various locations and will be model still be useful for another independent data set?,

(2) Is the assumption of normality and constant variance in error structure valid?,

(3) What are the uncertainties in the estimated parameters and are the parameters significant in the statistical sense?, and

(4) Is the overall regression significant in a statistical sense?

4.2 Statistical Tests

Various statistical tests are used to address the above concerns, some of these are briefly described here.

4.2.1 Model Performance

The squared multiple R represents the proportion of the variation in the dependent variable accounted for, or explained by, the linear model. The higher R square values indicate better performance of model and suitability of model in computing the dependent variable (Draper and Smith, 1981). The other related test to examine the suitability of the model for a new data set (from the same population) of dependent variable is the adjusted squared multiple R (Wilkinson, 1990). The adjusted squared multiple R will be smaller than squared multiple R values as the coefficients would be optimized for first data set.

The model performance is also judged by visual examination of the linear plot of observed and model computed dependent variables. The assumptions of normality and variance can be examined by plotting residuals (errors) and y_i, for all observations. A randomly distributed plot is indicative of constant variance and normal distribution.

4.2.2 Uncertainties in Parameters (Confidence Interval and Hypothesis testing)

A 95 % confidence interval indicates a range which will include the true value with 0.95 probability. The narrower the interval the better is the estimation. For correlated parameters, joint confidence region is a better representation of confidence in estimated parameters than the individual confidence interval as it considers parameters jointly responsible in calculating the dependent variables (Draper and Smith, 1981). Inclusion of zero in confidence interval/region signifies that zero is a possible value of the parameters. The other test frequently used for examining if estimated parameters are significantly different than zero(or any other value) is the hypothesis testing using t-statistics (Draper and Smith, 1981). If the observed t-value is higher than the critical t-value at a certain level of significance, it indicates the estimated parameter to be significantly different than zero.

4.2.3 Analysis of Variance and F-statistics

The analysis of variance (ANOVA) and the value of F-ratio is used for assessing the significance of regression. When the F-ratio is statistically significant, it implies that a significantly large amount of the variation in the data about the mean has been taken up by the regression equation (Draper and Smith, 1981). The above statistical investigations greatly assist in examining the model performance and quality of the estimated parameters.

4.3 Detection of Influential Cases

The technique of case analysis, in which the data are analyzed in detail with attention given to the role of each case in determining values of estimators and test statistics, is used in the detection of influential cases. The primary concerns of case analysis are two interrelated questions. First, how well the model used resembles the data actually observed. The basic statistic here will be the residual. If the fitted model does not give a set of residuals that seems reasonable, then some aspect of the model will be called into doubt. The second question of interest is the effect of each case on estimation and other aspects of aggregate analysis. In some data sets, for example, it may be that the observed aggregate statistics depend on one case in such a way that, if that case were detected, the outcome of the aggregate analysis would change. Such cases are termed as 'influential cases' and could be detected using two case statistics called leverage values and distance measures.

4.3.1 Leverage Value

The vector of residuals $\hat{e}$ is defined by

$$\begin{aligned} \hat{e} &= Y - \hat{Y} \\ &= Y - X(X^TX)^{-1}X^TY \\ &= [\, I - X(X^TX)^{-1}X^T]\, Y \end{aligned} \qquad (5)$$

The matrix defined by $X(X^TX)^{-1}X^T$ is very important in the study of case analysis, so we shall give it a name V, defined by

$$V = X(X^TX)^{-1}X^T \qquad (6)$$

Using this definition, the fitted values are given by

$$\hat{Y} = VY \qquad (7)$$

and the residuals are given by

$$\hat{e} = (1 - V)Y \qquad (8)$$

By assumption, the errors are uncorrelated random variables with zero means, and common variance σ^2. Using (5) the moments of $\hat{e}$ are given (Weisberg, 1980) as

$$\begin{aligned} E(\hat{e}) &= 0 \\ \mathrm{var}(\hat{e}) &= \sigma^2(1 - V) \end{aligned} \tag{9}$$

Like the errors, each of the residuals has zero mean, but they have different variances and they are not uncorrelated.

The elements of V, the v_{ij}'s, are given by the equation

$$v_{ij} = x_i^T(X^TX)^{-1}x_j \tag{10}$$

and, for the diagonal elements,

$$v_{ii} = x_i^T(X^TX)x_i \tag{11}$$

where x_i^T and x_j^T are, respectively, the ith row and the jth row of the data matrix X. From (9), the variance of ith residual is

$$\mathrm{var}(\hat{e}_i) = \sigma^2(1 - v_{ii}) \tag{12}$$

and the covariance between the ith and jth residual is

$$\mathrm{cov}(\hat{e}_i, \hat{e}_j) = -\sigma^2 v_{ij} \tag{13}$$

Also, the correlation between the ith and jth residual is

$$corr(\hat{e}_i, \hat{e}_j) = \frac{-v_{ij}}{(1-v_{ii})^{\frac{1}{2}}(1-v_{jj})^{\frac{1}{2}}} \tag{14}$$

It is clear that each v_{ii} must fall in the range between 0 and 1. The notion of v_{ii} giving a measure of how far the ith case is from the center of the data is central to case analysis. As can be seen from (12), cases with large values of v_{ii} will have small values for $var(\hat{e}_i)$; as v_{ii} gets closer to one, this variance will approach zero, and as this happens, regardless of the value of y_i observed for the ith case, it is nearly certain to get a residual for the ith case near zero; that is $\hat{y}_i \cong y_i$. Such a case can be very important in estimating parameters. Cases with large v_{ii} have been called high leverage cases (Weissberg, 1980). In multiple regression, v_{ii} measures the distance from the point x_i to the center of the data, and cases with unusual values for the independent variables will tend to have large values of v_{ii}.

4.3.2 Studentized Residual

As discussed above, $var(\hat{e}_i)$ will be small whenever v_{ii} is large, so cases with x_i near $\bar{x}$ will be fit poorly, and cases with x_i far from $\bar{x}$ will fit well. This is particularly undesirable because violations of a model may be most likely to occur under unusual conditions. The detection of those violations by simply examining the residuals is not possible. However, an improved set of residuals can be obtained by scaling so that cases with large v_{ii} get larger scaled residuals, and cases with smaller v_{ii} get smaller scaled residuals. Then, all the residuals in the analysis can be compared directly. One good way of doing scaling is to divide each of the residuals by an estimate of its standard deviation. Such scaled residuals are called Studentized residuals (Weisberg, 1980), and defined by

$$r_i = \frac{\hat{e}_i}{\sigma\sqrt{1-v_{ii}}}, \qquad i=1,2,\dots\dots\dots,n \tag{15}$$

The most important advantage of the Studentized residuals is that $var(r_i) = 1$ for all i, independent of both σ^2 and the v_{ij}'s, as long as the model is correct; when the model is incorrect, then $var(r_i)$ will generally not be constant over all i (Weisberg, 1980). As a guideline, the cases having their absolute value of r_i more than 3 (Draper and Smith, 1981) may be suspected to be outliers and hence their influence must be studied.

4.3.3 Cook's Distance

The residuals, or the Studentized residuals, are the most commonly used case statistics to measure the success or failure of fitting a model at each case. Now to declare a particular case to be an outlier, its impact up on the values of the model estimates should be checked. If this is found to be significant then that particular case may be called an outlier and must be removed from the data set ; however, these points could be the most important observations in the data set.

