Naturhistoriska riksmuseet

Some consequences using pooled samples versus individual samples and pooled samples with various relation between sampling error and uncertainty due to chemical analysis

Överenskommelse 212 0840 Diarienr 235-1775-08Mm

SWEDISH · MUSEUM · OF · NATURAL · HISTORY
The Department of Contaminant Research
P.O. Box 50007
SE-104 05 Stockholm

Some consequences using pooled samples versus individual samples and pooled samples with various relation between sampling error and uncertainty due to chemical analysis

Anders Bignert,
Department of Contaminant Research, Swedish Museum of Natural History.

Summary
Choosing an appropriate sampling strategy for environmental monitoring or storing samples in Environmental Specimen Banks for future analyses includes the important decision of using individual or pooled samples. A number of advantages using individual samples for temporal trend studies can be identified e.g. that information about sample variance is important in it self and changes in variance is often the first sign of a change in contaminant burden, freedom of choosing an appropriate central measure (for right skewed distributions i.e. geometric mean values or medians) whereas pooled samples will represent arithmetic means. Furthermore, individual sampling enables adjustments for confounding factors (e.g. fat content, age, size) and detection of extreme values. However, resources may be saved by using pooled samples, if the sample variance is dominated by small-scale differences in time or space or by genetic and/or physiological differences among individual biological samples rather than of instrumental errors at the chemical analyses. The statistical power at temporal or spatial studies is determined by the random/unexplained sample variation. The relation between the instrumental error and other sources of variation as well as the relation between the cost for chemical analysis and collection and preparation of samples will determine the number of individual samples in each pool and the number of pools that should be analysed to achieve high cost efficiency. Various scenarios of different number of individual samples, different number of pooled samples containing various number of individual specimens, different relation between instrumental error and other sources of sample variance have been compared by simulating random sampling from computer generated populations using realistic measures of variation from ongoing monitoring activities. The results may give guidance to the selection of a cost efficient sampling strategy

Introduction
Temporal trend studies within the national Swedish monitoring programmes for contaminants in biological samples have used a number of individual samples collected each year at the same station during the same sampling season.

A number of advantages using individual samples can be identified (e.g. Bignert *et al*, 1993):

1. Information about sample variance is important in it self and changes in variance is often the first sign of a change in contaminant burden, more sensitive than mean values. Furthermore, information about the maximum value could also be essential for a risk analysis (See Fig.1).

2. Freedom of choosing an appropriate central measure. Pooled samples will approximate arithmetic means. For right skewed distributions, geometric mean values are more appropriate.
3. Individual sampling enables adjustments for confounding factors (e.g. fat content, age, size) and detection of extreme values.

