Environ Monit Assess (2009) 156:73–90
DOI 10.1007/s10661-008-0464-6

Development and sensitivity analysis of a global drinking water quality index

C. J. Rickwood · G. M. Carr

Received: 21 January 2008 / Accepted: 26 June 2008 / Published online: 12 September 2008

Abstract The UNEP GEMS/Water Programme is the leading international agency responsible for the development of water quality indicators and maintains the only global database of water quality for inland waters (GEMStat). The protection of source water quality for domestic use (drinking water, abstraction etc) was identified by GEMS/Water as a priority for assessment. A composite index was developed to assess source water quality across a range of inland water types, globally, and over time. The approach for development was three-fold: (1) Select guidelines from the World Health Organisation that are appropriate in assessing global water quality for human health, (2) Select variables from GEMStat that have an appropriate guideline and reasonable global coverage, and (3) determine, on an annual basis, an overall index rating for each station using the water quality index equation endorsed by the Canadian Council of Ministers of the Environment. The index allowed measurements of the frequency and extent to which variables exceeded their respective WHO guidelines, at each individual monitoring station included within GEMStat, allowing both spatial and temporal assessment of global water quality. Development of the index was followed by preliminary sensitivity analysis and verification of the index against real water quality data.

Keywords Composite index · Drinking water · GEMS/Water · Global water quality · Vistula River

Introduction

Any number of water quality measurements can serve, and have already been used, as indicators of water quality; however, there is no single measure that can describe overall water quality for any one body of water, let alone at a global level. As such, a composite index that quantifies the extent to which a number of water quality measures deviate from normal, expected or ‘ideal’ concentrations may be more appropriate for summarizing water quality conditions across a range of inland water types and over time. To date, no such global index has been developed. However, such a global index of water quality is needed to assess changes in water quality over time and space and also to evaluate successes and failures of international treaties designed to protect aquatic resources. For example, a global index could be a valuable tool for tracking progress toward meeting the

C. J. Rickwood (✉) · G. M. Carr
UNEP GEMS/Water Programme c/o Place Vincent
Massey, 8th floor, 351 St Joseph Boulevard,
Gatineau, QC, K1A 0H3, Canada
e-mail: carrie.rickwood@usask.ca

Millennium Development Goals and the Plan of Implementation of the World Summit on Sustainable Development, as well as other internationally agreed goals and targets.

A number of countries have begun the process of developing composite indices of water quality to describe the state of their domestic waters, including the United States of America (Cude 2001), Taiwan (Liou et al. 2004), Argentina (Pesce and Wunderlin 2000), Australia (ISC 2005), Canada (Khan et al. 2003; Lumb et al. 2006; CCME 2001) and New Zealand (Smith 1989, 1990; Nagels et al. 2001). Similar to indices of economic strength, such as Gross National Product (GNP), these water quality indices take information from a number of sources and combine them to develop an overall snapshot of the state of the national system. In the case of inland waters, the information used to generate the indices typically consists of concentrations of a number of different water quality variables measured as part of routine national, regional, and local monitoring programmes. Just as in the case of economic indices (Statistics Canada 2002; Esty et al. 2006), there is considerable debate as to which measures should be included in the derivation of an index, and what type of information such a composite index is able to provide to the general public and to policy makers (Harkness 2004). Despite the debate, there is some agreement that water quality indices are useful tools for comparing water quality across systems and over time, and they can provide a benchmark for evaluating successes and failures of management strategies aimed at improving water quality.

There are many global water quality issues, and a number of priority issues of concern. One of these is safeguarding drinking water supplies. The protection of source water quality for domestic use (drinking water, abstraction etc) was identified by an experts' group (UNEP workshop, Vienna, Austria May 4th–6th 2005) as a priority for assessment (the full experts' recommendation report is available at: www.gemswater.org/publications/index-e.html). Assessment of source water quality was selected because of its importance to human health, because it could be conducted on a global scale, and because it could be developed using internationally recognized guidelines for drinking water quality, such as those from the World Health Organisation (WHO 2004).

The overall goal of this study is to report on the development, sensitivity analysis, and validation of a global index of source water quality. The water quality data used to analyze and validate the index calculations have been derived from GEMStat, an online global database of water quality for inland waters maintained by the UNEP GEMS/Water Programme (www.gemstat.org). The performance of the index at several geographic scales (global, regional, and local) is examined and limitations to the data and index are discussed along with suggestions for future development.

Index development

The approach for developing an index for global source drinking water quality has three parts: (1) *Guideline selection*: Selecting guidelines that are appropriate in assessing global water quality for human health, (2) *Variable selection*: selecting variables from GEMStat that have an appropriate guideline and have reasonable global coverage, (3) *Station selection*: selecting only stations that measure variables consistently on an annual basis.

Guideline selection

The first objective was to select water quality variables that could be associated with an existing drinking water quality guideline. As the goal was to develop a global index, the variables selected were based on those in the World Health Organisation's Drinking Water Guidelines (WHO 2004). The primary purpose of the WHO guidelines is the protection of public health by describing guideline values for constituents of water or indicators of water quality. These guidelines divide variables into two categories:

i. Health guidelines, which take into account chemical and radiological constituents that have the potential to have a direct adverse effect on human health; and
ii. Acceptability guidelines, which include variables that may not have any direct health effects but result in objectionable taste or

Table 1 Comparison of WHO drinking water guidelines (GLs) for selected variables against guidelins from the European Union (EU), United States (USEPA) and Australia

Variable	WHO	EU[a]	USEPA	Australia
Ammoniacal-N	1.5 mg/L	0.50 mg/L	No GL	0.50 mg/L
pH	6.5–8	No GL	6.5–8.5	6.5–8.5
Chloride	250 mg/L	250 mg/L	250 mg/L	250 mg/L
Iron	0.3 mg/L	0.2 mg/L	0.3 mg/L	0.3 mg/L
Lead	0.01 mg/L	0.01 mg/L	0.015 mg/L	0.01 mg/L
Arsenic	0.01 mg/L	0.01 mg/L	0.01 mg/L	0.007 mg/L
Copper	2.0 mg/L	2.0 mg/L	1.3 mg/L	2.0 mg/L
Faecal coliform bacteria	0 counts/100 mL	0 counts/100 mL	0 counts/100 mL	No GL

[a] WHO guidelines for drinking water were used as a basis for the standards for the EU Drinking Water Directive

> odour in the water. Water that is highly turbid, highly coloured or that has an objectionable taste or odour could lead the consumer to believe that the water is unsafe.

To assess the robustness of the guidelines proposed by WHO, comparisons with drinking water quality guidelines currently in place in the European Union, Australia and USA were conducted. For this comparison some of the most common variables measured and reported in GEMStat were selected (ammoniacal-N, pH, chloride, iron, lead, arsenic, copper and faecal coliform bacteria) (Table 1).

The guidelines for the variables selected compared well across nations and international agencies, with only small deviations from each other (Table 1). The only WHO guideline that was substantially higher (and, therefore, less stringent) than the others was ammoniacal-N (1.5 mg/L), when compared to the EU and Australian guideline of 0.5 mg/L. The WHO guideline for pH was slightly more conservative than US and Australian guidelines, whereas the EU guideline for iron was more stringent than WHO, US, or Australian guidelines. The US guideline for lead was less stringent than other agencies, but the US guideline for copper was the most stringent of all agency guidelines. Finally, the Australian guideline for arsenic was the most stringent of all agency guidelines.

It was concluded that based on the variables selected, WHO drinking water quality guidelines were representative of a number of national guidelines currently in place, and, therefore were selected for use in the development of a global index of source water quality. In addition, based on the categorisation of the parameters by WHO, three water quality indices (WQI) were developed (Table 2).

