Evaluation of river water quality monitoring stations of River Harmu by principal component analysis

Mrigendra Kumar 1, Ramakar Jha 2

1 Department of Civil Engineering, Adani Power Limited, Raipur, Chhattisgarh, India, - 493225;
2 Department of Civil Engineering, National Institute of Technology Patna, Bihar, India – 800005.

Abstract

The development of a surface water monitoring network is a critical element in the assessment, restoration, and protection of stream water quality. This study applied principal component analysis (PCA) and principal factor analysis (PFA) techniques to evaluate the effectiveness of the surface water quality-monitoring network in a river where the evaluated variables are monitoring stations. The objective was to identify monitoring stations that are important in assessing monthly variations of river water quality. Twelve stations used for monitoring physical, chemical, and biological parameters, located at the main stem of the upper zone of Harmu river, Jharkhand, were selected for the purpose of this study. Results show that 3 monitoring stations were identified as less important in explaining the monthly variance of the data set, and therefore could be the non-principal stations. Results reveal that total organic carbon, dissolved organic carbon, total nitrogen, dissolved nitrate and nitrite, orthophosphate, alkalinity, salinity, Mg, and Ca were the parameters that are most important in assessing variations of water quality in the river. This study suggests that PCA techniques are useful tools for identification of important surface water quality monitoring stations and parameters.

Keywords- Monitoring network; Principal component analysis; Surface water quality

1. Introduction

Pollution of surface water with toxic chemicals and excess nutrients, resulting from storm water runoff, upper surface zone leaching, and groundwater discharges, has been an issue of worldwide environmental concern. With an increased understanding of the importance of drinking water quality to public health and raw water quality to aquatic life, there is a great need to assess surface water quality. This is true for the upper zone of Harmu river, Jharkhand. Pollution of the UZHR with contaminants such as nutrients, hydro- carbons, pesticides, and heavy metals comes from both point and non-point sources. These sources are the results of surface runoff generated from urban, rural, and agricultural lands; discharge from ditches and creeks; groundwater seepage from malfunctioning septic tank systems; aquatic weed control and naturally occurring organic inputs; and atmospheric deposition. The degradation of water quality due to these contaminants has resulted in altered species composition and decreased the overall health of aquatic communities within the river basin (Campbell et al., 1993; Durell et al., 2001; Ouyanget al., 2002).

The primary objectives are to identify water quality problems, describe seasonal and spatial trends for developing qualitative and quantitative models of the riverine ecosystem, and determine permit compliance. Since its inception, the monitoring network has become one of the most critical efforts in assessment of surface water pollution in the UZHR and has been a significant resource for others working to prevent pollution of the river (Campbell et al., 1993). However, efforts to determine the effectiveness and efficiency of the monitoring network are still warranted. To this end, the principal component analysis (PCA) techniques were employed in this study.

PCA are multivariate statistical techniques used to identify important components or factors that explain most of the variances of a system. They are designed to reduce the number of variables to a small number of indices (i.e., principal components) while attempting to preserve the relationships present in the original data. The problems of indicator parameter or import monitoring station identification, data reduction and interpretation, and characteristic change in water quality parameters can be approached through the use of the PCA. Details for mastering the arts of PCA are published elsewhere (Manly, 1986; Davis, 1986; Wackernagel, 1995; Tabachnick and Fidell,2001).

In recent years, the PCA and PCF techniques have been applied to a variety of environmental issues, including evaluation of ground water monitoring wells, interpretation of groundwater hydrographs, examina- tion of spatial and temporal patterns of heavy metal contamination and identification of herbicide species related to hydrological conditions. Some examples of PCA and PCF applications in environmental practices are described below.

Measurements of water level in wells are a routine part of groundwater studies. Recently, Winter et al. (2000) applied the PCA techniques to investigate the areal distribution of various types of water level fluctuation patterns within a study area and to determine if fewer wells could be measured while still achieving effective long-term monitoring goals at four small, lake-watershed research sites in the USA. These authors found that the PCA technique was very useful in summarizing information from large data sets to select long-term monitoring wells, which would greatly reduce the cost of monitoring programs.

Additionally, the PCA technique has been used to estimate spatial and temporal patterns of heavy metal contamination (Shine et al., 1995); to investigate nutrient gradients within a eutrophic reservoir (Perkins and Underwood, 2000); and to identify the major herbicide compositions causing the observed data variations (Tauler et al., 2000). These studies have provided good examples of the effective application of PCA. However, there are few documented examples of the evaluation of the highly dynamic and complex surface water quality monitoring networks in river systems using the PCA or PFA technique.

The aims of this study are to demonstrate the application of these novel data reduction techniques (i.e., the PCA and PFA techniques) to evaluate the potential for reducing the number of ambient water quality monitoring stations located in the main stem of the UZHR for long-term monitoring purposes and to evaluate the importance of various water quality parameters. The specific objectives are to: (1) present detailed procedures on how to interpret PCA results, (2) identify the non-principal surface water quality monitoring stations, and (3) extract the parameters that are most important in assessing variations in river water quality.

2. Study area

The study area of river Harmu is categorized in three distinct broad physiographic namely Achaean Precambrian metamorphic rock, Tertiary, and Quaternary alluvium. Granite and quartzite are the major rock in the upper region and the deposition of sediments in the lower region of the Subarnarekha river basin, major rock types are and the lower section shows mainly vast deposition of sediments.

