JACOBS PUBLISHERS

OPEN ACCESS

Jacobs Journal of Civil Engineering

Research Article

Factor analysis to assess pollutant source apportionment and to investigate the relationship between catchment attributes and instream water quality

A. Azzellino[1*]

[1]D.I.C.A - Environmental Section - Politecnico di Milano, Piazza Leonardo da Vinci, 32 - 20133 Milano, Italy

*Corresponding author: Dr. A. Azzellino, D.I.C.A - Environmental Section - Politecnico di Milano, Piazza Leonardo da Vinci, 32 - 20133 Milano, Italy,Tel: +390223996431; Email: arianna.azzellino@polimi.it

Received: 08-10-2015

Accepted: 10-30-2015

Published: 11-12-2015

Abstract

The EU Water Framework Directive (WFD EC60/2000) requires that quality-flow compliance at a particular surface-water reach entail consideration of all upstream inputs, including contaminated land and groundwater contributions. Multivariate statistical techniques may improve our understanding of the pollutant sources affecting river quality. Aim of this study is to analyze the source apportionment and the groundwater contribution to the total pollutant load of Mella river. Factor Analysis (FA) was applied to a series of water quality measurements at seven monitoring sites, located upstream, in the middle and downstream the groundwater recharge area of the Mella river watershed. FA results in the upstream sites were completely different from the lowland stations that were strongly influenced by the groundwater contribution. In the upstream sites, in fact, the major pollutant source resulted to be the contribution of the Gobbia tributary which collects the industrial loads of the Val Trompia metallurgic consortium. On the other hand the groundwater was found to be the most significant pollutant source in the lowland sites. FA proved also useful to distinguish between sources of metals and chlorinated solvents.

Keywords: Factor Analysis; Groundwater Interactions With Surface Waters; Source Apportionment; Macro and Micropollutants

Introduction

The European Water Framework Directive (WFD: 2000/60/CE) defines a new logic in surface- and ground- water quality management for the European Union, promoting a river basin-approach rather than a local scale approach, stimulating a more integrated approach to mitigate and manage pollution at watershed scale. This is the reason why the analysis of water quality requires today more complex investigations and the identification of all the emission sources affecting water quality at catchment levels. The monitoring and quantification of point and nonpoint sources contributions to the global pollutant load is therefore a key issue for the implementation of management strategies. Diffuse loads may be significant either in wet-weather conditions (i.e. pollutants carried by surface runoff) or in dry-weather conditions (i.e. pollutants carried by subsurface runoff or due to the groundwater exchanges with surface waters). In this respect, the understanding of the interactions between surface waters and ground waters may be the basis for effective water resource management [1]. To fully understand the interactions between surface waters and ground waters, a sound and robust monitoring of surface- and ground water quality data is required. Instream measurements, that are very often instantaneous, can provide information about the total loads in a specific watershed, but do not provide insights about the source apportionment of pollutants, if they are not integrated with other investigative tools, such as mathematical models or statistical techniques. Although experiences are reported concerning the source apportionment of micropollutants (e.g. [2-11], most of the available literature concerns the monitoring either at the emission source [8] or in water bodies [12] of these substances. On the other side, many conceptual models have been developed for the surface-groundwater system [13-14]. Nevertheless, the effect of the groundwater interactions on surface water quality is hard to quantify. Multivariate statistical techniques (e.g. Factor Analysis) may support the understanding of these interactions at watershed scale [18, 19]. Aim of this study has been to apply Factor Analysis (hereinafter FA) to series of water quality measurements collected at seven monitoring sites, located within the Mella river watershed (Figure 1), in order to assess the source apportionment of different macro- and micropollutants.

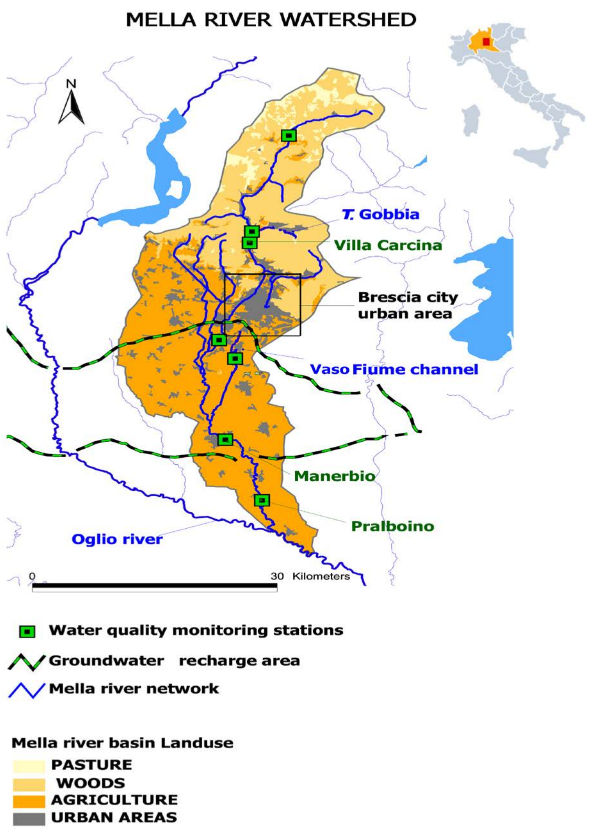


Figure 1. Mella river watershed study area. The water quality monitoring stations and the ground water recharge area are shown.

Materials And Methods

Study Area

The Mella river watershed is a basin of 1022 km^2 located in Northern Italy. The catchment is characterized in the upland portion by streams with relatively steep slopes and deep-incised channels, flowing across a largely forested region with some agricultural areas. On the contrary in the lowland region the Mella river passes through an urban area (Brescia city, about 200,000 inhabitants) and, downstream, through a very productive agricultural region (see Figure 1) characterized by the groundwater recharge area. At the basin closure, the Mella river mean flowrate is of about 30-40 m^3s^{-1}. In this study the effect of land use was also taken into consideration.

Land use data were extracted from DUSAF archive (ERSAF Lombardia, downloadable as open data from the Regione Lombardia geoportal, www.cartografia.regione.lombardia.it).

Input Data and treatment prior Factor Analysis

FA was applied to the instream measurements of several water quality parameters, analyzed both during dry and wet weather conditions (see Table 1).

All the measurements derive from to the monthly monitoring activity, conducted by ARPA, the Italian Regional Environmental Protection Agency, during the period 2001-2007 at 7 sampling stations within the Mella river watershed: 5 stations located on the Mella river main course and 2 stations located on the tributaries more significant in terms of polluting loads. Details about the used analytical methods can be found in the [20] which is the standard reference guide for the Italian Regional Environmental Protection Agencies. The stations located on the Mella river main course are the following: *Bovegno,* located in the upland region at about 10 km from the headwater, *Villa Carcina,* located in the upland region at about 33 km from the headwater, *Castelmella* located at about 50 km from the headwater, *Manerbio* located at about 70 km from the headwater, the latter two stations both in the watershed portion where the water table is at surface and constitutes a groundwater recharge area, and *Pralboino* located at 87 km from the headwater at few km to the river confluence into the Oglio river. The tributary stations are *Sarezzo* on Gobbia creek which carries the pollutant load of the Val Trompia area, one of the major metallurgic consortium of industries in Northern Italy, also characterized by several raw civil wastewater discharges, and *Flero* located on the Vaso Fiume channel which carries the load of the Brescia city urban area. The frequency of detection of all retained constituents was generally larger than 70 percent. The exceptions were the micropollutants such as the solvents and the metals for which nondetects may constitute more the 80 percent of the values. These constituents were retained in the analysis because of their importance to environmental processes of concern. The detection limit was assumed as substitution method for nondetects. The listwise deletion criterion was set as default to manage missing data while performing Factor analysis.

Statistical Analysis

FA was performed on the correlation matrix of the measurements [21]. All the statistical computations were made using the statistical package SPSS 22.0. Factor Analysis was obtained through a preliminary Principal Component Analysis (PCA) which extracted the eigenvalues and eigenvectors from the covariance matrix of the original variances. Factor analysis was chosen to reduce the contribution of the less significant parameters within each component, by extracting a new set of varifactors through rotating the axes defined by the PCA extraction. The Varimax rotation criterion was used to rotate the PCA axes allowing to maintain the axes orthogonality. The number of factors to be retained was chosen on the basis of the "eigenvalue higher than 1" criterion (i.e. all the factors that explained less than the variance of one of the original variables were discarded).

HCA was run based on the FA extracted varifactors and the Ward's method was used as clustering method.

where *i* and *j* refer to a couple of stations, and *k* to the considered parameters.

Results

FA was applied to the whole data set and to the seven subsets of data deriving from the specific monitoring station. The different extractions of varifactors are summarized in Table 2. As it can be observed, although with different number of varifactors all the performed analysis explained approximately the same amount of variance.

Table 1. Summary statistics of the water quality measurements available for the seven monitoring stations.

