



Strål  
säkerhets  
myndigheten

Swedish Radiation Safety Authority

Författare: Rolf Bergman

Forskning

# 2014:14

Multivariatanalys av radioaktivitetsdata från utsläpp till luft och vatten från Forsmarks Kraftgrupp AB och i omgivningen av Forsmark



## SSM perspektiv

### Bakgrund

De svenska kärntekniska anläggningarna genomför regelbundet mätningar av utsläppt radioaktivitet till luft och vatten samt mätningar av radioaktivitet i olika provslag i omgivningen. Omgivningskontrollen genomförs enligt ett program framtaget av dåvarande Statens strålskyddsinstitut, nuvarande Strålsäkerhetsmyndigheten (SSM). Det nuvarande omgivningskontrollprogrammet ska revideras av SSM.

### Syfte

Syftet med denna studie var att undersöka korrelationer mellan uppmätt utsläppt radioaktivitet från kärnkraftverket i Forsmark och uppmätta halter radioaktivitet i omgivningen av Co-58, Co-60, Cs-137, Fe-59, Mn-54 och Zn-65. Ett mål var att se hur olika provslag i omgivningen av Forsmark fungerar som indikatorer för den uppmätta utsläppta radioaktiviteten. Resultaten från denna studie kommer att tas hänsyn till vid den kommande revideringen av omgivningskontrollprogrammet.

### Resultat

Studien visar korrelation mellan uppmätt utsläppt radioaktivitet och uppmätt radioaktivitet i omgivningen för vattenrelaterade utsläpp med signifikanta modeller för nukliderna Co-58, Co-60, Cs-137, Mn-54 och Zn-65. Gällande luftrelaterade utsläpp kunde studien endast påvisa en signifikant modell för Co-60. För Fe-59 kunde inte några signifikanta modeller identifieras för varken luft- eller vattenutsläpp.

Studien har även påvisat ett stort behov av fungerande databaser med utsläpps- och omgivningsdata. I föreliggande studie har de använda radioaktivitetsvärdena hämtats från gamla databaser, sparade textfiler och pappersrapporter. Det har inte varit möjligt att kvalitetssäkra använda data under den begränsade tid som fanns till förfogande.

### Behov av ytterligare forskning

Denna studie analyserade data för sex specifika nuklider i utsläpp från kärnkraftverket i Forsmark och i omgivningen runt detta. Dessa resultat kan inte extrapoleras till att gälla andra anläggningar eller andra nuklider. Därför finns behov av att dels göra motsvarande studier för andra kärntekniska anläggningar, dels studera fler nuklider. En undersökning om vilka nuklider som är mest intressanta för korrelationsstudier bör föregå ett sådant projekt.

### Projekt information

Kontaktperson SSM: Charlotte Lager

Referens: SSM2013-2219





Strål  
säkerhets  
myndigheten

Swedish Radiation Safety Authority

Författare: Rolf Bergman  
Rolf Bergman Consulting, Uppsala

# 2014:14

Multivariatanalys av radioaktivitetsdata  
från utsläpp till luft och vatten från  
Forsmarks Kraftgrupp AB och i  
omgivningen av Forsmark

Datum: December 2013

Rapportnummer: 2014:14 ISSN:2000-0456

Tillgänglig på [www.stralsakerhetsmyndigheten.se](http://www.stralsakerhetsmyndigheten.se)

Denna rapport har tagits fram på uppdrag av Strålsäkerhetsmyndigheten, SSM. De slutsatser och synpunkter som presenteras i rapporten är författarens/författarnas och överensstämmer inte nödvändigtvis med SSM:s.

# Innehåll

<b>SAMMANFATTNING</b> .....	<b>3</b>
<b>1. BAKGRUND</b> .....	<b>4</b>
<b>2. SYFTE</b> .....	<b>5</b>
<b>3. DATA</b> .....	<b>6</b>
<b>4. ANALYSVERKTYG</b> .....	<b>7</b>
4.1. PCA.....	7
4.2. PLS.....	7
4.3. ANALYSGÅNG .....	9
<b>5. RESULTAT OCH DISKUSSION</b> .....	<b>10</b>
5.1. DATA .....	10
5.2. PCA – ALLA DATA.....	12
5.3. PCA – UTSLÄPPSDATA.....	14
5.4. PLS – ALLA DATA.....	15
5.5. PLS – UTSLÄPP TILL LUFT RESPEKTIVE VATTEN.....	16
5.6. PLS – PER NUKLID .....	17
5.6.1. <sup>58</sup> Co .....	17
Utsläpp till luft.....	17
Utsläpp till vatten .....	18
5.6.2. <sup>60</sup> Co .....	22
Utsläpp till luft.....	22
Utsläpp till vatten .....	23
5.6.3. <sup>137</sup> Cs .....	27
Utsläpp till luft.....	27
Utsläpp till vatten .....	27
5.6.4. <sup>59</sup> Fe .....	31
Utsläpp till luft.....	31
Utsläpp till vatten .....	32
5.6.5. <sup>54</sup> Mn .....	32
Utsläpp till luft.....	32
Utsläpp till vatten .....	32
5.6.6. <sup>65</sup> Zn .....	34
Utsläpp till luft.....	34
Utsläpp till vatten .....	34
5.7. PREDIKTIONSFÖRMÅGA OCH PREDIKTIONER .....	40
<b>6. REKOMMENDATIONER</b> .....	<b>42</b>
<b>7. SLUTSATSER</b> .....	<b>43</b>





# Sammanfattning

En multivariat utvärdering har gjorts av kontrollerade utsläpp av radioaktiva ämnen och aktivitet i vissa provslag (t.ex. vissa alger och fiskar) i omgivningen av Forsmark för perioden 1980-2012. De studerade nukliderna är  $^{58}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$ ,  $^{59}\text{Fe}$ ,  $^{54}\text{Mn}$  och  $^{65}\text{Zn}$ . Utsläppen sker kontrollerat till luft och vatten från de kärntekniska anläggningarna i Forsmark.

Multivariatanalys innebär att många variabler kan analyseras samtidigt. Detta medför också att de resultat som fås kan påvisa hur fler variabler (här provslag) samtidigt påverkas av t.ex. utsläpp av radioaktiva ämnen. Den genomförda multivariatanalysen visar att det går att beräkna storleken på utsläpp från Forsmark genom mätvärden från ett fåtal olika provslag (t.ex. blåstång, abborre och påväxtprover) vid någon eller några av de utvalda mätstationerna i omgivningarna av Forsmark. Resultaten visar också att det är möjligt att beräkna hur stor koncentration, av de här studerade radioaktiva nukliderna, som kan förväntas i ett antal olika provslag och mätstationer vid "normala" utsläpp.

De multivariata sambanden är framtagna ur ett stort antal variabler, ursprungligen 381 stycken där varje enskild variabel definieras av nuklid, mätstation och provslag (ex.  $^{58}\text{Co\_S101-Påväxt}$ ). Analysen visar att det för 4 av de 6 studerade nukliderna ( $^{58}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$  och  $^{65}\text{Zn}$ ) är tillräckligt att mäta på endast ett provslag vid någon mätstation för att få en mycket god uppfattning om storleken av ett kontrollerat utsläpp till vatten. För nukliden  $^{54}\text{Mn}$  behövs minst 2 provslag, för samma uppskattning. Ett undantag är  $^{59}\text{Fe}$  där inga säkra samband kunde påvisas. Detta beror förmodligen på ett för litet antal mätvärden från proverna i omgivningen för denna nuklid.

För utsläpp av radioaktiva nuklider till luft kunde bara motsvarande beräkningar göras för nukliden  $^{60}\text{Co}$  och då med hjälp av minst 2 provslag.

# 1. Bakgrund

Kärntekniska anläggningar är enligt lag skyldiga att genomföra årliga mätprogram för att kontrollera att verksamheten inte ger upphov till några miljöeffekter i närområdet. Även EU-kommissionen har utfärdat vissa riktlinjer för hur detta mätprogram (omgivningskontroll) ska vara utformat (96/29/Euratom). Permanenta omgivningskontrollprogram fastställdes i slutet av 1970-talet av dåvarande Statens strålskyddsinstitut (SSI) nuvarande Strålsäkerhetsmyndigheten (SSM). Myndigheten var redan från början ansvarig för utformningen av programmet i samarbete med Naturvårdsverket medan tillståndshavarna blev ansvariga för mätningar samt i viss utsträckning för provtagningarna. SSM genomför dock kontinuerligt egna mätningar på ett urval av de prover som samlas in. Programmet har återkommande reviderats, senast år 2004 (*Lindén SSI-R 2004:15*).

Mätningarna som ligger till grund för den aktuella rapporten började samlas in 1980 för kontrollerade utsläpp och det har sedan 1981 samlats ansevärliga mängder mätdata, från ett stort antal mätstationer och provslag. Den långt ifrån optimala organisationen av data har gjort det svårt att extrahera eventuell information om, och i vilken omfattning, de kontrollerade utsläppen till luft och vatten korrelerar dels med varandra, men också med den uppmätta aktiviteten i olika provslag i tid och rum.

Eftersom antalet variabler (namngivna enligt: nuklid\_mätstation-provslag) i föreliggande datamatrix, vida överstiger antalet observationer (varje år är en observation) blir det i praktiken omöjligt att behandla värdena med traditionell en-variabelsstatistik, som vanligtvis kräver betydligt fler observationer och det får inte saknas några observationer i tidsserierna. Dessutom måste jämförda variabler i det fallet också vara linjärt oberoende, vilket inte är fallet i den aktuella sammanställningen.

På grund av att datamatriken har ett fåtal observationer och ett stort antal variabler, är det lämpligt att utvärdera data med multivariata verktyg. På marknaden befintliga mjukvaror är konstruerade att hantera korrelerade variabler och data kan också saknas vilket kallas "missing data". Beräkningsalgoritmerna togs fram på 1930-talet i USA för att avslöja fusk vid totalisatorspel, vilket också lyckades, där antalet "missing data" ofta var så stort som 75 %.

Multivariatanalys av datamatriken har utförts med mjukvaran SIMCA-P+ 12.01 (Soft Independent Modeling of Class Analogies), från Umetrics i Umeå, [www.umetrics.com](http://www.umetrics.com).

## 2. Syfte

Syftet med denna multivariatanalys är att undersöka om och hur normala utsläpp av radionuklider till luft och vatten från den kärntekniska anläggningen i Forsmark korrelerar med uppmätta aktiviteter i olika provslag och vid olika provstationer. Analysen syftar också till att eventuellt ge myndigheten underlag för att revidera omgivningskontrollprogrammet.

Mot bakgrund av hur tidserierna ser ut, med årsmedelvärden av samtliga aktiviteter, är det inte möjligt att få en detaljerad uppfattning om spridningskinetiken. För det ändamålet krävs att mätningarna i omgivningen utförs samtidigt med mätningarna av utsläppen av radioaktiva ämnen och med samma periodicitet.

### 3. Data

Den aktuella datamatriken består av insamlade mätvärden för årliga utsläpp av nukliderna  $^{58}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$ ,  $^{59}\text{Fe}$ ,  $^{54}\text{Mn}$  och  $^{65}\text{Zn}$ , till luft (L) och vatten (V) [Bq/år] från de kärntekniska anläggningarna i Forsmark. Aktiviteten av respektive nuklid utgör medelvärden per varje år. Datamatriken innehåller också årsvisa medelvärden av mätdata för varje nuklid i olika provslag [Bq/kg], för luft respektive vatten vid olika mätstationer.

Mätstationer och provslag är de som omnämns i omgivningskontrollprogrammet.

Datamatriken är organiserad så att varje år utgör en observation, och varje variabel definieras av nuklid\_mätstation-provslag (ex.  $^{58}\text{Co}$ \_S101-Gädda). Det finns 33 observationer (1980-2012) och innan exkludering på grund av för få mätpunkter, totalt 381 variabler. En del variabler har bara ett fåtal observationer och kommer inte att kunna beaktas vid modellberäkningarna då dessa endast genererar brus. I praktiken har variabler direkt uteslutits när antalet observationer varit färre än 6 av 33 möjliga, dvs. ungefär 80 % "missing data".

## 4. Analysverktyg

Nedan följer några korta meningar om en del av de använda analysteknikerna och i rapporten frekvent använda begrepp. För detaljerad information hänvisas till [www.umetrics.com](http://www.umetrics.com). Även analysgången beskrivs översiktligt.

### 4.1. PCA

Principal Component Analysis (PCA) används ofta som ett första steg för att upptäcka korrelationer i en datamängd. Metoden är bra på att identifiera likheter/olikheter bland olika observationer samt att se grupperingar och extremvärden, s.k. "outliers". Den första principalkomponenten (PC1) visar den riktning i data som har den största variationen. Den näst största variationen, ortogonal mot föregående, visas av den andra principalkomponenten (PC2) och så vidare. PC1 visar på detta sätt en projektion av den eller de mest betydelsefulla variablerna i den studerade ursprungsmatrisen.

En "score plot" ger en karta över hur lika/olika de studerade observationerna är. Observationer som ligger nära varandra har mycket gemensamt i de ingående variablerna. På motsvarande sätt, om observationerna befinner sig långt ifrån varandra i "score plot" skiljer sig värdena åt i den ursprungliga variabeltabellen. I grafer med t.ex. PC1 mot PC2 visas ofta 95 % konfidensnivån, vilket blir en ellips i två dimensioner. Observationer som hamnar utanför denna ellips kan vara potentiella "outliers", men det finns också andra diagnostiska verktyg, i den använda mjukvaran, för att avslöja detta.

En "loading plot" (variabelkarta) visar en karta över hur de olika variablerna är korrelerade. "Loading plot" visar också varför de olika observationerna hamnar som de gör i "score plot". Variabler som ligger nära varandra har en positiv korrelation och befinner sig variablerna på var sin sida om origo är de negativt korrelerade. Exempel ges nedan under analysen.

### 4.2. PLS

Partial Least Square eller Projection to Latent Structures (PLS) används för att kvantifiera korrelationer mellan multivariata X och Y.

Mjukvaran gör först en PCA av Y och sedan av X, därefter söker algoritmerna efter en "bästa möjliga" anpassning av X till Y, inte nödvändigtvis i strikt minstakvadratmening.

"Score plot" och "loading plot" visar för PLS, på motsvarande sätt som i fallet för PCA, en karta över observationer och en över variabler. Det går att välja om alla Y ska analyseras samtidigt, eller om endast ett Y i taget ska analyseras. I den följande analysen har båda sätten använts, dock för mer detaljerad information analyseras varje enskild Y-variabel ( $Y_i$  är kontrollerade utsläpp av varje enskild nuklid (i)).

R<sup>2</sup> visar hur väl framräknade modeller anpassas till data. Förklaringsgraden, Q<sup>2</sup>, anger modellens prediktionsförmåga av nya Y med nya X-värden. Q<sup>2</sup> beräknas genom korsvalidering, som innebär att en eller flera observationer utesluts och en undermodell beräknas med resterande observationer. Undermodellen får ligga till grund för en prediktion av den nyss uteslutna observationen. När alla observationer på detta sätt har varit uteslutna en gång summeras osäkerheterna i prediktionerna och en medelosäkerhet kan beräknas som leder till Q<sup>2</sup> (se SIMCA-handboken). Modeller med värden

på Q2 som är mindre än 5 % anses vara slumpmässiga. Ett högre värde på Q2 medför inte automatiskt att modellen är signifikant.

En normalfördelningsplot visar om residuerna (differensen mellan observerat och predikterat värde) är normalfördelade. Grafen visar sannolikheten för residuerna som funktion av standardavvikelsen (std) per varje observation (se t.ex. Figur 20). Residuerna bör ligga i anslutning till den beräknade regressionslinjen. Om residuerna klart avviker från linjen bör man undersöka om det kan finnas någon bakomliggande förklaring. Observationer med mer än 3-4 std kan misstänkas vara "outliers", som om dessa ingår, resulterar i en modell som inte beskriver "verkligheten" korrekt.

Framräknade regressionskoefficienter för omgivningsvariabler, som används för att beskriva ett utsläpp, visas med ett osäkerhetsintervall som räknas fram ur korsvalideringen (vilket kallas "Jack-Knifing" inom statistiken). Koefficienterna är "scaled and centered". "Scaled" innebär en normering av alla variabler för att kunna jämföra variabler med annars stora skillnader i magnitud. "Centred" innebär att koefficienten är beräknad utifrån ett totalt medelvärde för alla variabler.

CV-ANOVA utgör ett **avgörande test** för att bedöma om den beräknade modellen (M) är signifikant eller bara slumpmässig. Det är naturligtvis bara signifikanta modeller som ska användas för vidare beräkningar. Icke signifikanta modeller kan ändå vara till nytta för att kanske ge ledtrådar till varför en förväntad regressionsmodell inte är signifikant. För definition av CV-ANOVA, se kopia nedan från SIMCA-handboken.

*CV-ANOVA, ANalysis Of VAriance testing of Cross-Validated predictive residuals, is a diagnostic tool for assessing the reliability of PLS, OPLS and O2PLS models introduced in SIMCA-P+ version 12. It is implemented for single-Y and multiple-Y models for the relation  $X \rightarrow Y$ . The diagnostic is based on an ANOVA assessment of the cross-validatory (CV) predictive residuals of a PLS, OPLS or O2PLS model. The advantages of using the CV-residuals are that no extra calculations are needed and that this procedure secures reasonably independent data and variance estimates.*

*Formally, ANOVA is a method to compare two models by the size of their residuals when fitted to the same data. In the regression context, the two models compared are:*

$$y_i = \text{constant} + d_i(1)$$

$$y_i = \text{constant} + bx_i + e_i(2)$$

*The ANOVA is then made on the size of the sum of squares,  $SS(d)$  and  $SS(e)$ , noting that they are not independent since the data underlying them ( $y$ ) are the same. In the current context, this means that we test whether the (PLS/OPLS/O2PLS) model has significantly smaller cross-validated predictive residuals than just the variation around the global average. In summary, the CV-ANOVA provides a significance test (hypothesis test) of the null hypothesis of equal residuals of the two compared models.*

En ytterligare visualisering av en regressionsmodells validitet ges i mjukvaran av "Validate Model". I denna test behålls X-matrisen intakt, men Y permuteras slumpvis, och för varje permutation beräknas modellenpassning, R2, och prediktionsförmåga, Q2. Dessa avsätts på den vertikala axeln, mot korrelationskoefficienten på den horisontella axeln för respektive ursprungligt Y-värde och predikterat Y-värde. Om R2- och Q2-värdena för de slumpmässigt permuterade modellerna är jämförbara eller större än för den ursprungliga modellen betraktas inte den "riktiga" modellen som valid (korrelationskoefficienten för ursprungsmodellen=1 och återfinns längst till höger i grafen). Exempel kan ses i figur 10 (signifikant modell) och figur 18 (icke signifikant modell).

## 4.3. Analysgång

- PCA på hela datamatrixen
- Eventuell förbehandling av data (ex. logaritmisk transformation för att göra data mer normalfördelade)
- Ny PCA på förbehandlade data
- PLS, inkluderande alla variabler. Y-variabler är utsläppen från källorna och X-variabler är alla omgivningsdata. Tanken är att med hjälp av X, finna regressionsmodeller för att kunna prediktera nya utsläpp, Y
- PCA och PLS på utsläpp till luft och till vatten, för att undersöka om dessa korrelerar och om det går att identifiera en "överbärning" av aktivitet från luftutsläpp till vatten
- Klassindelning av data per nuklid ( $y_i$ ) med tillhörande omgivningsdata
- PLS på varje enskilt  $y_i$ , där modellen "trimmas"/förenklas till högt Q2 och till signifikans
  - Normalfördelningsplot (för att se om det finns potentiella "outliers")
  - Plot av regressionskoefficienter med osäkerhetsintervall. Successiv reduktion av icke signifikanta variabler, där små koefficientvärden med stora osäkerheter exkluderas först, inför nästa modellberäkning. Vidare exkluderas negativt korrelerade omgivningsvariabler, då det förefaller orimligt att aktiviteten i omgivningen skulle minska då utsläppen ökar ("trimning")
  - Enklast möjliga modell med så få variabler som krävs för att modellen ska vara signifikant och med högt Q2-värde
- Graf av observerat mot predikerat värde ger en översikt över hur väl modellen i praktiken kan prediktera hur stort ett utsläpp har varit för att leda till mätt aktivitet i omgivningen
- Om modellen är signifikant med 1 omgivningsvariabel, kan modellen enkelt användas till prediktion av aktivitetsnivån i aktuell omgivningsvariabel från givet utsläpp. Det kan förekomma fler signifikanta modeller för respektive nuklid, med olika en-variabelslösningar/omgivningsvariabler, som förmodligen har olika Q2

# 5. Resultat och diskussion

## 5.1. Data

Det allmänna intrycket, under sammanställandet av datamatriken, var att frekvensen "missing data" per omgivningsvariabel, i de flesta fall är hög, oftast över 50 %, men med något undantag bara 18 %. Mätningar av utsläpp av  $^{59}\text{Fe}$  till luft har 85 % "missing data" vilket talar för att det blir svårt att finna någon signifikant korrelationsmodell med tillhörande omgivningsvariabler.

