

Variations and Spurious Correlations Related to DIN, DIP, TN, TP, DIN/DIP and TN/TP in Lakes, Rivers, and Marine Systems

Maria I. Stenström-Khalili and Lars Håkanson*

Department of Earth Sciences, Uppsala University, Villav. 16, SE-752 36 Uppsala, Sweden

Abstract: Inherent variations and uncertainties in empirical data from aquatic systems ultimately constrain approaches to predictions and possibilities to identify critical thresholds and points of no return. This work addresses coefficients of variations (CV) for total phosphorus (TP), total nitrogen (TN), dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP), DIN/DIP and TN/TP. We have collected data from 28 lakes, 34 rivers, and 38 stations in the Baltic Sea, Kattegat, the Danish Straits and Skagerrak. Key questions are: How do the CV-values vary among lakes, rivers and marine systems? Are there patterns in the CVs related to the trophic conditions? Are there seasonal patterns in the CVs? Selected results: DIN varies most of the studied nutrients. DIP displays a relatively homogenous variation with CVs of about 0.3 to 0.5. The variability in DIN governs the high uncertainty in the DIN/DIP-ratio (CV = 0.8 to 1.4). The CVs for TN/TP are much lower (0.4 to 0.5). The CVs for DIN and TN increase from the Bothnian Bay to the Skagerrak. There is a pattern of lower variabilities in oligotrophic lakes for DIN/DIP as compared to TN/TP. Results from an experiment to illustrate spurious correlations (using randomly generated data) are also given, which clearly show that for low-productive systems, one can find almost any ratio of DIN/DIP or TN/TP. This implies that one must be very cautious in interpretations of what DIN/DIP or TN/TP actually represent in low-productive systems.

Keywords: Coastal systems, Lakes, Rivers, Coefficients of variation, Phosphorus, Nitrogen, DIN, DIP, Redfield ratio, Sampling.

INTRODUCTION

Substances in the water column can be divided into two main parts, the dissolved phase and the particulate phase, relating to their fates and transport routes (pelagic versus benthic). The distribution (= partition = partitioning) coefficients of substances depend on the association to suspended particulate matter (SPM; see [1]). Particulate substances are, by definition, subject to gravitational sedimentation. Hence, they are to a high degree retained within a given system (e.g., a coastal area) and affect benthic habitats. The dissolved fraction, on the other hand, is more related to the pelagic pathways. Operationally, the particulate fraction is generally defined as the non-filterable remains on a filter. For such determinations, one would often use a pore size of 0.45 μm or pore sizes in the range from 0.2 to 0.9 μm (e.g., [2-4]). Filtration is often a justifiable method from many ecological and mass-balance modelling perspectives. Substances bound to colloids (i.e., bound to particles smaller than approximately 0.45 μm) are, hence, often operationally included in the dissolved fraction, although they are not truly dissolved in a chemical sense. Often, chemical fractions, such as phosphate, nitrite, nitrate, ammonium, DIN, DIP, DOC, etc. would often not correspond to the dissolved fraction as determined from filtration (DF = 1 - PF). This means that if SPM and the particulate fraction (PF) are operationally determined from filtration, the dissolved fraction derived from the chemical fractions (e.g., DIN = nitrate + nitrite + ammonium) is not the same as DF.

It is well established [5] that plankton cells have a typical atomic composition of $\text{C}_{106}\text{N}_{16}\text{P}$, which means that 16 times as many atoms (and 7.2 times as many grams) of N are needed than of P to produce phytoplankton. A general rule is that nutrient limitation is decided by the Redfield ratio, R, estimated from:

$$R = \text{TN/TP} \text{ or } R = \text{DIN/DIP} \quad (1)$$

TN = concentration of total nitrogen (mg/l)

TP = concentration of total phosphorus (mg/l)

DIN = concentration of dissolved inorganic nitrogen (mg/l)

DIP = concentration of dissolved inorganic P (mg/l).

One well-known and general approach to describe the affinity of all types of substances to carrier particles is by means of the partition coefficient, K_d (l/g dw; dw = dry weight). K_d is generally defined as the ratio of filter-retained to filter-passing concentrations as:

$$K_d = (C_{\text{part}}/\text{SPM})/C_{\text{diss}} \quad (2)$$

where SPM is the suspended particulate matter concentration (g dw/l), C_{diss} is the dissolved (filter-passing) concentration (g dw/l) and C_{part} is the particulate concentration (g dw/l). Physically, K_d describes the particle affinity and represents the equilibrium of numerous processes such as sorption onto particulate matter, precipitation and dissolution [6]. Examples of substances for which K_d have been determined are trace metals [7-10], organic micropollutants [11, 12], phosphorus [13] and radionuclides [14, 15]. There are, to the best of our knowledge, no operationally and well-tested algorithms available to predict K_d or the particulate fraction (PF) for nitrogen in aquatic systems.

*Address correspondence to this author at the Department of Earth Sciences, Uppsala University, Villav. 16, SE-752 36 Uppsala, Sweden; E-mail: Lars.Hakanson@geo.uu.se

The uncertainties of all variables evidently influence the statistical reliability of the data (e.g., the mean value and the standard deviation) and the predictive power of models. The coefficient of variation (CV = standard deviation/mean value) varies among different variables. For example, CV for PF varies less than CV for K_d and PF is therefore more suitable in, e.g., predictive mass-balance models [1]. On average CV for K_d is 3.0 times larger than CV for PF [16].

