Estimating averages and detecting trends in water quality data

L. SOMLYÓDY, J. PINTER, L. KONCSOS, I. HANACSEK
Water Resources Research Centre (VITUKI),
H-1453 Budapest, P.O.Box 27, Hungary
I. JUHASZ
National Environmental Protection Office (OKTH),
H-1365 Budapest, P.O.Box 732, Hungary

ABSTRACT The estimation of temporal averages and trends are significant elements of elaborating water quality management strategies. This paper reviews some relatively simple statistical techniques in the context of shallow lake eutrophication control. The techniques are illustrated by analyzing measurement data from the Lake Balaton region of Hungary.

INTRODUCTION

Water quality monitoring networks have been established and gradually extended during the last decades. It has been recognized that the usual uniform observation strategies often will not properly and cost effectively describe the changes of water quality. In other words, too few or too many measurements might be taken when monitoring objectives are not properly considered. General experience shows that these objectives are often rather loosely defined.

The objectives of water quality surveillance systems may be quite diverse. The most important objectives are the following ones:

(a) Process description in time at given monitoring station (mainly for research purposes);
(b) Estimation of averages, e.g. yearly or seasonal average pollutant and nutrient loads of rivers, lakes, etc.;
(c) Detection of water quality trends (a primary objective at national level);
(d) Analogues of (a) and (b) for characterizing spatial changes of water quality; and
(e) Joint consideration of temporal and spatial aspects, e.g. if we are interested in determining the total annual load of a lake which is carried by several tributaries of different dynamics.

It is noted that when solving the above listed tasks - which are often interdependent - a certain loss of information is allowed. Hence uncertainties are important elements of design and operation of monitoring networks.

In this paper we focus on problems (b) and (c) by showing the usefulness of relatively simple, existing statistical techniques. Objectives (a), (d) and (e) are discussed by Somlyody (1986), Pinter et al. (1985) and Pinter & Somlyody (1986) respectively. The methodology summarized in this paper will be illustrated by results related to the region of Lake Balaton, Hungary.
AVERAGE ESTIMATION

In this section well-known statistical results (Bayley & Hammersley, 1946; Cochran, 1962) are combined to yield mean estimates for finite samples.

Method

Consider a finite (total) sample of identically distributed, but not necessarily independent elements $y_1, \ldots, y_N$ with mean $\bar{y}_N = 1/N \sum_k y_k$; let $\sigma$ denote the variance of the individual sample elements (being the realizations of some random variable). Suppose that a random $n$-sample ($1 \leq n \leq N$) $y_1, \ldots, y_n$ is taken with replacement from the total sample. Then the random variable

$$\bar{y}_n = \frac{1}{n} \sum_{k=1}^n y_{ik}$$

is an unbiased estimator of $\bar{y}_N$, while its variance equals

$$\sigma_n = \frac{\sigma}{n} \left( \frac{N-n}{N} \right)$$

where

$$\sigma = \frac{1}{N-1} \sum_{k=1}^N (y_k - \bar{y})^2$$

The factor $(N-n)/N$ in (2a) is a finite sample size correction of the respective well known result from probability theory. Note that (2) is valid only when $y_1, \ldots, y_N$ are independent. In the case of dependent sample elements

$$\sigma^* = \sigma \left[ 1 + 2 \sum_{t=1}^N (1 - t/N) r(t) \right]$$

replaces $\sigma$ in (2), where $r(t)$ is the $t$-lag autocorrelation coefficient of the total sample. In other words, the effective sample size of dependent observations $y_1, \ldots, y_N$ equals $Na/\sigma^*$ (Bayley & Hammersley, 1946).

