

# Förslag till metod för utvärdering av övervakningsprogram i sötvatten



Claudia von Brömssen

SLU, institutionen för energi och teknik

Version 1.0

Rapport nr 127  
ISSN: 1654-9406

Institutionen för energi och teknik, SLU  
Box 7032  
750 07 Uppsala

<https://www.slu.se/institutioner/energi-teknik/>

*Omslagsillustration/omslagsfoto:* Bilden är skapad med AI genom Bing bildskapare, <https://www.bing.com/images/create?FORM=GDPGLP> , Promt: picture of freshwater systems like lakes, watercourses and groundwater, water chemistry, biology. Evaluation of data and monitoring program quality

Uppsala, januari 2024

## Förord

Denna rapport utgör redovisningen av ett projekt utfört på uppdrag av Havs- och vattenmyndigheten enligt överenskommelsen med diarienummer 1841-23, "Statistikstöd till vägledning om övervakning och inför revision av övervakningsprogram inom programområde Sötvatten".

## Innehållsförteckning

1. Introduktion.....	7
2. Statistiska metoder för utvärdering av övervakningsdata .....	7
2.1 Långsiktiga förändringar – sammanfattande nivå.....	7
2.1.1 Mann-Kendall-test.....	7
2.1.2 Linjär regression .....	8
2.1.3 Log-linjär regression .....	9
2.1.4 Icke-linjär regression .....	9
2.2 Långsiktiga förändringar – detaljerad nivå.....	9
2.2.1 Icke-linjär regression, generaliserade additiva modeller .....	9
2.2.2 Piecewise linjär regression .....	10
2.3 Långsiktiga förändringar – spatiala och tematiska komponenter.....	11
2.3.1 Kartor.....	11
2.3.2 Heatmaps .....	12
2.3.3 Trend screening plots .....	13
2.3.4 Regionala modeller.....	14
2.4 Att kvantifiera plötsliga förändringar .....	15
2.5 Att avgöra om en målnivå är uppnådd.....	15
2.5.1 Hypotestest för att avgöra måluppfyllelse .....	15
2.5.2 Konfidensintervall för att avgöra måluppfyllelse .....	16
2.5.3 Tillförlitlighet i måluppfyllelse .....	18
2.6 Att avgöra hur nära en målnivå man har kommit över tid .....	18
2.7 Att avgöra hur nära en målnivå man har kommit över tid – spatiala komponenter .....	19
3 Samband mellan kvalitetsfaktorer och andra miljövariabler .....	21
3.1 Som hjälpmedel i utvärdering av trender eller måluppfyllelse.....	21
3.1.1 Normalisering .....	21
3.1.2 Säsongsjustering.....	22
3.1.3 Objektspecifika referensvärden .....	22
3.2 För att identifiera orsaker .....	22
3.2.1 Regressionsmodeller .....	23
3.2.2 Multivariat statistik .....	24
3.2.3 Temporalt samvarierande trender .....	24
3.2.4 Geografiskt samvarierande trender .....	24
3.2.5 Tematiskt samvarierande trender.....	25
3.2.6. Förändringar i samband mellan en förklarande och en miljövariabel över tid.....	25

4. Kvalitetsindikatorer för övervakningsprogram .....	26
4.1 Generella kvalitetsindikatorer .....	26
4.1.1. Relevans.....	27
4.1.2 Tillförlitlighet och riktighet .....	27
4.1.3 Aktualitet och punktlighet.....	27
4.1.4 Tillgänglighet och tydlighet .....	28
4.1.5 Jämförbarhet och sammanvändbarhet .....	28
4.2. Kvalitetsindikatorer som kan beräknas genom enklare statistiska metoder .....	28
4.2.1. Riktighet.....	29
4.2.2. Stickprovsstorlek, precision och medelfel.....	29
4.2.3 Stickprovsstorlek och statistisk styrka.....	30
4.2.4 Stickprovsstorlek och teoretiska sannolikhetsberäkningar.....	31
4.2.5. Bestämning av nödvändig stickprovsstorlek .....	32
4.2.6. Minskning av variation genom normalisering.....	33
4.3 Utökad kvalitetsdeklaration .....	33
4.3.1 Att övervaka för framtiden.....	33
4.3.2 Samanvändning .....	34
4.3.3 Att förstå orsak och verkan .....	34
4.3.4 Forskning baserad på övervakningsdata .....	34
5. Förslag till metod för att utvärdera miljöövervakningsprogram.....	35
5.1 Förslag till metod för statistisk utvärdering av producerade data.....	35
5.2 Förslag till kvalitetsutvärdering av delprogram.....	35
6. Referenser .....	37
Appendix.....	40
A. Exempel på styrkeberäkningar .....	40
A.1 Med hjälp av en typserie.....	40
A.2 Med hjälp av generella värden – absolut förändring.....	41
A.3 Med hjälp av generella värden – relativ förändring.....	42
B. Simulerade sannolikheter för felklassningar av status.....	45
C. Varianter av screening/lasanga plot.....	46
D. Förslag till kvalitetsdeklaration .....	47
E. Centrala Begrepp.....	48
E.1 Noggrannhet – riktighet – precision.....	48
E.2 Variation, varians, standardavvikelse.....	48
E.3 Riktighet, bias, representativitet.....	48
E.4 Medelfel, ”standard error” .....	49

E.5 Samplingfördelning.....	49
E.6 Konfidensintervall, felmarginal .....	50
E.7 Hypotestest .....	51
E.8 Fel av första och andra slaget, statistisk styrka .....	52
E.9 Oberoende observationer vs temporala och spatiala autokorrelationer .....	54
E.10 Skeva fördelningar .....	54
E.11 Värden under detektions- eller kvantifieringsgräns.....	55

## 1. Introduktion

Övervakningsprogram behöver utvärderas regelbundet för att fastställa att de fortfarande är ändamålsenliga och att data som produceras i programmen kan användas för de syften som programmen har.

Huvudmålen med övervakningsprogram är vanligtvis att upptäcka temporala trender, det vill säga att identifiera om medelnivån av en serie eller en grupp av serier ökar eller minskar långsiktigt, samt att kunna bedöma miljötillstånd i förhållande till målnivå i ett objekt. Kapitel 2 och 3 presenterar olika statistiska metoder som ofta används inom miljöövervakningen för trendanalys och bedömning av miljötillstånd. Kapitel 4 diskuterar ett antal kvalitetsindikatorer som kan användas för att beskriva kvaliteten av ett övervakningsprogram. Ett förslag för utvärdering av miljöövervakningsprogram presenteras i kapitel 5.

## 2. Statistiska metoder för utvärdering av övervakningsdata

### 2.1 Långsiktiga förändringar – sammanfattande nivå

Att analysera långsiktiga förändringar, trender, på en sammanfattande nivå innebär att information komprimeras till ett eller några få numeriska värden. Ofta genom att genomföra ett trendtest för en specifik tidsperiod som resulterar i testets p-värde och en uppskattning av trendens (genomsnittliga) magnitud. Några vanliga metoder beskrivs nedan.

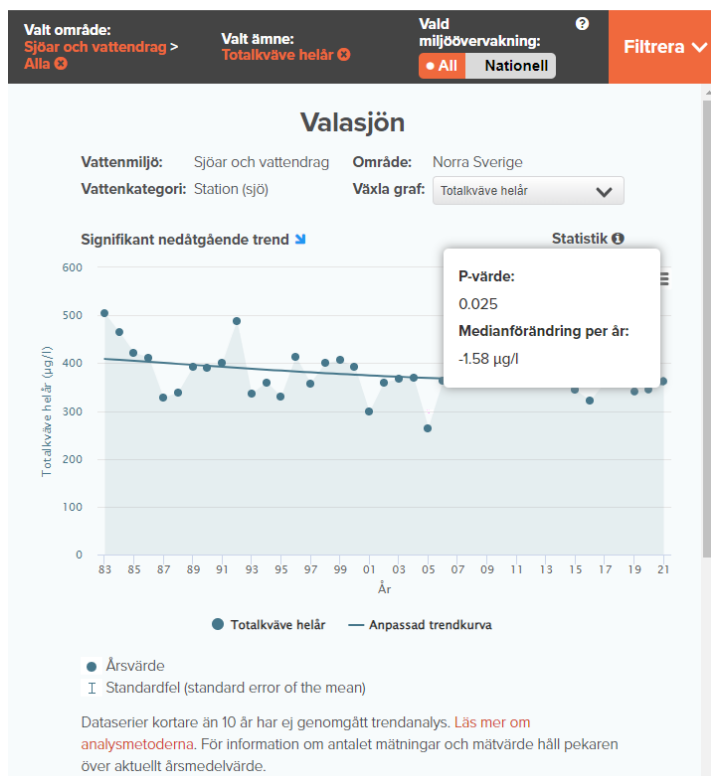
#### 2.1.1 Mann-Kendall-test

Mann-Kendall-test (Hirsch et al., 1982; Kendall and Gibbons, 1990; Mann, 1945) baseras på rangordning av data, det vill säga att de faktiska värdena inte används direkt utan enbart deras relativa ordning. Detta test kräver alltså inte att miljövariabeln följer en specifik fördelning, såsom normal eller log-normal. Inte heller värden under detektionsgräns och enstaka mycket avvikande värden (outliers) utgör några problem, då de bara översätts till lägsta respektive högsta rang. Mann-Kendall-tester kan beräknas på tidsserier som är korta (ca 10 årliga observationer eller fyra år av månatliga data) och kan kompletteras med skattningar av en medianförändring över tiden, som beräknas som Theil-Sen slope (Sen, 1968; Theil, 1959).

Om miljövariabeln provtas som månatliga värden kan ett "seasonal" Mann-Kendall-test användas (Hirsch and Slack, 1984). Säsongsvariationen tas hänsyn till genom att beräkna Mann-Kendall-test separat för varje månad, som sedan summeras. Ett "seasonal" Mann-Kendall-test kan även hantera att det finns temporal autokorrelation (avsnitt E.9) i data, vilket görs genom att beräkna korrelationer mellan månatliga Mann-Kendall-testvärden. Är data ännu mer frekventa kan även det tas hänsyn till, men det kräver mer förarbete. En liknande princip som för "seasonal" Mann-Kendall-tester tillåter också gemensamma analyser för ett flertal stationer (Helsel and Frans, 2006; Loftis et al., 1991).

Mann-Kendall-trendtesterna kan justeras för förklarande variabler som introducerar naturlig variation, till exempel avrinning. Detta görs med partiella Mann-Kendall-tester (Libiseller and Grimvall, 2002).

Mann-Kendall-test är enkla och fungerar bra i många olika sammanhang utan att kräva starka antaganden om data eller formen av trenden, vilket gör dem populära inom miljöstatistiken. Ett exempel på Mann-Kendall-test ges i Figur 1.



Figur1 : Mann-Kendall-test används för att beskriva tidstrender på webbsidan Sveriges Vattenmiljö (<https://www.sverigesvattenmiljo.se/>) . Mann-Kendall-test ges för tidstrenden i totalkväve i Valasjön mellan 1983 och 2021. P-värdet är 0,025 och visar på en signifikant nedåtgående trend. Theil-Sen slope beräknas till -1,58 µg/l. I genomsnitt per år minskar totalkväve alltså med 1,58 enheter.

### 2.1.2 Linjär regression

Linjär regression kan också användas för att modellera en genomsnittlig förändring över tid, samt för att testa om denna förändring skiljer sig signifikant från noll. Dessa modeller baseras på fler antaganden än Mann-Kendall-tester. Antaganden är:

- (i) trenden som observeras kan approximeras med en linje,
- (ii) modellens residualer följer approximativt en normalfördelning,
- (iii) variationen i data är konstant över hela tidsperioden, och
- (iv) observationerna är oberoende.

Om antaganden (i) och (ii) inte är uppfyllda kan ibland datatransformationer användas (se 2.1.3). Är variationen inte konstant (antagande (iii)) men följer en viss struktur kan denna struktur tas med i modellen, till exempel i en viktad regression. Eftersom analyserna baseras på en tidsserie, kommer antagande (iv) inte vara uppfyllt. För årliga data brukar autokorrelationen i data ofta vara tillräckligt liten för att kunna ignoreras. För data med högre provtagningsfrekvens modelleras autokorrelationen antingen i en mixad modell (mixed model) eller genom att använda generaliserade least squares-skattningar.



Linjär regression är en standardmodell och både grundmodellen och mixade modeller beskrivs i ett stort antal grundläggande statistikböcker. Eftersom det dock ofta inte rimligt att anta att trenden är linjär, speciellt inte om tidsserien är längre än några få år, så används inte linjär regression speciellt mycket inom miljöstatistiken.

### 2.1.3 Log-linjär regression

Log-linjär regression är en variant av linjär regression där trenden antas vara exponentiell istället för linjär. En sådan modell används när minskningen eller ökningen över tiden ska skattas som en konstant årlig procentuell förändring. Modellen anpassas vanligtvis genom en regressionsanalys där responsvariabeln log-transformeras före analysen. Antaganden som görs i modellen är då att feltermen i modellen är normalfördelad och har konstant varians. Det innebär att responsvariabeln på originalskalan antas ha en log-normal fördelning och att variansen ökar med ökat värde på responsvariabeln.

Istället för log-transformation kan även en icke-linjär regressionsmodell med specifik exponentiell form anpassas till data. Modellens residualer antas igen vara approximativt normalfördelade och med konstant varians.

Log-linjär regression används, till exempel, för trendanalyser i marin biota (Bignert et al., 2016).

### 2.1.4 Icke-linjär regression

Även icke-linjär regression kan användas för att ta fram sammanfattande mått av långsiktiga trender (Fryer and Nicholson, 1999), men är vanligare när trender studeras på en detaljerad nivå (se sektion 2.2)

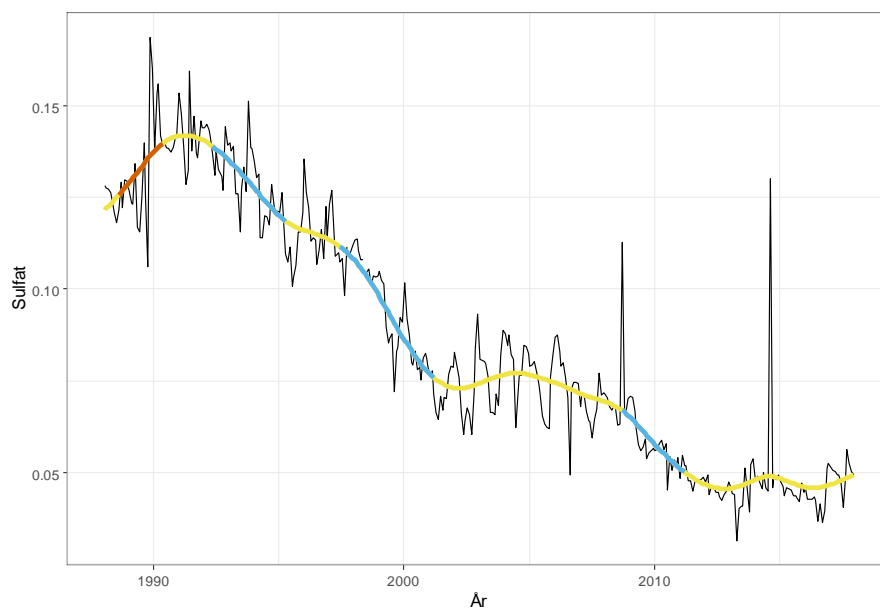
## 2.2 Långsiktiga förändringar – detaljerad nivå

Att analysera trender på en sammanfattande nivå kan verka attraktivt, då det ger svar på om en trend finns eller inte finns under en angiven tidsperiod och resultaten kan redovisas på ett enkelt sätt. Samtidigt förloras en del av informationen som finns i data, speciellt om tidsserierna är långa. Om det finns både upp- och nedgångar under den observerade tidsperioden kan valet av tidsperiod som ska testas vara avgörande för vilka resultat som erhålls. Istället kan mer komplexa trendfunktioner användas för att behålla mer av den tillgängliga informationen.

### 2.2.1 Icke-linjär regression, generaliserade additiva modeller

För att illustrera långsiktiga trender på en detaljerad nivå används ofta figurer med observerade data och en data-driven modell som förtydligar den temporala förändringen. Sådana modeller kräver inte att man i förhand avgör om trenden är linjär, log-linjär eller följer någon annan form, utan de anpassar sig till de data man har. Numera används mest generaliserade additiva modeller (Hastie and Tibshirani, 1986; Wood, 2017) för att modellera tidstrender, då de har bra matematiska egenskaper och kan användas i många olika sammanhang. För att förtydliga när signifikanta förändringar sker i en serie har derivator, deras osäkerhet och färgkodningar använts (Figur 2, Simpson, 2014, 2018; von Brömssen et al., 2021).

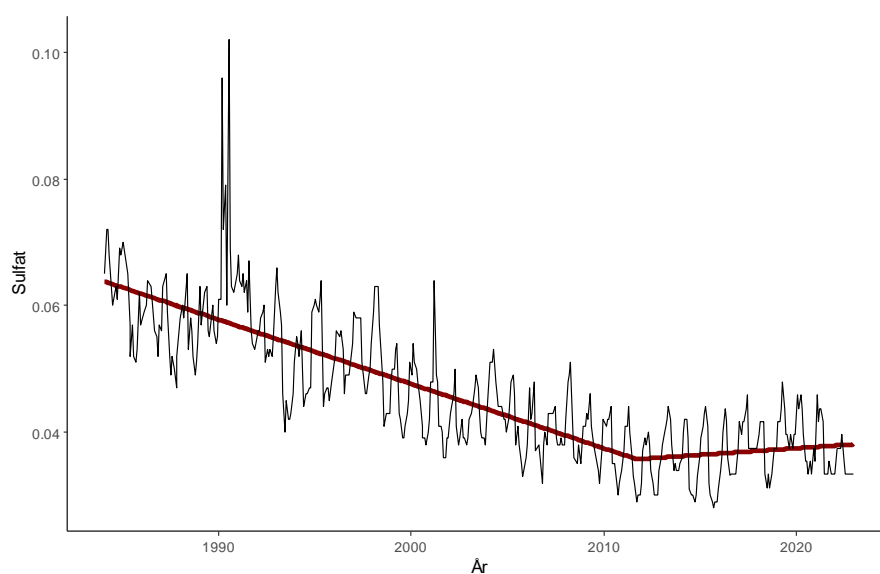
Nackdelen med icke-linjär regression är att frågan om ifall en förändring har skett inte längre kan sammanfattas i ett enda ja/nej-svar. Det går att beräkna ett p-värde som avgör om serien signifikant avviker från en horisontell linje någon gång under tidsperioden, men ett sådant är tämligen ointressant. Istället är det vanligt att göra en visuell presentation av observerade data och en trendkurva.



Figur 2: Trend av sulfat i Alsterälven skattad med en generaliserad additiv modell. Blå indikerar en signifikant nedåtgående trend, röd en signifikant uppåtgående trend och gul att ingen signifikant trend finns vid denna tidpunkt.

### 2.2.2 Piecewise linjär regression

Piecewise linjär, eller segmented, regression är en annan variant av linjär regression där trendlinjen tillåts ändra riktning eller lutning. Förändringen sker antingen vid en i förhand bestämd tidpunkt eller genom skattning av brytpunktens placering utifrån data. De två lutningarna skattas separat, men hänger vanligtvis ihop vid brytpunkten (Figur 3); men det är också möjligt att tillåta en stegvis förändring i nivån samtidigt som förändringen i trendlinjen. För korta serier blir en datadriven skattning av brytpunkten ofta osäker, då dataunderlaget är för litet för att både skatta separata lutningar och placeringen av brytpunkten. Till exempel kan enstaka nya tillkommande observationer ändra var skattningen för brytpunkten ligger och med det förändra resultat och tillhörande slutsatser avsevärt.



