academicJournals

Vol. 11(11), pp. 565-577, November 2017
DOI: 10.5897/AJEST2017.2353
Article Number: 56A14E166520
ISSN 1996-0786

http://www.academicjournals.org/AJEST

African Journal of Environmental Science and Technology

Full Length Research Paper

Spatial and temporal water quality dynamics of Awash River using multivariate statistical techniques

Amare Shiberu Keraga[1*], Zebene Kiflie[1] and Agizew Nigussie Engida[2]

[1]School of Chemical and Bio Engineering, Addis Ababa Institute Technology, Addis Ababa University, P. O. Box 385, Addis Ababa, Ethiopia.
[2]School of Civil and Environmental Engineering, Addis Ababa Institute Technology, Addis Ababa University, P. O. Box 385, Addis Ababa, Ethiopia.

Received 13 May, 2017; Accepted 26 August, 2017

Awash River has important socio-economic and ecological values in Ethiopia. On the contrary, it is prone to serious water pollution. This study aims to assess the spatial and temporal variation of water quality of the river. Means of the 9 years' (2005-2013) water quality dataset of 19 parameters from 10 stations in the basin were considered. After validating, normalizing and checking the sampling adequacy and internal consistency of the data, principal component analysis was computed and four principal components were generated. Factor loadings, correlations between variables and the principal factors as well as between sites and the principal factors were tabulated. Agglomerative hierarchical clustering done on the dataset resulted in four clusters based on similarity of water quality characteristics. The Mann-Kendall's two tailed trend test detected temporal trends for total hardness in February over all sites and for most parameters in the basin in the 9 years period. Spatial analysis of the 14 sampling sites of the basin showed that as one moves from upper to lower parts of the basin, electrical conductivity, total hardness and chloride decrease in the dry season. However, total hardness slightly increases and total dissolved solids, chloride, and sulfate decrease in the rainy season.

Key words: Agglomerative hierarchical clustering, Ethiopia, Mann Kendall trend, principal component analysis, water pollution.

INTRODUCTION

Surface water quality degradation and its spatial and temporal variation in developing countries like Ethiopia are becoming a threat to ecosystem services due to the rapid increase in population, climate change, industrialization, and the associated land use dynamics in the countries (Kithiia, 2012; Abbaspour, 2011; Davies and Simonovic, 2011). The variation is governed both by natural and anthropogenic factors including climate, types of soils and soil erosion, rocks, hydrology and surfaces through which it moves, agricultural land use, and sewage

*Corresponding author. E-mail: amare.shiberu@aau.edu.et. Tel: +251911048516. Fax: +25111 12 39 480.

discharge (Bu et al., 2010; Pejman et al., 2009).

Management gap of water quality is observed in Ethiopian River basins as witnessed by poor water quality which does not fit the intended water use requirements (Romilly and Gebremichael, 2011). For instance, the water quality of many rift valley lakes is very poor in terms of salinity and alkalinity and hence not suitable for irrigation, domestic or industrial purposes (HGL and GIRDC, 2009). Investigation of comprehensive and detailed water quality dynamics at a river basin scale has not been done in most parts of Ethiopia. There is also a spatial difference in water quality between the lakes (Tiruneh, 2005; HGL and GIRDC, 2009). In the characterization of the spatial and temporal variability of water quality parameters and sediment distribution in Lake Abaya, Gebremariam (2007) came up with a result that water temperature, pH, conductivity, dissolved oxygen (DO), total suspended solids (TSS), and total dissolved solids (TDS) at fixed stations in the lake varied respectively from 21.9 to 30°C; 8.8 to 9.3, 861 to 1162 $\mu Scm^{-}1$, 5.4 to 7.9 $mg.L^{-1}$, 4 to 404 $mg.L^{-1}$, and 618 to 1206 $mg.L^{-1}$.

Awash River is the most developed water system and hence it is the major component of Ethiopian economy in terms of serving as a water supply source for households, hydropower, industries, small to large-scale irrigation schemes of sugarcane, cotton, fruit, vegetable, and flower farms. Nowadays, there are, however, many factors threatening these advantages of the river, namely, population growth, expansion of irrigated area and deforestation in the upper basin (Taddesse et al., 2004). Moreover, the river receives untreated and uncontrolled domestic, industrial and agricultural wastes from the catchment directly along its course (Belay, 2009; Alemayehu, 2001; Awash Basin Authority [AwBA], 2014). The effect of surface water-groundwater interactions seen in the mixing of Lake Beseka with Awash River could result in quality change (Belay, 2009). Therefore, among the major Ethiopian rivers, these factors have made Awash River most prone to various types of serious pollution.

Consequently, effects of direct waterborne diseases to animals and human-beings and high concentration of heavy metals on vegetables and coliforms were observed. Little Akaki River, which is a tributary of Awash River is unfit for any intended use as it is loaded by high microbiological and chemical pollution. Water pollution in the basin is found to have contributed to the disappearance of aquatic species (Gebre and Rooijen, 2009). The irrigation water heavy metal content is shown to exceed the standard for irrigation water. The incidence of dental and skeletal fluorosis from high concentration of fluoride is well documented in its valleys (Reimann et al., 2003). Though high fluoride concentration is especially apparent in the Rift Valley Lakes basin, the problem is observed to have a negative impact on public health in the Awash Basin too. Few investigations also showed that nitrate levels are above 10 mg/L in the surface water, and according to Taddese et al. (2003), mean concentrations of heavy metals including manganese, chromium, nickel, lead, arsenic and zinc in Addis Ababa catchments are measurably higher in the soils irrigated by Akaki River (Taddese et al., 2003).

Studies in relation to the spatial and temporal variation in the basin have not been reported so far. However, investigating the spatial and temporal dynamics of Awash River water quality has a scientific and practical significance in that it fills the existing knowledge gap. Therefore, the objective of this study is to investigate the temporal and spatial water quality variation and detect trends of water quality of Awash River.

MATERIALS AND METHODS

Description of the study area

The basin extends from 7°53′N to 12°N and 37°57′E to 43°25′E in Ethiopia as shown in Figure 1. It covers a total land area of 113,304 km^2 of which 64,000 km^2 is in western section of the basin. This section of the basin drains to the Awash River or its tributaries. The remaining 49,304 km^2, most of which comprises the so-called Eastern Catchment drains into a desert area and does not contribute to the Awash River. The Awash River has a total length and an annual flow of 1250 km and 4.6 billion m^3 (BCM), respectively. The river originates at an elevation of about 3000 m in the central Ethiopian highlands near Ginchi town about 80 km west of Addis Ababa (Degefu et al., 2013; Tessema, 2011; Berhe et al., 2013).

The main physiographic units of Awash River basin are highlands, the main Ethiopian rift and Afar triangular depression, in which grassland, shrub-land, woodland and forests are the main units. The basin is endowed with several wetlands of various types as well as artificial and natural lakes. It is characterized by wide-ranging agro-climatic zones with varied ecological conditions. With extreme ranges of topography, vegetation, rainfall, temperature and soils, the basin extends from cold high mountain zones to both semi-desert lowlands (Gedion, 2009). Settled rain-fed agriculture is practiced with possible double cropping in areas receiving considerable spring rainfall (January to May) in addition to main summer rainfall.

Mean annual temperature ranges from 16.7 to 29°C and the mean monthly temperatures range from 9.6°C in the capital to 37°C around Lake Abe Area. Mean annual relative humidity in the basin varies from 60.2 to 49.7%. The mean annual wind speed is 0.9 m/s. The mean annual rainfall varies from about 1600 mm at Ankober to 160 mm at Asayita, according to Berhe et al. (2013). There are two major soil types in the catchment; the deep red clay soil, Nitosol, and the dark clay soil, Vertisol. The Nitosol is found in the upland areas, whereas the Vertisol is found in lowland areas with slopes ranging from 2 to 8% (Moreda, 1999).