Assume now that a certain estimation error $\Delta y = \sigma \bar{y}_N$ is acceptable. More precisely, one postulates that the error should remain less than $\Delta y$ with a high probability $p$. Then assuming that the central limit theorem is applicable, one obtains the approximation

$$P\left( \left| \bar{y}_n - \bar{y}_N \right| < \Delta y \right) = P\left[ \bar{y}_N(1 - \alpha) < \bar{y}_n < \bar{y}_N(1 + \alpha) \right]$$

$$= P\left( \frac{\alpha \bar{y}_N}{\sqrt{\sigma_n}} < \frac{\bar{y}_n - \bar{y}_N}{\sqrt{\sigma_n}} < \frac{\alpha \bar{y}_N}{\sqrt{\sigma_n}} \right) = \Phi \left( \frac{\alpha \bar{y}_N}{\sqrt{\sigma_n}} \right) - \Phi \left( - \frac{\alpha \bar{y}_N}{\sqrt{\sigma_n}} \right)$$

$$= 2 \left( \frac{\alpha \bar{y}_N}{\sqrt{\sigma_n}} \right) - 1 = p$$

Hence $\alpha \bar{y}_N/\sqrt{\sigma_n} = \phi^{-1}\left( (1 + p)/2 \right)$, where $\phi_p$ is the respective $(1 + p)/2$ quantile of the normal probability distribution function $\phi$. 


Consequently we have (Somlyody, 1984)

\[
n = \frac{N}{1 + N(\alpha)^2 \frac{\bar{y}_N^2}{\sigma}} \tag{5}
\]

Equation (5) shows that the necessary sample size is a monotonically decreasing function of both the relative variance \( \bar{y}_N/\sigma \) and the ratio \( \alpha/\sigma \). Note that for dependent samples \( \sigma \) is to be replaced by \( \sigma^* \), cf. equation (3).

Applications

The above simple observations lead to a number of useful conclusions concerning existing monitoring strategies. Consider, for example, the total phosphorus load of River Zala which is to be estimated with, say, 20% relative error (\( \alpha = 0.2 \)). Then, by (5), if \( \alpha = 0.2 \) is to be assured for shorter time periods (i.e. \( N \) is smaller), then the necessary sample size \( n \) will significantly increase. Table 1 illustrates this point. The length of considered time periods and respective sample sizes per year are shown (in the example \( p = 0.95 \), i.e. \( y_p \approx 1.96 \)). The data of the table are based on daily measurements taken in 1976-1979.

<table>
<thead>
<tr>
<th>Time period</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976-1979</td>
<td>17</td>
</tr>
<tr>
<td>a year</td>
<td>55</td>
</tr>
<tr>
<td>a given month (in each year)</td>
<td>108</td>
</tr>
<tr>
<td>a single month</td>
<td>156</td>
</tr>
</tbody>
</table>

According to the results above, 17 observations per year are sufficient for estimating long-range yearly means. The yearly average can be reasonably well estimated (\( \alpha = 0.2 \)) with 55 samples per year, while monthly means need 156 samples per year, i.e. sampling every second day. Note that the cited results are based on unavoidable model inaccuracies such as the supposition of random sampling with replacement.

While accepting the above generally valid conclusions, the stratification (seasonality) features may be very pronounced. For example, according to the varying monthly dynamics of the nutrient (pollution) load process, the necessary number of monthly measurements may vary significantly: this remark implies that stratified sampling will, in general, lead to cost effective
improvements of monitoring policies. This point is treated in
detail by Pinter et al. (1985), therefore we only remark here that
the introduction of stratified sampling requires detailed preliminary
measurement programmes, on the base of which the strata and the
respective stratified sample characteristics can be well estimated.
For illustrating this, note that e.g. monthly streamflow/pollution
load characteristics may vary largely between subsequent years,
therefore the determination of proper stratum characteristics is
rather difficult, even in the presence of up-to-date automatic
monitoring devices.

As the streamflow rate and respective nutrient load are, as a
rule, strongly related, monitoring strategies can also be based on
streamflow data. For example, on the basis of River Zala data we
found that the monthly relative deviations $\sigma / \overline{N}$ of total phosphorus,
nitrogen and suspended sediment can be related to those of flow.
For example

$$\left( \frac{\sigma}{\overline{N}_{TP}} \right) \approx 1.65 \left( \frac{\sigma}{\overline{N}_{Q}} \right)$$

(correlation factor $r = 0.88$)

It is also easy to see that for estimating $P$ and $N$ load averages with
an acceptable error of e.g. 25%, it is necessary to observe the
respective streamflows with a maximal error of 15%. It is noteworthy
that streamflow based monitoring leads to meaningful results, if non-
point source pollution and surface runoff are the primary components
of the watershed nutrient load processes. Point-source pollution
adds to the load average, while (being basically permanent) it
usually decreases its relative variance: hence, our method outlined
above usually results in a safe overestimate of the errors.