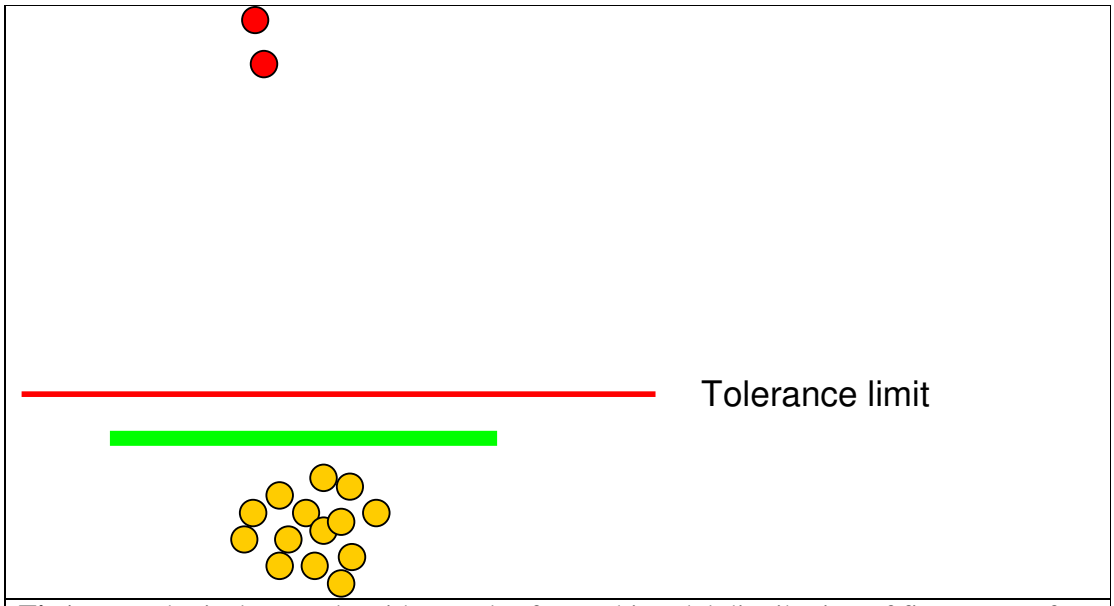


Fig1. Hypothetical example with samples from a bimodal distribution of figs, some of which is contaminated with e.g. aflatoxins hazardous for health. This risk can easily be disguised in a pooled sample, that may show a value below the accepted tolerance limit.

However, in cases where sampling and sample preparation are considerably cheaper than analytical costs and where the contribution from biological and physiological factors are much larger than the analytical error to the total random/unexplained variation, variation may be reduced at the same cost if pooled samples are used.

A reduction of the random/unexplained between-year variation implies that a smaller annual change can be detected or that a specified lowest trend that needs to be detected will be detected in a shorter period of time (Bignert *et al.*, 2004).

Methods

Samples from computer generated populations of normally or log-normally distributed values with added variance simulating both variances from sampling (genetic, physiological differences etc) and from the uncertainty at chemical analysis, were taken randomly for various combinations of magnitude from these sources. Sampling were repeated 10 times to simulate 10 years. Furthermore, the n of individual samples in a pool, n of pools per year and various distributions were also varied at the sampling. The whole procedure was repeated 1000 times and mean values were calculated. A regression line was fitted to the mean values (see Fig. 2 – 6).

Results and discussions

In Fig. 2, the probable between-year variation achieved at different sampling size (individual samples) are simulated. A total variance of, CV= 50% (normal distribution) was used. This is slightly higher than the average CV encountered from the time-series from the national monitoring program of biological samples from the marine environment Hg (35%)and CB-153 and TCDD-eqv (both 42%), about the same as for Pb (54%) and lower than for e.g. Cd (Bignert et al, 2008). The more individual samples that are taken each year the lower the random between-year variation can be expected. In the existing programs, 10 (dioxins, spring samples, cod and perch samples) -12 (all other) individual samples are collected and analyzed. According to the diagram, the expected variance (CV) is about 14% for 12 individual samples and slightly more than 15% for 10 individual samples per year.