In the context of case analysis, rather than comparing estimation techniques, it is important to study the change in the estimate of β when a case is to be deleted from the data. Viewing the estimate $\hat{\beta}$ from the full sample as a fixed point, let $\hat{\beta}_{-i}$ be the least squares estimate of β obtained from the regression using all the cases except the ith. An empirical version of the influence function is obtained by taking the difference between the full data estimate and the estimate using (n-1) cases (excluding the ith), $\hat{\beta}_{-i} - \beta$. A method of measuring the distance between these points is needed to judge whether the ith case has sufficient influence on the estimation of parameters and deletion of it would result in a substantially different conclusion. Cook's distance (D_i) given below, is used to obtain a confidence region for $\hat{\beta}$

$$D_i = \frac{(\hat{\beta}_{-i} - \hat{\beta})^T (X^T X)(\hat{\beta}_{-i} - \hat{\beta})}{(k+1)\sigma^2} \qquad (16)$$

where D_i is the Cook's distance, and (k+1) is the total number of parameters. D_i is compared to $F(k+1, N-k-1, 1-\alpha)$ for selected α; a large D_i denotes an influential ith observation.

5.0 DEVELOPMENT OF REGRESSION MODEL FOR CONDUCTIVITY MODELLING:

Regression analysis were performed in each case on both seasonal subsets pre monsoon and post monsoon.

1. Pre - monsoon (1987, 1995)
2. Post - monsoon (1987, 1995)

5.1 PRELIMINARY ANALYSIS

Prior to a statistical regression analysis of a data set, an initial filtrating of the data which consisted of a statistical analysis, a preliminary regression analysis, partial visual inspection of the data files and the creation of scatter plots revealed obvious data input errors. Once the identified input errors were removed, a general regression analysis assuming all water quality parameters as independent variables and electrical conductivity as dependent variables was made to identify any outliers on the basis of leverage value and studentized residual statistics.

Using the filtered data, correlation of each water quality constituent with conductivity matrices are obtained for both the data sets of premonsoon and post monsoon (Table 1 and 2).

To enhance the visualisation of the correlation matrix, Table 3 presents the square of correlation coefficient to indicate the contribution of individual water quality parameters in explaining the variation in the dependent variable for both the pre monsoon and post monsoon seasons. Since Mg hade no significant correlation (Table 3) with conductivity, this parameter was not considered any further for model formulation.

5.2 SELECTION OF INDEPENDENT VARIABLES FOR REGRESSION ANALYSIS:

To make the model useful for predictive purposes, one wants to include as many independent variables as possible so that reliable fitted values can be determined. Furthermore, since R^2

gives the proportion of the variation in the dependent variables that is explained by the fitted regression model, one obviously desires R^2 to be large. On the other hand, because of the effort involved in the monitoring of a large number of independent variables, there is interest in including as few independent variables as possible. The compromise between these extremes is what is usually called selecting the best regression variables and consequently the best model., There is no unique statistical procedure for doing this (Draper and Smith, 1981). However, there are many statistical procedures such as all possible regression, backward elimination, forward elimination in stepwise regression, ridge regression, principal component regression, and stagewise regression which may help in optimum model formulation (Draper and Smith, 1981; Montgomery and Peck, 1982) and Weisberg, 1980).

In the present study the best subset regression procedures have been used to select the best set of independent variables.

5.2.1 Best Subset Regression:

Using the R^2 information (e.g. Table 1), various best subsets of independent variables can be selected on the basis of proportion of variation explained in the dependent variable. For each subset the regression was assessed according to 1) the value of R^2 achieved, 2) the F value (defined in Equation 3) and 3) the number of observations used in developing the model. The model obtained from the large data set and achieving higher values of R^2 and F value will always be preferred. The above two criteria (R^2 and F values) which will be used in model selection are briefly described below:

5.2.2 R^2 Criterion:

R^2 value is used as a criterion for comparing models, a computing formula for R^2 is

$$R^2 = 1 - \frac{SSE}{SS_Y} = \frac{SSR}{SS_Y} \qquad (17)$$

with $SS_y = \Sigma(Y_i-Y)^2$; $SSE = \Sigma(Y_i-Y_i)^2$; $SSR = \Sigma(Y_i-Y)^2$

where Y is the average value of dependent variable and Y_i are the model computed values of the dependent variable.

A strong linear association between Y_i and Y_i yields a large value of R^2 and vice versa. Unfortunately, R^2 provides an inadequate criterion for subset model selection since, whenever comparing a subset model to a large model including the subset, the large model will always have an R^2 value as value, or larger than R^2 for the subset model. Thus the full model will always have the largest possible value of R^2. However, for a fixed number of independent variables (equal to k), R^2 can be used to compare different models with a large value of R^2 indicating the preferred model.

5.2.3 F-Value Criterion:

The F value is mathematically described as (Draper and Smith 1981):

$$F = \frac{N-k-1}{k} \cdot \frac{R^2}{1-R^2} \quad \text{------------------(18)}$$

From the above expression, it is clear that apart from the constant multiple [(N-k-1)/k], the F statistic is the ratio of the explained to the unexplained variation in Y_i. Therefore it is natural to say that the regression is significant only when the proportion of explained variation is large. This occurs when the F value is large.

The F statistic can also be used to compare any two models as long as all the independent variables in the smaller model are also included in the large model, i.e. the small model is a subset model of the large model.

As defined earlier, the residual sum of squares reflects the variation in the dependent variable that is not explained by the model. If the predictor variables which are not included in the subset model are important, then deleting them from the subset model should result in a significant increase in unexplained variation of Y_i . That is SSE_r , should become considerably large than SSE_f.A convenient test statistic (Weisberg, 1980) using this idea is:

$$F_{k-m,N-k-1} = \frac{(SSE_r - SSE_f)/(k-m)}{SSE_f/(N-k-1)} \qquad \text{---------(19)}$$

Where SSE (Equ.17) and SSE are the residual error sum of squares of the full model (containing k independent variables) and the subset model containing k-m independent variables (where m is the number of independent variables dropped from the full model) respectively. The larger model will be preferred when the $F_{k-m,N-k-1}$ statistic is sufficiently large. One reasonable rule should be to prefer the full model if $F_{k-m,N-k-1} > F^*$ where F^* is the αx100% point of the $F_{k-m,N-k-1}$ distribution. The choice of $\alpha = 0.05$ is typical (Weisberg, 1980).