De luftrelaterade omgivningsvariablerna innehåller generellt högre frekvens "missing data" än de för vatten. Mätningarna av luftrelaterade omgivningsvariabler är oftast utförda mellan 1992-2000, vilket motsvarar mindre än 30 % av antalet möjliga observationer och som försvårar modellbyggandet.

Variabler med färre än 6, av totalt 33, observationer ("missing data" > 80 %) har direkt uteslutits för vidare analys. Initialt gjordes undantag för  $^{59}\text{Fe}$ , som analyserades för att se om det skulle gå att få någon indikation om eventuell regressionsmodell, trots den alldeles för höga andelen "missing data".

Det är också viktigt att notera att risken för icke kausala korrelationer ökar med antalet variabler, **risken =  $1-0,95^k$** , där k är antalet variabler. Redan vid  $k=10$  är risken 40 % att det uppträder en slumpartad korrelation.

Variabler i omgivningen som är tagna ur vattnet (t.ex. vattenväxter och fiskar) har använts för att söka korrelationer till vattenutsläppen. På motsvarande sätt har landbaserade provslag (t.ex. mossa och landdjur) använts för att söka korrelationer till utsläppen till luft. Den fortsatta analysen kommer att visa argument för att det är mycket tveksamt att det finns någon signifikant korrelation i det aktuella datasetet mellan utsläppen till luft och vatten, vilket således ger stöd åt en uppdelning av analysen i en för luft och en för vatten. Vidare är antalet variabler för luft relativt få och består som nämnts av en hög andel "missing data". Detta försvårar i hög grad modellbyggandet.

En förklaring till en svag, om ens någon, korrelation mellan luft och vatten är kanske att större delen av utsläppen till luft sprids "åt alla håll" och därför blir mer utspädda och därmed ger ett mindre bidrag till omgivningens vatten. Förekomsten av aktivt material i de luftrelaterade provslagen är också mer beroende av meteorologiska förhållanden och dessutom är förekomst av t.ex. betesvallar och sallad säsongsbunden. Det är viktigt att notera att utsläppen till luft ofta är ungefär 2 tiopotenser lägre än de till vatten, vilket naturligtvis gör att bidraget från luftutsläpp bör vara lägre än bidraget från vattenutsläpp.

Det kan inte uteslutas att utsläpp av radioaktiva nuklider från våra grannländer på andra sidan av Östersjön också är en "felkälla" som försvårar möjligheterna att finna signifikanta korrelations samband för de kontrollerade utsläppen från Forsmark. Speciellt gäller det utsläppen av  $^{137}\text{Cs}$  från Forsmark, vilka blandas med en avklingande mängd från Tjernoby.

Utsläppen till vatten följer oftast samma vattenströmmar och verkar vara mer förutsägbara. Ytterligare omständigheter som kan leda till uteblivna signifikanta korrelationsmodeller kan vara att aktiviteten ligger nära detektionsgränsen. En stor andel "missing data" är inte heller stabiliserande för regressionsmodellerna.

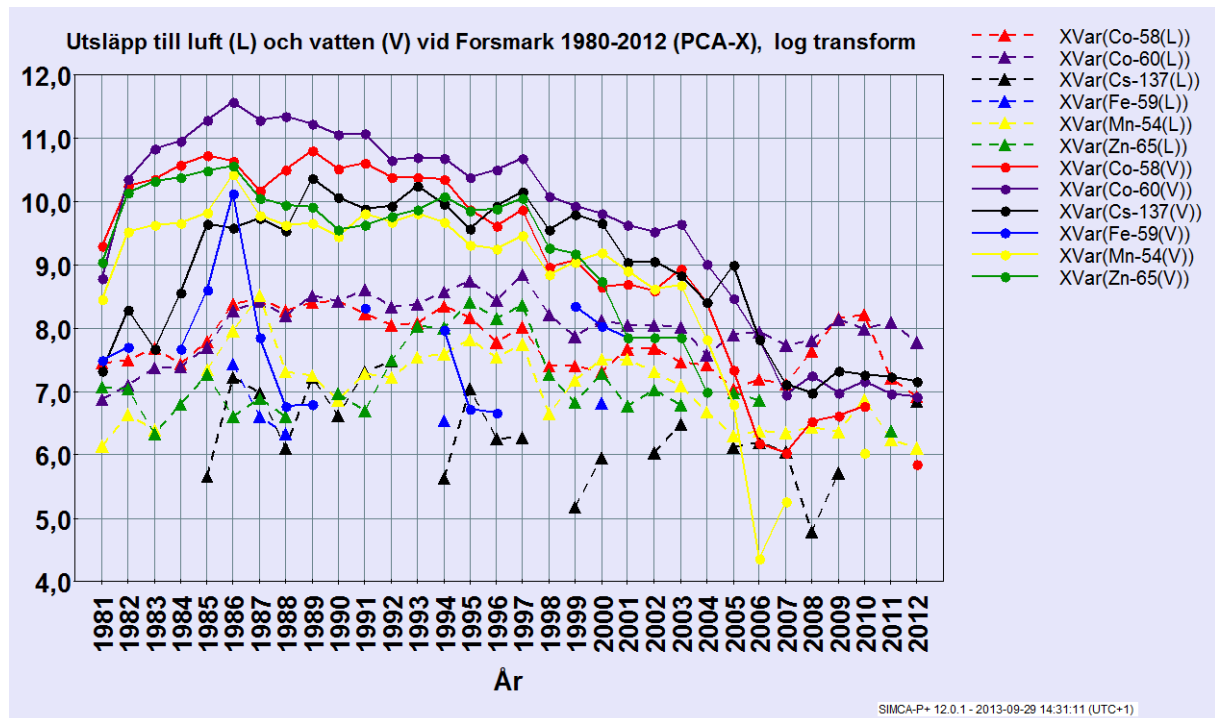
Innan den multivariata analysen startas kan det vara bra att studera hur mätvärdena är fördelade under respektive variabel. I mjukvaran finns enkla sätt att se om variablerna är normalfördelade,



vilket är önskvärt, eller om det finns skevhet, som borde transformeras för att komma närmare en normalfördelning. I den kommande analysen visas exempel på den förbättring av analysresultatet som fås då samtliga mätvärden transformeras med en enkel logaritmisk funktion. Alla analyser har i denna rapport genomförts på log-transformerade data.

Det är också viktigt att förstå att beräknade modeller inte behöver utgöra en "global sanning", vare sig det rör om signifikanta eller för den delen icke signifikanta modeller och variabler, utan samtliga modeller är relaterade till de data som legat till grund för den aktuella analysen och rapporten.

En första överblick på mätvärden (log-skala) av utsläppen till luft och vatten ses nedan i figur 1.



**Figur 1.** Årsmedelvärden för utsläpp till luft (L) och vatten (V). Heldragna linjer och runda prickar svarar mot utsläpp till vatten och streckad linje och trianglar svara mot utsläpp till luft. Observera log-skala.

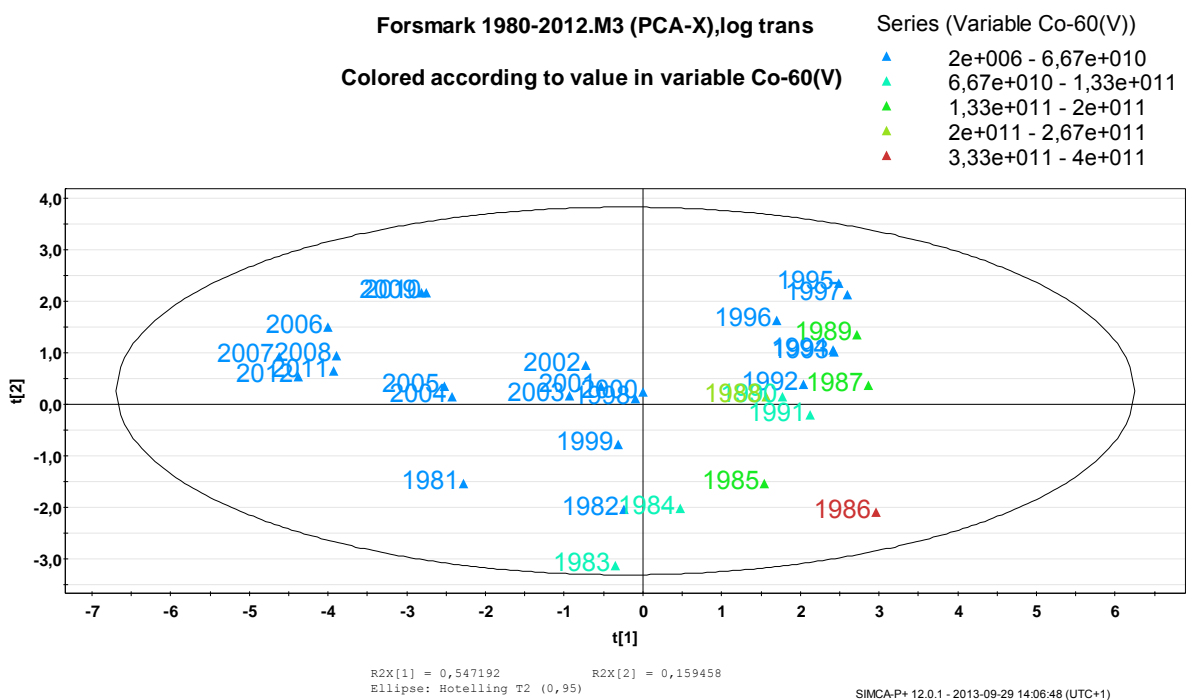
Mätvärdena för utsläpp till vatten, ses i den övre delen av figur 1, och är fram till 2000-talet mellan (2-4) tiopotenser högre än utsläppen till luft. Det ses också en nedåtgående trend i utsläppen, där det är tydligast för utsläppen till vatten. Utsläpp av <sup>59</sup>Fe till vatten (blå prickar) är relativt lågt jämfört med utsläpp av övriga nuklider till vatten, utom runt år 1986 då utsläppet ökar med 2 tiopotenser. För <sup>137</sup>Cs sker en markant ökning av kontrollerade utsläpp till vatten från år 1985, och den utsläppsnivån ligger kvar i mer än 10 år. Utsläppen av de båda Co-isotoperna till vatten har fram till 2012, vilket framgår av figur 1, reducerats med mer än 4 tiopotenser från de högsta mätta värdena i slutet av 1980-talet. Generellt ligger utsläppen till vatten, efter 2005 av de i rapporten studerade nukliderna, under eller i nivå med utsläppen till luft.

## 5.2. PCA – alla data

En PCA-modell på samtliga data (319 och 33 observationer variabler, efter att ett antal variabler exkluderats från de ursprungliga 381) utan någon förbehandling av data resulterade inte i någon signifikant principalkomponent (PC1). Vid kontroll av mätvärdenas fördelning per varje variabel, befanns flertalet variabler uppvisa en logaritmisk fördelning. Därför har data förbehandlats med en logaritmisk transformation, för att få en mer normalfördelad datamängd.

Efter transformationen blev resultatet, med samma data, en modell med 2 signifikanta principalkomponenter och en förklaringsgrad på 17 % (Q2=0,166). Den logaritmiska transformeringen resulterade således i att nästan 17 % av variationerna i data kan förklaras.

Alla följande analyser kommer att baseras på log-transformerade värden.

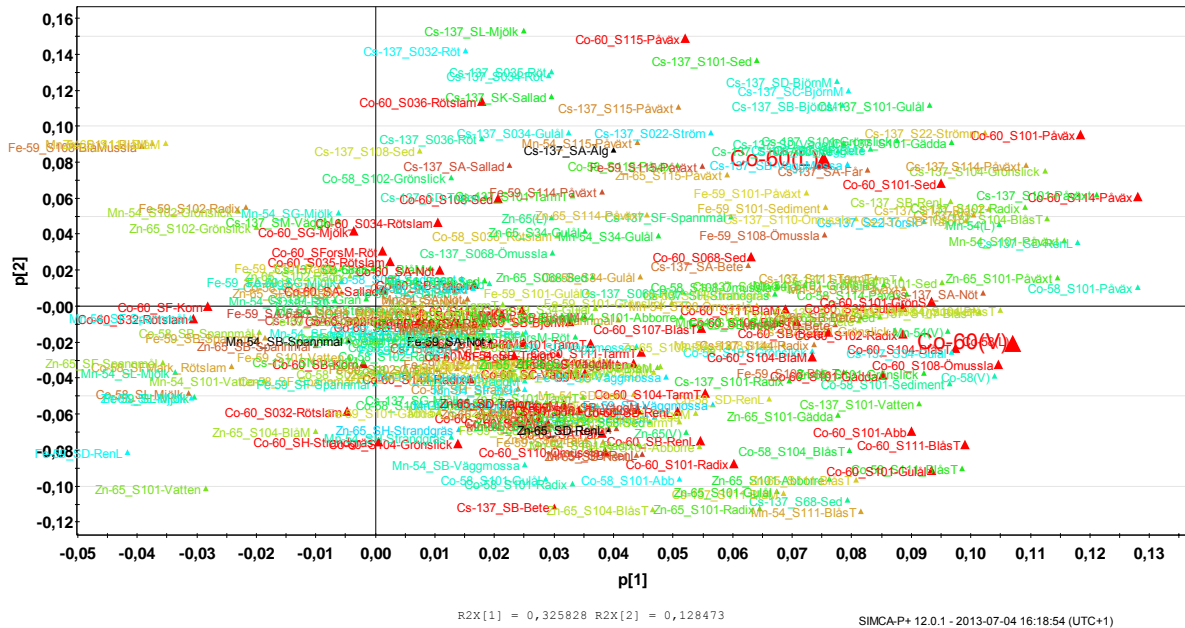


**Figur 2.** "Score plot" av samtliga observationer (år) med hänsyn till samtliga variabler utan uppdelning för luft och vatten. Storleken av utsläpp av, i detta exempel,  $^{60}\text{Co}$  till vatten har kodats med färg.

Denna "score plot", figur 2, visar för att ge några exempel på vad som går att uttolka, i detta initiala analyskedje. År 1986 (röd färg) var utsläppet av  $^{60}\text{Co}$  som högst, vilket också kan ses i figur 1. Efterföljande år minskar utsläppen av denna nuklid, enligt färgkodningen mot mörkblått, och mot slutet av 90-talet fortsätter den nedåtgående utsläppstrenden. Under åren (2006-2012) förändras inte utsläppen i någon större grad, vilket också kan utläsas i figur 2, där dessa observationer är grupperade nära varandra. Gruppering av observationer i "score plot", fås då de underliggande variablerna, här alla mätvärden av utsläpp och omgivningsvariabler, är likartade för observationerna.

I "score plot" visas om det finns likheter/olikheter i utsläpp och mätta aktiviteter per varje år. Årtal som ligger nära varandra har stora likheter och på motsvarande sätt om observationerna är långt ifrån varandra finns olikheter. I tillhörande "loading plot", figur 3, visas hur variablerna förhåller sig till varandra. I denna plot fås en första indikation om variablernas inbördes korrelationer och också en antydning om anledningen till observationernas placering i "score plot".

Forsmark 1980-2012, M3 (PCA-X), log trans  
Colored according to Var ID (Primary)



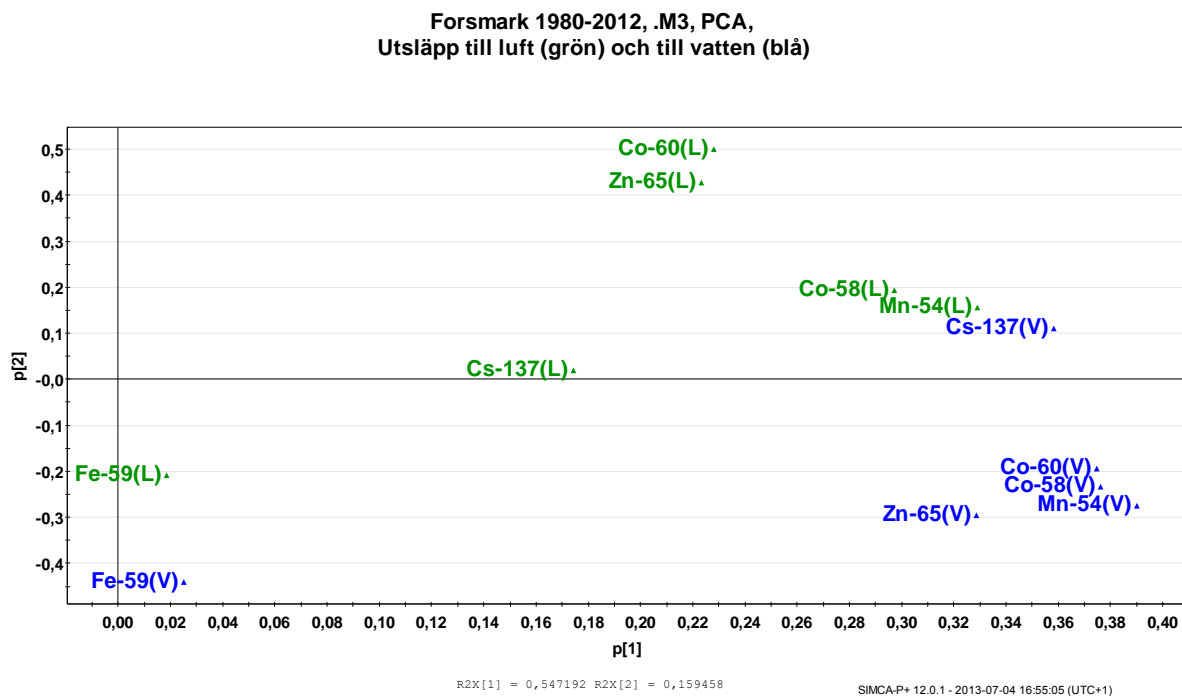
Figur 3. "Loading plot" av samtliga variabler. <sup>60</sup>Co-relaterade variabler är som exempel markerade med röd färg och utsläppen har markerats med större trianglar.

Figur 3 är alldeles för kompakt och tät för att ge någon detaljerad information, nästan det enda som går att tolka är att det finns ett stort antal korrelationer mellan variablerna (variabler som ligger nära varandra). I denna figur har alla <sup>60</sup>Co-relaterade variabler markerats med rött och utsläppen av <sup>60</sup>Co har markerats med lite större trianglar. Generellt gäller att observationerna i t.ex. 4:e kvadranten i "score plot", figur 2, ligger där de gör på grund av att de är positivt korrelerade med variablerna i 4:e kvadranten i figur 3, "loading plot". Således går det att till exempel få en antydning om, att observationen 1986, figur 2, är förknippad med utsläppet av <sup>60</sup>Co till vatten i 4:e kvadranten i figur 3. Det går också exempelvis i figur 3, i 4:e kvadranten, att få en indikation om en korrelation mellan utsläpp till vatten av <sup>60</sup>Co(V) och aktiviteten av Co-60\_S101-Abborre och Co-60\_S111-Blåstång, då dessa variabler ligger relativt nära varandra. Dessa korrelationer kommer, längre fram i analysen, att visa sig vara signifikanta och betydelsefulla.