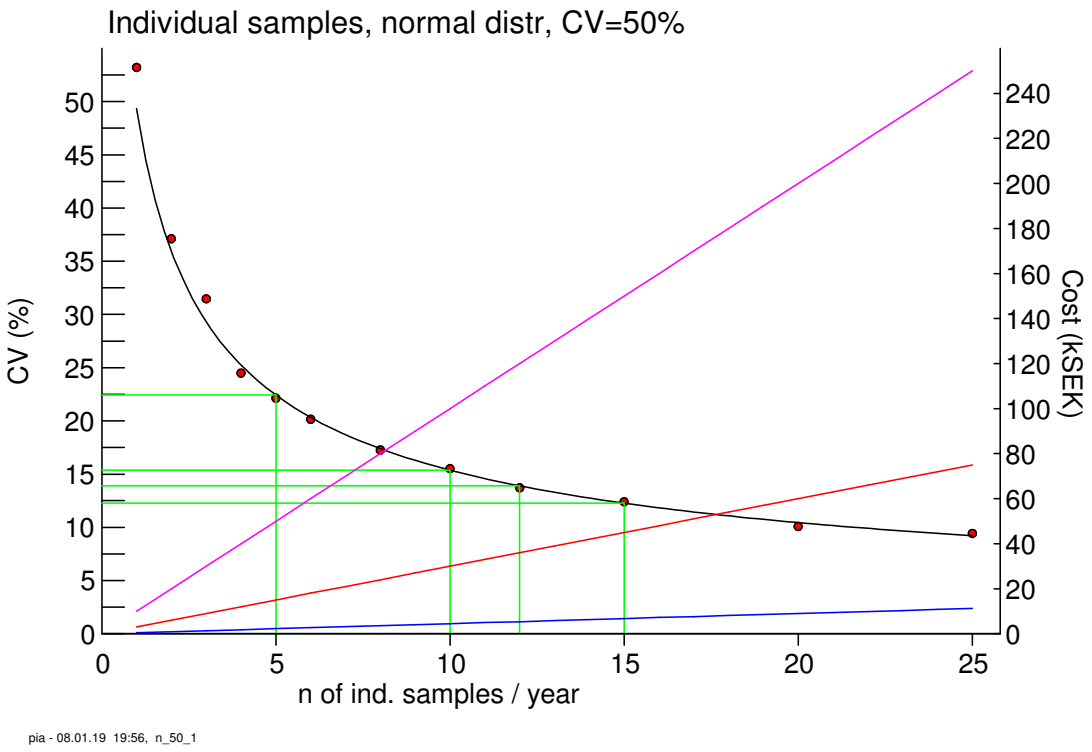
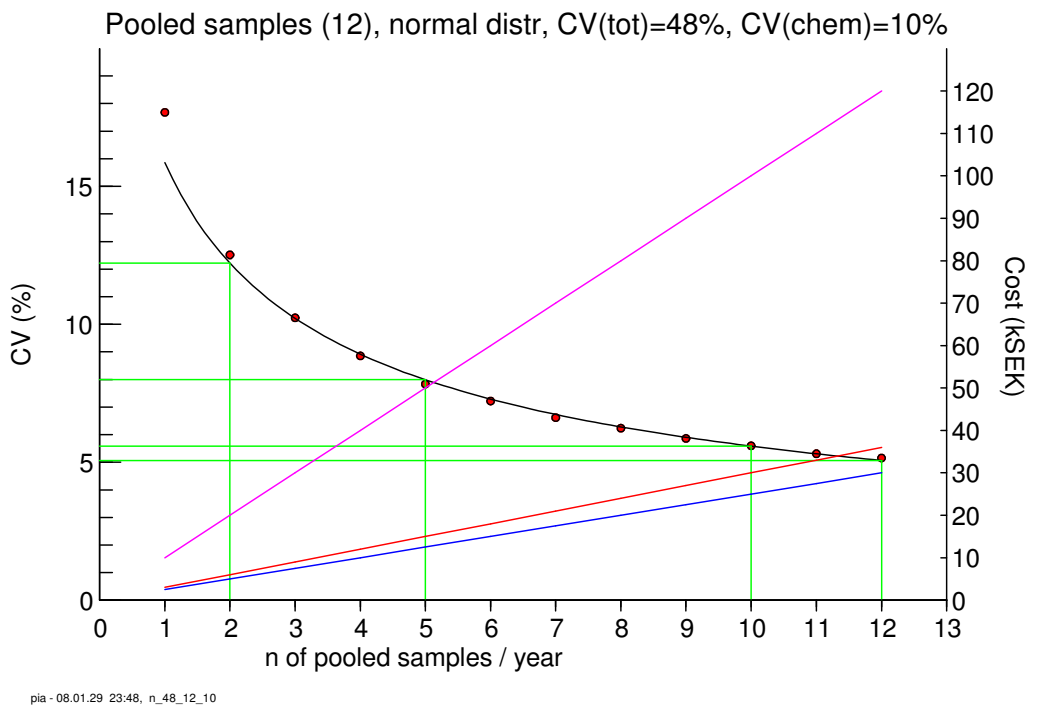