The basin has an estimated population of 14.8 million. Majority of this population are engaged in agriculture and animal husbandry. From 48 to 70% of the existing large-scale irrigated agriculture and more than 65% of the national industries are located in the basin (Tessema, 2011; AwBA, 2014). The relative surface water resource of the basin is about 4.65 BCM. It is the most developed and utilized basin since 77.4% of the irrigable land in the basin has been cultivated. Wide varieties of crops are cultivated ranging from

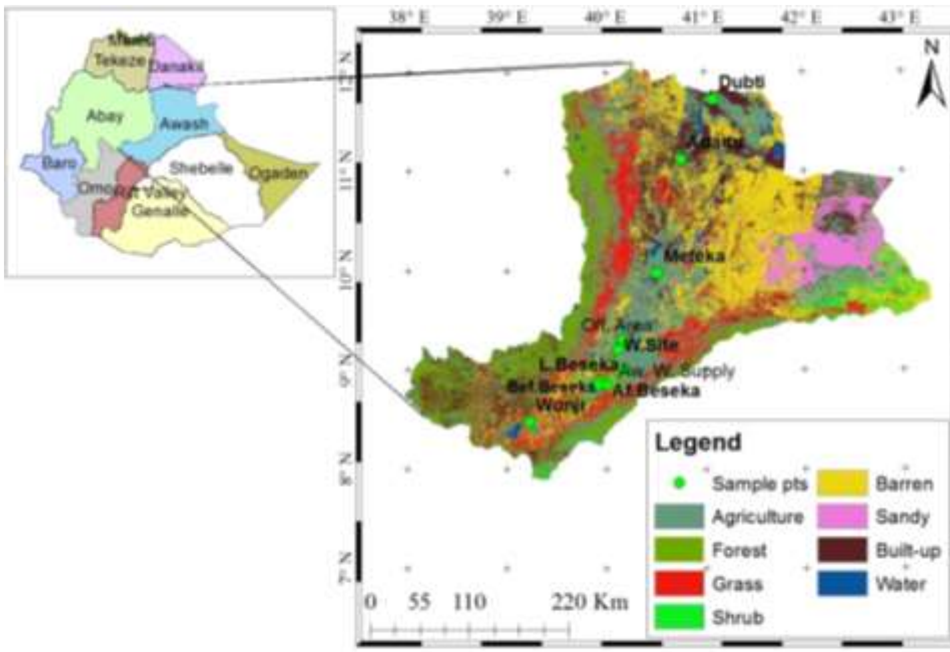

Figure 1. Location map of Awash River Basin with the sampling sites (Based on 2014 LU/LC data).

cereals, vegetables, flowers, cotton to perennial fruit orchards and sugarcane (AwBA, 2014).

Data collection and analysis

Dataset of 10 water quality monitoring stations, containing 19 parameters monitored monthly over 9 years (2005-2013) was collected from AwBA. The water quality parameters considered in this analysis, their abbreviations, their units and methods of analysis (following standard methods for the examination of water and wastewater) are summarized in Table 1 (APHA et al., 1998). The mean of the monthly measured 9-year data-set on the river water quality is summarized in Table 2.

Multivariate statistical techniques and data treatment

Multivariate statistical analysis approaches were used to analyze the water quality data. Such approaches are useful to deal with complex environmental dataset exposed to varying natural and anthropogenic factors and solve the associated problems without misinterpretation. They provide a means of handling large dataset with large number of variables by summarizing the redundancy, as well as reflecting, detecting and quantifying the multivariate nature of ecological data accurately (Hulya and Hayal, 2008; Wang et al., 2007). Principal component analysis (PCA) and cluster analysis (CA) are specifically useful for considering several related random environmental variables simultaneously, and thus for identifying a new, small set of uncorrelated variables that account for a large proportion of the total variance in the original variables (Wang et al., 2007). They were employed here to sort out the variables of water quality parameters and sampling stations.

Cluster analysis (CA)

CA is a multivariate procedure which classifies data based on placing of objects into more or less homogeneous groups. The main idea behind clustering is to combine identical sites as one cluster and group two clusters of the highest similarity as a new cluster, which in turn is combined with another most similar cluster as another new cluster and so on until all clusters become one cluster (Xu et al., 2012). Agglomerative hierarchical clustering (AHC) is the most common and iterative classification method in which clusters are formed sequentially by starting with the most similar pair of objects and forming higher clusters step by step (Singh et al., 2004). AHC was done by Euclidean distance proximity type and Ward's method of agglomeration on the dataset to assess the similarity among sampling sites. The Euclidean distance, which is used to quantify the similarities or differences between the two sites (sampling locations) i and j, was calculated using the formula below:

$$d_{ij}^2 = \sum_{k=1}^{m} (Z_{i,k} - Z_{j,k})^2 \qquad (1)$$

where d_{ij} is the Euclidean distance, $Z_{i,k}$ and $Z_{j,k}$ are variable k for objects i and j respectively, and m is the number of variables (Gibrilla et al., 2011; Singh et al., 2004).

Principal component analysis (PCA)

PCA is a very effective tool used to reduce the dimension of a data set consisting of a large number of inter-related variables by reducing the contribution of variables with minor significance, while retaining as much variability of the data set as possible and

Table 1. Water quality parameters, their units and methods of analysis.

Parameter	Abbreviation	Unit	Method
Turbidity	Turb	NTU	Turbid metric method
Total solids 105C	TS	mg/L	Gravimetric method
Total dissolved solid	TDS	mg/L	Gravimetric method
Electrical conductivity	EC	µS/cm	Calorimetric method
pH	pH	-	Calorimetric method
Ammonia	NH_3	mg/l NH_3	Aluminon method
Sodium	Na^+	mg/l Na^+	Flame photometer
Potassium	K^+	mg/l K^+	Flame photometer
Total hardness	TH	mg/l $CaCO_3$	Titrimetric method
Calcium	Ca^{2+}	mg/l Ca^{+2}	Titrimetric method
Magnesium	Mg^{2+}	mg/l Mg^{+2}	Periodate oxidation method
Total Iron	TFe	mg/l Fe^{+3}	Phenanthroline method
Fluoride	F^-	mg/l F^-	SPADNS method
Chloride	Cl^-	mg/l Cl^-	Argentometric method
Nitrate	NO_3^-	mg/l NO_3^{-2}	Cadmium reduction method
Alkalinity	Alkal	mg/l $CaCO_3$	Titrimetric method
Bicarbonate	HCO_3^-	mg/l HCO_3^-	Titrimetric method
Sulphate	SO_4^{2-}	mg/l SO_4^{-2}	Turbid metric method
Phosphate	PO_4^{3-}	mg/l PO_4^{3-}	Ascorbic acid, molybdate blue method

interpreting the total variability of the dataset. This is accomplished by transforming the data set into a small number of new set of variables called Principal Components (PCs). The PCs are orthogonal (non-correlated), linear combinations of the originally observed water quality data and are arranged in decreasing order of importance (Shrestha and Kazama, 2007; Singh et al., 2004). The PCs can be expressed as:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \cdots + a_{im}x_{mj} \quad (2)$$

where Z is the component score, a is the component loading, x is the measured value of variable, i is the component number, j is the sample number and m is the total number of variables (Muangthong, 2015; Shrestha and Kazama, 2007).

Mann Kendall trend test

Mann-Kendall (MK) trend test is a rank-based and non-parametric statistical test used to detect trends and assess the significance of trends in hydro-meteorological time series data such as water quality, streamflow, temperature, and precipitation. This test is also documented to be more suitable and powerful for detecting trends of non-normally distributed and censored data (Yue et al., 2002; Xu et al., 2012) such as that of Awash River. It can be used in place of the parametric linear regression analysis, which requires that the residuals from the fitted regression line be normally distributed. Basically, it separately calculates the test statistics and variance of water quality data in each season after which overall statistics is calculated (Xu et al., 2012).