DIN and DIP are generally poor predictors in contexts of nutrient limitation and in modelling of primary production due to their inherent uncertainty demonstrated by their high CVs in comparison to TN and TP [17, 18]. Table 1 gives introductory results (from [19]) from one bay in the Baltic Proper (Himmerfjärden) on this matter and it shows the CVs for DIN, DIP, TN and TP. Table 2 gives the number of samples of different fractions of N and P needed to estimate the annual mean with an error of 15 percent in the Himmerfjärden Bay, calculated from Eq. 3 (the sampling formula; see [20]). If the variability within an ecosystem is large, many samples must be at hand to obtain a given level of certainty in the mean value. Eq. 3 is derived from the basic definitions of the mean value, the standard deviation and the Student's t value, which expresses how many samples are required (n) in order to establish a mean value with a specified certainty:

$$n = (t \cdot CV/L)^2 + 1 \quad (3)$$

where t = Student's t, which specifies the probability level of the estimated mean (usually 95%; strictly, this approach is only valid for variables from normal frequency distributions). CV is the coefficient of variation within a given ecosystem. L is the level of error accepted in the mean value. For example, L = 0.1 implies 10% error so that the measured mean will be expected to lie within 10% of the expected mean with the probability assumed in determining t. Since one often determines the mean value with 95% certainty ($p = 0.05$), the t-value is often set to 1.96. The data in Table 2 for the Himmerfjärden Bay is based on the CV-values in Table 1 and Eq. 3. To determine the mean DIN-value with a 15% accepted error (L), one would need 154 samples! Typically, there would be 1 to 5 monthly samples available in many regular monitoring programs, so the actual uncertainty in the mean value would be very high. This restricts all interpretations of changes in the system often related to costly remedial measures as well as interpretations of co-variations and factors influencing variations in DIN (and other variables with high inherent CVs), and how such variations influence target bioindicators (such as chlorophyll, Secchi depth and cyanobacteria).

Paper [17] showed that concentrations of inorganic nutrients in the water may be low also when the supply is

high. The turnover rate of bioavailable nutrients is high and low levels of dissolved inorganic nutrients can be found even in highly productive waters. [17] also suggested that only when the levels of DIN are much higher than the levels of DIP (e.g., 100:1), it is unlikely that DIN is limiting and only if $DIN/DIP < 1$, it is unlikely that P is the limiting nutrient. It is also concluded that DIN and DIP are poor predictors of nutrient status in aquatic systems compared to TN and TP.

From this background, we would like to stress that the main objective of this work has been the conduction of a data-mining to compile CV-values for TN, TP, TN/TP, DIN, DIP, DIN/DIP from as many lakes, rivers and coastal systems as possible covering a wide gradient in trophic conditions (from oligotrophic to hypertrophic) to be able to provide a more general framework for analyses related to the very important concept of "limiting" nutrient. So, this work will specifically focus on the CV-values and patterns in CVs, because we are convinced that this would provide a more solid scientific foundation for analyses related to the dilemma to find the "limiting" nutrient, whether such analyses are done using statistical or causal methods.

Due to the definition of TN/TP, DIN/DIP and K_d as ratios, spurious correlations may contribute to observed correlations between these variables and environmental variables (e.g., [21-23]). This means that regressions to varying degrees may overestimate correlations with variables that are included or closely related to any of the variables in the given ratio. Here, we will illustrate spurious correlations, a fundamental problem in all analyses where ratios, such as DIN/DIP or TN/TP, are involved.

An important aspect related to spurious correlations concerns to the fact that both nutrients (N and P) appear in plankton cells. This means that one generally finds a marked co-variation between phosphorus and nitrogen concentrations in aquatic systems (see [24] and Eq. 4 from [25]):

$$\log(TN) = 0.70 \cdot \log(TP) + 1.61 \quad (4)$$

$(r^2 = 0.88; n = 58 \text{ coastal systems})$

The following section will first present the data used in this study, and the next section will give results on CV-values for the studied variables. The last section will present an experiment to illustrate the importance of spurious correlations in contexts where ratios, such as TN/TP, DIN/DIP and K_d , are used.

DATA

We have carried out a "data-mining" to access data on TN, TP, DIN and DIP from many possible sources and the aim has been to cover data from lakes, rivers and marine

Table 1. Mean Coefficients of Variation (CV) Calculated for the Himmerfjärden Bay from 1997 to 2006 (from [19])

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean
TN	0.13	0.11	0.14	0.16	0.15	0.12	0.10	0.09	0.10	0.11	0.14	0.24	0.13
DIN	0.30	0.26	0.47	1.49	1.20	1.52	1.27	1.50	1.39	0.99	0.59	0.42	0.95
TP	0.10	0.10	0.15	0.24	0.23	0.18	0.16	0.13	0.21	0.28	0.24	0.19	0.18
DIP	0.10	0.12	0.47	0.92	0.58	0.51	0.68	0.61	0.90	0.63	0.30	0.20	0.54

systems with a focus on the Baltic Sea, Kattegat and Skagerrak. For overall comparison, we have focused on data from the surface-water layer (< 10 m) and data from the growing season (May to September). For the conditions in the Baltic Sea, we have targeted on mean monthly data to get information also on seasonal variations.

