# Correlation Study and Regression Analysis of Water Quality Assessment of Nagpur City, India

**[1]Soni Chaubey and [2]Mohan Kumar Patil**

[1]PhD Research Scholar Mewar University NH-76 Gangrar, Chhittorghara Rajasthan, India
[2]Dr. Mohan kumar Patil, Senior Environment Professional, En Carp Solutions, Nagpur, Maharashtra, India

***Abstract-*** The present study has determined water quality assessment on the basis of physico-chemical analysis such as pH, Electrical Conductivity (EC), Turbidity (Turb) ,Total Dissolved Solids ( TDS), Total Hardness (TH), Calcium ($Ca^{2+}$), Magnesium ($Mg^{2+}$), Chloride ($Cl^-$), Sulphate ($SO_4$) ,Total Alkalinity (TA) and Dissolved Oxygen (DO) etc., using a statistical method. There is a relationship between variables which shows that one variable actually causes changes in another variable. In this paper, a statistical regression analysis method of fifty two data points of drinking water in different source at four fields (i.e, Khaparkheda Water Supply, Koradi Gram Panchayat, Koradi Devi Mandir and Bokara) for Nagpur city with hot and dry climate, in Maharashtra, center of India was carried out. Samples were collected during October 2013 to May 2014. The temperature of the water Samples is in the range 25-30°C. This technique was based on the study and calculating the correlation coefficients between various physicochemical parameters of drinking water. The results were further compared with drinking water quality standards issued by World Health Organization (WHO) and it was deduced that most of the water samples are not potable. The results proved to be a useful mean for rapid monitoring of water quality with the help of systematic calculations of correlation coefficient between water parameters and regression analysis.

***Index Terms***- Regression equation, Water quality parameters, Correlation coefficient method

## I. Introduction

Pollution is generally regarded as the result of the industrial revolution. The revolution introduced various industrial activities that rendered the environmental quality of the area concerned deteriorated. The activities further gave birth to different sources of pollution that are essential to be identified on the first hand for exploring the current status of the pollution in the area. On an average a human being consumes about two liters of water every day during his whole life period .. All segments of the environment are being polluted by various ways. However water pollution has been taken under inspection since water forms an integral part of life on earth. The most prominent factors that elevates the level of water pollution are exploding population, increasing industrialization and urbanization. Various treatment methods are adopted to raise the quality of drinking water. Water should be free from the various detoxifications such as Organic and Inorganic pollutants, Pesticides, Heavy metals etc. As well as all its parameter like pH, Dissolved Oxygen, Chloride, Total Dissolved Solid, Total Alkalinity, Calcium, Magnesium, Total Hardness, Nitrate and Electrical Conductivity should be within acceptable limit. A novel approach of regression method is adopted to assess quality of water.

## II. Materials and Methods

Drinking ground water samples were collected from four different fields for City namely: (a) Field I (Khaparkheda Water Supply) (b) Field II ( Koradi Gram Panchayat) (c) Field III (Koradi Devi Mandir) and (d) Field IV (Bokara) as in Table 1. The testing was done in the Hydrology project division water quality lab, Nagpur, Maharashtra.

### Collection, Preparation of Water Samples and Analysis

One liter water samples were taken in transparent plastic bottles. Each sample was characterized for water quality throughout for four permanent stations . Laboratory samples were stored in ice at 4°C until transport for analysis.

### Water Quality Analysis

The water samples were analyzed for physico-chemical (i.e. pH, free carbon dioxide, dissolved oxygen, biochemical oxygen demand, total alkalinity, hardness, calcium, magnesium, phosphate and ammonia). The pH was evaluated by EUTECH instrument pH-510. In the laboratory electrical conductivity was calculated with microprocessor based on the conductivity TDS meter. The temperature was recorded with laboratory thermometer. The above named instruments have been used in the limits of precise accuracy and chemicals used were of analytical grade. Statistical analysis of water samples from different sources as shown in Table 2.

