



Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Institutionen för vatten och miljö

Bakgrundskorrigerering av fosfatfosfor

-

Påverkan på tidserier vid förändring i analysmetoden för fosfatfosfor 2014

C. Von Brömssen, K. Wallman, J. Fölster

SLU, Vatten och miljö: Rapport 2019:11

Referera gärna till rapporten på följande sätt:

Von Brömssen, C., Wallman, K., Fölster, J. (2019) Bakgrundskorrigerig av fosfatfosfor – Påverkan på tidserier vid förändring i analysmetoden för fosfatfosfor 2014. Rapport/Sveriges lantbruksuniversitet, Institutionen för vatten och miljö (2019:11)

Kontakt

Karin.Wallman@slu.se

<http://www.slu.se/vatten-miljo>

Innehåll

Förord	4
1. Bakgrund	5
2. Modeller för bakgrundskorrigerig	5
2.1 Prediktionsmodell baserad på observationer efter maj 2016	6
2.2 Prediktionsmodell baserad på observationer efter juni 2014.....	8
2.3 Prediktionsmodell för turbiditet	10
3. Hur bakgrundskorrigeringen beräknas.....	12
4. Test av modellerad bakgrundskorrigerig	13
4.1 Färgeån	14
4.2 Alelyckan.....	15
4.3 Kalix älv Karlsborg	15
4.4 Ostvik.....	16
5. Slutsatser	16
6. Referenser	16

Förord

Vattenkemiska laboratoriet vid Institutionen för vatten och miljö har utfört analyser inom miljöövervakningen av sjöar och vattendrag i mer än 50 år. Utvecklingen av analysmetoder och analysinstrument under dessa 50 år har inneburit att man i bland måste byta metod. Detta sker dock på ett sådant sätt så att det säkerställer möjligheterna att kunna använda och tolka tidsserierna.

Vid metodbyte av fosfatfosfor 2014 introducerades en bakgrundskorrigeringsmetod av värdena vid analysen dvs. att störningar från partiklar och eventuell bakgrundsfärg dras från det uppmätta värdet. Vid valideringen av den nya metoden uppmärksammades att detta skulle kunna leda till stegvisa förändringar i tidserierna. Det beslöts därför att vi efter ett antal år skulle utvärdera denna påverkan på tidserierna för att se om vi kunde korrigera äldre data för felet från bakgrunden. Under de fem år som gått sedan metodförändringen har de bakgrundskorrigerade värdena rapporterats till kund samtidigt som okorrigerade värden sparats i databasen för att möjliggöra en utvärdering.

Institutionen för vatten och miljö har gett statistikern Claudia Von Brömssen på Institutionen för energi och teknik, SLU, i uppdrag att utvärdera resultaten för fosfatfosfor för att se om det går att korrigera äldre data för bakgrundskorrigeringen.

Karin Wallman

Biträdande laboratoriechef

Jens Fölster

Sektionschef geokemi

1. Bakgrund

Bakgrundskorrigerings för fosfatfosfor görs sedan juni 2014. För att kunna justera äldre serier kan statistiska modeller användas, där bakgrundskorrigerings predikteras med hjälp av andra tillgängliga variabler. Huvudvariabeln som kommer att användas för att modellera bakgrundskorrigerings är turbiditet, som dock i sin tur inte observerades för tidiga serier och då måste skattas ur differensen av filtrerad och ofiltrerad absorbans (absdiff). Den förslagna korrigerings testas på några serier.

Projektet består av tre steg:

1. Att definiera en prediktionsmodell där bakgrundskorrigerings förklaras med hjälp av framförallt turbiditet, samt ytterligare variabler om de ökar förklaringsgraden.
2. Att definiera en prediktionsmodell för turbiditet med hjälp av absdiff.
3. Att testa bakgrundskorrigerings av serier av fosfat med hjälp av modellerna framtagna i 1. och 2.

Bakgrundskorrigerings användes sedan juni 2014, men med olika instrument och med olika detektionsgränser.

Detektionsgränsen var 1 µg/l 201406-201604 (analys diskret analysator Gallery)

Sedan 201605 analyseras alla prov, där Gallery erhållit en halt mindre än rapporteringsgränsen (4 µg/l), om på ett segmenterat flödessystem (AA3 respektive QuAAtro). Instrumentet QuAAtro ersatte AA3 201903 men de båda metoderna är snarlika och har samma detektionsgräns.

Detektionsgränsen är 0,3 µg/l fr o m 201605 (analys AA3 eller QuAAtro)

2. Modeller för bakgrundskorrigerings

Två ansatser görs för att bygga en prediktionsmodell för bakgrundskorrigerings:

- a. Enbart värden från och med maj 2016 används.
- b. Alla data från och med juni 2014 används.

