



Strål  
säkerhets  
myndigheten

Swedish Radiation Safety Authority

Författare: Rolf Bergman

Forskning

# 2014:15

Multivariatanalys av radioaktivitetsdata från utsläpp till luft och vatten från Studsviksanläggningarna och i omgivningen av Studsvik



## 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 Studsviksanläggningarna och uppmätta halter radioaktivitet i omgivningen av Co-60, Cs-134, Cs-137, Eu-152, Eu-154 och Mn-54. Ett mål var att se hur olika provslag i omgivningen av Studsvik 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-60, Cs-134, Cs-137, Eu-152, Eu-154 och Mn-54. Gällande luftrelaterade utsläpp kunde studien påvisa signifikanta modeller för Cs-134 och Cs-137.

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 Studsviksanläggningarna 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

# 2014:15

Multivariatanalys av radioaktivitetsdata  
från utsläpp till luft och vatten från  
Studsviksanläggningarna och i  
omgivningen av Studsvik

Datum: December 2013

Rapportnummer: 2014:15 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>
<b>4.1. PCA</b> .....	<b>7</b>
<b>4.2. PLS</b> .....	<b>7</b>
<b>4.3. ANALYSGÅNG</b> .....	<b>9</b>
<b>5. RESULTAT OCH DISKUSSION</b> .....	<b>10</b>
<b>5.1. DATA</b> .....	<b>10</b>
<b>5.2. PCA – UTSLÄPPSDATA</b> .....	<b>13</b>
<b>5.3. PLS – ALLA DATA</b> .....	<b>14</b>
<b>5.4. PLS – UTSLÄPP TILL LUFT RESPEKTIVE VATTEN</b> .....	<b>15</b>
<b>5.5. PLS – PER NUKLID</b> .....	<b>16</b>
<b>5.6.1. <sup>60</sup>Co</b> .....	<b>16</b>
Utsläpp till luft.....	16
Utsläpp till vatten.....	16
<b>5.6.2. <sup>134</sup>Cs</b> .....	<b>20</b>
Utsläpp till luft.....	20
Utsläpp till vatten.....	22
<b>5.6.3. <sup>137</sup>Cs</b> .....	<b>24</b>
Utsläpp till luft.....	24
Utsläpp till vatten.....	26
<b>5.6.4. <sup>152</sup>Eu</b> .....	<b>30</b>
Utsläpp till luft.....	30
Utsläpp till vatten.....	30
<b>5.6.5. <sup>154</sup>Eu</b> .....	<b>31</b>
Utsläpp till luft.....	31
Utsläpp till vatten.....	31
<b>5.6.6. <sup>54</sup>Mn</b> .....	<b>34</b>
Utsläpp till luft.....	34
Utsläpp till vatten.....	34
<b>5.7. PREDIKTIONSFÖRMÅGA OCH PREDIKTIONER</b> .....	<b>36</b>
<b>6. REKOMMENDATIONER</b> .....	<b>38</b>
<b>7. SLUTSATSER</b> .....	<b>39</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 Studsvik för perioden 1977-2012. De studerade nukliderna är  $^{60}\text{Co}$ ,  $^{134}\text{Cs}$ ,  $^{137}\text{Cs}$ ,  $^{152}\text{Eu}$ ,  $^{154}\text{Eu}$ , och  $^{54}\text{Mn}$ . Utsläppen sker kontrollerat till luft och vatten från de kärntekniska anläggningarna i Studsvik.

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 Studsvik genom mätvärden från ett fåtal olika provslag (t.ex. blåstång, Sediment och påväxtprover) vid någon eller några av de utvalda mätstationerna i omgivningarna av Studsvik. 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 169 stycken där varje enskild variabel definieras av nuklid, mätstation och provslag (ex.  $^{137}\text{Cs}_{\text{S33-Blåstång}}$ ). Analysen visar att det för 5 av de 6 studerade nukliderna,  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$ ,  $^{152}\text{Eu}$ ,  $^{154}\text{Eu}$ , och  $^{54}\text{Mn}$ , 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  $^{134}\text{Cs}$  behövs minst 3 provslag, för samma uppskattning.