**Table 1 Physicochemical parameters of drinking water at studied source location for Nagpur city (Note: all parameters are in mg/l except pH, EC in µ mhos/cm and Turb in NTU)**

| Source No. | Field | pH | DO | EC U mhos | Turb | TA | TDS | TH | $Ca^{2+}$ | $Mg^{2+}$ | Cl | BOD | $NO_3^-$ | $SO_4$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S1 | I | 8.3 | 3.5 | 354 | 2 | 204 | 220 | 134 | 32.9 | 12.6 | 16 | 2.08 | 1.2 | 21.4 |
| S2 | | 7.4 | 4 | 859 | 4 | 296 | 542 | 336 | 92.2 | 25.8 | 58 | 3.0 | 2.4 | 57.1 |
| S3 | | 7.7 | 3.5 | 908 | 1 | 380 | 534 | 354 | 40.9 | 61.2 | 56 | 3.5 | 1.5 | 55.9 |
| S4 | | 8.36 | 5.46 | 372 | 1 | 120 | 304 | 102 | 30.8 | 26.3 | 90 | 2.05 | 1.6 | 59 |
| S1 | II | 7.4 | 4 | 859 | 4 | 296 | 542 | 336 | 92.2 | 25.8 | 58 | 3.0 | 1.4 | 57.1 |
| S2 | | 8.36 | 5.46 | 372 | 1 | 120 | 102 | 108 | 86.3 | 90 | 22 | 2.00 | 0.4 | 20 |
| S3 | | 7.32 | 5.36 | 492 | 1 | 142 | 170 | 32 | 21.8 | 24 | 30 | 2.8 | 1.0 | 25 |
| S4 | | 7.42 | 5.36 | 482 | 1 | 132 | 360 | 170 | 32 | 21.8 | 24 | 8.8 | 1.36 | 20 |
| S1 | III | 7.7 | 3.5 | 608 | 2 | 380 | 534 | 354 | 40.9 | 61.2 | 56 | 3.5 | 1.5 | 25.9 |
| S2 | | 7.62 | 5.36 | 492 | 1 | 142 | 170 | 32 | 21.8 | 24 | 30 | 2.8 | 1.60 | 36 |
| S3 | | 7.82 | 6.07 | 566 | 1 | 130 | 262 | 94 | 28 | 5.83 | 47 | 4.6 | 1.19 | 46 |
| S4 | | 8.39 | 5.82 | 362 | 1 | 78 | 200 | 100 | 25.6 | 8.74 | 26 | 4 | 0.264 | 30 |
| S1 | IV | 7.9 | 3 | 995 | 4 | 360 | 578 | 390 | 68.1 | 53.5 | 89 | 1.1 | 2.2 | 43 |
| S2 | | 8.2 | 6.8 | 488 | 1 | 60 | 396 | 100 | 24 | 4.8 | 27 | 6 | 1.1 | 12.6 |
| S3 | | 7.4 | 5.5 | 408 | 1 | 380 | 534 | 354 | 40.9 | 61.2 | 56 | 3.5 | 1.5 | 55.9 |
| S4 | | 0.72 | 4.2 | 759 | 4 | 396 | 442 | 436 | 92.2 | 25.8 | 58 | 3.0 | 1.4 | 55.1 |

**Table 2 Statistical analysis of water samples from different source**

| Parameter | Max | Min | Range | Mean | SD |
|---|---|---|---|---|---|
| pH | 8.39 | 7.2 | 1.19 | 7.780625 | 0.420863695 |
| DO | 6.8 | 3 | 3.8 | 4.805625 | 1.122639561 |
| EC | 995 | 354 | 641 | 586 | 218.5107167 |
| Turb | 4 | 1 | 3 | 1.875 | 1.310216267 |
| TA | 396 | 60 | 336 | 226 | 124.4679343 |
| TDS | 578 | 102 | 476 | 368.125 | 164.9783319 |
| TH | 436 | 32 | 404 | 214.5 | 143.1968342 |
| $Ca^{2+}$ | 92.2 | 21.8 | 70.4 | 48.1625 | 27.70453332 |
| $Mg^{2+}$ | 90 | 4.8 | 85.2 | 33.285625 | 24.60307731 |
| Cl | 90 | 16 | 74 | 46.4375 | 22.72727803 |
| BOD | 8.8 | 1.1 | 7.7 | 3.483125 | 1.819420123 |
| $NO_3^-$ | 2.4 | 0.264 | 2.136 | 1.350875 | 0.537747385 |
| $SO_4$ | 59 | 12.6 | 46.4 | 38.75 | 16.60517991 |

**2.2 Linear Regression Model**

Based on the literature review, the relationship of water quality parameters on each other in the samples of water analyzed was determined by determining correlation coefficients (r) by using the mathematical formula as given below. Let x and y be any two variables (water quality parameters in the present investigation) and n = number of observations. Then the correlation coefficient (r), between the variables x and y is given by the relation.

$$r = \frac{n\sum(x*y) - \sum x * \sum y}{[f(x)*f(y)]^{0.5}}$$

Where

$f(x)$=n $\sum (x^2) - (\sum x)^2$ $f(x)$=n$\sum(y^c)$-$(\sum y)^c$ and all the summations are to be taken from 1 to n.

