

Statistical Models and Water Quality Characterization of Source Waters in Nepal

Iswori L. Shrestha¹, Srijan L. Shrestha²

¹Nepal Environmental and Scientific Service (P) Ltd. (NESS), Kathmandu

²Central Department of Statistics, Tribhuvan University Kirtipur, Kathmandu

e-mail: srijan_shrestha@yahoo.com

Abstract

Available sample data of 270 water samples collected from different intake sources of Nepal is analyzed. Analysis showed that well waters were the hardest with more than 60% samples recording hardness above 180 mg/l whereas stream and spring waters showed only 20% and 10% in the same range, respectively. Total dissolved solid (TDS) contents in spring and stream waters were below 500 mg/l while wells waters showed 16% above 500 mg/l. More than 80% from all sources showed pH values within WHO limit. Electric conductivity (EC) was found very low in almost all samples from stream and spring waters. Only well waters showed about 55% with EC in between 500 and 1000 μ mhos/cm. Spring and stream waters were corrosive compared to well waters. In well waters 60% were found stable and 15% scalable. Two nonlinear statistical models, a log-linear and a Box-Cox transformed model are found appropriate for estimating total hardness and TDS, respectively with predictor variables like temperature, pH, and electric conductivity. Models are useful for the purpose of prediction when measurements are unavailable for total hardness and TDS.

Key words: Box-Cox model, log-linear model, total dissolved solid, total hardness, water quality

Introduction

Nepal is considered fortunate in possessing abundant water resources in the form of surface water as well as groundwater. Three classes of rivers / rivulets constitute the entire surface water systems of Nepal. They are: perennial snow-fed surface flows that originate in the Himalayas; perennial as well as spring-fed flows originating in midland mountain ranges; and hosts of small rivers in the terai plains originating in the southern Siwalik range of hills. The average combined run-off from all three classes of rivers is estimated to be in the range of 220 billions m^3 per year with an average national precipitation of 1530 mm per year and all Nepal rivers drain from north to south and

finally mix with the Ganges in India. The Ganges is known as the largest river of the Indian subcontinent with 2510 km length and the Nepal surface run-offs alone comprise 45% of its annual flow volume and 70% of its flow in dry season (WRS 2002).

There are seventeen river basins in the Nepalese river system covering a total drainage area of about 190,000 km^2 of which 22% lies in China and 5% is in India. Karnali, Sapt Gandaki and Sapta Koshi are the major river basins with their origins in the Himalayas and account for around

80% runoff. The Babai, West Rapti, Bagmati, Kamala and Kankai are medium river basins accounting for about 7% of the runoff. The southern rivers, with origins in the Siwalik Hill Range, are Bering, Balan, Khutiya, Pathraiya, Lal Bakaiya, Ratu, Sirsia, Manusmara and Banganga. These rivers are seasonal with little flows during non-monsoon periods. The Mahakali and Mechi rivers form the western and eastern frontiers with India (ICID 2010).

Regarding groundwater sources, the terai plains of Nepal together with inner terai region are considered as the most promising and productive unconfined aquifers of the sub-continent. It is estimated that the rechargeable groundwater in the terai is anywhere between 5.8 billions m^3 to

11 5 billion m^3 per year with the water table generally lying between 3 to 5 meters below ground level (UN 1995). The groundwater resources have not been fully assessed. Ongoing studies show that a good potential for groundwater extraction exists, especially in the southern lowland plains (terai) and inner valleys of the hilly and mountainous regions. Much of the terai physiographic region and some parts of Siwalik valleys are underlain by deep or shallow aquifers, many of which are suitable for exploitation as sources of irrigation water (Eoearth 2010). The abundance of internal renewable water resources in South Asian countries including Nepal is also reflected in the State of Environment in Asia and Pacific Report, 1995, UN, New York. The data is as follows:

Table 1. Total volume of water per capita per year (in m^3)

Country	Bhutan	Bangladesh	Nepal	Sri Lanka	Pakistan	India
<i>Total water per capita per year</i>	59380	11270	7760	2350	2280	1080

The report also says that if in a country, the renewable fresh water supply per capita per year falls below 1700 m^3 , the country begins to feel water stress meaning occasional or localized periodic water shortages. The above data shows Nepal's position is much above the threshold level of stress (WWF 2010).