Notera att den underliggande modellen, M3, bara har en förklaringsgrad på mindre än 17 % (Q2=0,166), samt att data inte är uppdelade på luft- respektive vattenutsläpp, så några långtgående slutsatser kan ännu inte dras, men vissa indikationer kan ändå observeras.

## 5.3. PCA – utsläppsdata

Innan PLS-analysen genomförs, för att undersöka korrelation mellan utsläpp (Y) och omgivningsdata (X) per nuklid, kontrolleras om och hur Y-variablerna är korrelerade. Frågeställningen är: Finns det någon koppling mellan luft- och vattenutsläpp, så att t.ex. hög aktivitet i utsläpp till luft också medför hög aktivitet i vattenutsläppet? Om ett signifikant samband kan beräknas, skulle det i princip vara tillräckligt att bara kontrollera det utsläpp som är mest lättillgängligt, och sedan uppskatta övriga med hjälp av korrelationsmodellen från PLS.



**Figur 4** "Score plot" av enbart utsläppen till luft (gröna) och vatten (blå).

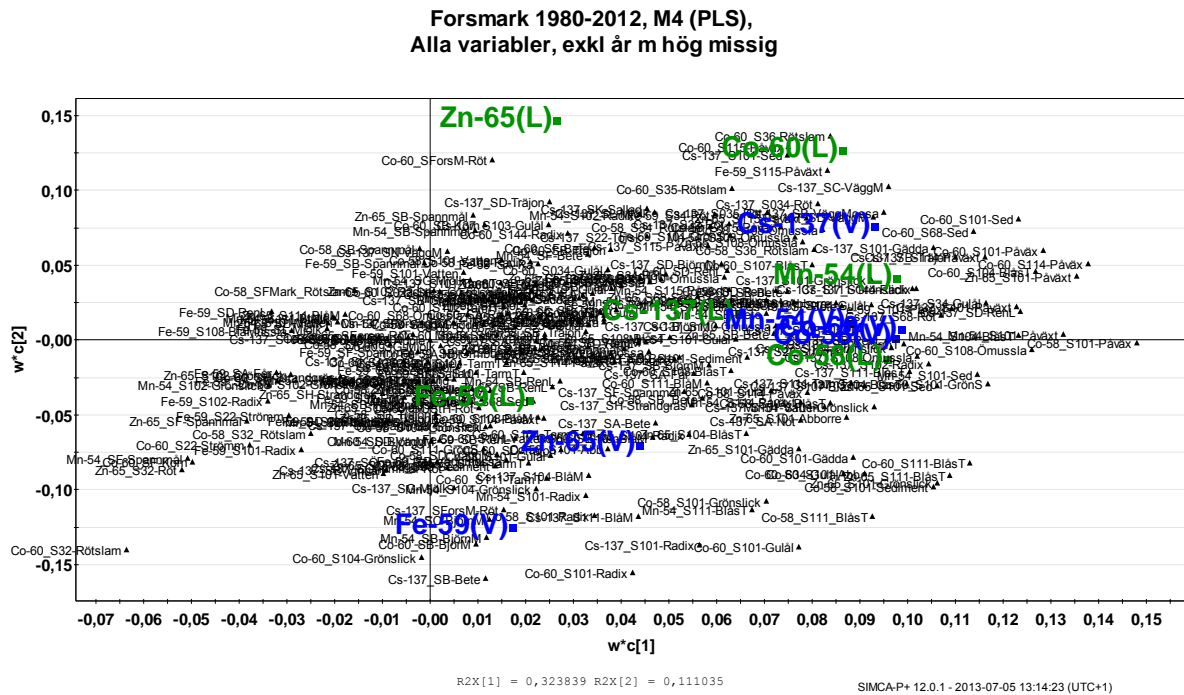
En indikation om korrelationsmönstret ses i figur 4. Det är en PCA som är gjord på enbart Y-variablerna, där andelen "missing data" är mycket låg, undantaget utsläppen av  $^{59}\text{Fe}$  till luft och till vatten med 85 % respektive 55 % "missing data". Det är viktigt att komma ihåg att korrelationsmönstret, i figur 4 för modell M3, speglar årsmedelvärdena av utsläppen till luft och vatten utan hänsyn taget till omgivningsdata. Om det finns en koppling mellan de olika utsläppen borde observationerna ligga nära varandra. Modell M3 har förklaringsgrad på knappt 17 % ( $Q^2 = 0,166$ ) längs den horisontella axeln (PC1) och längs den vertikala axeln (PC2) är  $Q^2$  praktiskt taget 0. Detta betyder att man kan bortse från eventuella skillnader i vertikalled.

Korrelationen mellan utsläppen till luft (grön) och vatten (blå) är mycket svag (luft och vatten ligger långt ifrån varandra för respektive nuklid i figur 4), däremot korrelerar utsläppen till vatten relativt väl med varandra och är samlade. För  $^{54}\text{Mn}$  och  $^{58}\text{Co}$  verkar det finnas en svag koppling mellan utsläppen till vatten och luft. Exempelvis ligger utsläppen till luft och vatten för  $^{60}\text{Co}$  längre från varandra i denna plot, vilket pekar på en lägre grad av korrelation.

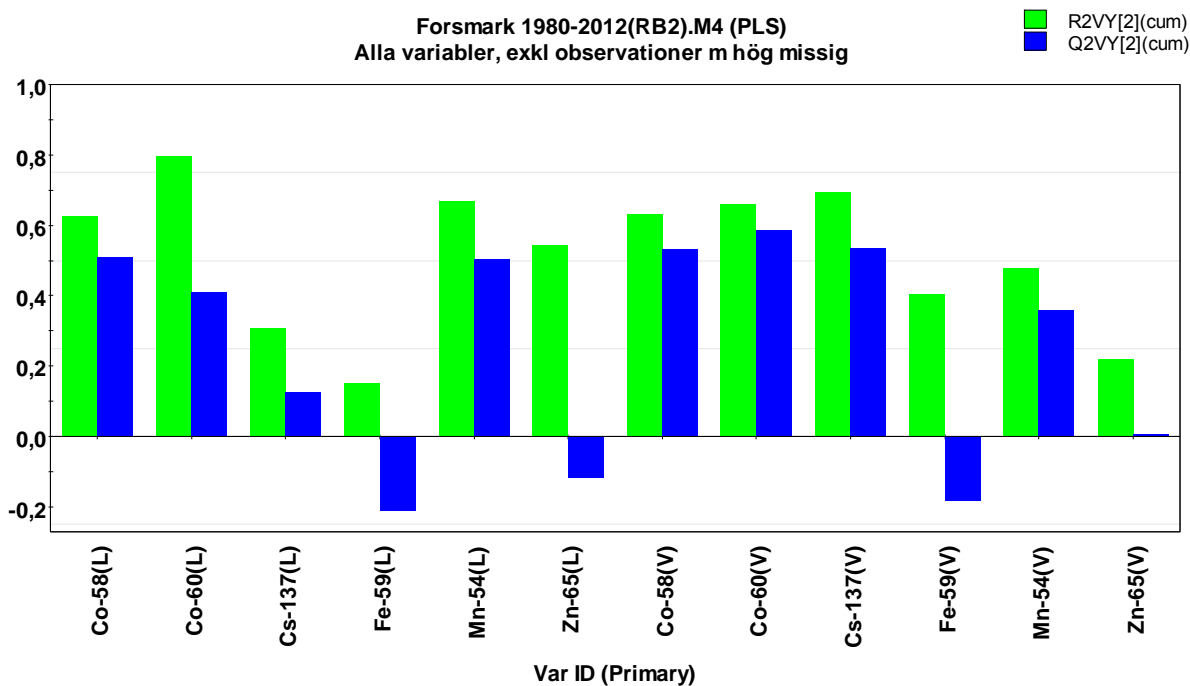
En avvikelse ses i figur 4, för  $^{59}\text{Fe}$ , där andelen "missing data" för både vatten- och luftutsläppen är mycket större än för övriga nuklider och man kan konstatera att det inte verkar finnas någon användbar information i mätdata för denna nuklid då "score"-värdena är mycket nära origo.

## 5.4. PLS – alla data

I figur 5 nedan visas en "loading plot" (variabelkarta) där samtliga utsläpp behandlats som Y-variabler och samtliga omgivningsvariabler som X-variabler. Modell M4 är inte signifikant, men mönstret av Y påminner om det i figur 4. Beräkningsalgoritmerna för regressionsmodellerna söker efter "bästa möjliga anpassning av X till Y" och då kommer en del variabler att, mot bättre vetande, uppenbarligen bli negativt korrelerade till utsläppen. Det är inte rimligt att ökade utsläpp leder till att den uppmätta aktiviteten i vissa omgivningsvariabler sjunker!



**Figur 5.** "Loading plot" där samtliga variabler har använts. Blå=utsläpp till vatten, grön=utsläpp till luft och svart=omgivningsvariabler.



**Figur 6.** En översikt av sambanden mellan X och Y för modell M4. Gröna staplar visar hur väl data kan anpassas (R2) och blå staplar visar prediktionsförmågan (Q2).

Även om modell M4 inte är signifikant så finns en indikation om den förväntade prediktionsförmågan för de enskilda nukliderna. Q2 är alltid lägre än R2 men kan inte vara negativ. Således förväntas inte <sup>59</sup>Fe, varken till luft- eller vattenrelaterade variabler, kunna ingå i signifikanta regressionsmodeller. Det bör också påpekas att antalet variabler fortfarande är stort, varför icke kausala korrelationer sannolikt finns med. För modell M4, är analysen ännu inte uppdelad mellan luft och vattenvariabler, vilket kan vara anledning till att några av sambanden för luft skenbart, av ovan nämnda anledning, ser ut att vara bra.

Den fortsatta analysen kommer att delas upp mellan utsläpp till luft respektive vatten för varje nuklid. Vidare kommer modellerna att trimmas till att om möjligt bli signifikanta och enbart baseras på signifikanta och positivt korrelerade variabler.

## 5.5. PLS – utsläpp till luft respektive vatten

Data har, förutom separation av luft- och vattenvariabler, klassindelats så att varje nuklid analyseras separat för att ge ett mer lättolkat resultat.

För att kvantifiera korrelationerna mellan utsläppt och mätt aktivitet i olika provslag vid olika mätstationer har PLS använts. Vanligtvis betraktas utsläppen som grunden till att aktivitet kan mätas i de olika provslagen i omgivningen och skulle därmed betraktas som X-variabler. I denna analys har i stället aktiviteten i provslagen vid de olika mätstationerna hanterats som X-variabler för att via regressionsanalys kunna prediktera utsläppen (Y). Fortsatt analys, med uppdelning av utsläpp till luft respektive vatten samt klassindelning för respektive nuklid och exkludering av icke signifikanta variabler, får visa om och hur prediktionsförmågan för modellerna kan förbättras. Modellerna måste också vara signifikanta enligt CV-ANOVA-testet.

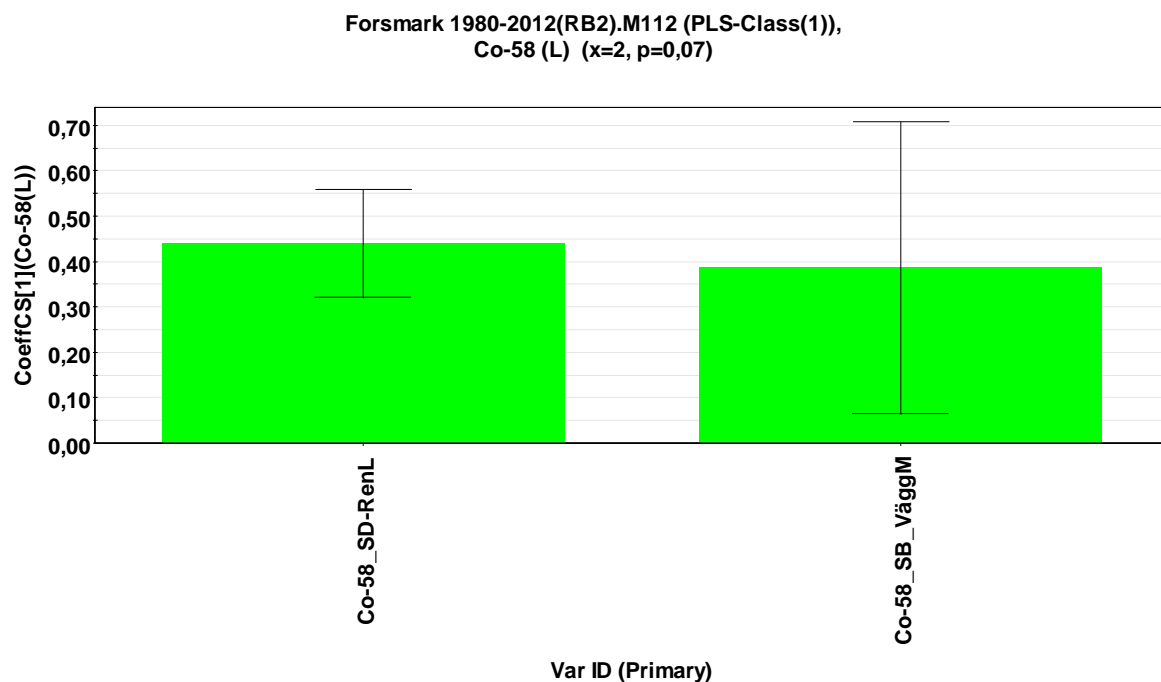
## 5.6. PLS – per nuklid

PLS på klassindelade data per varje nuklid med relaterade variabler för luft respektive vatten.

### 5.6.1. $^{58}\text{Co}$

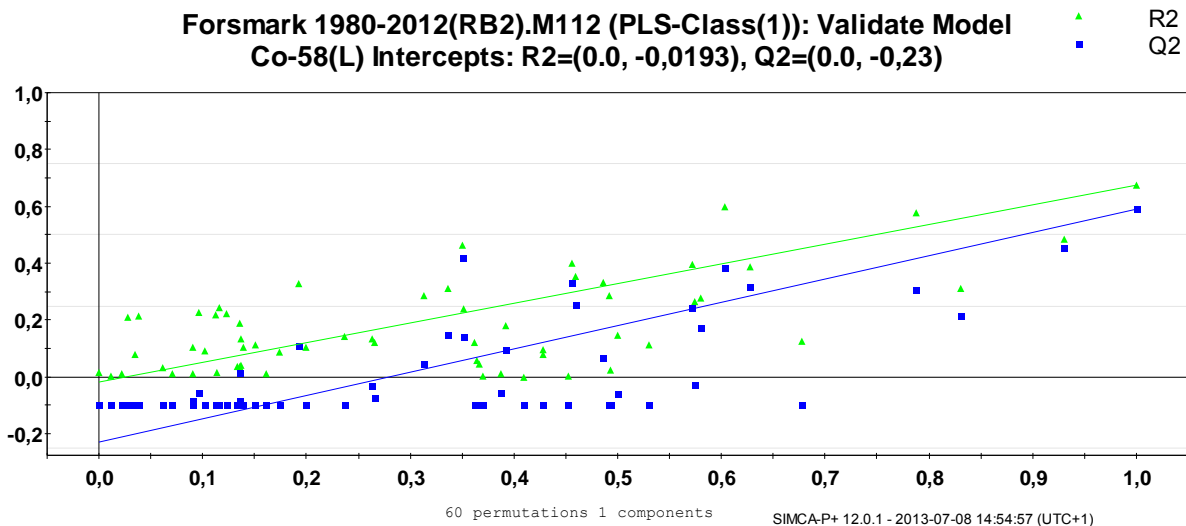
#### Utsläpp till luft

Det går inte att finna strikt signifikanta modeller, enligt CV-Anova-testet, som relaterar utsläpp av  $^{58}\text{Co}$  till luft med luftrelaterade omgivningsvariabler. Om en modell är signifikant så är  $p \leq 0,05$ , om den inte är strikt signifikant så ligger  $p$  i intervallet  $0,05-0,1$  och om den är icke signifikant så är  $p > 0,1$ . Här visas modell M112 i figur 7, för utsläpp av  $^{58}\text{Co}$  till luft, med 2 signifikanta variabler.



**Figur 7.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{58}\text{Co}$  till luft med 2 signifikanta variabler för modell M112, som inte är strikt signifikant ( $p=0,069$ ). Osäkerhetsintervallen är markerade i de gröna regressionskoefficienterna.

Den erhållna modellen (M112) baserad på Renlav och Väggmossa är inte strikt signifikant då CV-Anova-testet ger  $p=0,069$ . Nedan i figur 8 ges en antydning om varför M112 inte är strikt signifikant. Denna plot (kallad "Validate Model" i mjukvaran) är framtagen genom att slumpvis permutera Y-värdena och behålla X-matrisen intakt. För varje ny permutation av Y-vektorn avsätts beräknade  $R^2$  och  $Q^2$ , på den vertikala axeln, mot korrelationskoefficienterna (horisontella axeln) mellan ursprungliga Y-värden och permuterade Y-värden. Det framgår att det finns några av de 60 slumpvisa permutationerna som resulterar i nästan samma  $R^2$ - och  $Q^2$ -värden som de för den "riktiga" Y-vektorn med korrelationskoefficient=1 (längst till höger i figur 8), och därmed kan inte modellen betraktas som valid och robust.

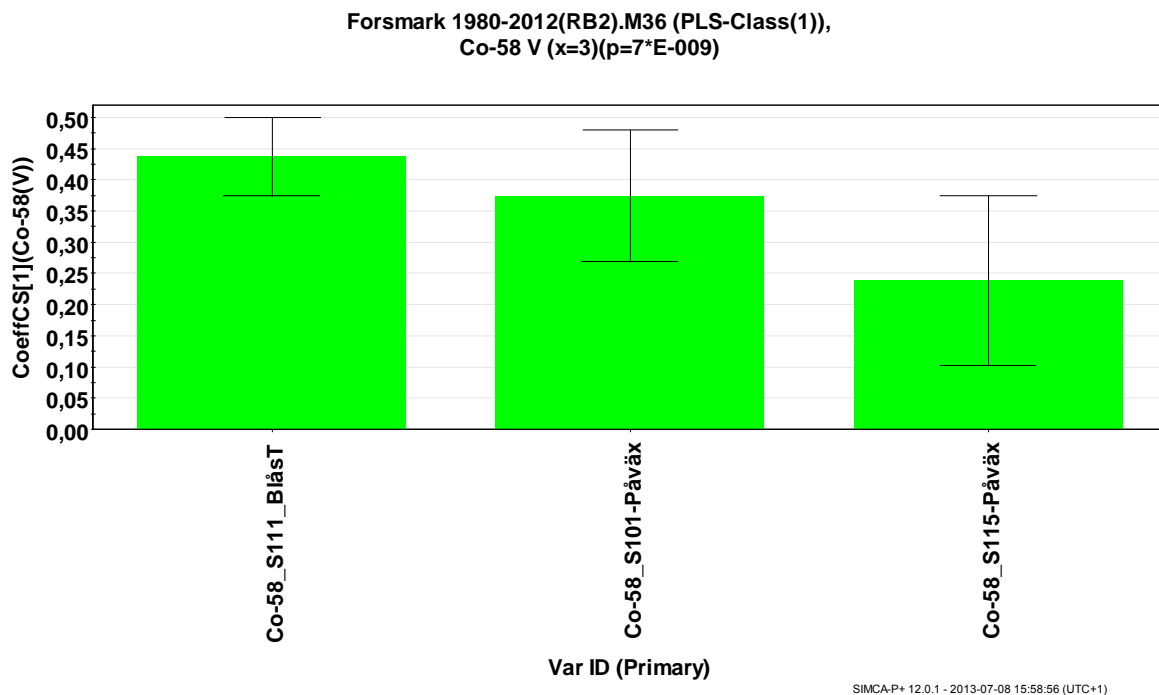


**Figur 8.** R2 och Q2 (vertikala axeln) avsatta för 60 slumpvisa permutationer av Y-värdena mot korrelationskoefficienter (horisontella axeln) mellan ursprungligt Y och permuterat Y.

Ytterligare förenkling av modellen till 1 omgivningsvariabel ger heller ingen strikt signifikant modell.