För utsläpp av radioaktiva nuklider till luft kunde bara motsvarande beräkningar göras för nukliden  $^{134}\text{Cs}$ , med hjälp av 1 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 datamatrixen 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 datamatrixen 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 de kärntekniska anläggningarna i Studsvik 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 tidsserierna 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 datamatriisen består av insamlade mätvärden för årliga utsläpp av nukliderna  $^{60}\text{Co}$ ,  $^{134}\text{Cs}$ ,  $^{137}\text{Cs}$ ,  $^{152}\text{Eu}$ ,  $^{154}\text{Eu}$  och  $^{54}\text{Mn}$  till luft (L) och vatten (V) [Bq/år] från de kärntekniska anläggningarna i Studsvik. Aktiviteten av respektive nuklid utgör medelvärden per varje år. Datamatriisen 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. Notera t.ex. att station 3 upphör efter 1995. Ett antal luftfilterstationer, Studsvik (A, B, C) har slagits samman för att generera ett samlat mätvärde. I något fall finns även en station D nämnd, men den finns inte redovisad på någon av kartorna för provtagningspunkterna. Aktivitetsmätningarna på luftfilter upphörde från 1995. Det förekommer också sporadiska mätningar vid ett flertal mätstationer och dessa kommer inte att kunna behandlas i den aktuella rapporten.

Datamatriisen är organiserad så att varje år utgör en observation, och varje variabel definieras av nuklid\_mätstation-provslag (ex.  $^{60}\text{Co}$ \_S2-Sediment). Kontrollerade utsläpp finns dokumenterade i denna rapport mellan 1977-2012,. Det finns 32 observationer (1981-2012) för mätvärden på omgivningsvariabler och antalet variabler som kan användas i föreliggande multivariatanalys är 169. Variabler med färre än 7 observationer, av totalt 32 möjliga (>80% missing data) har uteslutits i den primära datamatriisen.

## 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 om variablerna befinner sig 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)).

R2 visar hur väl framräknade modeller anpassas till data. Förklaringsgraden, Q2, anger modellens prediktionsförmåga av nya Y med nya X-värden. Q2 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 Q2 (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. 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 av "Validate Model" i mjukvaran. I denna test behålls X-matrisen intakt, men Y permuteras slumpvis, och för varje permutation beräknas modell Anpassning, R2, och prediktionsförmåga, Q2. Dessa avsätts på den vertikala axeln, mot korrelationskoefficienten på den horisontella axeln för respektive ursprungligt Y-värde och

predikerat Y-värde. Om R2 och Q2-värdena för de slumpmässigt permuterade modellerna är jämförbara eller större än 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 (inte helt robust, men signifikant modell) och figur 16 (robust och 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 och består av 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 många fall är hög, ofta över 50 %, men med något undantag bara 9 %.

Endast 1 mätvärde finns registrerat för utsläpp av <sup>152</sup>Eu till luft och för <sup>154</sup>Eu finns inget. För utsläpp av <sup>54</sup>Mn till luft finns totalt 8 registreringar av mätdata, varav 4 av 8 är gjorda under den undersökta perioden 1981-2012. Då multivariatanalysen kommer att göras av mätdata från den tidsperiod där det finns tidsmässigt överlappande data av utsläpp och av aktivitet i omgivningsvariabler (dvs. 1981-2012), kommer dessa tre ovan nämnda nukliderna inte att kunna bidra med någon information vid analysen av luftrelaterade variabler, då andelen ”missing data” är alldeles för hög. I tabell 1 visas utsläppsdelen av datamatriken. Det syns att de tre ovan nämnda nukliderna saknar väldigt mycket mätdata för utsläpp till luft (grön). Däremot är utsläppen till vatten (blå) bättre representerade, vilket då ökar möjligheterna att finna regressionsmodeller mellan dessa och aktiviteten i några omgivningsvariabler.