De lägsta värdena tas inte med i denna modellenpassning, dvs modellen baseras enbart på observationer som har en bakgrundskorrigerad halt över 0.8 µg/l. Eftersom modellerna beräknas på log₁₀-skala för alla variabler sorteras även observationer som har värde 0 för någon av de potentiella förklarande variablerna bort. Modellenpassningen görs för värden som representerar ytprov.

2.1 Prediktionsmodell baserad på observationer efter maj 2016

Genom grafer, stegvisa regressionsmodeller och expertkunskap identifierades följande mest intressanta förklarande variabler: Turbiditet, Absorbans F 420/5 och TOC. Modell 1a för värden från och med maj 2016 är en generell mixed modell med dessa förklarande variabler som fixa faktorer och en slumpmässig faktor som representerar station. Den slumpmässiga faktorn behövs då det flera observationer görs på varje station. Autokorrelationen inom serierna skattas inte.

Linear mixed-effects model fit by REML

Data: indata1

AIC	BIC	logLik
526.7663	562.2758	-257.3831

Random effects:

Formula: ~1 | StNamn

(Intercept) Residual

StdDev: 0.220311 0.2429271

Fixed effects: logBakg ~ logTOC + logAbsF + logTURB

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-0.9881390	0.11030520	2437	-8.95823	0.0000
logTOC	0.7462261	0.06957892	2437	10.72489	0.0000
logAbsF	-0.1270355	0.05006519	2437	-2.53740	0.0112
logTURB	0.6690266	0.01959041	2437	34.15072	0.0000

Correlation:

	(Intr)	logTOC	lgAbsF
logTOC	-0.969		
logAbsF	0.939	-0.867	
logTURB	0.035	-0.131	-0.006

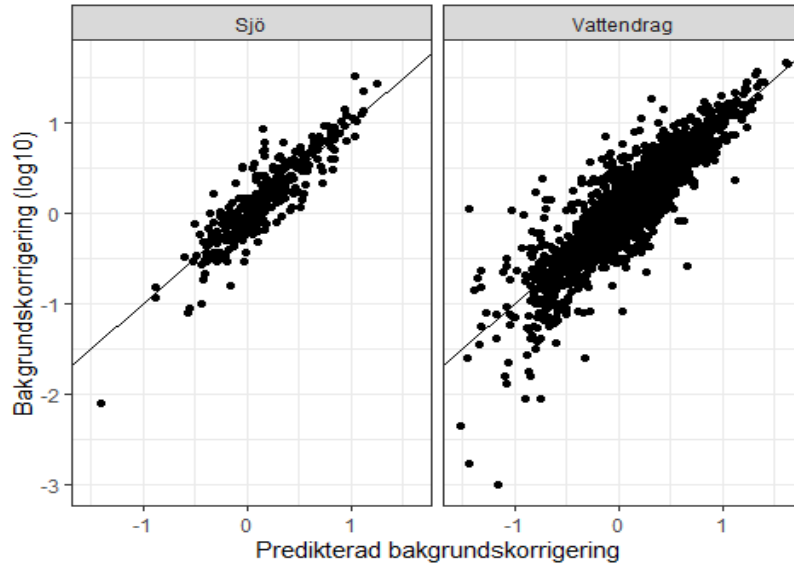
Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-7.56403345	-0.50341675	0.01330605	0.50513206	6.12344370

Number of Observations: 2751

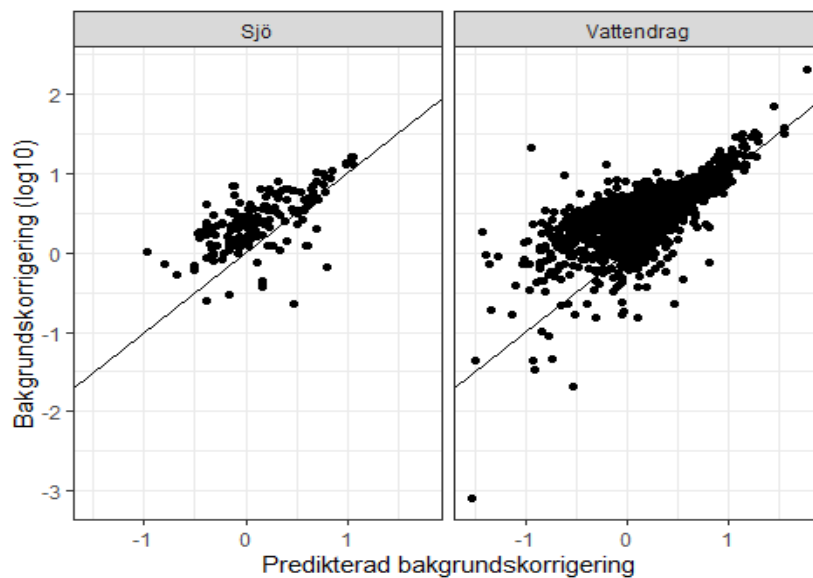
Number of Groups: 311

Resultat 1: Modellutskrift från en generell linjär mixed modell där nivån av den \log_{10} -transformerade bakgrundskorrigeringen modelleras med hjälp av TOC, absorbans och turbiditet (alla \log_{10} -transformerade). Station är inkluderat som slumpmässig faktor. Data från och med maj 2016. Modell 1a.