### Utsläpp till vatten

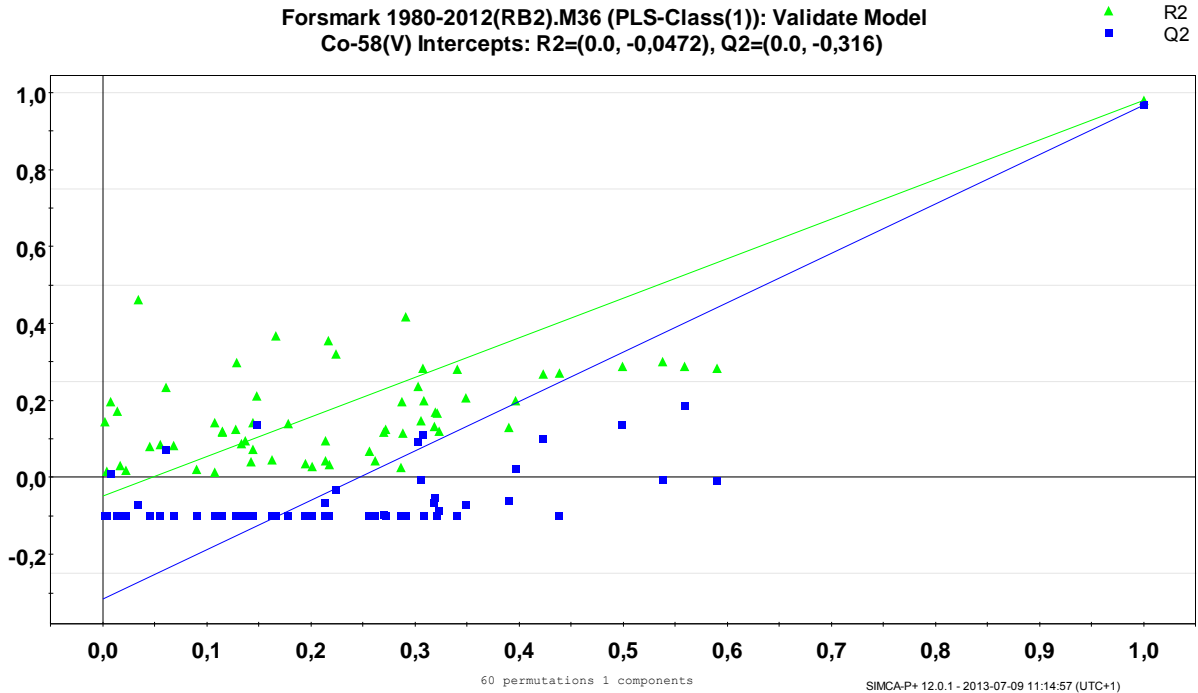
En signifikant modell, M36 ( $p=7 \cdot 10^{-9}$ ), med tre signifikanta variabler erhöles med en mycket hög prediktionsförmåga ( $Q2=0,967$ ). Detta betyder att variationer i normala utsläpp till vatten av  $^{58}\text{Co}$ , med de i figuren nedan visade omgivningsproverna, kan predikteras till c:a 97 %.



**Figur 9.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{58}\text{Co}$  till vatten med 3 signifikanta variabler för modell M36, som är signifikant.

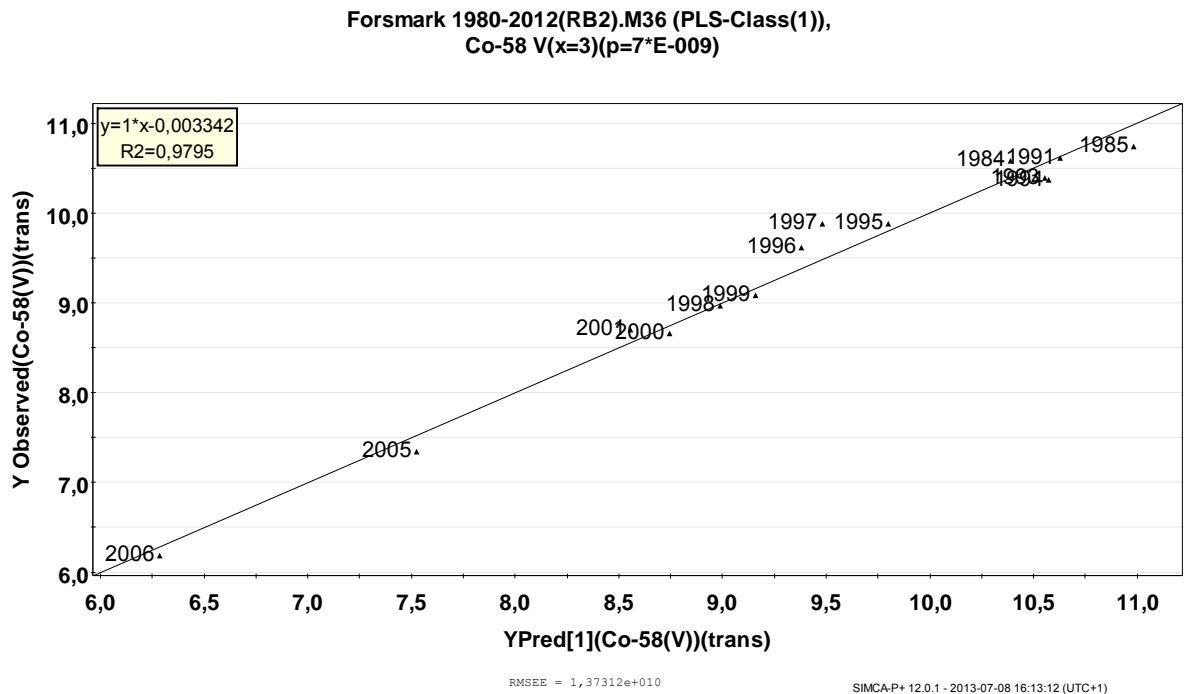
Jämför plot "validate model" i figur 10 nedan, för den signifikanta modellen M36, med den i figur 8 för den icke strikt signifikanta modellen M112. I figur 10 syns det att det inte finns någon konkurrerande, slumpvis permuterad, modell som kommer i närheten av R2 och Q2 som tillhör den ursprungliga valida modellen med korrelationskoefficient=1 (längst till höger i figur 10).





**Figur 10.** R2 och Q2 (vertikala axeln) avsatta för 60 slumpvisa permutationer av Y-värdena mot korrelationskoefficienter (horisontella axeln) mellan ursprunglig Y och permuterat Y.

Nedan i figur 11 är observerat utsläpp av <sup>58</sup>Co till vatten, avsatt mot predikerat utsläpp. Överensstämmelsen är extremt hög över nästan 5 tiopotenser i utsläpp! Observera att logaritmisk skala används i figuren.



**Figur 11.** Observerat mot predikerat utsläpp av <sup>58</sup>Co till vatten. Prediktionen är gjord med modell M36 med tre variabler. Observera log-skala.

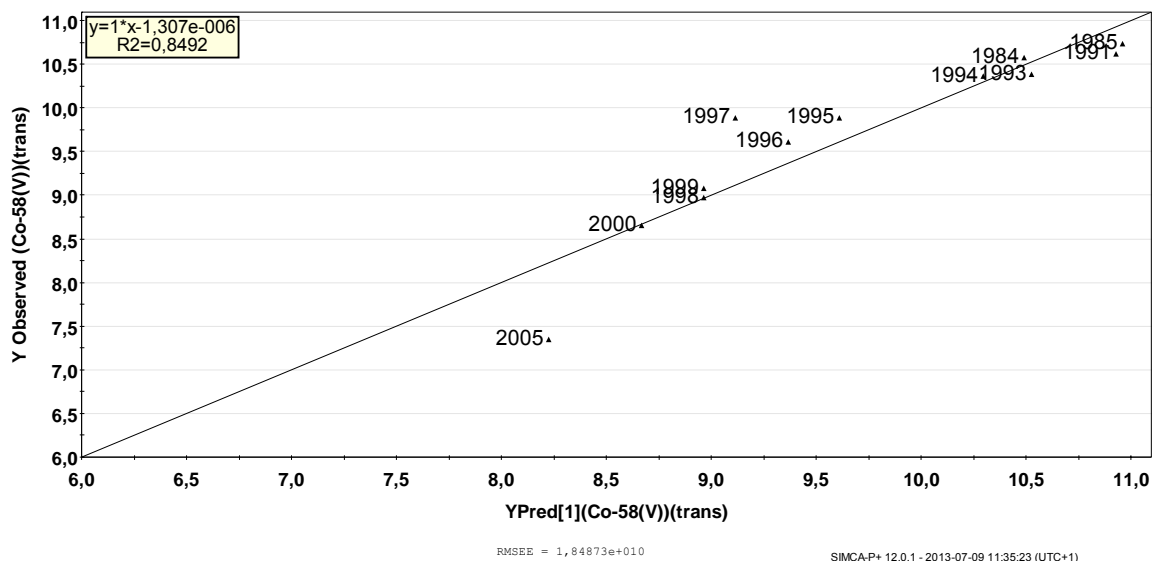
En ytterligare förenkling av modellen, M36, kan göras till endast 1 omgivningsvariabel. Det framgår av figur 9, att Blåstång vid mätstation 111 har den största regressionskoefficienten och det minsta

osäkerhetsintervallet, varför denna variabel primärt valdes till en modell. Prediktionsförmågan för modell M115,  $Q2=0,786$ , är trots bara en omgivningsvariabel väldigt hög.  $Q2$  har visserligen minskat från c:a 96 % med tre variabler, till c:a 79 % med endast en variabel. En ännu bättre modell kan fås, M176, där  $p=3,6 \cdot 10^{-5}$  och  $Q2=0,897$ , med  $^{58}\text{Co\_S101-Påväxt}$  som X-variabel. Den senare variabeln har något lägre frekvens med "missing data".

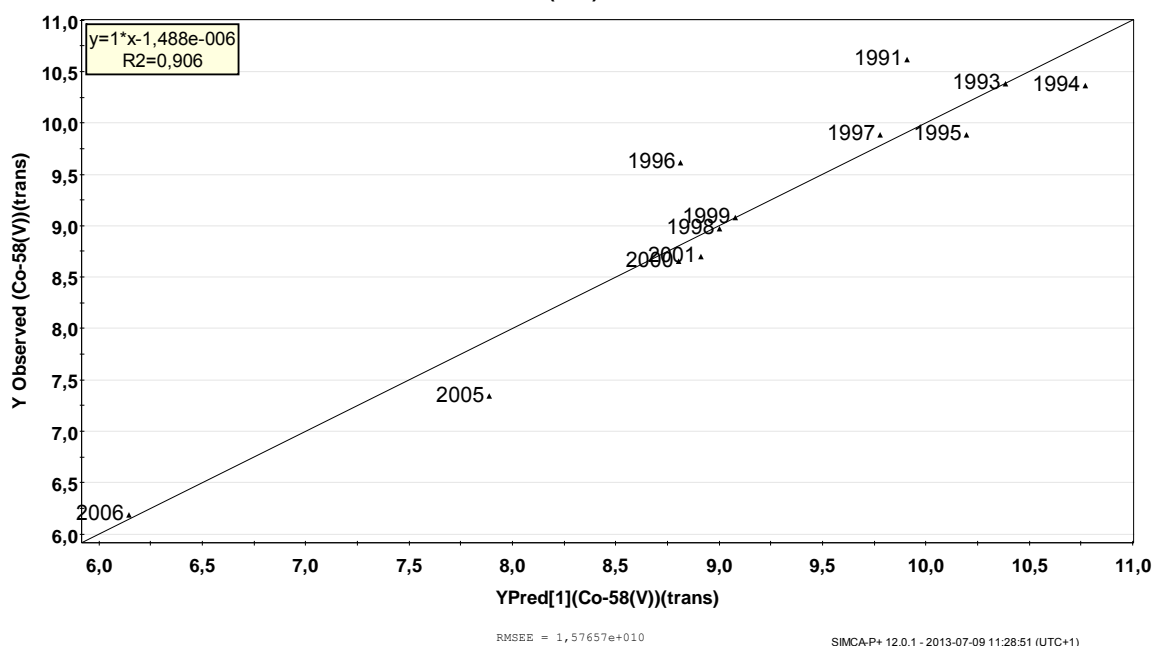
Den observante noterar, i figur 12a nedan, jämfört med i figur 11, att observationer senare än 2005 inte finns med, då det tyvärr saknas omgivningsmätningar efter 2005, för blåstång vid station 111. Figur 12b visar observerat mot predikterat, när variabeln S101-Påväxt har använts för prediktionen och där finns observationen 2006 representerad.

Den senare modellen, M176, kan "i normalfallet" prediktera variationer i utsläppen av  $^{58}\text{Co}$  till vatten till nästan 90 %, vilket är c:a 11 procentenheter bättre än för modell M115. Se tabell 1 nedan för en jämförelse mellan modellerna.

Forsmark 1980-2012(RB2).M115 (PLS-Class(1)),  
Co-58 V (x=S111-Blåstång, p=0,001, Q2=0,786)  
(12a)



Forsmark 1980-2012(RB2).M176 (PLS-Class(1)),  
Co-58(V) (x=S101-Påväxt, p=3,6\*E-5, Q2=0,897)  
(12b)



**Figur 12a och b.** Observerat mot predikterat utsläpp av <sup>58</sup>Co till vatten. Prediktionen är gjord med modell M115 (12a) med x-variabel \_S111-Blåstång och Modell M176 (12b) med x-variabel \_S101-Påväxt. Observera log-skala.

**Tabell 1** Sammanställning av prediktionsförmåga, Q2, och modellsignifikans, p, för 2 en-variabelmodeller för utsläpp av <sup>58</sup>Co till vatten.

Modell	Mätstation	Provslag	Q2	p
M115	S111	Blåstång	0,786	$9,8 \cdot 10^{-4}$
M176	S101	Påväxt	0,897	$3,6 \cdot 10^{-5}$

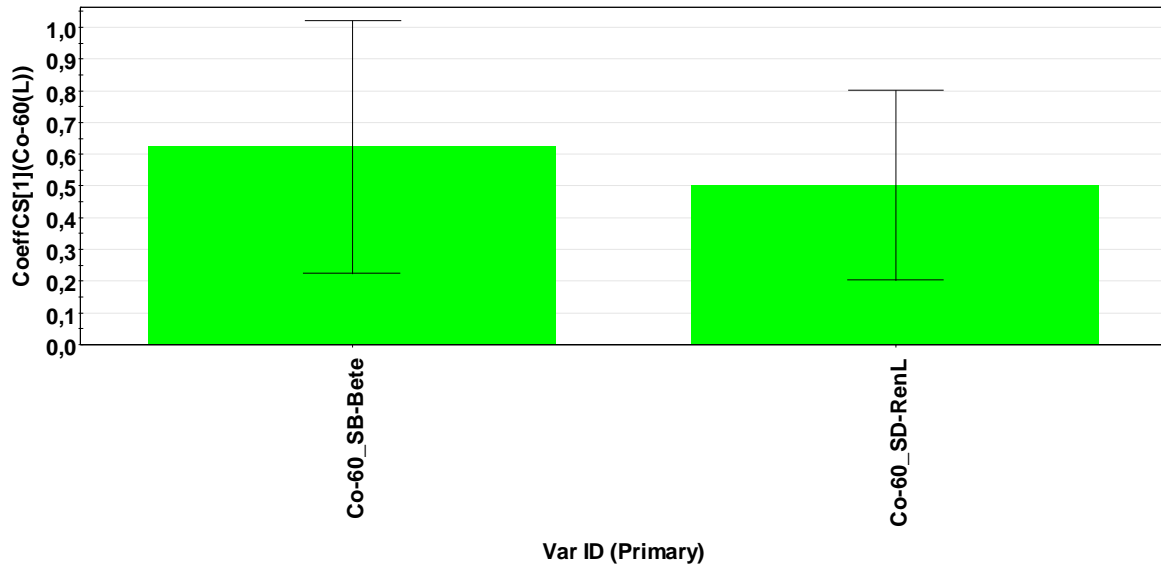
Alla möjliga en-variabelmodeller är inte testade, men de här redovisade modellerna var de som "föll ut" efter förenklingen som finns beskriven under "Analysgång".

## 5.6.2. $^{60}\text{Co}$

### Utsläpp till luft

Det går att finna en signifikant regressionsmodell, M122, med 2 signifikanta variabler, som relaterar utsläppen till luft med tillhörande luftvariabler. Prediktionsförmågan är relativt hög,  $Q^2=0,862$ .

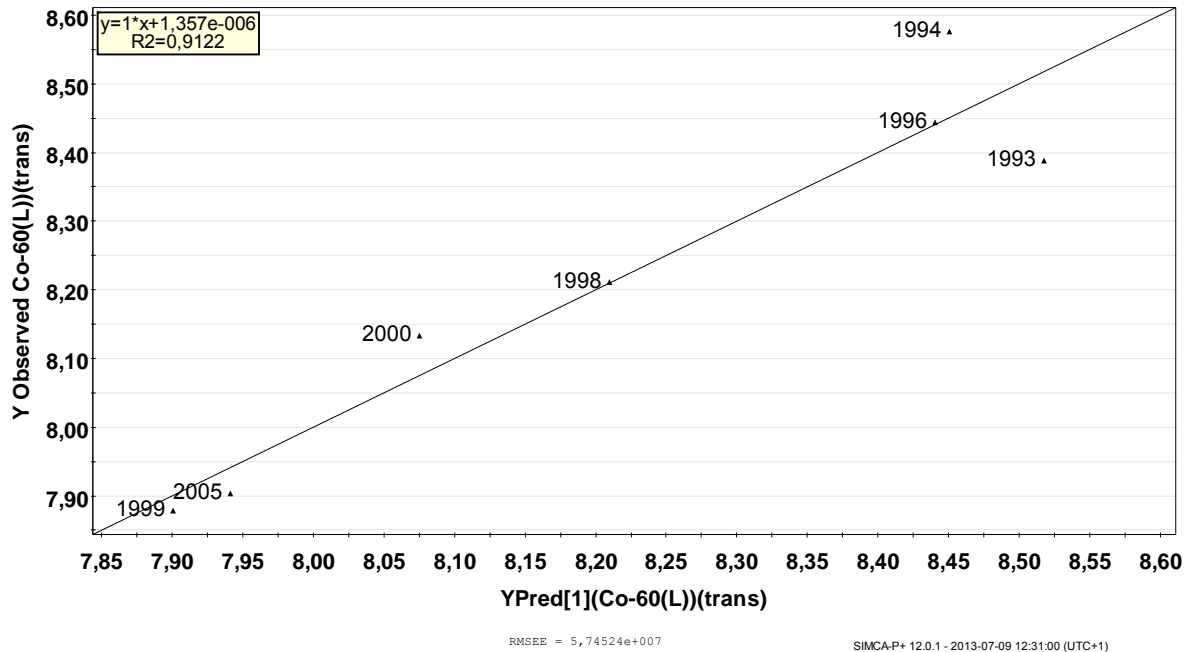
Forsmark 1980-2012(RB2).M122 (PLS-Class(2)),  
Co-60(L),  $x=2$ ,  $p=0,019$ ,  $Q^2=0,862$



**Figur 13.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{60}\text{Co}$  till luft med 2 signifikanta variabler för den signifikanta modellen M122.

Det framgår i figur 14 att det är väldigt få observationer för dessa 2 variabler. De enda luftrelaterade variablerna, för utsläpp av  $^{60}\text{Co}$  till luft, där det finns mätningar redovisade efter 2005, är röt slam vid stationerna 34, 35 och 36. Dessvärre bidrager dessa variabler inte med någon användbar information till regressionsmodellerna. Ingen luftvariabel för  $^{60}\text{Co(L)}$  hade lägre frekvens "missing data" än 64 %, vilket minskar sannolikheten att finna någon valid regressionsmodell.

Forsmark 1980-2012(RB2).M122 (PLS-Class(2)),  
Co-60 L, x=2, p=0,019, Q2=0,862



**Figur 14.** Observerat mot predikerat utsläpp av  $^{60}\text{Co}$  till luft. Prediktionen är gjord med modell M122. Observera log-skala.