**Tabell 1.** Data för utsläpp till luft (grön) och vatten (blå) för de 6 studerade nukliderna [Bq].

År	Utsläpp till luft						Utsläpp till vatten					
	Co-60	Cs-134	Cs-137	Eu-152	Eu-154	Mn-54	Co-60	Cs-134	Cs-137	Eu-152	Eu-154	Mn-54
1977	4,07E+10	7,67E+10	1,59E+11			5,40E+09	3,98E+10	3,71E+10	1,65E+11			2,97E+09
1978	5,18E+10	1,46E+11	2,22E+11			4,80E+09	4,99E+10	7,29E+10	2,20E+11			2,81E+11
1979	6,66E+10	1,08E+11	2,04E+11			3,00E+09	6,64E+10	7,85E+10	2,04E+11			2,45E+09
1980	3,41E+10	3,60E+10	1,37E+11			2,00E+09	3,41E+10	3,48E+10	1,04E+11			2,01E+09
1981							4,18E+10	5,13E+10	1,24E+11		8,70E+08	1,39E+09
1982							4,99E+10	6,14E+10	1,22E+11		1,20E+08	1,01E+09
1983		9,10E+05	7,57E+06				1,07E+11	3,03E+10	1,08E+11		5,90E+08	3,90E+08
1984	6,00E+06	1,83E+06	1,32E+07				9,85E+10	5,00E+10	1,50E+11			
1985	2,62E+07	3,55E+06	6,56E+07				4,25E+10	2,23E+10	8,73E+10			
1986	2,10E+06	2,65E+06	8,62E+07				1,81E+10	7,48E+09	3,40E+10			
1987	5,10E+06		1,27E+07				1,76E+10	2,27E+10	6,85E+10			
1988			7,10E+06				1,86E+09	6,98E+09	2,31E+10	9,50E+08		
1989	3,12E+07	7,54E+06	4,27E+07	3,00E+05		4,27E+06	8,81E+08	5,46E+09	1,42E+10	2,08E+10	1,02E+10	4,52E+07
1990	2,40E+06	2,34E+05	4,12E+07			1,52E+05	1,84E+09	3,24E+09	1,06E+10	2,59E+09	1,38E+09	1,33E+08
1991	4,18E+05	4,90E+04	2,11E+06			6,40E+03	2,46E+09	3,46E+09	1,69E+10	1,52E+09	8,94E+08	2,80E+08
1992	4,01E+05	9,30E+03	2,10E+06			4,64E+03	1,56E+10	2,64E+09	1,76E+10	9,65E+08	6,20E+08	1,93E+09
1993	1,90E+06	1,68E+04	2,20E+06				5,98E+10	4,36E+09	1,69E+10	6,48E+09	3,19E+09	5,39E+09
1994	1,10E+06	7,60E+03	7,80E+04				7,13E+09	4,95E+09	1,33E+10	1,42E+09	1,18E+09	2,73E+08
1995	1,43E+05	1,80E+05	2,73E+05				8,32E+09	6,68E+09	1,96E+10	6,79E+09	2,91E+09	1,44E+08
1996	7,30E+04						1,87E+10	7,45E+09	2,41E+10	1,37E+10	5,97E+09	1,10E+09
1997	9,80E+04		6,69E+03				2,68E+09	8,23E+09	1,88E+10	2,09E+09	1,72E+09	1,06E+08
1998	1,13E+05	1,10E+04	4,02E+04				1,44E+09	7,80E+09	1,42E+10	1,00E+09	1,06E+09	8,25E+07
1999	2,94E+05		3,40E+04				9,31E+08	3,86E+09	8,59E+09	3,26E+08	3,23E+08	3,84E+07
2000			6,70E+04				1,35E+10	1,60E+09	3,93E+09	1,69E+09	6,32E+08	2,19E+09
2001		6,70E+05	2,48E+06				4,92E+09	6,75E+09	1,19E+10	2,56E+08	1,33E+08	6,26E+08
2002	1,12E+05		7,59E+04				2,44E+09	2,01E+10	3,08E+10	1,50E+08	7,16E+07	4,06E+08
2003	9,82E+04	3,14E+04	1,68E+05				7,08E+09	2,83E+09	8,58E+09	2,33E+07	1,80E+07	7,67E+07
2004	2,32E+05		7,99E+03				9,18E+09	1,23E+09	4,24E+09	2,00E+08	8,45E+07	5,71E+08
2005	4,90E+04	2,50E+04	8,40E+04				2,42E+09	2,91E+08	1,33E+09	3,60E+06	4,70E+07	5,52E+07
2006	5,05E+05		1,64E+05				1,41E+09	4,12E+08	1,35E+09	1,40E+06	1,70E+07	3,90E+08
2007	9,50E+04		1,51E+05				1,10E+09	3,00E+08	9,68E+08		4,30E+06	1,70E+08
2008	1,66E+05		4,70E+04				3,86E+08	1,85E+08	1,23E+09	3,80E+05	1,20E+06	2,20E+07
2009	1,10E+05		8,47E+04				2,04E+09	1,02E+08	7,31E+08			2,90E+08
2010	3,80E+05	1,10E+03	3,39E+04				1,11E+09	9,20E+07	7,65E+08			2,21E+08
2011	4,20E+05		1,58E+05				2,13E+08	1,40E+08	4,84E+08			1,20E+06
2012	1,60E+06		1,64E+05				1,31E+08	1,20E+08	4,43E+08	1,90E+05		5,10E+04



Variabler med färre än 7 observationer, av totalt 32, ("missing data" >80 %) har direkt uteslutits för vidare analys, då sådana variabler mest genererar brus i modellerna.

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. Utsläppen till vatten följer oftast samma vattenströmmar och verkar vara mer förutsägbara.