Figur 1: Predikterade mot observerade bakgrundskorrigeringsnivåer för data från och med maj 2016.

Figur 1 visar prediktioner från modell 1a mot de observerade värdena för bakgrundskorrigerig. Eftersom data mellan juni 2014 och april 2016 inte har tagits med i denna anpassning kan det vara intressant att reda ut hur bra modellen lyckas prediktera de äldre värdena som då alltså inte ingick i modellanpassningen. Resultat för denna anpassning visas i Figur 2 nedan. Generellt leder modellen till en underskattning av bakgrundskorrigerig av äldre data, antagligen eftersom de äldre lägre värdena är bestämda med ett annat instrument (AA3).



Figur 2: Prediktioner för tiden juni 2014 till april 2016 med en modell anpassad för tiden maj 2016 till 2018.

2.2 Prediktionsmodell baserad på observationer efter juni 2014

Ett alternativ är att anpassa modellen för alla data sedan juni 2014.

```

Linear mixed-effects model fit by REML
Data: indata2
      AIC      BIC   logLik
2389.7 2428.031 -1188.85

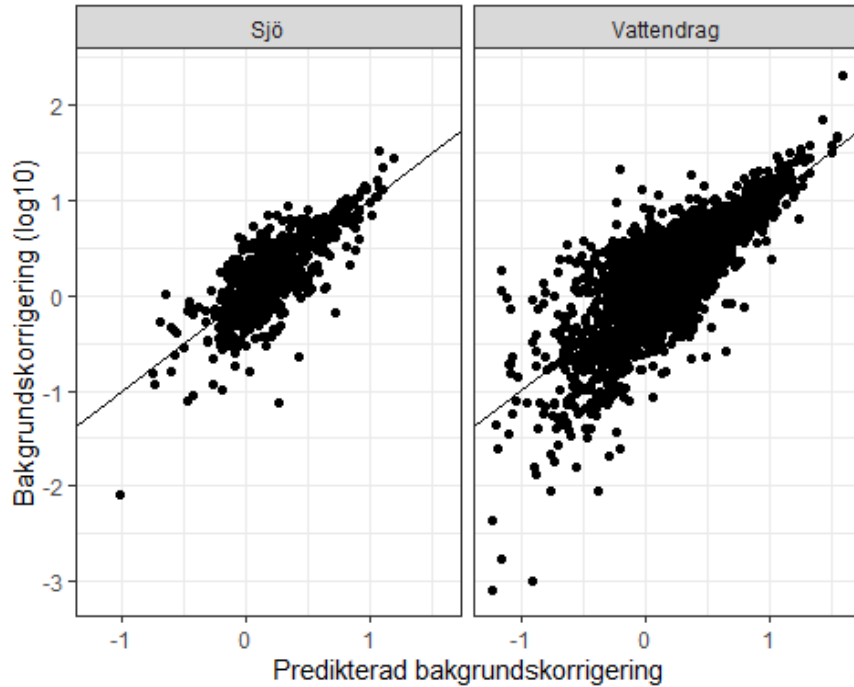
Random effects:
Formula: ~1 | StNamn
      (Intercept) Residual
StdDev:   0.2061643 0.2987335

Fixed effects: logBakg ~ logTOC + logAbsF + logTURB
              Value Std.Error   DF  t-value p-value
(Intercept) -0.5294431 0.09838399 3992 -5.381395  0.000
logTOC       0.5388144 0.06271073 3992  8.592061  0.000
logAbsF      -0.0462860 0.04392278 3992 -1.053803  0.292
logTURB      0.5560537 0.01855829 3992 29.962555  0.000
Correlation:
      (Intr) logTOC lgAbsF
logTOC  -0.966
logAbsF  0.938 -0.853
logTURB  0.038 -0.139 -0.042

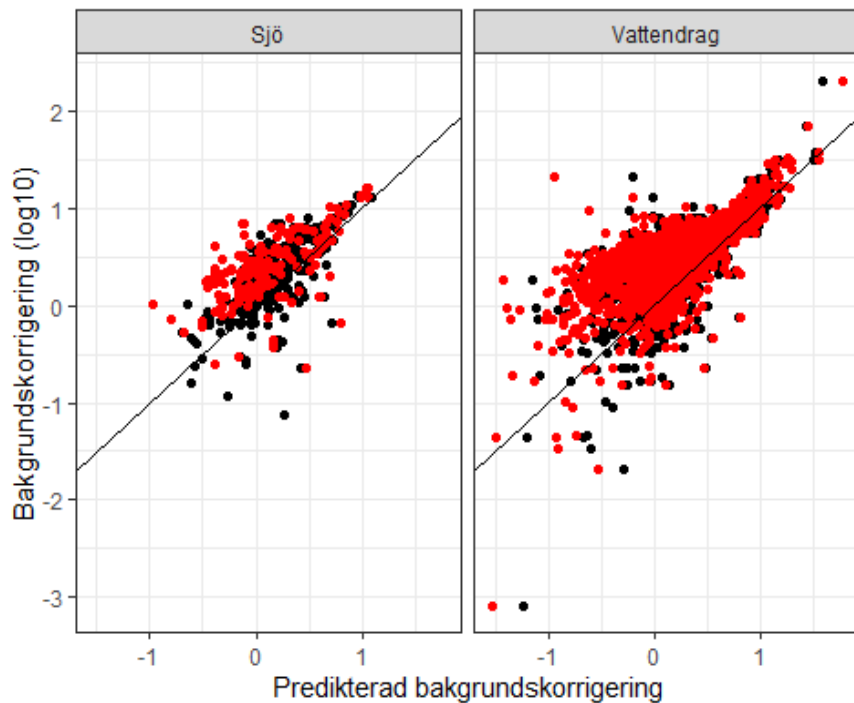
Standardized Within-Group Residuals:
      Min          Q1          Med          Q3          Max
-6.986262801 -0.546736648  0.008864092  0.550057637  5.090881550

Number of Observations: 4400
Number of Groups: 405
    
```