Nepal being a predominantly mountainous country with extremely rugged and difficult terrain, distribution of yearly precipitation is very much uneven. Again, Nepal lacks adequate number of large natural lakes to hold sufficient volume of surface flows. Thus, a large part of monsoon rains immediately drains out as surface run-offs. All of these point that adequate supply of safe drinking water has always been a problem to the most particularly, during dry season. With rapidly growing human population and poor water resource management practice, the available fresh water is not only dwindling fast, it is also getting

more and more polluted. Signs of water stress (scarcity syndrome) are already getting more evident by days in fresh water deficient regions; the worsening Katmandu Valley drinking water situation provides a glaring example.

Traditionally, regular water quality monitoring to protect the natural water bodies at sources has always been a very weak point in most water resource management policies applied in Nepal so far. The dominant focus has been more on water development than water management. In this study, samples of water quality test data and related information collected from different sources constituting stream, springs and different types of ground water wells and representing three typical ecological regions of Nepal - mountain; midland hills; and terai and inner terai have been compiled, statistically treated and results studied and presented.

Methodology

In this paper data analysis has been done to fulfill two main objectives. The first one is to reveal the state of water quality conditions of various typical source waters, which are used by the locals for various domestic and other purposes including drinking. In Nepal, about 53.4% of households used piped water, 28.6% used tube well water, 9.1% used well water, 6.5 % used spout water, 1.5% used rivers/streams, and 0.9% used other sources (CBS 2008). The other objective is to prepare suitable statistical models through which some essential water quality test parameter values could be predicted within a reasonably accepted accuracy. Since source waters tapped for rural water supplies generally lie at difficultly accessible locations far from supply area and sample taking, preservation, transportation to laboratories for testing involve high cost and time consuming, only parameters like temperature, electric conductivity, and pH level are measured at sampling sites. Other parameters such as hardness and total dissolved solids are measured only in laboratories. Moreover, instead of measuring these parameters, values of such parameters can be predicted by statistical models which can be a viable alternative as well. Such models based upon local water quality characteristics could prove to be a valuable tool for water quality management as a whole, in future.

Data

The total number of test samples taken from various source types representing the three main ecological regions of Nepal for the present analysis is 270 of which 71 samples were from stream; 93 from spring; and 106 from ground waters. Ecologically 22 (8.1%) samples were from mountain, 131 (48.5%) from hill, and 117 (43.3%) from the terai region. The district-wise coverage is as follows: Districts from mountain were Dolakha (4) and Dhading (18). Similarly, districts from hill were Sindhuli (19), Ramechhap (4), Sindhupalanchwok (19), Kavre (44), Bhaktapur (2), Kathmandu (2), Nuwakot (23), and

Makawanpur (18). Lastly, districts from terai were Sarlahi (9), Rautahat (7), Bara (3), Parsa (22), Chitwan (7), Nawalparasi (6), Rupendehi (9), Kapilbastu (18), Dang (2), Banke (10), and Bardia (24). Altogether, 21 districts were covered. Test parameters studied are: pH, temperature, electric conductivity (EC), total dissolved solid (TDS), total alkalinity (TA), total hardness (TH), Langlier Index (LI), and Ryzner stability index (RSI).

Data source

Water quality data has been taken from some available laboratory test reports of Nepal Environmental and Scientific Services (NESS) Pvt. Ltd., Kathmandu, Nepal. The tests being carried out under the water quality testing projects granted by Rural Water and Supply and Sanitation Fund Development Board (RWSSFDB) of Nepal Government and funded by World Bank, Kathmandu Office. RWSSFDB has been operating in Nepal for more than a decade now. Its main objective is to deliver sustainable health and hygiene benefits to rural population through development and or improvements of safe drinking water supply and sanitation facilities and also strengthen institutional capabilities to undertake and sustain such efforts [RWSSFDB 2009].

Data analysis

Data is analyzed through some frequency distributions, descriptive statistics, and a couple statistical models built for the predictions of total hardness and total dissolved solids. The models are built since spot measurements of hardness and total dissolved solids are most often difficult to conduct. Therefore, statistical models would be useful to estimate these parameters based upon other parameters which are easily measurable such as electric conductivity, temperature, and pH value. Statistical software package, SPSS (statistical package for the social sciences), version 17 was used to analyze the data.