The Harmu River is a tributary of Subarnarekha River that forms the main local river system in Jharkhand and passing through Ranchi City (Fig.1). The study area is located between 85º07’ to 85 º 34’ E longitudes and 23º11’ to 23º32’N at an average elevation 640 m above mean sea level. The Study area witnessed rapid development during past decades in terms of urbanization, industrialization, and also population increase substantially. The Harmu River which once provided water to habitation settled near its course has now turned into a waste dumping ground resulting in complete deterioration of water quality and aesthetics. Geologically, The river Harmu flows from west to north- east and it is the minor tributaries of Subarnarekha River. The Subarnarekha River and its tributaries constitute the local river system. During the survey it was found that forest cover in the bank of Harmu River has almost non-existent. In addition to there is a large and small slum encroachments along the bank of the river. This encroachment of bank of the river has a resulted into a drastic situation where the Harmu River has slowly dried up. Summer temperatures range from 20°C-42°C degrees whereas winter temperatures from 0°C-25°C degrees. The normal annual rainfall indicates that average rainfall is 1260 mm with maximum rainfall (90%) concentrated during monsoon months (June-September).

The Sampling Locations in Harmu river system for collecting water quality data is shown in Fig.2.

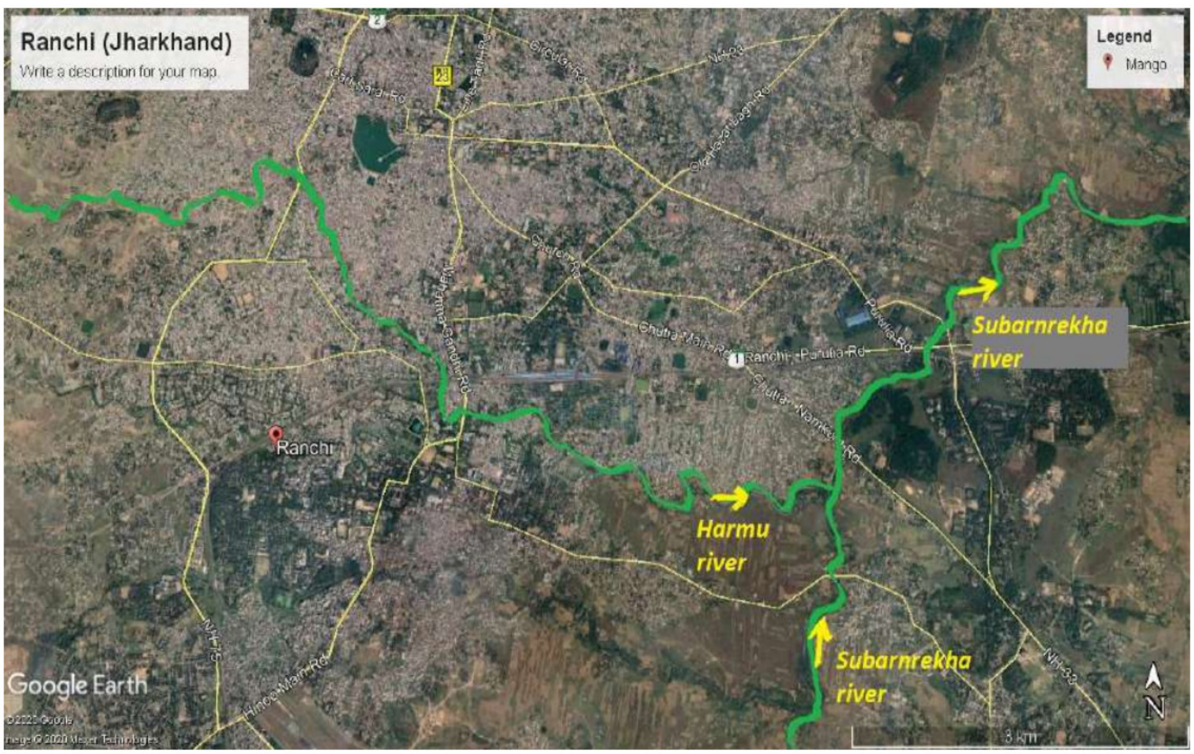


Fig.1 Location of Harmu River tributary of Subarnarekha River

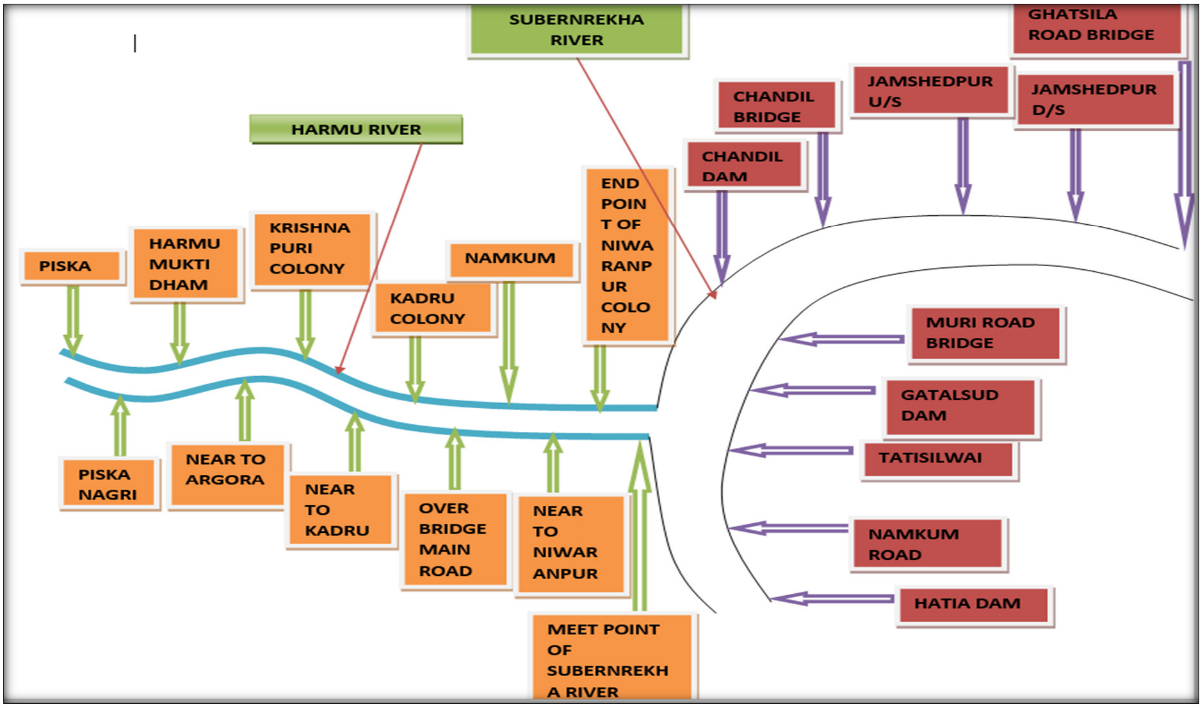


Fig.2 Sampling Point of Harmu River

Table-1

Water quality data from the monitoring stations located on the Harmu River used for this study