Resultat 2: Samma modell som i resultat 2, men anpassad på data från och med juni 2014. Modell 1b.



Figur 3: Prediktioner och observerade bakgrundskorrigeringar för perioden juni 2014 till 2018.

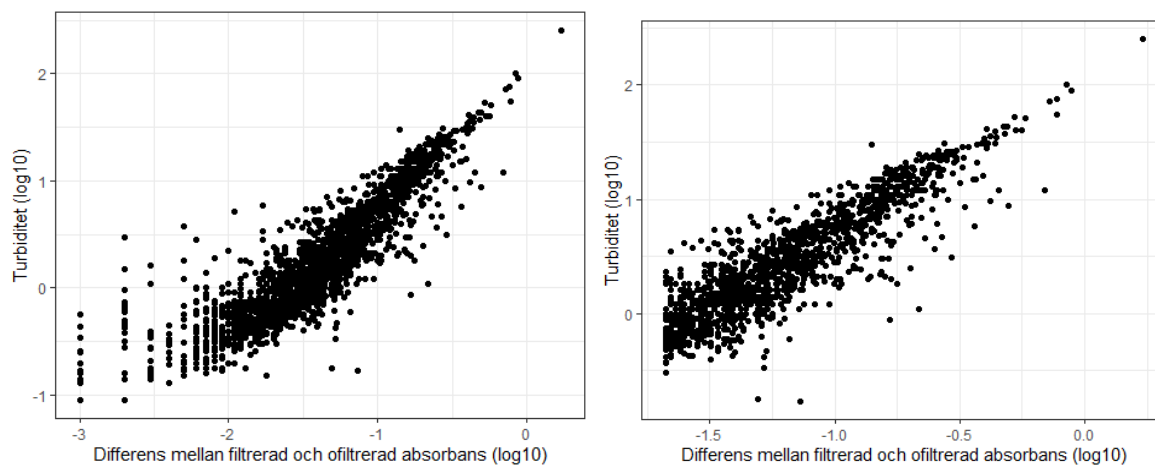


Figur 4: Jämförelse av prediktioner för perioden juni 2014 till april 2016 med modellerna 1a (anpassad på maj 2016 till 2018, röd) och 1b (anpassad för tiden juni 2014 till 2018, svart).

Prediktionerna blir ganska lika för de två modellerna, modell 1a och modell 1b (Figur 4). Modellparametrarna skiljer sig dock en del mellan modell 1a och modell 1b pga. att förklarande variabler i modellerna är korrelerade. Modellerna ska användas enbart med syfte att producera så bra prediktioner som möjligt. Vi drar slutsatsen att det inte spelar en någon avgörande roll vilken modell som används för bakgrundskorrigerig.

2.3 Prediktionsmodell för turbiditet

När turbiditet inte finns observerad, så kan den i sin tur predikteras med hjälp av differensen mellan filtrerat och ofiltrerat absorbans (absdiff).



Figur 5: Samband mellan absdiff och turbiditet (vänster) och samma samband när de lägsta värden för absdiff har tagits bort (höger)

I figur 5 ser man att låga värden avviker från ett linjärt samband. En möjlighet att hantera detta är att anpassa en kvadratisk eller kubisk modell eller att begränsa anpassningen till värden där t.ex. absdiff är större än 0.02 (Figur 5, höger). Här väljs det senare alternativet.