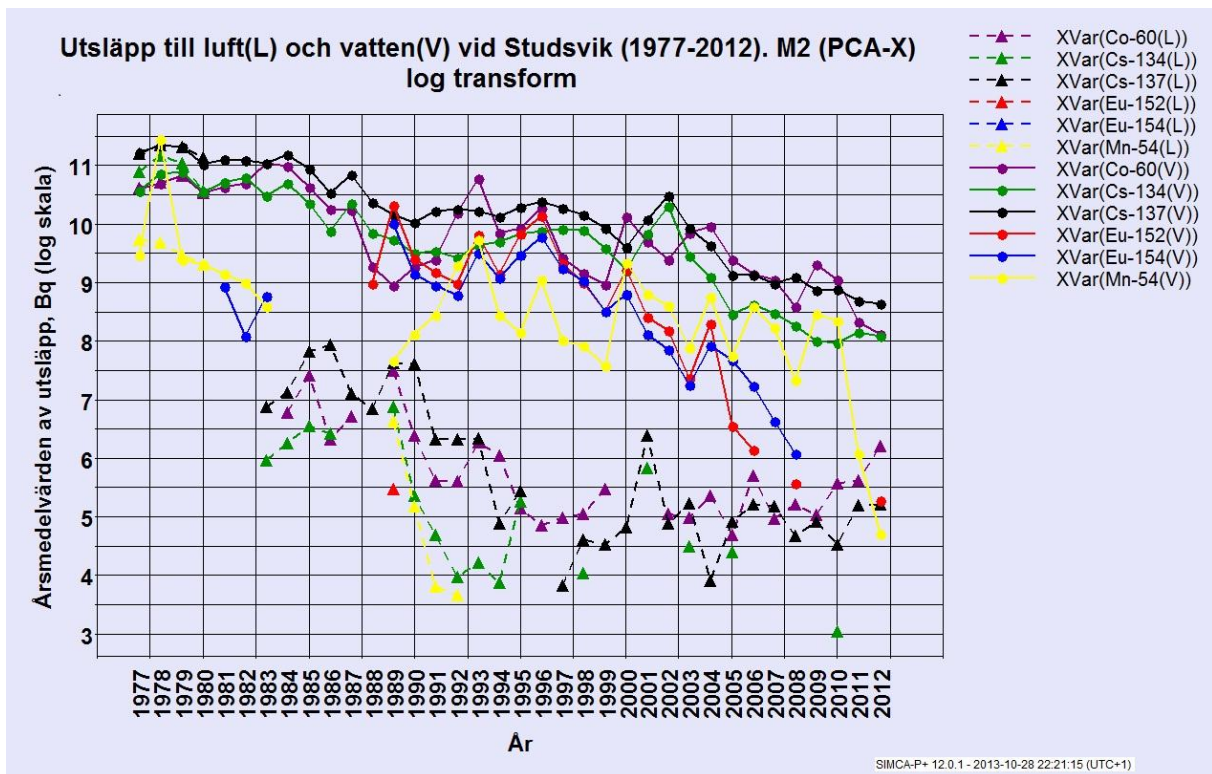
Det kan inte uteslutas att utsläpp av radioaktiva nuklider från våra grannländer på andra sidan av Östersjön också är en "felkälla" som kan påverka möjligheterna att finna signifikanta korrelations samband för de kontrollerade utsläppen från Studsvik. I motsvarande tidigare analyser av utsläppen från de kärntekniska anläggningarna i Oskarshamn och i Forsmark verkar speciellt utsläppen av  $^{137}\text{Cs}$  blandats med nedfallet från Tjernobyli. I figur 1 nedan syns ingen uppenbar "Tjernobyleffekt" i närheten av 1986.

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 mellan (2-4) tiopotenser högre än utsläppen till luft. Det ses också en nedåtgående trend i utsläppen, där det kanske är tydligast för utsläppen till vatten. Det framgår också att de båda Cs-nukliderna har väldigt likartade variationer i tiden. Utsläppen till vatten av de båda Eu-nukliderna och <sup>54</sup>Mn har sedan mitten av 2000-talet minskat med ungefär 3 tiopotenser.