Apr-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namkum m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	9.2	8	6.5	6.1	9.2	8.2	9	9	8	8.3	9.3	9.2
Tepm.	19.5	20	20.5	22	23.5	24	26	26	27.5	27.5	28	28
BOD	18.8	17.8	39	37	32.6	32.5	29.8	29	25	24	20	16
COD	67.5	74.3	61.3	55	54	53.3	51.7	52.5	48.3	64.8	67.7	71.7
EC	325	358	413	370	303	299	348	353	325	364	455	402
TDS	4004	4344	3592	3410	3171	3488	3418	3076	2891	3382	3247	3117
TSS	145.8	140	185	180	175	180	178.5	165.8	135.5	140.6	125.6	124
Hardness	195	200	275	265	280	285	190	185	255	260	285	280
FC	185	190	300	325	295	302	285	230	320	320	310	215
TC	1795	1851	1487	1823	1851	1851	1795	2511	1935	1627	1851	1459
DO	7.6	6.5	3.2	3.3	3.7	4.1	4.3	4.4	5.4	5.6	5.9	6.2

May-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	8.5	8.6	6	6.1	9	9.2	8	9	7.5	7.8	9	9.5
Tepm.	23	23.5	25	25.8	28.8	28	30.5	30.6	30	30	30.8	31
BOD	20.8	18	41.8	40.5	36.7	33	31	30.2	26.7	23.4	21.5	18
COD	37.8	32.7	76	73.6	66.7	60	56.4	54.9	48.5	42.5	39.1	32.7
EC	179	155	504	488	379	417	489	440	414	404	465	372
TDS	2524	2184	5072	4914	4570	5027	4926	4433	4167	4875	4680	4493
TSS	160	150	195	185	180	178	170.5	165	140	135	128	123.6
Hardness	205	195	280	275	290	280	192	180	250	255	280	285
FC	190	200	305	330	310	320	295	235	340	320	315	220
TC	1066	1122	1711	1851	1739	1795	1655	1788	1907	1795	1767	1234
DO	7	6.4	2.1	2.3	2.5	3.5	4	4.3	5	5.2	5.8	6

Jun-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	8	8.2	6.2	6.1	8.8	8.7	8.5	8.6	7.2	7.5	9.5	8
Tepm.	25	26	26	27.3	27.8	28	28.5	30	30.6	31	31	30.6
BOD	22.6	20.9	44.8	44.6	39	35.8	32.6	30	26.6	24	22	17.6
COD	45.2	41.8	89.6	89.2	78	71.6	65.2	60	53.2	48	44	35.2
EC	195	180	540	538	417	459	539	485	456	445	512	410
TDS	2742	2536	5436	5412	5033	5536	5425	4883	4590	5370	5155	4949
TSS	155	145.8	190	182.3	168	172	168.3	164	138.5	130	125	120.5
Hardness	210	205	295	285	300	290	198	185	265	260	290	295
FC	195	200	310	340	320	320	300	240	345	330	320	228
TC	1094	1122	1739	1907	1795	1795	1683	1826	1935	1851	1795	1279
DO	6.5	6	2	2.1	2.3	3.1	3.6	3.9	4.5	4.8	5.3	5.6

Jul-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	8	8.2	6	6.5	9	8.8	8	8.6	7.7	8	9	9
Tepm.	25	26	26	27.3	27.8	28	28.5	30	30.6	31	31	30.6
BOD	19.4	18	38.2	36.8	31	29.6	28.5	27.3	24.8	22.1	19.8	15.4
COD	25.9	24	50.9	49.1	41.3	39.5	38	36.4	33.1	29.5	26.4	20.5
EC	167	155	461	444	344	379	445	400	376	367	423	338
TDS	2354	2184	4635	4466	4153	4568	4477	4029	3787	4431	4254	4084
TSS	140.5	140	175	168	172	170.8	185	186	144	152.3	128	128
Hardness	180	180	225	225	286	274	183	184	265	265	290	290
FC	180	180	225	225	286	274	283	184	265	265	290	290
TC	1010	1010	1262	1262	1604	1537	1588	1400	1487	1487	1627	1627
DO	7.5	7.5	3.1	3.1	3.8	3.8	4	4.2	4.3	4.6	5	5.8

Aug-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	9	8.6	6.2	6.1	9	8.5	9	9	7.7	7.9	9.5	9.5
Tepm.	24	26	28	28.5	28.8	29	30.5	30.6	30	30	30.4	30.5
BOD	17	16	35	33	29	28.1	27.2	25.7	24.1	20.6	18.9	15
COD	22.7	21.3	46.7	44	38.7	37.5	36.3	34.3	32.1	27.5	25.2	20
EC	147	138	422	398	309	339	399	359	337	329	379	303
TDS	2063	1942	4247	4004	3724	4096	4014	3613	3396	3973	3814	3661
TSS	155	145.8	190	182.3	168	172	168.3	164	138.5	130	125	120.5
Hardness	185	206	265	245	286	274	183	184	265	265	295	295
FC	178	194	295	326	300	309	284	225	325	325	305	212
TC	999	1088	1655	1829	1683	1733	1593	1712	1823	1823	1711	1189
DO	7.5	7.5	3.1	3.1	3.8	3.8	4	4.2	4.3	4.6	5	5.8