5.3 SELECTION OF INDEPENDENT VARIABLES:

a) Saharanpur District:

i) Pre-monsoon period:

It is clear from table 3 that for premonsoon season, calcium is the best single variable explaining more than 57% variation in the conductivity levels. The other water quality parameters namely SO4, Hard, K, alk, Na, Cl and Na if taken alone as independent variable explain approximately 55%, 48%, 45%, 33%, 23%, 11% and 7% variation in the conductivity respectively. Now to increase the R^2 the various pairs of water quality parameters are attempted. It is evident from Table-4 that the pair consisting of (Ca & SO4) having larger R^2 and F value is the best model among the other pairs./ Further increasing the R^2 the various water quality parameters were added in the pair of Ca+ SO4+NO3) is selected as the best 3-parameter model. Similarly, the 4-parameter model, 5-parameter model and 6-parameter models were selected which are (Ca+SO4+NO3+Na), (Ca+SO4+NO3+Na+alk) and (Ca+SO4+NO3+Na+alk+K) respectively the various selected models are tabulated in Table - 5. Now, to choose the best model among the various models of table 5, one has to keep both things in mind that the selected model should have minimum number of explaining variables and maximum R^2 value. Now, there is trade off between the two criteria, the decision can be made on the basis F statistic explained in equation. The selection procedure is explained in Table . It is found that the set of variables (Ca+SO4+NO3) is the best set for explaining the variability in the conductivity of pre-monsoon season.

ii) Post-monsoon period:

Similar procedure has been followed to select the best set of variables for the post-monsoon season. The selection procedure is explained in Table 6 and Table 7. It is found that the subset comprising of (Cl+alk) being the best subset should be employed in the predictive model for post monsoon season.

b) Hardwar District :

Similar procedure has been adopted to carry out the regression modelling for Hardwar District. The selection procedure and evolvement of final regression models have been explained through Table 10 to Table 18.

6.0 MODEL PERFORMANCE AND DISCUSSION OF RESULTS:

The final models for pre-monsoon and post-monsson seasons for Saharanpur District are given by Equ. (20) and Equ. (21) respectively as:

Final model for Pre-monsoon:

CON = -44.54 + 6.84 Ca + 6.03 SO4 + 7.23 NO3

(R^2 = 80%, F = 24.187) ------------------------------(20)

Final model for Post-monsoon:

CON = 147.90 + 8.74 Cl + 1.05 Alk

(R^2 = 79%, F = 35.699) ------------------------------(21)

It may be seen that the statistical models developed in this research perform very well in computing the conductivity levels for both the seasons. The F-values for both the regression models indicate statistically significant regressions. Figure 1 presents the comparison of observed and model computed conductivity levels for the pre-monsoon data which suggests good agreement in observed and model computed conductivity levels. Plot (Figure 2) developed for post-monsoon season also indicate a good fit between the observed and model computed conductivity levels.

In Equ. (20) and (21), the first parameters in each model are, by far, of greater significance. The water quality parameters such as calcium and chloride play major roles in the prediction of conductivity levels. However, calcium is important in the pre-monsoon season and chloride is important in the post-monsoon season. This suggest that a single general model can not be developed for explaining the total variability in the overall conductivity data.

Furthermore, in the pre-monsoon season calcium, sulphate, and nitrate are the primary water quality parameters responsible for the conductivity which depends on the application of fertilizers, whereas, in the post-monsoon season calcium and chloride are the important parameters indicating the source of

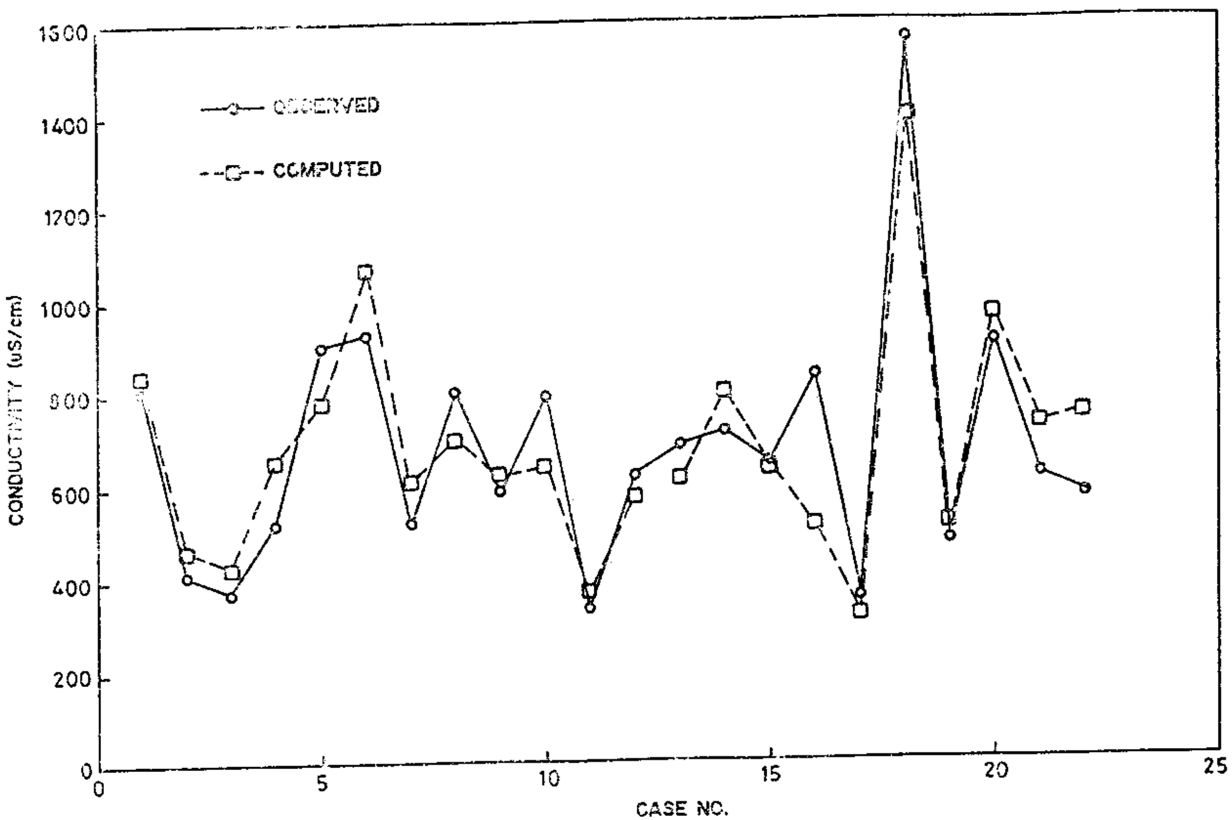

FIG. I COMPARISON OF OBSERVED AND COMPUTED CONDUCTIVITY LEVELS IN THE PRE-MONSOON SEASON FOR SAHARANPUR DISTRICT.