Assessment and validation of data errors and anomalies

Assessment of integrity and validity of a given dataset based on knowledge, experience and intuition is an important and initial step of any water quality data analysis to draw meaningful conclusions from a study (Rangeti et al., 2015). It could be realized that the water quality dataset of Awash River basin has lots of errors and anomalies that need to be validated. The anomalies observed in the dataset were outliers, missing values and censored data. Among the observational (box-plots, time series, histogram, ranked data plots and normal probability plots) and statistical (Grubbs, Dixon, Cochran's C test and Mendel's h and k statistics) techniques available to test outliers, Dixon test was preferred since the number of values to be tested was greater than 6 and less than 25 and since the test was intended to be undertaken on non-normal raw data (Rangeti et al., 2015). It is based on the ratio of the distance between the potential outlier value and its nearest value (Q_{gap}) to the range of the whole data set (Q_{range}), as shown in Equation 4.

$$Q_{exp} = \frac{Q_{gap}}{Q_{range}} \quad (3)$$

If Q_{gap} is large enough as compared to Q_{range}, then the value is considered as an outlier. Dixon test was run excluding first only lake Beseka and then both lake Beseka and after Beseka sites on the same dataset. This resulted in outliers of parameters including alkalinity, HCO_3^-, and SO_4^{2-} at Meteka; TDS at office area and PO_4^{3-} at Awash water supply since the two-tailed p-value (< 0.0001) was less than the significance level α=0.05. Running the test on each of the outlying parameters for the corresponding sites twice, the specific year and month, the effect was identified and appropraite correction was made. Therefore, some parameters values of specific months were not used to calculate the mean of the 9-years.

The outliers have come from technical and personal errors as the wrong calculation of average of the months' values, the wrong recording of the values like imperfect data entry while recording and transferring data, incorrect measurements due to equipment error,

Table 2. Mean values of water quality parameters for the ten sampling sites of Awash River basin during 2005-2013.

Par\site	Stat.	Dupti	Adaitu	Meteka	Off. Area	Weir Site	AwWSupp	Af. Bes	L. Bes	Bef. Bes	Wonji
Turb	mean	1003.89	1435.71	674.36	1537.61	871.43	1650.40	1199.75	40.80	795.05	233.09
	SD	1283.52	538.44	477.99	951.35	561.38	1257.08	715.72	20.44	464.34	69.02
TS	mean	1914.78	3551.92	2369.96	3116.61	2255.12	3438.40	2453.73	4273.50	2040.39	490.02
	SD	1641.66	1449.15	1338.07	1806.37	1122.53	1328.32	1420.73	613.61	1462.16	121.72
EC	mean	603.48	656.16	880.93	647.80	436.28	473.26	1164.12	5825.31	432.66	322.10
	SD	57.32	130.39	231.62	214.52	103.69	192.25	1062.65	606.16	202.72	35.42
pH	mean	7.98	8.32	8.23	7.88	7.90	7.90	8.03	9.42	7.72	8.25
	SD	0.17	0.61	0.61	0.19	0.23	0.27	0.46	0.17	0.38	0.87
NH_3	mean	0.51	0.88	0.48	0.76	0.67	0.75	0.94	0.66	0.73	0.65
	SD	0.23	0.82	0.17	0.41	0.24	0.26	0.59	0.29	0.24	0.33
Na^+	mean	90.72	101.66	154.75	108.34	65.14	75.24	279.27	1474.85	63.73	36.09
	SD	13.36	25.51	51.76	53.49	28.98	52.90	311.44	190.30	58.71	8.53
K^+	mean	6.06	7.35	11.11	9.70	9.56	9.04	11.77	55.55	10.31	6.93
	SD	0.67	1.37	2.96	3.74	4.20	2.78	4.07	8.39	4.27	0.70
Ca^{2+}	mean	35.83	32.65	31.87	33.42	31.19	29.06	26.71	6.25	31.22	28.55
	SD	6.69	4.07	1.89	4.74	1.82	4.02	6.37	1.82	1.60	1.12
Mg^{2+}	mean	8.35	7.08	9.68	7.67	6.51	6.56	6.77	2.25	6.25	5.37
	SD	2.64	1.03	2.32	4.24	2.07	3.61	4.57	1.62	2.86	1.33
TFe	mean	0.34	0.20	0.14	0.37	0.23	0.25	0.27	0.15	0.27	0.25
	SD	0.76	0.16	0.06	0.34	0.16	0.20	0.16	0.08	0.21	0.17
F^-	mean	1.39	1.45	2.12	2.07	2.01	1.43	5.65	25.43	1.67	1.49
	SD	0.39	0.49	0.31	0.27	1.08	0.49	6.27	9.64	0.60	0.38
Cl^-	mean	42.37	47.73	61.51	44.25	25.91	27.64	91.11	492.06	24.16	15.57
	SD	8.64	12.16	19.24	20.19	10.50	16.65	93.80	68.65	17.45	2.79
NO_3^-	mean	4.96	3.01	2.09	2.88	3.56	4.27	3.51	2.57	2.90	3.52
	SD	7.33	1.47	0.86	1.08	1.16	2.76	0.77	2.35	1.02	1.61
Alkal	mean	196.88	224.93	320.05	242.95	182.22	187.08	439.54	2253.48	183.53	135.47
	SD	25.70	35.84	79.76	72.33	46.90	75.58	372.52	396.43	86.70	18.14
SO_4^-	mean	55.21	49.65	57.04	41.79	28.50	32.27	94.78	497.84	21.99	12.72
	SD	13.61	13.10	24.05	22.73	16.15	21.17	101.47	116.34	16.97	6.34
PO_4^-	mean	0.45	0.80	0.62	0.79	0.48	0.54	0.86	2.68	0.46	0.50
	SD	0.18	0.50	0.22	1.11	0.13	0.16	0.51	0.37	0.17	0.28

SD: Standard deviation; Stat.: statistic; Off. Area: office area; AwWSupp: Awash water supply; Af. Bes: after Beseka; L. Bes: lake Beseka, Bef. Bes: before Beseka. Concentration units are in mg/L except turbidity (NTU), EC (μS/cm), and pH which is dimensionless.

Table 3. Eigenvalues.

	F1	F2	F3	F4	F5	F6	F7	F8	F9
Eigenvalue	9.02	2.61	1.62	1.05	0.78	0.42	0.34	0.12	0.04
Variability (%)	56.39	16.32	10.10	6.57	4.87	2.62	2.13	0.75	0.26
Cumulative %	56.39	72.71	82.81	89.38	94.24	96.86	98.99	99.74	100.00

loss of sample before analysis or else. In the raw water quality dataset, there were a number of parameters of missing values and some censored data such as 'nil' and 'trace'. However, they were ignored while calculating the mean. Finally, all these errors were corrected with an informed decision and hence the 9-year mean could be changed.

Testing normality of the data, except turbidity, TS, NH_3, TFe, pH and Mg, all other parameters were not normally distributed since the significant values of both Kolmogorov-Smirnov and Shapiro-Wilk tests were relatively large (>0.05) for the normal ones but close to zero for the rest. However, transformation into normality, according to Singh et al. (2004), increases the influence of variables whose variance is small, reduces the influence of those whose variance is large and also eliminates the influence of different units of measurement thereby making the data dimensionless. Therefore, the variables were transformed to normal by Box-Cox (Osborne, 2010) variable transformation using SPSS and XLSTAT. After transforming, their normality was checked and proved to conform to normality.