## 5.2. PCA – alla data

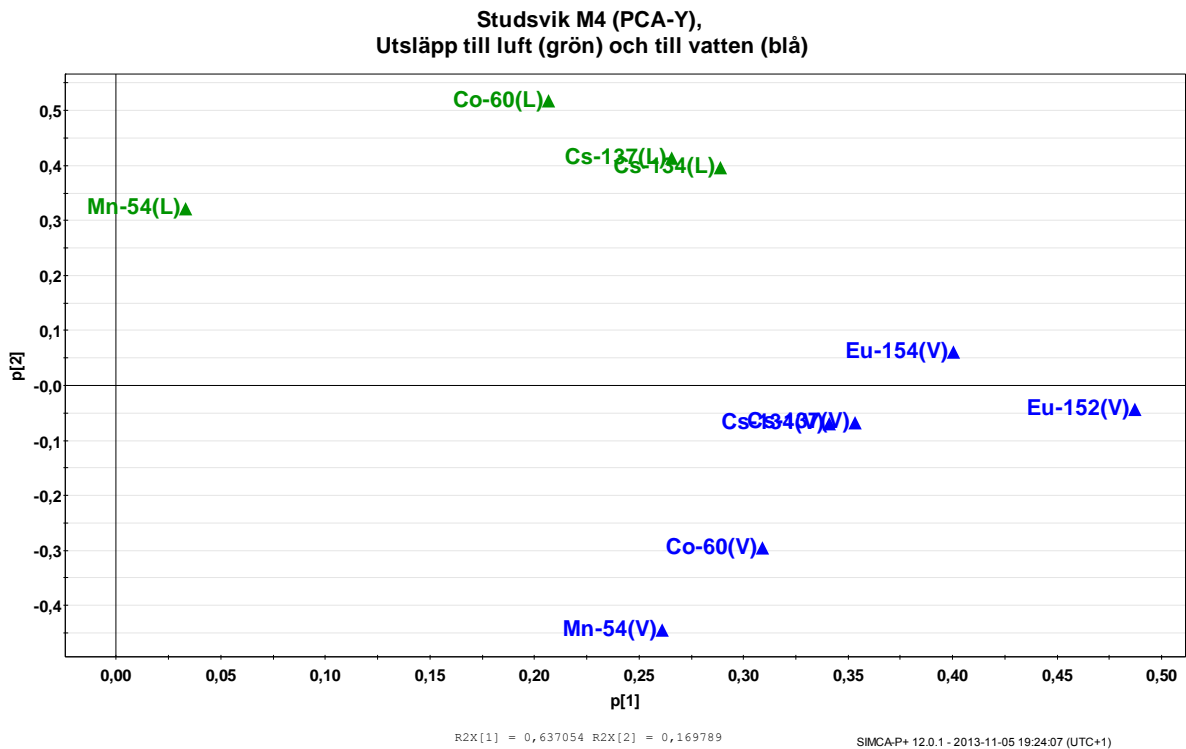
En PCA-modell på samtliga data, undantaget utsläpp mellan 1977-1981, har gjorts för perioden (1981-2012), (169 variabler och 32 observationer) utan någon förbehandling av data. Detta resulterade i en modell, M1, med en förklaringsgrad,  $Q^2=0,112$ , dvs. 11,2 % av variationerna i data kan förklaras. 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  $Q^2=0,291$ , dvs. förklaringsgraden har ökat till 29,1 %. Den logaritmiska transformeringen resulterade således i en modell som har nästan 3 gånger bättre förklaringsgrad.

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

## 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.



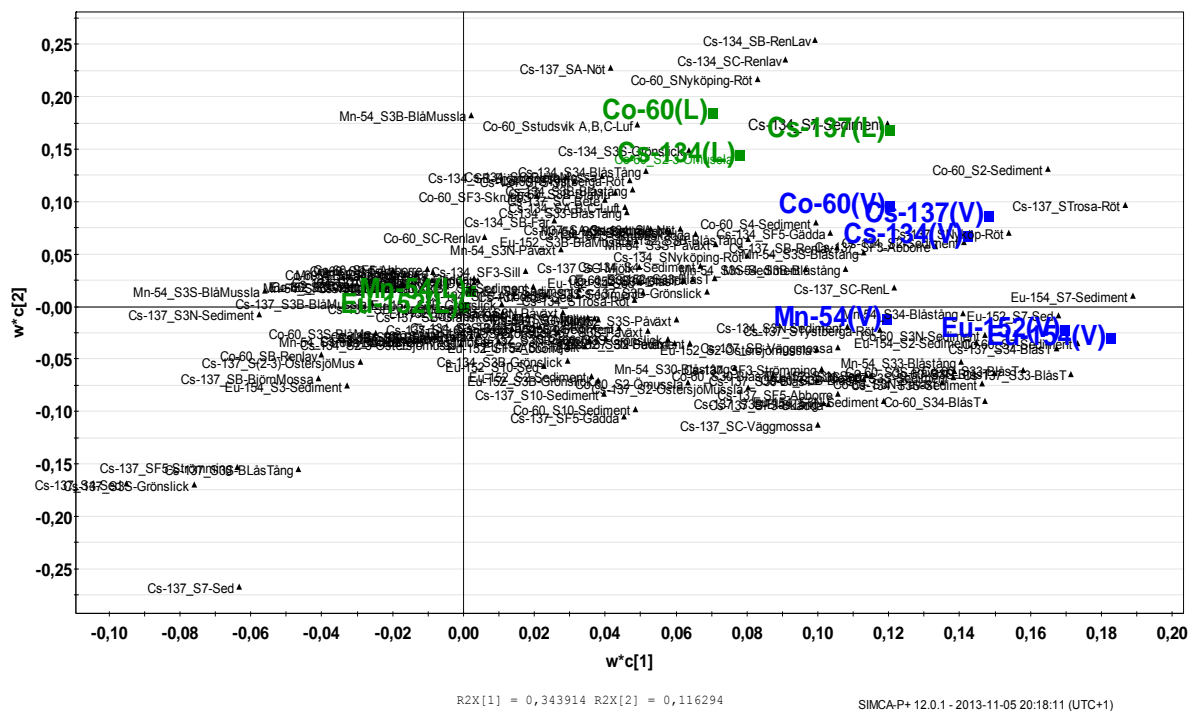
**Figur 2.** "Score plot" av enbart utsläppen till luft (gröna) och vatten (blå).