Sep-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	8.5	8.6	6	6.1	9	8.7	8	9	8	8.2	9	9.5
Tepm.	22	25.4	25.4	26	26	28	28.5	29	29	30	30.2	30.5
BOD	17	15.8	34	32.4	28.4	28.1	27	25	24	20.1	18.3	14.8
COD	22.7	21.1	45.3	43.2	37.9	37.5	36	33.3	32	26.8	24.4	19.7
EC	147	136	410	391	303	333	392	353	331	323	372	298
TDS	2063	1917	4126	3932	3657	4023	3943	3549	3336	3903	3747	3597
TSS	134.5	135	168	168	174	178.6	188.8	183.2	145.6	152.3	128.9	129
Hardness	180	180	225	225	286	274	183	184	265	265	290	290
FC	195	200	310	340	320	320	300	240	345	330	320	228
TC	1094	1122	1739	1907	1795	1795	1683	1826	1935	1851	1795	1279
DO	7.5	7.5	3.1	3.1	3.8	3.8	4	4.2	4.3	4.6	5	5.8

Oct-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	8	8.2	6.2	6.1	9	8.4	8.3	9	8.2	8.5	9.5	8
Tepm.	8	8.5	11	14	14	14.5	15	15.6	17	17.6	18	18
BOD	16	15.8	33	32	28	27.4	26	24.3	24	19.5	18	14.6
COD	21.3	21.1	44	42.7	37.3	36.5	34.7	32.4	32	26	24	19.5
EC	138	136	398	386	299	329	387	348	327	319	368	294
TDS	1942	1917	4004	3883	3611	3972	3893	3504	3294	3854	3700	3552
TSS	155	145.8	190	182.3	168	172	168.3	164	138.5	130	125	120.5
Hardness	185	185	200	200	286	274	182	185	260	263	285	285
FC	200	210	310	335	305	315	295	240	330	330	320	225
TC	1122	1178	1739	1879	1711	1767	1655	1826	1851	1851	1795	1262
DO	6.8	6.8	2.9	2.9	3.6	3.8	4	4.5	4.5	4.6	5.3	6

Nov-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namkum**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernrekha river**
pH	8	8.2	6.2	6.1	8.5	8.2	8	9	8	8.2	9.5	8
Tepm.	6	6.5	7	7	7.8	9	10.6	11	11.5	11.5	12	12
BOD	15.5	15.2	32	31	26	27	25	24	23.5	19.5	17.5	14
COD	20.7	20.3	42.7	41.3	34.7	36	33.3	32	31.3	26	23.3	18.7
EC	134	131	386	374	290	319	375	337	317	309	356	285
TDS	1881	1844	3883	3762	3499	3849	3772	3395	3191	3733	3584	3441
TSS	138	135	170	168	174	178	190	183.2	145	152.3	128.9	125
Hardness	265	235	345	310	300	305	240	195	280	275	300	300
FC	190	200	305	330	310	320	295	235	340	320	315	220
TC	1066	1122	1711	1851	1739	1795	1655	1788	1907	1795	1767	1234
DO	6.8	6.8	2.9	2.9	3.6	3.8	4	4.5	4.5	4.6	5.3	6

Dec-21	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namkum**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernrekha river**
pH	8.5	8.6	6	6.1	9	8.5	8	9	8.1	8.2	9	9.5
Tepm.	5.5	6	7	7	7.8	9	10.6	11	12	12	13	13.5
BOD	15.2	15.5	32	31.5	31	28.5	25	24.2	24	20.5	18.9	18
COD	20.3	20.7	42.7	42	41.3	38	33.3	32.3	32	27.3	25.2	24
EC	131	134	386	380	295	324	381	343	322	314	362	290
TDS	1844	1881	3883	3822	3554	3909	3831	3448	3241	3792	3640	3494
TSS	155	150	195	185	180	178	170.5	165	140	135	128	125
Hardness	260	190	255	245	300	274	183	184	265	265	285	280
FC	195	205	315	340	320	330	305	240	355	330	320	225
TC	1094	1150	1767	1907	1795	1851	1711	1826	1992	1851	1795	1262
DO	6.5	6.5	3.1	3.1	3.5	3.5	3.8	4	4.5	4.6	5.4	7

Jan-22	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namkum**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernrekha river**
pH	9.2	8	6.5	6.1	9.2	8	9	9	8.5	8.6	9.3	9.2
Tepm.	8	8.5	11	14	14	14.5	15	15.6	17	17.6	18	18
BOD	15	14.6	30	29	28	27.9	24	23.6	23	19.7	17	16
COD	20	19.5	40	38.7	37.3	37.2	32	31.5	30.7	26.3	22.7	21.3
EC	129	126	362	350	271	298	351	315	297	289	333	267
TDS	1820	1772	3640	3519	3273	3600	3528	3175	2985	3492	3352	3218
TSS	140	130.6	155	160	165	161	185	183.2	145.6	152.3	128.9	124.6
Hardness	285	230	260	255	290	285	240	200	275	280	285	280
FC	210	215	320	355	330	340	320	250	360	345	330	225
TC	1178	1206	1795	1992	1851	1907	1795	1903	2020	1935	1851	1262
DO	6.3	6.3	2.8	2.8	4.1	4.1	3.8	4.1	4.5	4.6	5.5	7