Before analysis, suitability of the data for PCA was checked in terms of sampling adequacy and internal consistency of the data using Kaiser-Meyer-Olkin (KMO) and Cronbach's alpha, respectively. They are measures of testing consistency of variable values by offering information on whether or not data could be modelled with PCA. If the KMO value is greater than 0.50, it can be said that a dataset can be factorized (Sen et al., 2016; Ghosh and Jintanapakanont, 2004). Accordingly, for this study KMO was found to be 0.563 (>0.5). Additionally, it resulted in an overall Cronbach's alpha test score of 0.78 verifying internal consistency of the data. Both values imply that the sample was adequate and the desired multivariate statistical analyses (PCA) could suitably be applied on the dataset.

RESULTS

Principal component analysis (PCA)

Removing the three redundant variables TDS, TH and HCO_3^- from the list, which were highly correlated with others, PCA on the normalized 16 variables was then computed to produce significant PCs and to further reduce the contribution of variables with minor significance. It was done using Pearson type of correlation automatically without fixing the number of components/factors to be generated because of the width of the study area.

The 16 variables were reduced by PCA to 9 factors of which the first four showed Eigen values greater than unity. These were retained as the principal (significant) components (Shrestha and Kazama, 2007), because they could explain about 89.4% of the total variation shown in Table 3. The table shows the sorted Eigen values (from large to small) and percentage of variability versus principal components.

A biplot of the variables and the sampling sites is shown by factors F1 against F2 in Figure 2. Both the plot and Table 4 indicate that F1 had a high and positive loading in TS, EC, pH, Na, K, F^-, Cl^-, SO_4^{2-} and PO_4^{3-}, which respectively were 0.69, 0.88, 0.79, 0.92, 0.85, 0.85, 0.89, 0.87 and 0.71. These large and positive loadings show strong linear correlation between the factors and parameters.

Cluster analysis

All the 10 sites of the basin are grouped by AHC into four statistically significant clusters at $(D_{link}/D_{max})\times100<20$. Results of application of the cluster analysis are best visualized by a dendrogram or binary tree as shown in Figure 3.

The dendrogram in Figure 3 clearly depicts grouping of the sites based on similarity of water quality characteristics. Accordingly, before Beseka, Weir site, Wonji and Dubti were grouped as cluster 1, while Adaitu, Office area, Awash water supply and after Beseka were grouped as cluster 2. However, Meteka and Lake Beseka are categorized as clusters 3 and 4, respectively.

Temporal trend analysis

Trend analysis of TDS, EC, pH, NH_3, Na, K, TH, F^-, Cl^-, and NO_3^- was performed in the 9 years period (2005 to 2013) by MK two tailed trend test for the 4 sites in the basin including Dubti, Office area, after Beseka and Wonji.

The analysis at 5% significant level in the dry season (December-February) of Dubti showed that except TH and F^-, none of the parameters were found to show any trend since their p values computed by exact method are greater than the significance level α=0.05. The test for the site revealed that TH had significant increasing and F^- decreasing trend since their respective computed p-values, 0.011 and 0.03, were less than the significant level α=0.05 (Figure 4a). Similarly, the analysis of the office area indicated that only TDS showed a trend, which is monotonic upward throughout the years of

Table 4. Factor loadings, correlations and % contribution of the variables to the PCs (a) and % contribution of observations to the PCs (b).

a) Loadings, correlations and % contribution of the variables to pcs

		F1	F2	F3	F4
Turb	Load&Corr	-0.29	**0.88**	0.21	-0.21
	% Contrib.	0.92	**29.71**	2.69	4.16
TS	Load&Corr	**0.73**	0.41	0.15	-0.05
	% Contrib.	5.84	6.33	1.35	0.20
EC	Load&Corr	**0.92**	0.29	-0.19	0.13
	% Contrib.	**9.45**	3.21	2.32	1.63
pH	Load&Corr	**0.79**	-0.42	-0.01	0.26
	% Contrib.	**6.85**	6.83	0.00	6.30
NH_3	Load&Corr	0.15	0.40	**0.79**	-0.31
	% Contrib.	0.24	6.20	**38.23**	9.41
Na^+	Load&Corr	**0.93**	0.31	-0.14	0.11
	% Contrib.	**9.53**	3.70	1.27	1.08
K^+	Load&Corr	**0.84**	-0.18	0.11	-0.29
	% Contrib.	7.78	1.28	0.76	**7.95**
Ca^{2+}	Load&Corr	**-0.72**	**0.50**	-0.42	-0.01
	% Contrib.	5.69	**9.67**	**10.89**	0.01
Mg^{2+}	Load&Corr	-0.48	**0.54**	-0.61	-0.25
	% Contrib.	2.52	**11.12**	**22.93**	6.14
TFe	Load&Corr	-0.52	0.48	0.25	0.34
	% Contrib.	3.03	8.86	3.98	**11.00**
F^-	Load&Corr	**0.85**	-0.09	0.08	-0.06
	% Contrib.	**8.08**	0.29	0.41	0.35
Cl^-	Load&Corr	**0.91**	0.32	-0.19	0.16
	% Contrib.	**9.21**	3.84	2.24	2.36
NO_3^-	Load&Corr	-0.54	0.25	0.36	**0.66**
	% Contrib.	3.25	2.37	8.07	**40.97**
Alkal	Load&Corr	**0.97**	0.19	-0.12	0.05
	% Contrib.	**10.32**	1.33	0.95	0.22
SO_4^{2-}	Load&Corr	**0.89**	0.34	-0.16	0.26
	% Contrib.	**8.69**	4.32	1.66	6.40
PO_4^{3-}	Load&Corr	**0.88**	0.16	0.19	-0.14
	% Contrib.	**8.61**	0.95	2.25	1.83

b) % Contribution of the observations to PCs

	F1	F2	F3	F4
Dupti	7.73	6.31	13.74	**59.21**
Adaitu	0.05	6.91	1.12	5.24
Meteka	1.95	0.81	**62.16**	15.61
Off.Area	0.02	16.16	0.48	2.68
W.Site	2.50	2.03	0.15	0.37
Aw.W.Supp.	2.08	3.01	8.70	0.01
Af.Bes.	6.51	8.88	8.61	0.12
L.Bes.	**59.52**	12.78	0.83	9.78
Bef.Bes.	3.93	2.61	0.54	6.72
Wonji	15.71	**40.51**	3.68	0.27

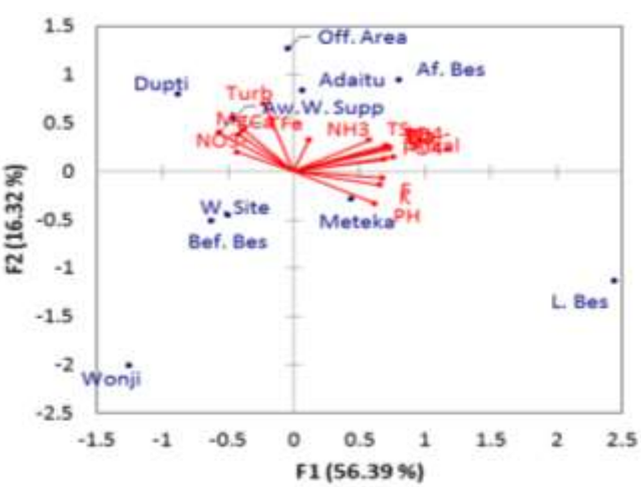

Figure 2. Biplot of sample sites and water quality variables (axes F1 and F2: 72.71%) obtained from principal component analysis.

consideration, and EC, NH_3, Cl^-, Na and K also showed increasing trends only after 2009 (Figure 4b), but the rest were all found to show no trend at all. At Wonji in the dry season, except TH, none of the parameters indicated a trend from 2006 to 2013; TH has shown a significant increasing trend in the period (Figure 4d). The test is undertaken in the wet season (June-August) of the parameters and most of them were found to show no trend. However, TH and K, respectively at Dubti and Wonji showed a decreasing and increasing trends as depicted in Figure 4c.