Table 2. Number of Samples Required to Determine the Mean Value with an Error (L) of 15 Percent of the Mean (with a 95% Certainty) Using Data from the Himmerfjärden Bay and the Sampling Formula (from [19])

	CV	n
TP	0.18	6
DIP	0.54	50
TN	0.13	3
DIN	0.95	154

From Table 3, which gives a compilation of the collected data, one can note that we have data from 28 lakes, 34 rivers sites, 20 measuring stations in the Baltic Sea and 18 stations outside the Baltic Sea in the Kattegat, The Danish Straits and Skagerrak. There are, e.g., 4963 individual data from these lakes, 3934 data from the river sites, 4172 data from the Baltic Sea and 5175 data from the Kattegat and Skagerrak. The data from Swedish lakes and rivers emanate from a data bank hosted by the Swedish University of Agricultural Sciences; data from the Baltic Sea and the Skagerrak, Kattegat and The Straits are from three different sources: SMHI [26], ICES [27] and Himmerfjärden [28]. The sampling programs in the Baltic Sea typically have a monthly data sampling frequency of 1 to 8 samples. The monitoring in Swedish lakes typically has a monthly data sampling frequency of 1 to 5 samples.

These data will be presented in the following text and the basic aim has been to try to find general patterns in the CV-values. How do the CVs vary among lakes, rivers and marine systems? Are there patterns in the CVs related to the trophic conditions? Are there seasonal patterns?

Table 4 gives a compilation of data from the individual lakes and marine areas. The messages in these data will be discussed in the subsequent section.

Table 3. CV-values for TN, TP, DIN, DIP, TN/TP and DIN/DIP from Different Aquatic Systems (Lakes, Rivers and Marine Systems)

	Period	CV _{TN}	CV _{DIN}	CV _{TP}	CV _{DIP}	CV _{DIN/DIP}	CV _{TN/TP}	n
All lakes, 28 lakes	1987-2006	0.24	0.64	0.43	0.61	0.84	0.41	4963
Lakes, CV _{DIN/DIP} -CV _{TN/TP} > 0.10 excluded, N=23	1987-2006	0.28	0.80	0.45	0.59	0.92	0.41	4250
Rivers, N=34 stations	1987-2006	0.36	0.70	0.48	0.64	0.87	0.49	3934
Baltic Sea, N=20 stations	1987-2001	0.12	0.57	0.28	0.68	1.11	0.36	4171
Skagerrak and Kattegat, N=18 stations	1987-2001	0.23	1.15	0.38	0.73	1.43	0.42	5175

The CVs in this table represent a compilation of data from Table 4. So, the CV of 0.24 for TN for the lakes represents the mean value for all 28 lakes in Table 4. For each individual lake, the CVs have been calculated from individual data from the growing season (May to September). The total number of individual data (n) for lakes is 4963 [from 36] and those data have been sampled in the period from 1987 to 2006.

RESULTS

Variations within and Among Systems

From the data in Table 3, one can first conclude that DIN is the nutrient fraction that varies most throughout the growing season in lakes, rivers, the Baltic Sea and in the Kattegat/Skagerrak area. The maximum variation in DIN is found in Skagerrak and smaller variations in lakes. DIP displays a relatively homogenous variation with CVs of about 0.59 to 0.73 in the different aquatic systems. From this, one can conclude that the variability in DIN seems to govern the high uncertainty (CV = 0.8 to 1.4) in DIN/DIP (Table 3). The CVs for TN/TP are much lower (0.4 to 0.5).

The CVs for DIN and TN seem to increase steadily (from 0.25 and 0.08 to 1.16 and 0.23, respectively) from the Bothnian Bay, the Bothnian Sea, and the Baltic Proper to the Skagerrak (see Table 4, lower part).

The relatively low CVs for TN and TP (0.08 and 0.27) in the Bothnian Bay might be related to the fact that the Bothnian Bay represents a rather enclosed and homogenous system opposite to the conditions in many rivers, the southern part of the Baltic Sea and the Kattegat/Skagerrak areas.

Distribution of Nutrient Fractions and Trophic Status of Lakes

From the more detailed results given in Table 4, one can see a clear pattern of lower variabilities in oligotrophic lakes for DIN/DIP as compared to TN/TP; the mean difference between these two CVs for the oligotrophic lakes is only 0.05. The same mean difference is 0.46 for mesotrophic lakes, 0.57 for eutrophic lakes, and as high as 0.85 on average for the marine systems. Comparing the CVs for DIN/DIP and TN/TP and selecting the lakes where the difference between the two ratios exceeds 10 percent result in an exclusion of all oligotrophic lakes in the data set and a clear threshold at the limit between mesotrophic and oligotrophic lakes (Table 3).

Paper [29] found the highest values of TN/TP in marine microalgae in oligotrophic waters and the lowest values in culture studies using nutrient-rich waters. The same principles seem to be valid for lakes, as shown in Fig. (1). We will discuss interpretations of the results in Fig. (1) more closely in a following section concerning spurious correlations. The point here is that the results in Fig. (1) are based on empirical data, so the scatter seen in Fig. (1) is real.

However, there are always problems to relate ratios (e.g., DIN/DIP in Fig. 1) to factors directly or indirectly related to DIN or DIP, such as chlorophyll in Fig. (1).

One can conclude that there are clear and interesting patterns in the CVs among these systems, as shown by the results in Tables 3 and 4.