	Abbreviations	Mean	Median	Std. Dev.	Minimum	Maximum	N
Streamflow (m^3 s^{-1})	Q	10.7	3.85	17.9	0.0	278.0	578
Dissolved Oxygen (mg l^{-1})	DO	8.9	9.4	2.9	0.0	15.9	578
DOsat (% saturation)	% DOsat	84.0	91	23.6	0.0	132.0	578
Biochemical Oxygen Demand (mg l^{-1})	BOD	3.5	2.9	2.6	0.3	34.3	565
Chemical Oxygen Demand (mg l^{-1})	COD	14.5	11.1	28.3	0.7	586.0	570
Escherichia coli - u.f.c.	E.coli	84367.0	54220	92792.4	0.0	1620900.0	584
N-NH4 (mg l^{-1})	N-NH4	0.9	0.535	1.1	0.0	7.7	570
N-NO3 (mg l^{-1})	N-NO3	3.3	1.99	2.4	0.3	11.1	569
Total phosphorus (mg l^{-1})	TP	0.4	0.275	0.7	0.0	10.5	570
Orthophosphate (mg l^{-1})	P-PO4	0.2	0.12	0.2	0.0	2.1	569
Total Nitrogen (mg l^{-1})	TN	5.6	4.952	4.7	0.5	60.3	518
pH	pH	7.6	7.52	0.5	6.4	9.7	578
Temperature (°C)	T°C	12.4	12.2	5.3	0.6	26.8	546
Conductivity (μS cm^{-1})	EC	523.2	510	210.9	3.4	1670.0	570
Hardness (mg l^{-1})	H	254.3	254	91.9	69.0	620.0	570
Total Suspended Solids (mg l^{-1})	TSS	13.3	4.925	42.4	0.2	800.0	570
Chlorides (mg l^{-1})	Cl^-	25.2	21.6	20.4	0.5	192.4	569
Sulfates (mg l^{-1})	SO_4^{-2}	49.7	52	18.1	0.0	110.0	569
Trichloromethane (μg l^{-1})	TCM	0.7	0.5	2.2	0.1	33.0	257
Trichloroethylene (μg l^{-1})	TCE	0.5	0.5	0.2	0.1	1.1	257
Tetrachloroethylene (μg l^{-1})	PCE	0.8	0.5	0.9	0.1	6.0	257
Iron (μg l^{-1})	Fe	97.3	47.55	128.2	4.1	958.0	166
Copper (μg l^{-1})	Cu	27.7	7	108.5	0.6	1336.0	288
Zinc (μg l^{-1})	Zn	81.1	31	285.8	1.0	3687.0	288
Cadmium (μg l^{-1})	Cd	1.6	1	1.7	0.0	5.0	295
Total Chromium (μg l^{-1})	Cr	22.8	5	102.9	0.1	1570.0	295
Mercury (μg l^{-1})	Hg	0.4	0.5	0.2	0.1	0.5	295
Nickel (μg l^{-1})	Ni	35.5	12.73	59.2	0.1	499.2	199
Lead (μg l^{-1})	Pb	7.7	5	16.8	0.0	184	211

That allowed to select few factors able to describe the whole data set with minimum loss of original information. Moreover, a Hierarchical Cluster Analysis (HCA, according to [21]) was used to analyze the similarities among the water quality profiles at the different monitoring stations, using the Euclidean Distance as distance metric (see Eq. 1).

$$d_2(\mathbf{x}_i, \mathbf{x}_j) = \sqrt[2]{\sum_{k=1}^{q} (\mathbf{x}_{ik} - \mathbf{x}_{jk})^2} \qquad \text{Eq.1}$$

Table 2. Summary of the FA extractions

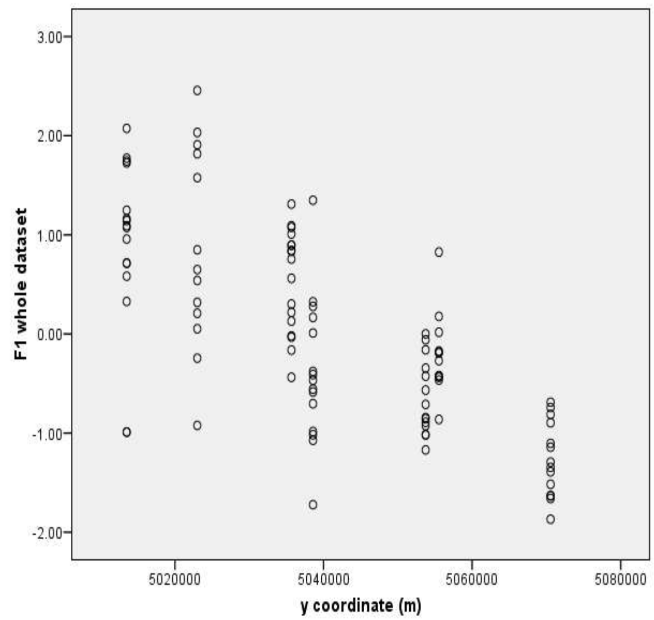
Monitoring station	Number of varifactors	Total explained variance	Sample size
Whole dataset	9	73.3%	101
	Mella river stations		
Bovegno	8	91.8%	14
Villa Carcina	8	88.9%	14
Castelmella	8	87.9%	15
Manerbio	7	87.7%	13
Pralboino	7	88.9%	16
	Mella tributaries		
Sarezzo – Gobbia creek	7	91.5%	12
Flero – Vaso Fiume	9	90.2%	17

By looking at the factor loadings matrix (i.e. the list of the correlation coefficients of the original variables with the extracted varifactors see Tables 3 to 10) it is possible to identify the most meaningful parameters within each component (i.e. factor loadings higher than 0.5 and lower than -0.5). It is also worthwhile to remind that parameters lying on the same varifactor may be reasonably attributed to the same origin.

Whole Watershed data set

As shown in Table 3 the first varifactor concerning the analysis of the whole dataset carries about 15 % of the whole information and is loaded by nitrates, conductivity, hardness and chlorides which is particularly relevant for this watershed due to the groundwater recharge to the surface water body which affects the lowland portion of the basin. Figure 2 shows the pattern of this first varifactor versus latitude: it can be observed how the varifactor increases inversely with latitude showing that the highest values in terms of nitrates, chlorides, conductivity and hardness correspond to the lowland stations. The second extracted varifactor explains about 12% of the total variance and accounts for pollutants which are typical of raw domestic wastewaters (i.e. N-NH4) and is inversely correlated with dissolved oxygen suggesting that the untreated wastewaters in this watershed are still the dominant controlling factor of the river oxygenation. The third varifactor accounts for the heavy metals (i.e. Cu, Ni and Cr) and explains a little less than 9% of the variance outlining how relevant is the impact of the metallurgic consortium of industries present in the watershed on the river water quality. The fourth varifactor accounts for COD, total phosphorus and suspended material and explains about 8% of the variance and is probably representative of the rainfall-driven pollution; the fifth varifactor is loaded by trichloroethylene and tetrachloroethylene and it explains 7.5% of the total variance. However it should be observed that in the used data series the detection limit threshold has changed in time from < 0.5 to <0.1 $\mu g\ l^{-1}$ which reasonably inflates the variance component of this varifactor.

Figure 2. Latitudinal pattern of the first varifactor extracted from the whole data set: it can be observed the inverse correlation with latitude which in this particular watershed oriented North to South shows how the highest scores of this varifactor occur in the lowland portion of the basin.



Besides the difference of detection limit the fourth varifactor being loaded also by pH suggests an inverse correlation of these parameters which is confirmed by their correlation coefficients (e.g. r: -0.23 and r: -0.26, P<0.001 n: 251; respectively for pH and tetrachloroethylene and pH and trichloroethylene). It is known that pH may affect the degradation of these pollutants [22] but in this dataset pH it appears to be a sort of tracer of the cultivated vs the uncultivated land uses (i.e. pH is inversely correlated with agricultural land use and directly correlated with grasslands and natural vegetation).

The seventh, eighth and ninth varifactors respectively account for BOD, Escherichia coli, trichloromethane and zinc variabilities and explain about 5% of the total variance.

Bovegno FA

As shown in Table 4 the first varifactor for the *Bovegno* subset is able alone to explain about 20% of the total variance and is loaded by orthophosphate, pH trichloromethane, trichloroethylene and tetrachloroethylene. However, it must be observed that the already mentioned change of detection limit of the micropollutants reasonably inflates its variance component. That is especially true for this monitoring station where all the value of the chlorinated solvents are nondetects. These micropolluntants were maintained in the analysis of this subset only to keep the homogeneity of parameters among the different FAs. The second and third varifactors explain roughly the same percent variance and respectively account for COD, TSS, zinc and conductivity, hardness and sulfates. The fourth varifactor explain about 10% of the total variance and accounts for the variability of some metals (e.g. Cd, Cu and Ni) showing that metal pollution in this watershed is relevant even at the upland sites. The fifth varifactor explains about 9% for the total variance and is loaded by streamflow and total phosphorus outlining the relationship of TP with high flow transport dynamics. The sixth varifactor explains little less than 9% and is loaded by ammoniacal nitrogen.

Seventh and eighth varifactors explain approximately the same percent of variance and respectively account for nitrates and chlorides, BOD and dissolved oxygen. It's interesting to observe that nitrates and chlorides lie on the same varifactor also in this subset although values are much lower than the values of the lowland stations.

Table 3. Factor loadings matrix of the whole data set. The factor loadings higher than 0.5 and lower than –0.5 are shown in bold.