Fig 2. Individual samples from a normal distributed population with a total variance (CV) of 50%. When 12 samples are taken the between-year variance is reduced to 14%. Green lines show expected variance at various sample sizes. Blue line, cost for sample preparation, red line for organochlorine analysis and purple cost for dioxin analysis (all costs in SEK).

In Fig. 3, the probable between-year variation achieved at different sampling size of pooled samples (with 10 individual samples in each pool) are simulated. The contribution of the uncertainty from chemical analysis is 10 %. Already at only 2 pooled samples (i.e. sample material from 20 individual specimens are analyzed but only two chemical analyses are carried out each year) the expected variance is about 13% i.e. lower than the target variance of 14%. At 5 pooled samples the expected CV is about 8.5%.



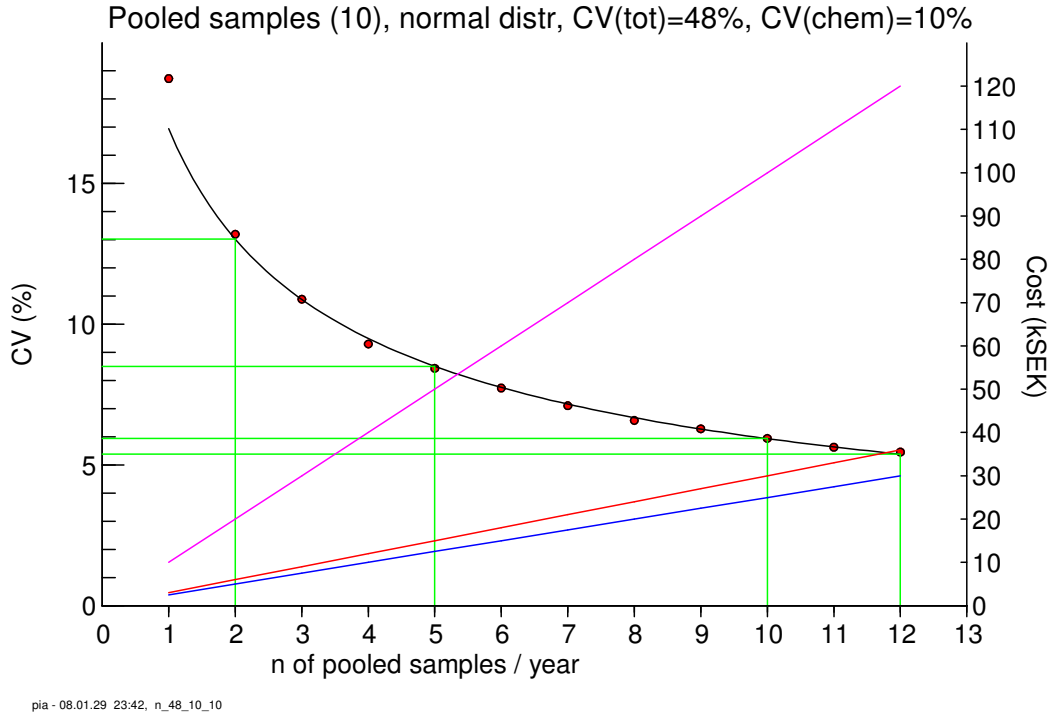


Fig 3. Pooled samples from a normal distributed population with a total variance (CV) of about 50%. Various numbers of pools with **A)** 12 and **B)**10 individual specimens in each pool are sampled for 10 years. When 2 pooled samples are taken the between-year variance is reduced to 12%(A) and 13%(B). Green lines show expected variance at various sample sizes. Blue line, cost for sample preparation, red line for organochlorine analysis and purple cost for dioxin analysis (all costs in SEK).