Temporal trend analyses of the annual average values of turbidity, TS, TDS, EC, pH, NH_3, Na, K, TH, Ca, Mg, TFe, F, Cl, NO_3^-, alkalinity, HCO_3^-, SO_4^{2-} and PO_4^{3-} were undertaken throughout the nine years period by MK two tailed trend test for Dubti, Adaitu, office area, Weir site, after Beseka, Beseka, before Beseka, and Wonji. The analysis at 5% significant level for Dubti indicated that only SO_4^{2-} and F^- have shown increasing and decreasing

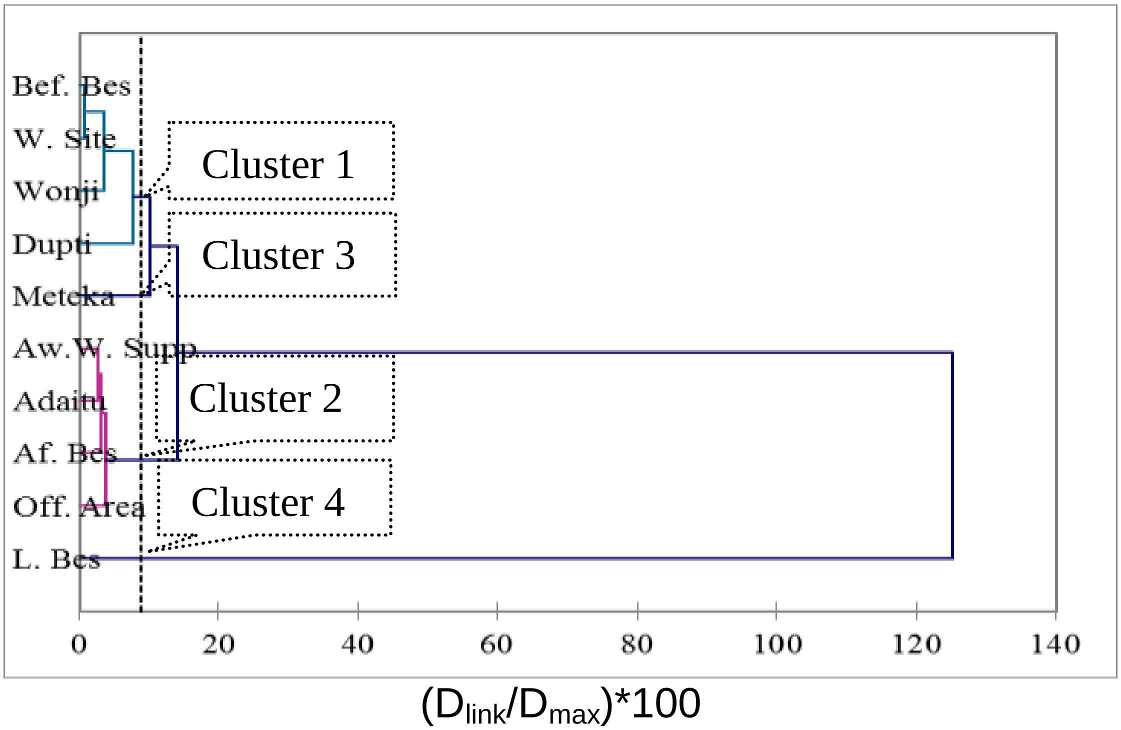

Figure 3. Dendrogram showing cluster analysis of sampling sites based on water quality characteristics of Awash River.

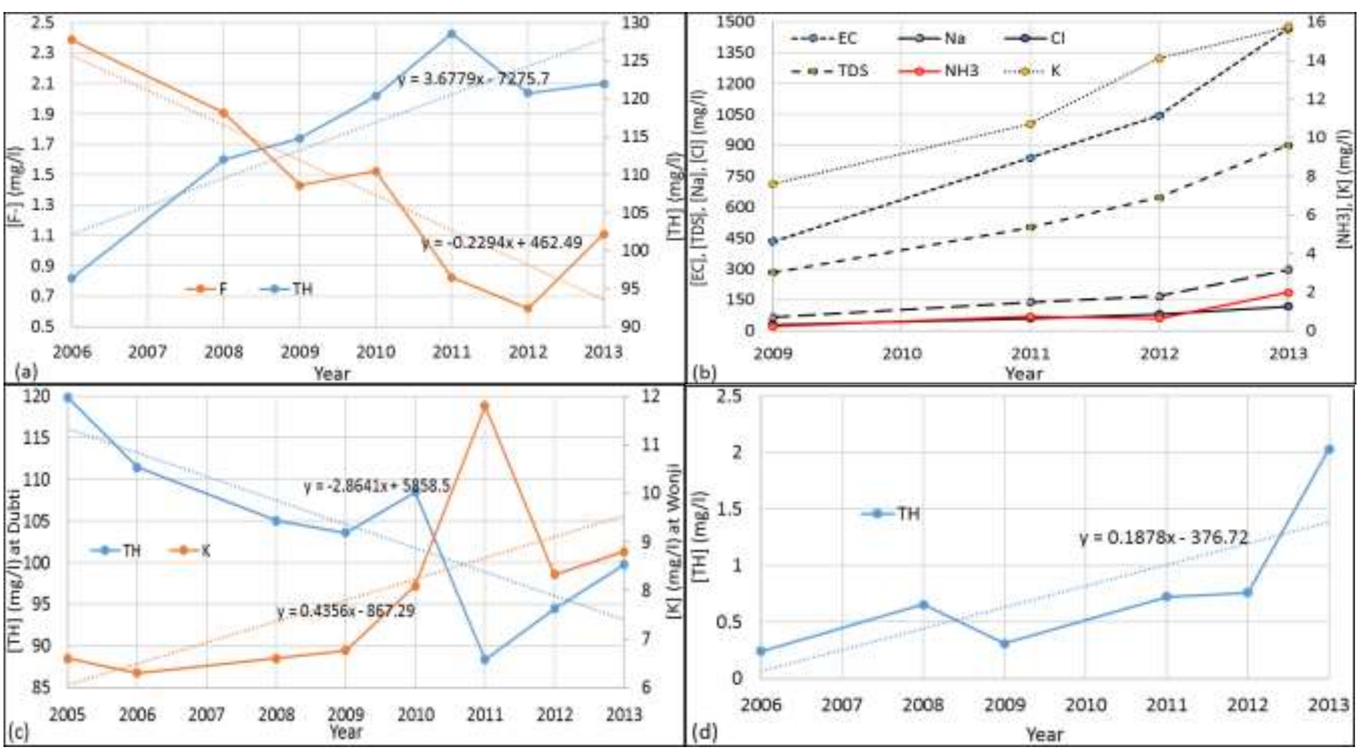

Figure 4. Temporal variation of TH and F^- at Dubti (a); EC, TDS, NH_3, Na, K, and Cl^- at the Office area (b); TH at Wonji (d) in the dry seasons and that of TH at Dubti and K at Wonji in the wet seasons (c).