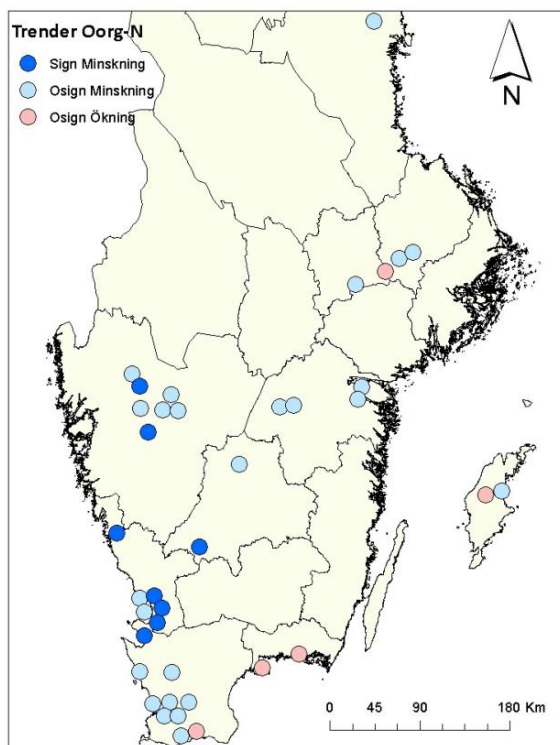
Figur 3: Piecewise linjär regression. Sulfat i Ljusnan. Brytpunkten är beräknad utifrån tillgängliga data och trenden antas vara linjär före och efter brytpunkten.

## 2.3 Långsiktiga förändringar – spatiala och tematiska komponenter

Eftersom påverkan på miljövariabler ofta är storskaliga, exempelvis genom klimatförändringar eller förändrad deposition, kan det vara intressant att se temporala förändringar i ett spatialt sammanhang. Även visualisering i tematiska sammanhang kan göras, till exempel genom att visa trender för flera stationer som har liknande bakgrundsförhållanden, såsom andel jord- eller skogsbruksmark.

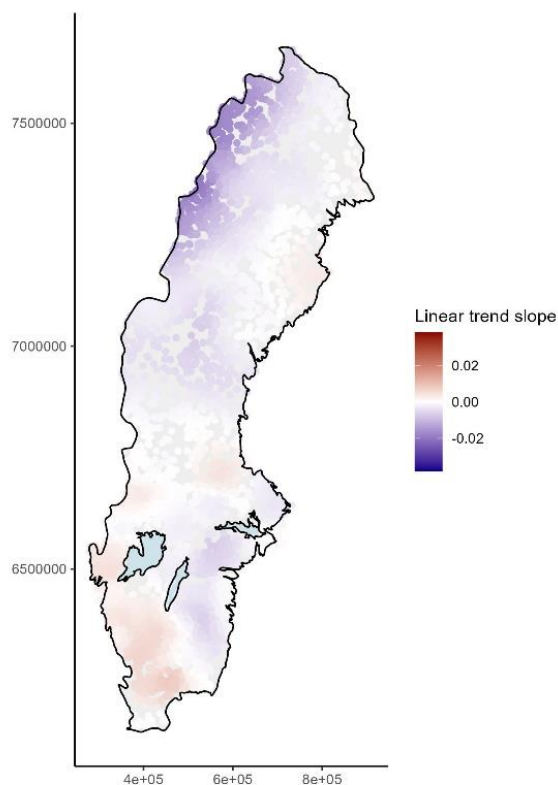
### 2.3.1 Kartor

Enkla sammanfattningar av data, såsom resultat från Mann-Kendall-test eller skattningar av genomsnittlig årlig förändring skattat genom Theil-Sen slope, linjär eller loglinjär regression kan presenteras i kartor för att illustrera den spatiala komponenten (Figur 4).



Figur 4: Signifikanta och icke-signifikanta förändringar i oorganiskt kväve. Bilden är framtagen av Institutionen för vatten och miljö, SLU (Havs- och vattenmyndigheten, 2020)

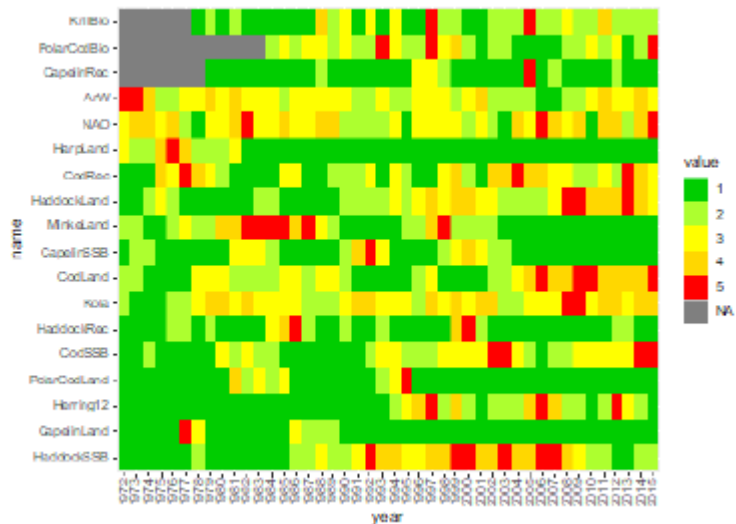
Om data finns med hög spatial upplösning och förväntade trender har en storskalig komponent orsakad av till exempel depositions- eller klimatförändringar kan rumsliga trender skattas med geografisk viktad regression (geographically weighted regression, GWR, (Brunsdon et al., 1998)) och visualiseras i en karta (Figur 5). Detta baseras på att stationer inom ett rumsligt fönster uppvisar liknande temporal trend. Skattningen av trenden vid medelpunkten för det geografiska fönstret görs genom en linjär regression, där enstaka observationer viktas med avseende på avståndet till fönstrets mittpunkt. Sedan flyttas det geografiska fönstret och en ny skattning görs för observationerna inom det nya fönstret. En fördel med denna metodik är att den även kan användas om observerade data har en ganska gles temporal upplösning med bara några enstaka observationer per station och tidsperiod (von Brömssen et al., 2023a).



Figur 5: Regionalt varierande trender skattat genom geografiskt viktade regressionsmodeller. Trender i TOC 2008-2021 baserat på data från omdrevssjöprogrammet (von Brömsen et al., 2023a).

### 2.3.2 Heatmaps

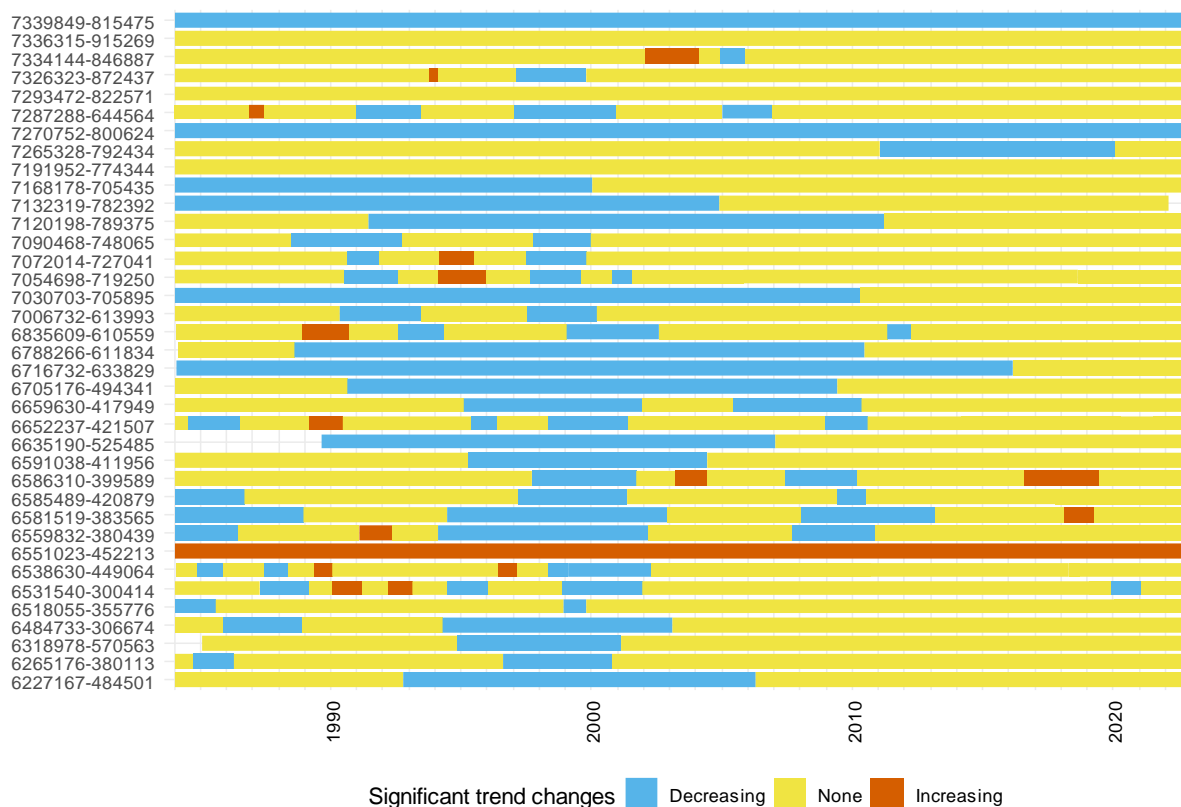
Nivåerna av en miljövariabel kan illustreras för flera serier samtidigt genom så kallade heatmaps. Heatmaps kräver en kategorisk variabel på x-axeln, vilket innebär att tidsvariabeln behöver delas in i separata månader eller år. För varje tidsperiod beräknas medelvärdet av miljövariabeln. Ett typiskt problem med dessa plottar är att nivåerna av en miljövariabel kan variera mycket mellan stationer eller mellan variabler, vilket gör att information om förändringar över tid är svårt att avläsa. Detta kan lösas genom medelvärdescentrering av alla serier före visualiseringen. Alternativt kan variabeln klassas i ett fåtal lika stora intervall från det lägsta till det högsta värdet i varje enskild tidsserie (Integrated trend analysis to support integrated ecosystem assessment; ICES, 2022). Istället för miljövariablernas nivåer visas då kategorierna i en heatmap (Figur 6).



Figur 6: Ett exempel på en tematisk heatmap för ICES data. År visas på x-axeln och olika variabler på y-axeln. Figuren ger visuell information om ifall nivåerna i de olika variablerna samvarierar och hur de förändras med tiden (ICES, 2022).

### 2.3.3 Trend screening plots

I stället för nivåerna kan det vara relevant att visualisera när under en tidsperiod förändringar i nivåer sker (trendkurvans lutning). Genom att använda principerna för att illustrera trender genom generaliserade additiva modeller (sektion 2.2.1) kan sammanfattande information om tidpunkter för förändring för ett antal olika stationer eller variabler presenteras. Stationerna kan ordnas för att ge information om hur trenderna varierar över landet, till exempel med en nord-syd-gradient, eller med hjälp av en tematisk gradient, såsom andel jordbruksmark eller skogsmark inom avrinningsområdet. Informationen om när signifikanta trender uppträder i de olika serierna visas i figurer som liknar heatmaps, men där tidsvariabeln i figuren är kontinuerlig, vanligtvis datum eller år. Dessa grafer kallas lasagnaplots eller screening plots (Figur 7, von Brömssen et al., 2021).

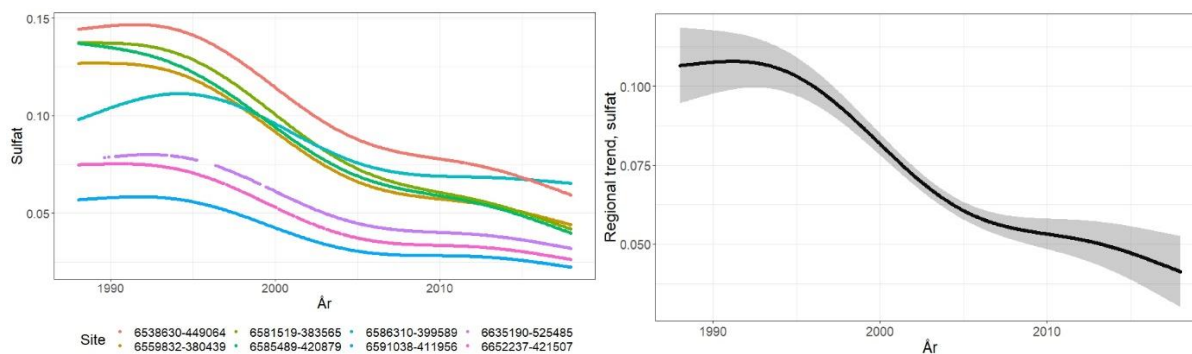


Figur 7: Spatial visualisering av signifikanta förändringar i sulfat. Stationerna är sorterade från nord (längst upp) till syd. Nedåtgående trender observeras i södra Sverige främst under 1990-talet och tidiga 2000-talet.

### 2.3.4 Regionala modeller

För ett större område eller en region kan en gemensam trendanalysmodell användas där alla objekt inom området ingår om det finns skäl att förvänta sig en trend som har ungefär samma storlek över hela området. Till exempel skulle deposition, klimateffekter eller storskaliga förändringar i markanvändning kunna leda till liknande förändringar i en vattenkvalitetsvariabel över ett större område.

En gemensam modell kan formuleras med antingen en linjär, en log-linjär eller en utjämnad trend och skattas i en mixad modell, där stationsnamn eller id ingår som en slumpmässig faktor. Exempel på ett antal olika sätt att kombinera ett flertal serier genom att använda utjämnade trender ges i Pedersen et al. (2019) och illustreras i Figur 8.



Figur 8: Individuella trendkurvor för åtta vattendrag i Värmland (vänster) och en gemensam skattad regional trend för samma serier (höger). Observera att de individuella trendlinjerna skattas med liknande utjämning och kan då likna varandra mer än de skulle göra om enskilda modeller skulle skattas för varje vattendrag. Figuren följer principerna beskrivna i von Brömssen et al. (2021)

## 2.4 Att kvantifiera plötsliga förändringar

Förutom gradvisa förändringar kan även plötsliga förändringar ske i en miljötidsserie till följd av, till exempel, omfattande åtgärder eller policyändringar. En sådan ökning eller minskning kan antingen vara (i) en snabb stegvis förändring av nivån, (ii) en kraftig förändring av trendriktningen eller magnituden eller (iii) båda.

Om förändringen är en snabb stegvis förändring vid en känd tidpunkt kan vanliga trendmodeller utökas med en indikator för tidpunkten när förändringen sker (se till exempel von Brömssen et al., 2018). Förändringar i trendlutningen kan fångas av generaliserade additiva modeller (sektion 2.2.1), som kan justeras på ett sätt så att de bättre fångar upp snabba förändringar. Även "piecewise" linjär regression (sektion 2.2.2) kan användas (Carstensen and Weydmann, 2012; HELCOM, 2013).

## 2.5 Att avgöra om en målnivå är uppnådd

### 2.5.1 Hypotestest för att avgöra måluppfyllelse

Hypotestest (se även avsnitt E.7) kan användas för att avgöra om det (okända) sanna värdet av en parameter avviker signifikant från ett givet värde. Till exempel kan det testas om det sanna medelvärdet av en miljövariabel (eller det sanna EK-värdet) ligger signifikant under (eller över) godmåttlig-gränsen som används för statusklassning.

Hypotestester baseras på en skattning och skattningens osäkerhet i termer av skattningens medelfel, samt skattningens samplingfördelning (avsnitt E.5). Dessutom behöver användaren definiera två hypoteser:

- (i) Nollhypotesen och
- (ii) Alternativhypotesen, som representerar det som ska visas.

Till exempel kan nollhypotesen vara att det sanna medelvärdet  $\mu$  av miljövariabeln ligger över godmåttlig-gränsen  $g_0$  ( $H_0: \mu \geq g_0$ ) och alternativhypotesen att det sanna medelvärdet  $\mu$  är lägre än godmåttlig-gränsen  $g_0$  ( $H_1: \mu < g_0$ ). Denna variant av test kallas "fail-safe", eftersom målet är att visa att statusen med statistisk säkerhet är god eller bättre (Carstensen, 2007; European Commission, 2003). Ett alternativ är att vända på testet och testa om miljövariabeln ligger signifikant sämre än godmåttlig-gränsen ("benefit-of-doubt"). Liknande tester har även kallats "green" och "brown" och visualiserats med hjälp av konfidensintervaller (Nyberg, 2016). När test eller konfidensintervall inte kan beräknas bedöms statusklassning ofta med "face-value" (Carstensen, 2007). Det betyder att det uppskattade värdet ensamt avgör vilken statusklass som fastställs. I regel genomförs hypoteser för

att testa måluppfyllelse med ett t-test eller motsvarande icke-parametriskt test, Wilcoxon signed rank test. Även konfidensintervall kan användas för hypotestest och ger samma svar så länge de är baserade på samma antaganden som motsvarande test.

Om endast få eller enskilda värden är tillgängliga för ett objekt måste ibland schablonvärden för variansen användas. Dessa tas fram med hjälp av (liknande) objekt där det finns tillräckligt med data. I WATERS-projektet föreslogs skattning av varianskomponenter genom en generell modell och detta resulterade i ett osäkerhetsbibliotek (Carstensen and Lindegarth, 2016), som kan användas när dataunderlaget inte tillåter egna variansskattningar.

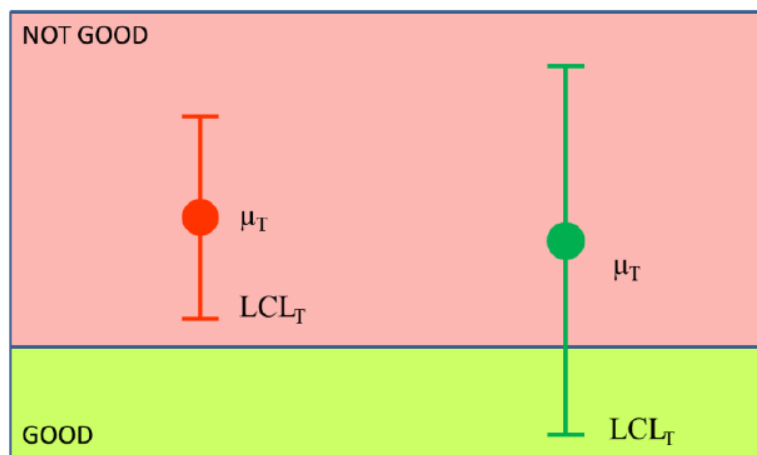
### 2.5.2 Konfidensintervall för att avgöra måluppfyllelse

Ett konfidensintervall (se även avsnitt E.6) baseras, precis som hypotestester, på en skattad parameter, till exempel ett medelvärde, skattningens varians och samplingfördelning.

Konfidensintervallet ger information om både var det sanna värdet skulle kunna ligga och hur osäker skattningen är (genom bredden av intervallet).