Researches have been carried out to build statistical models and predict water quality parameters like TDS

and hardness with independent parameters such as magnesium, calcium, sodium, etc. using techniques of regression and structural equation modeling (Chenini & Khemiri 2009). Studies have also been conducted for constructing water quality index using different parameters like pH, total hardness, calcium, magnesium, bicarbonate, chloride, nitrate, sulfate, total dissolved solids, iron, manganese and fluorides (Ramakrishnaiah, Sadashivaiah & Ranganna 2009). In the present data analysis, two statistical models are developed to estimate total hardness and TDS with different set of independent variables which are relatively much easier to measure at the sampling site. Before selecting the final models several alternative models including multiple linear and nonlinear models were examined for their suitability for prediction.

Results

Results are presented in the following three sections, namely general, descriptive measures, and statistical models.

General

General descriptions of the parameters such as total hardness, total dissolved solid, total alkalinity, pH values, electric conductivity, and water stability are provided below

Total hardness

Fig. 1 shows spring waters comparatively less hard as compared to stream and well waters. Well waters are found to be among the hardest with 64% of the samples falling in very hard category. In addition, 36 % of spring waters studied are found to be soft, the rest with varying hardness with percentage of very hard water samples having more than 180 ppm hardness being 10 only (Fig. 1) according to water hardness classification made by US Department of Interior and Water Quality Association, US. The classification categorizes water as soft for hardness in the range 0 – 17.1 mg/l, slightly hard soft for hardness in the range 17.1 – 60 mg/l, moderately hard for hardness in the range 60 – 120 mg/l, hard for hardness in the range 120 – 180 mg/l, and very hard for hardness 180 mg/l or above.

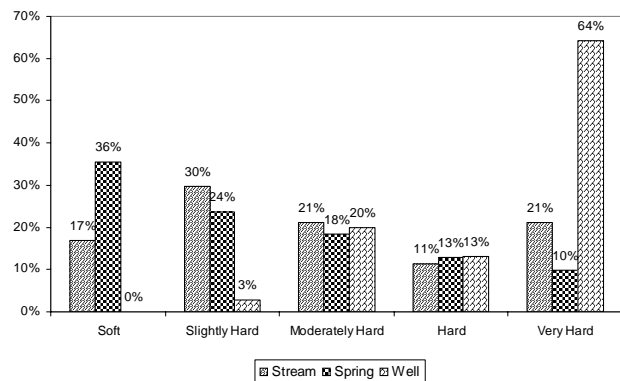


Fig.1. Percentage bar diagram of total hardness of source waters in Nepal

Total Dissolved Solids (TDS)

Fig. 2 shows all the samples collected from stream and spring intake sources had TDS values less than 500 mg/l where as samples taken from well waters showed 16% with values above 500 mg/l. This implies that only well waters have TDS values above WHO limit of 1000 mg/l for drinking in Nepal according to examined samples (Fig. 2).

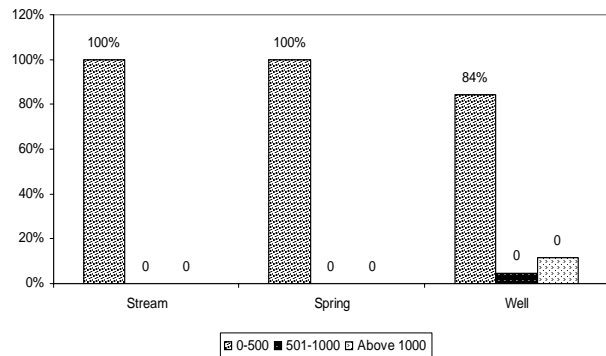


Fig.2. Percentage bar diagram of TDS of source waters in Nepal

pH Value

Fig. 3 shows very high percentage of water samples from all intake sources of Nepal within WHO limit (6.5 – 8.5), about 91%, 82%, and 81% for stream, spring and well water intake sources, respectively. Spring and well waters showed almost the same percentage (81%) of samples within the WHO limit. None of the samples showed pH value above WHO limit.

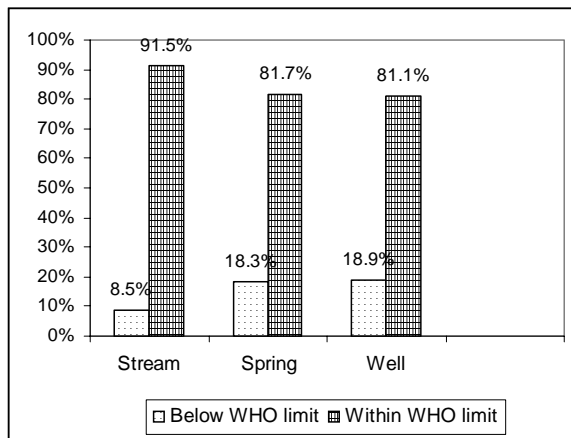


Fig.3. Percentage bar diagram of pH levels of source waters in Nepal

Electric conductivity

As shown in Fig. 4, very high percentage (87% or more) in the samples collected from stream and spring had EC less than or equal to 500 μ mhos/cm where as only 30% in well waters. In well waters more than 50% of the samples had values between 501 and 1000 μ mhos/cm and 2 samples had more than 2000 μ mhos/cm EC. Only 12 samples fell outside Nepal drinking water quality standard regarding EC (1500 μ mhos/cm) all of which were from well waters.