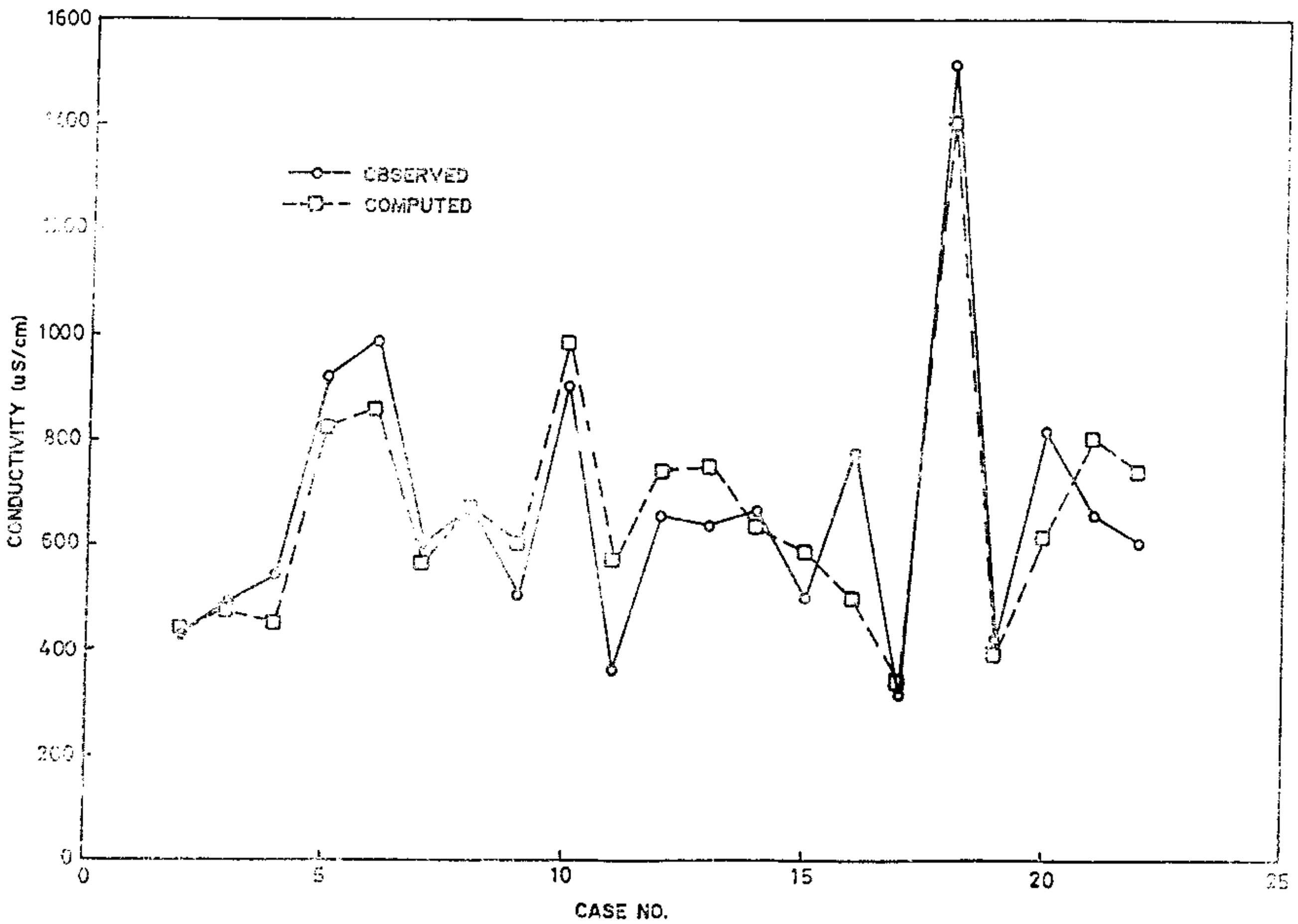

FIG. 2 COMPARISON OF OBSERVED AND COMPUTED CONDUTIVITY LEVELS IN THE POST-MONSOON SEASON FOR SAHARANPUR DISTRICT.

conductivity due to infiltration of salt-rich water which during infiltration dissolves a considerable amount of calcium.

The final models for Hardwar District are given as:

i) Pre-monsoon Season:

CON = -64.375 + 2.01 ALK + 5.206 SO4 + 2.682 Cl

(R^2 = 92%, F = 78.859) ------------------------------(22)

ii) Post-monsoon Season:

CON = 96.275 + 1.871 HARD + 6.057 K

(R^2 = 91%, F = 104.197) -----------------------------(23)

It is observed that both the models for Hardwar are statistically significant explaining more than 90% variability in the conductivity levels. Figure 3 and 4 presents the comparisons between the observed and model computed conductivity levels showing a good agreement between the observed and computed conductivity levels.

In the pre-monsoon season alkalinity is the most significant parameter explaining 76% variability of conductivity levels if taken alone as the explaining regression. While on the otherhand hardness explaining more than 83% conductivity variability alone is the most significant parameter in the post-monsoon season. Here Hardness may also be a function of more than one parameter e.g. bicarbonates, carbonates, sulphates and chlorides of calcium and magnesium. Hardness is selected as a parameter because it is a commonly measured water quality parameter knowing which the conductivity levels may be predicted. However, if one wants to take the basic parameter as regression , than one has to know hardness causing parameters.

The alkalinity and hardness are highly correlated parameters having more than 0.80 correlation coefficient. The reason behind it is that the carbonates and bicarbonates of calcium and sodium and magnesium are the common impurities that cause both alkalinity and hardness in groundwaters. However, there are certain chemical constituents which cause the increase in hardness without increasing the alkalinity levels. The fact is clear from Table-12 which shows that hardness R-square increases from 0.588 to 0.833 while alkalinity R-square decreases from