Variable selection and global and regional coverage of included variables

The variables listed under the WHO health and acceptability categories were selected from GEMStat for inclusion in the index (WHO 2004; Table 3). Refinement of the database was needed

Table 2 Description of indices with details of types of variables included and justification of the index calculation

Index	Parameters included	Justification
Drinking WQI (DWQI)	All parameters regardless of category	Gives an overall ‘big picture’ as to the quality of water
Acceptability WQI (AWQI)	Parameters listed under the acceptability category	Provides an assessment of the public’s perception of the quality of water, rather than specific health issues, as it assesses variables that cause unacceptable taste or odour in drinking water
Health WQI (HWQI)	Parameters listed under the health category	Provides a more relevant assessment of water quality as it includes only variables that have the potential to result in adverse health effects in humans

Table 3 List of all possible parameters, and their associated WHO guideline and whether they were measured in 20%, 35% and 50% of countries in all regions: Europe, Asia, Africa, Americas and Oceania

Variable	Guideline	Country coverage selection criteria		
(units of measurement)	(mg/L)	20%	35%	50%
Health variables				
2,4–D	0.03	LD	LD	LD
Aldicarb	0.01	LD	LD	LD
Aldrin and dieldrin	0.00003	LD	LD	LD
Antimony	0.02	LD	LD	LD
Arsenic	0.01	Y	LD	LD
Atrazine	0.002	LD	LD	LD
Barium	0.7	LD	LD	LD
Boron	0.5	Y	LD	LD
Cadmium	0.003	Y	LD	LD
Chromium	0.05	Y	LD	LD
Copper	2	Y	Y	Y
Cyanide	0.07	LD	LD	LD
DDT and metabolites	0.001	LD	LD	LD
Endrin	0.0006	LD	LD	LD
Fluoride	1.5	Y	Y	Y
Lead	0.01	Y	Y	LD
Lindane	0.002	LD	LD	LD
Manganese	0.4	Y	Y	Y
Mercury	0.001	Y	LD	LD
Nickel	0.02	LD	LD	LD
Nitrate	50	Y	Y	Y
Nitrite	3	Y	Y	LD
Selenium	0.01	LD	LD	LD
Acceptability variables				
Aluminium	0.1	LD	LD	LD
Ammoniacal-N	1.5	Y	Y	Y
Chloride	250	Y	Y	Y
Hardness	200	LD	LD	LD
Hydrogen sulphide	0.05	LD	LD	LD
Iron	0.3	Y	Y	Y
pH (pH units)	6.5, 8[a]	Y	Y	Y
Sodium	200	Y	Y	Y
Sulphate	250	Y	Y	LD
Total dissolved solids	600	LD	LD	LD
Turbidity (NTU)	5	LD	LD	LD
Zinc	3	Y	Y	LD

Y Complete data, *LD* lack of data, *NTU* nephelometric turbidity units

[a] Minimum (6.5) and maximum (8) guidelines for pH

to ensure that the variables included in the index calculations were similar; that there was a minimum level of similarity in the variables included to calculate an index value for each region: Asia, Africa, Americas, Europe and Oceania. Therefore, we determined criteria for the percent coverage of variables by countries within each of the regions.

We first assessed the distribution of variables under 20%, 35% and 50% regional coverage limits, meaning that each variable had to have been measured in either 20%, 35% or 50% of countries within each region (Table 3). The distribution of acceptability variables was quite consistent at all three criteria levels, with five variables measured in at least 50% of countries within each region, and seven variables meeting the 35% and 20% regional country coverage criteria. The country coverage for health-related variables was less comprehensive than the acceptability variables, with only four variables having been measured in 50% of countries in each region, compared to 11 variables that were measured in 20% of countries in each region. Since the inclusion of more variables should yield a more relevant water quality index, the 20% criterion was selected as a global distribution guideline for the development of the indices. Thus, each variable included in the index had to have been measured in at least 20% of countries in each of the major regions.

Initial calculations of the water quality index included faecal coliform bacteria (FCB) because they were commonly reported in GEMStat and, more importantly, their presence is directly linked to the prevalence of waterborne diseases in humans. However, on initial calculation it was revealed that FCB was strongly influencing the index result (UNEP GEMS/Water Programme 2007). It would seem that because of the stringent guideline set by WHO (0 counts per 100 mL), FCB was consistently in exceedance and far outweighed any other health variable included in the index equation. Due to this heavy influence on the index it was not possible to compare stations that had FCB with those that had not. It was decided that removal of FCB was required from the health criteria and separate assessment of FCB would be conducted. As a result, the current indices did not include any measurement of microbial data, which is an important aspect in analysing the safety of drinking water. The index developed here assesses chemical contamination, and as such the inclusion of microbial measurements into the current index, or development of a microbial index, will be assessed separately.

Station selection

Following selection of water quality variables based on water quality guideline availability, as well as based on global coverage of the different water quality variables, the database used for index calculation was further refined to only include data from stations where monitoring of several variables was consistent over time. Thus, data used for deriving an index were only selected from stations where at least four variables were measured per year, and, that each of these variables was measured at least four times per year, hence, the 'Four by Four' (4 × 4) rule. This rule ensures that only stations that regularly monitor variables were included and follows recommendations made by the Canadian Council of Ministers for the Environment (CCME 2001).

Following the decision to use the WHO guidelines and selection of variables, observations, and monitoring stations to be used in index derivation, the three indices (DWQI, HWQI, AWQI) were developed. Separate databases were generated for the derivation of each index and, for the purposes of this investigation, the selection rules as outlined above were applied to each database (Drinking, Health and Acceptability), such that variables, stations, and records in each database met all of the following criteria:

1) Variables included were measured in at least 20% of countries in each major global region: Asia, Africa, Europe, Oceania and Americas; and,
2) Stations included measured at least four variables, four times per year (4 × 4 rule).

It is important to stress that while these indices provided an overall picture of the quality of a body of water, they can not be relied upon to definitely determine if a water source is safe for drinking. Primarily because of a lack of available monitoring data, there are a number of variables that were not included in the indices that could still adversely affect the safety (from a human health perspective) or acceptability of water for drinking.

Derivation and application of the index

The equation used for the derivation of the indices reported here was based on the water quality index (WQI) endorsed by the Canadian Council of Ministers of the Environment (CCME 2001). The CWQI model was selected as it requires the use of a benchmark or guideline which allowed us to compare values to the World Health Organisation's Drinking Water Quality Guidelines (WHO 2004). Instead of normalizing observed values to subjective rating curves, the Canadian Water Quality Index (CWQI) compares observations to a benchmark, where the benchmark may be a water quality standard or site specific background concentration (CCME 2001; Khan et al. 2003; Lumb et al. 2006). The CWQI quantifies for one station, over a predetermined period of time (typically 1 year), the number of parameters that exceed a benchmark, the number of records in a dataset that exceed a benchmark, and the magnitude of exceedance of the benchmark. The index is flexible in terms of the benchmarks that are used for calculation, and depends on the information required from the index. By applying the Canadian index model and WHO guidelines to our data set we were able to develop an index that uses globally accepted guidelines for drinking water.

The index allows measurements of the frequency and extent to which variables exceed their respective guidelines at each monitoring station. The index was calculated on an annual basis resulting in an overall rating for each station per year. This allowed both the spatial and temporal assessment of global water quality.

A full description of the index calculation has been outlined by CCME (2001, 2005). In brief, the Canadian Water Quality Index (CWQI) equation is calculated using three factors as follows:

$$\mathrm{WQI} = 100 - \left(\frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732}\right)$$

Where:

F_1 represents *Scope*: The percentage of variables that exceed the guideline;

F_2 represents *Frequency*: The percentage of individual tests within each variable that exceeded the guideline;

F_3 represents *Amplitude*: The extent (excursion) to which the failed test exceeds the guideline; and The constant, 1.732, is a scaling factor (square root of three) to ensure the index varies between 0 and 100.