Den anpassade modellen och predikterade värden finns nedan (Resultat 3) och illustreras i Figur 6.

```
Linear mixed-effects model fit by REML
```

```
Data: turb_absdiff1
      AIC      BIC logLik
-1029.68 -1008.494 518.84
```

```
Random effects:
```

```
Formula: ~1 | StNamn
```

```

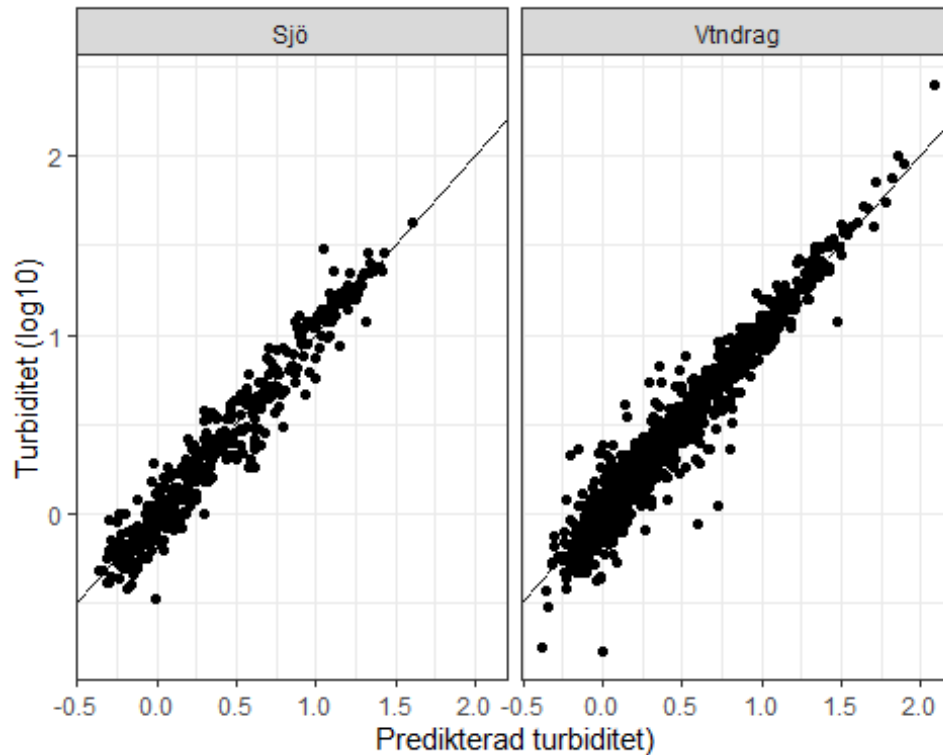
(Intercept) Residual
StdDev:      0.178274 0.130645

Fixed effects: logturb ~ log_absdiff
              Value Std.Error   DF  t-value p-value
(Intercept)  1.722901 0.02213079 1073  77.85084    0
log_absdiff  1.088730 0.01616119 1073  67.36697    0
Correlation:
      (Intr)
log_absdiff 0.897

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-5.90314316 -0.44553191  0.02731459  0.48296635  4.04018780

Number of Observations: 1477
Number of Groups: 403
    
```

Resultat 3: En prediktionsmodell för \log_{10} -transformerat turbiditet baserad på den \log_{10} -transformerade differensen mellan filtrerat och ofiltrerat absorbans.



Figur 6: Observerade och predikerade värden för \log_{10} -transformerad turbiditet för sjöar och vattendrag.

3. Hur bakgrundskorrigeringen beräknas

Följande varianter på beräkningar föreslås för beräkning av bakgrundskorrigerig. Observera att beräkningarna görs på log₁₀-skala och att de måste tillbakatransformeras ($10^{\log_{10}(\text{bakgkorr})}$) innan korrigerig kan dras från de uppmätta värdena på originalskala.

- a) Om TOC, absorbans och turbiditet finns tillgängligt görs bakgrundskorrigeringen med hjälp av följande modell (model1b):

$$\begin{aligned} \log_{10}(\text{bakgkorr}) \\ = -0.529 + 0.539 \cdot \log_{10}(\text{TOC}) - 0.0463 \cdot \log_{10}(\text{AbsF}) + 0.556 \cdot \log_{10}(\text{TURB}) \end{aligned}$$

- b) Om turbiditet inte finns tillgängligt kan den i sin tur skattas genom:

$$\log_{10}(\text{TURB}) = 1.723 + 1.089 \cdot \log_{10}(\text{absdiff})$$

där absdiff är differensen mellan ofiltrerat och filtrerat. Det ger modellen:

$$\begin{aligned} \log_{10}(\text{bakgkorr}) \\ = -0.529 + 0.539 \cdot \log_{10}(\text{TOC}) - 0.0463 \cdot \log_{10}(\text{AbsF}) + 0.556 \cdot (1.723 + 1.089 \\ \cdot \log_{10}(\text{absdiff})) \end{aligned}$$

som blir:

$$\begin{aligned} \log_{10}(\text{bakgkorr}) \\ = 0.429 + 0.539 \cdot \log_{10}(\text{TOC}) - 0.0463 \cdot \log_{10}(\text{AbsF}) + 0.605 \cdot \log_{10}(\text{absdiff}) \end{aligned}$$