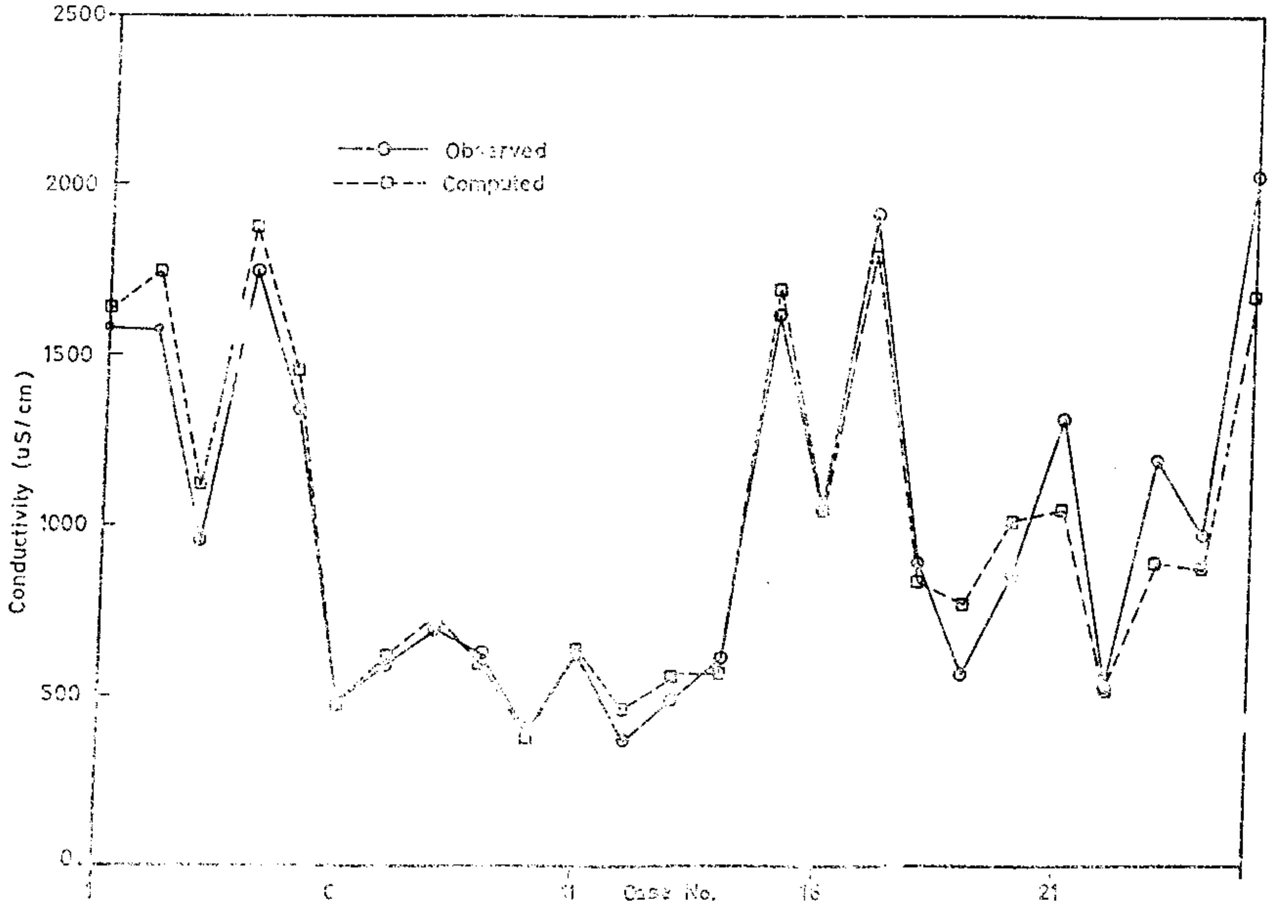

FIG. 3: COMPARISON OF OBSERVED AND COMPUTED CONDUCTIVITY LEVELS IN THE PREMONSOON SEASON FOR HARDWAR DISTRICT.

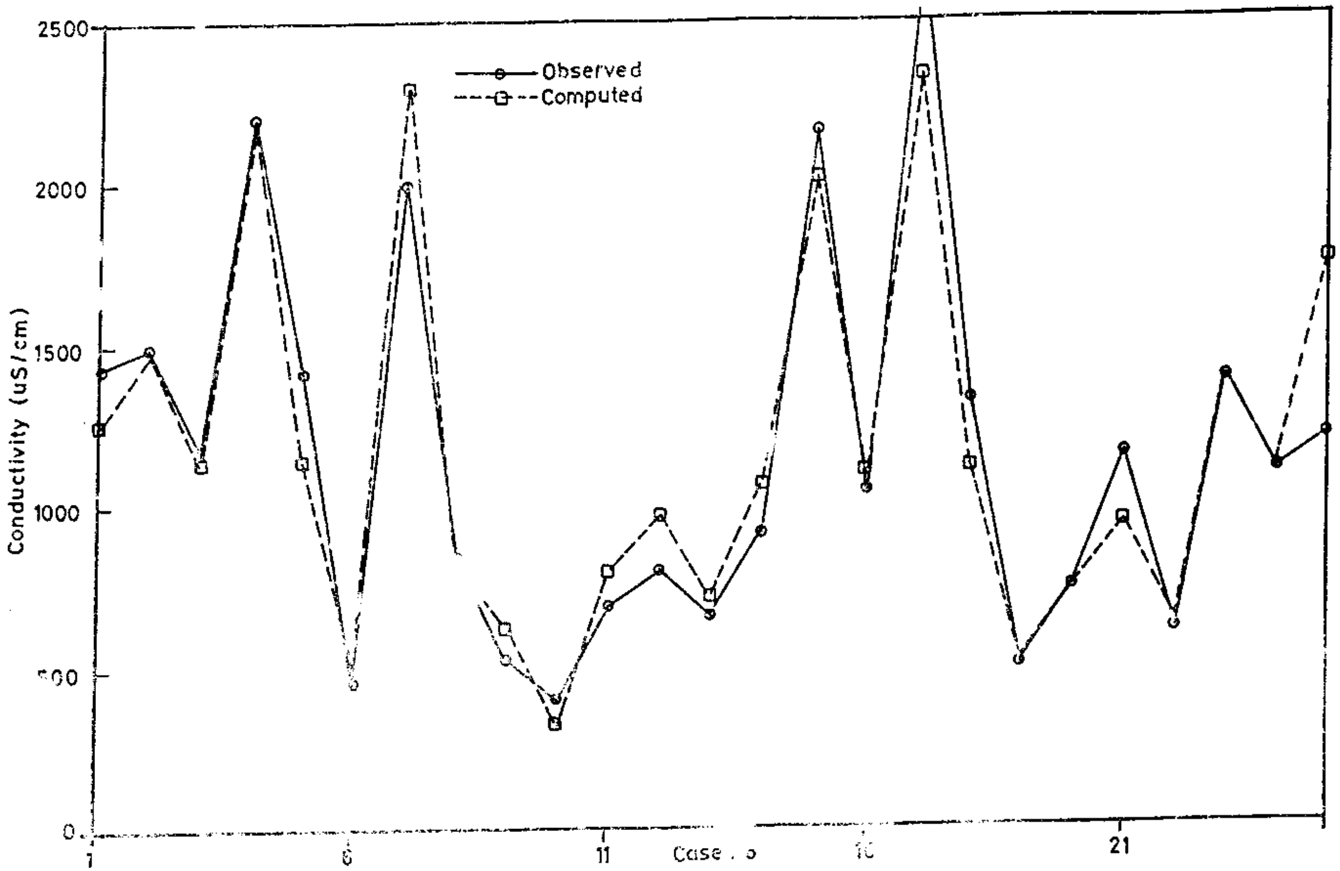

FIG. 4: COMPARISON OF OBSERVED AND COMPUTED CONDUCTIVITY LEVELS IN THE POSTMONSOON SEASON FOR HARDWAR DISTRICT

0.764 to 0.531.

It is noticed that there is not a single model which can be used to predict the conductivity levels. The variation in conductivity not only varies from site to site but also it varies from season to season. However, it is observed that alkalinity and hardness are the parameters which can be used to predict conductivity to a great extent. Because both the alkalinity and hardness are commonly measured water quality parameters, it is suggested both of them should be used to estimate the suitability of ground water with respect to irrigational and other beneficial water uses.

7.0 CONCLUSIONS:

Useful regression models for predicting conductivity concentrations using other constituents were developed for both pre monsoon and post monsoon seasons. As both of the regression models for pre-monsoon and post-monsoon are successful in explaining about 80% of the variation in the conductivity levels, for Saharanpur district and 90% variability in conductivity levels in Hardwar district, the developed models may be used for the prediction of missing observed values. However, the variability of the results from one season to another indicates that a general model can not be derived to predict the total phosphorous concentration for both the seasons. It is evident that in the pre monsoon season the model comprising the subset of (Ca+SO4+NO3) is entirely different from the subset of variables (Cl+alk). Whereas for Saharanpur District both the sets are equally good in explaining the variability in conductivity of about 80%.

Similar fact is also seen for Hardwar District in which (Alk+SO4+Cl) is the best subset for pre-monsoon season and (hard+K) is the best subset for post-monsoon season. This finding is consistent with the knowledge that the major portions of conductivity are influenced by the previously migration pathways at the time and conductivity portion are generated due to the limnological transformations which depend upon fertilizers application, geological formations of the region and other physical conditions at the observation locations.