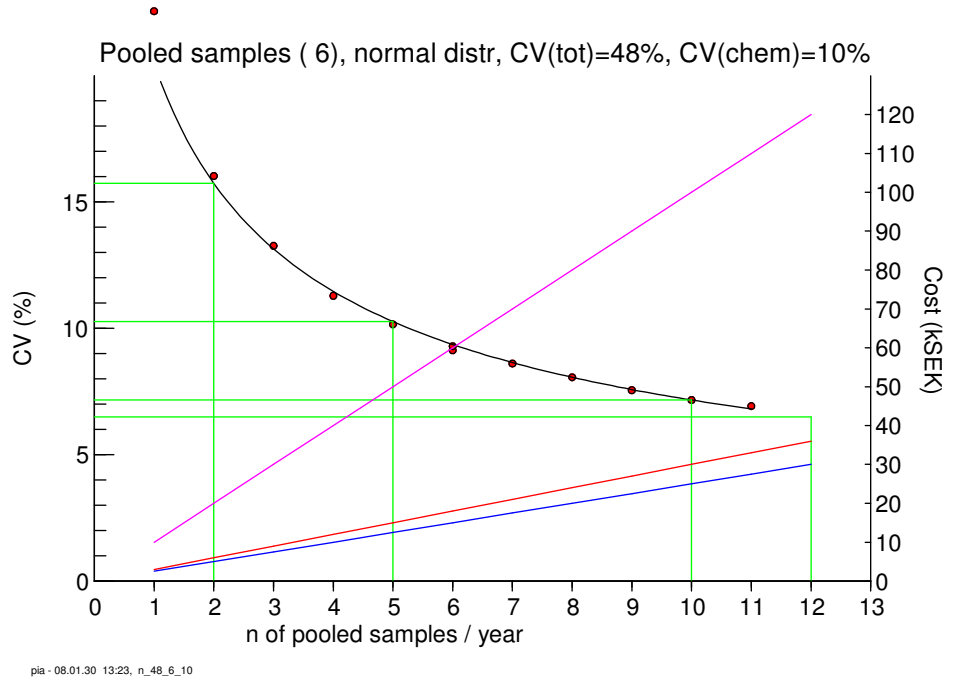


Fig 4. Pooled samples from a normal distributed population with a total variance (CV) of about 50%. Various numbers of pools with 6 individual specimens in each pool are sampled for 10 years. At least 3 pooled samples each year are required to come below the target CV of 14%. Green lines show expected variance at various sample sizes. Blue line, cost for sample preparation, red line for organochlorine analysis and purple cost for dioxin analysis (all costs in SEK).

In Fig. 4, the probable between-year variation achieved at different sampling size of pooled samples (with 6 individual samples in each pool) are simulated. The contribution of the uncertainty from chemical analysis is 10 %. At least 3 pooled samples each year are required to come below the target CV of 14% (i.e. sample material from 18 individual specimens are analyzed and 3 chemical analyses are carried out each year) the expected variance is about 13% i.e. lower than the target variance of 14%. At 5 pooled samples the expected CV is about 10%.