Seasonal Variations in CVs in the Baltic Sea

The sampling program in the Himmerfjärden Bay on the Swedish side of the Baltic Proper is probably one of the most

comprehensive long-term monitoring program for coastal areas in the Baltic Sea [30-32]. The data from the Himmerfjärden Bay have been used here to exemplify the variation in the monthly error (L) for the mean values for DIN/DIP and TN/TP (see Table 5 and Fig. 2).

From Fig. (2), one can note the seasonal patterns for the error (L, calculated from the CVs in Table 6 and the number of measurements each month). The CVs for DIN/DIP in this bay are very high during the growing season with peak values of about 2 in April and October (see Table 6). The

Table 4. Values on Trophic Level and Mean CV-Values Using Data from the Growing Season (May-Sep.) from Swedish Lakes, the Baltic Sea and the Danish Straits from 1987-2006

Lake	Trophic		CV _{TN}	CV _{DIN}	CV _{TP}	CV _{DIP}	CV for DIN/TN	CV for DIP/TP	CV _{DIN/DIP}	CV _{TN/TP}	CV _{DIN/DIP} / CV _{TN/TP}	n	Mean CV _{DIN/DIP} / CV _{TN/TP}
	Status	Chl											
Vättern, Edeskovarnaån	Oligo	1	0.11	0.06	0.31	0.53	0.73	0.27	0.35	0.36	-0.01	132	
Vättern, Jungrun	Oligo	1	0.12	0.06	0.42	0.66	0.72	0.33	0.43	0.34	0.09	132	
Tjulträsk	Oligo	1	0.36	0.43	0.38	0.49	0.19	0.29	0.62	0.52	0.10	97	
Härsvatten	Oligo	1	0.25	0.37	0.47	0.55	0.48	0.28	0.49	0.53	-0.04	201	
Vänern, Tärnan	Meso	2	0.20	0.43	0.38	0.50	0.73	0.25	0.39	0.30	0.09	151	0.05
Vänern, Megrundet	Meso	2	0.34	0.77	0.46	0.52	0.71	0.25	0.45	0.27	0.18	149	
Allgjuttern	Meso	2	0.22	0.90	0.51	0.66	0.13	0.25	1.14	0.49	0.65	207	
Stensjön	Meso	2	0.24	0.64	0.36	0.50	0.12	0.29	0.80	0.44	0.36	287	
Tväringen	Meso	3	0.36	0.43	0.38	0.49	0.10	0.33	1.02	0.47	0.55	108	
Remmarsjön	Meso	3	0.23	0.52	0.32	0.52	0.11	0.27	0.70	0.42	0.28	199	
Övre Skärsjön	Meso	3	0.21	0.41	0.35	0.58	0.40	0.29	0.76	0.34	0.42	216	
Stora Skärsjön	Meso	4	0.34	1.24	0.46	0.59	0.24	0.25	1.21	0.51	0.70	225	
Bränträsket	Meso	4	0.26	0.49	0.30	0.46	0.09	0.27	0.63	0.37	0.26	118	
Mäsen	Meso	4	0.22	0.68	0.42	0.53	0.08	0.14	0.86	0.41	0.45	122	
Brunnsjön	Meso	4	0.28	0.62	0.60	0.70	0.26	0.31	0.64	0.43	0.21	293	
Hagasjön	Meso	5	0.32	0.77	0.43	0.52	0.29	0.21	1.31	0.46	0.85	126	
Skärgölen	Meso	5	0.28	0.87	0.40	0.53	0.06	0.21	1.06	0.44	0.62	138	
Humsjön	Meso	6	0.32	1.31	0.45	0.47	0.22	0.20	1.34	0.43	0.91	98	0.46
Mälaren, Prästfjärden	Eu	6	0.33	0.72	0.39	0.59	0.32	0.22	0.68	0.29	0.39	151	
Fiolen	Eu	6	0.22	0.84	0.45	0.72	0.20	0.15	0.92	0.40	0.52	286	
Fräcksjön	Eu	6	0.24	0.58	0.30	0.50	0.47	0.25	0.98	0.43	0.55	208	
Bysjön	Eu	7	0.31	0.93	1.05	1.56	0.18	0.21	1.11	0.37	0.74	141	
Hjälmaren	Eu	8	0.28	1.11	0.50	0.53	0.14	0.32	0.80	0.35	0.45	76	
Mälaren, Blacken	Eu	10	0.30	0.46	0.37	0.52	0.52	0.32	0.53	0.30	0.23	316	
Stora Envättern	Eu	11	0.26	1.20	0.44	0.58	0.14	0.18	1.32	0.45	0.87	288	
Rotehogstjärnen	Eu	11	0.23	1.12	0.49	0.51	0.18	0.18	1.03	0.40	0.63	284	
Storasjön	Eu	16	0.28	0.91	0.44	0.53	0.07	0.17	1.12	0.42	0.70	104	
Älgsjön	Eu	16	0.23	0.80	0.38	0.56	0.06	0.19	0.74	0.51	0.23	110	0.57
Bothnian Bay	Oligo	2	0.08	0.25	0.27	0.62	0.14	0.11	0.65	0.41	0.24	486	
Bothnian Sea	Oligo	2	0.14	0.7	0.31	0.87	0.05	0.16	1.32	0.36	0.96	1022	
Himmerfjärden	Meso	3	0.11	1.48	0.18	0.66	0.02	0.12	1.53	0.15	1.38	516	
Baltic Proper	Meso	3	0.14	0.74	0.25	0.54	0.02	0.28	1.36	0.31	1.05	2663	
Kattegat & Sounds	Meso	3	0.24	1.13	0.39	0.74	0.06	0.26	1.28	0.44	0.85	4346	
Skagerrak	Oligo	2	0.23	1.16	0.37	0.72	0.04	0.18	1.58	0.4	1.17	829	0.85