It is observed that the variation in conductivity causing parameters not only varies from one season to another but also it varies from one site to another. However, it can be concluded that alkalinity and hardness are the two parameters which could be used in most of the cases to predict the conductivity of ground water to assess its salinity effects.

Table 1 : Correlation Coefficient between Water Quality Parameters for Pre-monsoon in Saharanpur District.

	Alk	Hard	Cl	SO4	NO3	Na	K	Ca	Mg
Alk	1.0								
Hard	0.668	1.0							
Cl	0.152	0.405	1.0						
SO4	0.540	0.665	0.009	1.0					
NO3	0.006	0.114	0.638	-0.058	1.0				
Na	0.875	0.363	0.084	0.485	0.008	1.0			
K	0.128	0.461	0.362	0.389	0.475	0.071	1.0		
Ca	0.541	0.826	0.214	0.602	-0.043	0.236	0.522	1.0	
Mg	0.324	0.459	0.406	0.259	0.294	0.237	0.029	-0.080	1.0

Table 2 : Correlation Coefficient between Water Quality Parameters for Post-monsoon, Saharanpur District.

	Alk	Hard	Cl	SO4	NO3	Na	K	Ca	Mg
Alk	1.0								
Hard	0.622	1.0							
Cl	0.400	0.678	1.0						
SO4	0.406	0.683	0.784	1.0					
NO3	0.098	0.348	0.664	0.417	1.0				
Na	0.847	0.250	0.344	0.385	0.050	1.0			
K	0.231	0.567	0.753	0.517	0.696	0.071	1.0		
Ca	0.438	0.878	0.564	0.538	0.401	0.068	0.613	1.0	
Mg	0.358	0.228	0.260	0.301	0.039	0.364	0.053	0.250	1.0

Table 3 : R^2 of Water Quality parameters with Conductivity, Saharanpur District.

	N	Alk	Hard	Cl	SO4	NO3	Na	K	Ca	Mg
Pre-monsoon	22	0.33	0.484	0.108	0.549	0.069	0.229	0.452	0.579	0.014
Post-monsoon	22	0.43	0.540	0.660	0.558	0.354	0.319	0.469	0.430	0.025

Table 4 : Various combinations of models and their statistics for Pre-monsoon Season, Saharanpur District.

Number of variables in model	Variables	R^2	F-value
2	Ca, Alk	0.617	15.305
	Ca, Hard	0.594	13.875
	Ca, Cl	0.608	14.734
	Ca, SO4 *	0.705	22.675
	Ca, NO3	0.667	19.028
	Ca, Na	0.675	19.695
	Ca, K	0.683	20.495
3	Ca, SO4, Alk	0.712	14.833
	Ca, SO4, Hard	0.705	14.335
	Ca, SO4, Cl	0.756	18.575
	Ca, SO4, NO3 *	0.801	24.187
	Ca, SO4, Na	0.733	16.507
	Ca, SO4, K	0.786	22.021
4	Ca, SO4, NO3, Alk	0.806	17.633
	Ca, SO4, NO3, Hard	0.820	19.299
	Ca, SO4, NO3, Cl	0.802	17.204
	Ca, SO4, NO3, Na *	0.826	20.148
	Ca, SO4, NO3, K	0.817	18.958
5	Ca, SO4, NO3, Na, Alk *	0.857	19.125
	Ca, SO4, NO3, Na, Hard	0.853	18.509
	Ca, SO4, NO3, Na, Cl	0.826	17.175
	Ca, SO4, NO3, Na, K	0.817	18.958
6	Ca,SO4, NO3, Na, Alk, Hard	0.862	15.569
	Ca,SO4, NO3, Na, Alk, Cl	0.857	14.942

Table 5 : Selected sets/subsets Candidate for possible model independent variables for Pre-monsoon season, Saharanpur District.

Number of variables	Set of independent	N	R^2	F-value	SSE
6	Ca+SO4+NO3+Na+Alk+K	22	0.863	15.793	207412.336
5	Ca+SO4+NO3+Na+Alk	22	0.857	19.125	217557.272
4	Ca+SO4+NO3+Na	22	0.826	20.148	264397.132
3	Ca+SO4+NO3	22	0.801	24.187	301685.506
2	Ca+SO4	22	0.705	22.675	448145.655
1	Ca	22	0.579	27.551	638389.534

Table 6: Various combinations of models and their statistics for Post-monsoon, Saharanpur District.

Number of variables in model	Variables	R^2	F-value
2	Cl, Alk *	0.790	35.699
	Cl, Hard	0.724	24.930
	Cl, SO4	0.693	21.413
	Cl, NO3	0.667	19.010
	Cl, Na	0.754	29.080
	Cl, SO4	0.693	21.410
	Cl, K	0.674	19.620
	Cl, Ca	0.719	24.240
3	Cl, Alk, Hard	0.795	23.328
	Cl, Alk, SO4	0.804	24.653
	Cl, Alk, NO3 *	0.817	26.869
	Cl, Alk, Na	0.790	22.576
	Cl, Alk, K	0.813	26.168
	Cl, Alk , Ca	0.811	25.672
4	Cl, Alk, NO3, Hard	0.825	20.033
	Cl, Alk, NO3, SO4 *	0.841	22.556
	Cl, Alk, NO3, Na	0.818	19.152
	Cl, Alk, NO3, K	0.827	20.309
	Cl, Alk, NO3, Ca	0.833	21.213
5	Cl, Alk, NO3, SO4, Hard	0.843	17.235
	Cl, Alk, NO3, SO4, Na	0.842	17.011
	Cl, Alk, NO3, SO4, K *	0.854	18.766
	Cl, Alk, NO3, SO4, Ca	0.851	18.315
6	Cl, Alk, NO3, SO4, K, Hard	0.854	14.674
	Cl, Alk, NO3, SO4, K, Na *	0.858	15.080
	Cl, Alk, NO3, SO4, K, Ca	0.858	15.060

Table 7 : Selected sets/subsets Candidate for possible model independent variables for Post-monsoon season, Saharanpur District.

Number of variables (k)	Set of independent	N	R^2	F-value	SSE
6	Cl+Alk+NO3+SO4+K+Na	22	0.858	15.080	202938.825
5	Cl+Alk+NO3+SO4+K	22	0.854	18.766	207899.147
4	Cl+Alk+NO3+SO4	22	0.841	22.556	226264.836
3	Cl+Alk+NO3	22	0.817	26.168	260505.277
2	Cl+Alk	22	0.790	35.699	299953.059
1	Cl	22	0.661	39.083	483084.530

Table 8 : Selection of Model Variables on the basis of F-Statistics, for Pre-monsoon Season, Saharanpur District.

Full model with k-parameters			Reduced model with (k-m) Parameters		k-m	N-k-1	$F_{k-m,N-k-1}$	F^* (α=0.05)	Prefe-red Model
Model	N	SSE	Model	SSE					
Ca+SO4+NO3+Na+Alk+K	22	207412	Ca+SO4+NO3+Na+Alk	217557	5	15	0.147	2.27	Reduced
Ca+SO4+NO3+Na+Alk	22	217557	Ca+SO4+NO3+Na	264397	4	16	0.861	2.33	Reduced
Ca+SO4+NO3+Na	22	264397	Ca+SO4+NO3	301685	3	17	0.799	2.44	Reduced
Ca+SO4+NO3	22	301685	Ca+SO4	448145	2	18	4.37	2.62	Full
Ca+SO4+NO3	22	301685	Ca	638389	1	19	21.21	2.99	Full

Table 9 : Selection of Model Variables on the basis of F-Statistics for Post-monsoon Season, Saharanpur District.