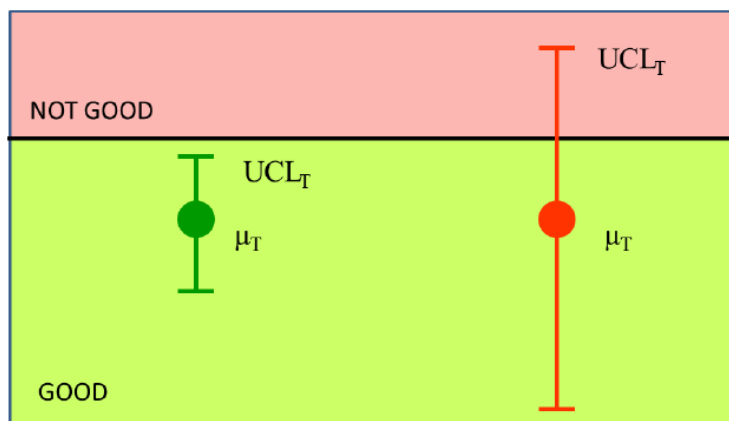
#### Exempel på användning av konfidensintervall för statusklassning:

(i) Konfidensintervall har använts av Nyberg (2016) för att illustrera hypotestest, då kallade "brown test" och "green test". Målet var att visa om en miljövariabel ligger signifikant över eller under en kvalitetsstandard eller klassgräns eller om det inte är möjligt att uttala sig om det (Figur 9 och Figur 10)



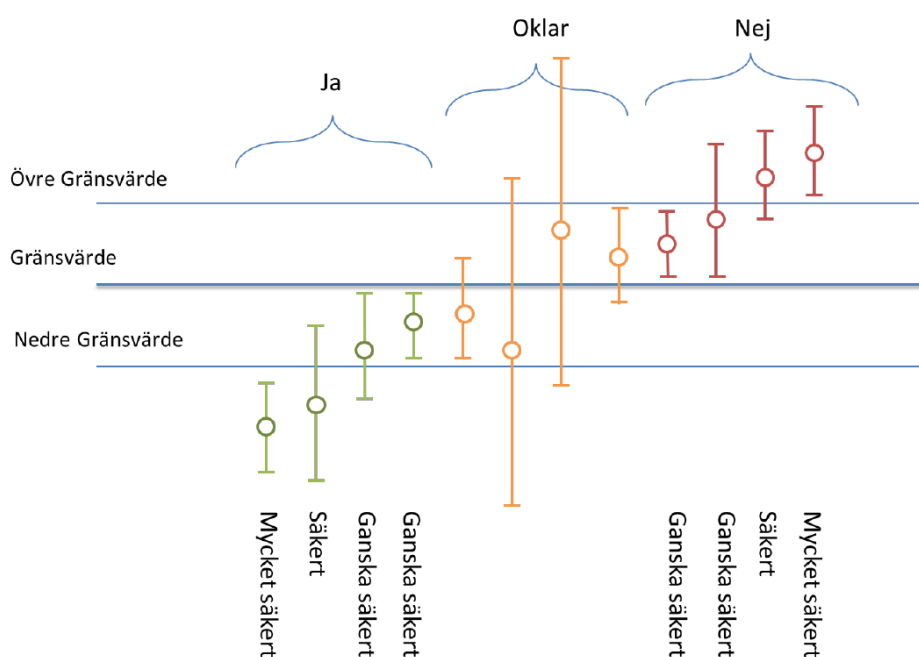
Figur 9: Tagen från Nyberg (2016). "Brown test" för att testa om en miljövariabel ligger inom området som klassas som dåligt ("not good"). Endast om hela konfidensintervallet ligger ovanför klassgränsen bedöms testet som signifikant.





Figur 10: Tagen från Nyberg (2016). "Green test" för att testa om en miljövariabel ligger inom området som klassas som bra ("good"). Endast om hela konfidensintervallet ligger nedanför klassgränsen bedöms testet som signifikant.

(ii) Ett liknande koncept med fler steg presenterades för klassning av totalfosfor i sjöar och vattendrag (Fölster and von Brömssen, 2012). I denna tillämpning är även gränsvärdet osäkert, då det tas fram med hjälp av en regressionsmodell. Likt Nyberg föreslås grönt (konfidensintervallet i sin helhet under gränsvärdet) och rött/brunt (i sin helhet över gränsvärdet), samt oklar/inget resultat om konfidensintervallet täcker gränsvärdet. Utöver det föreslås en gradering för säkerheten av resultatet som baseras på avståndet mellan det skattade värdet och klassgränsen där även klassgränsens osäkerhet kan användas (Figur 11).



Figur 11: Tagen från Fölster and von Brömssen (2012). Bedömning av om en miljövariabel ligger inom området god (grön) eller måttlig (röd), samt när ingen bedömning kan göras (orange), baserat på konfidensintervall. Ju längre hela konfidensintervallet ligger från klassgränsen, desto säkrare bedöms resultatet.

### 2.5.3 Tillförlitlighet i måluppfyllelse

Med hjälp av hypotestest och konfidensintervall kan det avgöras om en nivå är statistiskt säkert uppnådd eller inte. I praktiken är dataunderlaget i statusklassning ofta litet, vilket leder till breda konfidensintervall och icke-signifikanta testresultat. Det är vanligt att använda en kombination av "face-value" och test eller konfidensintervall för att gradera bedömningsresultaten. Ett exempel på detta gavs redan i Figur 11, där statusen bedöms som god (eller bättre) om konfidensintervallet i sin helhet ligger under gränsvärdet och som måttlig (eller sämre) om konfidensintervallet i sin helhet ligger över. För dessa fall kan klassningen anses säker. För resterande fall används det uppskattade värdet för statusklassningen men notera att resultatet inte är säkert och objektet behöver övervakas ytterligare (Tabell 1).

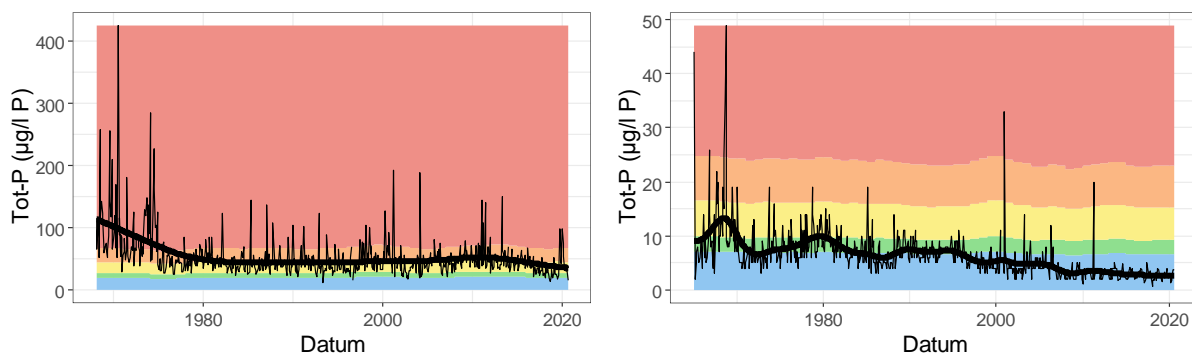
Om resultatet visar övervakningsbehov kan ytterligare provtagning göras. Konfidensintervallets bredd minskar och möjligheten att få signifikanta resultat ökar med stickprovsstorleken så länge ytterligare oberoende observationer kan samlas in under liknande förhållanden. Det är speciellt fördelaktigt om det finns få observationer från början, då det observerade medelvärdet måste anses osäkert och variationen lätt kan underskattas om stickprovet är litet. Om det sanna medelvärdet av miljövariabeln ligger nära klassgränsen kan det dock krävas orimligt stora datamängder för att avgöra om det faktiskt ligger under eller över denna gräns. I praktiken innebär det att det måste göras en bedömning av om ytterligare datainsamling är ekonomiskt försvarbart. En sådan bedömning görs med hjälp av observerade data, det uppskattade medelvärdet och antalet observationer som redan finns tillgängliga (se även sektion 4.2.4).

Tabell 1: Bedömningsresultat och tillförlitlighet av statusklassning baserat på "face value" samt resultat av hypotestest eller konfidensintervall.

"Face-value"	Hypotestest	Konfidensintervall	Tillförlitlighet
Uppskattat värde under gränsvärdet	Signifikant	Täcker inte gränsvärdet	Säker, Frias
Uppskattat värde under gränsvärdet	Inte signifikant	Täcker gränsvärdet	Osäker, Övervakningsbehov
Uppskattat värde över gränsvärdet	Inte signifikant	Täcker gränsvärdet	Osäker, Övervakningsbehov
Uppskattat värde över gränsvärdet	Signifikant	Täcker inte gränsvärdet	Säker, Förbättringsbehov

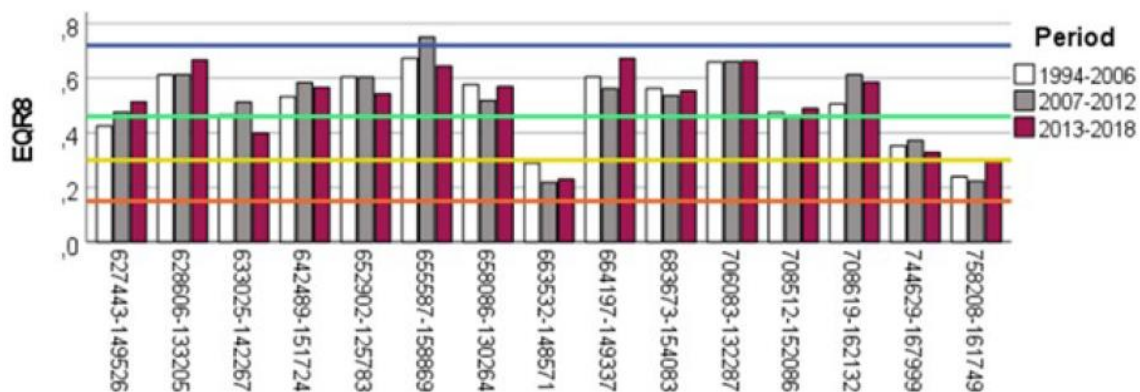
### 2.6 Att avgöra hur nära en målnivå man har kommit över tid

När samma statusbedömningar ska göras över tid kan det vara aktuellt att även följa dessa över tid. Ibland baseras statusbedömningar på data som finns som tidsserie och en visualisering av nivåerna i relation till klassgränser kan då göras i en gemensam figur (Havs- och vattenmyndigheten, 2020). För totalfosfor baseras klassgränserna på ett antal bakgrundsvariabler som i sin tur varierar över tid. Klassgränserna kan då visas som bakgrund till observerade nivåer av totalfosfor (Figur 12). Liknande visualiseringskoncept används inom HELCOM för att visa hur hög näringsämnestransporten till Östersjön är jämfört med *maximum allowable input* (MAI; HELCOM, 2015).



Figur 12: Trender i totalfosfor för Visman Nybble (vänster) och Skellefte älv (höger) i relation till de ekologiska statusklassningarna, representerade som 5-års glidande medelvärden.

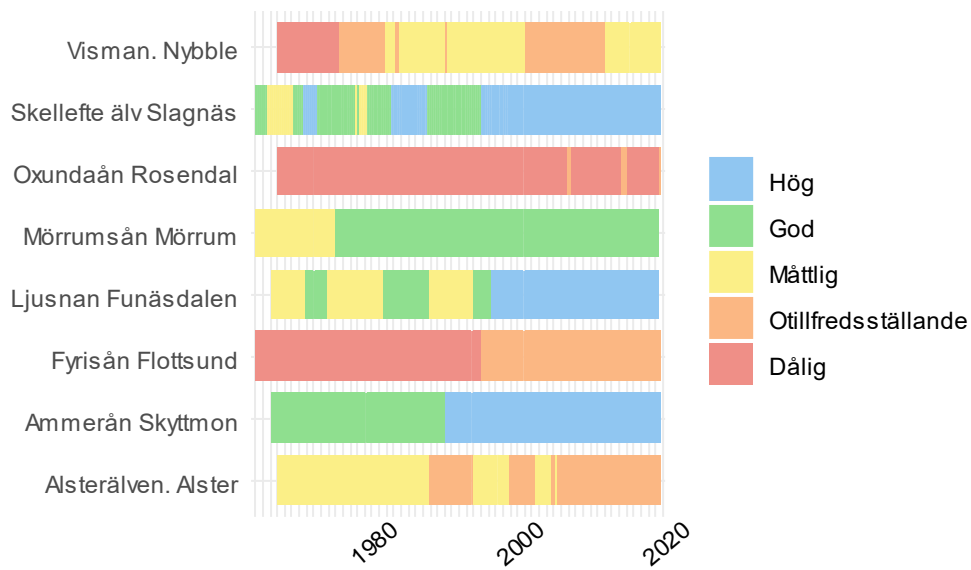
Om statusbedömning bara görs per förvaltningscykel kan dessa bedömningar visas i stapeldiagram, som till exempel för bedömningar av fiskindex som visas för tre bedömningsperioder (1994-2006, 2007-2012, 2013-2018; Figur 13; (Holmgren, 2021))



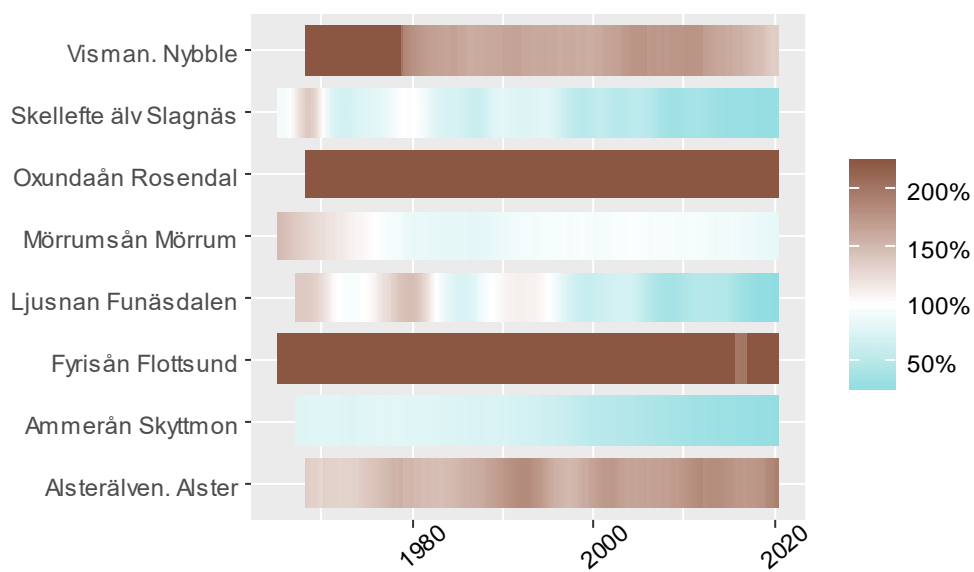
Figur 13: Klassgränser och fiskindex EQR8 på olika provplatser och för tre bedömningsperioder.

## 2.7 Att avgöra hur nära en målnivå man har kommit över tid – spatiala komponenter

När tidsupplösningen är god kan flera serier presenteras samtidigt genom att använda färgkodning som motsvarar statusklassningen (Figur 14) eller som relativ nivå jämfört med värdet av gränsen mellan god och måttlig status (Figur 15). Med konstanta klassgränser kan även presentationen i stapeldiagram användas (Figur 13).



Figur 14: Totalfosfor för åtta serier och deras motsvarande statusklassning för olika tidpunkter.



Figur 15: Nivå av totalfosfor i relation till gränsen mellan god och måttlig status över tid.

## 3 Samband mellan kvalitetsfaktorer och andra miljövariabler

### 3.1 Som hjälpmedel i utvärdering av trender eller måluppfyllelse

#### 3.1.1 Normalisering

För att testa om ett mål är uppnått eller om en förändring över tid är statistiskt signifikant, är det tre aspekter som är del i beräkningarna:

- (i) den sanna skillnaden mellan målet och verkligheten eller den sanna förändringen över tid,
- (ii) variationen i data och
- (iii) stickprovsstorleken.

Om den sanna skillnaden eller förändringen är liten eller om stickprovet är litet blir det alltid svårt att uppnå signifikanta resultat. En möjlighet att förbättra analysen är då att försöka minska variationen i data genom att inkludera förklarande bakgrundsvariabler som leder till naturlig variation i data. Detta kallas ofta normalisering och målet är att presentera medelvärden eller trender om dessa bakgrundsvariabler hade varit konstanta över tiden eller rummet.

Den vanligaste tillämpningen är så kallad flödesnormalisering. För vattenkvalitetsvariabler i vattendrag är nivåerna ofta påverkade av avrinningen vid samma tidpunkt. Att justera för mängden avrinning görs för att minska variationen och skapa en tidsserie som representerar en medelflödesnivå. Målet är att tydliggöra trender som orsakas av andra faktorer än förändringar i avrinningen.

Flödesnormalisering kan göras på olika sätt, men baseras ofta på att sambandet mellan miljövariabeln och den förklarande variabeln kan antas vara linjärt. En modell som ofta används är

$$y_i = \mu_i + \beta \cdot Q_i + \varepsilon_i$$

Där  $y_i$  är miljö-/responsvariabeln,  $Q_i$  är avrinningen och  $\beta$  anger lutningen av regressions sambandet, alltså hur mycket koncentrationen ökar när avrinningen ökar med en enhet.  $\mu_i$  anger den genomsnittliga koncentrationen vid tidpunkt  $i$ .  $\mu_i$  kan antingen antas vara konstant eller representera gradvisa förändringar i responsvariabeln över tiden, det vill säga en trend. Vanligtvis är det fördelaktigt att inkludera en trendskattning i en normaliseringsmodell, då det ger en robustare skattning av  $\beta$  om det finns en trend i responsvariabeln, samtidigt som skattningen inte blir sämre om ingen trend finns (von Brömssen et al., 2023b). En utjämnad trend kan skattas i en generell additiv modell tillsammans med flödesparametern  $\beta$ , som antas vara konstant över tiden.

Oberoende av modellen beräknas de normaliserade värdena som

$$y_{Ni} = y_i - \hat{\beta} \cdot (Q_i - \bar{Q}),$$

där  $\bar{Q}$  representerar medelflödesnivån,  $\hat{\beta}$  är den skattade regressionskoefficienten för flöde och  $y_{Ni}$  står för de normaliserade observationerna av miljövariabeln.  $y_{Ni}$  används i vidare beräkningar av trender eller för att testa om mål är uppfyllda.

Om fördelningen av responsvariabeln är skev, som ofta är fallet för koncentrationer, log-transformeras både den och variabeln för avrinning innan en modell anpassas. För att analysera trender kan flödesnormaliserade värden på log-skala användas direkt. För att kunna använda samma värden för att bedöma måluppfyllelse behöver de först tillbakatransformeras, vilket dock inte är helt oproblemiskt, då det tillbakatransformerade medelvärdet kommer att avvika från det observerade

medelvärdet<sup>1</sup>. Resultaten är biased och underskattar vanligtvis medelvärdet på originalnivå. En ytterligare justering kan göras för att minimera denna bias (HELCOM, 2019)

För miljögifter i organismer har normalisering genom längd, ålder, fett- eller torrinnehåll studerats (Nyberg et al., 2020).

### 3.1.2 Säsongjustering

Många miljövariabler uppvisar säsongvariation. För att undvika att denna, ofta ointressanta, variation påverkar statistiska slutsatser kan man:

- Enbart provta en gång om året, under den säsong som är mest relevant för variabeln, till exempel under sommaren eller under höstomblandningen.
- Beräkna ett årsmedelvärde som används i vidare beräkningar. Observera att det måste finnas lika många observationer för varje år, som är utspridda på samma sätt över året (månatliga eller specifika månader) så att årsmedelvärdena blir jämförbara.
- Inkludera information om månad eller säsong i (trend-)modellen. Modellen utökas då med antingen en kategorisk variabel som indikerar månad eller säsong eller, om provtagningsfrekvensen är tillräckligt hög, en utjämnad cyklisk kurva som beskriver säsongvariationen. För Mann-Kendall-trendtester kan seasonal Mann-Kendall-tester (Hirsch and Slack, 1984) användas när månads- eller säongsdata är tillgängliga.

### 3.1.3 Objektspecifika referensvärden

Normalisering/justering med hjälp av förklarande variabler är svårare när målet är statusklassning, då datamaterialet ofta är för litet för att det ska vara möjligt att skatta robusta samband mellan variablerna. En möjlighet att ändå korrigera för variabler som kan påverka variationen i målvariablerna är att beräkna referensvärden baserat på variabler som beskriver relevanta bakgrundstillstånd för vattenförekomsten eller den specifika stationen.

För näringsämnen i sjöar och vattendrag beräknas objektspecifika referensvärden baserat på ett dataset av opåverkade referenssjöar (Fölster et al., 2021). Förklarande variabler inkluderar vanligtvis absorbans, koncentration av sulfat, koncentration av magnesium och någon eller några variabler som beskriver avrinningsområdet, såsom procent skogsmark eller procent sankmark. Dessutom ingår sjödjup i beräkning av referensvärden för sjöar.

Även för fiske i sjöar används sjöspecifika referensvärden (Holmgren, 2021), som beräknas utifrån sjöarnas höjd över havet, sjöarea, maximalt djup, position i förhållande till högsta kustlinjen och lufttemperaturens årsmedelvärde och amplitud (Havs- och vattenmyndigheten, 2018).

## 3.2 För att identifiera orsaker

Kausala samband kan vanligtvis inte visas med övervakningsdata, dels eftersom data inte samlas in för detta syfte och dels eftersom övervakningsdata inte har tillräckligt fin upplösning i tid och rum för att kunna beskriva komplexa samband och processer. Däremot kan övervakningsdata användas för att skatta och visualisera korrelationer och samband för att generera hypoteser kring vilka processer som kan ligga bakom förändringar. Framtagna hypoteser kan studeras närmare i fältstudier eller genom specifika projekt.