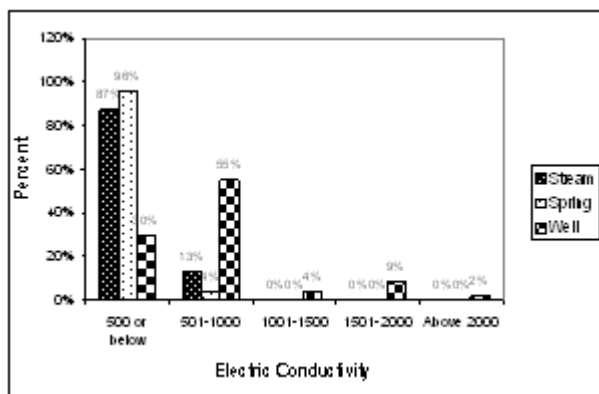


Fig.4. Percentage bar diagram of electric conductivity of source waters in Nepal

Total alkalinity

The distributional pattern of alkalinity measurements of source waters of Nepal shows similarity with that of electric conductivity measurements. Stream and

spring source waters showed more than 83% of the samples with alkalinity 200 or less (mg/l) while well waters had about half (42 %) in the same range. In well waters a substantial percent (46.%) of the samples measured alkalinity between 201 mg/l and 400 mg/l. Very few samples showed alkalinity above 400 mg/l in all the three source waters of Nepal (13) (Fig. 5)

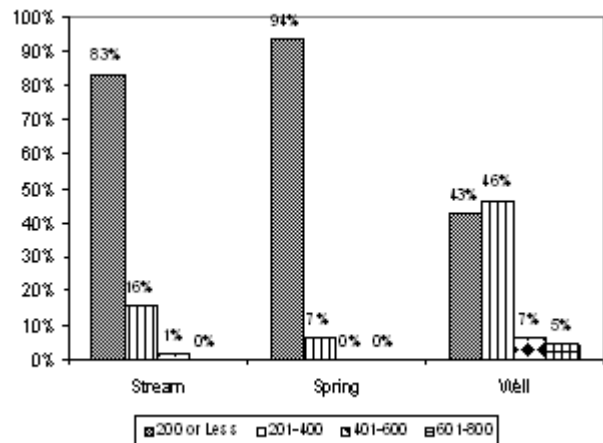


Fig.5. Percentage bar diagram of alkalinity of source waters in Nepal

Water stability

Fig. 6 shows stream and spring source waters more corrosive, 59.2% and 74.4 %, respectively than well waters (26.4%). 59.4% of well water samples were stable while 20% to 30% were stable for stream and spring waters, respectively. The scalable samples were low for all the water sources examined with lowest measured for spring waters (2.2%). In the analysis, water is categorized as corrosive, stable, and scalable if Langelier Saturation Index is less than -0.5, between -0.5 and 0.5, and above 0.5, respectively. Langelier Saturation Index (LSI) is taken as:

$$pH - pH_s$$

where $pH_s = (9.3+TF+TDSF-HF-AF)$ and TF is the temperature factor, TDSF is the TDS factor, HF is the hardness factor, and AF is the alkalinity factor.

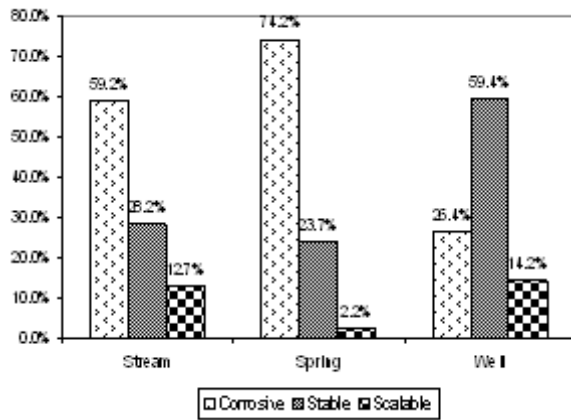


Fig.6. Percentage bar diagram of water stability of source waters in Nepal

Frequency distribution of water samples can also be done using Ryznar Stability Index (RSI). The index is given by:

$$RSI = 2 \text{pH} - \text{pH}_s$$

Water is corrosive, slightly corrosive, stable, and scalable if $RSI < 6$, $6 \leq RSI < 7$, $7 \leq RSI < 8$, and $RSI \geq 8$. According to RSI about 60% of the well water samples are found stable where as in stream and spring waters about 48% and 65% were corrosive, respectively (Fig. 7).