	1	2	3	4	5	6	7	8	9
Streamflow - $m^3 s^{-1}$	-0.089	0.040	-0.152	0.008	0.046	**-0.783**	0.140	-0.078	-0.088
Dissolved Oxygen ($mg\ l^{-1}$)	-0.341	**-0.755**	0.145	-0.326	-0.054	0.040	0.299	0.019	-0.076
Dissolved Oxygen - % sat.	-0.194	**-0.854**	0.136	-0.218	-0.158	0.046	0.130	-0.062	-0.054
BOD_5 ($mg\ l^{-1}$)	0.091	-0.091	0.098	0.122	-0.055	-0.114	**0.723**	-0.201	0.031
COD ($mg\ l^{-1}$)	0.116	0.321	0.003	**0.556**	0.105	-0.001	0.418	0.258	0.003
Escherichia coli - u.f.c.	0.022	0.012	0.047	0.126	0.049	0.009	-0.152	**0.814**	0.133
N-NH4 ($mg\ l^{-1}$)	0.125	**0.611**	0.240	0.114	0.306	0.377	0.275	-0.184	0.066
N-NO3 ($mg\ l^{-1}$)	**0.858**	0.087	-0.139	-0.068	-0.023	-0.251	0.092	0.087	-0.007
Total phosphorus ($mg\ l^{-1}$)	0.055	0.067	-0.002	**0.755**	0.178	0.252	-0.112	-0.066	-0.084
Orthophosphate ($mg\ l^{-1}$)	0.306	0.416	-0.149	0.335	0.198	0.439	0.255	-0.072	-0.063
Total Nitrogen ($mg\ l^{-1}$)	**0.515**	**0.647**	0.070	-0.161	0.076	-0.128	0.097	0.090	-0.071
pH	-0.256	-0.175	0.068	-0.131	**-0.568**	0.165	-0.176	-0.224	0.206
T - °C	0.492	0.149	-0.115	0.440	-0.218	0.021	-0.450	-0.252	0.054
Conductivity ($\mu S\ cm^{-1}$)	**0.811**	0.283	-0.052	0.055	0.160	0.220	0.077	-0.129	0.083
Hardness ($mg\ l^{-1}$)	**0.819**	0.075	0.050	0.156	-0.001	0.142	-0.128	-0.004	0.082
Suspended material ($mg\ l^{-1}$)	-0.004	0.072	-0.166	**0.683**	0.015	-0.171	0.188	0.150	0.072
Cl- ($mg\ l^{-1}$)	**0.732**	0.295	0.010	0.129	0.225	0.194	0.134	0.193	0.175
SO4-2 ($mg\ l^{-1}$)	0.495	0.109	-0.340	-0.126	0.072	0.228	0.051	0.478	-0.100
Trichloromethane ($\mu g\ l^{-1}$)	0.282	-0.089	-0.130	0.051	0.236	-0.058	-0.218	-0.064	**0.728**
Trichloroethylene ($\mu g\ l^{-1}$)	-0.079	0.005	0.214	0.085	**0.822**	-0.028	-0.027	-0.036	0.211
Tetrachloroethylene ($\mu g\ l^{-1}$)	0.234	0.332	-0.190	0.084	**0.637**	0.302	-0.143	0.056	0.024
Cu ($\mu g\ l^{-1}$)	-0.107	-0.060	**0.767**	-0.043	0.093	0.082	0.114	-0.033	0.125
Zn ($\mu g\ l^{-1}$)	-0.018	0.165	0.147	-0.056	-0.076	0.181	0.266	0.249	**0.749**
Cd ($\mu g\ l^{-1}$)	0.055	-0.536	0.304	0.402	0.333	-0.159	-0.103	-0.234	0.057
Total Cr ($\mu g\ l^{-1}$)	0.030	-0.042	**0.683**	-0.070	0.066	-0.054	-0.023	0.065	-0.150
Ni ($\mu g\ l^{-1}$)	-0.060	-0.027	**0.830**	-0.040	-0.094	0.141	0.050	-0.044	0.056
Eigenvalue	**3.9**	**3.1**	**2.3**	**2.1**	**1.9**	**1.5**	**1.5**	**1.4**	**1.3**
% of Variance	**14.9**	**11.8**	**8.9**	**8.3**	**7.5**	**5.8**	**5.7**	**5.3**	**5.2**
Cumulative % of Variance	**14.9**	**26.7**	**35.6**	**43.8**	**51.3**	**57.1**	**62.8**	**68.2**	**73.4**

Table 4. Factor loadings matrix of the *Bovegno* subset. The factor loadings higher than 0.5 and lower than –0.5 are shown in bold.

	1	2	3	4	5	6	7	8
Streamflow ($m^3 s^{-1}$)	0.137	-0.063	-0.032	0.111	**0.859**	-0.08	-0.348	-0.074
Dissolved Oxygen ($mg\ l^{-1}$)	0.48	0.485	-0.326	-0.126	-0.184	-0.424	-0.06	0.408
Dissolved Oxygen - % sat.	0.398	0.151	-0.03	0.091	0.242	-0.336	0.191	**0.68**
BOD_5 ($mg\ l^{-1}$)	0.248	-0.041	-0.205	0.144	-0.078	0.062	-0.079	**0.832**
COD ($mg\ l^{-1}$)	0.175	**0.852**	-0.059	-0.121	0.369	0.129	-0.139	0.195
Escherichia coli - u.f.c.	-0.375	0.2	**0.58**	**-0.516**	0.013	0.101	0.092	-0.304
N-NH4 ($mg\ l^{-1}$)	0.039	0.114	0.179	-0.107	-0.086	**0.892**	0.24	-0.105
N-NO3 ($mg\ l^{-1}$)	0.054	-0.065	-0.137	-0.004	-0.131	0.222	**0.889**	-0.09
Total phosphorus ($mg\ l^{-1}$)	0.243	-0.077	0.243	0.374	**0.741**	0.184	0.253	0.012
Orthophosphate ($mg\ l^{-1}$)	**-0.798**	0.113	0.149	0.203	-0.212	0.202	-0.062	-0.111
Total Nitrogen ($mg\ l^{-1}$)	0.163	-0.019	-0.232	**-0.922**	-0.122	0.177	-0.032	-0.1
pH	**-0.597**	-0.216	0.046	0.062	**-0.691**	0.149	-0.216	-0.153
T - °C	-0.335	**-0.574**	0.417	0.24	0.3	0.39	0.124	-0.045
Conductivity ($\mu S\ cm^{-1}$)	-0.109	0.002	**0.946**	0.092	0.089	0.089	0.184	-0.099
Hardness ($mg\ l^{-1}$)	0.138	-0.169	**0.708**	0.098	-0.002	0.271	0.017	**0.553**
Suspended material ($mg\ l^{-1}$)	-0.013	**0.967**	0.032	0.069	-0.034	0.168	-0.013	-0.032
Cl^- ($mg\ l^{-1}$)	0.144	-0.052	0.138	0.122	0.089	0.017	**0.913**	0.092
SO_4^{-2} ($mg\ l^{-1}$))	-0.293	0.105	**0.847**	-0.025	-0.055	-0.071	-0.296	-0.225
Trichloromethane ($\mu g\ l^{-1}$)	**0.631**	0.091	-0.069	0.069	0.066	**0.546**	-0.006	0.194
Trichloroethylene ($\mu g\ l^{-1}$)	**0.898**	0.145	-0.057	-0.047	0.127	0.1	0.047	0.253
Tetrachloroethylene ($\mu g\ l^{-1}$)	**0.898**	0.145	-0.057	-0.047	0.127	0.1	0.047	0.253
Cu ($\mu g\ l^{-1}$)	**0.737**	0.113	-0.188	**0.529**	0.056	-0.069	0.127	-0.097
Zn ($\mu g\ l^{-1}$)	0.112	**0.949**	0.114	0.154	-0.096	-0.056	0.026	-0.107
Cd ($\mu g\ l^{-1}$)	0.027	0.11	-0.095	**0.722**	0.242	**0.546**	0.126	0.204
Total Cr ($\mu g\ l^{-1}$)	0.442	0.234	-0.426	0.118	0.426	0.27	0.244	0.308
Ni ($\mu g\ l^{-1}$)	**0.725**	0.108	-0.171	**0.565**	0.102	0.091	0.085	0.002
Eigenvalue	**5.3**	**3.5**	**3.3**	**2.6**	**2.5**	**2.3**	**2.2**	**2.2**
% of Variance	**20.2**	**13.3**	**12.6**	**10.2**	**9.6**	**8.8**	**8.6**	**8.5**
Cumulative % of Variance	**20.2**	**33.6**	**46.1**	**56.3**	**65.9**	**74.6**	**83.2**	**91.8**

Villa Carcina and Sarezzo FA

Villa Carcina is located immediately downstream the Gobbia tributary confluence. As already mentioned, Gobbia river is heavily polluted since it carries the pollutant load of the Val Trompia area, characterized by several raw civil wastewater discharges and by the pollutant load of one of the major metallurgic consortium of industries in Northern Italy. It is therefore not surprising that the varifactors extracted from this subset differ quite markedly from the varifactors deriving from Bovegno site (Table 5).

Although explaining roughly the same percent variance (i.e. 19.5%) than the first varifactor of Bovegno subset, the first varifactor extracted from the Villa Carcina subset is loaded by streamflow, conductivity, total chromium and nickel. Particularly the sign of the factor loadings of these parameter suggests an inverse relationship of conductivity, total chromium and nickel with streamflow.

Table 5. Factor loadings matrix of the *Villa Carcina* subset. The factor loadings higher than 0.5 and lower than –0.5 are shown in bold.