WQI designations

The index equation generates a number between 0 and 100, with 0 indicating poor and 100 indicating excellent water quality. Within this range, designations have been set by CCME (2005) to classify water quality as “poor” (0–44), “marginal” (45–64), “fair” (65–79), “good” (80–94) or “excellent” (95–100). These classifications were adopted here for illustrative purposes. Validation and expert opinion should be determined before designations are applied, but we would propose applying treatment descriptions to each designation in future development of the index. For example, applying descriptions as to the level of:

1) Removal processes—pre-treatment, flocculation, sedimentation, coagulation and filtration of water for drinking
2) Inactivation processes—primary or secondary disinfection of water for drinking

Global water quality index

Once the indices were calculated, index values were placed into their designations and the proportion of stations within each designation and region were calculated. For ease of illustration, index values for DWQI only were grouped into good–excellent, fair, and marginal–poor. The proportions within each designation were then plotted over time for each region (Fig. 1). A general improvement in water quality in the Americas could be detected over time with a reduction in the proportion of stations categorized as fair, marginal or poor. The pattern was less clear in Asia, with fluctuations in the proportion of stations designated as good or excellent between 0.4 and 1 over time but with a high proportion of stations with a good or excellent designation in the most recent sampling years (2003–2006). In Africa, the proportion of stations designated good, or excellent, decreased with a corresponding increase in the proportion of stations designated fair, suggesting a slight deterioration in water quality with time. In all three of these regions (Africa, Asia and Americas) it would seem that the proportion of good–excellent stations tended to decrease as the number of monitoring stations with calculated index values increased. This is an interesting observation that could indicate that the more stations included in the assessment of these regions results in a more uniform distribution of stations within each category. In Europe, there was very little deviation from the good,

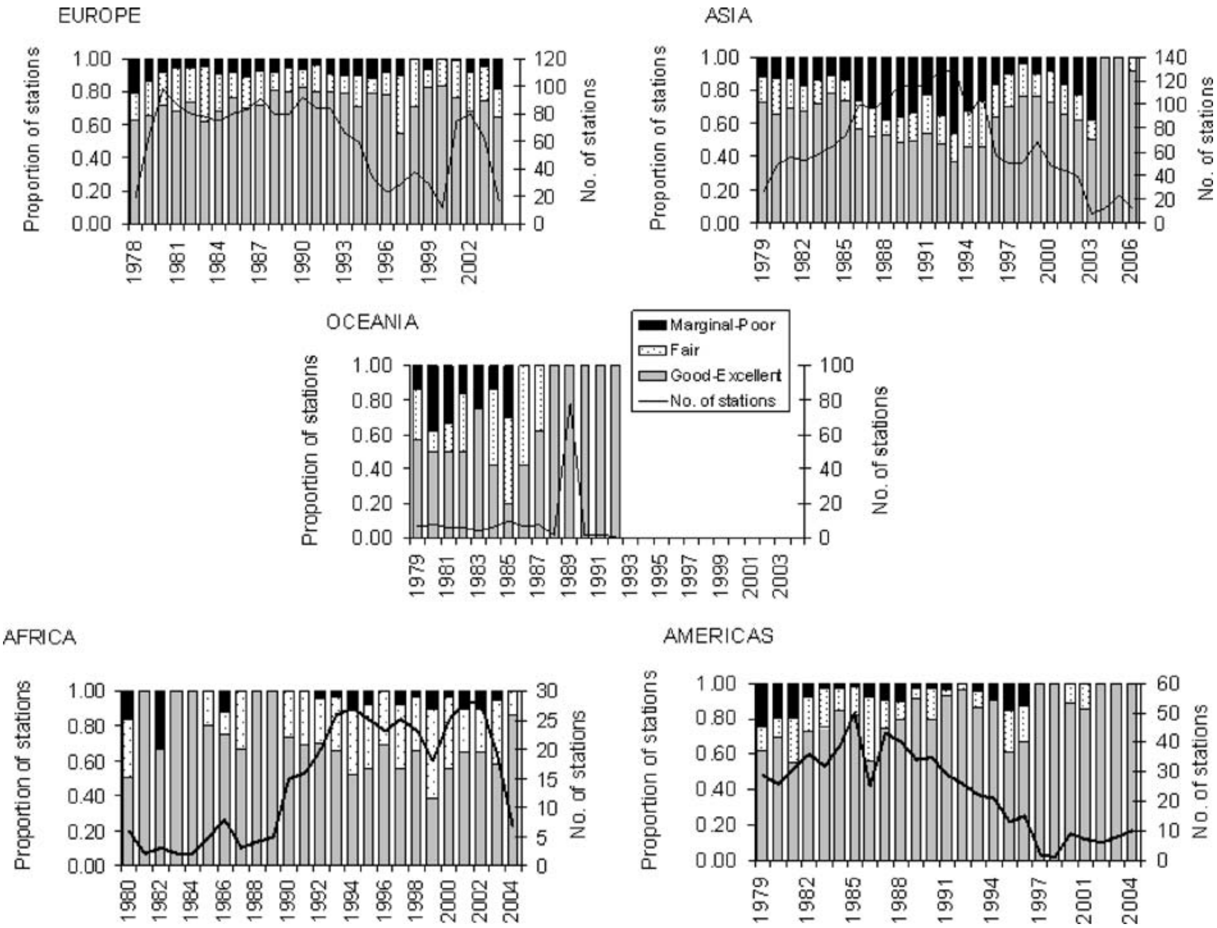


Fig. 1 Temporal trends in water quality (DWQI, HWQI and AWQI) for the five major regions (Africa, Americas, Asia, Europe and Oceania). *Bars* represent proportion of stations within each region designated poor-marginal, fair, and, good–excellent. *Line* indicates number of stations included in the calculation

or excellent, designations of monitoring stations over time, with a fairly consistent number of stations designated poor–fair since 1978. This fairly consistent result does not seem to be affected by the drop in number of stations between 1992 and 2001, indicating that the stations included within this time-period seemed to be fairly representative of water quality within this region. Unfortunately, the lack of data for Oceania did not allow a full temporal assessment; however, after 1985 the proportion of stations designated good–excellent increased. Again, similar to Europe, the number of stations did not seem to have any bearing on the designations, with a large increase in number of stations in 1989 with no corresponding change in station designations.

The development of the three indices outlined in this report allowed assessment of water quality not only temporally on a station-by-station basis but also spatially across different regions, countries and/or watersheds. Sensitivity analysis and validation of the indices also allowed for assessment of the suitability of the WHO guidelines and, specifically, whether they were too stringent, or, whether additional variables needed to be included that were not assessed by WHO. With this in mind the following sections focus not only on providing both a regional and watershed assessment of drinking water quality but also on investigating variable and/or guideline sensitivity for the purposes of improving the database and/or index calculation.

Table 4 Number of stations, listed by country, for which an HWQI and AWQI were calculated for the year 2002

Countries	HWQI	AWQI
Morocco	6	6
Argentina	5	7
Japan	11	13
Republic of Korea	1	1
Belgium	37	18
Poland	6	6
Switzerland	2	6
South Africa	–	24
India	–	24
Pakistan	–	5
Russian Federation	–	34

Sensitivity analysis

Sensitivity analysis was conducted to assess the influence on index results of the variables included in the calculation of each of the three indices. This entailed removing each variable from the index calculation and comparing the reduced indices to the original index that included all measured variables. The objective was to observe whether the removal of any one variable changed the index so much that it was no longer correlated with the original indices. The sensitivity analysis was conducted on the HWQI and AWQI only. DWQI was omitted from this analysis since any variable that influenced either HWQI or AWQI would also automatically influence DWQI, since both HWQI and AWQI are sub-indices of DWQI.

Sensitivity analysis was conducted on monitoring data from 2002, the most recent year in which data for all variables were available (Table 4). A total of 68 stations had an HWQI designation and 144 stations had an AWQI designation in 2002. Approximately 45% of stations were classed as excellent and 25% were classed as good for health aspects (HWQI) in 2002 (Fig. 2). Less than 2% of stations with an HWQI were classed as poor in the same year. For AWQI, approximately 30% of stations were classed as good with less than 15% of stations classed as excellent, while the majority (40%) of stations were classed as fair in 2002 (Fig. 2).

To determine which variables had greatest influence on index results, the variables contributing to each index were selected for the sensitivity analysis. Once each of the variables were removed, the HWQI and AWQI were recalculated and plotted against the original index (Fig. 3). For AWQI (Fig. 3a), the removal of pH increased the number of stations designated as excellent from approximately 15% to 50% and reduced the number of stations designated as marginal from approximately 40% to <10%. Correlation analysis showed that all indices were significantly inter-correlated regardless of which variable was removed ($p < 0.001$). However, pH showed the least strong, but still highly significant ($r = 0.832$, $P < 0.001$), correlation to AWQI which follows the pattern expected from Fig. 3a.