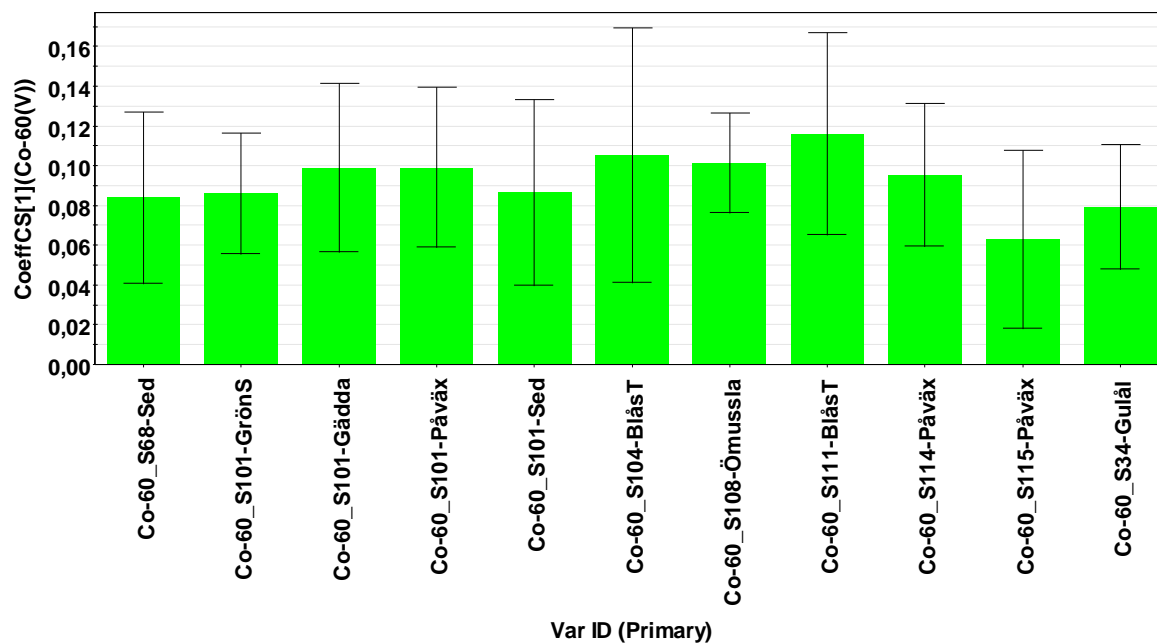
Ingen av de två variablerna i figur 13, ger någon strikt signifikant en-variabelmodell, vilket troligen kan förklaras av att antalet observationer är för få.

### Utsläpp till vatten

Mätningarna av  $^{60}\text{Co}$  i omgivningsvariablerna har, för den aktuella studien, en relativt hög frekvens med i bästa fall bara 52 % "missing data". En signifikant modell, M50, med 11 omgivningsvariabler,  $p=3,2 \cdot 10^{-9}$  och  $Q2=0,859$ , kan användas till prediktion av gjorda utsläpp av  $^{60}\text{Co}$  till vatten, och det förväntade resultatet borde nås till c:a 86 %. Regressionskoefficienterna för modell M50 visas i figur 15.

Modell M50 med 11 signifikanta omgivningsvariabler kommer att vara resurskrävande när den ska användas till prediktion av utsläpp av  $^{60}\text{Co}$  till vatten.

Forsmark 1980-2012(RB2).M50 (PLS-Class(2)), Co-60 V  
(x=11, p=3e-9, Q2=0,859)

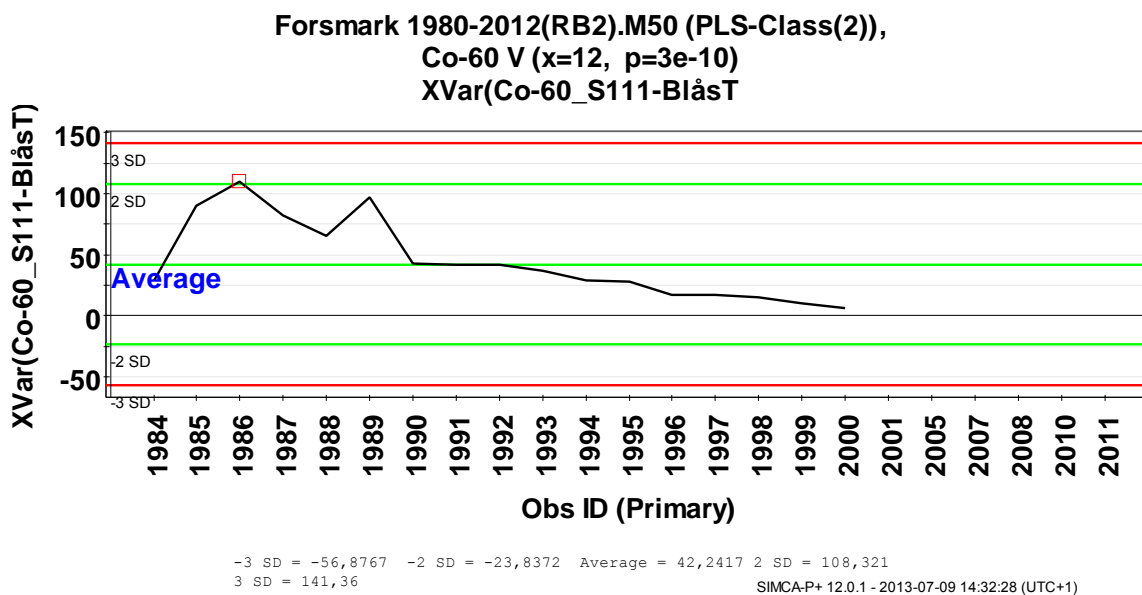
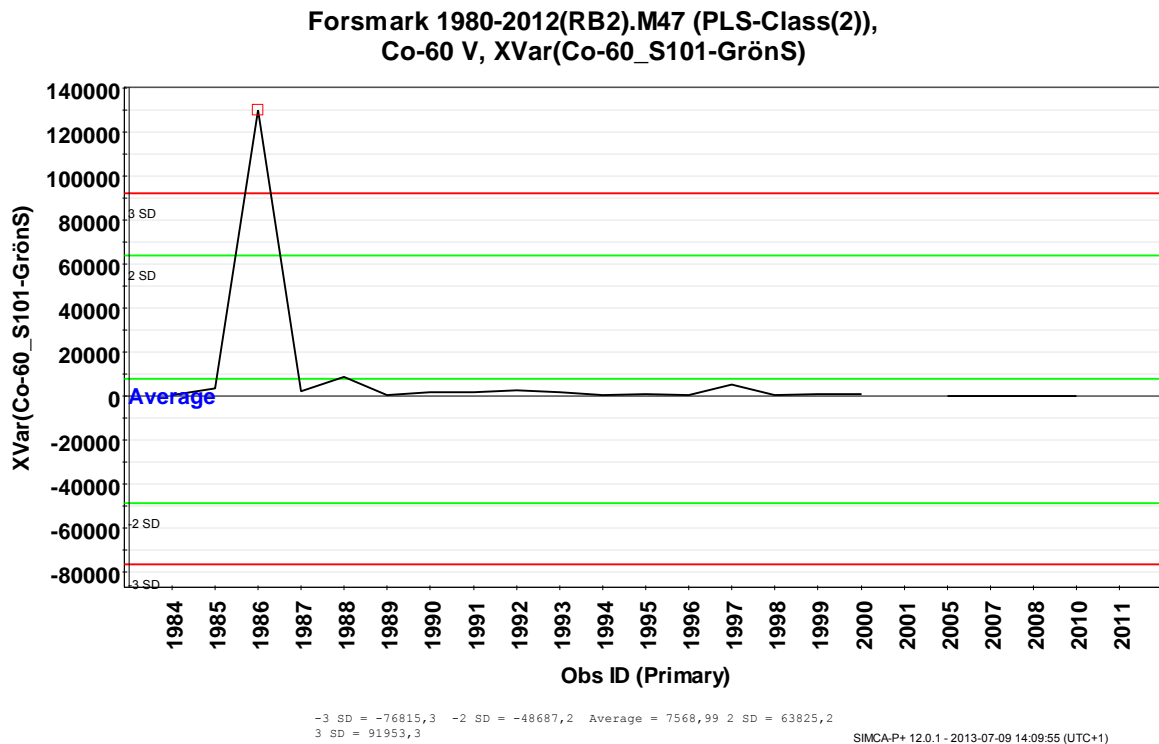


SIMCA-P+ 12.0.1 - 2013-07-09 14:58:03 (UTC+1)

**Figur 15.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av <sup>60</sup>Co till vatten med 11 signifikanta variabler för den signifikanta modellen M50.

Exempel på några intresanata observanda för ett par av variabler ses nedan i figurerna 16a och 16b. Mätvärdena är här avsatt mot respektive år och det visas också hur många standardavvikelser mätvärdena ligger i förhållande till medelvärdet. I den övre ploten, Co-60\_S101-Grönslick ses ett extremvärde 1986.

I figur 16b visas förhöjda värden för Co-60\_S111-Blåstång år 1986 till 1989, dock inte extrema



**Figur 16a och b.** Mätvärdena är här avsatt mot respektive år och det visas också hur många standardavvikelse mätvärdena ligger i förhållande till medelvärdet.

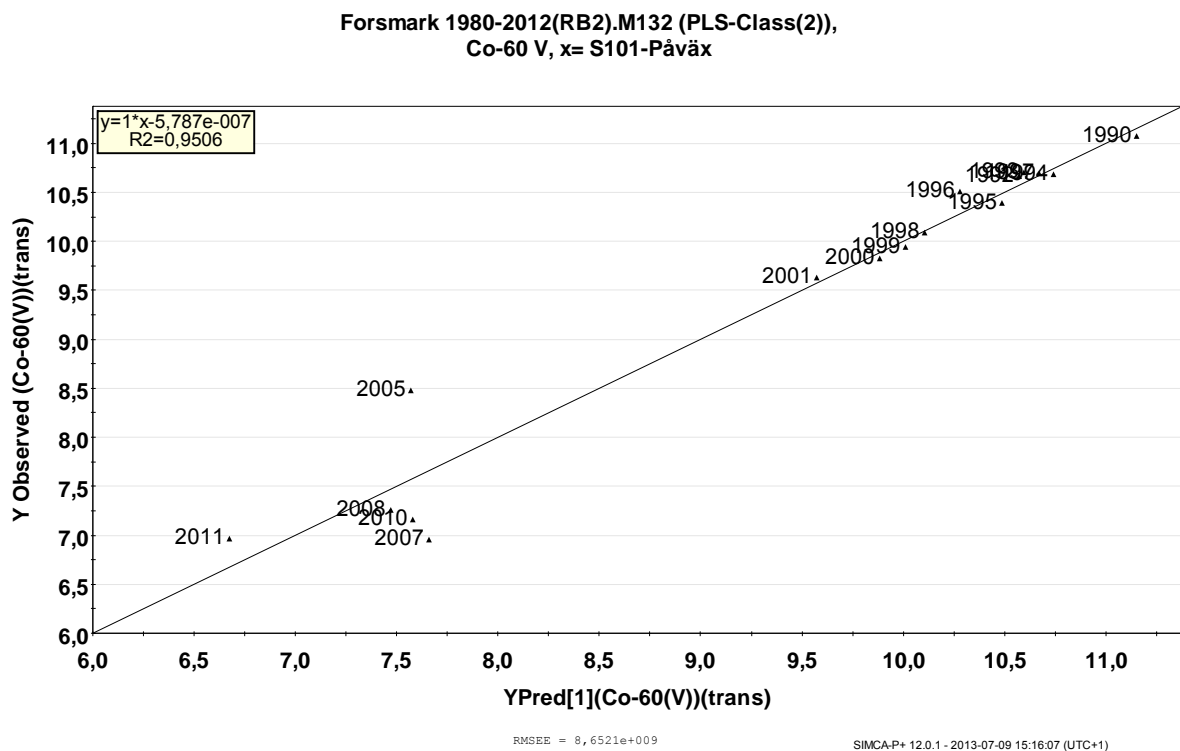
Det går också att identifiera några signifikanta en-variabelslösningar för prediktion av utsläppen av <sup>60</sup>Co till vatten, vilka har liknande Q2 eller till och med något högre Q2 än för modell M50. Vilken modell man än väljer kommer denna att vara betydligt mindre resurskrävande än med modell M50. Modellerna M132 och M134 har dessutom bättre prediktionsförmåga än den för M50!

Nedan i tabellen redovisas tre signifikanta modeller, med mycket höga Q2. Det bör också påpekas att alla tänkbara variabler inte är testade, men att de här valda har högst frekvens av mätvärden.

**Tabell 2.** Sammanställning av prediktionsförmåga, Q2, och modellsignifikans, p, för 3 en-variabelmodeller för utsläpp av <sup>60</sup>Co till vatten.

Modell	Station	Provslag	Q2	p
M132	101	Påväxtprov	0,936	$1,7 \cdot 10^{-8}$
M134	111	Blåstång	0,908	$1,8 \cdot 10^{-7}$
M131	114	Påväxtprov	0,828	$6,3 \cdot 10^{-5}$

I normalfallet kan det således vara tillräckligt att kontrollera en av de i tabell 2 redovisade omgivningsvariablerna för att kunna prediktera storleken på utsläpp av <sup>60</sup>Co till vatten, med enligt tabell 2 förväntad mycket hög precision. I figur 17 visas ett exempel för modell M132, på prediktionsförmågan. Observera att prediktionerna spänner över mer än 4 tiopotenser.



**Figur 17.** Observerat mot predikerat utsläpp av <sup>60</sup>Co till vatten. Prediktionen är gjord med modell M132. Observera log-skala.

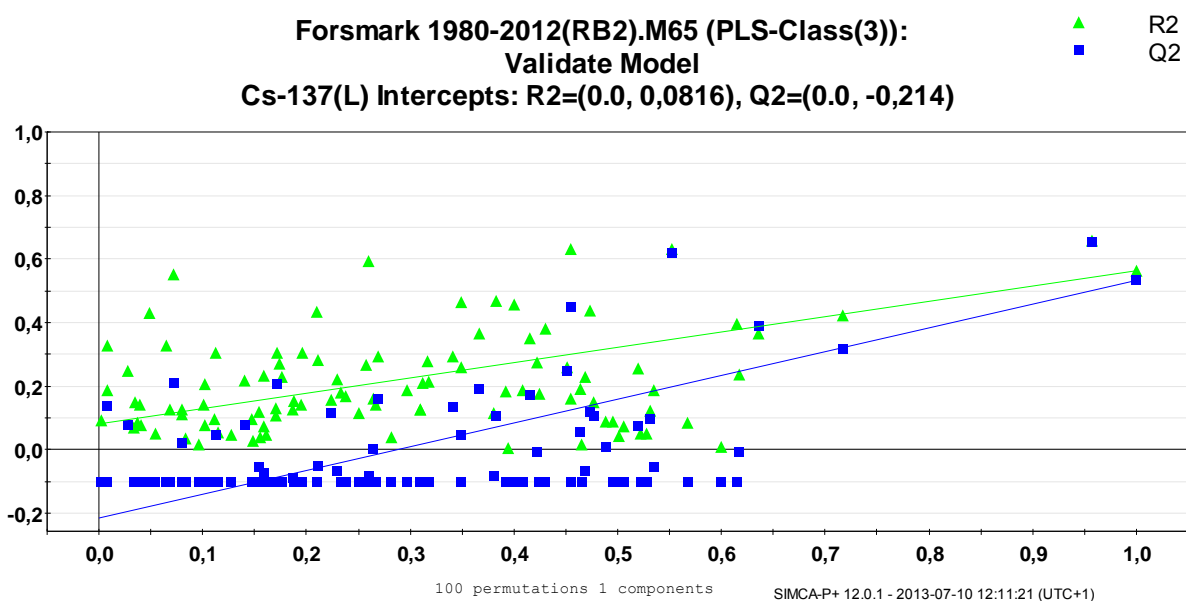


### 5.6.3. <sup>137</sup>Cs

#### Utsläpp till luft

Det går inte att finna någon strikt signifikant modell som kan relatera luftutsläppen av <sup>137</sup>Cs med de mätvärden som finns för de luftrelaterade omgivningsvariablerna. Detta trots relativt bra kontinuitet i mätserierna av omgivningsvariablerna. Det finns en stark misstanke om att <sup>137</sup>Cs från Tjernobyli, fördärvar regressionen mellan kontrollerade utsläpp och uppmätta aktiviteter i omgivningsvariablerna.

Modell M65, med 2 variabler (station B-Björnmossa och station Forsmark-Röt slam) har  $p=0,10$  och  $Q2=0,534$ , kommer närmast signifikans, av här testade modeller (p-värdet mindre än 0,05 för signifikant modell). Det finns dock slumpvisa permutationer av Y (testet "Validate Model") som resulterar i ett Q2 som är högre än det för originalmodellen, och därmed kan inte modellen betraktas som valid (se figur 18).

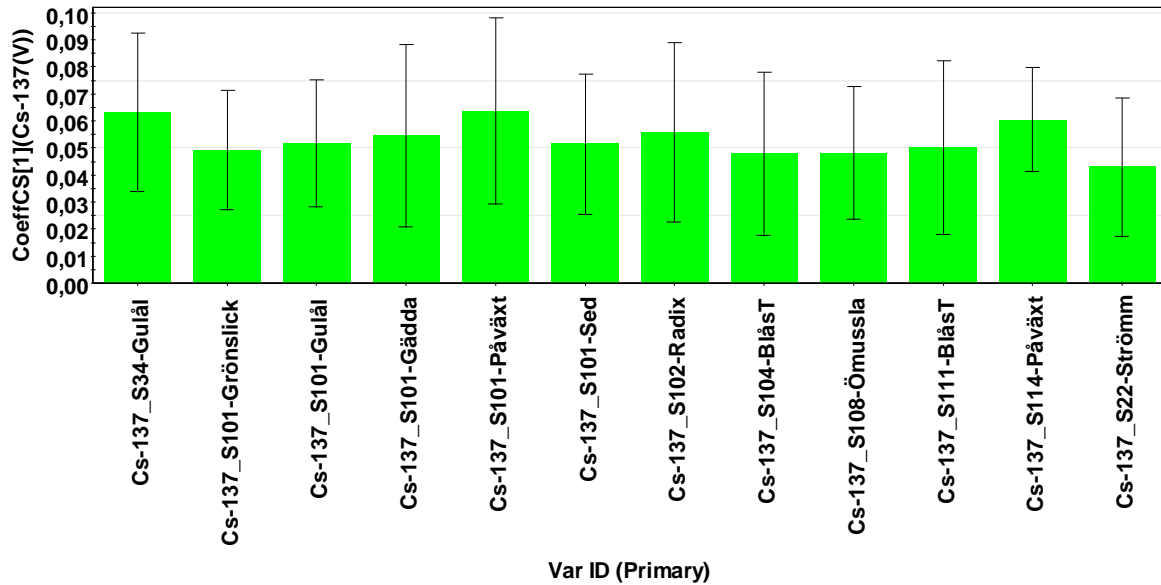


**Figur 18.** R2 och Q2 (vertikala axeln) avsatta för 100 slumpvisa permutationer av Y-värden mot korrelationskoefficienter (horisontella axeln) mellan ursprunglig Y och permuterat Y.

#### Utsläpp till vatten

En signifikant modell, M137, för utsläpp till vatten kan ses i figur 19. Denna modell är baserad på 11 omgivningsvariabler och har en prediktionsförmåga på 37,9 % ( $Q2=0,379$ ). Således en resurskrävande modell där många mätningar måste göras utan att få en speciellt bra modell för prediktion av de gjorda utsläppen. Det visar sig i den fortsatta analysen att, mätningar på omgivningsvariabler gjorda år 1986 och i viss mån också några efterföljande år, i flera fall är så extrema att det blir svårt att finna modeller med hög prediktionsförmåga. Det verkar troligt att <sup>137</sup>Cs från Tjernobyli utgör en betydande del av det mätta värdet och därför kommer modellerna inte bara att omfatta de kontrollerade utsläppen från anläggningen i Forsmark. Dock kommer modellerna att beskriva förhållanden som råder nu och ett antal år framåt, men med en icke helt entydig korrelation mellan kontrollerade utsläpp och mätt aktivitet i omgivningsvariablerna. Försök att beräkna, och kompensera för, "Tjernobylinslaget" av <sup>137</sup>Cs verkar vanskligt då det finns en skillnad mellan fysikalisk halveringstid och ekologisk halveringstid mellan olika provslag.