Feb-22	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	9.5	9	6.8	6	9	8.8	8	9	8.5	8.5	9	8.5
Tepm.	9	8.5	11	14.5	14.8	14.8	15	15.6	17	17.6	19	19
BOD	19	18.5	40.6	38.9	33	32.5	30	29.6	25	24	20.6	15.8
COD	25.3	24.7	54.1	51.9	44	43.3	40	39.5	33.3	32	27.5	21.1
EC	164	160	490	469	364	400	470	423	398	388	447	358
TDS	2306	2245	4927	4720	4390	4829	4732	4259	4003	4684	4497	4317
TSS	145.8	140	185	180	175	180	178.5	165.8	135.5	140.6	125.6	124
Hardness	185	206	265	245	286	274	183	184	265	265	295	295
FC	185	206	265	245	286	274	315	220	295	290	290	230
TC	1038	1156	1487	1374	1604	1537	1767	1674	1655	1627	1627	1290
DO	5.8	5.8	2.5	2.5	3.8	4.3	3.9	4.6	5	5.4	5.8	6.2

Mar-22	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	9.2	8	6.5	6.1	9.2	7.8	9	9	8.3	8.4	9.3	9.2
Tepm.	16	16.8	17	18	18.9	20	20	22	22.5	24	25	25.6
BOD	20	19.8	41	40.6	34.6	33	30	29.6	25	24	21	15.9
COD	30.8	30.5	63.1	62.5	53.2	50.8	46.2	45.5	38.5	36.9	32.3	24.5
EC	173	171	494	490	380	418	491	442	415	405	466	373
TDS	2427	2403	4975	4927	4582	5040	4939	4445	4178	4888	4692	4504
TSS	140.5	140	175	168	172	170.8	185	186	144	152.3	128	128
Hardness	290	260	275	255	290	285	240	200	275	280	285	295
FC	200	210	310	335	305	315	295	240	330	330	320	225
TC	1122	1178	1739	1879	1711	1767	1655	1826	1851	1851	1795	1262
DO	6	5.8	2.8	2.9	3.4	3.9	4	4.1	4.7	5.4	5.8	6.2

Apr-22	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	8.5	8.6	6	6.1	9	8.8	8	9	8	8.1	9	9.5
Tepm.	19.5	20	20.5	22	23.5	24	26	26	27.5	27.5	28	28
BOD	22.5	20.9	44	43	38.6	35.8	32.6	30	26.6	24.8	22.6	18.9
COD	37.5	34.8	73.3	71.7	64.3	59.7	54.3	50	44.3	41.3	37.7	31.5
EC	194	180	530	518	402	442	520	468	440	429	494	396
TDS	2730	2536	5339	5218	4853	5338	5231	4708	4426	5178	4971	4772
TSS	160	150	195	185	180	178	170.5	165	140	135	128	123.6
Hardness	195	200	275	265	280	285	190	185	255	260	285	280
FC	195	200	275	265	280	285	290	185	255	260	285	240
TC	1094	1122	1543	1487	1571	1599	1627	1408	1431	1459	1599	1346
DO	5.6	5.3	2.3	2.3	3	3.5	3.7	4	4.5	5	5.5	5.6

May-22	Site-1	Site-2	Site-3	Site-4	Site-5	Site-6	Site-7	Site-8	Site-9	Site-10	Site-11	Site-12
Station	**Piska**	**Piska Nagri**	**Harmu Mukti Dham**	**Near to Argora**	**Krishna puri Colony**	**Near to Kadru**	**Kadru colony**	**Over Bridge Main Road**	**Namku m**	**Near to Niwaran pur colony**	**End point of Niwaran pur colony**	**Meet Point of Subernr ekha river**
pH	8	8.2	6	6.5	9	8.4	8	8.6	8	8.4	9	9
Tepm.	25	26	26	27.3	27.8	28	28.5	30	30.6	31	31	30.6
BOD	23	22.3	44.8	44.6	39	36.8	33.6	31.2	27	25.3	23	19
COD	41.8	40.5	81.5	81.1	70.9	66.9	61.1	56.7	49.1	46	41.8	34.5
EC	198	192	540	538	417	459	539	485	456	445	512	410
TDS	2791	2706	5436	5412	5033	5536	5425	4883	4590	5370	5155	4949
TSS	155	145.8	190	182.3	168	172	168.3	164	138.5	130	125	120.5
Hardness	210	205	295	285	300	290	198	185	265	260	290	295
FC	195	200	310	340	320	320	300	240	345	330	320	280
TC	1094	1122	1739	1907	1795	1795	1683	1826	1935	1851	1795	1571
DO	5.3	5	2.3	2.2	2.7	3.2	3.3	3.4	3.8	4.6	5	5.6

Data collected are used to support a number of scientific and management investigations, including (1) identifying water quality problem areas (2) determining the amount, or mass load, of pollutants entering rivers from tributaries and point sources; (3) estimating daily, seasonal, and long- term water quality trends; (4) assessing the ability of bestmanagement practices to improve non-point source pollution; (5) examining how water quality affects the plants and animals living within the river; and (6) investigating how water quality varies within different reaches of the river. In this study, 12 water quality monitoring stations located in the main stem of the UZHR were selected for analysis (Fig. 1). The physical and chemical parameters collected from those 12 monitoring stations and used in this study are givenin Table 1.

3. **Analysis Procedure**

PCA was first performed in this study to identify the potential for reducing the number of monitoring stations. This analysis investigated the annual variations in water quality parameters measured from the ambient water quality monitoring stations of the UZHR over a 1 year time period (April 2021- May 2022). Eleven water quality parameters from the 12 stations were examined in thisstudy (Table 1). The procedures used for PCA are described below.

3.1. Selection of water quality data

The ambient water quality monitoring databases from the main stem of the LSJR were used in this study. Station locations are where conditions are the most representative and homogeneous, away from transitional areas such as point source mixing zones and near-shore regions. Some stations are sampled daily or monthly and a couple of stations are sampled seasonally due to budget constraints. Timing of sample collection isroutine and not intended to capture any specific flow or rainfall events.