For HWQI (Fig. 3b), the removal of lead and arsenic had the greatest impact on station designations. Removal of arsenic resulted in higher index scores with more stations categorized as good or excellent and fewer stations categorized as poor, marginal or fair. Similarly, the removal of lead re-

Fig. 2 The number of stations (percentage of total stations globally) in 2002, categorised as poor, marginal, fair, good and excellent for both HWQI and AWQI

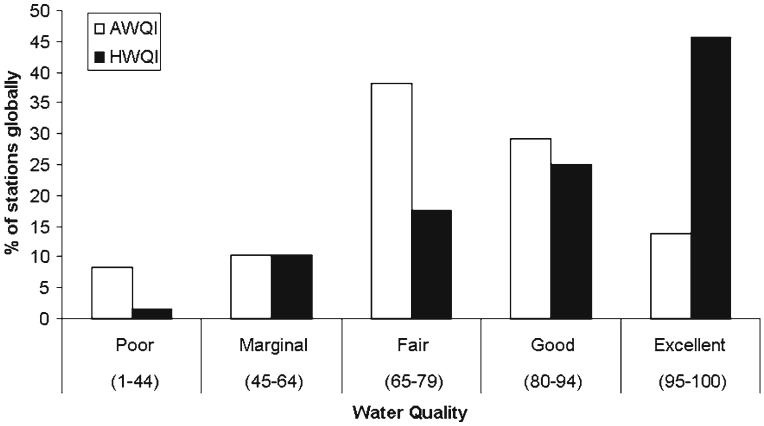
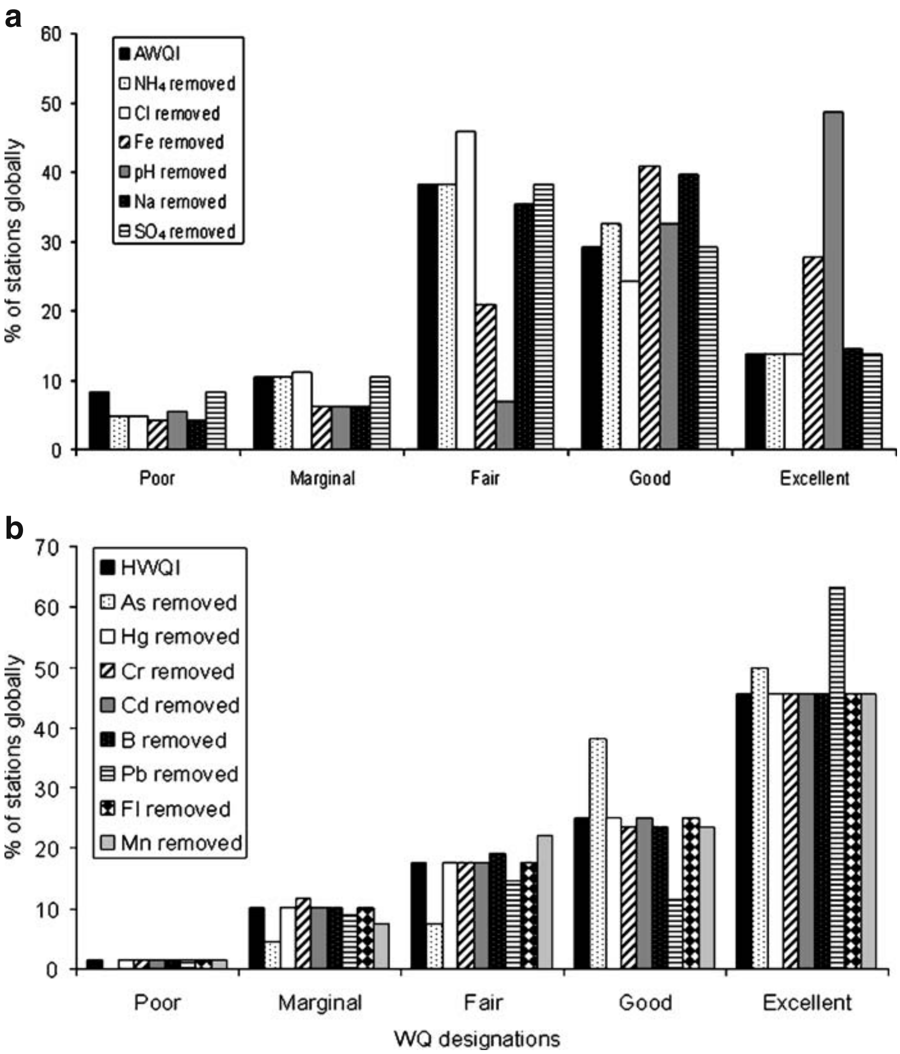


Fig. 3 Designation of stations in 2002 (displayed as a percentage of the total number of stations) for **a** AWQI [$n = 144$ stations] and **b** HWQI [$n = 68$ stations], and the contributions of each variable to their respective index



duced the amount of stations designated good and increased the amount of excellent from 45% to 63%. However, the removal of these variables did not significantly change the HWQI designations (i.e. Pearson's correlation analysis was not significantly different ($p < 0.001$)). This would suggest that the HWQI was not particularly sensitive to any one variable. It was concluded that no one variable was strongly influencing the HWQI, suggesting that the index value was robust regardless of the variables that were included.

Conclusions from sensitivity analysis

The Pearson's correlation matrix for both HWQI and AWQI revealed that regardless of which variable was removed, the indices were still significantly correlated. This would indicate that both indices were not strongly driven by one particular variable, but rather by the combination of all variables.

Index evaluation: Vistula River, Poland

Overview

To validate the three indices against real water quality monitoring data both from GEMStat and from published literature, a station-level assessment of the indices was conducted. The stations selected for validation were chosen within the same river basin, to allow assessment on both a temporal and spatial (upstream to downstream) scale. The river basin selected was the Vistula River in Poland. This river was of particular interest for this study as the water quality within the Vistula River basin has deteriorated during the last 30 years due to both urban and industrial growth (Laenen and Dunnette 1997). Despite showing a slight improvement in the last decade, the issue of water quality is still of concern as it is used to supply drinking water for many communities; there are approximately 30 surface-water intakes within the river basin. The major issue with regards to water quality is that most of the heavy industry is located in the upper part of the river, polluting the river from upstream to downstream (Laenen and Dunnette 1997).

The Vistula River is the longest river in Poland, spanning 1,047 km and draining an area of 194,424 km^2. Vistula flows from south to north, originating at Barania Góra (1,220 m high) in the Beskidy Mountains and discharging into the Vistula Lagoon and Gdañsk Bay of the Baltic Sea (Fig. 4). Water quality monitoring stations are located in three of the large Polish cities through which the Vistula River flows (Fig. 4).

The first objective was to assess the water quality (overall, health and acceptability) of the Vistula River over time. An overview of the temporal trends of all three indices at each station on the Vistula River is illustrated in Fig. 4. Two clear patterns were observed:

1) Spatial: at each time point the quality of water improved from upstream (Krakow) to downstream (Tczew); and
2) Temporal: over time the quality of water, with respect to health and overall drinking water quality variables, improved in the upstream and midstream sites. The quality of water with respect to acceptability improved at all three sites from 1992 to 1997, with little change in acceptability after 1997.

In addition, the trends in DWQI, HWQI and AWQI correspond well at each site along the