CVs for TN/TP are lower and do not show the same temporal pattern (the average monthly CV is 0.17; see Table 6). This means that the error is very large for DIN/DIP, approaching 300% in October and L is higher than 200% in several summer months (Fig. 2). This is logical: The high CVs for DIN/DIP during the growing season reflect the fact that DIN and DIP represent the bioavailable fractions of the nutrients participating in fast and dynamic reactions concerning biouptake and retention of the nutrients (typical turnover times for phytoplankton are in the order of 2.5 to 4 days; see [33]). The bacterioplankton will decompose the dead phytoplankton and this will release (regenerate) the bioavailable forms. Typical turnover times for bacterioplankton are 2 to 3.5 days [33]. Total phosphorus and total nitrogen determined from water samples include by definition TP and TN in dead and living phytoplankton and bacterioplankton. So, TN and TP in the water is the pool for DIN and DIP. The transfer of DIN to TN and DIP to TP and vice versa *via* biouptake and remobilisation is, thus, very quick.

Table 5. Error (L, see Eq. 3) of the Mean Value Calculated by Using the CV-Values from the Baltic Sea (1987-2001) and with a Sampling Frequency Equal to the Sampling Frequency of the Himmerfjärden Bay (Accessible to Us)

Month	n	L _{DIN/DIP}	L _{TN/TP}
1	1	0.66	0.45
2	1	0.51	0.36
3	3	0.56	0.23
4	4	0.88	0.35
5	3	1.08	0.32
6	2	1.43	0.40
7	2	1.23	0.33
8	2	1.64	0.41
9	3	1.08	0.30
10	2	1.09	0.37
11	2	0.61	0.35
12	1	0.57	0.47

The error curves for the two ratios for the data from the Baltic Sea are shown in Fig. (3), as a comparison to the results from the Himmerfjärden Bay in Fig. (2). The CV-values for DIN and DIP are also very high in the Baltic Sea, as shown in Table 4, and the seasonal pattern in the error is quite similar in the two figures, because the CVs are high in the summer and the number of analyses are relatively low.

If the CVs are high, many samples are required to obtain reliable mean or median values (Eq. 3). Table 5 is meant to demonstrate this for the CVs for DIN/DIP and TN/TP. This table also gives the number of data from Himmerfjärden Bay accessible to us and the error (L) in the mean. Generally, one would like to have reliable mean values with errors smaller than, say 15% to 25% of the mean. In this monitoring program, for DIN/DIP, the error is generally over 100% in

the growing season, with the maximum value of 164% for DIN/DIP in August. This has evident bearings on the conclusions that may be drawn related to the role of the DIN/DIP-ratio in this bay, and the same principles are applied to all aquatic systems!

Table 6. Monthly CV-Values for TN, DIN, TP, DIP, DIN/DIP and TN/TP in Himmerfjärden Bay

Month	CV _{TN}	CV _{DIN}	CV _{TP}	CV _{DIP}	CV _{DIN/DIP}	CV _{TN/TP}
Jan	0.13	0.30	0.10	0.10	0.57	0.12
Feb	0.11	0.26	0.09	0.12	0.54	0.13
Mar	0.14	0.47	0.15	0.47	1.05	0.16
Apr	0.16	1.49	0.24	0.92	2.01	0.27
May	0.15	1.20	0.23	0.58	1.38	0.19
Jun	0.12	1.52	0.18	0.51	1.62	0.17
Jul	0.10	1.27	0.16	0.68	1.53	0.11
Aug	0.09	1.50	0.13	0.61	1.58	0.12
Sep	0.10	1.39	0.21	0.90	1.52	0.16
Oct	0.11	0.99	0.28	0.63	1.90	0.24
Nov	0.14	0.59	0.24	0.30	0.77	0.17
Dec	0.24	0.42	0.19	0.20	0.62	0.20

Fig. (4) gives similar results based on the CVs for DIN/DIP and TN/TP using data from the Baltic Sea. From the CVs in Table 4, one would generally require about 8 samples to determine the mean TN/TP-ratio with an error (L) smaller than 15%, and 250 samples to get the same reliability in the mean value for DIN/DIP in August. Few (if any) regular monitoring programs that we know are structured to account for the natural variabilities in the measured variables expressed by their CV-values. This is a major flaw and it means that it is often difficult to use the data for the basic purpose of monitoring, e.g., to detect whether significant ongoing changes take place or not and to carry out the relevant actions if this is called for.