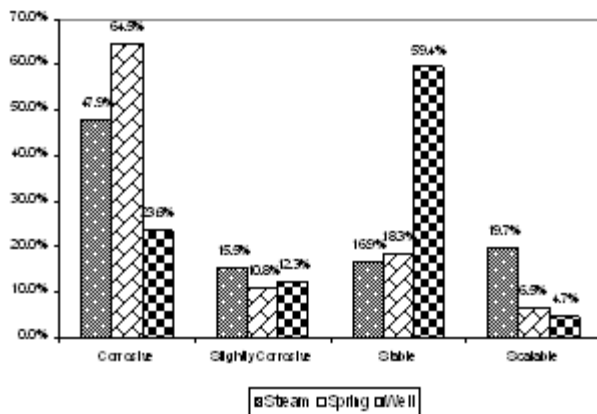


Fig.7. Percentage bar diagram of water stability of source waters in Nepal using Ryznar stability index

Descriptive measures

Table 2 shows some descriptive measures on central tendency by arithmetic mean and variability by standard deviation (SD) and coefficient of variation (CV) of the measured water quality parameters. High variations in the measures are observed regarding coefficient of variations within sources except for pH values which are less than 10%. The mean hardness, alkalinity, TDS, electric conductivity levels, and Langelier index are found to be highest for well waters and lowest for spring waters. In contrast, mean pH value is found to be highest for stream and lowest for spring waters. If Boxplots of Langelier saturation index and Ryznar stability index are viewed, median values are found to be lowest for spring waters and highest for well waters considering both the indexes. The plots also indicate skewed distributions for all the source waters with highest skewness with few outliers observed for well waters.

Statistical models

Two distinct statistical models are developed to estimate total hardness and TDS with different set of independent variables. Before selecting the final models several alternative models were examined for their suitability. The following table (Table 3) shows different alternative models. The initial predictor variables considered for modeling are temperature, electric conductivity, pH value, and categorical contrast variables separating intake water sources.

For estimating total hardness, log-linear model is found to be appropriate whereas for estimating TDS variance stabilizing Box-Cox model is found to be appropriate. To choose the best model different model adequacy tests were used including goodness of fit, normality, residual analysis, multicollinearity, and heteroscedasticity (Montgomery, Peck & Vining 2003).

Total hardness

The log-linear model is found appropriate and used. The model used is

where x_1 = electric conductivity, x_2 = pH value, y is total hardness, and ϵ is the error term.

The model can be reduced to a linear model by taking logarithms as follows.

$$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \varepsilon$$

The parameter estimates of the linear form are given in Table 3. Conductivity and pH values are found to be positively correlated with hardness of water and are highly significant. The fitted model is as follows.

$$y = 0.017284 e^{2.895043(x_1)} e^{3.781043(x_2)}$$

Table 3. Parameter estimates for total hardness

Model	Un-standardized Coefficients		t	Sig.
	B	Std. Error		
(Constant)	-4.058	.472	-8.596	.000
Ln (conductivity)	1.063	.018	59.504	.000
Ln (pH)	1.330	.261	5.097	.000

Model adequacy tests

Goodness of fit of the log-linear model is assessed by adjusted R² which is found to be very high (0.95). Residual analysis shows errors moderately normally distributed (judged by p-p plot). The scatter plot between standardized residuals and standardized predicted values shows the plots randomly distributed and evenly scattered. Variance inflation factor is used to assess the presence of multicollinearity which shows values equal to 1.26 which indicates absence of multicollinearity in the model. There are only two marginal outliers whose values are less than 3.2 though the model itself is built by deleting some 5 significant outliers. All these tests suggest that the model is appropriate.

TDS

Box-Cox model is selected with λ = 0.5 after considering several other values of λ. The dependent variable is transformed using the Box-Cox transformation for variance stabilization (Greene 2003).