If the numerical value of the correlation coefficient between two variables x and y is fairly large, it implies that these two variables are highly correlated. In such cases, it is feasible to try a linear relation of the form

y = Ax +B ----------(1)

To correlate x and y, the constant A and B are to be determined by fitting the experimental data on the variables x and y to equation (1). According to the well-known method of least squares, the value of constants A and B are given by the relations

And $B = \bar{y} - A\bar{x}$ .....(2)

$$A = \frac{n\sum(x*y) - \sum x * \sum y}{n\sum(x-\bar{x})^2}$$

Where $\bar{x} = \frac{\sum x}{n}$ $\bar{y} = \frac{\sum y}{n}$

By using these relations, with the help of Microsoft Excel finding the values of correlation **coefficients (r)** which has been given below in Table 3.

**Table 3 Correlation coefficients (r) among various water quality parameters**

| | pH | DO | EC | Turb | TA | TDS | TH | $Ca^{2+}$ | $Mg^{2+}$ | Cl | BOD | $NO_3^-$ | $SO_4$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pH | 1 | | | | | | | | | | | | |
| DO | 0.245 | 1 | | | | | | | | | | | |
| EC | -0.486 | -0.664 | 1 | | | | | | | | | | |
| Turb | -0.39 | -0.665 | 0.739 | 1 | | | | | | | | | |
| TA | -0.549 | -0.768 | 0.675 | 0.598 | 1 | | | | | | | | |
| TDS | -0.458 | -0.546 | 0.742 | 0.59 | 0.794 | 1 | | | | | | | |
| TH | -0.484 | -0.695 | 0.732 | 0.703 | 0.935 | 0.886 | 1 | | | | | | |
| $Ca^{2+}$ | -0.292 | -0.474 | 0.575 | 0.796 | 0.509 | 0.411 | 0.641 | 1 | | | | | |
| $Mg^{2+}$ | -0.017 | -0.355 | 0.167 | 0.001 | 0.481 | 0.22 | 0.406 | 0.379 | 1 | | | | |
| Cl | -0.158 | -0.413 | 0.566 | 0.472 | 0.552 | 0.646 | 0.584 | 0.327 | 0.232 | 1 | | | |
| BOD | -0.187 | 0.47 | -0.198 | -0.36 | -0.326 | 0.02 | -0.172 | -0.353 | -0.364 | 0.539 | 1 | | |
| $NO_3^-$ | -0.461 | -0.536 | 0.65 | 0.567 | 0.569 | 0.691 | 0.547 | 0.279 | 0.038 | -0.174 | -0.36 | 1 | |
| $SO_4$ | -0.386 | -0.317 | 0.521 | 0.434 | 0.566 | 0.528 | 0.558 | 0.418 | 0.093 | -0.414 | 0.659 | 0.751 | 1 |
| Strong 1 Moderate 17 Weak 28 Negative 31 | | | | | | | | | | | | | |

The correlation coefficient (r) measures the degree of association that exists between two variables, one taken as dependent variable. The greater the value of regression coefficient, the better is the fit and more useful the regression variables (***Daraigan Sami G.,***2011).Correlation is the mutual relationship between two variables. Direct correlation exists when increase or decrease in the value of one parameter is associated with a corresponding increase or decrease in the value of other parameter (**K. Jothivenkatachalam,** 2010). In this study, the numerical values of correlation coefficient (r) for the thirteen water quality parameters are tabulated in Table 3.

## III. Result and Discussions

In the studied area, water used for drinking purposes should be colourless, odourless and free from slight turbidity and excess salts. The taste of the water is slightly brackish at some of the locations. The temperature of the water is in the range 25-30° C. The important physico-chemical characteristics of\ analyzed water samples viz., Mean and Standard Deviation (SD) have been presented in Table-2 and the values are compared with standard parameters inTable-3. It shows that variation among the measured values of these parameters at different locations is not too high and variation range is very narrow.

The regression equation was used as a mathematical tool to calculate different dependent characteristics of water quality by substituting the values for the independent parameters in the equations. The regression analyses carried out for the water quality parameters found to have better and higher level of significance in their correlation coefficient as shown in Table 4.