Data from these stations were collected at different times of day and/or different days of the year for each parameter. Plots of all of the data in Excel spreadsheets show that they are not normally distributed and are positively skewed. For the purpose of this analysis, the monthly median values for each parameter were used. The choice of the median values rather the mean values was based upon the fact that the measured parameter values are very skewed. In general, when this is the case, the median is a better measurement than the mean (Anderson and Sclove, 1986).

In this study, a 1-year time period was selected based on the following reasons: (1) no complete data set is available to include all of the water quality parameters used in this study beyond the 1-year period. In other words, although some parameters have been measured for a period of 15 years, others have only recently been added; and (2) the PCA requires no missing values in a data set.

3.2 Selection of computation method

PAST is free software for scientific data analysis, with functions for data manipulation, plotting, univariate and multivariate statistics, ecological analysis, time series and spatial analysis, morph metrics and stratigraphy. PAST is a practical tool designed to help you analyze scientific data by calculating statistical indicators and drawing plots? The statistics section includes univariate and multivariate analysis methods such as variance analysis, interclass correlation or canonical correspondence. package (Version 4.03), developed by (Oyvind Hammer). was employed to perform principal component analyses. This software has the PAST modules that can perform the analyses.

Mathematically, PCA normally involve the following five major steps: (1) start by coding the variables x_1, x_2, y , x_p to have zero means and unit variance, i.e., standardization of the measurements to ensure that they all have equal weights in the analysis;

(2) calculate the covariance matrix C; (3) find the eigenvalues l_1; l_2; …; l_p and the corresponding Eigen- vectors a1, a2, y, a_p; (4) discard any components that only account for a small proportion of the variation in data sets (Manly, 1986); This conservative criterion was selected because the study area was large and the river system was highly non-linear and dynamic. Stations that do not have any factors with correlation coefficients greater than this value were considered asnon-principle stations.

4. Results and discussion

4.1. Principal component analysis

In a PCA, the number of components is equal to the number of variables. However, a component is not only comprised of a single variable but rather all of the variables used in a study. For example, there are 12 variables (stations) used in this study, which produce 12 components. In each component, there are 12 variables (or stations) as shown in Eq. (1) below. The PCA results showed that of the 12 components, the first component accounted for about 95.6% and the second component accounted for about 6.5% of the total variance in the data set. These two components together accounted for about 99.1% of the total variance and the rest of the 10 components only accounted for about 0.8%. Therefore, our discussions should focus only on the first two components.

From the eigenvectors obtained in the PCA, the first component, Z_1, can be given as

Z_1 ¼ 0:21x_1 þ 0:21x_2 þ 0:20x_3 þ 0:22x_4 þ 0:22x_5
þ 0:21x_6 þ 0:21x_7 þ 0:20x_8 þ 0:22x_9 þ 0:20x_{10}
þ 0:20x_{11} þ 0:22x_{12} - Eq.(1)

Where x is the monitoring station, the subscripts denote the station numbers, and the coefficients are the eigenvectors. This component had almost equal loadings (i.e., similar coefficient values in Eq. (1) on all variables and therefore is a measure of overall performance of the stations.

It is apparent that Z_1 has an extremely high correlation with the measured data as it accounts for 95.6% of the data variance, which would indicate that only one major source of data variation is present. This finding is somewhat different from other studies where many more components are needed to explain the same amount of variance (Bengraine and Marhaba, 2003).

Similarly, the second component can be expressed as

Z_2 ¼ — 0:23x_1 — 0:181x_2 — 0:41x_3 — 0:04x_4 — 0:09x_5
P 0:27x_6 þ 0:26x_7 þ 0:36x_8 þ 0:12x_9 þ 0:37x_{10}
— 0:36x_{11} — 0:08x_{12} - Eq. (2)

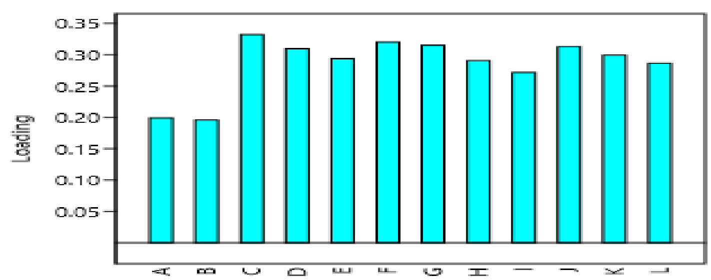
This equation shows that the second component, Z_2, will be high if x_2 to x_6, x_9 and x_{12} are high but x_1 to x_5 and x_{10} are low. Hence, Z_2 represents a difference among the stations. The low coefficients of x variables such as those associated with x_4 and x_{10} mean that the values of these variables have little effect on Z_2.

A graphical representation of the first two component loadings is given in Fig. 2. This diagram was constructed using the eigenvectors from the first two components. It becomes clear that the first component had similar loadings (eigenvectors) for all of the monitoring stations and therefore this component represents the overall performance of all the monitoring stations (Fig. 2A), while the second component measured the difference among the stations (Fig. 2B).

4.2 Extraction of important monitoring stations

Although the PCA results identified two principal components accounting for 99.1% of the annual variance in the dataset, they did not indicate which monitoring stations contributed most to this variance. To address this, a Principal Factor Analysis (PFA) was conducted. As with PCA, the number of factors in PFA equals the number of variables in this case, 12 monitoring stations—resulting in 22 factors. The Eigen value criterion used to retain principal factors was 4×10−64 \times 10^ {-6}4×10−6 (the SAS default), leading to the selection of 12 factors. PFA results showed that Factor 1 accounted for 94.7% and Factor 2 for 4.5% of the total variance, matching the 99.1% explained variance obtained from PCA.