Consider now the effect of $\alpha$ on the necessary sample size. For
simplicity's sake, let $N = 365$ (daily measurements), and let
$\overline{N}/\sigma = 1$ (this ratio is rather reasonable for TP and TN loads in our
case). Then for $p = 0.95$ we have

$$n(\alpha) = \frac{365}{1 + 365(\frac{\alpha}{1.96})^2}$$

From (6) one obtains that $n(0.2) \approx 76$, $n(0.5) \approx 15$, $n(0.7) \approx 8$, i.e.
for mean estimations the sample size may be very significantly
reduced.

Uncertainties in mean estimates
As expected, less frequent observations lead to increasing
uncertainties. By (5), the error of the mean estimation equals

$$\alpha = \frac{\sigma}{\overline{N}} \left( \frac{N - n}{nN} \right)^{\frac{1}{2}}$$

Figure 1 shows the effects of different sampling strategies on the
yielded errors. The displayed results are based on Monte Carlo
simulations (Somlyody, 1984) and are in good agreement with the
theory (Cochran, 1962). The figure reflects the fact that the
estimation error of TP decreases in a hyperbolic manner, as a
function of the sample size $n$: hence with small samples a slight increase in sampling effort will yield substantial improvements in accuracy. As a matter of course, the above observations are valid not only in connection with the sampling criteria of TP load estimation, but - in a qualitative sense - also for other water quality components.

DETECTION OF WATER QUALITY TRENDS

This section is devoted to several statistical tests for recognizing trends and their applications (Lettenmaier, 1976; Pinter & Koncsos, 1985).

Method

In most cases trends are modelled by some step function (representing an abrupt water quality change) or a linear function (modelling gradual, uniform change of water quality). There exist several known procedures, such as Student's $t$-statistic, Mann-Whitney's and Spearman's tests, which are suitable theoretical tools for detecting such trends. It is worth while to emphasize, however, that the above classical tests are directly applicable only for analyzing independent time series data: therefore suitable extensions are necessary for handling the dependent observations which occur in practice (Bayley & Hammersley, 1946; Lettenmaier, 1976).

The probability of recognizing existing trends is shown by the power of the respective statistics. Based on the investigation of Lettenmaier (1976), for step trends this power depends basically on...
$N_T = t_r \sqrt{N/(2\sigma)}$: here $t_r$ is the step size (to be recognized), $N$ is the effective total sample size and $\sigma$ is the standard deviation as estimated from the sample, supposing that the trend indeed occurs. A similar relationship holds for linear trends: $N_T = t_r N/(2\sqrt{3}\sigma)$ will determine the power of the statistics, where $t_r = N t_Q$ is the overall magnitude of the linear trend to be detected. According to the Monte Carlo results of Lettenmaier (1976), in the case of dependent autocorrelated observations, the total sample size $N$ is to be replaced by

$$N^* = N/[1 + 2 \sum_{t=1}^{N} (1 - t/N)r(t)]$$

Equation (8) implies that for strongly autocorrelated observation sequences $N^*$ may be significantly less than $N$. For example if $r(t) = 0.2^t$ as in the AR(1) model, then $N^*$ varies between 0.65 $N$ and 0.75 $N$. Consequently the power of the mentioned statistical tests becomes smaller than for independent observations: that, in turn, leads to increasing error in recognizing trends.

It is to be emphasized that this situation cannot be rectified by simply increasing the sampling frequency as this leads to increasing autocorrelation. Actually, as it is not difficult to show (Lettenmaier, 1976), there exists a maximum effective number of independent samples, $N_{\text{max}}$, which may be collected in a specified time period $N_O$. For AR(1) models with $r(t) = \rho^t$ ($0 < \rho < 1$), there holds

$$N_{\text{max}}^* = N_O \frac{(\ln \rho)^2}{2 \rho N_O - N_O \ln \rho - 1}$$

(9)

Using (9), one can state that the power of the above statistics is bounded by $t_r N_{\text{max}}^*/(2\sigma)$ for step trend and by $t_r N_{\text{max}}^*/(2\sqrt{3}\sigma)$ for linear trends. If we specify a priori the test confidence levels as, say, 95%, then $N_T = 4$ and $N_T' = 4$ would assure the "safe" recognition of existing trends. Given the finiteness of $N_{\text{max}}^*$ there follows that only sufficiently large trends can be recognized with an acceptable likelihood - whatsoever is the actual sample size! This fact implies that detection of trends is to be declared cautiously, always specifying the probability of possible misjudgement.