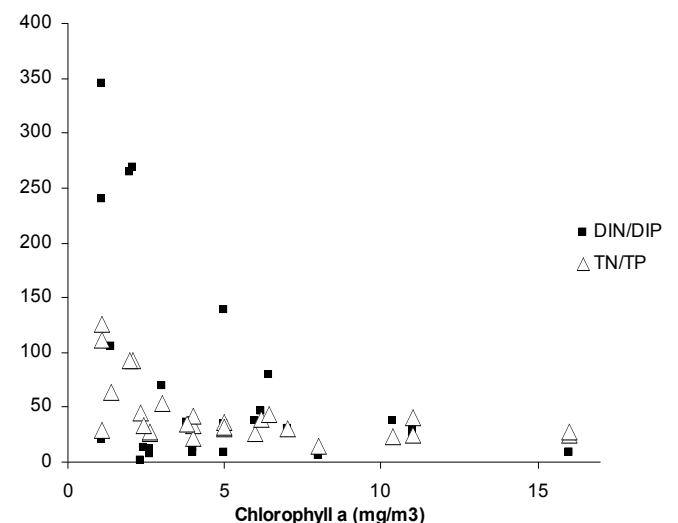


Fig. (1). DIN/DIP and TN/TP plotted on the y-axis against chlorophyll using data from the 28 lakes.

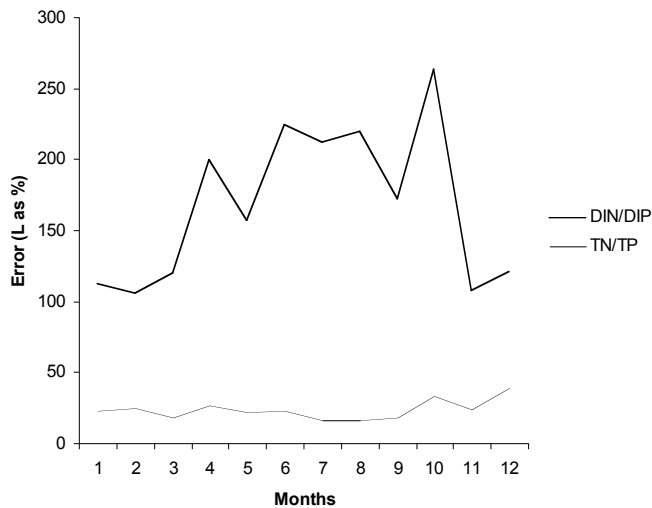


Fig. (2). Error (L) of the mean monthly value for DIN/DIP and TN/TP calculated from the number of samples (n) from the Himmerfjärden Bay.

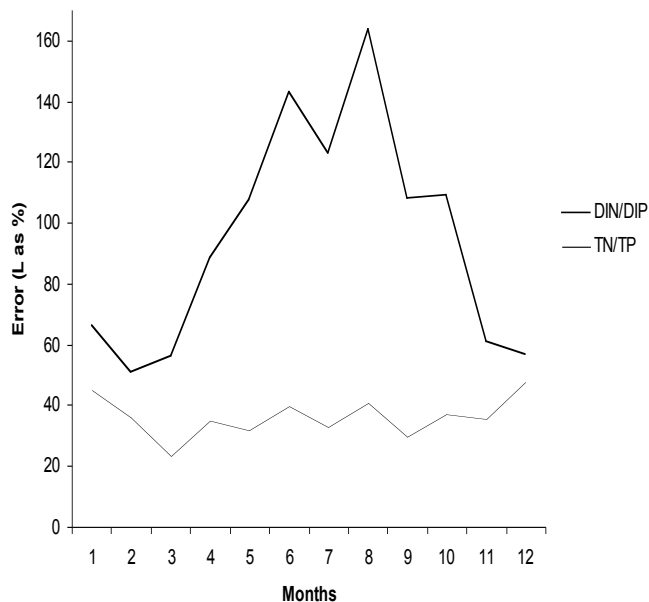


Fig. (3). Error (L) of mean for DIN/DIP and TN/TP using data from the Baltic Sea, 1987 to 2001.

Ratios and Spurious Correlations

Results from an experiment to illustrate the relationship between any arbitrary ratio (u/x) and x are given in Fig. (5). Here, two independent series of 10344 data (n) called x and u have been randomly generated, and Fig. (5) shows the relationship between the ratio u/x on the y-axis and x. The equation of the best-fit regression line is given in Fig. (5) and the obtained coefficient of determination (r^2) is 0.57 (which is highly significant; $p < 0.001$) for these randomly generated data. The figure clearly shows that for situations/ systems with low x- values, e.g., low-productive systems, one can find almost any ratio of, e.g., DIN/DIP or TN/TP. It also implies that one must be very cautious in interpretations of what DIN/DIP or TN/TP actually represent in low-productive systems with low values of DIN, DIP, TN and TP. This is true for all substances, but of greatest importance for substances with high inherent CV-values, such as DIN,

DIP and DIN/DIP. This is true also in interpretations of the relative role of different ratios for variations in, e.g., concentrations of cyanobacteria and chlorophyll and in regressions involving ratios.

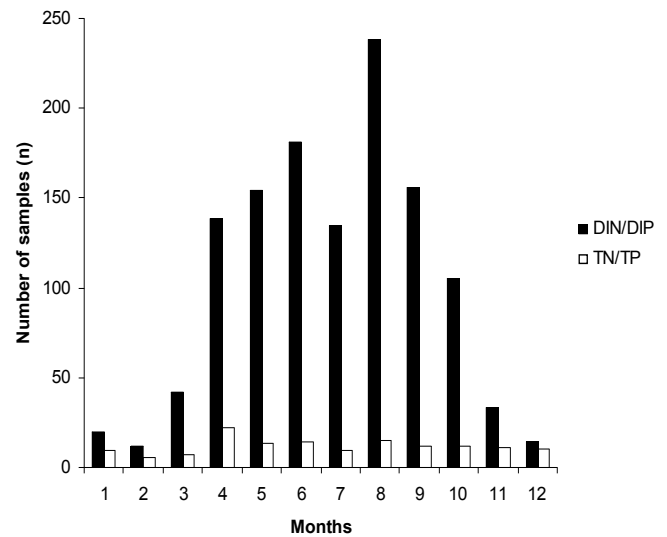


Fig. (4). Number of samples required to calculate monthly mean values of DIN/DIP and TN/TP if the error (L) is set to be 15 percent of the mean value (with a 95% certainty). Calculated using the data from the Baltic Sea, 1987 to 2001.