---

<sup>1</sup> Det är möjligt att justera tillbakatransformerade medelvärden för denna systematiska minskning som sker via transformeringen. Vanligtvis återskapas dock ändå inte värden på samma medelnivå som före transformeringen. Ett huvudskäl för transformering är vanligtvis att den skeva fördelningen gör att ett medelvärde ändå inte är representativt för de flesta observationerna.

### 3.2.1 Regressionsmodeller

En regressionsmodell kan inkludera en eller flera förklarande variabler. Med hjälp av regressionsmodeller kan direkta samband (vid samma tidpunkt) mellan olika variabler studeras. En möjlig modell skulle kunna vara genom en additiv modell (GAM)

$$y_i = s(t_i) + \beta \cdot x_i + \varepsilon_i$$

där  $s(t_i)$  står för en utjämnad tidstrend och  $\beta \cdot x_i$  beskriver sambandet till en potentiell förklarande variabel  $x$ . Sambandet kan vara linjärt, som givet i formeln, ha någon annan specificerad form, till exempel exponentiell, eller skattas genom en utjämning, det vill säga ett icke-linjärt samband med okänd form. Modellen kan utökas för att inkludera temporal autokorrelation (avsnitt E.9) i feltermen. Flera förklarande variabler kan läggas till i samma modell.

Linjära mixade (additiva) modeller (sektion 2.3.4) kan användas för att undersöka potentiella förklarande variabler för flera stationer samtidigt om sambanden antas vara liknande för alla inkluderade stationer.

För datamaterial med många observationer används ibland regressionsträd eller random forest för att studera påverkan av ett stort antal variabler på en miljövariabel. Fördelen med dessa modeller är att de identifierar komplexa samband, inklusive interaktioner. Däremot är det inte möjligt att inkludera temporal eller rumslig autokorrelation i modellen. Det är ett mindre problem om data är väl utspridda i rummet (för spatiala dataset) eller tiden (för tidsserier), men kan ge missvisande resultat om data är klustrade i rum eller tid, alltså om provtagningen är tätare i vissa områden eller under vissa tidsperioder. Tidstrender kan inkluderas i modellerna genom att tidsvariabeln ingår bland de förklarande variablerna. Den främsta begränsningen för användandet av random forests är dock att miljöövervakningsprogrammen generellt inte genererar tillräckligt med (oberoende) data för att identifiera komplexa samband, samt att modellernas främsta syfte är prediktering. Modellerna genererar inte heller någon modellekvation. Detta innebär att variabler som har ett samband med responsvariabeln kan identifieras, men den exakta formen hos sambandet kan inte beskrivas.

Regressionsmodeller är användbara om den omedelbara verkan av förklarande variabler på en miljövariabel ska skattas. När påverkansprocesser är mer komplexa brukar det vara svårare att bygga bra men enkla regressionsmodeller. Relevanta aspekter som bör tas hänsyn till vid modellering av potentiell påverkan är:

- Att det kan finnas tidsfördröjda effekter.
- Att trender eller nivåer som observeras är påverkade av input och processer som ligger uppströms (både tidsfördröjda effekter och inkomplett information).
- Att påverkan som leder till långfristiga förändringar lätt överskuggas av korttidsvariation som orsakas av till exempel väder eller säsongfluktuationer.
- Att påverkan eller samband mellan variabler kan förändras över tiden, det vill säga att de inte alltid kan anses vara konstanta.
- Att bakgrundsvariabler kan uppvisa korrelationer till varandra (multikollinearitet) vilket gör det svårare att identifiera och särskilja effekterna av enskilda variabler. Till exempel kan det finnas korrelationer mellan temperatur och nederbörd (det är vanligare med antingen varmt och torrt eller svalt och regnigt väder än de andra kombinationerna).

#### 3.2.1.1 Resultat från mekanistiska modeller som indata i regressionsmodeller

I viss mån kan deterministiska modeller användas för att representera komplexa processer i bakgrundsvariabler. Modelloutput kan sedan användas för att beskriva variation i observerade data och förtydliga trender som inte kan förklaras av kända och i modellen inkluderade processer.

Modeller för läckage av näringsämnen och rainfall-runoff modeller har testats främst för målet att normalisera observerade data. Resultaten är dock känsliga för modellernas kvalitet och syfte och deras kalibrering (Letcher et al., 2001; Libiseller, 2003). Komplexiteten i att använda deterministiska, mekanistiska modeller för att beskriva naturlig variation och begränsningen av användbarheten av resultaten gör att sådana analyser sällan görs.

### 3.2.2 Multivariat statistik

Multivariat statistik är en gren inom statistiken som tillåter studium av många, möjligen högt korrelerade, variabler samtidigt. Metoder som principalkomponentanalys (PCA), partial least squares (PLS) eller non-metric multidimensional scaling (NMDS) kan användas för att visualisera samband mellan variabler genom att identifiera bakomliggande strukturer. I samband med trendanalys används dessa metoder för att söka kopplingar mellan en variabel som kvantifierar trenden på en station, till exempel genom en regressionskoefficient eller en Theil-Sen-slope, och ett antal variabler som beskriver egenskaper hos stationen eller avrinningsområdet.

Det är också möjligt att extrahera faktorer ur modellerna för att använda dessa som förklarande variabler i regressionsmodeller, till exempel i syfte att minska naturlig variation i data och förtydliga trender som beror av annat än de inkluderade bakgrundsvariablerna.

### 3.2.3 Temporalt samvarierande trender

Genom gemensam visualisering av trender i en miljövariabel och en eller flera potentiella förklarande variabler kan mönster i den gemensamma utvecklingen över tid undersökas. Detta kan till exempel göras med liknande visualiseringar som beskrivits i sektion 2.3.3 men genom att presentera olika variabler istället för olika stationer. Dessa grafer kan också underlätta identifiering av eventuella tidsförskjutningar som kan ge insyn i bakomliggande processer.

Korrelation mellan olika långtidstrender skulle kunna vara intressant att studera. Det är dock ofta svårt att göra på ett pålitligt sätt då det som kvantifieras snarare är den kortvariga samvariationen, som kanske styrs av väderförhållanden eller säsongsvariation. Att normalisera eller säsongjustera tidsserier förbättrar möjligheterna att studera korrelationer mellan tidstrender. En annan möjlighet att illustrera samvariation mellan tidstrender är att begränsa analyserna till års(medel)värden och därmed exkludera mycket av korttidsvariationen. Till exempel Asmala et al. (2019) har använd årsmedelvärden för att studera sambandet mellan trender i organiskt kol och andra variabler.

### 3.2.4 Geografiskt samvarierande trender

När den spatiala upplösningen tillåter det kan tidstrender tolkas i sin geografiska kontext. För detta kan kartor användas (sektion 2.3.1). Antingen kan trendkoefficienter beräknas per station och plottas i kartor eller så kan trenderna beräknas för geografiska fönster, så kallade geografiskt viktade regressionsmodeller. Använder man det senare är det viktigt att komma ihåg att man redan har genomfört en rumslig utjämning av observerade trender. Enskilda punkter i kartan kan då inte kopplas till lokala bakgrundsvariabler, utan bedömningen måste ske för större geografiska områden.

Lokala förklarande variabler kan också inkluderas i geografiska trendmodeller för att åstadkomma en skattning av trender som är justerade för den valda förklarande variabeln. Med hjälp av sådana "multiscale" geografiskt viktade regressionsmodeller (Comber et al., 2023; Fotheringham et al., 2017) kan det undersökas om trenderna i en variabel fortfarande är synliga när relevanta förklarande variabler har inkluderats. Då dessa modeller fortfarande är ganska nya och beräkningstunga behöver det forskas mer om deras användbarhet i miljöövervakningen.



### 3.2.5 Tematiskt samvarierande trender

Tematisk samvariation kan ske genom olika grupperingar eller samvisualisering av olika serier, som nämnts tidigare. Trendanalyserna kan göras gemensamt för en grupp av stationer som har liknande belastning med mixade linjära modeller (sektion 2.3.4).

### 3.2.6. Förändringar i samband mellan en förklarande och en miljövariabel över tid

Att förstå hur relationen mellan två variabler förändras över tid kan också bidra till förståelse av vilka källor och vilka transportvägar som är viktiga. När relationen mellan till exempel näringsämneskoncentrationer och avrinning ändras kan det tyda på förändringar i förhållandet mellan punkt- och diffusa föroreningskällor eller så kan det vara ett tecken på effekter av åtgärdsprogram. Några exempel på hur sådana tidsvarierande samband kan skattas beskrivs av von Brömssen et al. (2023b), men ofta kommer förändringar i samband mellan variabler att vara små och svåra att studera.

## 4. Kvalitetsindikatorer för övervakningsprogram

Kvaliteten hos ett övervakningsprogram kan definieras med avseende på olika dimensioner. Generella kvalitetsindikatorer kan definieras i syfte att datainsamlingen och datahanteringen ska uppnå kvalitet genom en väldefinierad process, där olika kvalitetsdimensioner diskuteras och förbättringsinsatser genomförs när det är nödvändigt (Biemer and Lyberg, 2003; European statistical systems, 2019). I ett större perspektiv utgör en kvalitetsdeklaration en del av ett cykliskt arbetssätt (SCB, 2023) som illustreras i Figur 16, nämligen utvärderingen av utfallet, som ska ge återkoppling till kommande år. Huvudsyftet med kvalitetsdeklarationen och andra utvärderingar av utfallskvaliteten är att underlätta diskussioner kring hur övervakningsprogrammen ska utvecklas över tid för att möta nuvarande och framtida behov.



Figur 16: Ett cykliskt arbetssätt för att garantera kvaliteten hos officiell statistik (SCB, 2023)

För närvarande görs kvalitetsdeklarationer inom miljöövervakningen för sötvatten enbart för officiella statistikprodukter, som är specifika undersökningar med ett mycket väl definierat och smalt syfte. I avsnitt 4.1 utgår vi ifrån de begrepp som används i kvalitetsdeklarationerna för officiell statistik, men kvalitetsdeklarationer för program har inte tidigare gjorts inom området och tydliga exempel på hur sådana kan tas fram saknas. Förslag på – och en instruktion för – en kvalitetsdeklaration för övervakningsprogram i sötvatten behöver utvärderas och anpassas efter att den har testats på några program.

I avsnitt 4.2 och 4.3 presenteras ytterligare kvalitetsaspekter som kan beräknas genom enklare statistiska metoder, samt en utökad kvalitetsdeklaration som inkluderar ytterligare faktorer som kan användas för att bedöma kvaliteten hos ett övervakningsprogram.

### 4.1 Generella kvalitetsindikatorer

Kvalitetsdeklarationer tas redan fram för officiella statistikprodukter inom Havs- och vattenmyndighetens ansvarsområde, såsom transport av näringsämnen till havet, fisk i sjöar, och olika ämnen i grundvatten<sup>2</sup>. De kvalitetsdimensioner som då beaktas är relevans, tillförlitlighet och riktighet, aktualitet och punktlighet, tillgänglighet och tydlighet, samt jämförbarhet och sammanvändbarhet. Var och en av dessa dimensioner beskrivs kortfattat i följande avsnitt, med exempel på hur de skulle kunna användas för ett övervakningsprogram istället för en specifik statistikprodukt.

<sup>2</sup> <https://www.havochvatten.se/data-kartor-och-rapporter/data-och-statistik/officiell-statistik/officiell-statistik--havs--och-vattenmiljo.html>

#### 4.1.1. Relevans

Relevansen hos ett övervakningsprogram handlar om att programmet genererar data som tillåter en utvärdering av programmets huvudsakliga syfte. Detta syfte kan vara analys av trender eller en bedömning av tillståndet i miljön. Programmets relevans ska också bedömas i enlighet med de lagkrav som ställs på övervakningsprogrammet.

#### 4.1.2 Tillförlitlighet och riktighet

Riktighet handlar om hur väl observerade variabler i ett program representerar verkligheten, och är en av huvudkomponenterna i hur tillförlitliga analyserna gjorda på programmets data blir. För att bedöma riktigheten kan man exempelvis fokusera på följande aspekter:

- att rätt objekt provtas för att representera en målpopulation. Målpopulationen kan till exempel vara alla sjöar med liknande belastning och förutsättningar.
- att rätt provplats väljs ut för att representera objektet.
- att provet tas vid rätt tillfälle, speciellt om kvalitetsfaktorer uppvisar stor säsongsvariation eller om observerade värden kan variera starkt av andra skäl.
- att tillräckligt många prov tas under året om målet är att beräkna årsmedelvärden, årstransporter eller att kvantifiera säsongsvariation.

Målpopulationen, valet av platser och tider, samt provtagningsfrekvensen, kommer att variera mellan olika övervakningsprogram beroende på programmets syfte och ekonomi och på målpopulationens tillgänglighet. Vilka objekt, provplatser och tidpunkter som är representativa för övervakningsmålet bör utvärderas regelbundet med hänsyn till de syften som programmet har och de förändringar som sker över tid. Utvärderingen är avhängig en tydlig beskrivning av syftet med programmet.

För trendanalyser spelar riktighet mindre roll, då bedömningen baseras på relativa förändringar. Däremot är det naturligtvis fortfarande viktigt att välja stationer som kan representera de tidsförändringar som ska kvantifieras.

#### **Exempel på beskrivning av tillförlitlighet och riktighet**

I kvalitetsdeklarationen för officiell statistik över tillförsel av kväve och fosfor till Östersjön (Havs- och vattenmyndigheten, 2022a) beskrivs målpopulationen som "alla Sveriges flodmynningar och stora punktkällor med direktutsläpp till havet". Inget urval av stationer görs utan alla flodmynningar undersöks.

För statistikprodukten "fisk i sjöar" (Havs- och vattenmyndigheten, 2022b) utgör målpopulationen mindre svenska sjöar som är relativt opåverkade av punkutsläpp. Ett urval av sjöar görs som en delmängd av de sjöar som ingår i övervakningsprogrammet trendsjöar och av sjöar som har en mindre sjöarea (10-730 ha). Praktiska och ekonomiska aspekter på provtagningen påverkar också valet av val av sjöar.

#### 4.1.3 Aktualitet och punktlighet

Aktualitet och punktlighet handlar om hur ofta och när statistik eller data blir tillgängliga för intressenter.

För produkter inom officiell statistikproduktion är en viktig kvalitetsaspekt att den beräknade statistiken levereras inom en viss tidsram för att garantera att den är aktuell. Vanligtvis redovisas hur långt efter referensperiodens slut statistiken publicerades, hur ofta den publiceras och om publiceringen gjordes enligt tidsplan.

För övervakningsprogram generellt kan aktualitet avse tidsramen tills data blir tillgängliga för myndigheter eller allmänheten genom publicering i databaser efter rimlig kvalitetsgranskning.

#### 4.1.4 Tillgänglighet och tydlighet

Tillgänglighet och tydlighet handlar om möjligheten att få tillgång till statistiken eller data, samt möjligheten att få tillgång till relevanta metadata.

För officiell statistik beskrivs hur statistiken görs tillgänglig inklusive dess dokumentation, vanligtvis som en webbsida. Det finns också en tidsplan för när statistikprodukten ska vara tillgänglig och vad statistiken ska redovisa.

För övervakningsprogram generellt kan det vara aktuellt att beskriva hur data tillgängliggörs för de användare och inom de användningsområden som beskrivningen av delprogrammen anger. Exempel på användare är olika myndigheter, länsstyrelser, kommuner eller forskningssamhället. Tillgänglighet kan diskuteras utifrån databasens funktionalitet och öppenhet, samt utifrån hur snabbt data blir tillgängliga via datavärden. Tydlighet kan avse att information om mätmetoder, enheter och ändringar i datainsamlingsprocessen är lätt tillgängliga. Det inkluderar även tillgång till metadata som beskriver de utvalda objekten, både för att resultat av dataanalyser ska kunna kopplas till bakgrundsvariabler och för att det ska vara möjligt att på ett lättare sätt utvärdera att de objekt som övervakas (fortfarande) är representativa för programmets syften.

#### 4.1.5 Jämförbarhet och sammanvändbarhet

Jämförbarhet handlar om möjligheten att göra jämförelser mellan olika referenstider eller olika grupper. För att tidstrender ska kunna beräknas måste data vara jämförbara inom en tidsserie. Det innebär att samma metod för provtagning ska användas; kemiska och biologiska analyser ska genomföras på likartat sätt för varje prov; provtagning ska ske vid samma plats inom en sjö eller ett vattendrag, och så vidare. Jämförbarhet är viktigt för att man ska kunna dra slutsatsen att observerade förändringar avser förändringar i miljön och inte artificiella nivåskillnader på grund av förändringar i provtagningen eller mättekniken. Om förändringar i metodik blir nödvändiga är det viktigt att planera dessa i god tid för att garantera kontinuitet i tidsserien, till exempel genom att under en överlappande tidsperiod studera om de olika metodikerna leder till systematiska skillnader i producerade data.

För data som samlas in i syfte att användas för statusklassning behöver man väga jämförbarhet mot riktighet och provtagningskvalitet. Om nyare och bättre provtagningsmetoder finns tillgängliga så kan det vara rimligt att byta även om jämförbarheten över tiden blir lidande.

Jämförbarhet kan också vara viktigt mellan olika program. Provplatser, provvariabler och mätmetoder bör överensstämma när samma eller liknande variabler mäts i olika program. Endast om data är jämförbara kan data producerade i olika program sammanvändas.

Sammanvändbarhet handlar om möjligheten att kombinera och tillsammans analysera data från olika övervakningsprogram. Sammanvändbarhet handlar precis som jämförbarhet om att använda standardiserade beteckningar (till exempel namn och enheter för mätta variabler) och – om möjligt – gemensamma stationsregister.

## 4.2. Kvalitetsindikatorer som kan beräknas genom enklare statistiska metoder

När data eller tidsserier ska bedömas efterfrågas ofta statistiska beräkningar för att kvantifiera olika aspekter. Ett antal möjliga frågeställningar redovisas nedan. Det är viktigt att komma ihåg att alla dessa beräkningar är teoretiska, det vill säga att de kan vara till hjälp för att förstå datamaterialets

eller övervakningsprogrammets potential, men de kan inte i sig själva avgöra om programmet är bra eller relevant.

#### 4.2.1. Riktighet

Det enda kvantitativa sättet att bedöma om en mätning eller medelvärdesskattning är riktig, alltså inte avviker systematiskt från det sanna värdet, är genom att göra en omfattande undersökning där alla eller nästan alla objekt i en population ingår. En sådan undersökning är vanligtvis inte genomförbar. Riktighet måste därför bedömas utifrån diskussionen kring tillförlitlighet och riktighet (4.1.2).

#### 4.2.2. Stickprovsstorlek, precision och medelfel

Precision avser hur mycket observationerna varierar och mäts genom stickprovsvariansen eller stickprovsstandardavvikelsen för en utvald miljövariabel. Precisionen blir högre för avgränsade frågeställningar, till exempel om provtagningen kan koncentreras till en utvald säsong.

Precision kan avse både grunddata (hur mycket variation som finns i insamlade data, det vill säga stickprovsvariansen) eller en skattning som beräknas på data (hur mycket denna skattning skulle variera om datainsamling gjordes om). Precisionen hos en skattning beräknas med hjälp av stickprovsvariansen och stickprovsstorlek och anges genom medelfel. Olika typer av skattningar (medelvärde, andel, regressionskoefficient, andelar med flera) har olika formler för hur medelfelet beräknas. I följande avsnitt används huvudsakligen medelvärde som ett exempel. Koncepten är liknande för andra skattningar.

Precisionen hos ett medelvärde kvantifieras som medelfel (eng.: standard error of the mean) enligt följande formel:

$$SE(\bar{x}) = \sqrt{\frac{s^2}{n}} ,$$

där  $\bar{x}$  är medelvärdet,  $s^2$  är stickprovsvariansen och  $n$  är stickprovsstorleken. Medelfelet blir alltid mindre ju fler observationer<sup>3</sup> som finns.

Observera att skattningens medelfel enbart anger hur mycket skattningen skulle variera om nya liknande datainsamlingar skulle utföras. Den kvantifierar inte om skattningen är riktig, alltså fri från systematiska fel.