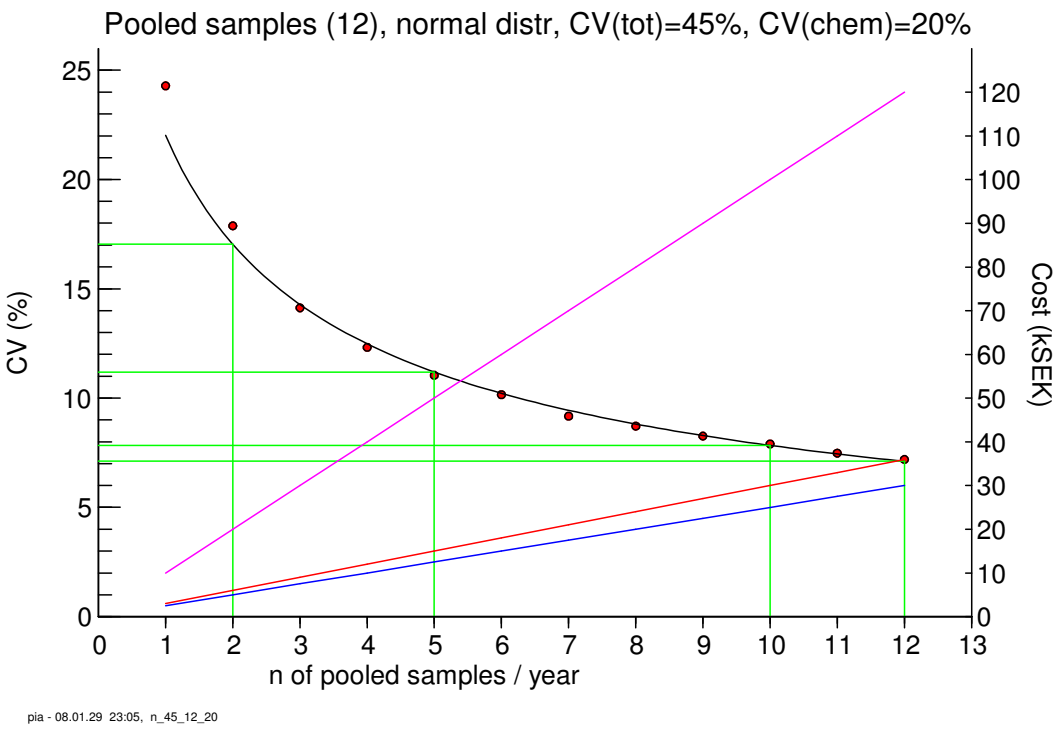


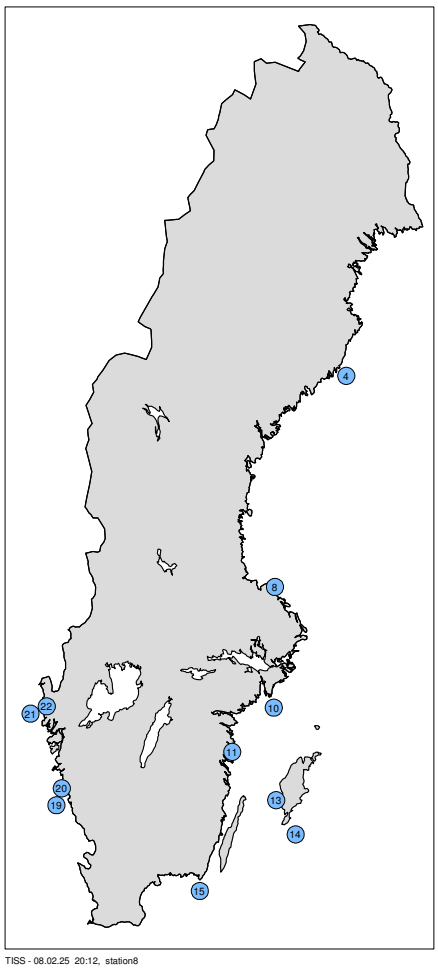
Fig 5. Pooled samples from a normal distributed population with a total variance (CV) of about 50%. The analytical uncertainty is increased to 20% compared to the 10 % in the previous figures. Various numbers of pools with 12 individual specimens in each pool are sampled for 10 years. At least 3 pooled samples each year are required to get approximately the target CV of 14%.4 pools should suffice for a CV of approximately 12.5%. Green lines show expected variance at various sample sizes. Blue line, cost for sample preparation, red line for organochlorine analysis and purple cost for dioxin analysis (all costs in SEK).