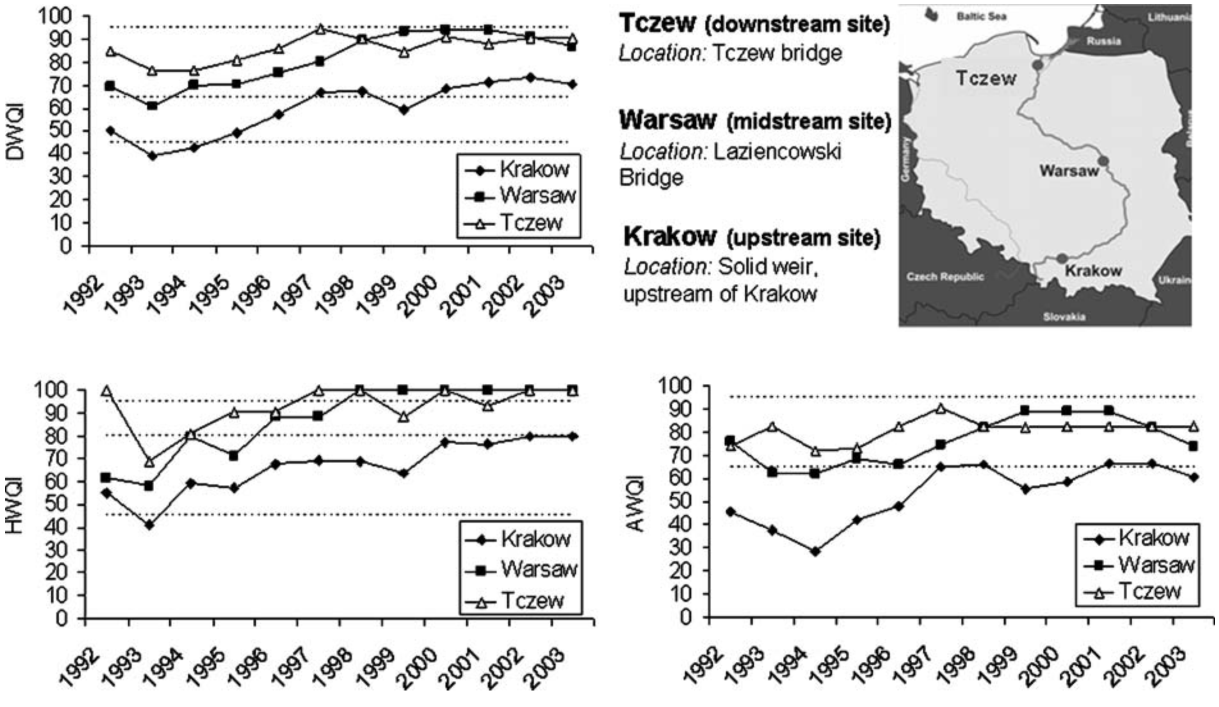


Fig. 4 DWQI, HWQI and AWQI at three sites along the Vistula River (Tczew, Warsaw and Krakow) between 1992 and 2003. *Dashed lines* correspond to index designations where: 0–44 = poor, 45–64 = marginal, 65–79 = fair, 80–94 = good, 95–100 = excellent)

Table 5 Pearsons correlation matrix for DWQI and AWQI against the contributing variables at Krakow, Warsaw and Tczew

Variable	Krakow (upstream)		Warsaw (midstream)		Tczew (downstream)	
	DWQI	AWQI	DWQI	AWQI	DWQI	AWQI
Ammoniacal-N	−0.902**	−0.828**	−0.894**	−0.867**	−0.238	−0.245
Chloride	−0.831*	−0.791*	−0.724	−0.802*	−0.533	−0.326
Iron	−0.916***	−0.875**	−0.829*	−0.658	−0.648	−0.810**
Sodium	−0.828*	−0.790*	−0.719	−0.807*	−0.464	−0.269
pH	−0.095	−0.062	0.469	−0.534	0.157	−0.400
Sulphate	−0.887**	−0.841**	−0.954***	−0.782*	−0.683	−0.557
Zinc	−0.188	−0.025	−0.734	−0.590	−0.273	−0.060

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Vistula River. This was especially true at Tczew and Krakow where the trends over time are very similar. The water quality at Warsaw was designated as excellent for HWQI but marginal to good for AWQI. As a result, the DWQI fell between the two resulting in a fair to good rating.

To understand the temporal patterns in the indices, the variables that contributed to each index over time were examined (Tables 5 and 6). Assessment of the variables was conducted on a site-by-site basis and correlation analysis of the variables against the index value was conducted for each station over time.

Results

The correlation analysis conducted with raw data and the index values are presented in Tables 5 and 6 for all three indices. A summary of the results is presented on a site by site basis as follows.

Krakow (upstream)

The HWQI and AWQI had a number of variables that consistently exceeded guideline values at Krakow. Lead and cadmium exceeded guidelines in all years in the HWQI, with mercury exceeding between 1992 and 1995. Chloride, ammoniacal-N and sodium exceeded guidelines in all years in the AWQI with iron failing to meet the guideline between 1992 and 1996. When the raw data were compared statistically to the indices, ammoniacal-N, chloride, iron and sodium were all significantly correlated with AWQI (Tables 5 and 6). For HWQI, chromium, manganese and mercury were significantly correlated with the index. Although cadmium was not significantly correlated, it still demonstrated a good relationship with HWQI (Tables 5 and 6). When the two indices were combined into the DWQI similar significant results were observed (Tables 5 and 6).

Table 6 Pearsons correlation matrix for DWQI and HWQI against the contributing variables at Krakow, Warsaw and Tczew

Variable	Krakow (upstream)		Warsaw (midstream)		Tczew (downstream)	
	DWQI	HWQI	DWQI	HWQI	DWQI	HWQI
Cadmium	−0.686	−0.725	−0.578	−0.757*	−0.694	−0.883*
Chromium	−0.676	−0.807*	−0.809*	−0.931	−0.608	0.423
Copper	−0.364	−0.168	−0.749*	−0.820	0.104	0.049
Lead	−0.076	−0.099	−0.856**	−0.929***	−0.590	−0.800*
Manganese	−0.933***	−0.917***	−0.879**	−0.850**	−0.731	−0.590
Mercury	−0.855*	−0.864**	−0.690	−0.863	−0.887**	−0.763
Nitrate	−0.567	−0.830	0.722	LD	−0.353	−0.391
Nitrite	−0.571	−0.572	0.228	LD	−0.356	−0.428

LD lack of data to complete correlation analysis

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Warsaw (midstream)

Lead was the predominant variable in exceedance for HWQI, and, ammoniacal-N, iron, pH and chloride were consistently in exceedance for AWQI. Interestingly, no exceedances were measured past 1998 for any of the health variables which resulted in an excellent rating for HWQI.

Cadmium, lead and manganese were all significantly correlated with HWQI (Tables 5 and 6). The correlation with mercury was not significant; however, a good relationship was still observed ($r = -0.863$, $p = 0.137$). Ammoniacal-N, chloride and iron were all significantly correlated with AWQI (Tables 5 and 6). pH was not significantly correlated even though it was in exceedance of the maximum guideline at various times throughout the monitoring period; however, when the extent of exceedance is assessed, at no point does pH exceed 8.5. Therefore, the lack of significance with AWQI was probably reflective of the small deviations in pH values between 1992 to 2003. When the two indices were combined into the DWQI similar significant results were observed (Tables 5 and 6).

Tczew (downstream)

The variables that exceeded the guideline at Tczew were cadmium, lead and mercury which were all significantly correlated with the HWQI (Tables 5 and 6). AWQI did not show such a strong correlation to all of the contributing variables as in the previous sites i.e. only iron was significantly correlated (Tables 5 and 6). Mercury, manganese and cadmium were significantly correlated with DWQI (Tables 5 and 6). The correlation results for this index were not as reflective of the HWQI or AWQI as they were in the previous two sites.

Discussion

A clear spatial gradient was observed in concentrations of ammoniacal-N, sodium and chloride,