	1	2	3	4	5	6	7	8
Streamflow (m^3 s^{-1})	**-0.892**	-0.127	-0.003	0.314	-0.086	-0.101	0.064	-0.055
Dissolved Oxygen (mg l^{-1})	-0.479	-0.440	0.489	0.453	0.223	0.027	-0.115	-0.021
Dissolved Oxygen - % sat.	-0.023	-0.443	0.368	**0.565**	0.372	-0.158	0.231	0.166
BOD_5 (mg l^{-1})	-0.322	**0.756**	-0.115	0.279	0.307	-0.142	0.123	-0.078
COD (mg l^{-1})	0.101	**0.917**	0.020	-0.064	-0.156	0.133	-0.096	-0.116
Escherichia coli - u.f.c.	-0.187	-0.211	0.430	0.052	**0.698**	0.302	0.056	0.150
N-NH4 (mg l^{-1})	0.224	**0.821**	0.083	-0.194	-0.082	-0.047	-0.146	0.313
N-NO3 (mg l^{-1})	-0.443	-0.220	-0.111	0.037	0.002	0.026	0.089	-0.754
Total phosphorus (mg l^{-1})	0.396	-0.243	-0.218	0.199	0.304	-0.072	**0.648**	0.138
Orthophosphate (mg l^{-1})	0.291	**0.723**	-0.045	-0.060	**-0.580**	-0.100	0.087	0.037
Total Nitrogen (mg l^{-1})	-0.019	-0.042	0.008	-0.205	0.208	-0.092	**-0.864**	0.029
pH	-0.068	-0.028	-0.221	0.026	**0.866**	-0.214	-0.165	0.005
T - °C	**0.566**	0.453	-0.389	-0.122	-0.134	-0.296	0.425	0.060
Conductivity (μS cm^{-1})	**0.896**	0.130	0.273	-0.096	-0.123	0.171	0.059	0.031
Hardness (mg l^{-1})	**0.907**	0.056	0.000	0.169	-0.032	0.204	0.096	**0.029**
Suspended material (mg l^{-1})	-0.214	0.306	-0.439	0.389	-0.498	-0.180	-0.445	-0.002
Cl^- (mg l^{-1})	0.246	0.099	**0.688**	-0.075	0.115	0.018	-0.419	-0.138
SO_4^{-2} (mg l^{-1}))	0.174	-0.015	0.146	**-0.898**	-0.050	-0.110	-0.138	0.047
Trichloromethane (μg l^{-1})	-0.178	-0.117	-0.210	0.311	0.108	-0.096	0.144	**0.733**
Trichloroethylene (μg l^{-1})	0.143	-0.110	0.138	0.323	0.061	**0.895**	-0.029	0.116
Tetrachloroethylene (μg l^{-1})	-0.063	0.103	-0.100	-0.270	-0.039	**0.865**	0.060	-0.243
Cu (μg l^{-1})	0.312	-0.065	**0.626**	-0.084	-0.229	**0.589**	0.108	-0.035
Zn (μg l^{-1})	0.023	-0.009	**0.912**	-0.074	0.004	0.014	-0.011	-0.003
Cd (μg l^{-1})	0.253	-0.058	-0.076	**0.773**	-0.098	-0.073	0.144	0.346
Total Cr (μg l^{-1})	**0.853**	-0.035	0.135	0.130	-0.121	-0.216	0.064	0.059
Ni (μg l^{-1})	**0.608**	-0.033	**0.618**	-0.023	-0.373	-0.068	0.066	0.009
Eigenvalue	**5.1**	**3.5**	**3.3**	**2.8**	**2.6**	**2.4**	**2.0**	**1.5**
% of Variance	**19.5**	**13.6**	**12.7**	**10.6**	**9.8**	**9.2**	**7.6**	**5.9**
Cumulative % of Variance	**19.5**	**33.1**	**45.8**	**56.4**	**66.3**	**75.5**	**83.1**	**89.0**

The second varifactor explains roughly the 13% of the total variance and accounts for pollutants which are typical of the untreated domestic wastewater (i.e. BOD, COD, ammonia and orthophosphate) which, in this watershed, are still far from being insignificant. The third varifactor explains little less than 13% and is loaded by chlorides, copper, zinc and nickel. The fourth varifactor explains 10% of the total variance and is loaded by sulfates and cadmium. The fifth varifactor explains little less than 10% of the total variance and is loaded by pH, and *Escherichia coli*. The sixth varifactor explains about 9% of the total variance and is loaded by trichloroethylene, tetrachloroethylene and copper.

The seventh and eighth varifactors explain respectively 7.6% and 6% of the total variance and are loaded by total phosphorus and nitrogen, nitrates and trichloromethane. So in this subset nitrates and chlorides load different varifactors suggesting the existence of sources other than the groundwater for both pollutants. Besides the obvious differences observed with respect to the Bovegno subset it is interesting to compare the Villa Carcina FA results with the FA results of the Sarezzo subset (Table 6).

Table 6. Factor loadings matrix of the *Sarezzo* subset. The factor loadings higher than 0.5 and lower than −0.5 are shown in bold.

	1	**2**	**3**	**4**	**5**	**6**	**7**
Streamflow (m^3 s^{-1})	-0.335	**0.510**	-0.066	0.058	0.097	**-0.736**	-0.007
Dissolved Oxygen (mg l^{-1})	0.058	**0.743**	-0.081	-0.116	**0.599**	-0.089	-0.085
Dissolved Oxygen - % sat.	0.269	0.077	0.105	-0.181	**0.865**	0.016	-0.170
BOD_5 (mg l^{-1})	0.159	-0.112	0.458	-0.056	**0.606**	0.160	0.465
COD (mg l^{-1})	0.364	0.199	-0.111	0.233	-0.165	**0.633**	-0.091
Escherichia coli - u.f.c.	-0.189	**0.694**	-0.499	0.152	-0.020	0.227	0.002
N-NH4 (mg l^{-1})	0.011	0.337	0.234	0.477	0.002	0.247	**0.668**
N-NO3 (mg l^{-1})	-0.012	0.142	-0.029	-0.254	-0.039	**-0.920**	-0.093
Total phosphorus (mg l^{-1})	0.236	**-0.818**	-0.152	0.016	0.262	0.092	0.074
Orthophosphate (mg l^{-1})	0.252	0.475	-0.133	-0.103	0.153	0.080	**-0.759**
Total Nitrogen (mg l^{-1})	0.151	0.329	0.016	**0.716**	-0.484	-0.014	0.315
pH	-0.044	0.189	**-0.640**	0.307	-0.208	-0.102	**0.566**
T - °C	0.223	**-0.911**	-0.010	-0.125	0.058	0.146	-0.089
Conductivity (μS cm^{-1})	0.005	-0.157	0.121	**0.918**	0.056	0.295	-0.014
Hardness (mg l^{-1})	0.044	-0.215	-0.254	**0.882**	0.002	0.114	0.230
Suspended material (mg l^{-1})	0.431	**0.559**	-0.400	0.033	0.029	-0.275	-0.191
Cl^- (mg l^{-1})	0.168	0.361	0.016	**0.892**	-0.097	0.011	0.032
SO_4^{-2} (mg l^{-1}))	-0.206	0.170	**0.659**	**0.517**	-0.314	0.082	-0.281
Trichloromethane (μg l^{-1})	**0.979**	-0.074	0.107	0.052	0.047	0.106	-0.013
Trichloroethylene (μg l^{-1})	**0.979**	-0.074	0.107	0.052	0.047	0.106	-0.013
Tetrachloroethylene (μg l^{-1})	**0.931**	-0.049	0.313	0.041	0.044	0.073	0.056
Cu (μg l^{-1})	0.120	0.191	0.394	-0.099	**-0.741**	0.273	0.014
Zn (μg l^{-1})	**-0.908**	0.171	0.333	-0.018	-0.102	-0.027	0.063
Cd (μg l^{-1})	**0.971**	-0.094	-0.115	0.060	0.047	0.134	-0.085
Total Cr (μg l^{-1})	0.084	-0.082	**0.901**	0.026	0.016	0.023	0.184
Ni (μg l^{-1})	-0.004	-0.015	**0.964**	-0.009	-0.048	-0.072	0.085
Eigenvalue	**5.4**	**4.0**	**3.8**	**3.8**	**2.6**	**2.3**	**1.9**
% of Variance	**20.8**	**15.3**	**14.7**	**14.5**	**9.9**	**8.8**	**7.5**
Cumulative % of Variance	**20.8**	**36.1**	**50.8**	**65.3**	**75.2**	**84.0**	**91.5**

As Table 6 clearly shows the Gobbia river at Sarezzo station is dominated by effluents, both domestic and metallurgic. Although the varifactors are not exactly the same than the ones of the Villa Carcina subset, they present similar associations between pollutants loading the same varifactor (e.g. total chromium and nickel, trichloroethylene and tetrachloroethylene, ammonia and orthophosphate).

Castelmella and Manerbio FA

These two stations are in the watershed portion where the water table is at surface and it constitutes a recharge area for the surface water body. They present similar characteristics as shown by FA results (see Table 7 and Table 8).

Table 7. Factor loadings matrix of the *Castelmella* subset. The factor loadings higher than 0.5 and lower than –0.5 are shown in bold.