Forsmark 1980-2012(RB2).M137 (PLS-Class(3)),  
Cs-137 (V), x=11, p=0,0085 , Q2=0,379



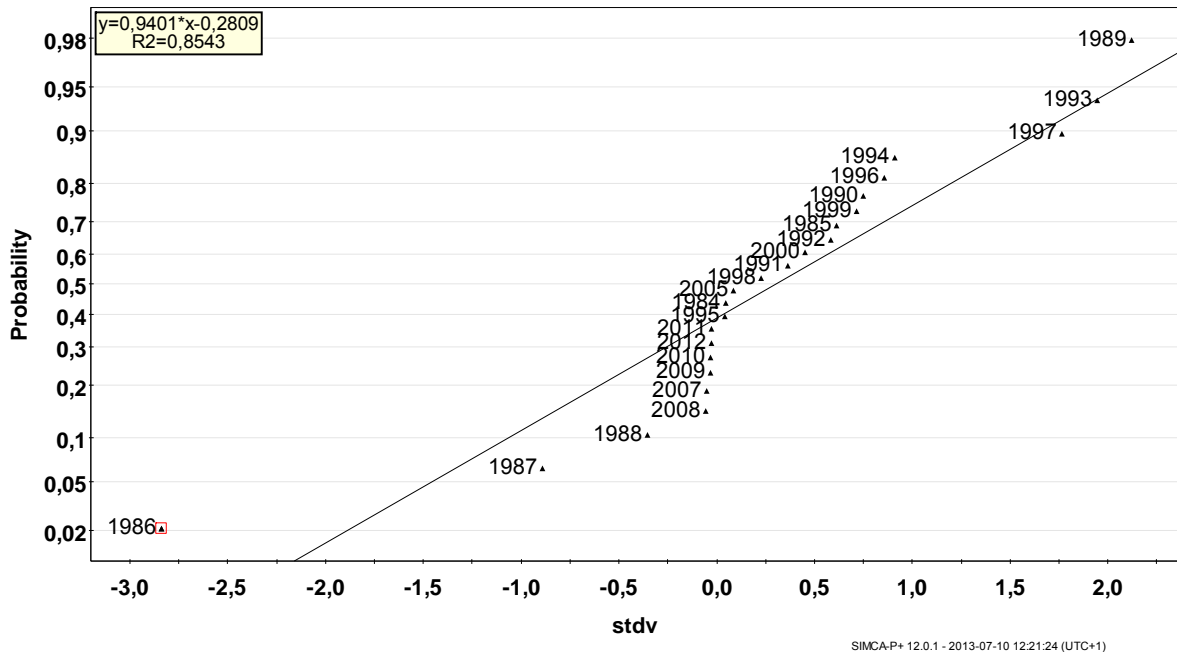
SIMCA-P+ 12.0.1 - 2013-07-12 12:34:59 (UTC+1)

**Figur 19.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av <sup>137</sup>Cs till vatten med 11 signifikanta variabler för den signifikanta modellen M137.

Av figur 19 ovan, kan det tyckas vara möjligt att skapa ett flertal olika en-variabelsmodeller, med så många signifikanta variabler. Förenklingen till en modell med en variabel kommer att göras i första hand efter hög frekvens mätdata per variabel, utspridd fördelning över tidsintervallet och utan extrema värden.

I figur 20 syns det att observationen 1986 är avvikande från de flesta övriga observationerna, och med vetskap om att det kontrollerade utsläppet av <sup>137</sup>Cs från Forsmark fick ett ytterligare bidrag från Tjernobyli bör denna observation uteslutas ur de fortsatta beräkningarna. Det visar sig att regressionsmodellerna blir "bättre" när även några efterföljande år utesluts.

Forsmark 1980-2012(RB2).M136 (PLS-Class(3)),  
Cs-137 (V), x=12, p=0,008 , Q2=0,368  
Normal Probability

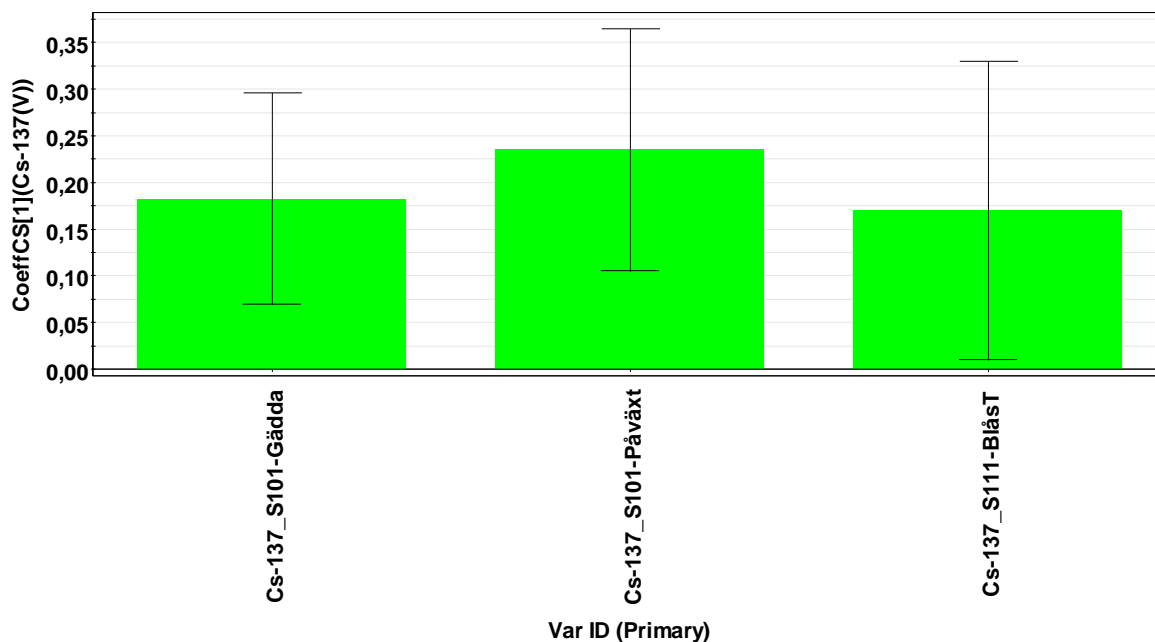


**Figur 20.** Normalfördelningsplot visar om residuerna (differens mellan observerat och predikterat värde) är normalfördelade.

När residuerna är normalfördelade ses dessa nära regressionslinjen. Observationen 1986 (märkt med rött i figur 20) avviker från linjen och dessutom är residualen nästan 3 standardavvikelser, vilket medför att man bör utreda om det kan finnas någon bakomliggande orsak till att denna observation inte riktigt ansluter till "normalfördelnings-linjen". Det är högst sannolikt att utsläppen från Tjernoby 1986 medför att denna observation inte passar in i mönstret av de övriga kontrollerade utsläppen och därför avviker något.

I modell M143 (figur 21) är observationer från 1986 till och med 1990 exkluderade. M143 är icke strikt signifikant, vilket förmodligen beror på att "Tjernoby-störningarna" har påverkat resultatet även efter 1990, även om den ekologiska halveringstiden är kortare än den fysikaliska.

Forsmark 1980-2012(RB2).M143 (PLS-Class(3)),  
Cs-137 (V), x=3, p=0,058



SIMCAP+ 12.0.1 - 2013-07-10 12:52:53 (UTC+1)

**Figur 21.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av <sup>137</sup>Cs till vatten med 3 signifikanta variabler för den icke strikt signifikanta modellen M143 (p=0,058).

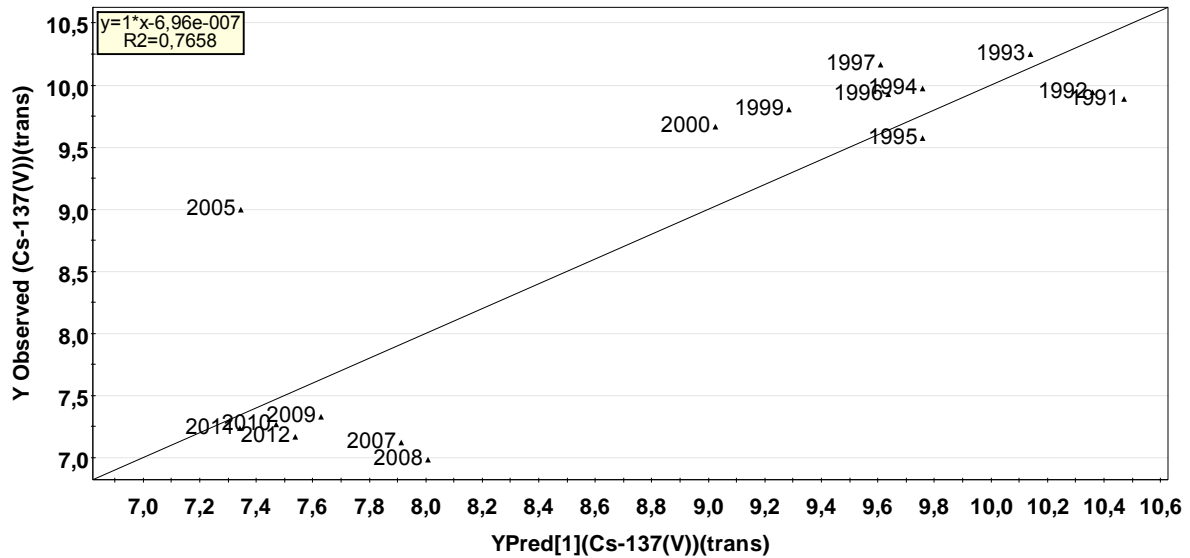
I tabell 3 ses prediktionsförmåga, Q2 och modellsignifikans, p, för en-variabelmodellerna baserade på de variablerna som visas i figur 21. Den klart bästa prediktionen av utsläpp av <sup>137</sup>Cs till vatten fås med modell M145. De tre omgivningsvariabler i tabell 3 har alla låg andel "missing data".

**Tabell 3** Sammanställning av prediktionsförmåga, Q2, och modellsignifikans, p, för 3 en-variabelmodeller för utsläpp av <sup>137</sup>Cs till vatten.

Modell	Mätstation	Provslag	Q2	p
M145	101	Påväxt	0,729	0,0002
M146	111	Blåstång	0,254	0,11(NS)
M147	101	Gädda	0,361	0,035

I figur 22 visas observerat mot predikerat utsläpp för modell, M145. Observation 2005, som avviker mest, underskattas med nästan två tiopotenser trots att residualen (se figur 23) inte är extrem.

Forsmark 1980-2012(RB2).M145 (PLS-Class(3)),  
Cs-137 (V), x= S101-Påväx, p=0,0002, Q2=0,729

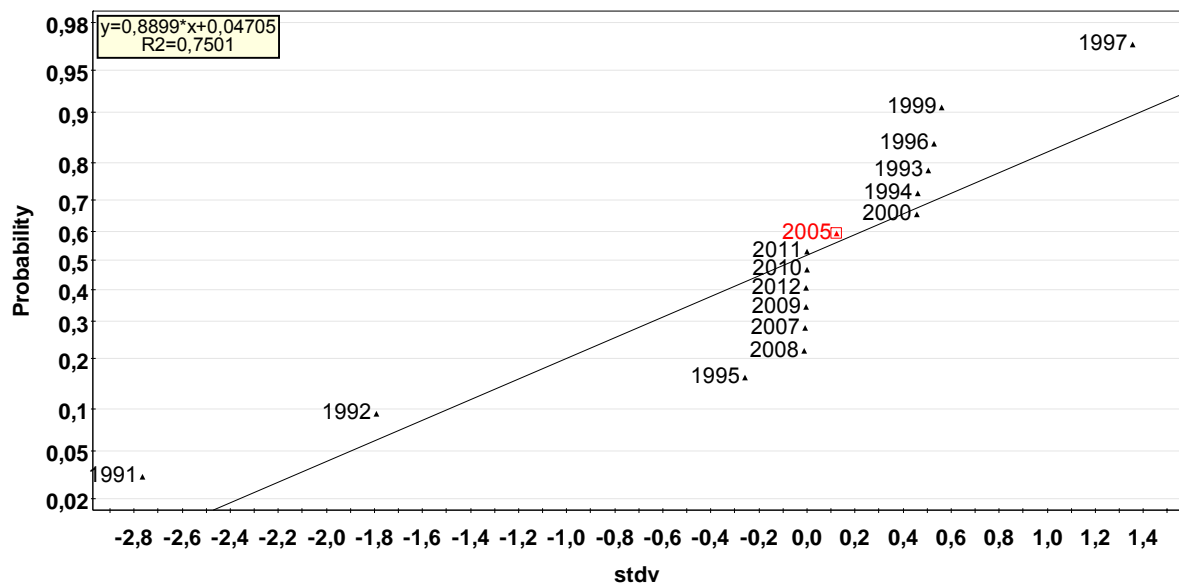


RMSEE = 7,89358e+009

SIMCA-P+ 12.0.1 - 2013-07-10 13:46:19 (UTC+1)

Figur 22. Observerat mot predikerat utsläpp av <sup>137</sup>Cs till vatten. Prediktionen är gjord med modell M145. Observera log-skala.

Forsmark 1980-2012(RB2).M145 (PLS-Class(3)),  
Cs-137 (V), x= S101-Påväx, p=0,0002, Q2=0,729



SIMCA-P+ 12.0.1 - 2013-07-10 13:59:53 (UTC+1)

Figur 23. Normalfördelningsplot visar om residuerna (differens mellan observerat och predikerat värde) är normalfördelade.

### 5.6.4. <sup>59</sup>Fe

#### Utsläpp till luft

Det finns endast 5 observationer/årsmedelvärden under tiden 1980-2012 av utsläpp av <sup>59</sup>Fe till luft, vilket medför att det svårligen går att finna någon korrelationsmodell mot luftrelaterade omgivningsvariabler.

## Utsläpp till vatten

Det går inte att finna någon signifikant regressionsmodell för utsläppen till vatten heller. Detta beror på att huvuddelen av utsläppen till vatten är registrerade före 1991 och endast 6 observationer vid senare tidpunkter. Vidare finns inga omgivningsdata redovisade före 1992, vilket leder till att det i detta fall inte går att finna några signifikanta regressionsmodeller.

Den tidiga modellen, M4 (figur 6), som inte hade separerat luft- och vattenvariabler gav en indikation om att det troligtvis skulle bli svårt att finna några signifikanta modeller, både för luft och för vatten.

### 5.6.5. $^{54}\text{Mn}$

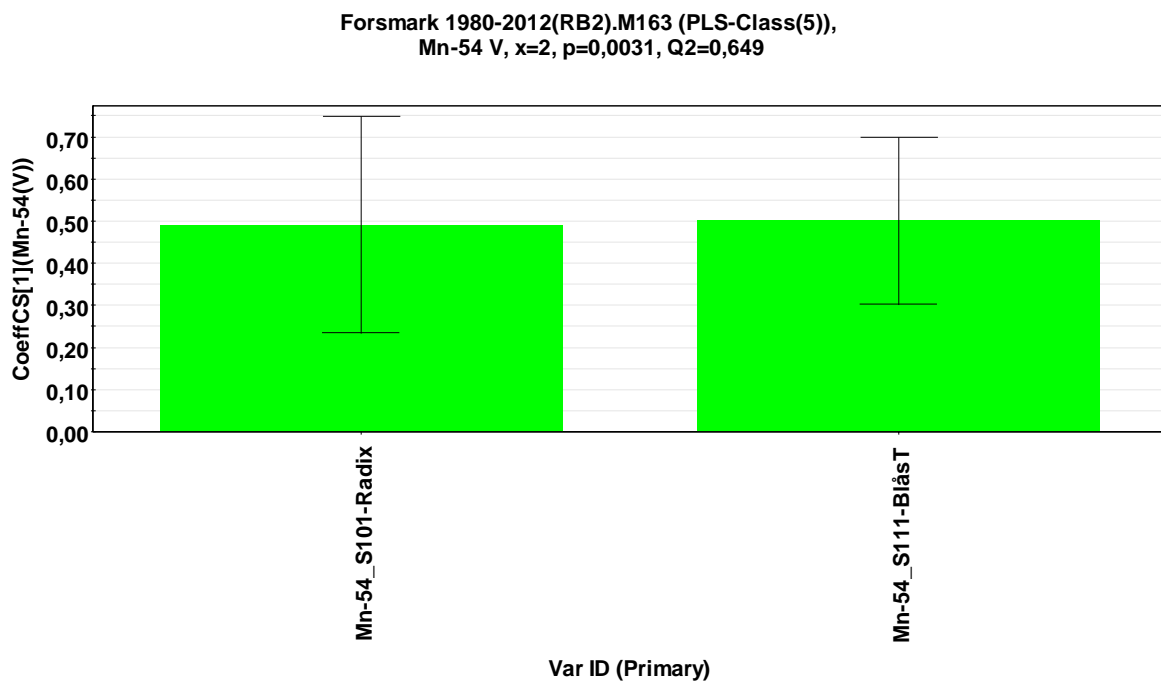
#### Utsläpp till luft

Trots att det med utgångspunkt i figur 6 (översikt av X/Y-samband för de olika nukliderna) verkade troligt att det skulle finnas säkra modeller för  $^{54}\text{Mn}$ , går det inte att finna något signifikant samband med befintliga data.

#### Utsläpp till vatten

En observation från datamatrixen visar att mätningarna vid station 101-Grönslick 1986, 1992 och 1993 visar betydligt högre värden än de från omgivande årtal.

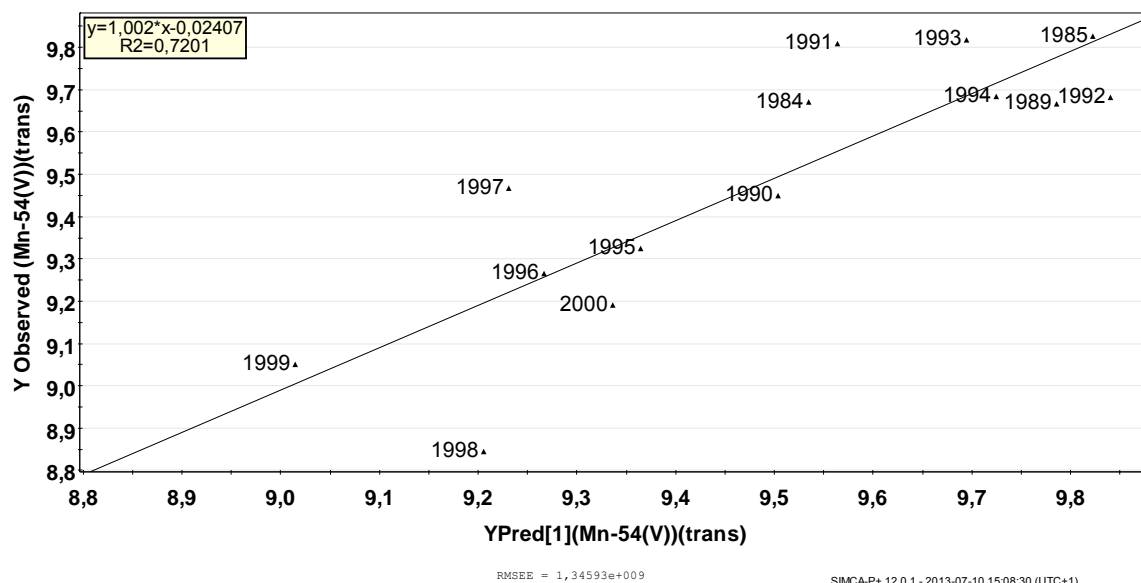
Trots försök med olika kombinationer av variabler verkar det bara resultera i 1 signifikant modell, M163, med 2 variabler.



**Figur 24.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{54}\text{Mn}$  till vatten med 2 signifikanta variabler, för den signifikanta modellen M163 ( $p=0,0031$ ).