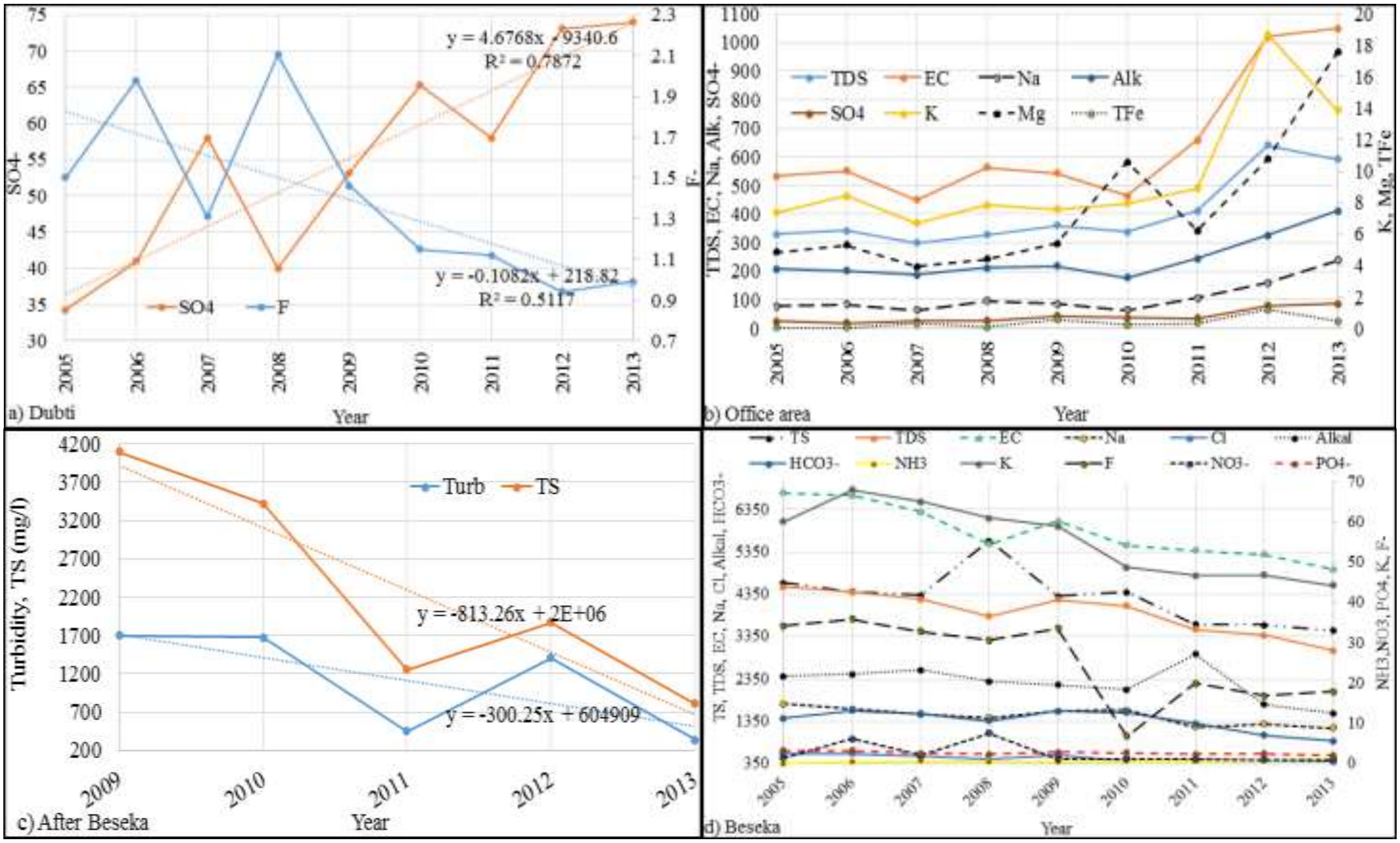

Figure 5. Trend analysis of the water quality data in the nine years' period at Dubti, Office area, After Beseka, and Beseka.

trends, respectively (Figure 5a), while all others did not as their p-values, computed by the exact and approximation methods, were greater than 0.05. A similar analysis at the office area indicated that TDS, EC, Na, K, Mg, TFe, alkalinity and SO_4^{2-} showed an increasing trend since their computed p-value was lower than the significance level.

The test at Beseka revealed that TS, TDS, EC, NH_3, Na, K, F^-, Cl^-, NO_3^-, alkalinity, HCO_3^- and PO_4^{3-} have shown trends since their p-values were smaller than alpha=0.05. At Beseka, all except NH_3 were decreasing. While both turbidity and TS showed a decreasing trend at Beseka, the rest were found to show no trend at all. On the other hand, performing MK test on NH_3, K, Mg, TH and SO_4^{2-} showed an increasing trend, while F^- showed a decreasing trend at Wonji as their p-values is lower than the significance level alpha=0.05, others did not show at all.

Before Beseka, TDS, EC, NH_3, Na, K, TH, TFe, Cl^-, alkalinity, and HCO_3^- showed a trend from 2005 to 2013, while at Weir site, TDS, EC, NH_3, Na, K, Mg, TFe, Cl^-, alkalinity, HCO_3^-, SO_4^{2-}, and PO_4^{3-} showed a trend. With a similar argument MK test indicated at Adaitu that only EC, NH_3, HCO_3^-, and SO_4^{2-} showed trends (Figure 5).

Spatial trend analysis

Spatial analysis of water quality parameters was assessed at the 14 sampling sites of Awash River Basin. From slopes of trendlines of graphs of the water quality indicators, it can be seen that as one moves from the upper to lower parts of basin in the dry season (October-January), EC, TH and Cl^- levels were observed to be decreasing (Figure 6). From the trendlines of the graphs, one can also observe that TH was slightly increasing while TDS, Cl^-, and SO_4^{2-} were decreasing in the same direction in the rainy season (June-September). It can also be seen from the graph that among the sites in both dry and wet seasons, Cl and EC/TDS/SO_4^{2-} were maximized, respectively, at Beseka and before Beseka. At Beseka in both seasons, TH showed a trend opposite to that of EC/TDS/SO_4^{2-}, that is, it revealed an absolute minimum. The most important sites responsible for the spatial variation were Beseka, before Beseka and Sodere spring since it was there, where for instance, EC, TDS,

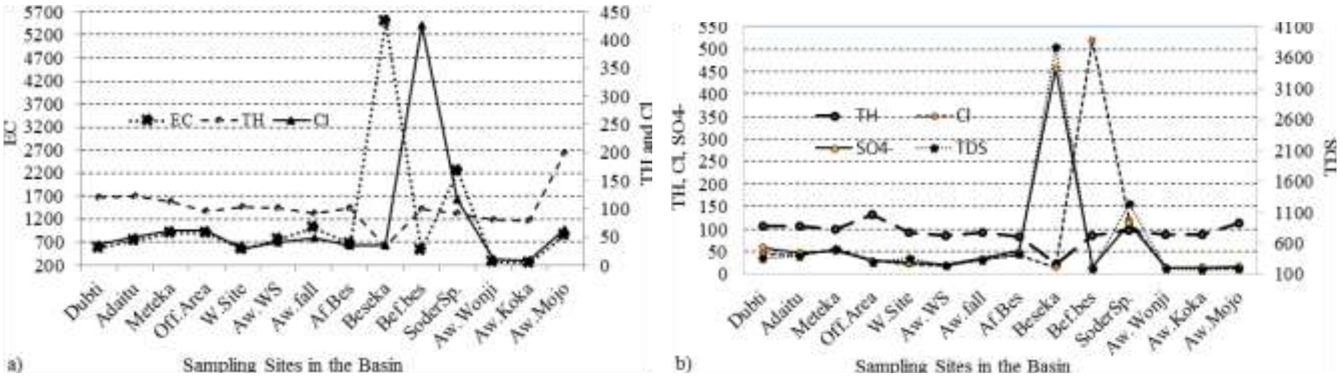

Figure 6. Spatial variation of EC, TH, and Cl in the dry (a) and TH, Cl, SO42- and TDS in the wet (b) seasons of Awash River basin.

TH, Cl, and SO_4^{2-} showed significant variations.