- c) Om TOC inte finns tillgängligt kan bakgrundskorrigerig istället göras med följande modeller:

$$\log_{10}(\text{bakgkorr}) = 0.297 + 0.286 \cdot \log_{10}(\text{AbsF}) + 0.558 \cdot \log_{10}(\text{TURB})$$

eller när turbiditet inte finns:

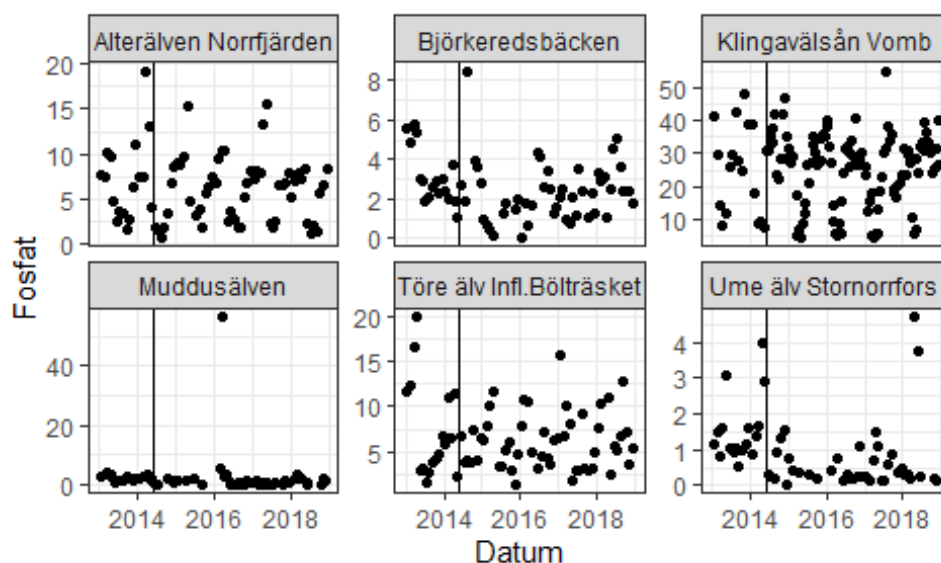
$$\log_{10}(\text{bakgkorr}) = 0.99 + 0.286 \cdot \log_{10}(\text{AbsF}) + 0.627 \cdot \log_{10}(\text{absdiff})$$

4. Test av modellerad bakgrundskorrigerigering

Några trendvattendrag väljs ut för att testa modellen för bakgrundskorrigerigeringen för äldre data, dvs för en tidsperiod där den verkliga bakgrundskorrigerigeringen inte är känd. Observationer mellan 2010 och maj 2014 korrigeras med modell 1b, dvs genom att använda turbiditet, TOC och absorbans.

När alla serier är bakgrundsjusterade undersöker vi om man kan upptäcka hopp i serierna vid positionerna där metodbyte sker, dvs juni 2014. För att det förväntade värdet i serien kan ändra sig över tiden används även en trendkomponent modellerad med en generaliserad additiv modell (Wood, 2017).

De flesta tidserierna visar inte några hopp i den visuella bedömningen (Figur 7)



Figur 7: Några trendvattendrag med indikationer när beräkningsmetoder ändras juni 2014.

P-värden för det skattade hoppet tas fram för alla trendvattendrag. P-värden ska användas enbart som indikation, eftersom modelleringen t.ex. inte har tagit hänsyn till autokorrelation eller säsongvariation i data, som antagligen leder till en underskattning av p-värdena. Det i sin tur leder till för många signifikanta hopp visas (för många falska positiva resultat).

Totalt är 19 av 71 hopp signifikanta, 13 uppvisar en minskning vid tidpunkt juni 2014 och 6 en uppgång (Resultat 4). Om ingen bakgrundskorrigerigering görs så är 40 hopp signifikanta (2 uppåt och 38 nedåt). Som nämnt är det antagligen är en överskattning av antal signifikanta hopp.