En indikation om korrelationsmönstret ses i figur 2. Det är en PCA som är gjord på enbart utsläppen, Y-variablerna, där andelen "missing data" är relativt låg, undantaget utsläppen av <sup>152</sup>Eu och <sup>154</sup>Eu till luft som har respektive en eller helt saknar mätdata. Det är viktigt att komma ihåg att korrelationsmönstret, i figur 2 för modell M4, speglar årsmedelvärdena av utsläppen till luft och vatten utan hänsyn taget till omgivningsdata. Om det finns en koppling mellan de olika utsläppen skulle observationerna ligga nära varandra. Exempelvis skulle **Co-60 (V)** ligga nära **Co-60(L)**, vilket inte är fallet i denna figur. Luft- och vattenutsläppen är klart separerade i figur 2. Vidare ses, att t.ex. de båda Cs-nukliderna har väldigt lika utsläppsmönster, både till luft och vatten. Mn-54(L) korrelerar inte med något annat utsläpp på grund av den stora andelen "missing data" (skulle egentligen inte ingått i analysen då andelen "missing data" = 88 %). Således föreligger inte någon betydande korrelation mellan utsläppen till luft och vatten, i det föreliggande datasetet, varför den fortsatta analysen kommer att göras separat för luft- respektive vattenrelaterade variabler per nuklid.