	1	2	3	4	5	6	7	8
Streamflow (m^3 s^{-1})	**-0.741**	-0.127	0.094	**-0.508**	0.037	0.000	-0.086	-0.227
Dissolved Oxygen (mg l^{-1})	0.194	0.199	-0.055	**0.874**	-0.014	-0.097	-0.081	-0.178
Dissolved Oxygen - % sat.	0.362	0.086	-0.201	**0.793**	-0.095	-0.021	0.001	0.198
BOD_5 (mg l^{-1})	0.053	**0.597**	**-0.575**	-0.057	-0.016	0.205	-0.302	-0.175
COD (mg l^{-1})	0.134	-0.031	-0.259	0.054	0.070	**0.888**	0.065	0.038
Escherichia coli - u.f.c.	0.007	0.423	0.236	-0.381	-0.367	0.024	**0.525**	0.097
N-NH4 (mg l^{-1})	0.209	0.336	0.299	0.180	0.055	**0.655**	0.278	-0.321
N-NO3 (mg l^{-1})	0.092	-0.075	0.093	0.037	**0.934**	-0.051	-0.112	0.004
Total phosphorus (mg l^{-1})	0.048	-0.021	0.356	-0.134	-0.133	**0.796**	-0.076	0.323
Orthophosphate (mg l^{-1})	**0.839**	-0.112	0.243	0.046	-0.160	-0.232	0.050	-0.272
Total Nitrogen (mg l^{-1})	-0.016	-0.003	0.086	-0.153	**0.937**	-0.005	0.052	-0.090
pH	-0.216	-0.238	-0.438	-0.371	0.068	-0.281	0.364	0.386
T - °C	0.260	-0.333	-0.136	-0.352	-0.141	0.207	0.046	**0.753**
Conductivity (µS cm^{-1})	**0.854**	0.147	-0.071	0.075	-0.159	0.179	0.232	0.176
Hardness (mg l^{-1})	0.118	0.330	0.293	0.219	-0.082	-0.041	0.099	**0.786**
Suspended material (mg l^{-1})	-0.401	-0.038	-0.121	-0.237	-0.066	**0.699**	-0.252	-0.110
Cl^- (mg l^{-1})	**0.768**	0.034	0.087	0.038	0.423	0.142	0.248	0.171
SO_4^{-2} (mg l^{-1})	**0.742**	0.234	0.002	0.398	0.407	0.086	-0.049	0.064
Trichloromethane (µg l^{-1})	0.408	-0.226	0.159	-0.092	0.001	-0.169	**0.774**	0.232
Trichloroethylene (µg l^{-1})	0.081	0.161	**0.836**	-0.099	0.157	0.036	0.294	0.000
Tetrachloroethylene (µg l^{-1})	0.120	-0.034	0.026	0.043	0.001	0.057	**0.837**	-0.049
Cu (µg l^{-1})	0.424	**0.717**	0.314	0.173	-0.097	-0.171	-0.236	-0.084
Zn (µg l^{-1})	-0.034	**0.948**	-0.033	0.113	-0.054	-0.058	-0.140	0.208
Cd (µg l^{-1})	0.023	-0.016	**0.917**	-0.095	0.148	-0.016	-0.080	0.063
Total Cr (µg l^{-1})	-0.112	-0.055	0.271	**0.515**	**0.552**	0.010	0.053	-0.170
Ni (µg l^{-1})	0.079	**0.930**	0.040	0.120	0.016	0.125	0.166	-0.090
Eigenvalue	**4.0**	**3.4**	**2.9**	**2.8**	**2.7**	**2.7**	**2.3**	**2.0**
% of Variance	**15.6**	**13.3**	**11.0**	**10.6**	**10.5**	**10.4**	**8.9**	**7.7**
Cumulative % of Variance	**15.6**	**28.8**	**39.8**	**50.4**	**60.9**	**71.3**	**80.2**	**87.9**

Although both the number of extracted varifactors and their composition in terms of factor loadings are not exactly the same, Castelmella and Manerbio FA present a first varifactor loaded by sulfates, chlorides and inversely by streamflow, a second varifactor loaded, among others, by copper, zinc and nickel, a third varifactor loaded by cadmium and trichloroethylene, a varifactor, the sixth concerning the Castelmella subset and the forth concerning Manerbio subset, which is loaded among others by COD and total phosphorus and roughly explain the 10% of the total variance.

On the other hand, Castelmella and Manerbio varifactors are different concerning the following characteristics: nitrates do not correlate with chlorides at Castelmella station while they do at Manerbio station; ammoniacal nitrogen correlates with copper, zinc and nickel and loads their same varifactor at Manerbio station,

Table 8. Factor loadings matrix of the *Manerbio* subset. The factor loadings higher than 0.5 and lower than –0.5 are shown in bold.

	1	2	3	4	5	6	7
Streamflow ($m^3\ s^{-1}$)	**-0.699**	0.120	0.332	0.041	0.130	**-0.518**	-0.120
Dissolved Oxygen ($mg\ l^{-1}$)	-0.422	**0.717**	0.383	0.010	0.120	-0.300	0.129
Dissolved Oxygen - % sat.	-0.284	-0.101	**0.541**	0.206	-0.109	-0.162	0.284
BOD_5 ($mg\ l^{-1}$)	-0.299	0.159	-0.047	-0.057	**0.827**	-0.370	-0.074
COD ($mg\ l^{-1}$)	0.023	-0.163	-0.079	**0.878**	0.316	0.130	-0.041
Escherichia coli - u.f.c.	0.233	0.020	0.068	0.050	-0.006	**0.948**	-0.012
N-NH4 ($mg\ l^{-1}$)	-0.053	**0.879**	-0.182	-0.187	0.296	0.006	0.156
N-NO3 ($mg\ l^{-1}$)	**0.937**	-0.157	-0.088	-0.164	-0.236	0.026	-0.075
Total phosphorus ($mg\ l^{-1}$)	-0.345	-0.109	0.263	**0.820**	-0.113	0.052	0.169
Orthophosphate ($mg\ l^{-1}$)	0.034	0.330	0.093	0.066	-0.069	0.084	**0.921**
Total Nitrogen ($mg\ l^{-1}$)	**0.523**	0.290	-0.234	-0.203	-0.308	0.196	**-0.586**
pH	**-0.802**	0.209	-0.446	0.066	-0.087	-0.127	-0.142
T - °C	0.291	**-0.869**	-0.170	0.165	-0.234	0.148	-0.002
Conductivity ($\mu S\ cm^{-1}$)	0.432	-0.153	-0.090	**-0.815**	0.035	0.082	-0.026
Hardness ($mg\ l^{-1}$)	0.287	-0.487	**-0.553**	-0.113	-0.488	-0.138	0.147
Suspended material ($mg\ l^{-1}$)	-0.294	-0.185	**0.540**	0.047	0.377	-0.409	-0.274
Cl^- ($mg\ l^{-1}$)	**0.889**	-0.193	0.076	-0.259	0.027	0.218	0.056
SO_4^{-2} ($mg\ l^{-1}$)	**0.893**	0.050	-0.250	-0.206	-0.062	0.234	0.016
Trichloromethane ($\mu g\ l^{-1}$)	0.295	-0.556	0.347	-0.409	0.047	0.263	0.021
Trichloroethylene ($\mu g\ l^{-1}$)	-0.019	0.200	**0.861**	0.006	-0.163	0.198	-0.011
Tetrachloroethylene ($\mu g\ l^{-1}$)	**0.713**	-0.156	-0.132	-0.065	-0.164	-0.234	-0.203
Cu ($\mu g\ l^{-1}$)	0.017	**0.578**	0.300	-0.162	-0.088	0.149	0.490
Zn ($\mu g\ l^{-1}$)	0.174	**0.544**	-0.376	0.128	-0.012	0.175	0.001
Cd ($\mu g\ l^{-1}$)	0.109	-0.123	**0.808**	0.089	-0.008	-0.074	0.464
Total Cr ($\mu g\ l^{-1}$)	0.002	0.233	-0.123	0.156	**0.907**	0.180	0.051
Ni ($\mu g\ l^{-1}$)	-0.222	**0.777**	0.023	0.124	0.019	0.543	0.100
Eigenvalue	**4.0**	**3.4**	**2.9**	**2.8**	**2.7**	**2.7**	**2.3**
% of Variance	**15.6**	**13.3**	**11.0**	**10.6**	**10.5**	**10.4**	**8.9**
Cumulative % of Variance	**15.6**	**28.8**	**39.8**	**50.4**	**60.9**	**71.3**	**80.2**

whereas it loads the same varifactor as COD, total phosphorus and suspended solids at Castelmella station; tetrachloroethylene load the first varifactor at Manerbio station whereas it loads the same varifactor of trichloromethane at Castelmella.

It is also worthwhile to compare these varifactors with those extracted from the Flero station on Vaso Fiume, the artificial channel which receives part of the effluent of the wastewater treatment plant serving the industrial city of Brescia (i.e. Verziano WWTP) which flows into Mella river few kilometers upstream the Manerbio station.

Table 9. Factor loadings matrix of the *Flero-Vaso* Fiume subset. The factor loadings higher than 0.5 and lower than –0.5 are shown in bold.