Om den statistik man är intresserad av är en skattning i en modell, till exempel regressionskoefficienten i en linjär eller log-linjär regression, så används variansen av residualerna<sup>4</sup> för att kvantifiera osäkerheten istället för stickprovsvariansen i beräkningen av medelfelet. När skattningen avser en andel finns specifika formler för andelens varians.

#### Exempel på precision som kvalitetsmått

I CIS guidance 7 (European Commission, 2003) definieras precision som halva bredden hos ett konfidensintervall, det vill säga som medelfel. Det baseras i så fall på att det redan i förväg är bestämt att den statistik som ska tas fram är ett konfidensintervall. Konfidensintervallet kan då gälla

---

<sup>3</sup> Observationerna ska vara oberoende, alltså uppvisa liten eller ingen temporal eller spatial variation, för att formeln ska vara korrekt.

<sup>4</sup> Residualerna är de individuella skillnaderna mellan observationerna och värdet som predikteras ur modellen. De representerar hur mycket av variationen i responsvariabeln som inte kunde förklaras av modellen.

- ett medelvärde, som kan antas vara normalfördelat eller följer annan lämplig teoretisk fördelning (utan transformation av data).
- en andel, som kan antas vara normalfördelat eller följa annan lämplig teoretisk fördelning.
- en regressionskoefficient i, till exempel, en linjär eller log-linjär regression
- en annan skattning i en relevant statistisk modell

Det är alltså underförstått att den statistiska metoden är fastställd i samband med att ett krav på (halva) bredden av ett konfidensintervall ställs.

För ett medelvärde kan konfidensintervallet beräknas som

$$\bar{x} \pm t_{1-\frac{\alpha}{2}, n-1} \cdot \sqrt{\frac{s^2}{n}}$$

och halva bredden hos konfidensintervallet är då

$$t_{1-\frac{\alpha}{2}, n-1} \cdot \sqrt{\frac{s^2}{n}}$$

där  $\sqrt{\frac{s^2}{n}}$  är medelvärdets medelfel och  $t_{1-\frac{\alpha}{2}}$  är en kvantil i t-fördelningen med n-1 frihetsgrader.  $1 - \alpha$  betecknar konfidensnivån. När stickprovsstorleken ökar blir medelfelet mindre och t-fördelningen smalare – konfidensintervallet blir smalare.

#### 4.2.3 Stickprovsstorlek och statistisk styrka

Ofta görs statistiska hypotestest i utvärderingen av data, till exempel testas om en trend är signifikant eller om det sanna medelvärdet för en miljövariabel ligger signifikant över eller under gränsen för god/måttlig status. Statistiska styrkeanalyser är kopplade till hypotestest och ger svar på hur ofta signifikanta resultat kan förväntas om en studie – hypotetiskt – skulle genomföras ett stort antal gånger med samma villkor, det vill säga med samma målpopulation, samma provtagningsfrekvens och samma mätmetodik.

Styrkeanalyser används främst i planeringen av nya datainsamlingar med specifika krav på vad producerade data ska kunna användas till. Om, till exempel, huvudsyftet för ett program är att kunna upptäcka en linjär trend på 5 enheter under en tidsperiod av 10 år, så är det rimligt att genomföra styrkeanalyser för att beräkna sannolikheten att påvisa en sådan förändring om en sådan förändring faktiskt äger rum. För att kunna göra beräkningarna behövs en god uppfattning om hur stor naturlig variation som förväntas i den relevanta miljövariabeln. Ofta bestäms styrkan genom datasimuleringar, men för enkla frågeställningar kan den också beräknas analytiskt.

Sannolikheten (styrkan) att kunna upptäcka en specifik förändring med en viss statistisk metodik beror på stickprovsstorleken, variationen i data och hur stor den förväntade förändringen av miljövariabeln är. Den beräknade sannolikheten avser hur ofta motsvarande statistiska test skulle vara signifikant om datainsamlingen upprepades med samma villkor och under samma naturliga variation ett stort antal gånger. Sannolikheten är inte applicerbar på ett enskilt fall.

Styrkeberäkningar görs enklast för parametriska statistiska metoder, såsom linjär eller log-linjär regression eller test för medelvärden eller andelar. Även för icke-parametriska metoder, så som Mann-Kendall-test, kan styrkan simuleras. För mer komplexa modeller, till exempel additiva modeller

där det inte är en specifik parameter som ska skattas, och där det inte heller finns något enskilt test att beräkna styrkan för, är styrkeanalyser vanligtvis inte relevanta.

#### **Exempel på styrkeanalyser i miljöanalysen:**

(i) Grandin (2006) beskriver hur styrkeberäkningar för regionala trendanalyser i sötvattensprogram kan genomföras. Som analysmetod använder han en mixad modell (GLMM) där år är fix faktor och övervakningsstation en slumpmässig faktor. Genom att man använder den fixa faktorn år som en kategorisk variabel behöver trenden inte vara linjär eller av någon annan förbestämd form, vilket dock gör det svårare att kvantifiera trenden med hjälp av en enskild parameter. Den slumpmässiga faktorn övervakningsstation tillåter att de olika stationerna inom en region har olika genomsnittsnivåer för den studerade variabeln (olika intercept). Modellen antar att trenderna inom regionen är liknande, det vill säga att alla stationer uppvisar ungefär samma utveckling över tid. Varianserna (den slumpmässiga variationen och variationen mellan stationerna) bestämdes ur befintliga övervakningsdata.

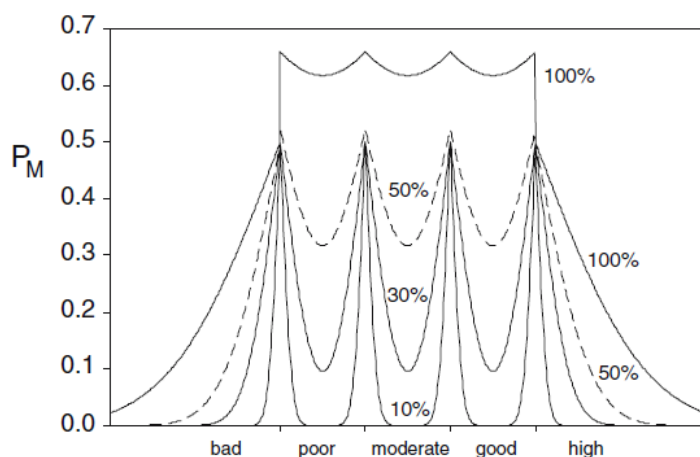
(ii) Statistisk styrka har också använts för att jämföra olika utformning (designer) hos övervakningsprogram, till exempel för att välja mellan "serially alternating"-designer (som liknar omdrevsdesignen i Sverige) och andra designer som innehåller en blandning av omdrev och stationer som övervakas oftare (Starceвич et al., 2018; Urquhart and Kincaid, 1999). Som statistisk modell använd vanligtvis en GLMM med gemensam linjär trend för alla stationer och slumpmässig intercept för att fånga upp nivåskillnader mellan stationerna. Styrkan som beräknas avser en viss förändring för den regionala, gemensamma trenden.

(iii) Några beräkningar av styrka under olika förutsättningar beskrivs i Appendix A.

#### **4.2.4 Stickprovsstorlek och teoretiska sannolikhetsberäkningar**

Teoretiska sannolikhetsberäkningar för måluppfyllelse illustreras av Clarke and Hering (2006). Sannolikheterna avser sannolikheten att ett observerat medelvärde av en miljövariabel ligger i olika statusklassningar givet att det sanna medelvärdet av miljövariabeln och dess variation är känt. Klassningen görs med hjälp av face-value: om det observerade medelvärdet ligger i klass 'god' så blir statusklassningen 'god'. Fördelningen av det observerade medelvärdet följer samplingfördelningen (se avsnitt E.5).

Om det sanna medelvärdet ligger exakt på en klassgräns är sannolikheten att det klassas i någon av de intilliggande klasserna 50% förutom om variationen är stor. Om det sanna medelvärdet ligger i mitten av en klass blir sannolikheten att felaktigt välja en annan klass utifrån det observerade värdet mindre (Figur 17). Eftersom det sanna värdet av miljövariabeln inte är känt kan principen inte användas för att beräkna sannolikheter i verkliga undersökningar, men illustrationen ger en bild av hur felklassningssannolikheterna ökar ju närmare det sanna värdet kommer en klassgräns och för olika antaganden om den naturliga variationen. Exempel på simulerade felklassningssannolikheter finns i avsnitt B i Appendix.



Figur 17: Från Clarke and Hering (2006) "Errors and uncertainties in bioassessments methods". Det sanna medelvärdet av en miljövariabel visas på x-axeln. När detta sanna värde ligger nära en klassgräns är det svårt att avgöra den sanna statusklassen – sannolikheten att felklassa blir hög. Om det sanna värdet ligger i klassmitten blir sannolikheten att felklassa lägre. Även stickprovsvariansen, eller medelvärdets medelfel spelar roll.

#### 4.2.5. Bestämning av nödvändig stickprovsstorlek

Ibland kan det vara önskvärt att anpassa stickprovsstorleken för en befintlig undersökning eller bestämma stickprovsstorleken för en kommande undersökning. Vi kan skilja mellan två olika fall:

- (i) stickprovsstorleken måste öka för att kunna uppnå önskade nivåer på medelfel, felmarginal eller styrka
- (ii) stickprovsstorleken måste minska baserat på ekonomiska eller praktiska grundförutsättningar.

##### 4.2.5.1 Ökad eller minskad stickprovsstorlek för trendanalyser

För trendanalyser är det ofta svårt att öka stickprovsstorleken, då det önskade medelfelet eller styrkan avser en skattning om förändring inom ett visst tidsintervall. Om till exempel trendanalysen avser en förändring under ett 10-årsintervall baserat på årsmedelvärden så kan stickprovsstorleken inte vara större än 10. Trendanalyser som avser mindre än 10 observationer är vanligtvis inte att rekommendera<sup>5</sup>, då trendskattningar i sådana fall har mycket variation och lätt kan bli vilseledande. Dessutom är resultaten sällan statistiskt signifikant när de baseras på små underlag.

En möjlighet är att öka tidsseriens upplösning, till exempel genom att analysera månadsvärden istället för årsvärden om det är möjligt. Med fler observationer kan analyserna få högre styrka och skattningarna lägre medelfel. På grund av säsongsvariation, ökad naturlig variation och ökad temporal autokorrelation blir vinsten dock inte nödvändigtvis stor.

Om trenderna inom ett geografiskt område kan förväntas vara likartade kan styrkan eller skattningarnas medelfel förbättras genom att man analyserar flera stationer tillsammans. Det är då viktigt att välja en modell som kan ta hand om den systematiska variationen mellan stationerna, till exempel en linjär mixad modell.

<sup>5</sup> Mann-Kendall-test ger korrekta resultat endast om de baseras på minst 10 årliga observationer eller 4 år av månatliga observationer. Linjära regressioner rekommenderas vanligtvis inte heller att användas med mycket små dataset, då det lätt leder till överanpassning.



#### *4.2.5.2 Ökad eller minskad stickprovsstorlek för statusklassning*

För att statusklassa objekt kan stickprovsstorleken anpassas enklare genom att ytterligare prover tas när stickprovet från början var för litet för att uppnå önskat medelfel eller styrka. Det är viktigt att hålla sig till samma provtagningsprotokoll, det vill säga välja relevanta säsonger eller ytterligare provplatser utvalda enligt samma principer som tidigare.

#### *4.2.5.3 Andra aspekter att beakta när stickprovsstorleken ska anpassas*

Många övervakningsprogram producerar data som inte bara används för att analysera trender eller statusklassa (huvudsyftena), utan även används i andra sammanhang. Om stickprovsstorleken inom ett program kan minskas, samtidigt som det ändå är möjligt att göra de analyser som utgör programmets huvudsyfte, bör man även göra en konsekvensanalys för eventuella viktiga andra tillämpningar. Ett exempel skulle kunna vara att trender i näringsämnen kan upptäckas lika bra om data samlas in varannan månad istället för varje månad. En sådan förändring påverkar dock möjligheten att beräkna bland annat årstransporter, där det är viktigt att ha tillräckligt frekventa underlagsdata för att kunna göra en pålitlig skattning av nivån, inte bara av relativ förändring.

#### *4.2.6. Minskning av variation genom normalisering*

För att uppnå önskad nivå av medelfel eller styrka kan även normalisering övervägas (sektion 3.1.1). För att normalisering ska kunna genomföras måste det redan finnas ett tillräckligt stort dataunderlag för att kunna etablera normaliseringsmodellen.

### *4.3 Utökad kvalitetsdeklaration*

#### *4.3.1 Att övervaka för framtiden*

Förutsättningarna för att genomföra övervakning och att övervaka rätt sak ändras över tiden. Nya frågeställningar och nya fokusområden blir aktuella och tillgänglig metodik förändras. Därför är det viktigt att inte bara fundera på om ett övervakningsprogram uppfyller kvalitetskraven just nu, utan även kartlägga vilka aspekter som kommer att förändras inom de närmaste åren. Några möjliga diskussionspunkter finns i följande avsnitt.

##### *4.3.1.1 Metodförändringar för kemiska eller biologiska analysmetod*

Kemiska eller biologiska analysmetoder utvecklas för att ge bättre eller mer robusta svar. Medan sådan förändring vanligtvis är bra, behöver byte av metod planeras och om förändringen är väsentlig behöver den gamla och nya metoden användas gemensamt under en tidsperiod för att verifiera att resultaten är likvärdiga, garantera kontinuitet och jämförbarhet i tidsserier och för att möjliggöra att gå tillbaka till den gamla metoden om det finns problem i omsättningen av den nya.

##### *4.3.1.2 Förändrade frågeställningar*

I framtiden kan fokus för miljöövervakningen skifta. Att tidigt kunna ta hänsyn till nytillkomna frågor och planera för hur ett befintligt program skulle kunna leverera relevanta data för att studera dessa underlättar för framtiden. Till exempel kan variabler eller områden som i dagsläget är nedprioriterade bli viktigare i framtiden, eller nya variabler kan behöva observeras. Om större fokus behöver läggas på extremvärden i miljövariabler, snarare än medelvärden, kan provtagningsfrekvensen behöva anpassas för att täcka fler av scenarierna där mycket höga eller låga värden observeras.

##### *4.3.1.3 Förändringar i bakgrundsvariabler som påverkar insamling eller tolkning av data*

En utmaning för miljöövervakningen är att bakgrundstillståndet i de övervakade områdena kan komma att ändras över tid. Mindre förändringar sker hela tiden: markanvändningen i avrinningsområdet förändras, föroreningskällor tillkommer i en tidigare opåverkad sjö eller i ett

vattendrag. Dessa behöver bedömas och ibland behöver stationer uteslutas eller provtagningsplatsen anpassas. Denna process ingår i den vanliga bedömningen av att stationerna ska vara representativa.

Mer generella förändringar av bakgrundsvariabler kan förväntas till följd av klimatförändringar. Till exempel görs viss övervakning under utvalda månader eller säsonger för att kunna observera tillståndet under vissa förutsättningar, såsom under snösmältning, under sommaren eller vid höstomblandningen. Eftersom det inte är möjligt att veta när dessa förutsättningar inträffar i praktiken under ett visst år tas proverna under en månad där det är mest sannolikt att påträffa dem. När klimatförändringarna gör att det blir mindre snö som smälter tidigare eller påverkar när algblomningen börjar och hur länge den pågår förändras även förutsättningarna för övervakningen. Beroende på målet med övervakningen kan det vara relevant att fundera över att förändra tidpunkterna för mätningar för att uppnå jämförbarhet över åren. Det är dock en svår fråga då jämförbarhet i tiden även avser att det mäts vid samma tidpunkt varje år.

En ytterligare förändring av bakgrundstillståndet, även den kopplad till klimatförändringar, är att mer extrema väderförhållanden observeras mer ofta. En sådan förändring leder till större variation i insamlade data, både inom och mellan åren. För övervakningsprogrammet kan det innebära att det blir svårare att hitta signifikanta trender i tidsseriernas medelnivå på grund av den ökade variationen. Dessutom kan andra statistiska frågeställningar bli viktigare, såsom att analysera trender i de lägsta eller högsta värden som observeras, speciellt om dessa utgör situationer som har negativ påverkan på, till exempel, biodiversiteten eller fiskbeståndet.

#### 4.3.2 Samanvändning

Samanvändbarhet, alltså möjligheten att använda data från olika program i gemensamma analyser, har redan nämnts tidigare. Samanvändning syftar på att sådana gemensamma analyser faktiskt genomförs. Detta kan vara i jämförande studier, där trender i liknande variabler som samlas in i olika program studeras, eller att data från ett program utgör viktiga underlagsdata eller hjälpinformation för analyser som görs inom andra program.

#### 4.3.3 Att förstå orsak och verkan

Vanligtvis är övervakningsprogram inte dimensionerade för att kunna förstå vad som orsakar förändringar eller avvikelser från god status. I den mån program ändå kan användas för att ge underlag för att generera eller studera hypoteser om orsaker bör det lyftas fram och beskrivas. I många fall kommer det att vara en fråga om gemensam visualisering av ett flertal serier eller stationer för att förstå temporala eller spatiala samband, som i sin tur kan studeras vidare i fältstudier eller specialiserade projekt.

#### 4.3.4 Forskning baserad på övervakningsdata

Om forskning bedrivs baserat på data från övervakningsprogram bör även det lyftas i en utökad kvalitetsdeklaration. Medan forskning aldrig kan vara ett huvudsyfte när ett övervakningsprogram genomförs, så ger ny kunskap genom forskning även underlag för förbättring av befintliga program. Att data används av många olika aktörer gör också att dessa data kvalitetsgranskas ytterligare, till exempel kan olika utvärderingsmetoder lyfta fram avvikande värden som inte observeras i grunddata. Att övervakningsdata förekommer i forskningen bekräftar även att dessa datas tillgänglighet och kvalitet är hög.

## 5. Förslag till metod för att utvärdera miljöövervakningsprogram

En utvärdering av miljöövervakningsprogram består av (minst) två olika delar:

- a) En statistisk utvärdering av de data som produceras av programmet.
- b) En utvärdering av om programmet är ändamålsenligt och anpassat till de krav som ställs på det.

### 5.1 Förslag till metod för statistisk utvärdering av producerade data

Hur en statistisk utvärdering av data ska genomföras beror bland annat på målet med övervakningen, den temporala och spatiala upplösningen av data, typiska dataegenskaper hos variabler (till exempel teoretisk fördelning, förekomst av extremvärden och förekomsten av värden under detektionsgräns) och antaganden om samband eller trender (linjära, monotona, icke-monotona). Utvärderingen av data måste alltså anpassas till de data som samlas in för att vara meningsfull och informativ för den grundläggande frågeställningen. Det är rekommenderat att en sådan utvärdering görs regelbundet.

Eftersom den optimala utvärderingen av data görs med olika metoder för olika övervakningsprogram och presenteras genom olika typer av figurer och tabeller är resultaten ofta inte lätt jämförbara och kan inte tolkas över programgränser. Det är därför en fördel att även använda generella och standardiserade metoder som är samma eller liknande för flera delprogram, antingen genom en detaljerad analys av flera program samtidigt eller enligt en förbestämd metodik i separata utvärderingar.

Används standardiserad och förbestämd statistisk metodik ska den fokusera på att visualisera och beskriva data på ett sätt som är jämförbart över olika delprogram, även om datastrukturen skiljer sig, och som utgår från så få strikta antaganden som möjligt. Det innebär att det, till exempel, inte är rimligt att utgå ifrån att alla serier i alla delprogram uppvisar en linjär trend. Att göra linjära regressionsanalyser för alla serier skulle då snarare ge missvisande resultat som inte går att tolka. Analyserna ska alltså, i den mån det är möjligt, vara "data-drivna", alltså baserade enbart på data och enklare sammanfattningar. En sådan standardiserad analys är främst av intresse för data som är insamlad i tidsserieform eller med tät rumslig upplösning. Program som ger underlag för statusklassningar av specifika objekt kan oftast inte standardiseras mellan delprogrammen, då ansatserna väsentligt skiljer sig åt.

Screening/lasagna plot (sektion 2.3.3) och dess varianter lämpar sig för att beskriva data i de flesta delprogrammen i programområdet Sötvatten. De kan anpassas till data med månatliga, säsongsvisa eller årliga data. Om det finns få trender i data kan en indelning i tre klasser (signifikant upp, signifikant ned och ingen signifikant trend) vara för grov. I så fall kan en variant som illustrera trender med hjälp av p-värden användas (Figur C1 i appendix).