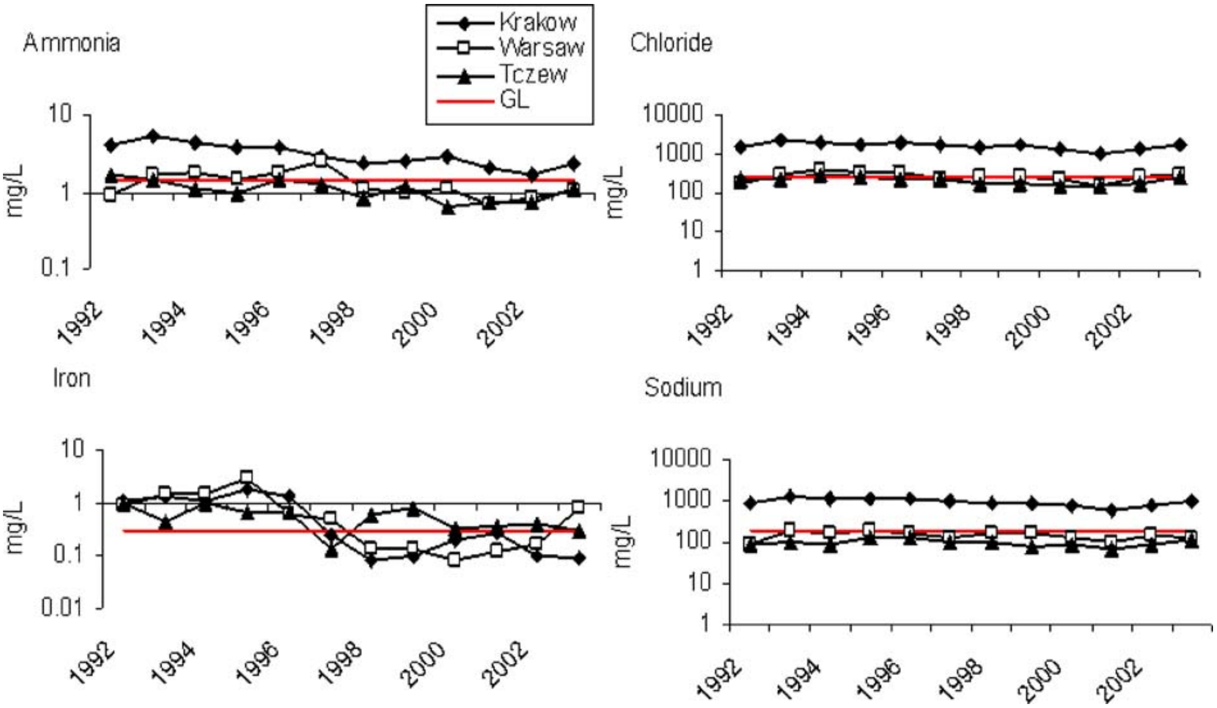


Fig. 5 Concentrations of ammoniacal-N, chloride, sodium and iron (mg/L) at three sites along the Vistula River (Tczew, Warsaw and Krakow) between 1992 and 2003. *Solid line* represents drinking water guideline

where the concentrations decreased from upstream to downstream; all three variables consistently exceeded their respective guideline at Krakow but not at the other two stations (Fig. 5). Furthermore, general decreases in the concentrations of iron and ammoniacal-N between 1992 and 2003 corresponded to increases in AWQI over the same time period at all three sites.

The spatial patterns in water quality in the Vistula River can be attributed to both anthropogenic and geological influences. High, but decreasing, ammoniacal-N concentrations at Krakow and, to a lesser extent, at Warsaw, most likely reflect municipal sewage contamination/inputs (Krzyzanowski and Slaski 2005; Buszewski and Kowalkowski 2003; Laenen and Dunnette 1997). In contrast, the major sources for chlorides and sulphates in the Vistula are both geological and anthropogenic (Laenen and Dunnette 1997). The Vistula River flows through industrialized and highly urbanized regions of Poland, and receives major inputs from industrial pollution and saline mine waters. For example, the hard coal mines in southern Poland discharge into the upper part of the Vistula River, upstream of Krakow. Kowalkowski et al. (2007) reported that the mines discharge approximately 9,000 tonnes of sulphates and chlorides per day, two thirds of which are carried off by the Vistula River. This would account for the elevated sodium and chloride concentrations observed at Krakow, which were well above guideline concentrations and approximately one order of magnitude higher than those observed at Warsaw and Tczew. Although sulphates did not exceed the guideline at any time or station, there was still a significant inverse relationship observed between sulphates and the AWQI and DWQI at Krakow and Warsaw (Tables 5 and 6). This would suggest that a decrease in overall pollutant loadings occurred at both a spatial (upstream to downstream) and temporal (1992 to 2003) level, which was reflected in the index values.

Discharge is likely to play a factor in downstream changes in water quality, since the concentrations of most chemical constituents is approximately inversely proportional to water flow (Laenen and Dunnette 1997). On an annual basis, AWQI was significantly correlated with discharge ($r = 0.772$, $p < 0.001$), suggesting that the improvement in water quality from upstream to downstream was, in part, due to a dilution effect. Chloride, sodium and sulphates, three of the main drivers of the AWQI, also were significantly negatively correlated to discharge (Fig. 6), supporting the dilution hypothesis.

Metals were the main variables that exceeded guidelines in the HWQI at all three sites (Fig. 6). HWQI tended to improve over time, which was in direct response to a reduction in concentrations of metals at all sites (i.e. cadmium and mercury decreased at all three sites, lead decreased at both Tczew and Warsaw, manganese decreased at both Krakow and Warsaw and chromium decreased at Krakow; Fig. 7). Temporally, it would seem

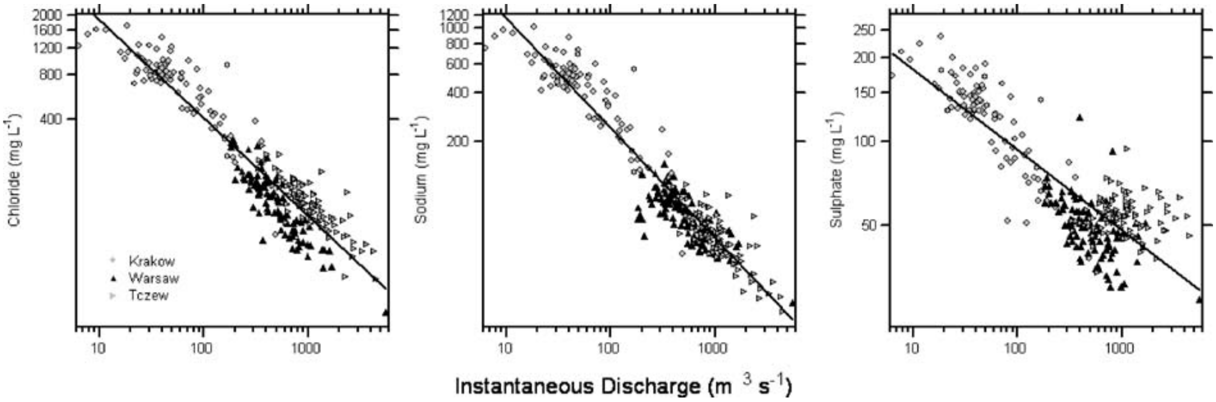


Fig. 6 Instantaneous discharge (m^3/s) and concentrations in sodium (mg/L), chloride (mg/L) and sulphate (mg/L) at all three sites (Krakow, Warsaw and Tczew) along the Vistula River between 2000 and 2003