	1	2	3	4	5	6	7	8	9
Streamflow ($m^3 s^{-1}$)	**0.733**	0.315	0.182	-0.112	-0.137	-0.188	-0.028	0.165	0.153
Dissolved Oxygen ($mg l^{-1}$)	**-0.886**	0.020	0.104	0.061	0.170	0.107	-0.184	-0.232	0.027
Dissolved Oxygen - % sat.	**-0.859**	0.124	0.199	-0.006	0.261	-0.026	-0.227	-0.099	0.023
BOD_5 ($mg l^{-1}$)	0.110	-0.108	0.049	-0.116	0.177	0.000	-0.060	-0.204	**0.838**
COD ($mg l^{-1}$)	0.054	-0.051	0.071	0.226	**-0.640**	0.149	0.483	0.474	0.004
Escherichia coli - u.f.c.	-0.113	-0.043	0.177	**0.955**	-0.074	-0.007	0.041	-0.004	-0.024
N-NH4 ($mg l^{-1}$)	0.314	-0.109	-0.234	-0.287	-0.116	0.278	**0.736**	-0.239	0.113
N-NO3 ($mg l^{-1}$)	0.110	0.270	**0.810**	-0.017	0.039	-0.219	-0.168	0.238	-0.152
Total phosphorus ($mg l^{-1}$)	-0.222	-0.010	**-0.640**	0.123	0.286	0.017	0.194	0.483	-0.148
Orthophosphate ($mg l^{-1}$)	-0.215	-0.126	-0.102	-0.451	-0.101	**-0.524**	0.467	0.144	-0.279
Total Nitrogen ($mg l^{-1}$)	0.475	-0.172	-0.105	-0.086	**-0.710**	-0.070	0.271	-0.251	-0.165
pH	0.036	**0.906**	-0.047	-0.161	-0.258	0.040	-0.102	0.031	0.085
T - °C	0.319	0.254	-0.097	-0.348	0.047	-0.398	0.160	**0.585**	-0.081
Conductivity ($\mu S cm^{-1}$)	-0.365	0.045	0.141	**-0.828**	-0.143	0.069	0.169	0.082	0.007
Hardness ($mg l^{-1}$)	-0.431	0.177	-0.247	0.216	0.084	-0.128	0.086	0.266	**0.694**
Suspended material ($mg l^{-1}$)	0.126	-0.094	0.117	-0.069	0.164	-0.052	-0.060	**0.865**	-0.043
Cl^- ($mg l^{-1}$)	-0.007	0.057	**0.789**	0.172	0.407	-0.070	0.317	0.016	0.028
SO_4^{-2} ($mg l^{-1}$)	-0.466	-0.126	**0.787**	0.114	-0.276	-0.128	0.007	0.021	-0.029
Trichloromethane ($\mu g l^{-1}$)	0.185	**0.930**	-0.021	0.045	0.164	0.013	-0.094	0.061	-0.038
Trichloroethylene ($\mu g l^{-1}$)	-0.002	0.115	-0.082	-0.031	-0.036	**0.929**	0.092	-0.077	0.070
Tetrachloroethylene ($\mu g l^{-1}$)	-0.059	-0.121	-0.314	-0.077	0.086	**0.747**	-0.125	0.011	-0.282
Cu ($\mu g l^{-1}$)	**0.759**	0.265	0.081	0.320	0.219	0.302	0.179	-0.126	-0.082
Zn ($\mu g l^{-1}$)	0.037	**0.892**	0.271	0.079	-0.024	-0.003	0.237	-0.148	-0.070
Cd ($\mu g l^{-1}$)	-0.093	-0.231	-0.060	0.078	**0.892**	0.075	-0.052	0.252	0.191
Total Cr ($\mu g l^{-1}$)	**0.768**	0.235	0.095	0.447	0.041	0.039	-0.007	-0.121	-0.064
Ni ($\mu g l^{-1}$)	0.330	0.115	0.150	0.020	-0.176	-0.229	**0.788**	0.140	-0.063
Eigenvalue	**4.5**	**3.1**	**2.8**	**2.6**	**2.5**	**2.3**	**2.1**	**2.1**	**1.5**
% of Variance	**17.3**	**11.9**	**10.9**	**9.8**	**9.4**	**8.7**	**8.2**	**8.1**	**5.9**
Cumulative % of Variance	**17.3**	**29.2**	**40.1**	**49.9**	**59.3**	**68.0**	**76.2**	**84.3**	**90.2**

As shown in Table 9, Flero station shares same characteristics with the varifactors extracted from the Manerbio subset.

Pralboino

Pralboino is the most downstream monitoring station of Mella river and it presents some similar characteristics with the upstream station of Manerbio (see Table 10).

And particularly: ammonia and nickel load the same varifactor as trichloromethane and zinc, as it was shown for the Manerbio subset.

As shown for Manerbio station, the first varifactor which explains about 23% of total variance, is loaded by conductivity, suspended material, sulfates, chlorides and nitrates, being also inversely correlated with streamflow.

Table 10. Factor loadings matrix of the *Pralboino* subset. The factor loadings higher than 0.5 and lower than –0.5 are shown in bold.

	1	2	3	4	5	6	7
Streamflow ($m^3\ s^{-1}$)	**-0.797**	-0.124	0.025	0.064	-0.136	0.281	-0.082
Dissolved Oxygen ($mg\ l^{-1}$)	0.104	0.167	-0.014	**0.953**	0.145	0.155	-0.036
Dissolved Oxygen - % sat.	0.238	0.001	0.004	**0.917**	0.261	-0.049	0.018
BOD_5 ($mg\ l^{-1}$)	-0.066	**0.878**	0.246	0.166	0.007	0.023	0.121
COD ($mg\ l^{-1}$)	-0.418	**0.702**	0.107	0.200	-0.154	-0.364	-0.087
Escherichia coli - u.f.c.	0.163	0.187	0.138	**-0.523**	0.393	-0.053	**0.555**
N-NH4 ($mg\ l^{-1}$)	0.103	**0.848**	0.000	-0.174	-0.290	0.223	0.228
N-NO3 ($mg\ l^{-1}$)	**0.941**	0.016	0.012	0.155	-0.222	0.079	-0.028
Total phosphorus ($mg\ l^{-1}$)	-0.356	0.064	0.240	0.133	**0.846**	-0.053	0.075
Orthophosphate ($mg\ l^{-1}$)	0.202	**0.868**	-0.071	0.096	0.105	0.039	-0.160
Total Nitrogen ($mg\ l^{-1}$)	0.128	-0.062	0.059	**-0.566**	**-0.561**	0.451	0.177
pH	0.157	-0.272	0.005	0.299	**0.794**	0.193	-0.020
T - °C	0.305	**-0.511**	0.147	-0.325	0.270	**-0.580**	0.083
Conductivity ($\mu S\ cm^{-1}$)	**0.910**	-0.072	-0.084	0.081	-0.072	-0.129	-0.085
Hardness ($mg\ l^{-1}$)	0.751	-0.338	-0.091	0.155	0.206	-0.074	0.416
Suspended material ($mg\ l^{-1}$)	**-0.687**	0.126	0.168	0.013	0.073	-0.387	0.169
Cl^- ($mg\ l^{-1}$)	**0.920**	0.148	-0.009	0.077	0.045	0.052	0.028
SO_4^{-2} ($mg\ l^{-1}$)	**0.943**	-0.002	-0.145	0.017	-0.083	0.054	0.161
Trichloromethane ($\mu g\ l^{-1}$)	-0.069	-0.068	**0.851**	0.048	-0.043	-0.293	0.279
Trichloroethylene ($\mu g\ l^{-1}$)	-0.140	0.059	**0.919**	0.003	0.027	0.013	0.168
Tetrachloroethylene ($\mu g\ l^{-1}$)	0.082	0.005	0.367	-0.039	-0.083	-0.156	**0.851**
Cu ($\mu g\ l^{-1}$)	-0.284	**0.655**	**0.564**	-0.355	0.078	0.120	0.081
Zn ($\mu g\ l^{-1}$)	0.214	**0.550**	0.471	-0.033	0.173	0.307	-0.200
Cd ($\mu g\ l^{-1}$)	0.147	0.146	**0.630**	0.383	0.289	**-0.517**	-0.102
Total Cr ($\mu g\ l^{-1}$)	-0.193	0.211	**0.858**	-0.148	0.116	-0.077	0.030
Ni ($\mu g\ l^{-1}$)	0.000	0.184	-0.191	0.065	0.149	**0.886**	-0.175
Eigenvalue	**5.9**	**4.2**	**3.7**	**3.0**	**2.4**	**2.3**	**1.6**
% of Variance	**22.8**	**16.0**	**14.1**	**11.7**	**9.2**	**9.0**	**6.2**
Cumulative % of Variance	**22.8**	**38.7**	**52.8**	**64.5**	**73.7**	**82.7**	**88.9**

Moreover, copper, zinc and ammonia in both stations load the same varifactor although with a varifactor composition which is not exactly the same. These characteristics outline also for this station the dominance of the groundwater recharge effect and the relationship of copper and zinc with ammonia that is a pollutant typical of the untreated wastewater.

Similarities among monitoring stations and correlation with land use

As described in the methods, similarities among stations has been studied through Hierarchical Cluster Analysis. As shown by the HCA dendrogram (Figure 3) HCA, applied to the varifactors extracted pooling the whole watershed dataset, allowed to identify seven natural clusters of data.

Cluster 4: high polluting load in terms of ammonia, tetrachloroethylene and trichloroethylene and streamflow lower than the average;

Cluster 6: high groundwater recharge effect and tetrachloroethylene and trichloroethylene slightly higher than the average, all the other characteristics being slightly below the average.

As shown in Figure 5 these characteristics refer to selected monitoring stations: being cluster 2 and cluster 3 dominated by the upstream stations such as Bovegno, Villa Carcina, Castelmella and the Sarezzo station on Gobbia tributary concerning only cluster 3; cluster 4 is dominated by the Flero station conditions and cluster 6 is dominated by the lowland stations Manerbio and Pralboino. Remarkably, cluster 1 which generically refers to high streamflow conditions, is composed almost equally by data deriving from all the main Mella river stations.