Full model with k-parameters			Reduced model with (k-m) Parameters		k-m	N-k-1	$F_{k-m,N-k-1}$	F^* (α=0.05)	Prefe-red Model
Model	N	SSE	Model	SSE					
Cl+Alk+NO3+SO4+K+Na	22	202938	Cl+Alk+NO3+SO4+K	207899	5	15	0.073	2.27	Reduced
Cl+Alk+NO3+SO4+K	22	207899	Cl+Alk+NO3+SO4	226264	4	16	0.353	2.33	Reduced
Cl+Alk+NO3+SO4	22	226264	Cl+Alk+NO3	260505	3	17	0.857	2.44	Reduced
Cl+Alk+NO3	22	260505	Cl+Alk	299953	2	18	1.36	2.62	Reduced
Cl+Alk	22	299953	Cl	483084	1	19	11.60	2.99	Full

Table 10 : Correlation Coefficient between Water Quality Parameters for Pre-monsoon, Hardwar District.

	Alk	Hard	Cl	SO4	PO4	Na	K	Ca	Mg
Alk	1.0								
Hard	0.808	1.0							
Cl	0.718	0.766	1.0						
SO4	0.023	0.229	0.306	1.0					
PO4	0.822	0.524	0.522	-0.081	1.0				
Na	0.800	0.671	0.865	0.158	0.697	1.0			
K	0.681	0.400	0.523	0.094	0.746	0.585	1.0		
Ca	0.659	0.849	0.693	0.342	0.318	0.537	0.204	1.0	
Mg	0.721	0.708	0.557	0.062	0.639	0.597	0.596	0.266	1.0

Table 11 : Correlation Coefficient between Water Quality Parameters for Post-monsoon, Hardwar District.

	Alk	Hard	Cl	SO4	PO4	Na	K	Ca	Mg
Alk	1.0								
Hard	0.755	1.0							
Cl	0.302	0.706	1.0						
SO4	0.503	0.870	0.674	1.0					
PO4	0.469	0.205	0.199	-0.050	1.0				
Na	0.516	0.453	0.578	0.443	0.275	1.0			
K	0.557	0.408	0.449	0.342	0.667	0.556	1.0		
Ca	0.766	0.948	0.651	0.791	0.280	0.476	0.370	1.0	
Mg	-0.232	-0.112	0.028	0.037	-0.266	-0.170	0.028	-0.404	1.0

Table 12 : R^2 of Water Quality parameters with Conductivity, Hardwar District.

	N	Alk	Hard	Cl	SO4	PO4	Na	K	Ca	Mg
Pre-monsoon	25	0.764	0.588	0.707	0.138	0.408	0.674	0.468	0.546	0.334
Post-monsoon	25	0.531	0.833	0.626	0.621	0.121	0.383	0.378	0.716	0.001

Table 13 : Various combinations of models and their statistics for Pre-monsoon Season, Hardwar District.

Number of variables in model	Variables	R^2	F-value
2	Alk, Ca	0.811	47.20
	Alk, Cl	0.858	66.33
	Alk, Hard	0.775	37.82
	Alk, K	0.779	38.67
	Alk, Mg	0.769	36.69
	Alk, Na	0.805	45.38
	Alk, PO4	0.783	39.74
	Alk, SO4 *	0.887	86.72
2	Ca, Cl	0.754	33.71
	Ca, Hard	0.616	17.68
	Ca, K *	0.843	59.17
	Ca, Mg	0.703	26.05
	Ca, Na	0.799	43.62
	Ca, PO4	0.728	29.39
	Ca, SO4	0.562	14.13
2	Cl, Hard	0.744	31.91
	Cl, K	0.789	41.12
	Cl, Mg	0.725	28.93
	Cl, Na	0.741	31.54
	Cl, PO4	0.762	35.18
	Cl, SO4	0.721	28.47
2	Hard, K	0.758	34.51
	Hard, Mg	0.591	15.91
	Hard, PO4	0.666	21.97
	Hard, SO4	0.629	18.68
2	Na, SO4	0.733	30.27
	Na, PO4	0.682	23.60
	Na, Mg	0.685	23.60
	Na, K	0.737	30.78
	Na, Hard	0.759	34.59
3	Alk, SO4, Ca	0.893	58.17
	Alk, SO4, Cl *	0.918	78.859
	Alk, SO4, Hard	0.888	55.47

	Alk, SO4, K	0.895	59.39
	Alk, SO4, Mg	0.897	60.93
	Alk, SO4, Na	0.903	65.11
	Alk, SO4, PO4	0.894	58.82
4	Alk, SO4, Cl, Ca	0.919	56.66
	Alk, SO4, Cl, Hard	0.927	63.52
	Alk, SO4, Cl, K	0.925	61.54
	Alk, SO4, Cl, Mg *	0.930	66.629
	Alk, SO4, Cl, Na	0.919	56.36
	Alk, SO4, Cl, PO4	0.922	59.15
5	Alk, SO4, Cl, Mg, Ca	0.932	51.97
	Alk, SO4, Cl, Mg, Hard	0.934	53.85
	Alk, SO4, Cl, Mg, K *	0.941	60.536
	Alk, SO4, Cl, Mg, Na	0.930	50.667
	Alk, SO4, Cl, Mg, PO4	0.932	52.28
6	Alk, SO4, Cl, Mg, K, Ca	0.941	48.21
	Alk, SO4, Cl, Mg, K, Hard	0.941	47.82
	Alk, SO4, Cl, Mg, K, Na	0.941	47.79
	Alk, SO4, Cl, Mg, K, PO4*	0.952	59.74

Table 14 : Selected sets/subsets Candidate for possible model independent variables for Pre-monsoon season, Hardwar District.

Number of variables	Set of independent	N	R^2	F-value	SSR	SSE
6	Alk+SO4+Cl+Mg+K+PO4	25	0.952	59.74	5886967	295652
5	Alk+SO4+Cl+Mg+K	25	0.941	60.54	5817443	365176
4	Alk+SO4+Cl+Mg	25	0.930	66.63	5751048	431571
3	Alk+SO4+Cl	25	0.918	78.859	5675854	504065
2	Alk+SO4	25	0.887	86.72	5486664	695955
1	Alk	25	0.764	74.38	4722341	1460278

Table 15 : Various combinations of models and their statistics for Post-monsoon, Hardwar District.