The new dependent variable is:

The model can be used to estimate total hardness of water for some specified values of electric conductivity and pH values. Variables found insignificant are temperature and categorical contrast variables separating the intake water sources.

The initial predictor variables considered for modeling are temperature, electric conductivity, pH value, categorical contrast variables separating the intake water sources.

$$\frac{y^\lambda - 1}{\lambda} \text{ when } \lambda \neq 0$$

$$\ln(y) \text{ when } \lambda = 0$$

The selected model is:

$$\frac{y^\lambda - 1}{\lambda} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

where x₁ = electric conductivity, x₂ = temperature, x₃ = pH value, y is TDS, and ε is the error term.

The fitted model parameter estimates are shown in Table 4 which shows that the parameter estimates are statistically significant (<0.05) and all the independent variables are positively correlated with TDS. The model can be used to estimate TDS of water for some specified values of electric conductivity, temperature (°C), and pH values. The variables examined but excluded from the model are the categorical contrast variables separating the intake water sources. Clearly in both of the developed models the predictions of total hardness and TDS were not significantly affected by the type of intake sources.

Table 4. Parameter estimates for TDS

Model Parameters	Un-standardized Coefficients		t	Sig.
	B	Std. Error		
Constant	-14.529	3.078	-4.721	.000
Electric Conductivity	.030	.001	46.566	.000
Temperature	.135	.061	2.233	.026
pH Value	3.391	.389	8.711	.000

Model adequacy tests

Goodness of fit of the log-linear model is assessed by adjusted R^2 which is found to be very high (0.94). Residual analysis shows errors moderately normally distributed (judged by p-p plot). The scatter plot between standardized residuals and standardized predicted values shows the plots randomly distributed and evenly scattered. Variance inflation factor is used to assess the presence of multicollinearity which shows values less than 1.8 which indicates absence of multicollinearity in the model. There are only four marginal outliers whose values are less than 3.8 though the model itself is built by deleting some 3 significant outliers. All these tests suggest that the model is appropriate.

Discussion

Data analysis of some 270 water samples of Nepal taken from different intake sources was carried out for their characterization and statistical model building. Two statistical models were developed for the purpose of estimation of total hardness and TDS. The models are suitable for their predictions in the absence of measurements. Hardness predicting model is the log-linear model whereas TDS predicting model is the Box-Cox variance stabilizing model. The predictor variables found statistically significant are EC, pH value, and temperature (in TDS model only) which are easy for spot measurements. All these variables are found to be positively associated with hardness and TDS.

Acknowledgements

The authors would like to express their gratitude and thanks to Kul Bahadur Shahi, Chemist, former staff of

NESS (P) Ltd., Kathmandu, Mrs. Anita Shakya, Chemist, former staff of NESS (P) Ltd., Kathmandu, and Mr. Sunil Babu Khatri, Chemist, Staff of NESS (P) Ltd., Kathmandu, and other concerned NESS employees for their involvement in laboratory tests of water samples. Deep appreciation goes to all those concerned in RWSSFDB, Kathmandu and World Bank, Kathmandu office for providing fund to generate data.

References

- CBS 2008. *Environment statistics of Nepal*. CBS Publication, Kathmandu, Nepal.
- Chenini, I. and S. Khemiri. 2009. Evaluation of ground water quality using multiple linear regression and structural equation modeling. *International Journal of Environmental Science and Technology* 6:509-519.
- Eoearth 2010. [Electronic resource]. http://www.eoearth.org/article/Water_profile_of_Nepal
- Greene, W. H. 2003. *Econometric analysis*. Pearson Education Inc, Singapore.
- ICID. 2010. [Electronic resource] http://www.icid.org/v_nepal.pdf
- Montgomery, D. C., E. A. Perck, and G. G. Vining. 2003. *Introduction to Linear Regression Analysis*. John Wiley and Sons, Singapore.
- Ramakrishnaiah, R., C. Sadashivaiah and G. Ranganna. 2009. Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. *E-Journal of Chemistry* 6:523-530.
- RWSSFDB 2009. Annual report on rural water supply and sanitation project, Rural Water Supply and Sanitation Fund Development Board, Kathmandu, Nepal.
- UN. 1995. *State of environment in Asia and Pacific report*. UN, New York.
- WRS. 2002. National water plan 2004. Water resource strategy. Water and Energy Commission Secretariat, Nepal.
- WWF. 2010. [Electronic resource]. <http://www.water-vision.org>