Figure 3. Dendrogram obtained based on the Hierarchical Cluster Analyis applied to the varifactors extracted for the whole data set: seven natural clusters of data can be observed

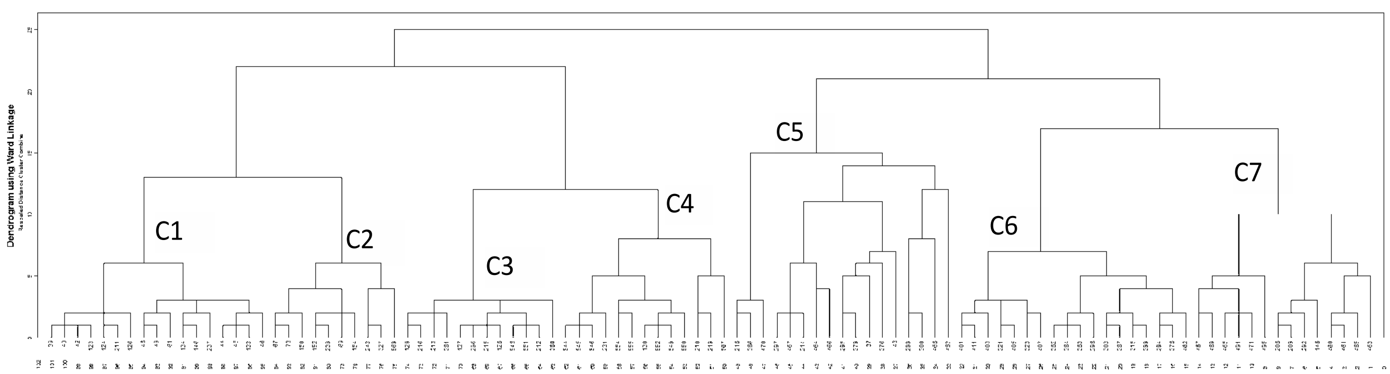


Figure 4 shows the cluster characteristics in terms of varifactors while Figure 5 shows the cluster composition with respect to the monitoring stations. It can be observed that two of the seven clusters (i.e. 5 and 7) are clearly made of outliers: cluster 5 being characterized by extremely high ammonia concentrations and cluster 7 has very high groundwater contribution (mostly nitrates and chlorides), *Escherichia coli* counts and trichloromethane and zinc.

Besides the outliers, the varifactors suggest the clusters have the following characteristics:

Cluster 1: streamflow conditions higher the average, high polluting load in terms of BOD, COD, SS and TP and E.coli;

Cluster 2: all characteristics around the average or below the average, (e.g. streamflow);

Cluster 3: copper, chromium and nickel higher than the average, all the other characteristics being around or slightly below the average;

Figure 4. Cluster characteristics in terms of varifactors. F1: groundwater recharge effect (N-NO3, Cl, SO4), F2: N-NH4 and DO, F3: Cu, Cr and Ni, F4: COD, TP and TSS, F5: PCE, TCE, F6: Q, F7: BOD, F8: E.coli, F9: TCM and Zn

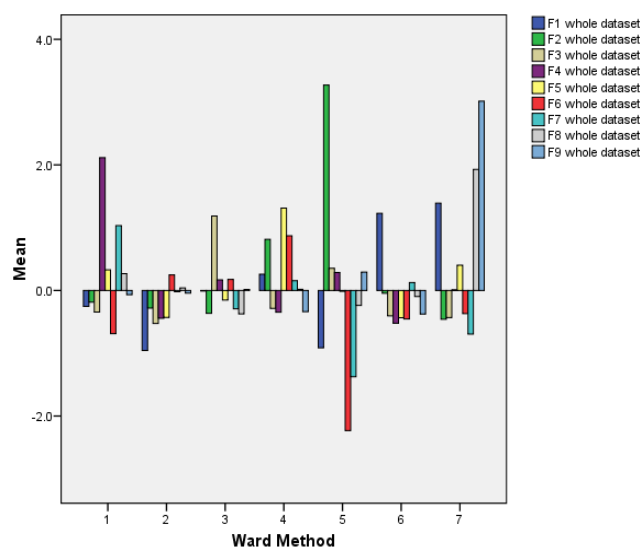
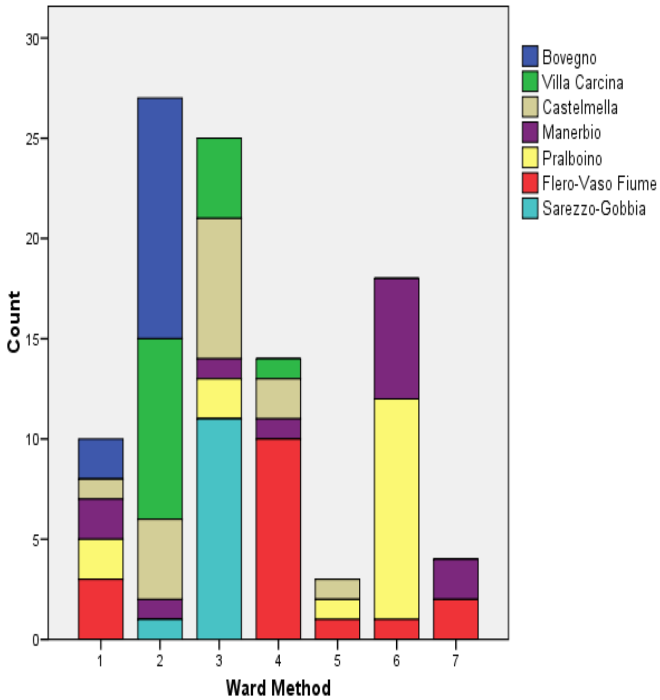


Figure 5. Cluster composition in terms of monitoring stations.



It is also interesting to investigate the relationship of these cluster of water quality profiles with land use. Table 11 shows the correlation analysis between the varifactors and the type of land use. The analysis shows clearly that only some varifactors strongly correlate with land use. Particularly F1, which is loaded by N-NO_3, Cl, SO_4) is directly correlated with cultivated areas and inversely correlated with grasslands and woodlands, F3 which is loaded by Cr, Cu and Ni is significantly correlated with urban areas and F6 which is loaded by streamflow (with an inverse relationship) is inversely correlated with the cultivated areas (then directly correlated with streamflow) and directly correlated with grasslands for pasture and woodlands (then inversely correlated with streamflow). This latter relationship is the direct consequence of the watershed characteristics with the cultivated land area being dominant in the lowland portion of the watershed and grasslands and woodlands being dominant in the upland portion of the watershed.

Table 11. Correlation matrix between land use and varifactors

		% Urban areas	% Cultivated areas	% Grasslands	% Woodlands, Natural vegetation	% Uncultivated land
% Urban areas	Pearson's r	1				
	Sig.					
	N	423				
% Cultivated areas	Pearson's r	.418(**)	1			
	Sig.	.000				
	N	423				
% Grasslands	Pearson's r	-.569(**)	-.972(**)	1		
	Sig.	.000	.000			
	N	423	423	423		
% Woodlands, Natural vegetation	Pearson's r	-.591(**)	-.979(**)	.982(**)	1	
	Sig.	.000	.000	.000		
	N	423	423	423	423	
% Uncultivated land	Pearson's r	-.019	-.429(**)	.519(**)	.350(**)	1
	Sig.	.693	.000	.000	.000	
	N	423	423	423	423	423
F1 whole dataset	Pearson's r	.239(*)	.746(**)	-.742(**)	-.709(**)	-.508(**)
	Sig.	.043	.000	.000	.000	.000
	N	72	72	72	72	72
F2 whole dataset	Pearson's r	.064	.137	-.147	-.133	-.115
	Sig.	.593	.250	.219	.264	.334
	N	72	72	72	72	72
F3 whole dataset	Pearson's r	.499(**)	.044	-.172	-.141	-.123
	Sig.	.000	.715	.148	.237	.302
	N	72	72	72	72	72
F4 whole dataset	Pearson's r	.189	.151	-.173	-.176	-.044
	Sig.	.112	.205	.146	.139	.714
	N	72	72	72	72	72
F5 whole dataset	Pearson's r	-.072	-.146	.152	.144	.096
	Sig.	.546	.222	.204	.228	.420
	N	72	72	72	72	72
F6 whole dataset	Pearson's r	-.011	-.341(**)	.294(*)	.309(**)	.088
	Sig.	.929	.003	.012	.008	.463
	N	72	72	72	72	72
F7 whole dataset	Pearson's r	-.035	.176	-.144	-.150	-.061
	Sig.	.771	.139	.227	.208	.612
	N	72	72	72	72	72
F8 whole dataset	Pearson's r	-.079	-.124	.157	.120	.226
	Sig.	.512	.300	.187	.317	.056
	N	72	72	72	72	72
F9 whole dataset	Pearson's r	-.091	-.015	.028	.034	-.037
	Sig.	.447	.902	.818	.775	.755
	N	72	72	72	72	72

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Discussion

Various researchers [23], [24], [25], [5] among others) have suggested the land-use and soil property variation can be the most sensitive factors that control the non-point pollution, the majority of these studies being based on the combined interpretation of chemical data with statistical techniques, GIS and mathematical models and focusing mostly on nitrogen and agriculture-related pollutants. Our findings confirm that land use is one of the major discriminant of water quality profiles and, although land uses themselves may be also correlated to each other as it happens in this watershed, it has been shown that some specific pollutants (such as N-NO_3, Cl, SO 4) clearly correlate with cultivated areas while others (such as Cr, Cu and Ni) correlate with urban areas. According to Lu *et al.*, 2015,chloride is an ideal indicator of sewage and manure input and dilution because the chemical behavior of chloride in natural water is conservative and its concentration can change only by mixing within the river system. Particularly, Cl sources in water from various areas generally include natural sources (dissolution of minerals), agricultural chemicals (i.e., KCl), animal waste, septic effluent, and road salt.