DISCUSSION

As shown in Table 4, the analysis of PCA revealed that the first component, explaining about 56.39% of the variance, was strongly and positively correlated with TS, EC, pH, Na, K, F^-, Cl^-, alkalinity, SO_4^{2-} and PO_4^{3-} and negatively to Ca and TFe. The second component of 16.32% variability explanatory had a relatively high and positive loading on turbidity, calcium and magnesium. The third and the fourth ones were positively related respectively to NH_3 and NO_3^-. From the table of correlation (percentage contributions) of variables to the four principal components, it could be observed that F1 is affected mostly by TS, EC, pH, Na, K, F^-, Cl^-, alkalinity, SO_4^{2-} and PO_4^{3-}; F2 by turbidity, Ca and Mg; F3 by NH_3 and F4 by NO_3^-.

When contribution of each of the sites to the factors is examined with reference to Table 4, the largest percentage of about 59.5 for F1 was taken by Lake Beseka. This is consistent with the respective factor loading values (0.69, 0.88, 0.79, 0.92, 0.85, 0.85, 0.89, 0.87 and 0.71) of TS, EC, pH, F^-, Cl^-, Na, K, SO_4^{2-} and PO_4^{3-} of the water being taken from the sites. Similarly, it can also be seen from Table 4 that among the remaining sampling sites F2 was affected mostly by Wonji, F3 by Meteka, and F4 by Dubti and hence these sites are considered as the most important pollution generators. However, the pH values of the 10 stations in Table 2 showed generally less variation with the range being only from 7.72 to 9.42.

Grouping of the before Beseka, Weir site, Wonji and Dubti into cluster 1 seems to be reasonable as these sites look like that they all are in the vicinity of specialized sugarcane state farms. The fact that Adaitu, office area, Awash water supply and after Beseka are grouped as cluster 2 is consistent with the ground truth that the sites receive waste contributing to turbidity, TS and NH_3 from the dominantly urban and bare-lands. Meteka seems to be unique in that it receives waste from small subsistence-oriented and diversified agricultural and rural areas. Similarly, Lake Beseka is unique in that the lake water quality is by far different from others as it shows the highest values of almost all parameters except for turbidity, TH, Ca, Mg, TFe and NO_3^-. This is confirmed by the study of Dinka (2016) who has urged in his conclusion to avoid even the contact of Lake Basaka water to crops and productive soil in the region because of its pollution.

The MK test at Beseka resulted in the fact that TS, TDS, EC, Na, K, F^-, Cl^-, NO_3^-, alkalinity, HCO_3^- and PO_4^{3-} showed decreasing trends. This finding is in agreement with that of Dinka (2017) who found that in the previous two decades (1960-1980) water quality parameters especially ionic concentrations had shown decreasing trends in response to the fast increasing volume of the lake, which has a dilution effect although his finding has unexpectedly shown stability of most and even increasing trend of other parameters post 2000. Throughout the years from 2005 to 2013, almost all parameters except NH_3 in Figure 5d were showing a temporal decreasing trend. This is suggested to be due to ever increasing volume of the lake from time to time which could dilute the water more (Alemayehu et al., 2006; Dinka, 2017).

On the other hand, EC (salinity-determinant) has shown a spatially increasing trend from upper to middle and then decreasing afterwards (Figure 6), though it was concluded by Taddese et al. (2003) that salinity is generally increasing from the upper to lower basin. The state of EC being maximum in the middle seems, however, to be reasonable since there are Lake Beseka and Sodere spring here having high EC values. Cl has also shown a similar trend as EC throughout the basin while TH showed decreasing trend in the upper and

Table 5. Percentages of land uses of the basin in 2000 and 2014.

LU Classses	2000		2014	
	Area (ha)	Percent	Area (ha)	Percent
Agriculture	2419874	21.3%	3233580	28.1%
Forest	1567171	13.6%	1290576	11.2%
Grassland	1599050	13.9%	1390025	12.1%
Shrubland	1546467	13.4%	1590415	13.8%
Barrenland	3052142	26.5%	2631510	22.8%
Sandy/Exposed rock	944903	8.2%	754349	6.5%
Builtup area	261478	2.3%	541663	4.7%
Waterbodies	102724	0.9%	85190	0.7%

increasing trend in the middle and lower sub-basins. Most other pollutants (discharged into the River from the upper basin) are also seen to decrease their concentraion in the downstream (except for EC, TDS, SO_4^-, and Cl^- at lake Beseka and Sodere hot spring). This might be due to the natural purification process taking place in the course of the river and the lesser amount of waste discharged into the river in the middle and lower basins though the climate is worsening in the downstream tending to concentrate the pollutants while concurrently diminishing the volume of the river due to diversion for different purposes in the lower part.

Comparison of water quality with the land uses in the basin

The 2000 and 2014 SRTM images of the basin have been layer stacked, mosaicked and classified by maximum likelihood of supervised classification technique. Then area in hectare (ha) and percentage statistics for the classified images of each land use class in the two years around which the water quality data were collected have been computed. Table 5 clearly shows why some parameters like SO_4^-, TH, EC, Na, Cl^-, K, and NH_3 increase monotonically at some sites and why the rest vary in the other sites of the basin temporally, which is corresponding to the land use changes in the respective sites. For instance, agriculture and built-up areas have significantly increased in the basin from 2000 to 2014. The associated pollution by agro-chemicals, nutrients, and hardness resulting from urbanization and agricultural intensification was observed.

From the 2014 land use map (Figure 1), the land is more degradable as one moves from upper to lower basin. The land in the lower part of the basin is mostly bare, sandy, rocky and the rest is covered by shrubs, which is an indication of an arid zone. In the middle part of the basin, there are Lake Beseka and Sodere hot spring, both of which are located in a tectonically active Main Ethiopian Rift region and discharge to Awash River. Lake Beseka is not only the fastest growing unlike other lakes in the region but also unique in its water quality characteristics (Dinka, 2017; Alemayehu et al., 2006). The exceptional pollution of this lake and at the site just before it was found in this study by parameters as EC, TDS, SO_4^- and Cl^-, is in line with these findings. This is found to be due to the underlying anthropogenic (increased discharge of the hot springs and discharge of huge amount of irrigation wastewater upstream of the lake, discharges from factory and domestic sewage), natural (weathering of rocks, soil erosion, sediment loading, deposition of animal and plant debris, and solution of minerals in the basin), climatic and geologic factors (Goerner et al., 2009; Dinka, 2017).

Conclusion

Spatial-temporal water-quality analyses usually involve huge multi-dimensional data that need multivariate statistical methods. The PCA resulted in four principal components representing the whole dataset and most parameters are shown by the analysis to vary spatially. It could also identify the most contributing sites and parameters for the principal components. The most sensitive site for the variation is found to be Lake Beseka for which appropriate management need to be sought. Here, agglomerative hierarchical cluster analysis is used to group the ten sampling sites into four clusters pertaining to water quality characteristics. The seasonal MK trend test detected that most of the parameters show temporal as well as spatial trends. If special attention is not payed to the water quality parameters that show a monotonic increasing trends such as EC, TDS, Na, alkalinity, SO_4^{2-}, NH_3, K, and Cl at the Office area; K, Mg, SO_4^{2-} at Wonji; and TH in all the sites, the water quality of the river in particular and the basin in general will deteriorate to the extent that it will not be fit for any intended uses. Thus, the multivariate statistical techniques were proven to be excellent exploratory tools in the analysis and interpretation of the complex dataset on water quality and in understanding their temporal and spatial variations.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

The authors would like to thank schools of Chemical and Bio-Engineering and Civil and Environmental Engineering of Addis Ababa University for their financial, technical and scientific support. They are also very grateful to the general manager of Awash basin authority, Mr Getachew Gizaw, for mobilizing other staffs to provide the long-term water quality data and all others in the authority for their cooperation. The corresponding author would also like to acknowledge Technology Institute of Hawassa University for its sponsorship and provision of study leave.

REFERENCES

Abbaspour S (2011). Water quality in developing countries, south Asia, South Africa, water quality management and activities that cause water pollution. IPCBEE, 15:94-102.