Table 1 presents the rotated factor correlation coefficients for all stations. A coefficient greater than 0.75 (75%) was considered significant, a conservative threshold chosen due to the large, highly non-linear, and dynamic nature of the river system. Stations with correlation coefficients below this value for all factors were classified as non-principal.



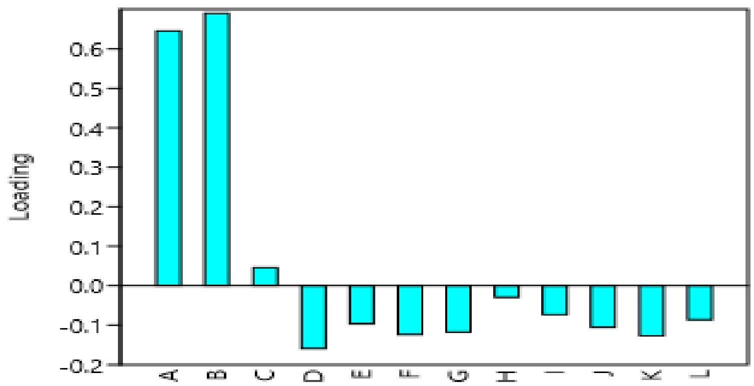


Fig. 3 Component loadings for the first component (A) and the second component (B).

4.3 Validation of PCA results

Before applying the above findings, their scientific reliability must be validated using independent methods. One approach is to compare water quality data analyses with and without the three non-principal stations. In this study, two cases were developed for comparison:

Case 1: Data from only the principal stations were used to develop four regression relationships:

(1) dissolved organic carbon (DOC) vs. water color;

(2) Chlorophyll *a* vs. total phosphorus (TP);

(3) biochemical oxygen demand (BOD) vs. total organic carbon (TOC); and

(4) chlorophyll *a* vs. total dissolved nitrogen (TDN).

Case 2: Data from all stations (principal and non-principal) were used to develop the same four relationships.

The two cases were then compared to determine whether including the three non-principal stations improved the regression fits. For example, comparison of DOC–water color relationships (Fig. 3) showed that adding the non-principal stations did not improve curve fitting. The regression using all 12 stations had an R2R^2R2 value of mentioned in (Fig. 4) relationships, where regressions based on principal stations consistently yielded slightly better R2R^2R2 values than those including all stations.

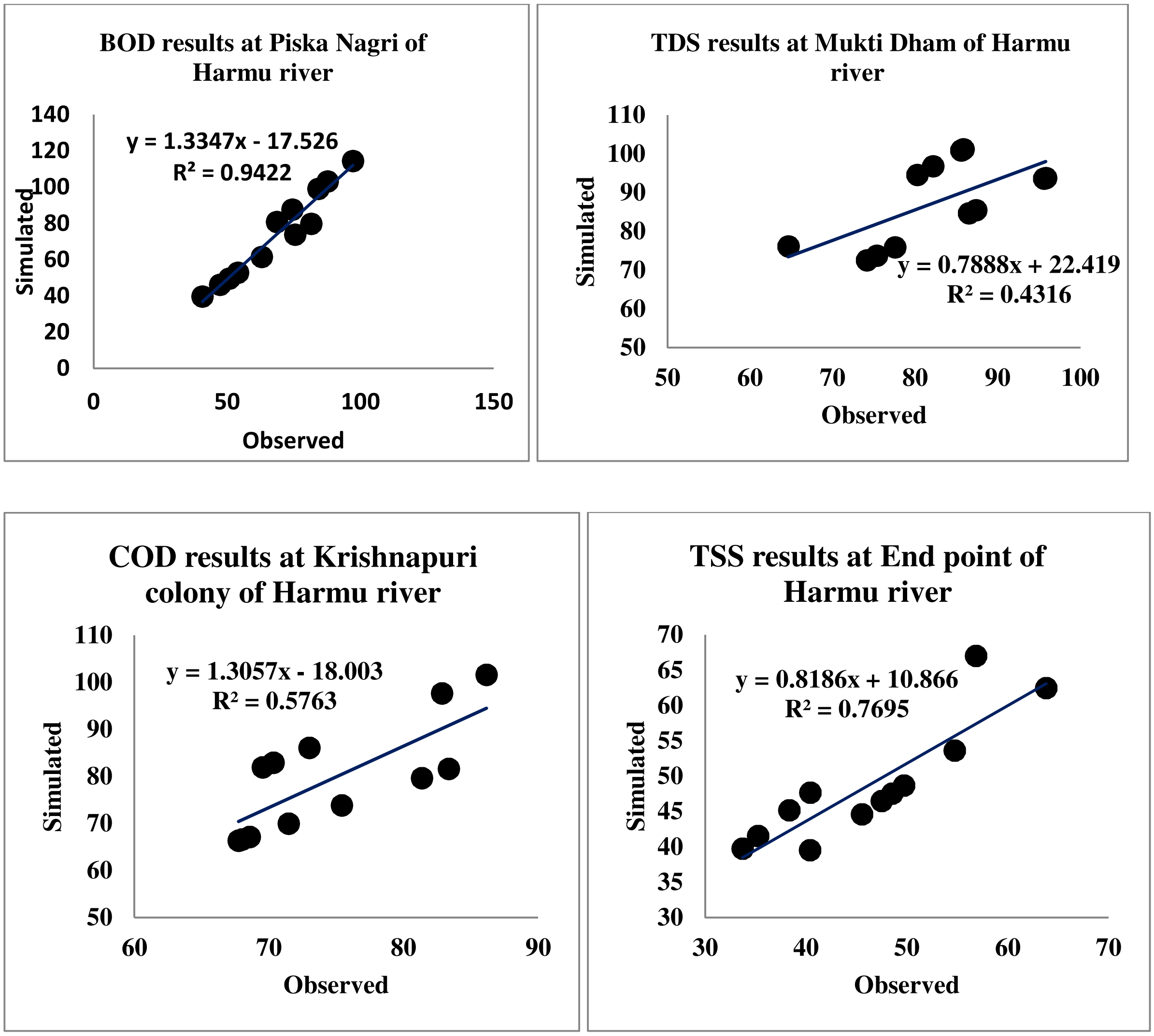


Fig.4 Relationship between the principal stations (A), all stations (B)

4.4 Identification of important water quality parameter

Characterizing changes in surface water quality is essential for assessing the potential impacts of natural or anthropogenic point and non-point pollution sources on ecosystem health. In this study, 12 commonly used surface water quality parameters site for Florida watersheds were selected to evaluate variations in LSJR water quality. A Principal Factor Analysis (PFA) was applied to identify the parameters most influential in explaining these variations.