In the current study it is evident from the Table 4 given below that distribution of pH ,dissolved oxygen DO, electrical conductivity EC, total hardness TH, calcium $Ca^{2+}$, magnesium $Mg^{2+}$, sulphate $SO_4$ , and chloride $Cl^-$ were significantly correlated (R > .41) with total dissolved solids ( TDS) and $Cl^-$ were significantly correlated (R > 0.32) with electrical conductivity EC . A considerably low correlation was observed between EC and $Cl^-$ (R=0.32) and $Ca^{2+}$ and EC (R=0.33). A high correlation value was observed between TH and TA (R=0.87).

The low correlation value is found between electrical conductivity and dissolved oxygen (R=0.44), Turb and DO (R=0.44), TA and DO (R=0.59), TH and DO (R=0.48). The pH, DO and BOD are negatively correlated with most of the water parameters. The results showed that regression relations have the same correlation coefficients, as: (i) EC and DO, Turb and DO (R=0.44) (ii) $Ca^{2+}$ and TH, Cl and TDS (R=0.41). Finally, it can be concluded that the correlation studies of the water quality parameters have great significance in the study of water resources.

## IV. Conclusion

The statistical regression analysis has been found to be a highly useful technique. Finding linear correlation between various physicochemical water parameters can be treated as a unique step ahead towards the drinking water quality management. The mathematical models used to access water quality involve two parameters to describe realistic groundwater situations. This technique has been proven as a very useful tool for monitoring drinking water and has a good accuracy. A significant relationship obtained from a systematic correlation and regression in this study has been established among different pairs of physico-chemical parameters. The method of linear correlation has been found to a significant approach to get an idea of quality of the ground water by determining a few parameters experimentally. It can be concluded that the dissolved oxygen and electrical conductivity are important physicochemical of drinking water quality parameters, because they are correlated with most of the water quality parameters. This study has revealed the facts that all the physicochemical parameters of drinking water in Nagpur city are correlated in some or the other ways. The study could be more enhanced by studying groundwater quality movement in the near future.

**Table 4 Linear correlation coefficient and regression equation for some pairs of parameters which have significant value of correlation**

| Pairs of parameters | Regression coefficient | | Regression equation | R square | P-value | F-value |
|---|---|---|---|---|---|---|
| | A | B | | | | |
| EC- DO | 129 | 1207 | EC = 1207 - 129 DO | 44.1% | 0.005 | 11.03 |
| Turb-DO | - 0.776 | 5.61 | Turb = 5.61 - 0.776 DO | 44.1% | 0.005 | 11.12 |
| Turb-EC | 0.00443 | -0.722 | Turb =- 0.722 +0.00443EC | 54.6% | 0.001 | 16.85 |
| TA-DO | - 85.2 | 635 | TA = 635 - 85.2 DO | 59.1% | 0.001 | 20.19 |
| TA-EC | 0.384 | 0.7 | TA = 0.7 + 0.384 EC | 45.6% | 0.004 | 11.72 |
| TA- Turb | 56.9 | 119 | TA = 119 + 56.9 Turb | 35.8% | 0.014 | 7.81 |
| TDS-EC | 0.560 | 39.8 | TDS = 39.8 + 0.560 EC | 55.1% | 0.001 | 17.16 |
| TDS-TA | 1.05 | 130 | TDS = 130 + 1.05 TA | 63.0% | 0.000 | 23.80 |
| TH-DO | - 88.7 | 641 | TH = 641 - 88.7 DO | 48.3% | 0.003 | 13.09 |
| TH-EC | 0.479 | - 66.4 | TH = - 66.4 + 0.479 EC | 53.5% | 0.001 | 16.12 |
| TH-Turb | 76.9 | 70.4 | TH = 70.4 + 76.9 Turb | 49.4% | 0.002 | 13.69 |
| TH-TA | 1.08 | - 28.6 | TH = - 28.6 + 1.08 TA | 87.4% | 0.000 | 97.35 |
| TH-TDS | 0.769 | - 68.5 | TH = - 68.5 + 0.769 TDS | 78.5% | 0.000 | 50.97 |
| $Ca^{2+}$ - EC | 0.0728 | 5.5 | Ca2+ = 5.5 + 0.0728 EC | 33.0% | 0.020 | 6.90 |
| $Ca^{2+}$ Turb | 16.8 | 16.6 | Ca2+ = 16.6 + 16.8 Turb | 63.4% | 0.000 | 24.28 |
| $Ca^{2+}$ -TH | 0.124 | 21.6 | Ca2+ = 21.6 + 0.124 TH | 41.1% | 0.007 | 9.77 |
| Cl- EC | 0.0589 | 11.9 | Cl = 11.9 + 0.0589 EC | 32.1% | 0.022 | 6.60 |
| Cl-TDS | 0.0889 | 13.7 | Cl = 13.7 + 0.0889 TDS | 41.1% | 0.007 | 10.01 |
| Cl-TH | 0.0926 | 26.6 | Cl = 26.6 + 0.0926 TH | 34.1% | 0.018 | 7.23 |