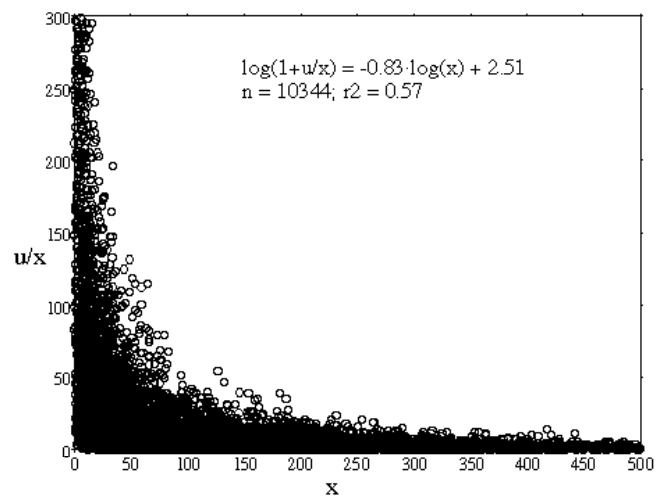


Fig. (5). Results from a random number experiment to illustrate the relationship between an arbitrary ratio (u/x) and an arbitrary x-variable. The figure gives the best-fit regression line. This experiment is based on 10344 (n) randomly generated data which show log-normal normal frequency distributions and ranges similar to phosphorus and nitrogen in aquatic systems. The coefficient of determination (r^2) is 0.57 between u/x and x (from the regression in the diagram) for these random data.

CONCLUDING COMMENTS

TN, TP, TN/TP and DIN/DIP are key variables in practically all contexts of aquatic sciences and management dealing with eutrophication, primary production, the issue of “limiting” nutrient, and predictions of standard bioindicators (such as the concentration of chlorophyll-a, cyanobacteria, oxygen concentrations and variables expressing water clarity). The focus in this work is set at the ecosystem scale,

i.e., on the conditions in entire coastal areas and lakes and on seasonal/monthly changes (rather than on variations at smaller temporal scales) and on comparative studies (lakes, rivers and marine areas). So, the focus is on the conditions in larger areas, which may be reflected in the data from several sample sites. The ecosystem scale is also the scale of main interests in water management and in discussions on remedial measures and strategies.

This work has presented CV-values and patterns in CV-values. These results may be considered as a stepping stone in future causal analyses and interpretations on the reasons for these patterns in variations, e.g., the very high CVs for DIN during the summer period, the very high CVs for DIN/DIP as compared to TN/TP, and the very high CVs for DIN/DIP in the marine areas as compared to the oligotrophic lakes. In the literature, there are many papers discussing how the relationship between DIN/DIP and TN/TP vary among systems of different trophic level neglecting the hazards of spurious correlations [34]. We have demonstrated that relationships involving ratios may be inherently and mathematically spurious, which means that such data and results based on such data should be regarded with due reservations. High CVs imply that one must take many samples (n , "small n ") to obtain a representative mean value for the given system, and it also means that one must take samples from many systems (N , "large N ") in regressions where the aim is to find more generally how a given x -variable with a high CV-value may influence a target y -variable, which could also have a high inherent CV-value [35].

Given the inherently high CV-values of many of these key variables in contexts of eutrophication studies, it must be stressed that more samples than generally taken in most regular monitoring programs are needed, if scientific unassailable conclusions are to be made concerning interrelationships among the variables and to produce scientifically meaningful information to detect critical ecosystem changes and threshold values.