The selection of 11 parameters, rather than the full set of 12, was due to a limitation of the SAS software. When the number of variables (e.g., water quality parameters) exceeds the number of observations (e.g., monitoring stations), singularity issues arise in estimating covariance and correlation matrices in PCA or PFA, leading to unstable solutions. Some studies have reported that, for PCA, the number of observations must exceed the number of variables to ensure solution stability (Yu et al., 1998). In contrast, other studies have shown that PCA can be applied to any type of data matrix, regardless of the relative number of variables and observations (Golub & van Loan, 1989). This apparent discrepancy is likely due to differences in the solution algorithms employed in these studies.

5. Conclusions

This study aimed to evaluate the ambient water quality monitoring stations located along the main stem of the Harmu River and, if necessary, refine the network based on scientific evidence. Findings indicated that the efficiency and cost-effectiveness of the monitoring network could be improved by reducing the number of stations from 22 to 10. Such a reduction could yield substantial cost savings without compromising the collection of essential surface water quality data. It should be noted, however, that the analysis was based solely on 3-year annual median values of water quality parameters. Before making any final decision to eliminate stations from the Harmu Mukti Dham network, PCA and PFA analyses should be conducted over a longer time frame (i.e., more than three years), assuming sufficient data are available. In addition, temporal variations—such as seasonal data—should also be examined. These analyses would be particularly valuable if budget constraints necessitate streamlining the existing network. Further research is also recommended to identify the key physical, chemical, and biological parameters most effective in predicting seasonal variations in surface water quality across the Piska Nagari monitoring network. Such work could also assess the potential for reducing the number of parameters routinely measured.

6. References

- Anderson, T.W., Sclove, S.L., 1986. The Statistical Analysis of Data, second ed. The Scientific Press.
- Bengraine, K., Marhaba, T.F., 2003. Using principal component analysis to monitor spatial and temporal changes in water quality. J. Hazard. Mater. 100, 179–195.
- Campbell, D., Bergman, M., Brody, R., Keller, A., Livingston- Way, P., Morris, F., Watkins, B., 1993. SWIM Plan for the Lower St. Johns River Basin. St. River Water Management District, Palatka, Florida.
- Canfield Jr., D.E., Linda, S.B., Hodgson, L.M., 1984. Relations between color and some limnological characteristics of Florida lakes. Water Resour. Bull. 20, 323–329.
- Davis, J.C., 1986. Statistical and data analysis in geology, second ed. John Wiley & Sons, New York.

- Durell, G.S., Seavey, J.A., Higman, J., 2001. Sediment Quality in the Lower St. Johns River and Cedar-Ortega River Basin: Chemical Contaminant Characteristics. Battelle, 397 Washington Street, Duxbury, MA.
- Gangopadhyay, S., Gupta, A.D., Nachabe, M.H., 2001. Evaluation of ground water monitoring network by principal component analysis. Ground Water 39, 181–191.
- Golub, G.H., van Loan, C.F., 1989. Matrix Computations, 2nd Ed. The John Hopkins University Press, Baltimore.
- Haven, K.E., 2003. Phosphorus-Algal bloom relationships in large lakes of south Florida: Implications for establishingnutrient criteria. Lake Reservoir Manage. 9, 222–228.
- Manly, B.F.J., 1986. Multivariate Statistical Methods: A Primer. Chapman & Hall, London.
- Ouyang, Y., Higman, J., Thompson, J., O'Toole, T., Campbell, D., 2002. Characterization and spatial distribution of heavy metals in sediment from cedar and ortega rivers Basin. J. Contam. Hydrol. 54, 19–35.
- Perkins, R.G., Underwood, G.J.C., 2000. Gradients of chlorophyll a and water chemistry alongan eutrophic reservoir with determination of the limitingnutrient by in situ nutrient addition. Water Res. 34, 713–724.
- SAS Institute Inc., 1999. SAS Proprietary Software Release 8.2. SAS Institute, Cary, NC.
- Shine, J.P., Ika, R.V., Ford, T.E., 1995. Multivariate statistical examination of spatial and temporal patterns of heavy metal contamination in New Bedford Harbor marine sediments. Environ. Sci. Technol. 29, 1781–1788.
- Tabachnick, B.G., Fidell, L.S., 2001. UsingMultivariate Statistics. Allyn and Bacon, Boston, London. Tauler, R., Barcelo, D., Thurman, E.M., 2000. Multivariate correlation between concentrations of selected herbicides and derivatives in outflows from selected US Midwestern reservoirs. Environ. Sci. Technol. 34, 3307–3314.
- Wackernagel, H., 1995. Multivariate Geostatistics. An Introduction With Applications. Springer, New York and London.
- Winter, T.C., Mallory, S.E., Allen, T.R., Rosenberry, D.O., 2000. The use of principal component analysis for interpreting ground water hydrographs. Ground Water 38, 234–246.
- Yu, J.C., Quinn, J.T., Dufournaud, C.M., Harrington, J.J., Roger, P.P., Lohani, B.N., 1998. Effective dimensionality of environmental indicators: a principal component analysis with bootstrap confidence intervals. J. Environ. Manage. 53,
  101–119.