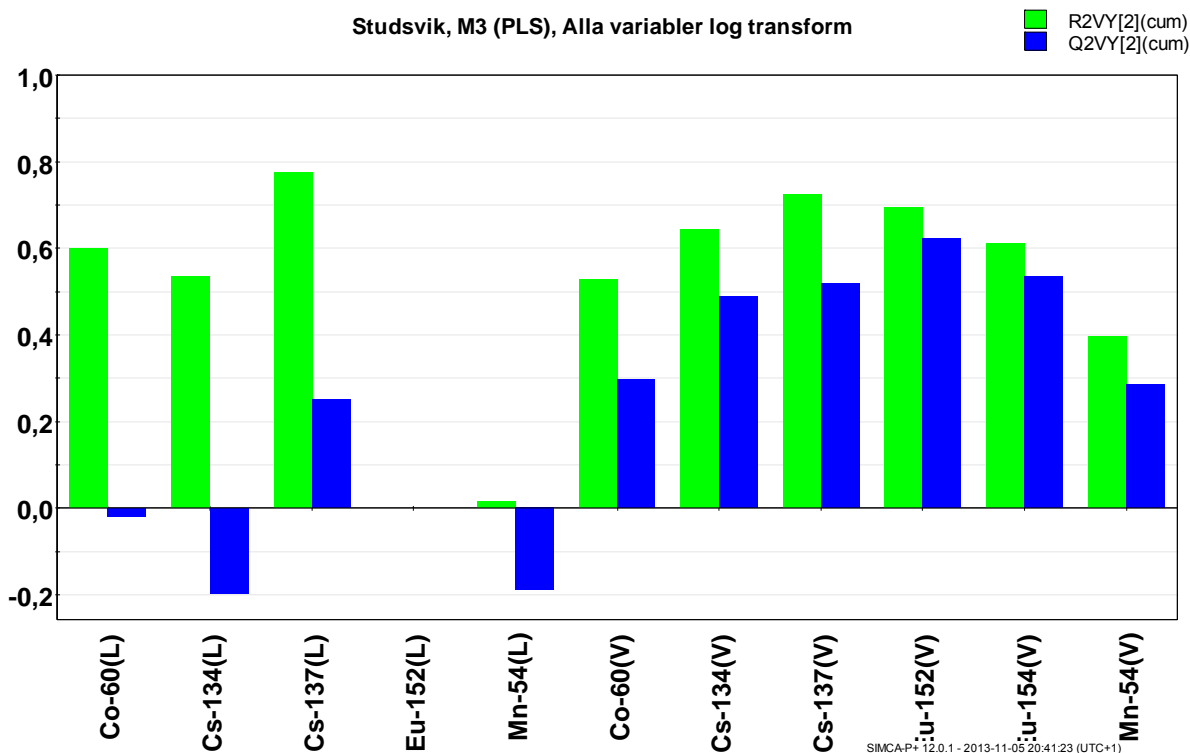
## 5.4. PLS – alla data

I figur 3 nedan visas en "loading plot" (variabelkarta) där samtliga utsläpp behandlats som Y-variabler och samtliga omgivningsvariabler (svarta) som X-variabler. Modell M3 är inte signifikant, men korrelationsmönstret, grönt till blått, av Y påminner om det som erhöles med PCA i figur 2. De svarta omgivningsvariablerna ligger alldeles för tätt för att någon detaljerad information skall kunna utvinnas. 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 (negativa score-värden i figur 3). <sup>152</sup>Eu och <sup>54</sup>Mn ligger nära origo, vilket inte går att se i denna figur, och det betyder att det inte verkar finnas någon korrelation mellan utsläpp och aktiviteter i omgivningsvariablerna för dessa nuklider. Den fortsatta nuklidvisa analysen får utvisa om detta är en korrekt observation när signifikanta modeller eftersträvas och studeras.

Studsvik, M3 (PLS). Alla variabler log transform



**Figur 3.** "Loading plot" där samtliga variabler har använts. Blå=utsläpp till vatten, grön=utsläpp till luft och svart=omgivningsvariabler.



**Figur 4.** En översikt av sambanden mellan X och Y för modell M3. Gröna staplar visar hur väl data kan anpassas (R2) och blå staplar visar prediktionsförmågan (Q2).

I figur 4 visas en översikt av hur väl utsläppen av de olika nukliderna kan beskrivas med hjälp av omgivningsvariablerna. Även om modell M3 inte är signifikant så finns en indikation om den förväntade prediktionsförmågan för de enskilda nukliderna. Således ser det i detta tidiga skede av analysen ut som att <sup>137</sup>Cs skulle kunna ge en signifikant regressionsmodell för luftrelaterade variabler. Modellbyggandet för de vattenrelaterade variablerna ser betydligt gynnsammare ut, då både anpassning till data, R2 (grön), och prediktionsförmåga, Q2 (blå), har relativt höga värden. Q2 är alltid lägre än R2 men kan inte vara negativ. Det bör understrykas att antalet variabler fortfarande är stort, varför icke kausala korrelationer sannolikt finns med i modellerna. Viktigt också att notera att för modell M3 är analysen ännu inte uppdelad mellan luft och vattenvariabler.

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 (se punkt 4. Analysverktyg).