Applications

For illustrating the outlined theoretical results, several numerical examples are detailed below: these examples are not connected with actual measurement data.

Denote by $J$ and $n$ the number of observed years and the number of (independent!) observations per year. Then the recognition of linear trends depends primarily on
Estimating averages and detecting trends

\[ N^*_T = \frac{t_r}{\sqrt{\sigma}} \sqrt{\frac{N}{12}} = \frac{t_r}{\sqrt{\sigma}} \sqrt{\frac{Jn}{12}} \]  

(10a)

whence

\[ Jn = 12 N^*_T \frac{(\sqrt{\sigma})^2}{t_r} \]  

(10b)

As noted before, given a confidence level for assuring low probability mis-specification of non-existent trends, \( N^*_T \) unambiguously defines the power of the test, i.e. the probability of detecting existing trends. Therefore, depending on the ratio \( t_r/\sigma \) and the desired test power, the necessary number of observations can be determined as shown in Table 2. Consider

<table>
<thead>
<tr>
<th>( t_r/\sqrt{\sigma} )</th>
<th>Confidence level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.15</td>
<td>1200</td>
<td>4800</td>
<td>10 800</td>
<td>19 200</td>
</tr>
<tr>
<td>0.2</td>
<td>0.45</td>
<td>300</td>
<td>1200</td>
<td>2 700</td>
<td>4 800</td>
</tr>
<tr>
<td>0.5</td>
<td>0.80</td>
<td>48</td>
<td>192</td>
<td>432</td>
<td>768</td>
</tr>
<tr>
<td>1.0</td>
<td>0.98</td>
<td>12</td>
<td>48</td>
<td>108</td>
<td>192</td>
</tr>
</tbody>
</table>

some examples from this table. If \( t_r/\sqrt{\sigma} = 0.1 \), then for almost sure trend recognition 19 200 (indepeddent!) sample elements are required, and even for detecting the trend with a low probability of 0.15 a sample size of 1200 would be needed. Further, for \( t_r/\sqrt{\sigma} = 0.5 \) (a rather high value) at least 432 observations (i.e. weekly observations for about 8 years!) would be necessary for recognizing the trend. These findings show that (a) trend detection often requires far more observations than average estimations (b) probably the standard bi-weekly or monthly measurements are insufficient to judge existence or non-existence of trends, unless they are extremely marked (see also Schilperoort & Groot, 1983). Considering the fact that in practice after some years of a given trend of water quality evolution, the pattern may change, it may be very hard (if not impossible) to detect short range water quality changes. The frequent dependence of subsequent observations increases the above methodological difficulties: hence, great care is to be taken when interpreting statistical statements about the occurrence of trends and their relative magnitude.
CONCLUSIONS

In this paper several interrelated issues of management oriented water quality monitoring were analyzed. Based on the summarized theoretical background and the presented examples, the following conclusions can be drawn:

(a) Uncertainties (both inherent and monitoring implied) play a key role in the analysis of existing monitoring networks. Therefore, even relatively simple concepts and methods of mathematical statistics can provide valuable insight into standard monitoring practice, making it possible to propose significant improvements.

(b) For estimating yearly averages of different water quality components or pollutant loads, standard monitoring results can be well used in most cases. In principle, stratification could significantly improve the consideration of seasonality, hence leading to cost effective sampling procedures; however, its proper realization is not easy because of the year-to-year stochastic changes of the hydrological regime.

(c) The statistically sound detection of trends needs, as a rule, more observations than estimates of average (and more than taken in most monitoring networks operated at present). Due to the underlying statistical background of trend detection, in many practical cases there exists an upper bound on the number of available independent observations, and hence a lower bound on the size of reliably observable trends.

REFERENCES