### 5.2 Förslag till kvalitetsutvärdering av delprogram

En kvalitetsdeklaration för ett övervakningsprogram bör göras regelbundet på ett liknande sätt som redan nu är brukligt för officiell statistik. Då övervakningsprogrammen är mer varierande och har större omfattning än en vanlig produkt i den officiella statistiken, behöver strukturen hos en kvalitetsdeklaration vara lite mer flexibel. Ett förslag till upplägg av kvalitetsdeklaration finns i avsnitt D i appendix.

En kvalitetsdeklaration syftar främst till att fungera som diskussionsunderlag för förväntningar och krav på programmet. Programansvarig myndighet förväntas göra tydligt gällande vilka (lag-)krav enskilda program ska uppfylla och vilka syften programmen har. Utförare har detaljkunskap kring

övervakade objekt och områden, kvalitetssäkring och publicering av data, och kunskap om aktuella metod- och miljöfrågor. En nedskrivna kvalitetsdeklaration kan underlätta ett utbyte av information och gör samma information tillgänglig för utförare av närliggande övervakningsprogram.

I samband med utvärderingen av programmen bör man även se över syftena med delprogrammen för att se om de behöver uppdateras. Till exempel behöver aspekter som behoven av att framtidssäkra programmen diskuteras, inte minst i skenet av hur förändrat och extremare väder kan påverka både insamlade data och vad som bör observeras.

## 6. Referenser

- Asmala, E., Carstensen, J., Råike, A., 2019. Multiple anthropogenic drivers behind upward trends in organic carbon concentrations in boreal rivers. *Environ. Res. Lett.* 14, 124018.  
<https://doi.org/10.1088/1748-9326/ab4fa9>
- Biemer, P.P., Lyberg, L.E., 2003. The Evolution of Survey Process Quality, in: *Introduction to Survey Quality*. John Wiley & Sons, Ltd, pp. 1–25. <https://doi.org/10.1002/0471458740.ch1>
- Bignert, A., Danielsson, S., Faxneld, S., Nyberg, E., 2016. Comments Concerning the National Swedish Contaminant Monitoring Programme in Marine Biota, 2016 (No. 5:2016). Swedish Museum of Natural History.
- Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically Weighted Regression. *J. R. Stat. Soc. Ser. Stat.* 47, 431–443. <https://doi.org/10.1111/1467-9884.00145>
- Carstensen, J., 2007. Statistical principles for ecological status classification of Water Framework Directive monitoring data. *Mar. Pollut. Bull., Implementation of the Water Framework Directive in European marine waters* 55, 3–15.  
<https://doi.org/10.1016/j.marpolbul.2006.08.016>
- Carstensen, J., Lindegarth, M., 2016. Confidence in ecological indicators: A framework for quantifying uncertainty components from monitoring data. *Ecol. Indic.* 67, 306–317.  
<https://doi.org/10.1016/j.ecolind.2016.03.002>
- Carstensen, J., Weydmann, A., 2012. Tipping points in the arctic: eyeballing or statistical significance? *Ambio* 41, 34–43. <https://doi.org/10.1007/s13280-011-0223-8>
- Clarke, R.T., Hering, D., 2006. Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. *Hydrobiologia* 566, 433–439. <https://doi.org/10.1007/s10750-006-0079-2>
- Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., Harris, P., 2023. A Route Map for Successful Applications of Geographically Weighted Regression. *Geogr. Anal.* 55, 155–178. <https://doi.org/10.1111/gean.12316>
- European Commission, 2003. Common Implementation Strategy for the Water Framework Directive, Guidance document n. 7 (No. 2000/60/EC).
- European statistical systems, 2019. Quality Assurance Framework of the European Statistical System.
- Fölster, J., Markensten, H., Sandström, S., Widén Nilsson, E., 2021. Förslag till bedömningsgrunder för näringsämnen i sjöar och vattendrag. Rapp. Sver. Lantbruksuniversitet Institutionen För Vatten Och Miljö.
- Fölster, J., von Brömssen, C., 2012. Osäkerhet i statusklassning - Näringsämnen i sötvatten i skogslandskapet (Institutionsrapport). Institutionen för vatten och miljö.
- Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale Geographically Weighted Regression (MGWR). *Ann. Am. Assoc. Geogr.* 107, 1247–1265.  
<https://doi.org/10.1080/24694452.2017.1352480>
- Fryer, R.J., Nicholson, M.D., 1999. Using smoothers for comprehensive assessments of contaminant time series in marine biota. *Ices J. Mar. Sci.* 56, 779–790.  
<https://doi.org/10.1006/jmsc.1999.0499>
- Grandin, U., 2006. Statistisk analys av möjligheter att kunna upptäcka regionala trender i de nuvarande programmen för nationell övervakning av sötvatten. (No. 2006:18). Department of Environmental Assessment, Swedish University of Agricultural Sciences.
- Hastie, T., Tibshirani, R., 1986. Generalized Additive Models. *Stat. Sci.* 1, 297–310.  
<https://doi.org/10.1214/ss/1177013604>
- Havs- och vattenmyndigheten, 2022a. Kvalitetsdeklaration Kust och hav - Miljötillstånd: Tillförsel av kväve och fosfor till kusten.
- Havs- och vattenmyndigheten, 2022b. KVALITETSDEKLARATION Miljötillståndet i Sötvatten: Fisk i sjöar.

- Havs- och vattenmyndigheten, 2020. Metod för uppföljning av miljötillståndet i vattenförekomster.
- Havs- och vattenmyndigheten, 2018. Fisk i sjöar - vägledning för statusklassificering (No. 2018:36), Havs- och vattenmyndighetens rapport. Havs- och vattenmyndigheten.
- HELCOM, 2019. HELCOM Guidelines for the annual and periodical compilation and reporting of waterborne pollution inputs to the Baltic Sea (PLC-Water).
- HELCOM, 2015. Strategy for future HELCOM assessments of inputs of nutrients and selected hazardous substances.
- HELCOM, 2013. Approaches and methods for eutrophication target setting in the Baltic Sea region (No. No 133), Baltic Sea Environment Proceedings.
- Helsel, D.R., Frans, L.M., 2006. Regional Kendall Test for Trend. *Environ. Sci. Technol.* 40, 4066–4073. <https://doi.org/10.1021/es051650b>
- Hirsch, R.M., Slack, J.R., 1984. A Nonparametric Trend Test for Seasonal Data With Serial Dependence. *Water Resour. Res.* 20, 727–732. <https://doi.org/10.1029/WR020i006p00727>
- Hirsch, R.M., Slack, J.R., Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* 18, 107–121. <https://doi.org/10.1029/WR018i001p00107>
- Holmgren, K., 2021. Provfiske i trendsjöar : historia, nuläge och framtida behov. *Aqua Rep.*
- ICES, 2022. Third Workshop on Integrated Trend Analysis to Support Integrated Ecosystem Assessment (WKINTRA3) (report). ICES Scientific Reports. <https://doi.org/10.17895/ices.pub.19398317.v1>
- Kendall, M., Gibbons, J.D., 1990. Rank Correlation Methods, 5 edition. ed. Oxford University Press, London : New York, NY.
- Letcher, R.A., Schreider, S.Yu., Jakeman, A.J., Neal, B.P., Nathan, R.J., 2001. Methods for the analysis of trends in streamflow response due to changes in catchment condition. *Environmetrics* 12, 613–630. <https://doi.org/10.1002/env.486>
- Libiseller, C., 2003. Considering Meteorological Variation in Assessments of Environmental Quality Trends, Linköping Studies in Statistics. Univ. Linköping.
- Libiseller, C., Grimvall, A., 2002. Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. *Environmetrics* 13, 71–84. <https://doi.org/10.1002/env.507>
- Loftis, J.C., Taylor, C.H., Chapman, P.L., 1991. Multivariate tests for trend in water quality. *Water Resour. Res.* 27, 1419–1429. <https://doi.org/10.1029/90WR01802>
- Mann, H.B., 1945. Nonparametric Tests Against Trend. *Econometrica* 13, 245–259. <https://doi.org/10.2307/1907187>
- Nyberg, E., 2016. Improved Assessment in Environmental Monitoring of POPs : Using monitoring data from the aquatic ecosystem and human milk.
- Nyberg, E., Bigner, A., Danielsson, S., Ek, C., Faxneld, S., Winkens Pütz, K., Sundbom, M., 2020. Adjustments for confounders in individual and pooled samples from the Swedish national monitoring of contaminants in freshwater fish. Swedish Museum of Natural History.
- Pedersen, E.J., Miller, D.L., Simpson, G.L., Ross, N., 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ* 7, e6876. <https://doi.org/10.7717/peerj.6876>
- SCB, 2023. Utvärdering av den officiella statistikens kvalitet, handbok version 3.2.
- Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* 63, 1379–1389. <https://doi.org/10.2307/2285891>
- Simpson, G., 2014. Confidence intervals for derivatives of splines in GAMs. *Bottom Heap*.
- Simpson, G.L., 2018. Modelling palaeoecological time series using generalized additive models. <https://doi.org/10.1101/322248>
- Starcevich, L.A.H., McDonald, T., Chung-MacCoubrey, A., Heard, A., Nesmith, J., Philippi, T., 2018. Trend estimation for complex survey designs of water chemistry indicators from Sierra Nevada Lakes. *Environ. Monit. Assess.* 190, 596. <https://doi.org/10.1007/s10661-018-6963-1>
- Theil, H., 1959. A rank-invariant method of linear and polynomial regression analysis. I, II, III. *Nederl Akad Wetensch* 53, 386–392.
- Urquhart, N.S., Kincaid, T.M., 1999. Designs for Detecting Trend from Repeated Surveys of Ecological Resources. *J. Agric. Biol. Environ. Stat.* 4, 404–414. <https://doi.org/10.2307/1400498>

- von Brömssen, C., Betnér, S., Fölster, J., Eklöf, K., 2021. A toolbox for visualizing trends in large-scale environmental data. *Environ. Model. Softw.* 136, 104949. <https://doi.org/10.1016/j.envsoft.2020.104949>
- von Brömssen, C., Fölster, J., Eklöf, K., 2023a. Temporal trend evaluation in monitoring programs with high spatial resolution and low temporal resolution using geographically weighted regression models. *Environ. Monit. Assess.* 195, 547. <https://doi.org/10.1007/s10661-023-11172-2>
- von Brömssen, C., Fölster, J., Futter, M., McEwan, K., 2018. Statistical models for evaluating suspected artefacts in long-term environmental monitoring data. *Environ. Monit. Assess.* 190, 558. <https://doi.org/10.1007/s10661-018-6900-3>
- von Brömssen, C., Fölster, J., Kyllmar, K., Bierzoza, M., 2023b. Modeling Complex Concentration-Discharge Relationships with Generalized Additive Models. *Environ. Model. Assess.* <https://doi.org/10.1007/s10666-023-09915-z>
- Wood, S.N., 2017. *Generalized additive models: an introduction with R*, Second edition. ed, Chapman & Hall/CRC texts in statistical science. CRC Press/Taylor & Francis Group, Boca Raton.

## Appendix

### A. Exempel på styrkeberäkningar

#### A.1 Med hjälp av en typserie

Styrkeberäkningar kan användas för att beräkna sannolikheterna att upptäcka en trend för en specifikt beskriven situation. Här utgår vi ifrån att trenderna ska utvärderas med linjär regression som statistisk metod. Eftersom serierna som ska analyseras oftast har både säsongvariation och temporal autokorrelation använder vi en typserie för att extrahera realistiska mönster för dessa egenskaper. I exemplet nedan används tidsserien för totalkväve i Alterälven. Simuleringarna avser 10 år av månadsdata och avser sannolikheten att upptäcka trender av olika storhet under denna period. Proceduren är att:

- (i) Extrahera säsongvariation, skattningar på relevanta temporala autokorrelationer och residualvariansen ur typserien. Eventuell trend i typserien skattas samtidigt eller räknas bort i förväg för att inte påverka skattningar av andra egenskaper.
- (ii) Simulera en serie av slumpstal med önskad längd och med samma autokorrelationsstruktur som typserien. Addera säsongvariationen till serien.
- (iii) Lägg till trender av olika storlek till serien (i exemplet: ingen förändring, 5, 10, 20 och 40 enheter/år).
- (iv) Generera 1000 serier med var och en av de olika trenderna och med nya slumpstal för varje serie.
- (v) Analysera trenderna i serierna med linjär regression. Beräkna hur ofta trender upptäcks för olika storlekar av trender.

I exemplet jämför vi styrkan när underlagsdata är:

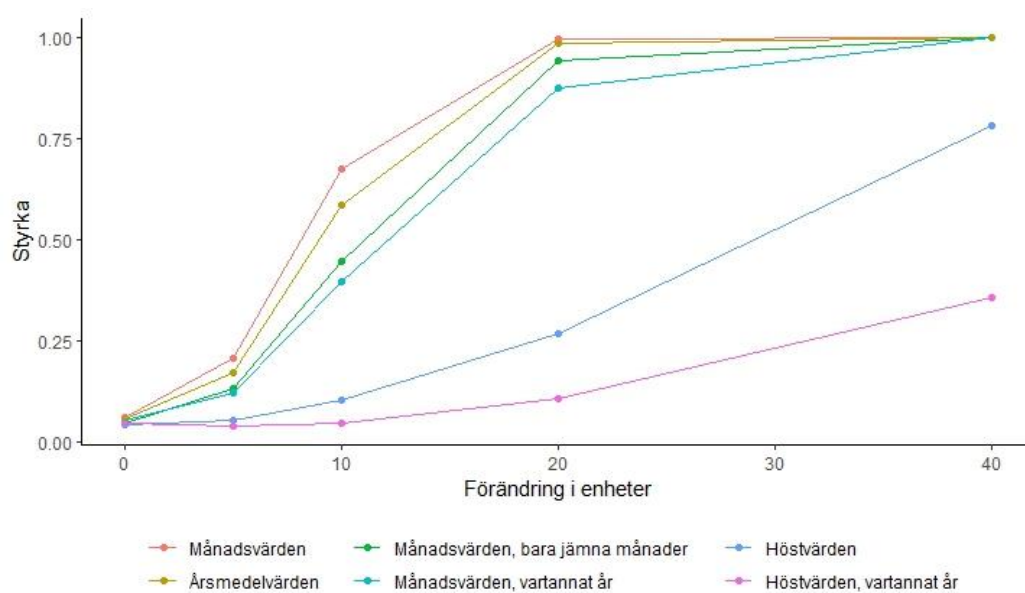
- a. 10 år av månadsdata
- b. 10 år av data för varannan månad, jämna månader
- c. 10 år av årsmedelvärden (baserat på månadsdata)
- d. 10 år av månadsvärden, men enbart observationer vartannat år
- e. 10 år av en mätning per år (höstvärde, oktober)
- f. 10 år av en mätning vartannat år

Observera att även om vi anger att vi analyserar enbart jämna månader eller enbart höstvärden, så är inte resultaten direkt tillämpliga i praktiken, då säsongvariation skattas som en jämn funktion och den artificiella trenden läggs på i efterhand. I praktiken kan det vara så att olika trender observeras under olika säsonger eller att jämna månader tillhandahåller mer information än udda, kanske för att vårfloden ligger under en sådan. De visade simuleringarna skulle visa samma resultat om udda månader eller någon annan enstaka månad valdes. Beräkningarna baseras på Alterälvens egenskaper. För andra vattendrag kan resultaten vara annorlunda.

I figur A1 ses styrkan för dessa olika typer av underlagsdata. När styrkan är 1 upptäcks alla trender (i 1000 simuleringar). Det är tydligt att styrkan är högst när analysen baseras på månatliga data, antingen genom direkt analys av dessa eller när data omräknas till årsmedelvärden. Även om observationer bara tas varannan månad eller om bara vartannat år provas med månatliga observationer är styrkan fortfarande förhållandevis hög, speciellt om förändringarna per år är stora. Baseras analyserna på enskilda observationer per år eller ännu mer sällan sjunker styrkan avsevärt.



Alterälven har en ungefärlig nivå på 489 µg/l N. En förändring på 5 enheter per år, eller 50 enheter per 10 år, motsvarar en förändring på 10% under hela perioden. En förändring på 40 enheter per år eller 400 enheter per 10 år motsvarar en förändring på 82% under hela perioden.



Figur A1: Styrkan (sannolikheten) att detektera en trend av olika storlekar (x-axeln; 0, 5, 10, 20 och 40 enheter) för serier med olika tidsupplösning.

## A.2 Med hjälp av generella värden – absolut förändring

För att få en mer generell överblick över styrka i olika serier inom ett övervakningsprogram kan typiska egenskaper, såsom variation och medelvärden, extraheras ur alla serier. I följande simulering används variation i serier av totalkväve i övervakningsprogrammen för trendvattendrag och flodmyningar. För att undvika säsongvariation och autokorrelationer baseras simuleringarna på årsmedelvärden för varje station. Proceduren är:

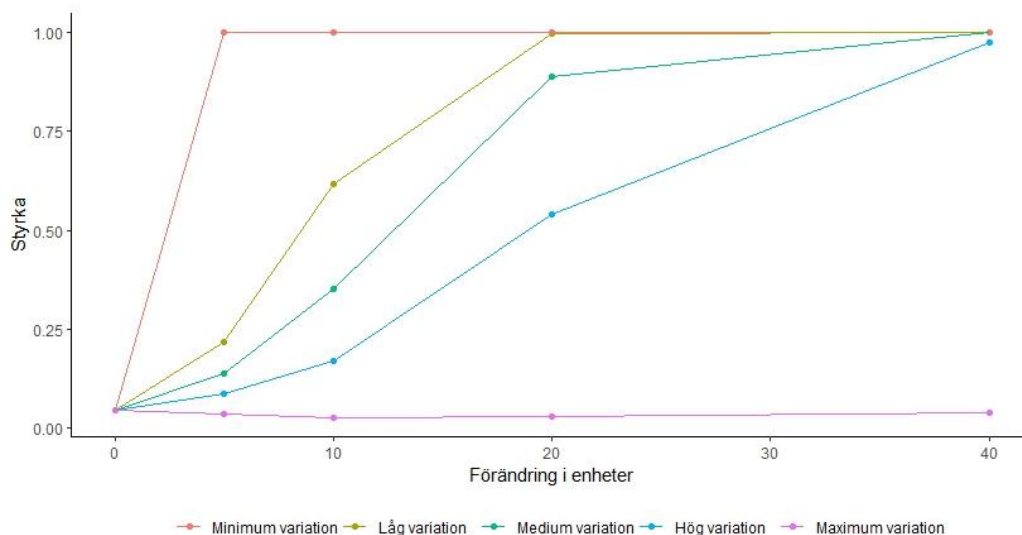
- (i) Skatta trenden i varje serie och extrahera residualvariansen, det vill säga skatta variationen i serien om ingen trend skulle finnas.
- (ii) Kvantifiera hur stor variation som finns i övervakningsprogrammet, till exempel genom att använda kvantiler för de residualvarianser eller residualstandardavvikelser som producerades i (i).
- (iii) Generera 1000 serier som har olika storlekar av variation och representera 10 år av årliga medelvärden.
- (iv) Lägg till trender av olika storlek till serien (i exemplet: ingen förändring, 5, 10, 20 och 40 enheter/år).
- (v) Analysera trenderna i serierna med linjär regression. Beräkna hur ofta trender upptäckts för olika storlekar av trender och olika hög variation.

Tabell A1 ger kvantiler för medelvärden av totalkväve i trend- och flodmynningsprogrammet och kvantiler för residualstandardavvikelse för samma serier efter att eventuella trender har tagits bort. Minsta variationen observeras i Vapstälven och störst variation i Råån vid station Helsingborg. 25% av stationerna har en variation som är mindre än 34,6 och 75% har en variation mindre än 78,6 enheter.

Stationsmedelvärdet påverkar inte styrkan hos trendanalyserna i dessa simuleringar. Variation i olika storlekar leder till stora skillnader i hur ofta trender kan upptäckas, alltså hur hög styrkan är (Figur A2). För Vapstälven, som har minst variation, kommer även små förändringar att kunna upptäckas, medan variationen i Råån är så pass stor att även förändringar med 40 enheter per år inte kommer att hittas. För mer än 50% av serierna kommer trender av en storleksordning av 20 enheter eller mer per år hittas med bra styrka (över 0,8).