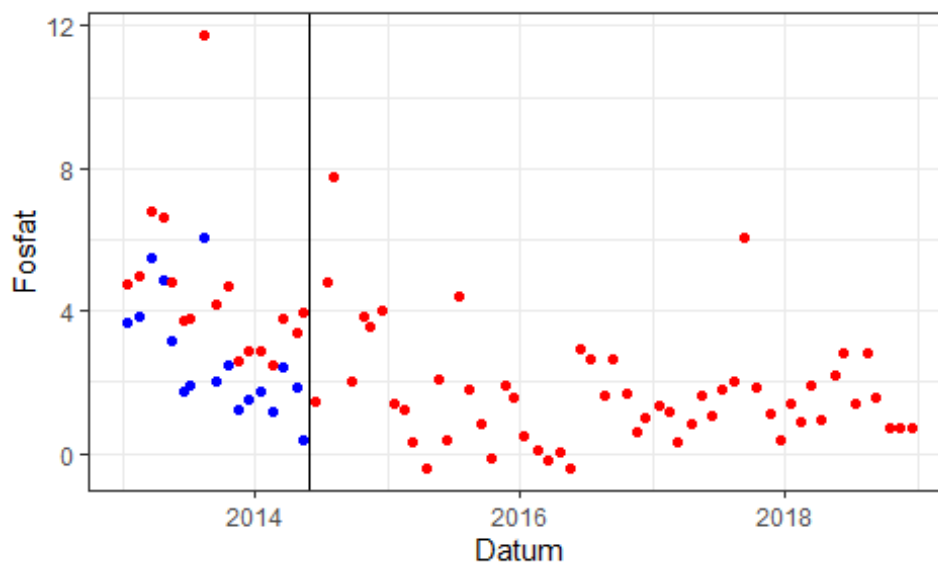
För att undersöka när hopp upptäcks tittar vi närmare på några av serierna (Figur 8 –Figur 11).

```
# A tibble: 19 x 4
```

StNamn	fosfatmean	`P-value 2014`	jump2
<chr>	<dbl>	<dbl>	<dbl>
1 Pite älv Bölebyn	1.96	0.0045	-2.87
2 Färgeån	2.06	0.0019	2.88
3 Gavleån Gävle stadspark	2.22	0.0121	-2.07
4 Rickleån Utl	2.31	0.0022	-1.63
5 Nordre älv, Gullö	2.43	0.0041	-3.27
6 Björkeredsbäcken	2.53	0.0196	2.30
7 Ätran Falkenberg	2.85	0.0424	-2.73
8 Öre älv Torrböle	2.90	0.0148	-2.78
9 Akkarjåkkå	2.95	0.0045	-2.32
10 Alelyckan	3.32	0.000600	-4.85
11 Kalix älv Karlsborg	3.60	0.02	-3.38
12 Lögde älv Lögdeå	4.11	0.0161	-1.86
13 Helgeån Hammarsjön	5.02	0.0027	-5.05
14 Viskan Åsbro	5.76	0.0121	-4.66
15 Bjurforsbäcken	6.43	0.0047	14.2
16 Ostvik	7.03	0.0015	-11.4
17 Tolångaån Tolånga	36.0	0.0435	22.2
18 Kävlingeån Högsmölla	38.7	0.0373	21.7
19 Skivarpsån Skivarp	76.8	0.0015	70.2

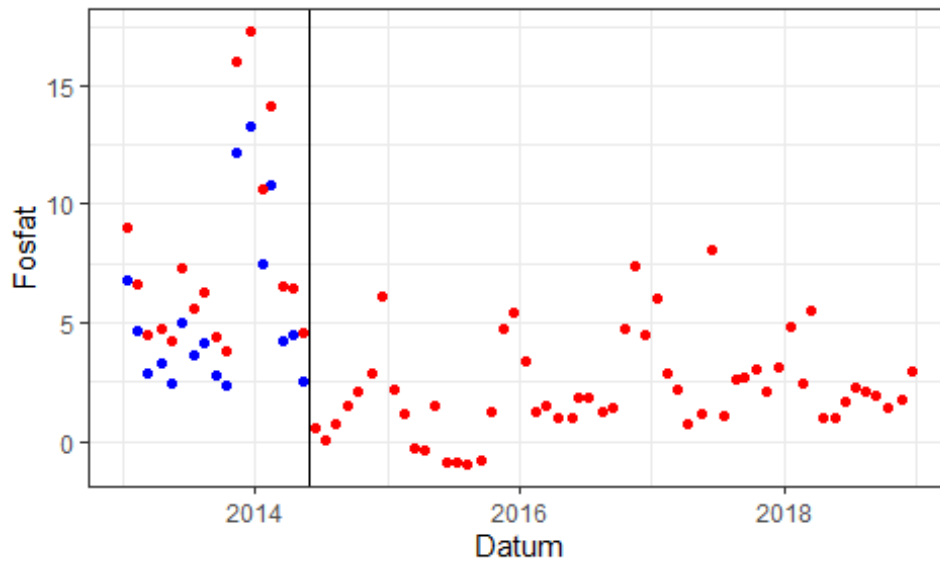
Resultat 4: Vattendrag som uppvisar ett hopp juni 2014.