REFERENCES

- [1] Håkanson L. Suspended particulate matter in lakes, rivers and marine systems. The Blackburn Press: New Jersey 2006; p. 331.
- [2] Seritti A, Petrsini A, Ferrara R, Barghigiani C. A contribution to the determination of reactive and total mercury in seawater. *Environ Technol Lett* 1980; 1: 50.
- [3] Bloom NS, Effler SW. Seasonal variability in the mercury speciation of Onondaga lake (New York). *Wat Air Soil Pollut* 1990; 53: 251-65.
- [4] Mason RP, Reinfelder JR, Morel FMM. Bioaccumulation of mercury and methylmercury. *Wat Air Soil Pollut* 1995; 80: 915-921.
- [5] Redfield AC, Ketchum BH, Richards FA. The influence of organisms on the composition of sea-water. In: Hill N, Ed. *The Sea* 2. Interscience, New York 1963; pp. 26-77.
- [6] Salomons W, Förstner U. *Metals in the Hydrocycle*. Springer, Heidelberg 1984; 349 p.
- [7] Honeyman BD, Santschi PH. Metals in aquatic systems. *Environ Sci Technol* 1988; 22: 862-871.
- [8] Balls PW. The partition of trace metals between dissolved and particulate phases in European coastal waters: a compilation of field data and comparison with laboratory studies. *Neth J Sea Res* 1989; 23: 7-14.
- [9] Benoit G. Evidence of the particle concentration effect for lead and other metals in fresh waters based on ultra-clean technique analyses. *Geochim Cosmochim Acta* 1995; 59: 2677-2687.
- [10] Turner A. Trace-metal partitioning in estuaries: importance of salinity and particle concentration. *Mar Chem* 1996; 54: 27-39.
- [11] Turner JV. In: Eriksson E, Ed. Hydrogeochemical evolution and variability of groundwater in experimental basins of S.W. Western Australia. Hydrochemical balances of freshwater systems. IAHS-AISH Publication 1984; No 150: pp. 41-54.
- [12] Zhou JL, Fileman TW, Evans S, *et al.* The partition of fluoranthene and pyrene between suspended particles and dissolved phase in the Humber Estuary: a study of the controlling factors. *Sci Tot Environ* 1999; 244: 305-321.
- [13] Håkanson L. Water pollution - methods and criteria to rank, model and remediate chemical threats to aquatic ecosystems. Backhuys Publishers, Leiden 1999; p. 299.
- [14] Santschi PH, Honeyman BD. In: Vernet JP, Ed. Radioisotopes as tracers for the interactions between trace elements, colloids and particles in natural waters. Heavy metals in the environment, Elsevier, Amsterdam 1991; pp. 229-246.
- [15] Carroll J, Harms IH. Uncertainty analysis of partition coefficients in a radionuclide transport model. *Wat Res* 1999; 33: 2617-2626.
- [16] Johansson H, Lindström M, Håkanson L. On the modelling of particulate and dissolved distributions of substances in aquatic ecosystems – sedimentological and ecological interactions. *Ecol Modelling* 2001; 137: 225-240.
- [17] Dodds KW. Misuse of inorganic N and soluble reactive P concentrations to indicate nutrient status of surface waters. *J N Am Benthol Soc* 2003; 22: 171-181.
- [18] Håkanson L, Bryhn AC, Hytteborn JA. On the issue of limiting nutrient and predictions of bluegreen algae in aquatic systems. *Sci Tot Environ* 2007; 379: 89-108.
- [19] Khalili M. Salt, water and nutrient fluxes to Himmerfjärden bay. Master thesis, Dept of Earth Sciences, Uppsala Univ. 2007.
- [20] Håkanson L. Sediment sampling in different aquatic environments: statistical aspects. *Water Resour Res* 1984; 20: 41-46.
- [21] Kenney BC. Beware of spurious self-correlations! *Wat Resour Res* 1982; 18: 1041-1048.
- [22] Jackson DA, Harvey HH, Somers KM. Ratios in aquatic sciences: Statistical shortcomings with mean depth and the morpoedaphic index. *Can J Fish Aquat Sci* 1990; 47: 1788-1795.
- [23] Krambeck HJ. Application and abuse of statistical methods in mathematical modelling in limnology. *Ecol Model* 1995; 78: 7-15.
- [24] Wallin M, Håkanson L, Persson J. Load models for nutrients in coastal areas, especially from fish farms (in Swedish with English summary). *Nordiska ministerrådet* 1992: 502, Copenhagen 1992; p. 207.
- [25] Håkanson L, Eklund J. A dynamic mass balance for phosphorus fluxes and concentrations in coastal areas. *Ecol Res* 2007; 22: 296-320.
- [26] SMHI 2007. [updated 2007 Sep 04] Available from: www.smhi.se
- [27] ICES 2007. [updated 2007 Sep 04] Available from: www.ices.dk
- [28] Himmerfjärden 2007. [updated 2007 Sep 04] Available from: www2.ecology.su.se/dbHFJ/index.htm, 2007-09-04
- [29] Geider RJ, La Roche J. Redfield revisited: variability of C: N: P in marine microalgae and its biochemical basis. *Eur J Phycol* 2002; 37: 1-17.
- [30] Elmgren R, Larsson U, Eds. Himmerfjärden. Förändringar i ett näringsbelastat kustecosystem i Östersjön. Swedish EPA, Stockholm 1997.
- [31] Larsson U, Hajdu S, Walve J, Andersson A, Larsson P, Edler L. *Bedömningsgrunder för kust och hav. Växtplankton, näringsämnen, klorofyll och siktdjup*. 2006.
- [32] Boesch, D, Hecky R, O'Melia C, Schindler D, Seitzinger S. *Eutrophication of Swedish Seas. Report 5509*, Swedish EPA, Stockholm 2006
- [33] Håkanson L, Bouillon V. *The Lake Foodweb - modelling predation and abiotic/biotic interactions*. Backhuys Publishers, Leiden 2002; p. 344.

- [34] Jeppesen E, Søndergaard M, Jensen SP, *et al.* Lake responses to reduced nutrient loading – an analysis of contemporary long-term data from 35 case studies. *Freshwater Biology* 2005; 50: 1747-1771.
- [35] Håkanson L. A data reduction exercise to detect threshold samples for regression models to predict key water variables. *Int Rev Hydrobiol* 2007; 92: 84-97.
- [36] SLU 2007. [updated 2007 Sep 04] Available from: www.ma.slu.se

Received: November 4, 2008

Revised: December 3, 2008

Accepted: December 3, 2008

© Stenström-Khalili and Håkanson; Licensee *Bentham Open*

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.