Tabell A1. Minimum, maximum, 25%-, 50%- och 75%-kvantilen för medelvärden och residualstandardavvikelse

	Stationsmedelvärde	Residualstandardavvikelse för stationerna
Minimum	87,2	5,62
25%, låg	308,4	34,6
50%, medium	461,5	50,5
75%, hög	864	78,6
Maximum	6904,8	1015,5



Figur A2. Styrkan (sannolikheten) att detektera en trend av olika storlekar (x-axeln; 0, 5, 10, 20 och 40 enheter) för serier med olika variation.

### A.3 Med hjälp av generella värden – relativ förändring

För många variabler varierar nivån kraftigt mellan olika objekt och att studera absoluta förändringar kan vara ointressant. Med en liknande ansats som i A.2. kan relativa förändringar studeras.

Variationen kan då kvantifieras som variationskoefficient ( $cv = \frac{\text{standardavvikelse}}{\text{medelvärde}}$ ).

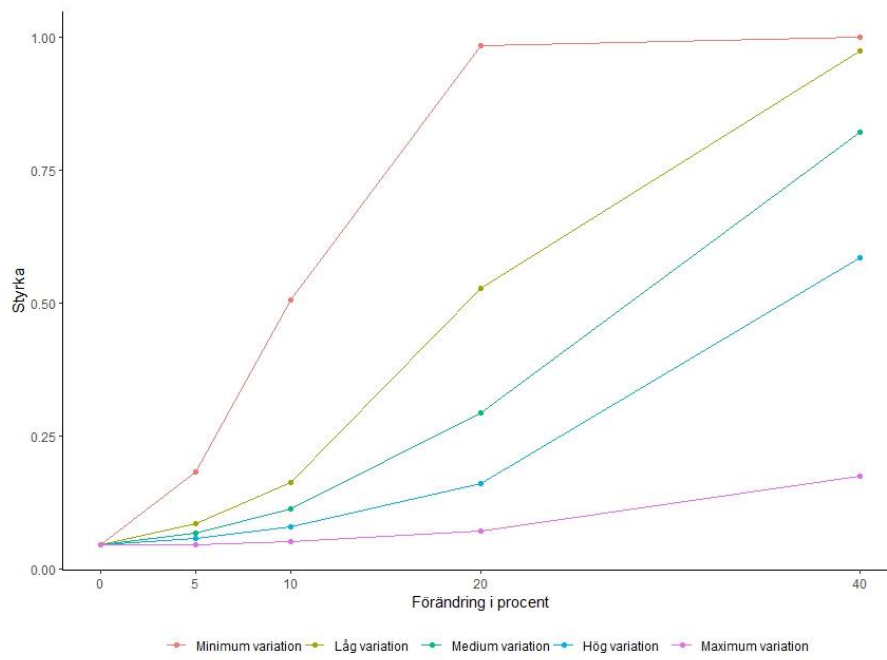
- (i) Utifrån data för den relevanta miljövariabeln i övervakningsprogrammet, beräkna årsmedelvärden för varje station.
- (ii) Skatta trenden i varje serie av årsmedelvärden och extrahera residualvariansen, det vill säga skatta variationen i serien om ingen trend skulle finnas. Beräkna residualstandardavvikelsen och dividera den med seriens medelvärde för att bestämma variationskoefficienten.
- (iii) Kvantifiera hur stor variation som finns i övervakningsprogrammet, till exempel genom att använda kvantiler för de variationskoefficienter som producerades i (ii).
- (iv) Generera 1000 serier som har olika relativa storlekar av variation och representera 10 år av årsmedelvärden.
- (v) Lägg till trender av olika storlek till serien (i exemplet: ingen förändring, 5%, 10%, 20% eller 40% per 10-årsperiod).
- (vi) Analysera trenderna i serierna med linjär regression. Beräkna hur ofta trender upptäcks för olika storlekar av trender och olika hög variationskoefficient.

Tabell A2 ger kvantiler för variationskoefficienterna i programmen. Minsta variationskoefficient observeras i Loån, med en standardavvikelse som är 4% av seriens medelvärde, och störst variationskoefficient observeras i Ljusnan, Ljusne Strömmar, där standardavvikelsen är 30% av medelvärdet. 25% av stationerna har en variationskoefficient som är mindre än 0,08 och 75% har en variationskoefficient mindre än 0,15.

Variationskoefficienter i olika storlekar leder till stora skillnader i hur ofta trender kan upptäckas, alltså hur hög styrkan är (Figur A3). För Loån, kan förändringar av 20% under en 10-årsperiod upptäckas med hög styrka, medan det för de flesta stationer behövs trender av uppemot 40% under en 10-årsperiod.

Tabell A2. Minimum, maximum, 25%-, 50%- och 75%-kvantilen för variationskoefficienten.

	Variationskoefficient
Minimum	0,04
25%, låg	0,08
50%, medium	0,11
75%, hög	0,15
Maximum	0,3



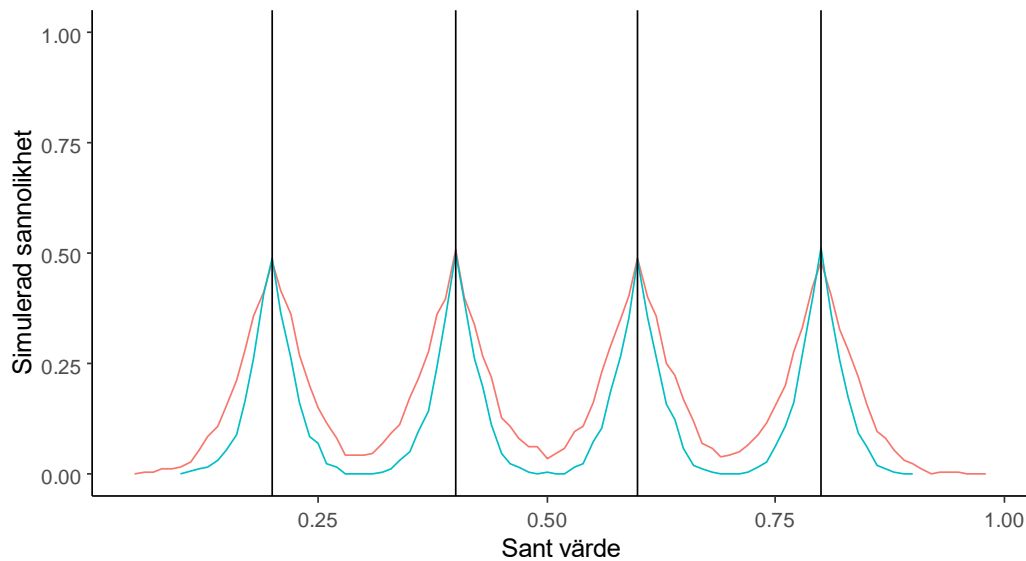
Figur A3: Styrka för att upptäcka trender med relativ variation skattat ur övervakningsprogrammen flodmynningar och trendvattendrag.

## B. Simulerade sannolikheter för felklassningar av status

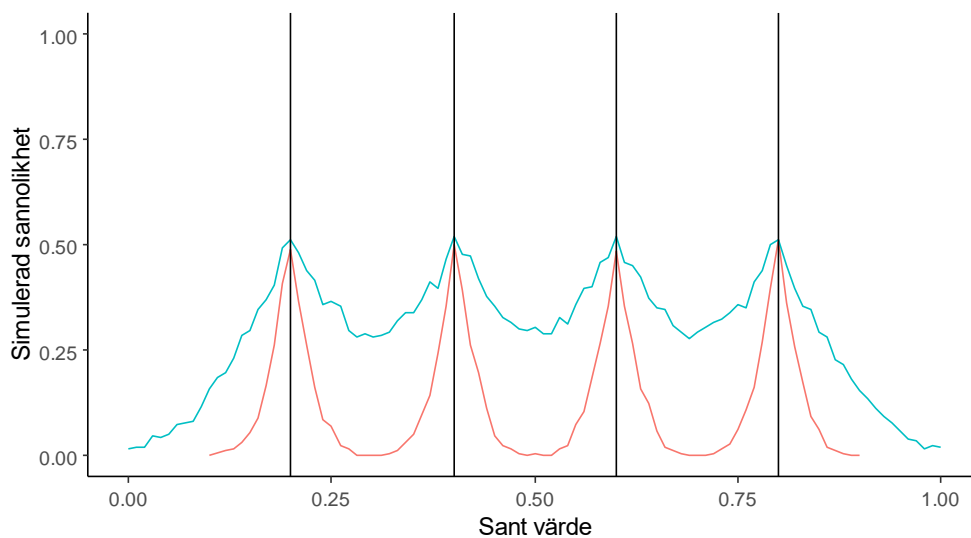
Sannolikheter för felklassificering av status simuleras genom att anta att:

- Statusklassning görs genom "face-value"
- Klassgränserna är fasta och ligger vid värdena 0,2, 0,4, 0,6 och 0,8
- Stickprovsstandardavvikelsen antas vara 0,1 eller 0,3 för olika simuleringar
- Stickprovsstorleken antas vara 4 eller 10 för olika simuleringar.

Den teoretiska sannolikheten för felklassificeringar ökar när stickprovsstorleken är mindre (Figur B1) och när variansen är större (Figur B2).

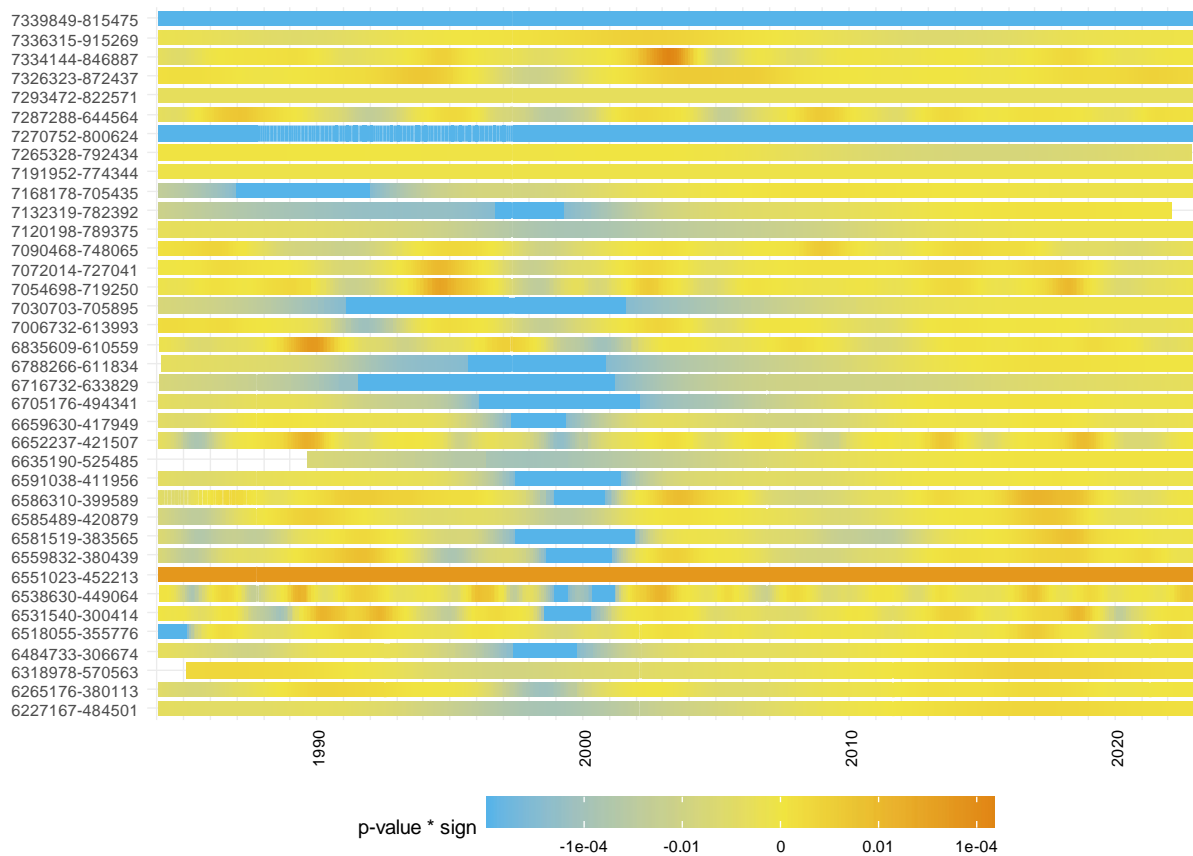


Figur B1: Simulerade sannolikheter att missklassificera status när "face-value" används. Baserat på medelvärde av 4 observationer (röd) och medelvärde av 10 observationer (blå) och en standardavvikelse på 0,1.



Figur B2: Simulerade sannolikheter att missklassificera status när "face-value" används. Baserat på medelvärde av 10 observationer och en standardavvikelse som är 0,1 (röd) och 0,3 (blå).

### C. Varianter av screening/lasanga plot



Figur C1: Spatial visualisering av punktvisa p-värden för förändringar i sulfat. Stationerna är sorterade från nord (längst upp) till syd.

## D. Förslag till kvalitetsdeklaration

### Kvalitetsdeklaration för övervakningsprogram

1. Programmets syfte och relevans
  - 1.1. Ändamål och informationsbehov
2. Programmets tillförlitlighet
  - 2.1. Tillförlitlighet och riktighet
  - 2.2. Eventuella förbättringsbehov
3. Datatillgångens aktualitet och punktlighet
  - 3.1. Framställningstid för kvalitetssäkrad data
  - 3.2. Frekvens av publicerad data
  - 3.3. Eventuella förbättringsbehov
4. Datatillgänglighet och tydlighet
  - 4.1. Tillgång till data
  - 4.2. Tillgång till, och tydlighet hos beskrivning av mätvariabler, mätmetoder, analysmetoder,...
  - 4.3. Tillgång till och tydlighet hos metadata
  - 4.4. Eventuella förbättringsbehov
5. Jämförbarhet och sammanvändbarhet
  - 5.1. Jämförbarhet över tid
  - 5.2. Jämförbarhet över grupper
  - 5.3. Sammanvändbarhet över olika delprogram inom samma programområde
  - 5.4. Eventuella förbättringsbehov
6. Utökad kvalitetsdeklaration
  - 6.1. Tänkbara nödvändiga förändringar i programmet
    - 6.1.1. Kopplat till förändringar i metodik
    - 6.1.2. Kopplat till miljöfrågor som nyligen har fått ökat intresse
    - 6.1.3. Kopplat till förändringar i förutsättningar
    - 6.1.4. Kopplat till programmets mål och syfte
  - 6.2. Sammanvändning
    - 6.2.1. Nuvarande sammanvändning med data från andra övervakningsprogram inom och utanför programområdet
    - 6.2.2. Eventuella förbättringsbehov
  - 6.3. Belysning av orsak och verkan
    - 6.3.1. Nuvarande möjligheter att använda data för att undersöka samband mellan påverkande faktorer och miljötillstånd
    - 6.3.2. Eventuella förbättringsbehov
  - 6.4. Användning av data i forskning
7. Statistiska sammanställningar (om relevant)
  - 7.1. Simulerade styrke- eller sannolikhetsberäkningar

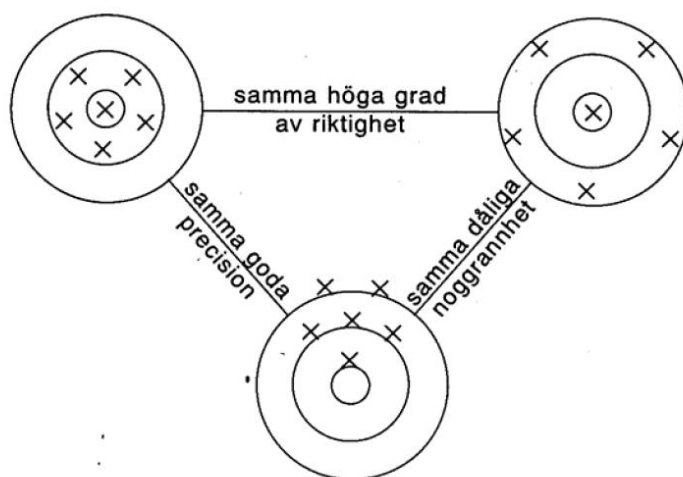
## E. Centrala Begrepp

### E.1 Noggrannhet – riktighet – precision

Noggrannheten (eng.: accuracy) hos ett stickprov beror på både precision (eng.: precision) och riktighet (eng.: unbiasedness, trueness).

Precisionen kvantifieras genom variation kring stickprovets medelvärde, till exempel genom stickprovsvariansen eller stickprovsstandardavvikelsen. Låg precision innebär att de enskilda observationerna uppvisar stor spridning, medan hög precision betyder att alla mätningar leder till liknande värden.

Riktighet avgör hur nära de observerade värdena ligger det sanna värdet. Vanligtvis känner vi inte till det sanna värdet och riktigheten går då inte att kvantifiera. Däremot kan det diskuteras om det finns en risk att ett stickprov eller urval av stationer är biased, det vill säga avviker systematiskt från det sanna värdet. Begreppet är nära knutet till representativitet av data.



Figur E1: Noggrannhet – riktighet – precision. Från rapporten "Vad är rätt och vad är fel?", Mikael Lilje, Lantmäteriet.

### E.2 Variation, varians, standardavvikelse

Precision bedöms genom en kvantifiering av variationen i ett dataset, till exempel genom beräkning av stickprovsvarians, med hjälp av observationernas genomsnittliga kvadratiska avstånd till medelvärdet.

$$s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

eller genom stickprovsstandardavvikelse (kvadratroten ur stickprovsvariansen).

För statistiska modeller används en liknande skattning som kvantifierar den variation som inte förklaras av modellen. Denna variation kallas ofta residualvariation.

### E.3 Riktighet, bias, representativitet

Riktighet hos skattningen kan uppnås genom att välja stationer, provplatser och tidpunkter som representerar det som ska övervakas, målpopulationen. Om urvalet av platser och tider inte är



representativt introduceras bias, som innebär att skattningen systematiskt avviker från det okända sanna värdet som ska observeras.

#### E.4 Medelfel, "standard error"

När skattningar, till exempel trender eller årsmedelvärden, beräknas anges förutom denna skattning ofta också ett mått på variation. En sådan kan vara stickprovsvariansen, stickprovsstandardavvikelsen eller residualvariansen (E.2) eller, vanligare, medelfelet (eng.: standard error) som tillhör den aktuella skattningen. Beräkningen av medelfelet hos en skattning baseras på stickprovsvariansen eller residualvariansen, samt stickprovsstorleken.

För ett medelvärde beräknas medelfelet ("standard error of the mean") som

$$SEM = \sqrt{\frac{s^2}{n}},$$

alltså kvadratroten ur stickprovsvariansen dividerat genom antal observationer som medelvärdet beräknas på.

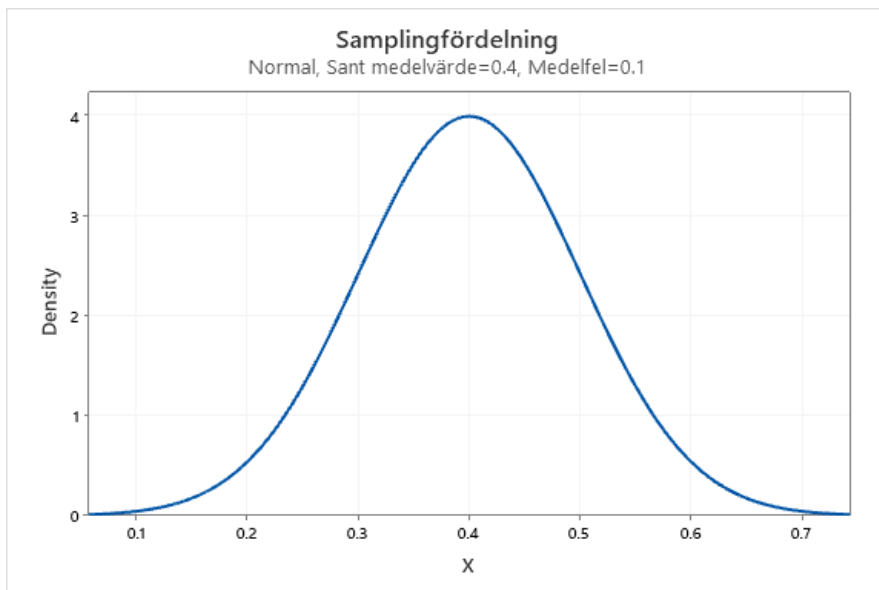
För en trendska skattning skattas medelfelet på ett liknande sätt men med en annan formel ur residualvariansen, som är variationen kring den anpassade trendlinjen eller -kurvan.