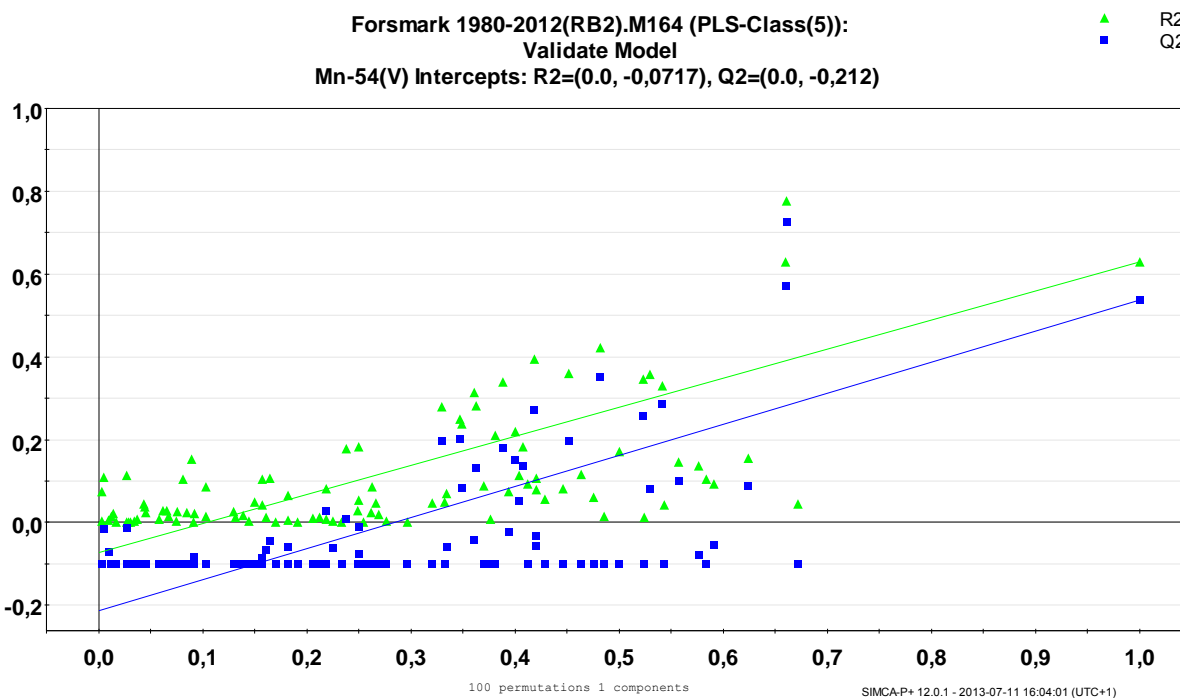
I figur 25 har observerat utsläpp avsett mot predikterat och det bör observeras att grafen "bara" spänner över en tiopotens, dock med relativt hög precision.

Forsmark 1980-2012(RB2).M163 (PLS-Class(5)),  
Mn-54 (V), x=2, p=0,0031, Q2=0,649



Figur 25. Observerat mot predikerat utsläpp av <sup>54</sup>Mn till vatten. Prediktionen är gjord med modell M163. Observera log-skala.

Den förenklade modellen M164, med 1 variabel, är inte strikt signifikant,  $p=0,067$  och enligt "validate model"-testet, efter 100 permutationer, finns det några slumpmässiga permutationer, som ger lika bra eller bättre prediktionsförmåga än den med den ursprungliga Y-vektorn med korrelationskoefficienten=1.



Figur 26. R2 och Q2 (vertikala axeln), för modell M164, avsatta mot korrelationskoefficienter (horisontella axeln) mellan ursprunglig Y och permuterat Y för 100 slumpvisa permutationer av Y-värden.

Det går således inte att finna någon en-variabelmodell för utsläpp av <sup>54</sup>Mn till vatten med föreliggande data.

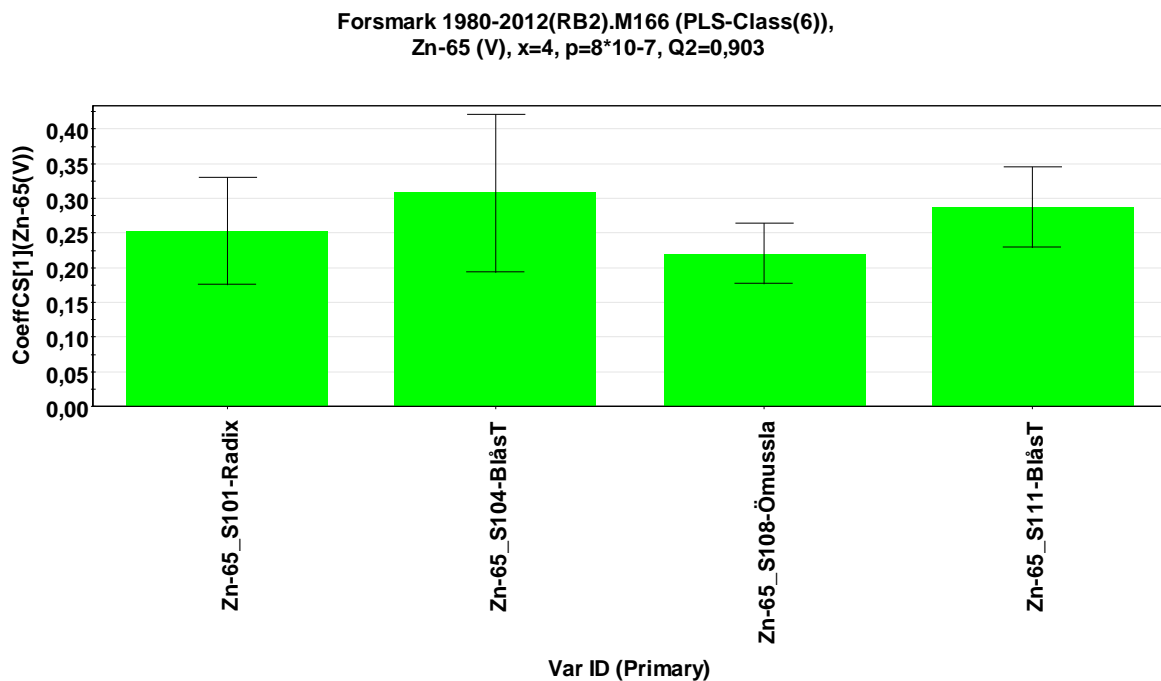
### 5.6.6. <sup>65</sup>Zn

#### Utsläpp till luft

Det går inte att finna någon signifikant modell med befintliga data. Omgivningsvariablerna är i de flesta fall mätta mellan åren 1993-2000. Det finns inga mätningar gjorda efter 2006. Andelen "missing data" är mycket hög och oftast högre än 73 %.

#### Utsläpp till vatten

Andelen "missing data" är relativt hög, men bättre än för utsläpp till luft. Det finns endast en mätning efter 2006 bland alla "vattenvariabler"! Detta till trots går det att finna några signifikanta modeller med riktigt hög prediktionsförmåga. Nedan visas regressionskoefficienterna för modell M166 som baseras på 4 omgivningsvariabler, där  $Q^2=0,903$  och  $p=8 \cdot 10^{-7}$ .

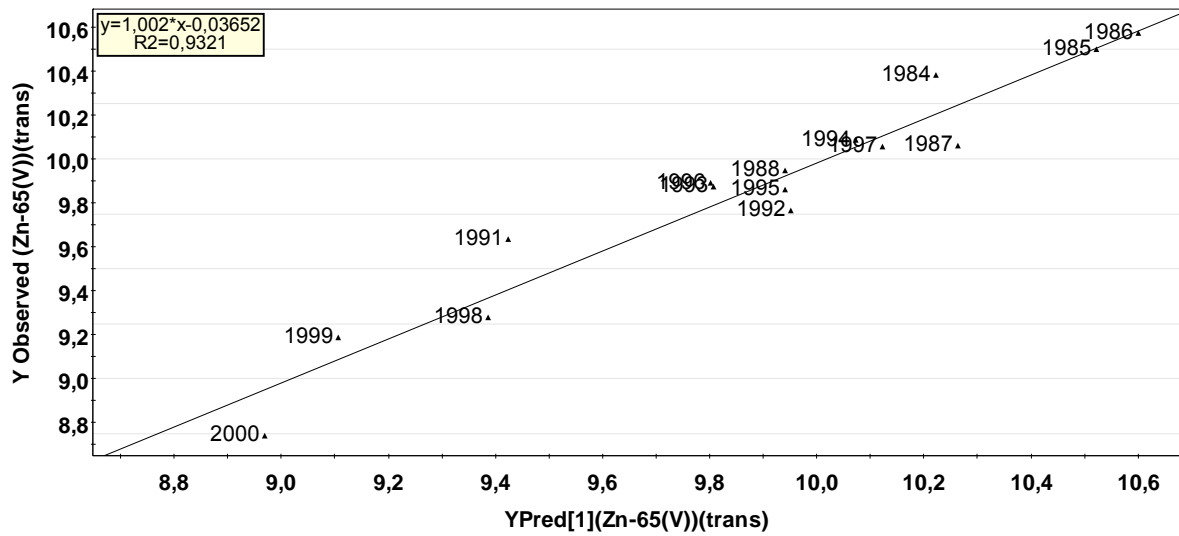


**Figur 27.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av <sup>65</sup>Zn till vatten med 4 signifikanta variabler, för den signifikanta modellen M166 ( $p=8 \cdot 10^{-7}$ ,  $Q^2=0,903$ ).

Observerat mot predikerat utsläpp för modell M166, ses i figur 28.



Forsmark 1980-2012(RB2).M166 (PLS-Class(6)),  
 Zn-65, (V)  $x=4$ ,  $p=8 \cdot 10^{-7}$ ,  $Q^2=0,903$

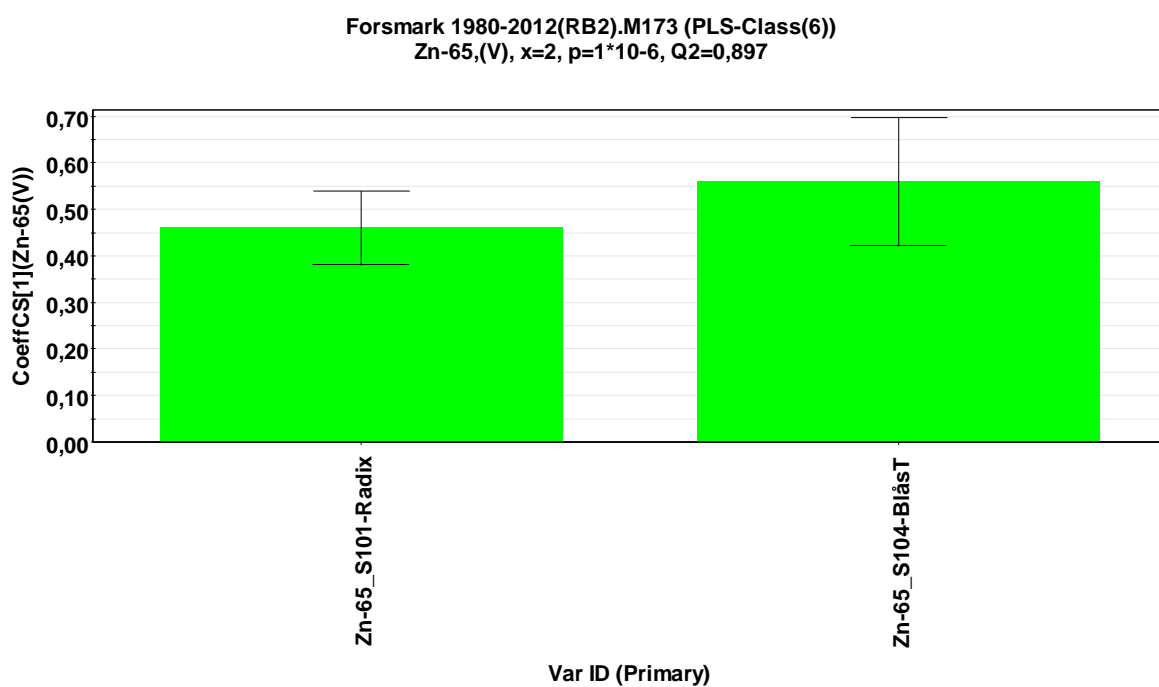
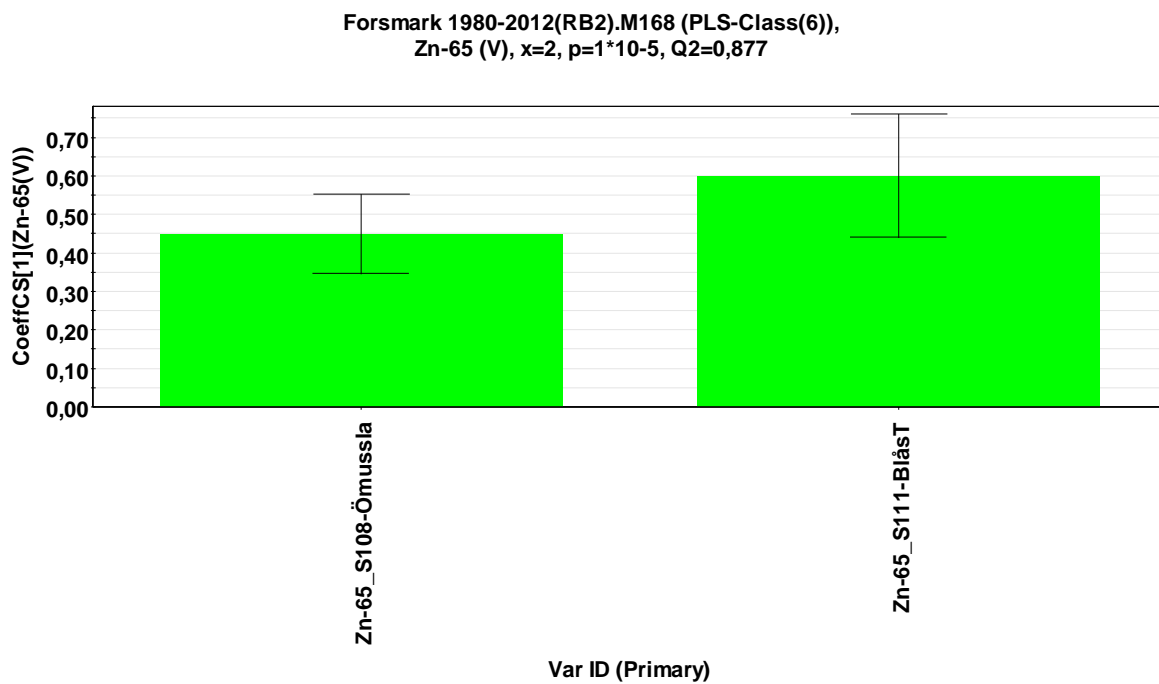


RMSEE = 3,17838e+009

SIMCA-P+ 12.0.1 - 2013-07-11 12:41:56 (UTC+1)

**Figur 28.** Observerat mot predikterat utsläpp av  $^{65}\text{Zn}$  till vatten. Prediktionen är gjord med modell M166. Observera log-skala

Förenklig av M166 leder till 2 signifikanta två-variabelmodeller, M168 och M173. Se figur 29 a och b.



**Figur 29a och b.** Plot av signifikanta regressionskoefficienter (scaled & centered) för utsläpp av  $^{65}\text{Zn}$  till vatten. **a)** M168, ( $p=9,7 \cdot 10^{-6}$ ,  $Q_2=0,877$ ) **b)** M173 ( $p=1,2 \cdot 10^{-6}$ ,  $Q_2=0,897$ ).

Ytterligare förenkling av M168 leder till bara en signifikant modell M170, med variabeln, Zn-65\_S11-Blåstång ( $p=3 \cdot 10^{-5}$ ) trots att osäkerheten för S108-Östersjömussla är mindre än den för S111-Blåstång blir inte en-variabelmodellen signifikant ( $p=0,22$ ). Prediktionsförmågan för M170 är mycket hög ( $Q_2=0,853$ ) vilket är nästan lika bra som modellen M166, med 4 signifikanta variabler.

En annan möjlig förenkling av modell M166 leder också till en signifikant modell M173, med 2 signifikanta omgivningsvariabler.

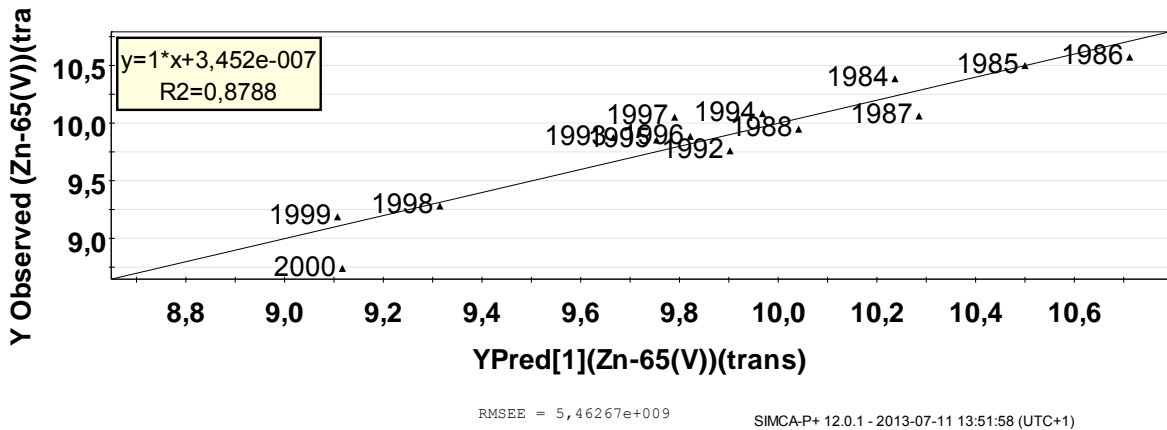
Således kan 3 signifikanta modeller med 1 variabel observeras där prediktionsförmågan Q2 skiljer något, men är hög.

Det bör påpekas att alla tänkbara kombinationer av omgivningsvariabler i ursprungsdata inte är testade. Det som eftersträvas vid förenklingen av modellerna är signifikans hos modeller och variabler, normalfördelade residuer i prediktionsfelen, hög frekvens mätdata per variabel och uteslutning av extrema avvikande mätvärden eller extrema år (ses i normalfördelningsplot där residualen är större än 3 standardavvikelser).

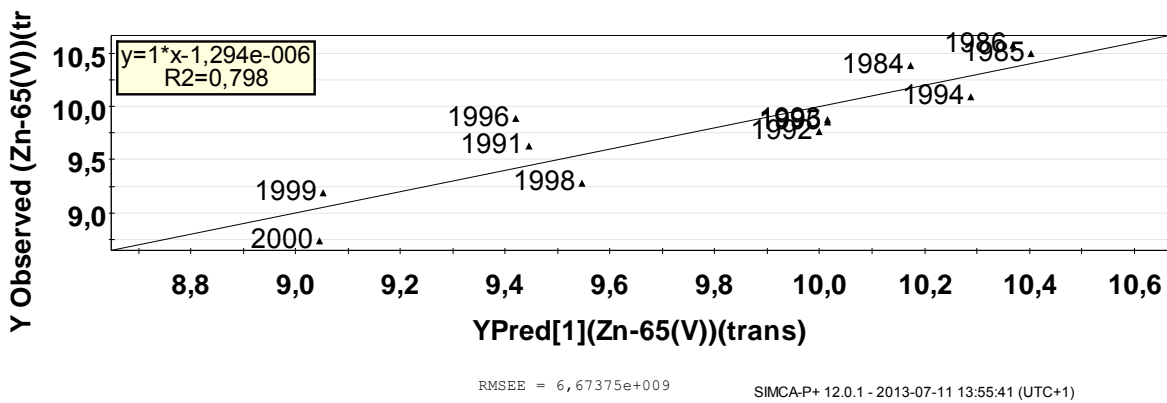
**Tabell 4.** Sammanställning av prediktionsförmåga, Q2, och modellsignifikans, p, för 3 en-variabel modeller för utsläpp av  $^{65}\text{Zn}$  till vatten.