Number of variables in model	Variables	R^2	F-value
2	Hard, Alk	0.838	56.849
	Hard, Ca	0.838	57.000
	Hard, Cl	0.878	78.140
	Hard, K	0.905	104.197
	Hard, Mg	0.840	57.870
	Hard, Na	0.888	86.860
	Hard, PO4	0.862	68.500
	Hard, SO4	0.834	55.460
2	Ca, Cl	0.816	48.625
	Ca, K	0.821	50.380
	Ca, Alk	0.731	29.950
	Ca, Mg	0.835	55.480
	Ca, Na	0.776	38.130
	Ca, PO4	0.729	29.580
	Ca, SO4	0.753	33.548
2	Cl, Alk *	0.890	89.000
	Cl, K	0.710	26.910
	Cl, Mg	0.630	18.540
	Cl, Na	0.665	21.815
	Cl, PO4	0.664	21.700
	Cl, SO4	0.744	32.000
2	SO4, PO4	0.772	37.190
	SO4, Na	0.711	27.100
	SO4, Mg	0.623	18.210
	SO4, K	0.756	34.040
	SO4, Alk	0.769	36.540

3	Hard, K, Alk	0.908	68.720
	Hard, K, Ca	0.907	68.070
	Hard, K, Cl *	0.925	86.750
	Hard, K, Mg	0.908	68.790
	Hard, K, Na	0.920	80.250
	Hard, K, PO4	0.905	66.390
	Hard, K, SO4	0.905	66.330
3	Cl, Alk, Hard	0.912	72.390
	Cl, Alk, K	0.892	57.720
	Cl, Alk, SO4	0.909	70.240
4	Hard, K, Cl, Alk	0.929	65.33
	Hard, K, Cl, Ca	0.927	63.24
	Hard, K, Cl, Mg	0.927	63.17
	Hard, K, Cl, Na *	0.932	68.02
	Hard, K, Cl, PO4	0.925	61.99
	Hard, K, Cl, SO4	0.926	62.92
5	Hard, K, Cl, Na, Alk	0.932	52.04
	Hard, K, Cl, Na, Ca *	0.935	54.70
	Hard, K, Cl, Na, Mg	0.935	54.65
	Hard, K, Cl, Na, PO4	0.932	51.86
	Hard, K, Cl, Na, SO4	0.933	52.98
6	Hard, K, Cl, Na, Ca, Alk	0.936	44.11
	Hard, K, Cl, Na, Ca, Mg	0.935	43.24
	Hard, K, Cl, Na, Ca, PO4	0.937	44.74
	Hard, K, Cl, Na, Ca, SO4 *	0.938	45.44

Table 16 : Selected sets/subsets Candidate for possible model independent variables for Post-monsoon season, Hardwar District.

Number of variables	Set of independent	N	R^2	F-value	SSR	SSE
6	Hard+K+Cl+Na+Ca+SO4	25	0.939	45.44	8011112	528878
5	Hard+K+Cl+Na+Ca	25	0.935	54.70	7985301	554689
4	Hard+K+Cl+Na	25	0.932	68.02	7955233	584757
3	Hard+K+Cl	25	0.925	86.75	7902331	637659
2	Hard+K	25	0.905	104.19	7724517	815473
1	Hard	25	0.834	115.81	7124995	1414995

Table 17 : Selection of Model Variables on the basis of F-Statistics, Hardwar District for Pre-monsoon season.

Full model with k-parameters			Reduced model with (k-m) Parameters		k-m	N-k-1	$F_{k-m,N-k-1}$	F^*	Prefe-red Model
Model	N	SSE	Model	SSE				(α=0.05)	
Alk+SO4+Cl+Mg+K+PO4	25	295652	Alk+SO4+Cl+Mg+K	365176	5	18	0.846	2.77	Reduced
Alk+SO4+Cl+Mg+K	25	365176	Alk+SO4+Cl+Mg	431571	4	19	0.864	2.90	Reduced
Alk+SO4+Cl+Mg	25	431571	Alk+SO4+Cl	504065	3	20	1.12	3.10	Reduced
Alk+SO4+Cl	25	504065	Alk+SO4	448145	2	21	3.99	3.47	Full
Alk+SO4+Cl	25	504065	Alk	1460278	1	21	39.84	4.32	Full

Table 18 : Selection of Model Variables on the basis of F-Statistics for Post-monsoon Season, Hardwar District for Post-monsoon season.

Full model with k-parameters			Reduced model with (k-m) Parameters		k-m	N-k-1	$F_{k-m,N-k-1}$	F^*	Prefe-red Model
Model	N	SSE	Model	SSE				(α=0.05)	
Hard+K+Cl+Na+Ca+SO4	25	528878	Hard+K+Cl+Na+Ca	554689	5	18	0.176	2.77	Reduced
Hard+K+Cl+Na+Ca	25	554689	Hard+K+Cl+Na	584757	4	19	0.258	2.90	Reduced
Hard+K+Cl+Na	25	584757	Hard+K+Cl	637659	3	20	0.603	3.10	Reduced
Hard+K+Cl	25	637659	Hard+K	815473	2	21	2.928	3.47	Reduced
Hard+K	25	815473	Hard	1414995	1	22	16.174	4.30	Full

REFERENCES:

1. Bouwer, H. (1978), Groundwater contamination. McGraw-Hill Book Co., Inc., New York, N.Y.

2. Canter, L. W., and Knox, R. C. (1985), Septic tank system effects on groundwater quality. Lewis Publishers, Inc. , Chelsea, Mich.

3. Draper, N. R. and Smith, H. (1981), Applied Regression Analysis. Second Edition, John Wiley and Sons, New York.

4. Foster, M. D. (1942), Chemistry of Groundwater in Hydrology (O.E. Meinzer, ed.), pp.646-655, McGraw Hill, New York.

5. Gower, A.M. (1980). Water Quality in a Catchment Ecosystem, John Wiley and Sons Ltd., New York,N.Y., USA.

6. Jain, C. K., Ram, D. and Bhatia, K. K. S. (1996). Ground Water Quality Monitoring and Evaluation in District Hardwar, UP, Tech. Report, National Institute of Hydrology, Roorkee.

7. Maliwal, G.L. (1968), Effect of irrigation waters of different qualities on soils and crops. Ph.D. Thesis, Udaipur University, Udaipur.

8. Mehta, K.K.(1970), Studies on the effect of irrigation water on soil properties and crop growth with special reference to boron. Ph.D. Thesis, Udaipur University, Udaipur.

9. Montgomery, D. and Peck, E. (1982), Linear Regression Analysis., Wiley, New York, N.Y.

10. Novotny, V. and Chesters G. (1981), Handbook of Non-point Pollution, Van Nostrand, New York, N.Y.

11. Paliwal, K.V.(1972), Irrigation with saline water, IARI, Monograph-2, Water Technology Centre, IARI, New Delhi.

12. Sandhu, S. S., Warren, W. J. and Nelson, P.(1977), Trace Inorganics in rural potable water and their correlation to possible sources. Water Res. 12(4)., 257-261.

13. Singh, R. P., Thakur, K. S., Singh, V. and Simon, S. K.(1979), Technical Memorandum No. 69, Ground Water Directorate of UP, Roorkee.

14. Kumar, Sudhir, Jain, C. K. and Bhatia, K. K. S. (1988). Ground Water Quality Variations in Saharanpur District (UP), Tech. Report no. TR-50, National Institute of Hydrology, Roorkee.

15. Weisberg, S. (1980), Applied Linear Regression. John Wiley and Sons, New York, N.Y.

DIRECTOR : DR. S. M. SETH

DIVISIONAL HEAD : DR. K. K. S. BHATIA

STUDY GROUP : SH. ADITYA TYAGI
SH. M. K. SHARMA
DR. K. K. S. BHATIA