Medelfelen blir alltid mindre ju fler observationer som finns tillgängliga, då stickprovsstorleken  $n$  ingår i formeln. Det innebär att skattningarna blir mer stabila, eller mindre variabla, ju fler observationer de baseras på. För tidsseriesdata leder dock högre provtagningsfrekvens till högre temporal autokorrelation i data vilket måste beaktas (se avsnitt I).

#### E.5 Samplingfördelning

I den frekventistiska statistiken (det som vanligtvis lärs ut i statistikkurser) är utgångspunkten att det finns ett sant parametervärde som är konstant och okänt. Statistiska metoder kan då användas för att uttala sig om vad som är sannolikt för detta okända parametervärde. I exemplet nedan används medelvärdet, men liknande resonemang görs för andra skattningar, till exempel regressionskoefficienter.

För att beräkna statistiska hypotestester behövs, utöver det skattade medelvärdet och medelvärdets medelfel, även antagandet om en teoretisk sannolikhetsfördelning, en så kallad samplingfördelning. Samplingfördelningen representerar en teoretisk fördelning för vad som kan förväntas i en provtagningsituation och måste inkludera ett antagande om det sanna – okända – medelvärdet i populationen (som anges i nollhypotesen). Anta, till exempel, att det i nollhypotesen antas att det sanna värdet är 0,4. 10 observationer samlas in och det beräknade medelfelet blir 0,1. Då är samplingfördelningen den som visas i Figur E2. Samplingfördelningen visar fördelningen för medelvärdet beräknade ur 10 observationer ifall vi skulle göra många identiska provtagningar. Vi kan förvänta oss observerade medelvärdet så små som 0.1 eller så stora som 0.7. Denna fördelning visar alltså hur stor variation i *medelvärdet* vi kan få enbart baserat på slumpen, när det sanna värdet är 0.4. Samplingfördelningens bredd beror på stickprovsstorleken – ju fler observationer ett medelvärde baseras på desto smalare blir samplingfördelningen.



Figur E2: Samplingfördelning centrerat kring det sanna medelvärdet av en miljövariabel, här 0,4. Fördelningen illustrerar vilka värden rimligen kan förväntas för det observerade medelvärdet av 10 observationer om data samlas in enligt vanliga principer för att uppnå representativa och oberoende observationer.

#### E.6 Konfidensintervall, felmarginal

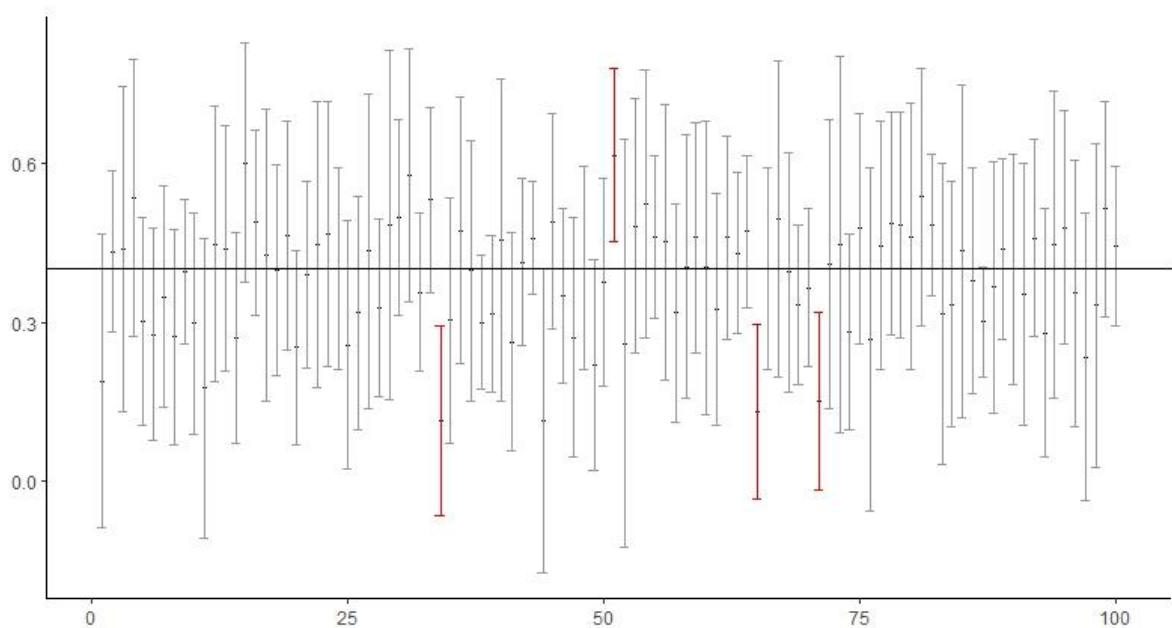
Statistisk konfidens beräknas inte, utan bestäms av statistikanvändaren när ett konfidensintervall ska beräknas för, till exempel, ett medelvärde eller en trendskattning. Den anges som konfidensgrad och fastställs utan hänsyn till data, oftast till 95%, och benämns med  $1 - \alpha$ . Ett konfidensintervall med konfidensgrad 95% är ett intervall inom vilket det sanna medelvärdet eller den sanna trendskattningen förväntas ligga i 95% av alla oberoende datainsamlingar gjorda med samma villkor och med samma stickprovsstorlek.

Ett konfidensintervall beräknas med hjälp av skattningen och skattningens medelfel, samt antagandet om en teoretisk fördelning (samplingfördelning, se avsnitt E.5). För ett medelvärde beräknas konfidensintervallet som

$$\bar{x} \pm t_{1-\frac{\alpha}{2}, n-1} \cdot \sqrt{\frac{s^2}{n}},$$

där  $\bar{x}$  är stickprovsmedelvärdet,  $\sqrt{\frac{s^2}{n}}$  är medelvärdets medelfel och  $t_{1-\frac{\alpha}{2}}$  är tagen från en teoretisk fördelning som ofta är lämplig för medelvärden (t-fördelningen ersätter normalfördelningen när variansen i data måste skattas,  $n-1$  betecknar t-fördelningens frihetsgrader). I denna term ingår även konfidensgraden  $1 - \alpha$ . Halva konfidensintervallet ( $t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2}{n}}$ ) benämns som felmarginal.

Ett konfidensintervall är alltså ett intervall som ligger runt det observerade medelvärdet i stickprovet. Om 100 stickprov tas och 95-procentiga konfidensintervall beräknas för varje stickprov, så förväntas 95 av konfidensintervallen täcka det sanna okända populationsmedelvärdet (Figur E3).



Figur E3: Illustration av konfidensintervallet genom en simulering. Anta att det sanna okända populationsmedelvärdet är 0,4 och 100 oberoende stickprov dras. 95 av de beräknade konfidensintervallen förväntas täcka det sanna värdet. I vår simulering blev det dock 96 – de fyra som inte täcker det sanna värdet syns i rött.

### E.7 Hypotestest

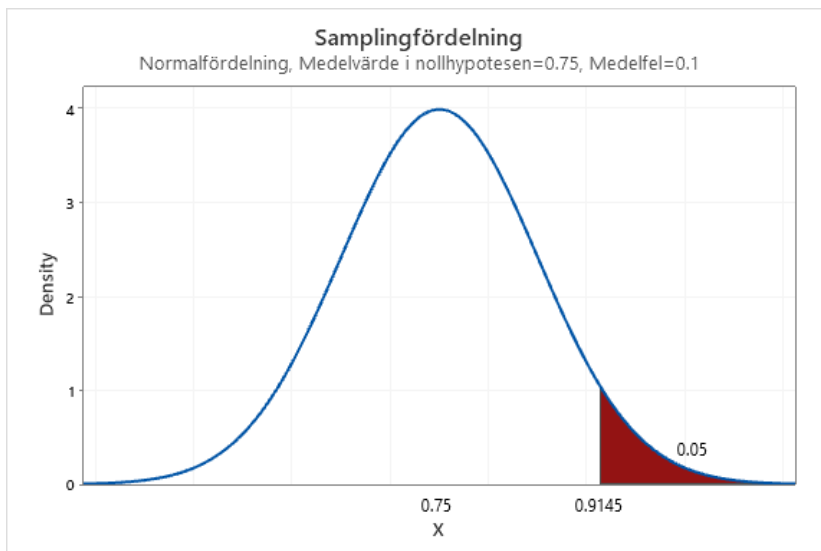
Ett hypotestest används för att avgöra om det sanna okända värdet av en parameter avviker signifikant från ett givet värde. För detta formuleras två hypoteser, en nollhypotes och en alternativhypotes. Till exempel kan det vara intressant att testa om det sanna medelvärdet i en population är större än 0,75. Då blir nollhypotesen ( $H_0$ ) och alternativhypotesen ( $H_1$ ) följande:

$$H_0: \mu = 0,75$$

$$H_1: \mu > 0,75.$$

Statistikanvändaren väljer en signifikansnivå ( $\alpha$ ), ofta 5%, och beräknar en teststatistika som jämförs med vad som kan förväntas om nollhypotesen gäller. Statistisk signifikans innebär att nollhypotesen förkastas, alltså att det sanna medelvärdet är signifikant större än 0.75 i vårt exempel (Figur E4).

Observera att statistiska test kan genomföras enkelsidigt eller dubbelsidigt. Om testet (och alternativhypotesen) är dubbelsidigt förkastar vi nollhypotesen om det observerade värdet ligger bland de 5% mest extrema utfallen i fördelningen, alltså både de mest extrema höga och låga värdena).



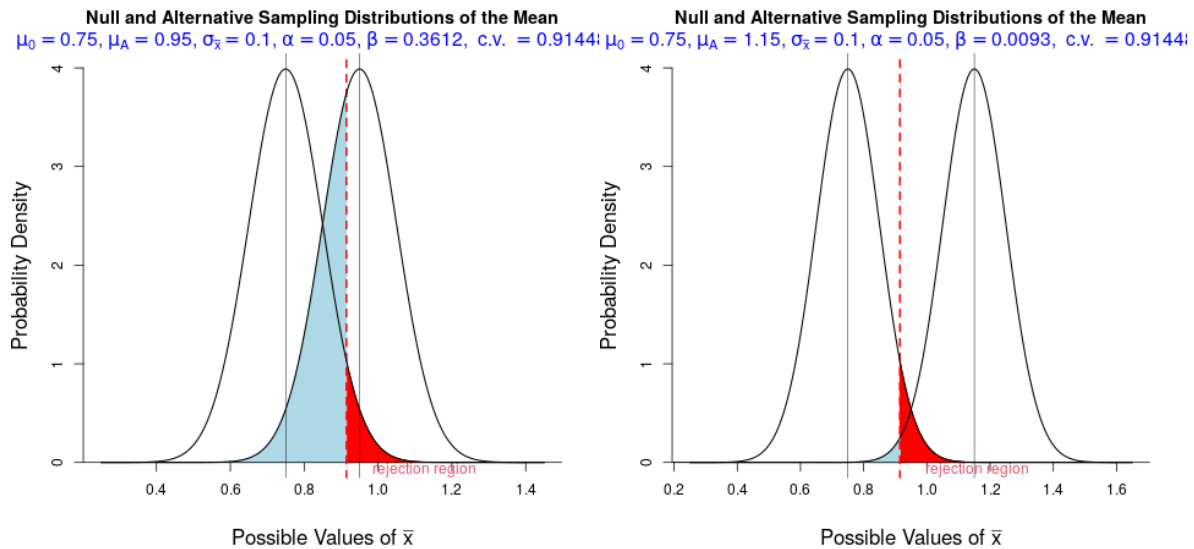
Figur E4: Ett statistiskt hypotestest baseras på en samplingfördelning som är centrerad kring det värde vi antar i nollhypotesen (i vårt exempel 0,75). Nollhypotesen förkastas om det observerade värdet ligger bland de 5% högsta utfallen i denna fördelning (rött område). I vårt fall betyder det att om vi observerar ett medelvärde som är 0,9145 eller högre så kan nollhypotesen förkastas.

#### E.8 Fel av första och andra slaget, statistisk styrka

Vid beräkning av ett hypotestest kan felet för första och andra slaget uppstå. Risker för dessa fel kan teoretiskt bestämmas för generella situationer givet att tillräckligt med information finns tillgängligt. Risker kan inte beräknas för ett specifikt test. Det går alltså inte att för ett genomfört test säga att just i detta fall har något av dessa två fel inträffat.

Typ I fel (eng.: type I error) uppstår när nollhypotesen förkastas, fast den egentligen är korrekt. Risker för att göra denna typ av fel fastställs med hjälp av signifikansnivån  $\alpha$  och bestäms därmed av statistikanvändaren. Detta innebär att statistikanvändaren är villig att ta risken att dra fel slutsats (falsk positiv) i  $\alpha\%$  (ofta 5%) av alla statistiska hypotestest som genomförs.

Typ II fel (eng.: type II error) uppstår när alternativhypotesen är korrekt, men det statistiska testet felaktigt inte förkastar nollhypotesen (falsk negativ). Beräkning av risken för denna typ av fel är mer komplicerat, eftersom den kräver, förutom signifikansnivån  $\alpha$ , även information om stickprovsvariansen, stickprovsstorleken och ett specifikt värde för alternativhypotesen, det vill säga en kvantifiering av hur fel nollhypotesen är. Typ II fel benämns ofta med  $\beta$  och illustreras i Figur E5.



Figur E5: Illustration av typ II-felet. Anta att vi testar nollhypotesen  $H_0: \mu = 0.75$ . I verkligheten avviker populationsmedelvärdet från 0.75. Till vänster: Det sanna medelvärdet är 0.95. Till höger: Det sanna medelvärdet är 1.15. För båda figurerna visar det blåa området sannolikheten för typ II-felet ( $\beta$ ) – nollhypotesen förkastas inte trots att den är fel. Felet blir mindre ju mer det sanna värdet skiljer sig från värdet i nollhypotesen. Figuren producerades med hjälp av websidan <https://shiny.rit.albany.edu/stat/betaprob/> publicerad av Psychology Department vid University at Albany.

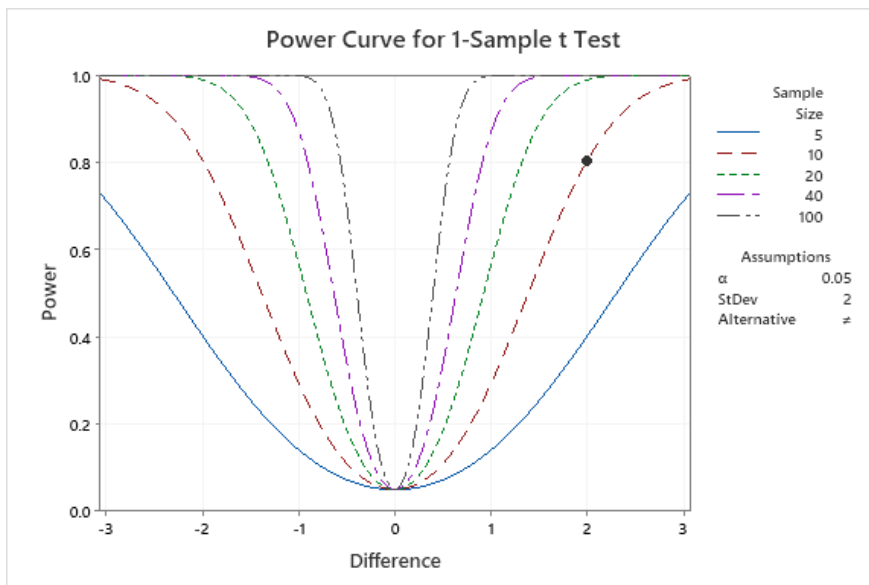
Styrkan hos ett test (eng: power) är nära knuten till felet av andra slaget och beräknas som  $1-\beta$ . Styrkan anger sannolikheten att nollhypotesen förkastas om den faktiskt är felaktig.

För att kunna beräkna den statistiska styrkan för ett test behövs:

- en nollhypotes som ligger till grund för testet (som i sin tur beror på vald statistisk metod, till exempel t-test eller regression)
- en specifik alternativhypotes för vilken styrkan ska beräknas
- en skattning eller gissning av den föreliggande stickprovsvariationen i data

Styrkan kan, som exempel, beräknas för en frågeställning som "Hur stor är sannolikheten att kunna påvisa en skillnad av 1 enhet från värdet 5 när ett t-test används och stickprovsstandardavvikelsen är 2 enheter?".

Styrkan kan också representeras i form av styrkekurvor, där styrkan presenteras för olika storlekar på skillnader eller trender (effektstorleken) och för olika stickprovsstorlekar (Figur E6). Ur styrkekurvan kan avläsas hur många observationer som behövs för att uppnå en viss styrka (ofta 0,8) för att upptäcka en viss skillnad. I exemplet nedan syns olika kurvor för olika stickprovsstorlekar. När 10 observationer görs kan en skillnad på 2 enheter upptäckas med en styrka av 0,8. För att kunna påvisa en skillnad på 0,5 enheter behövs minst 100 observationer för att uppnå samma styrka.



Figur E6: Ett exempel på styrkekurvor. Effektstorleken (difference) ges på x-axeln, styrkan (Power) på y-axeln. Olika stickprovsstorlekar ges i olika färger.

### E.9 Oberoende observationer vs temporala och spatiala autokorrelationer

För statistiska beräkningar är det viktigt att stickprovsvariansen (eller residualvariansen) kan bestämmas korrekt. Grundläggande i statistiken är att de observationer som beräkningen baseras på ska vara oberoende, det vill säga att varje observation bidrar med helt ny information. I miljösammanhang är detta sällan uppfyllt. Data samlas antingen in som tidsserie, där observationerna uppvisar ett beroende i tiden, eller med en rumslig komponent, alltså ett spatialt beroende. Sådana temporala eller spatiala autokorrelationer försvårar korrekta uppskattningar av varianser. Om ingen hänsyn tas till befintliga beroenden underskattas variationen, vilket leder till att konfidensintervall blir för snäva och hypotestesten blir för ofta felaktigt signifikanta.

Finns det tillräckligt med data kan den temporala autokorrelationen inkorporeras i modellerna och variansen kan skattas korrekt. För tidsserier av vattenkvalitet är det också vanligt att anta att observationer som ligger ett år eller mer ifrån varandra uppvisar så liten autokorrelation att den är försumbar och inte behöver tas hänsyn till.

### E.10 Skeva fördelningar

Ofta när statistiska test och konfidensintervall för medelvärden beräknas, baseras beräkningarna på antagandet att medelvärdet är normalfördelat. Detta kan uppnås på två sätt:

- Om den observerade variabeln följer en normalfördelning så är även medelvärdet normalfördelat.
- Om medelvärdet baseras på minst 20 (oberoende) observationer så är medelvärdet normalfördelat.

Om medelvärdet beräknas på få observationer som inte kan antas vara normalfördelade finns möjligheten att:

- Använda ickeparametriska test. Då testas vanligtvis medianen istället för medelvärdet.
- Transformera data för att uppnå en mer symmetrisk fördelning.

### E.11 Värden under detektions- eller kvantifieringsgräns

Värden under en detektions- eller kvantifieringsgräns kan användas för trendanalyser om gränsen inte förändras över tid. Oftast ersätts dessa värden med halva gränsvärdet och behandlas som vanliga observationer i den statistiska beräkningen. Om detektions- eller kvantifieringsgränsen ändras över tid kan värdena ersättas med halva värdet av högsta förekommande gräns. För långa serier kan detta innebära en förlust av viktig information och man bör också fundera över om trendanalysen måste göras på hela serien eller om en kortare tidsperiod kan räcka.