Alemayehu T (2001). The impact of uncontrolled waste disposal on surface water quality in Addis Ababa, Ethiopia. *SINET.* Ethiop. J. Sci. 24(1):93-104.

Alemayehu T, Ayenew T, Kebede S (2006). Hydrogeochemical and lake level changes in the Ethiopian Rift. J. Hydrol. 316(1):290-300.

APHA (American Public Health Association), AWWA (American Water Works Association) & WEF (Water Environment Federation) (1998). Standard methods for the examination of water and wastewater: selected analytical methods approved and cited by the United States Environmental Protection Agency. 18th ed., Washington, DC.

AwBA (Awash Basin Authority) (2014). Strategic River Basin Plan for Awash Basin, Melka Werer, Ethiopia.

Belay EA (2009). Growing lake with growing problems: integrated hydrogeological investigation on Lake Beseka, ZEF, Germany.

Berhe FT, Melesse AM, Hailu D, Sileshi Y (2013). MODSIM-based water allocation modeling of Awash River Basin, Ethiopia. Catena, 109:118-128.

Bu H, Tan X, Li S, Zhang Q (2010). Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. Ecotoxicol. Environ. Saf. 73(5):907-913.

Davies EG, Simonovic SP (2011). Global water resources modeling with an integrated model of the social–economic–environmental system. Adv. Water Resour. 34(6):684-700.

Degefu F, Lakew A, Tigabu Y, Teshome K (2013). The water quality degradation of upper Awash River, Ethiopia. Ethiop. J. Environ. Stud. Manag. 6(1):58-66.

Dinka MO (2016). Quality Composition and Irrigation Suitability of Various Surface Water and Groundwater Sources at Matahara Plain. Water Resour. 43(4):677-689. ISSN 0097_8078.

Dinka MO (2017). Analyzing the temporal water quality dynamics of Lake Basaka, Central Rift Valley of Ethiopia, International Conference on Energy Engineering and Environmental Protection(EEEP2016) IOP Publishing IOP Conf. Series: Earth and Environmental Science 52(1):012057.

Gebre G, Rooijen DV (2009). Urban water pollution and irrigated vegetable farming in Addis Ababa. InWater, sanitation and hygiene: sustainable development and multisectoral approaches. Proceedings of the 34th WEDC International Conference, United Nations Conference Centre, Addis Ababa, Ethiopia, 18-22 May. Water, Engineering and Development Centre (WEDC) Loughborough University of Technology pp. 216-221.

Gebremariam B (2007). Basin scale sedimentary and water quality responses to external forcing in Lake Abaya, southern Ethiopian Rift Valley, Doctoral dissertation, University of Siegen, Germany.

Gedion T (2009). Surface Water–Groundwater Interactions and Effects of Irrigation on Water and Soil Resources in the Awash Valley (Doctoral dissertation, AAU).

Ghosh S, Jintanapakanont J (2004). Identifying and assessing the critical risk factors in an underground rail project in Thailand: A factor analysis approach. Int. J. Project Manag. 22(8):633-643.

Gibrilla A, Bam EKP, Adomako D, Ganyaglo S, Osae S, Akiti TT, Kebede S, Achoribo E, Ahialey E, Ayanu G, Agyeman EK (2011). Application of water quality index (WQI) and multivariate analysis for groundwater quality assessment of the Birimian and Cape Coast Granitoid Complex: Densu River Basin of Ghana. Water Quality, Exposure and Health 3(2):63-78.

Goerner A, Jolie E, Gloaguen R (2009). Non-climatic growth of the saline Lake Beseka, Main Ethiopian Rift. J. Arid Environ. 73(3):287-295.

HGL (Halcrow Group Limited), GIRDC (Generation Integrated Rural Development Consultants) (2009). The Federal Democratic Republic of Ethiopia, Rift Valley Lakes Basin Integrated Resources Development Master Plan Study Project, Phase 2 Final Report, Part I Master Plan, Volume 1: Master Plan and Stakeholder Consultation, Ministry of Water, Irrigation and Energy.

Hulya B, Hayal B (2008). Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. Environ. Geol. 54(2):275-282.

Kithiia SM (2012). Water Quality Degradation Trends in Kenya over the Last Decade. Water Quality Monitoring and Assessment, 509, ISBN: 978-953-51-0486-5, Postgraduate Programme in Hydrology, Department of Geography and Environmental Studies, University of Nairobi, Nairobi, Kenya.

Moreda F (1999). Conceptual rainfall-runoff models for different time steps with special consideration for semi-arid and arid catchments. Laboratory of Hydrology Faculty of Applied Sciences, VUB Pleinlaan, 2, 1050.

Muangthong S (2015). Assessment of surface water quality using multivariate statistical techniques: A case study of the Nampong River Basin, Thailand. J. Ind. Technol. 11(1):25-37.

Osborne JW (2010). Improving your data transformations: Applying the Box-Cox transformation. Practical Assessment, Research & Evaluation, 15(12):1-9.

Pejman AH, Bidhendi GN, Karbassi AR, Mehrdadi N, Bidhendi ME (2009). Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. Int. J. Environ. Sci. Technol. 6(3):467-476.

Rangeti I, Dzwairo B, Barratt GJ, Otieno FA (2015). Validity and Errors in Water Quality Data-A Review. Research and Practices in Water Quality. Durban University of Technology, Durban, South Africa, pp. 95-112.

Reimann C, Bjorvatn K, Frengstad B, Melaku Z, Tekle-Haimanot R, Siewers U (2003). Drinking water quality in the Ethiopian section of the East African Rift Valley I - data and health aspects. Sci. Total Environ. 311(1):65-80.

Romilly TG, Gebremichael M (2011). Evaluation of satellite rainfall estimates over Ethiopian River basins. Hydrol. Earth Syst. Sci. 15(5):1505-1514.

Sen S, Yilmaz A, Temel S (2016). Adaptation of the Attitude toward the Subject of Chemistry Inventory (ASCI) into Turkish. J. Educ. Training Stud. 4(8):27-33.

Shrestha S, Kazama F (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River basin, Japan. Environ. Modelling & Software, 22(4):464-475.

Singh KP, Malik A, Mohan D, Sinha S (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) - a case study. Water Res. 38(18):3980-3992.

Taddese G, Sonder K, Peden D (2003). The water of the Awash River Basin: a future challenge to Ethiopia. International Livestock Research Institute, Addis Ababa.

Taddesse G, McCornick PG, Peden D (2004). Economic importance and environmental challenges of the Awash River Basin to Ethiopia. In Proceedings of Water Rights and Related Water Supply Issues, Water Management Conference, United States Committee on Irrigation and Drainage, Salt Lake City, Utah, USA, pp. 257-268.

Tessema SM (2011). Hydrological Modelling as a Tool for Sustainable Water Resources Management: A Case Study of the Awash River Basin, Licentiate Thesis Environmental Management and Assessment Research Group Department of Land and Water Resources Engineering, Royal Institute of Technology, Stockholm, Sweden.

Tiruneh AT (2005). Water Quality Monitoring in Lake Abaya and Lake Chamo Region. *PhD Thesis*, University of Siegen, Germany.

Wang XL, Lu YL, Han JY, He GZ, Wang TY (2007). Identification of anthropogenic influences on water quality of Rivers in Taihu watershed. J. Environ. Sci. 19(4):475-481.

Xu HS, Xu, ZX, Wu W, Tang FF (2012). Assessment and spatio-temporal variation analysis of water quality in the Zhangweinan River Basin, China. The 18th Biennial Conference of International Society for Ecological Modelling. Procedia Environ. Sci. 13:1641-1652.

Yue S, Pilon P, Cavadias G (2002). Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. J. Hydrol. 259(1):254-271.