Table 1. Summary of results. Between-year variances expressed as the Coeficient of Variance (CV) and costs for organochlorine- (ClC), dioxin-analysis and sample preparation respectively in SEK.

Scenario	**CV (%)**	**ClC**	**Dioxines**	**Samp.prep.**
Ind. Sampl. 10, 50%	15.4	34500	104500	4500
Ind. Sampl. 12, 50%0	13.9	41400	125400	5400
Ind. Sampl. 15, 50%0	12.3	51750	156750	6750
Pooled 12, 48+10, 1	17.5	6000	13000	3000
Pooled 12, 48+10, 2	12	12000	26000	6000
Pooled 12, 48+10, 3	11	18000	39000	9000
Pooled 12, 48+10, 4	8.5	24000	52000	12000
Pooled 12, 45+20, 1	24	6000	13000	3000
Pooled 12, 45+20, 2	18	12000	26000	6000
Pooled 12, 45+20, 3	15	18000	39000	9000
Pooled 12, 45+20, 4	12.5	24000	52000	12000

Examples:

1) Provided that the analytical error is low (10%) we need only two pools (of 12 individuals each) to obtain a CV of 12% corresponding to 15 individual samples. This implies that about 40.000:- SEK is saved for one year at one site, analyzing PCB's, DDTs, HCH's and HCB. .
2) If the analytical error is larger (20%) and we want to avoid a larger CV than 15%. Then we have to analyze 3 pooled samples. This implies that around 65.500 SEK is saved



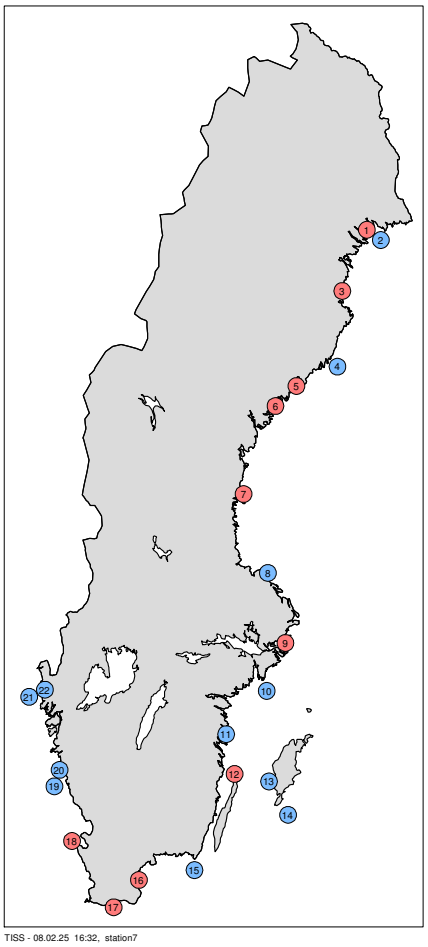


Fig. 7. Illustrates how the resources gained from a change from individual (**A**) to pooled **(B)** samples can be used to increase the station density.

Acknowledgement

This study was financed by the Swedish Environmental Protection Agency.

References:

Bignert A, Göthberg A, Jensen S, Litzen K, Odsjö T, Olsson M. & Reutergårdh L. 1993. The need for adequate biological sampling in ecotoxicological investigations: a retrospective study of twenty years pollution monitoring. *The Science of the Total Environment, 128 p.121-139.*

Bignert A., Riget F, Braune B., Outridge P., Wilson S. 2004. Recent temporal trend monitoring of mercury in Arctic biota – how powerful are the existing datasets? *J. Environ. Monit, 6, 351 – 355).*

Bignert, A, Danielsson S., Strandmark A., Nyberg E., Asplund L., Eriksson U., Berger U., Wilander A., **Haglund P.** 2008. Comments Concerning the National Swedish Contaminant Monitoring Programme in Marine Biota. Report to the Swedish Environmental Protection Agency, 2008-04-01. 142 pp.