Chemical fertilizers generally have high nitrogen contents with low chloride contents. Sewage effluents have high Cl contents and low ratios of NO_3/Cl. Elevated nitrate contents can be observed in water samples with agricultural inputs. The values of NO_3/Cl may decrease if de-nitrification removes nitrate or if plants take up the nutrients.

Thus, the ratio of NO_3 and the Cl concentrations may provide additional information to help distinguish the different input sources. The performed analysis showed how the correlation between nitrates and chlorides changes among the monitoring stations within the Mella watershed. Consistently, the ratios of NO_3/Cl were found significantly different among the water quality profiles of the 5 main clusters (Kruskal-Wallis chi-square: 16.68; df: 4 P<0.01, , see Fig.6) identified in this study which were also characterized by significant differences in terms of land use. Particularly, the ratios of NO3/Cl were found to decrease as the percent of cultivated areas increased. These results support the hypothesis that the NO3/Cl ratios may help to distinguish the different pollutant sources which can be reasonably correlated with specific land use attributes.

Conclusions

Factor analysis was proven effective to distinguish between anthropogenic and geogenic factors influencing the instream water quality patterns.

FA enabled to outline the effect of the dominant pollution source at different sites along the Mella river, highlighting the effect of the Gobbia and Vaso Fiume tributaries and the effect of the groundwater recharge area on Mella river water quality. FA results confirmed that groundwater is a significant source of N-NO_3 for Mella river although it does not appear to be a source for other pollutants. On the other hand, heavy metals and the chlorinated solvents showed variable correlations at different monitoring stations, suggesting different polluting sources. The groundwater varifactor and the varifactor correlated with metals showed also a clear relationship with land use.

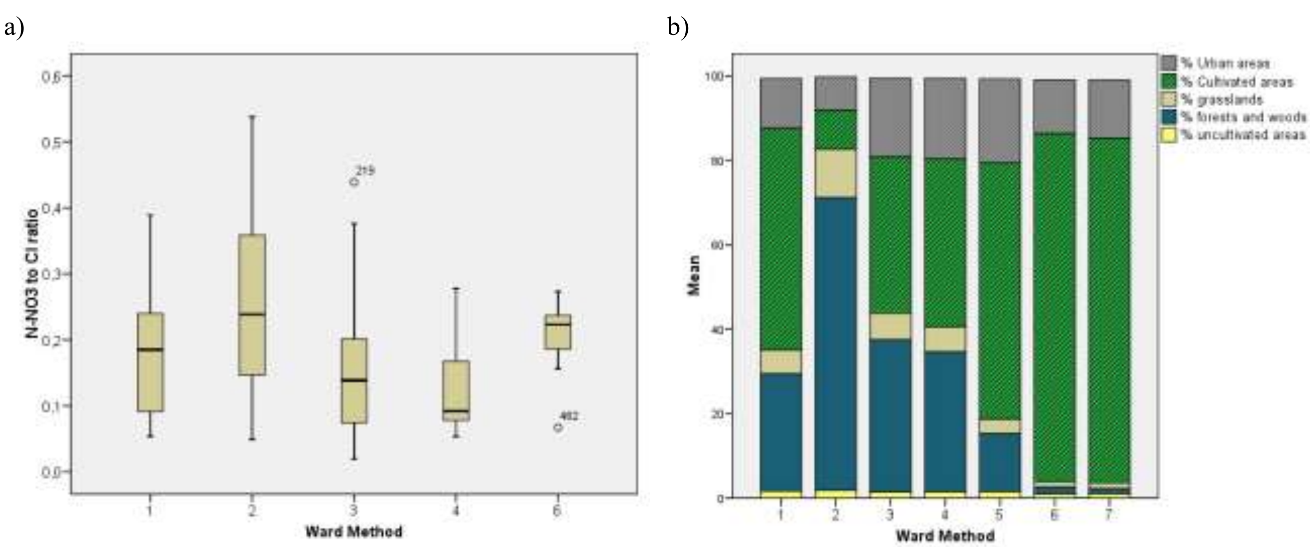


Figure 6. a) NO3/Cl ratios of the 5 main water quality profile clusters; b) Cluster composition in terms of land uses.

References:

1. Sophocleous M. Interactions between groundwater and surface water: the state of the science. Hydrogeology Journal. 2002, 10: 52-67.

2. Simeonov V, Stratis J.A, Samara C, Zachariadis G, Voutsa D, et al. Assessment of the surface water quality in Northern Greece. Water Research. 2003, 37: 4119-4124.

3. Pekey H, Karakaş D, Bakoğlu M. Source apportionement of trace metals in surface waters of a polluted stream using multivariate statistical analyses. Marine Pollution Bulletin. 2004, 49: 809-818.

4. Asher W.E, Luo W, Campo K.W, Bender D.A, Robinson K.W, et al. Application of a source apportionment model in consider ation of volatile organic compounds in an urban stream. Environmental Toxicology and Chemistry. 2007, 26(8): 1606-1613.

5. Jeanneau L, Faure P, Montarges-Pelletier E. Quantitative multimolecular marker approach to investigate the spatial variability of the transfer of pollution from the Fensch River to the Moselle River (France). Science of the Total Environment. 2008, 389: 503-513.

6. Gevaert V, Verdonck F, Benedetti L, De Keyser W, De Baets B. Evaluating the usefulness of dynamic pollutant fate models for implementing the EU Water Framework Directive. Chemosphere. 2009, 76: 27-35.

7. Ki S.J, Kang J-H, Lee Y.G, Lee Y.S, Sthiannopkao S, et al. Statistical assessment for spatio-temporal water quality in Angkor, Cambodia. Water Science and Technology. 2009, 59(11): 2167-2178.

8. Chon H-S, Ohandja D-G, Voulvoulis N. Implementation of E.U. Water Framework Directive: source assessment of metallic substances at catchment levels. Journal of Environmental Monitoring. 2010, 12: 36-47.

9. Tobiszewski M, Tsakovski S, Simeonov V, Namiesnik J. Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. Chemosphere. 2010, 80: 740-746.

10. Tobiszewski M, Namiesnik J. PAH diagnostic ratios for the identification of pollution emission sources. Environmental Pollution. 2012, 162: 110-119.

11. Lu L, Cheng H, Pu X, Liu X, Cheng Q. Nitrate behaviors and source apportionment in an aquatic system from a watershed with intensive agricultural activities. Environ. Sci Processes Impacts. 2015, 17: 131-144.

12. Joint Research Centre - European Commission Directorate - and Sustainability Indicators and methods for the ecological status assessment under the Water Framework Directive. Linkages between chemical and biological quality of surface waters. Edited by Angelo G. Solimini, Ana Cristina Cardoso and Anna-Stiina Heiskanen, 2006.

13. Cho J, Mostaghimi S, Kang M.S. Development and application of a modeling approach for surface water and groundwater interaction. Agricultural Water Management. 2010, 97: 123-130.

14. Schilling C, Zessner M, Blaschke A.P, Gutknecht D, Kroiss H. Groundwater protection and diffuse nitrogen emissions to surface waters - Which catchment areas have to be considered?. Water Science and Technology. Water Supply. 2007, 7(3): 103-110.

15. Ebel B.A, Mirus B.B, Heppner C.S, VanderKwaak J.E, Loague K. First-order exchange coefficient coupling for simulating surface water-groundwater interactions: parameter sensitivity and consistency with a physics-based approach. Hydrological Processes. 2009, 23: 1949-1959.

16. Martin C, Aquilina L, Gascuel-Odoux C, Molénat J, Faucheux M et al. Seasonal and interannual variations of nitrate and chloride in stream waters related to spatial and temporal patterns of groundwater concentrations in agricultural catchments. Hydrological Processes. 2004, 18: 1237-1254.

17. Sulis M, Meyerhoff S.B, Paniconi C, Maxwell R.M, Putti M et al. A comparison of two physics-based numerical models for simulating surface water–groundwater interactions. Advances in Water Resources. 2010, 33: 456-467.

18. Menció A, Mas-Pla J. Assessment by multivariate analysis of groundwater-surface water interactions in urbanized Mediterranean streams. Journal of Hydrology. 2008, 352: 355-366.

19. Azzellino A, Salvetti R, Vismara R, Bonomo L. Combined use of the EPA-QUAL2E simulation model and factor analysis to assess the source apportionment of point and non point loads of nutrients to surface waters. Science of the Total Environment. 2008, 371: 214-222.

20. APAT, IRSA CNR. Metodi Analitici per le Acque, APAT Manuali e Linee Guida 29/2003.

21. Afifi A, Clark V. Computer-Aided Multivariate Analysis. Texts in Statistical Science. Chapman & Hall. 1996.

22. Zhuang P, Pavlostathis SG. Effect of temperature, pH and electron donor on the microbial reductive dechlorination of chloroalkenes. Chemosphere. 1995, 31(6): 3537-3548.

23. Tran Ngoc Han, Gin Karina Yew-Hoong, Ngo Huu Hao. Fecal pollution source tracking toolbox for identification, evaluation and characterization of fecal contamination in receiving urban surface waters and groundwater. Science of the Total Environment.2015, 538: 38-57.

24. Mehdi B, Lehner B, Gombault C, Michaud A, Beaudin I et al. Simulated impacts of climate change and agricultural land use change on surface water quality with and without adaptation management strategies. Agriculture Ecosystems & Environment. 2015, 213: 47-60.

25. Ouyang W, Huang H, Hao F, Guo B. Synergistic impacts of land-use change and soil property variation on non-point source nitrogen pollution in a freeze-thaw area. Journal of Hydrology. 2013, 495: 126-134.