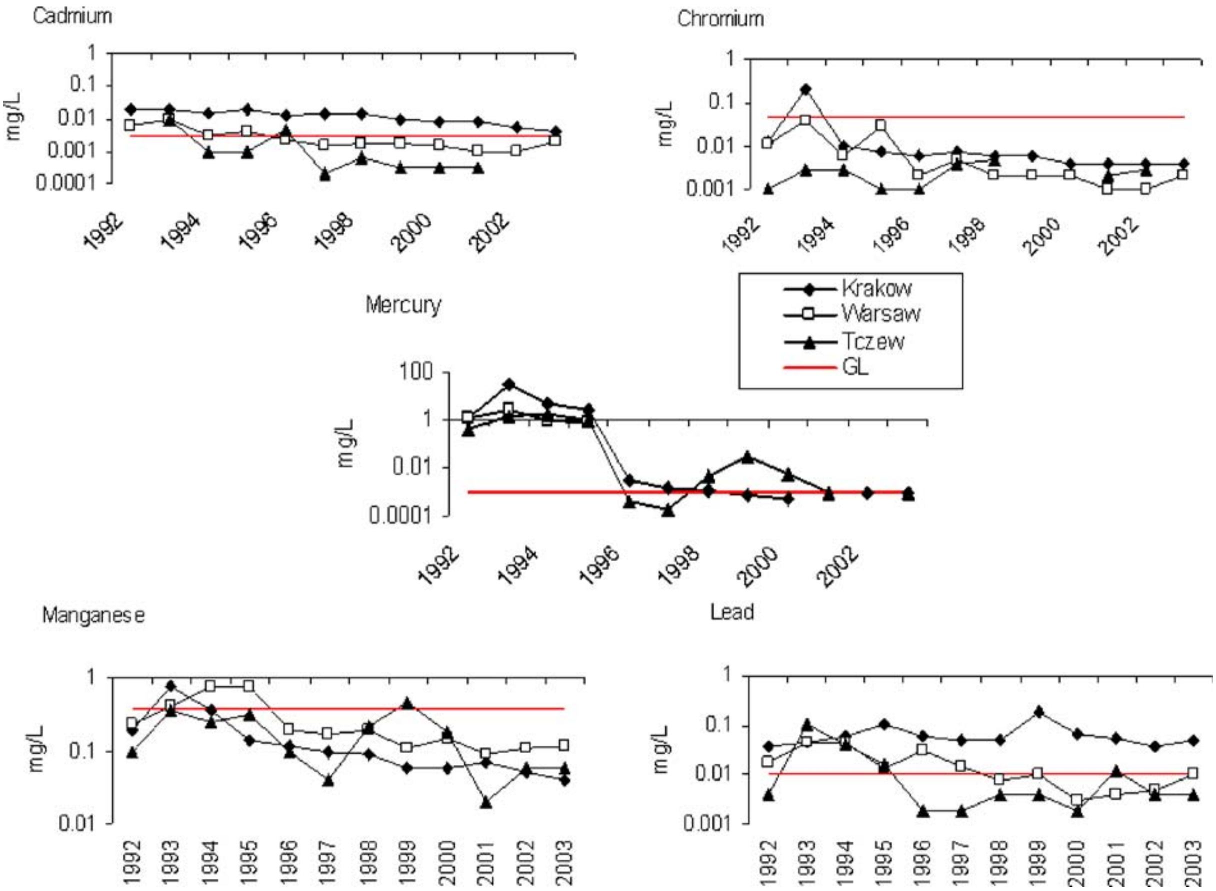


Fig. 7 Concentrations of cadmium, chromium, mercury, manganese and lead (mg/L) at three sites along the Vistula River (Tczew, Warsaw and Krakow) between 1992 and 2003. *Solid line* represents drinking water guideline

that metal contamination has been improving in the Vistula river at all three sites. This improvement could possibly be due to improved treatment or a reduction in the amount of industrial discharge. Spatially, both lead and cadmium decreased in concentration from upstream to downstream and consistently exceeded the guideline at the upstream site (Fig. 7). Again, this decrease in concentration could have been due to dilution. Similar to AWQI, HWQI was significantly correlated with discharge ($r = 0.711$, $p < 0.001$), suggesting that the major source of metals is originating from upstream. This corresponds to previous observations that suggest the main source for contaminants are the coal mines located in the upper part of the Vistula River basin, as well as the zinc and lead mines in the Przemsza River, a main tributary to the Vistula River upstream of Krakow (Gueguen and Dominik 2003). Even though the concentrations of various metals decreased from upstream to downstream, cadmium and lead still remained above the guideline value. High concentrations of heavy metals in water including cadmium, lead and chromium have been reported in the Bay of Gdansk, and were attributed to the direct discharge of the Vistula River (Pempkowiak et al. 2006). In addition, Beldowski and Pempkowiak (2007) have also reported the Vistula River is the main source of mercury into the Gdansk bay; our results correspond well with these observations. Overall, we could conclude that the HWQI was reflective of the spatial and temporal changes in concentrations of variables, specifically metals, from the upstream to downstream sites.

When the two indices were combined into the DWQI, similar results were observed at both Warsaw and Krakow, indicating DWQI was representative of AWQI and HWQI at these sites (Fig. 4, Tables 5 and 6). However, the temporal pattern of AWQI and HWQI were reflective of very different issues within the river. AWQI was reflective of the salinity and ammoniacal-N issues, and HWQI was reflective of metal contamination. By comparison only cadmium was significantly correlated with the DWQI at Tczew, which was not reflective of the AWQI and HWQI results. It would seem that when the indices are combined at this site, the correlation with variables was lost. This is a concern, especially if DWQI was the only index used to accurately reflect the status of water quality in the Vistula River. HWQI and AWQI, it seems, were more sensitive than the DWQI to pick up on the slight deviations in the variables reported at this site, where guideline exceedances were less extreme than those at the mid and upstream sites. Hence, the HWQI and AWQI were more reflective of the situation within the Vistula River at this site. This improved sensitivity justifies our decision to split the overall DWQI into two indices, the HWQI and AWQI.

In summary, the significant correlations between the variables in that exceeded guidelines and the indices related well to previous studies conducted within the Vistula River, with high metal content reflected in the HWQI, and, high chloride and ammoniacal-N content reflected in the AWQI. Most importantly the variables correlated well with their respective indices, suggesting that both HWQI and AWQI were truly reflective of the different variables included within their calculation. On the whole, the indices were a good reflection of water quality in the Vistula River. The DWQI, while corresponding well to the other two since it demonstrated similar temporal and spatial patterns, did not seem to be as sensitive or as descriptive as the HWQI or AWQI. This was especially true for the downstream site at Tczew, where the correlations of the DWQI to the respective variables were lower. The HWQI and AWQI were more reflective of the exceeded variables. This highlights the importance of the different indices. The DWQI can provide a very broad overview of the situation, but HWQI and AWQI were far more sensitive to deviations from the guideline and more descriptive in specific issues of concern. This is not to say that the DWQI was not useful, more that the decision to use either indices is dependent on the type of question or analysis that is being conducted.

One of the most interesting observations was the gradients in different variables from upstream to downstream. Specifically, certain metals, namely cadmium, copper, lead and zinc decreased in concentration from upstream to downstream with a corresponding increase in HWQI. Ammoniacal-N, sodium, sulphates and chloride also demonstrated a decrease from upstream to downstream with a corresponding increase in AWQI. When both AWQI and HWQI were assessed against instantaneous discharge, it was clear that the improvement in water quality from upstream to downstream, was, in part, due to a dilution effect.

Therefore, not only were these indices reflective of the real data on a temporal basis, but also on a spatial scale from upstream to downstream. In conclusion, this case study has demonstrated the usefulness and sensitivity of the indices developed and it is recommended that further development and application of the indices be made.

Summary and future directions

This paper outlines the development and application of one of the first water quality indices designed for use at the global scale. The need to develop a global water quality index was identified by UNEP as a priority for assessment and, as such, this study marks an important starting point for future work.

In the development of this report, we identified a number of advantages of not only the index itself but also the methodology used to calculate the index. Firstly, the index examines water quality on a station by station basis, giving a value for a specific year and location. This allows assessment both spatially, at the drainage basin, regional and national level, and temporally, on a station by station level if historical data are available. The index is also flexible in terms of the variables that are included within the calculation. For example, water quality monitoring variables could be chosen for inclusion in the index to reflect effects on aquatic biodiversity, or rates of eutrophication in inland waters, or simply to assess the overall health of an aquatic system. There are certain basic monitoring criteria that must be met: specifically, the variables must be measured consistently at a specific site and the variables measured should be related to a specific end-point for comparison (e.g. a benchmark or guideline concentration needs to be identified). Once variables have been selected and benchmark concentrations have been selected for comparison, the index calculation is fairly simple. The flexibility and ease of computation of the index is a great advantage for stakeholders worldwide, such as water managers using this index.

The index also examines deviations from globally recognized guidelines i.e. World Health Organisation Drinking Water Guidelines, and as such does not assume a preconceived notion of what is good. Most water quality indices rely on normalizing data variable by variable according to expected concentrations and some interpretation of 'good' versus 'bad' concentrations. For example, Pesce and Wunderlin (2000) normalized their data to a common scale, Stambuk-Giljanovik (2003) calculated a weighted average index from normalized values, and, Tsegaye et al. (2006) standardized values to the maximum concentration for each variable. The overall index values calculated in this report were directly related to set guidelines and, as such, are sensitive to deviations from these guidelines and reflective of real water quality conditions. Designations of index values into good or bad categories was attempted within this report, however these are illustrative and can be adapted on a case by case basis.

Although a significant effort has been made in the development of an index to assess global drinking water quality, there are a number of issues that need to be addressed for their future development.

Firstly, in the development of this index it was necessary to establish specific criteria for inclusion of variables into the calculation. A number of variables are outlined by the World Health Organisation as having either health or acceptability guidelines related to drinking water. However, it was not possible to include all of these variables as there were inconsistencies in measurements on a global basis. Because of this, we established certain criteria for inclusion of variables and monitoring data in the index. Specifically, the variables included had to have some representation in all the five major regions in the world (Europe, Asia, Oceania, Africa and Americas), and, stations within these regions were only selected if they measured at least four of these variables, four times a year. Of course, by including these rules we limited our analysis to only those monitoring stations that had adequate replication and appropriate geographical representation. This resulted in two issues. Firstly, the variables included in the index were not all encompassing of the variables outlined by WHO in their guidelines for drinking water. As such, certain priority pollutants may have been missed: for example in areas of intense agriculture, the inclusion of certain pesticides would be essential but were not included in this analysis. Secondly, because we accepted stations that had measured at least four variables out of the total variables selected, index calculations for certain stations would be based on variables that were not included in the index calculations for others. Although the variables included would be related to either health effects of acceptability issues, comparing among stations may not always be appropriate if the variables measured are very different. With these issues in mind, future development should include an assessment to prioritize pollutants to establish a framework for development of a DWQI, HWQI or AWQI for individual countries or areas. This would include establishing relevant variables that must be measured and included in the calculation of the index.