<b>Modell</b>	<b>Station</b>	<b>Provslag</b>	<b>Q2</b>	<b>p</b>
M170	111	Blåstång	0,853	$3 \cdot 10^{-5}$
M174	101	Radix	0,757	$2 \cdot 10^{-3}$
M175	104	Blåstång	0,937	$6 \cdot 10^{-8}$

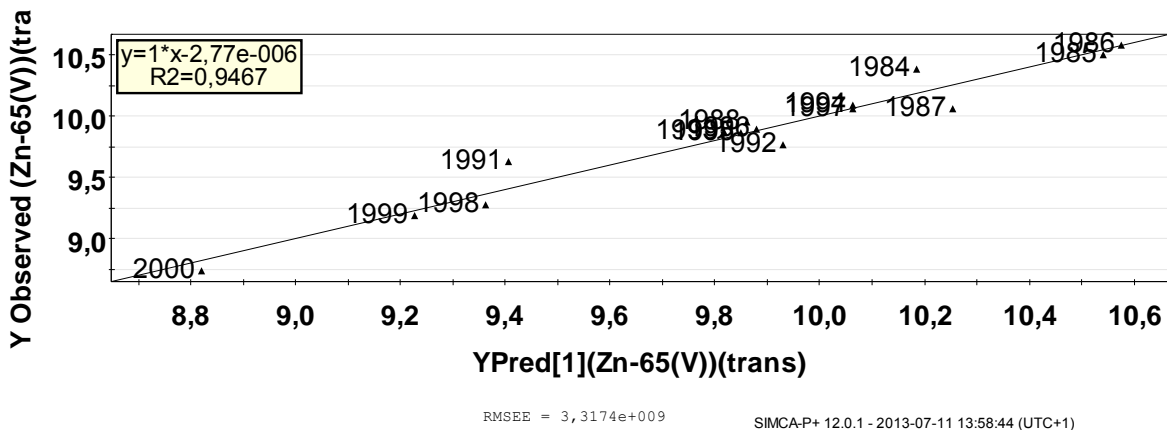
**Forsmark 1980-2012(RB2).M170 (PLS-Class(6))**  
**Zn-65 (V), x=111- Blåstång, p=3\*10-5, Q2=0,853**



**Forsmark 1980-2012(RB2).M174 (PLS-Class(6))**  
**Zn-65 (V), x=101-Radix, p=0,002, Q2=0,757**

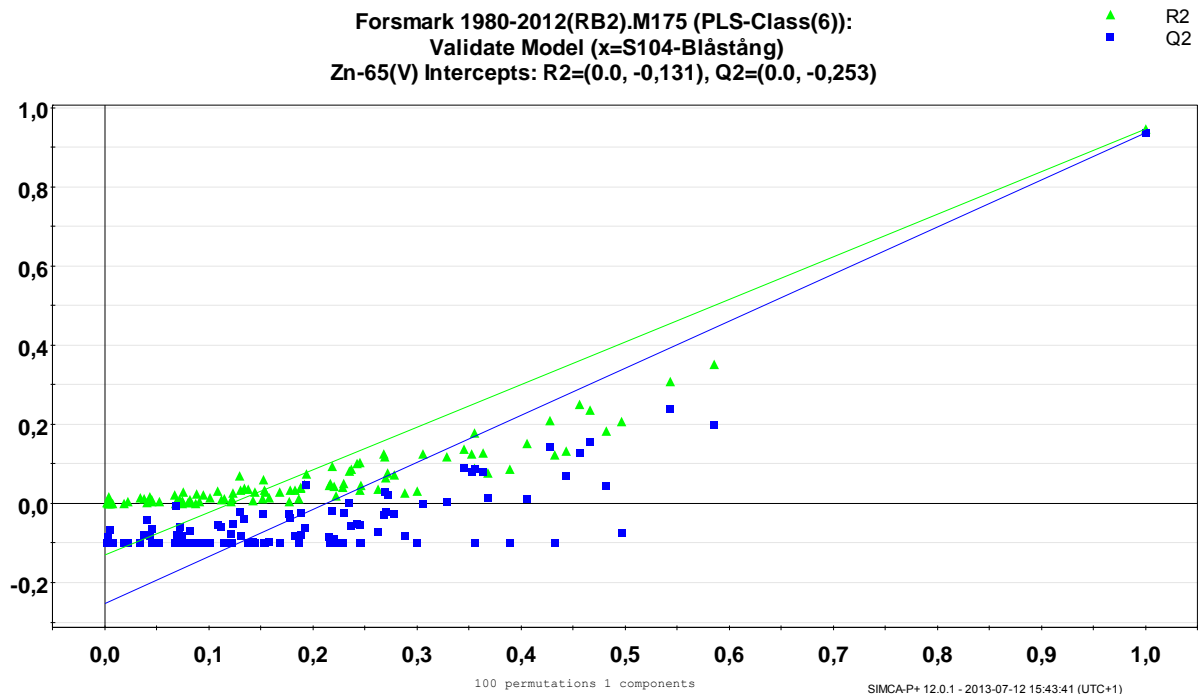


**Forsmark 1980-2012(RB2),M175 (PLS-Class(6))**  
**Zn-65,(V), x=104-Blåstång, p=3\*10-5, Q2=0,937**



**Figur 30a, b och c.** Observerat mot predikerat utsläpp av <sup>65</sup>Zn till vatten. Prediktionen är gjord med modellerna a) M170 b) M173 och c) M175. Observera log-skala.

Det syns ingen större skillnad mellan de tre modellerna när det gäller prediktion av kontrollerade utsläpp av  $^{65}\text{Zn}$  till vatten. Det har därför liten betydelse för resultatet vilken variabel som väljs, men S104-Blåstång ger ändå det bästa resultatet. Att modell M175 är trovärdig och inte kommit fram av en slump, visas i figur 31, där ingen av de slumpvis genererade modeller får ett  $Q^2 > 0,3$  och där modell M175 har ett  $Q^2 = 0,937$ .



**Figur 31.** R2 och Q2 (vertikala axeln), för modell M175 (S104-Blåstång) avsatt mot korrelationskoefficienter (horisontella axeln) mellan ursprunglig Y och permuterat Y för 100 slumpvisa permutationer av Y-värden.

De väldigt övertygande resultaten för de slutliga modellerna för utsläpp av  $^{65}\text{Zn}$  till vatten visar att den tidiga modellen M4 som visas i figur 6, inte var helt tillförlitlig, innan data analyserades klassvis.

## 5.7. Prediktionsförmåga och prediktioner

Nedan följer en sammanställning av prediktionsförmåga och prediktioner för signifikanta PLS-modeller med en eller fler variabler för varje nuklid.

**Tabell 5.** Sammanställning av prediktionsförmåga för utsläppen till vatten av respektive nuklid för modeller med flera (n) variabler mot modeller med 1 variabel. För <sup>59</sup>Fe finns ingen signifikant modell till vare sig luft eller vatten. Observera att det är endast för <sup>60</sup>Co(L) som det finns en signifikant modell, med 2 variabler, för utsläpp till luft och luftrelaterade omgivningsvariabler.

Nuklid	Modell	Antal variabler (n)	Q2_n (%)	Modell	1 variabel	Q2_1 (%)
<sup>58</sup> Co(V)	M36	3	96,7	M176	S101-Påväxt	89,7
<sup>60</sup> Co(V)	M50	11	85,9	M132	S101-Påväxt	93,6
<sup>137</sup> Cs(V)	M137	11	37,9	M145	S101-Påväxt	72,9
<sup>59</sup> Fe(V)	-	-	-	-	-	-
<sup>54</sup> Mn(V)	M163	2	64,9	-	-	-
<sup>65</sup> Zn(V)	M166	4	90,3	M175	S104-Blåstång	93,7
<sup>58</sup> Co(L)	-	-	-	-	-	-
<sup>60</sup> Co(L)	M122	2	86,2	-	-	-
<sup>137</sup> Cs(L)	-	-	-	-	-	-
<sup>59</sup> Fe(L)	-	-	-	-	-	-
<sup>54</sup> Mn(L)	-	-	-	-	-	-
<sup>65</sup> Zn(L)	-	-	-	-	-	-

**Tabell 6.** Regressionskoefficienter (logaritmisk skala) för modeller med 1 signifikant variabel

Modell	M176	M132	M145			M175
Nuklid	<sup>58</sup> Co(V)	<sup>60</sup> Co(V)	<sup>137</sup> Cs(V)	<sup>59</sup> Fe(V)	<sup>54</sup> Mn(V)	<sup>65</sup> Zn(V)
Konstant (B)	5,77849	4,67772	0,573919	-	-	7,65916
Koefficient (k)	1,60059	1,64372	3,11325	-	-	1,17745

I tabell 6 avviker regressionskoefficienterna för <sup>137</sup>Cs något, från mönstret av de övriga nukliderna. Modeller för <sup>137</sup>Cs ger heller inte riktigt lika bra resultat som för övriga signifikanta nuklid-modeller. Det är troligt att det i den mätta aktiviteten för <sup>137</sup>Cs, i de olika provslagen, också inkluderar bidrag från Tjernobyl.

För <sup>59</sup>Fe saknas signifikanta regressionsmodeller både för utsläpp till luft och till vatten. För utsläpp av <sup>54</sup>Mn till vatten saknas signifikant en-variabelmodell.

Observera också att det inte, för någon nuklid, finns någon signifikant en-variabelmodell för utsläpp till luft, men det finns 1 signifikant modell med 2 variabler för utsläpp av <sup>60</sup>Co till luft.

Regressionssambandet kan skrivas:

$$\log(Y) = k \cdot \log(X) + B$$

där  $Y$ =utsläpp/nuklid,  $X$ =uppmätt aktivitet/omgivningsvariabel och  $k$  och  $B$  är konstanter för varje regressionsmodell. Följaktligen kan aktiviteten för varje variabel, uppskattas från känt utsläpp med följande modell:

$$\log(X) = (\log(Y) - B)/k$$

I tabell 7 visas prediktionsresultat som gäller för år 1995. Detta år valdes på grund av att det inte saknas några värden för de kontrollerade utsläppen, men också för att omgivningsvariablerna generellt inte har så mycket "missing data" detta år.

**Tabell 7.** Mätta och predikterade värden för utsläpp till vatten och vattenrelaterade omgivningsaktivitet under år 1995. För  $^{59}\text{Fe}$  och  $^{54}\text{Mn}$  finns inga signifikanta modeller. Predikterat utsläpp baseras på mätt aktivitet i respektive omgivningsvariabel. Predikterad aktivitet för en omgivningsvariabel baseras på mätta utsläpp till vatten eller luft.

Nuklid	Mätt utsläpp (Bq) 1995	Predikterat utsläpp	Mätt på en omgivningsvariabel (Bq) 1995	Predikterat för en omgivningsvariabel	Variabel
$^{58}\text{Co}$	7,62E+09	1,58E+10	577	366	101-Påväxt
$^{60}\text{Co}$	2,47E+10	3,08E+10	3428	2997	101-Påväxt
$^{137}\text{Cs}$	3,72E+09	5,74E+09	892	776	101-Påväxt
$^{59}\text{Fe}$	5,41E+06	-	-	-	-
$^{54}\text{Mn}$	2,12E+09	-	-	-	-
$^{65}\text{Zn}$	7,24E+09	7,07E+09	73	74	104-Båstång

Prediktioner av utsläppen i tabell 7 visar god överensstämmelse med faktiskt mätta värden. Prediktionsfelen är klart mindre än en tiopotens. Bäst överensstämmelse ses för  $^{65}\text{Zn}$ , i överensstämmelse med vad som indikeras av respektive modells Q2-värde i tabell 5.

Det bör påpekas att de signifikanta modellerna i denna studie bör ge bra prediktioner för utsläpp som ligger inom det utsläppsintervall, som legat till grund för "modellbygget". Observera också att modellerna kan användas för varje enskilt utsläpp och inte bara för årsmedelvärden. Därmed ska det inte förväntas att extrapolationer långt utanför detta intervall behöver vara trovärdiga. Modellerna kommer således att kunna användas för prediktioner vid "normala" utsläpp och därmed också signalera när utsläppen inte är som "de brukar vara".

## 6. Rekommendationer

- Denna analys har visat på nödvändigheten av att ha välorganiserade data och kontinuitet i mätningar av kontrollerade utsläpp och mätta aktiviteter i omgivningsprover. Detta är viktigt för att dels kunna följa eventuella förändringar i utsläpp och mätt aktivitet i omgivningsvariablerna, dels möjliggöra och säkerställa signifikanta regressionssamband mellan utsläpp till luft och vatten och relaterade mätvärden från omgivningsprover. Det är också viktigt för att kunna göra trovärdiga prediktioner för, å ena sidan storleken på utsläpp genom mätning av omgivningsvariabler och å andra sidan för att uppskatta aktivitet i visst provslag från kontrollerade utsläpp.
- Fortsättningsvis bättre kontinuitet i mätserierna där även värden under detektionsgräns noteras för omgivningsvariabler, för både luft och vatten.
- Ur ett "regressionsperspektiv" är det bättre att fokusera på färre mätstationer och provslag, för att öka kontinuiteten. Det finns förstås andra bevekelsegrunder för myndigheten att rekommendera mätningar på många stationer och provslag.
- Vattenrelaterade omgivningsvariabler där det i nuläget finns relativt hög frekvens med mätvärden, och som är intressanta för regressionssambanden är t.ex. påväxtprov, blåstång, gädda och abborre.
- Luftrelaterade omgivningsvariabler visar i denna studie generellt mycket låg mätfrekvens. Endast för  $^{60}\text{Co(L)}$  erhöles en signifikant modell med 2 variabler (Bete och Renlav). Det är klokt att fortsätta att fokusera på dessa provslag och kanske med tillägg av Vagg- och Björnmossa, som också visar lite högre mätfrekvens än de flesta övriga "luftvariablerna".
- För att få en uppfattning om spridningskinetiken från utsläpp till omgivningen, bör ett kontrollerat utsläpp följas upp med mätning av aktiviteten i några olika provslag och mätstationer med regelbunden frekvens. Mätningarna kanske bör göras veckovis eller varje månad efter ett kontrollerat utsläpp. Viktigt dock att prover tas "samtidigt". Dessa mätdata kan även kombineras med meteorologiska data (t.ex. förhärskande vindriktning och vindhastighet, vattentemperatur och strömningshastighet etc.). Därefter skulle en multivariat tidsserieanalys kunna ge en uppfattning om kinetik av upptag/deponering till olika provslag\_provstationer.



## 7. Slutsatser

- Data som analyserats med multivariata verktyg (PCA och PLS) visar att rapporteringen av de kontrollerade utsläppen av radioaktiva nuklider ( $^{58}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$ ,  $^{59}\text{Fe}$ ,  $^{54}\text{Mn}$  och  $^{65}\text{Zn}$ ) från Forsmark till luft och vatten har hög kontinuitet i tidsserierna 1980-2012. Undantaget  $^{59}\text{Fe}$  där endast utsläpp till vatten har en användbar kontinuitet av mätvärden.
- De kontrollerade utsläppen av radioaktiva ämnen till vatten tenderar att minska med tiden.
- Aktiviteten av utsläppen till luft är i medeltal oftast lägre än aktiviteten av utsläpp till vatten fram till ungefär 2005. Därefter är utsläppen till luft lika stora eller större än de till vatten. Minskningen av luftutsläppen är inte lika tydlig som i "vattenfallet".
- Insamlade mätvärden på omgivningsvariabler, olika provslag och mätstationer, har dessvärre, med några undantag, mycket låg frekvens. Flest antal mätningar är gjorda mellan 1993-2000.  $^{137}\text{Cs}$  utgör ett undantag där det finns högre frekvens mätdata även under perioden 2005-2012, där det för övriga nuklider bara finns sporadiska observationer.
- Multivariatanalysen visar att det finns kvantifierbara samband mellan utsläpp till vatten och till vattenrelaterade omgivningsvariabler för 5 av de 6 studerade nukliderna, men endast ett samband för luftrelaterat utsläpp,  $^{60}\text{Co(L)}$ .
- För nukliderna  $^{58}\text{Co}$ ,  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$  och  $^{65}\text{Zn}$  finns signifikanta vattenrelaterade en-variabelsamband.
- PLS-modeller med 1 signifikant variabel har, i denna studie, i 3 av 4 fall, bättre prediktionsförmåga,  $Q^2$ , än de fall där modellerna består av fler signifikanta variabler.
- Storleken på ett utsläpp ( $Y$ ) kan predikteras genom mätning av aktiviteten i en eller flera omgivningsvariabler. De använda PLS-modellerna, med 1 eller flera variabler, såväl som omgivningsvariablerna, måste vara signifikanta. Observera att inte alla omgivningsvariabler i denna rapport kan användas.
- Utsläppen av  $^{58}\text{Co}$  och  $^{60}\text{Co}$  till vatten kan predikteras med hög precision över mer än 4 tiopotenser i aktivitet.
- Aktiviteten i en omgivningsvariabel ( $X$ ) kan beräknas med signifikanta PLS-modeller med 1 signifikant variabel vid känt utsläpp. Efter att regressionsuttrycket för utsläppet omformats till  $\log(X) = (\log(Y) - B)/k$ , där  $Y$ =mätt utsläpp/nuklid, och  $k$  och  $B$  är konstanter för varje nuklid, provslag och mätstation. För signifikanta modeller fås relativt god överensstämmelse mellan observerat och predikerat värde. Prediktionsfelen är betydligt mindre än en tiopotens.
- Multivariatanalysen visar således att det, ur ett mättekniskt perspektiv, räcker att mäta aktiviteten i något/några provslag vid någon enstaka mätstation för att kunna uppskatta storleken på det gjorda utsläppet.
- Vidare kräver signifikanta PLS-modeller med bara 1 signifikant variabel mindre resurser och arbetsinsatser än modeller med fler variabler. Analytiskt/praktiskt bibehålls eller till och med förbättras prediktionsförmågan,  $Q^2$ , vilket bör tas i beaktning när kostnad, nytta och "omvärldens" uppfattning om säkerhet ska vägas samman.
- De signifikanta modellerna i denna studie kan användas för att prediktera kontrollerade utsläpp utifrån mätningar i omgivningen, och vidare kan aktiviteter i omgivningsvariabler predikteras utifrån uppmätta utsläpp. Observera också att modellerna kan användas för varje enskilt utsläpp och inte bara för årsmedelvärden Detta gäller för utsläpp som ligger inom det utsläppsintervall som legat till grund för modellerna. Modellerna borde alltså kunna ge en indikation om avvikelser vid stora prediktionsfel, som t.ex. ett diffust läckage.







2014:14

Strålsäkerhetsmyndigheten har ett samlat ansvar för att samhället är strålsäkert. Vi arbetar för att uppnå strålsäkerhet inom en rad områden: kärnkraft, sjukvård samt kommersiella produkter och tjänster. Dessutom arbetar vi med skydd mot naturlig strålning och för att höja strålsäkerheten internationellt.

Myndigheten verkar pådrivande och förebyggande för att skydda människor och miljö från oönskade effekter av strålning, nu och i framtiden. Vi ger ut föreskrifter och kontrollerar genom tillsyn att de efterlevs, vi stödjer forskning, utbildar, informerar och ger råd. Verksamheter med strålning kräver i många fall tillstånd från myndigheten. Vi har krisberedskap dygnet runt för att kunna begränsa effekterna av olyckor med strålning och av avsiktlig spridning av radioaktiva ämnen. Vi deltar i internationella samarbeten för att öka strålsäkerheten och finansierar projekt som syftar till att höja strålsäkerheten i vissa östeuropeiska länder.

Strålsäkerhetsmyndigheten sorterar under Miljödepartementet. Hos oss arbetar drygt 250 personer med kompetens inom teknik, naturvetenskap, beteendevetenskap, juridik, ekonomi och kommunikation. Myndigheten är certifierad inom kvalitet, miljö och arbetsmiljö.

**Strålsäkerhetsmyndigheten**  
**Swedish Radiation Safety Authority**

SE-171 16 Stockholm  
Solna strandväg 96

**Tel:** +46 8 799 40 00  
**Fax:** +46 8 799 40 10

**E-mail:** [registrator@ssm.se](mailto:registrator@ssm.se)  
**Web:** [stralsakerhetsmyndigheten.se](http://stralsakerhetsmyndigheten.se)