**Table 5 Comparison of ground water quality with drinking water standards**

| Parameter | USPH | WHO | European standard | ICMR | Present study |
|---|---|---|---|---|---|
| pH | 6.0-8.5 | 6.5-9.2 | 6.5-8.5 | 6.5-8.5 | 7.2-8.39 |
| EC, µ mho/cm | 300 | 300 | 400 | - | 995-354 |
| TDS | 500 | 500 | 500 | 500-1500 | 578-102 |
| Total Hardness | 500 | - | - | 300 | 436-32 |
| Calcium | 100 | 75 | 100 | 75 | 92.2-21.8 |
| Magnesium | 30 | 50 | - | 50 | 90-4.8 |
| Chloride | 250 | 200 | 250 | 250 | 90-16 |
| Total Acidity | - | - | - | - | 4-37 |
| Total Alkalinity | - | - | - | - | 396-60 |
| Sulphate | 250 | 200 | - | 200 | 59-12.6 |

USPH-United State of Public Drinking water Standard

WHO- World Health organization

ICMR-Indian Council of Medical Research

## REFERENCES

[1] Anita Bhatnagar and Pooja Devi, 2012. "Applications of correlation and regression analysis in assessing lentic water quality: a case study at Brahmsarovar Kurukshetra, India", International Journal of Environmental Sciences, Volume 3, No 2, Department of Zoology, Kurukshetra University, Kurukshetra, India.

[2] Soni Chaubey and Mohan Kumar Patil, 2015. "Applications of Artificial Neural Network in Assessing Water Quality: A Case Study", International Journal of Current Research Vol. 7, Issue, 01, pp.11403-11407, India

[3] El-shafie A., M. Mukhlisin, Ali. A, Najah and M. R. Taha, 2011. "Performance of artificial neural network and regression techniques for rainfall-runoff prediction", International Journal of the Physical Sciences, Vol. 6(8), pp. ISSN 1992 - 1950 1Department of Civil and Structural Engineering, University Kebangsaan Malaysia, Malaysia.

[4] Kavita Gupta, S. C. Verma, Meena Thakur and Aakriti Chauhan, 2014. "Impact of Land Uses on Surface Water Quality and Associated Aquatic Insects at Parwanoo Area of Solan District of Himachal Pradesh, India", International Journal of Bio-resource and Stress Management, 5(3):427-431 Dept. of Environmental Science, Dr Y S Parmar University of Horticulture and Forestry, Nauni, Solan, HP ,India

[5] Sami G. Daraigan , Ahmed S. Wahdain , Ahmed S. BaMosa and Manal H. Obid 2011"Linear correlation analysis study of drinking water quality data for Al -MukallaCity, Hadhramout, Yemen" International Journal of Environmental Sciences, Science and Technology (HUST),Mukalla , Hadhramout, Yemen

[6] Purushottam J. Puri, M.K.N. Yenkie, D. G. Battalwar, Nilesh V. Gandhare and Dewanand B. Dhanorkar (2010) "Study and Interpretation of Physico-Chemical Characteristic of Lake Water Quality in Nagpur City (India)", Department of Chemistry, L.I.T., RTM, Nagpur University, Nagpur, India.

[7] K. Jothivenkatachalam, A. Nithya and S. Chandra Mohan (2010)"Correlation Analysis of Drinking Water Quality in And Around Perur Block of Coimbatore District, Tamil Nadu, India", .rasayanjournal Vol.3, No.4 (2010), 649-654 Department of Chemistry, Anna University of Technology Tiruchirappalli,

## AUTHORS

**First Author** – Soni Chaubey, PhD Research Scholar Mewar University NH-76 Gangrar, Chhittorghara Rajasthan, India
**Second Author** – Mohan Kumar Patil, Dr. Mohan kumar Patil, Senior Environment Professional, En Carp Solutions, Nagpur, Maharashtra, India