Secondly, the WHO guidelines used in this study are specifically for drinking water, and as such do not represent guidelines for source water. The data in GEMStat are predominantly measurements of natural surface waters or groundwater, and therefore there are limitations in the variables chosen. We only selected variables that were considered to have a human health or acceptability issue in water with no further treatment required. Variables such as biochemical oxygen demand (BOD) or dissolved oxygen (DO) are not recognised variables affecting drinking water. However, these variables may cause considerable problems if that water body is intended as a source to be treated for drinking since increased aeration or biological treatment of the water may be required to establish its suitability for drinking. As it stands, the use of the WHO guidelines for drinking water was a useful tool to assess water quality against stringent guidelines and will be useful as an assessment tool for situations where the water body is used as drinking water, with little or no treatment. It is recommended, however, that further development of the guideline include variables that will affect the quality of water as a source for drinking. In this regard, it would also be useful to gain expert opinion with regards to establishing treatment level categories for the index values, as opposed to the current good, fair or bad designations used in this report.

Finally, it was noted that the DWQI was not as sensitive in revealing temporal and spatial trends as the HWQI or AWQI. It was thought that the significance of individual variables were lost due to the higher number included within the indices, possibly obscuring information. If this is the case then it is recommended that future investigations compare different methods of calculation. For example, using the 'minimum operator' approach outlined by Smith (1989, 1990).

In conclusion, the drinking water quality index developed made a significant contribution to the issue of global water quality assessment. Validation and sensitivity analysis revealed the index was reflective of the real data obtained from GEMStat, and allowed assessment of water quality on both a temporal and spatial scale. A number of advantages of using this method were highlighted. In addition, important suggestions were made for the development of the index, which is now in progress.

Acknowledgements We would like to thank the following people for their valued suggestions and advice in the development of this manuscript. Richard Robarts, UNEP GEMS/Water Programme, Canada; Sabrina Barker, UNEP GEMS/Water Programme, Kenya; R.M. Bhardwaj, Ministry of Environment & Forests 'Parivesh Bhawan', India; Salif Diop, United Nations Environment Programme, Kenya; Oscar E. Natale, Instituto Nacional de Agua, Argentina; Harold van Niekerk, Department of Water Affairs and Forestry, South Africa; Alexander de Sherbinin, CIESIN, Columbia University, USA; Leendert Vijselaar, DACAAR, Afghanistan; Monique Dubé, University of Saskatchewan, Canada; Ilia Bounari, Division de la Qualité et Pollution des Eaux, Niger; Ali Abdul-Zahra Zebon Al-Lami Ministry of Environment, Iraq.

References

Beldowski, J., & Pempkowiak, J. (2007). Mercury transformations in marine coastal sediments as derived from mercury concentration and speciation changes along source/sink transport pathway (Southern Baltic). *Estuarine, Coastal and Shelf Science, 72*, 370–378. doi:10.1016/j.ecss.2006.10.007.

Buszewski, B., & Kowalkowski, T. (2003). Polands environment—past, present and future state of the environment in the Vistula and Odra river basins. *Environmental Science and Pollution Research, 10*, 343–349.

CCME (2001). Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, user's manual. In *Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg, Manitoba*. Retrieved on March 29, 2007, from http://www.ccme.ca/assets/pdf/wqi_usermanualfctsht_e.pdf.

CCME (2005). *Canadian environmental sustainability indicators. Freshwater quality indicator: Data sources and methods*. Catalogue no. 16-256-XIE. Retrieved on March 29, 2007, from http://www.statcan.ca/bsolc/english/bsolc?catno=16-256-XIE#formatdisp.

Cude, C. G. (2001). Oregon water quality index: A tool for evaluating water quality management effectiveness. *Journal of the American Water Research Association, 37*, 125–137.

Esty, D. C., Levy, M. A., Srebotnjak, T., de Sherbinin, A., Kim, C. H., & Anderson, B. (2006). *Pilot 2006 environmental performance index*. New Haven, USA: Yale Center for Environmental Law & Policy.

Gueguen, C., & Dominik, J. (2003). Partitioning of trace metals between particulate, colloidal and truly dissolved fractions in a polluted river: The Upper Vistula River (Poland). *Applied Geochemistry, 18*, 457–470.

Harkness, S. (2004). *Social and political indicators of human well-being*. United Nations University: World Institute for Development Economics Research (WIDER), Research paper No. 2004/33, i–22 (May 2004).

ISC (2005). *Index of stream condition: The second benchmark of Victorian River*. Victorian Government

Department of Sustainability and Environment, Melbourne, August 2005. Retrieved on March 29, 2007 from www.vicwaterdata.net.

Khan, F., Husain, T., & Lumb, A. (2003). Water quality evaluation and trend analysis in selected watersheds of the Atlantic region of Canada. *Environmental Monitoring and Assessment, 88*, 221–242.

Kowalkowski, T., Gaszala-kopciuch, M., Kosobucki, P., Krupczynska, K., Ligor, T., & Buszewski, B. (2007). Organic and inorganic pollution of the Vistula River basin. *Journal of Environmental Science and Health Part A, 42*, 421–426.

Krzyzanowski, D. A., & Slaski, M. W. (2005). *Case study: Vistula River—evaluation of the Vistula River catchment area management. Science and politics of international freshwater management 2003/04*. Zurich: Swiss Federal Institute of Environmental Science and Technology.

Laenen, A., & Dunnette, D. (1997). *River quality: Dynamics and restoration*. New York, USA: CRC.

Liou, S. M., Lo, S. L., & Wang, S. H. (2004). A generalised water quality index for Taiwan. *Environmental Monitoring and Assessment, 96*, 35–32.

Lumb, A., Halliwell, D., & Sharma, T. (2006). Application of the CCME water quality index to monitor water quality: A case study of the Mackenzie River Basin, Canada. *Environmental Monitoring and Assessment, 113*, 411–429.

Nagels, J. W., Davies-Colley, R. J., & Smith, D. G. (2001). A water quality index for contact recreation in New Zealand. *Water Science and Technology, 43*, 285–292.

Pempkowiak, J., Walkusz-Miotik, J., Beldowski, J., & Walkusz, W. (2006). Heavy metals in zooplankton from the Southern Baltic. *Chemosphere, 62*, 1697–1708.

Pesce, S. F., & Wunderlin, D. A. (2000). Use of water quality indices to verify the impact of Cordoba City (Argentina) on Suquia River. *Water Research, 34*, 2915–2926.

Smith, D. G. (1989). A new form of water quality index for rivers and streams. *Water Science and Technology, 21*(2), 123–127.

Smith, D. G. (1990). A better water quality index for rivers and streams. *Water Research, 10*, 1237–1244.

Stambuk-Giljanovik, N. (2003). Comparison of Dalmation water evaluation indices. *Water Environment Research, 75*, 388–405.

Statistics Canada (2002). *Gross domestic product by industry: Sources and methods*. Catalogue no. 15-547-XIE. Ottawa, Canada: Industry Measures and Analysis Division, Statistics Canada.

Tsegaye, T., Sheppard, D., Islam, K. R., Johnson, A., Tadesse, W., & Atalay, A. (2006). Development of chemical index as a measure of in-stream water quality in response to land-use and land cover changes. *Water, Air, and Soil Pollution, 174*, 161–179.

UNEP GEMS/Water Programme (2007). *Report: Global drinking water quality index development and sensitivity analysis*. United Nations Environment Programme GlobEnvironment al Monitoring System (GEMS)/Water Programme. Retrieved on January 9, 2007, from www.gemswater.org.

WHO (2004). *Guidelines for drinking-water quality (3rd ed., Vol. 1): Recommendations*. Geneva: World Health Organisation.