4.1 Färgeån



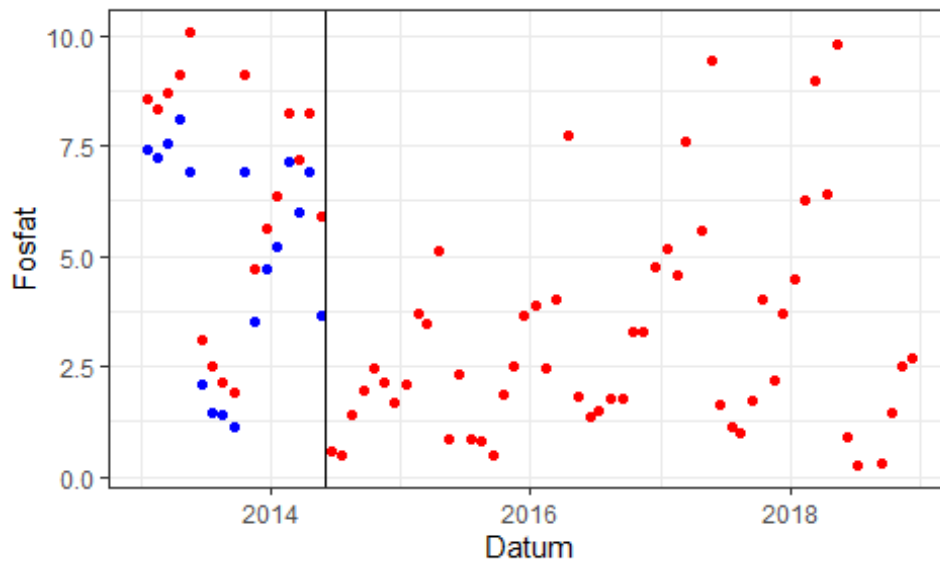
Figur 8: Fosfat, station Färgeån, Rapporterade värden (röd) och värden före juni 2014 efter bakgrundskorrigerig (blå).

4.2 Alelyckan



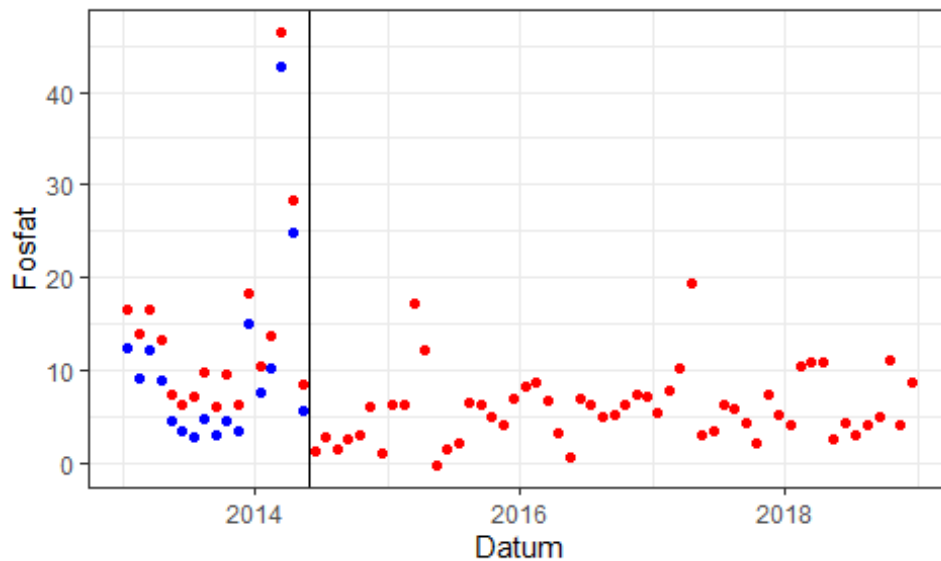
Figur 9: Fosfat, station Alelyckan, Rapporterade värden (röd) och värden före juni 2014 efter bakgrundskorrigerigering (blå).

4.3 Kalix älv Karlsborg



Figur 10: Fosfat, station Kalix älv Karlsborg, Rapporterade värden (röd) och värden före juni 2014 efter bakgrundskorrigerigering (blå).

4.4 Ostvik



Figur 11: Fosfat, station Ostvik, Rapporterade värden (röd) och värden före juni 2014 efter bakgrundskorrigerig (blå).

5. Slutsatser

I denna rapport föreslår vi en justering av äldre fosfatvärden som liknar den bakgrundsjustering som görs idag i samband med den kemiska analysen. Den föreslagna justeringen är baserad på statistiska regressionsmodeller där de förklarande variablerna TOC, absorbans och turbiditet används. Modelleringsresultat visar att bakgrundskorrigeringen kan till viss del bestämmas ur dessa förklarande variabler och att de predikterade värdena inte avviker systematisk från den faktiska bakgrundskorrigeringen som görs.

Det är svårt att avgöra hur bra gamla serier kan justeras med hjälp av dessa modeller. Studien av några tidserier som uppvisar signifikanta hopp under juni 2014 visar att detta hopp är mindre när den statistiska bakgrundskorrigeringen görs. Även antalet signifikanta hopp minskar till hälften när bakgrundskorrigeringen görs med de föreslagna modellerna.

Samtidigt kan enbart en del av variationen i den faktiska bakgrundskorrigeringen förklaras med hjälpvariablerna. Statistiskt justerade serier kommer alltid vara mer osäkra än bestämningarna efter juni 2014, där bakgrundskorrigerig görs direkt i den kemiska analysen.

6. Referenser

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.