## 5.6. PLS – per nuklid

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

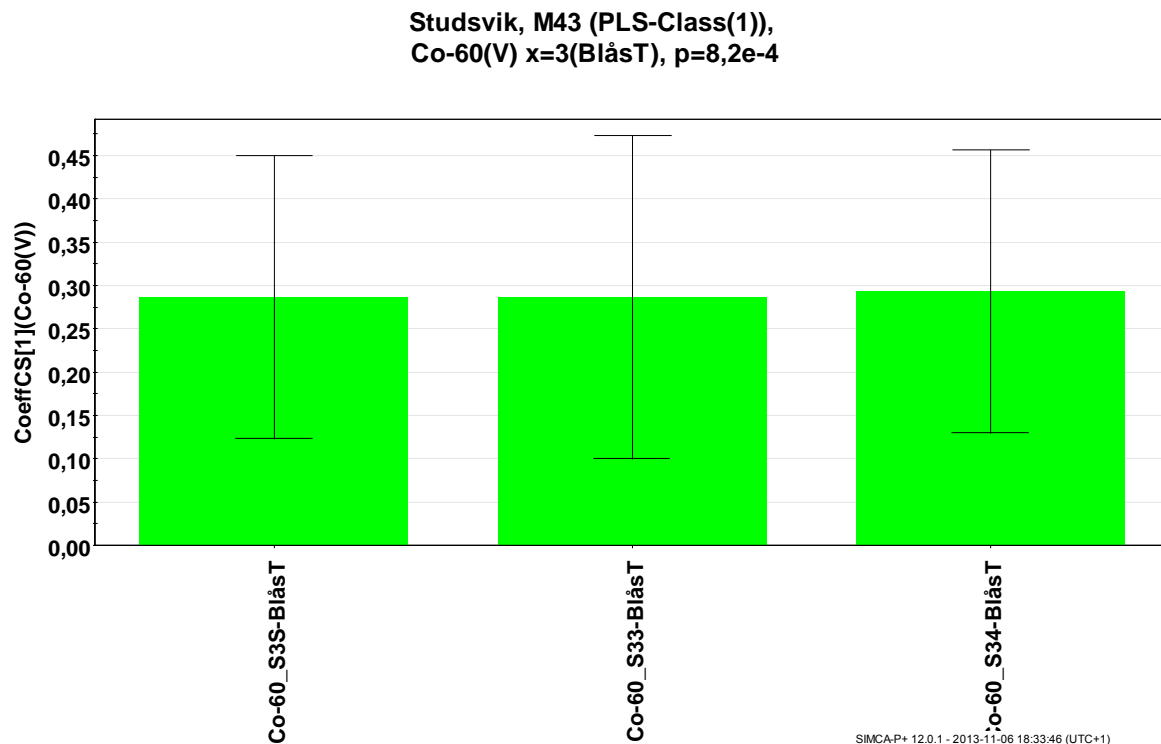
### 5.6.1. $^{60}\text{Co}$

#### Utsläpp till luft

Det går inte att finna någon signifikant modell, enligt CV-Anova-testet, som relaterar utsläpp av  $^{60}\text{Co}$  till luft med luftrelaterade omgivningsvariabler. Om en modell är signifikant skall CV-Anova-testet returnera ett  $p \leq 0,05$ , om den inte är strikt signifikant ligger p i intervallet 0,05-0,1 och om den är icke signifikant så är  $p > 0,1$ . Endast 10 luftrelaterade omgivningsvariabler finns registrerade för  $^{60}\text{Co}$ , och den lägsta andelen "missing data" är så hög som 59 %.

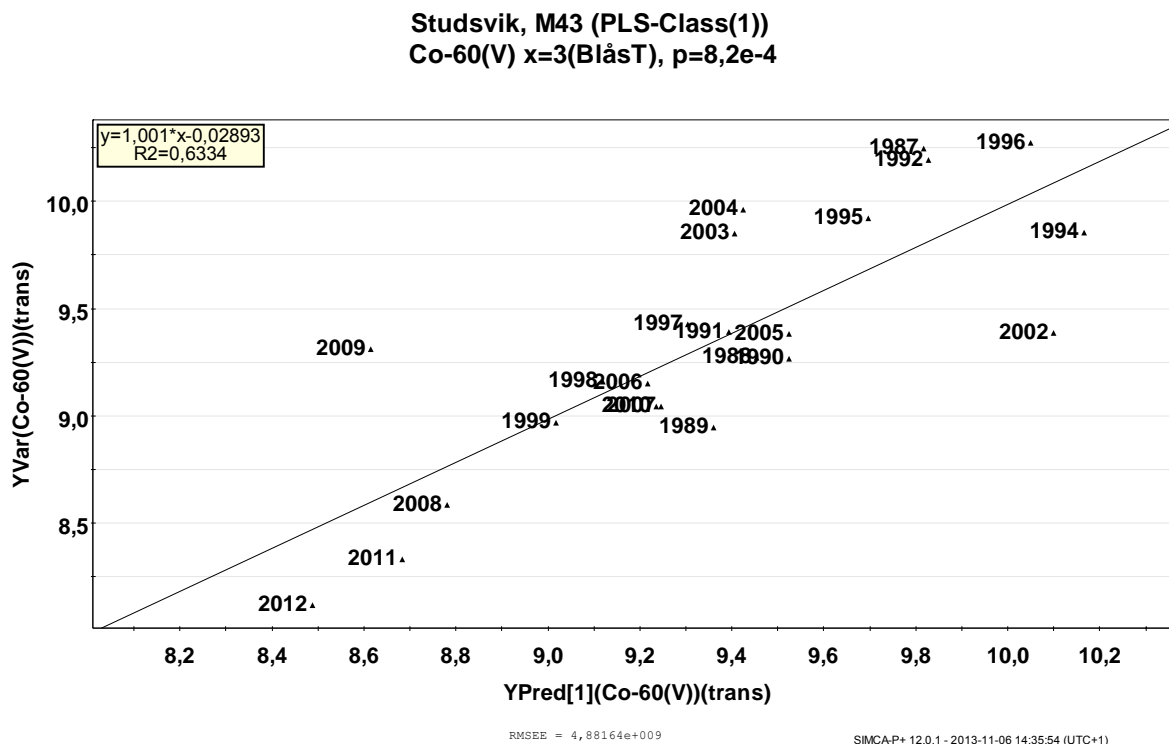
#### Utsläpp till vatten

En signifikant modell, M43 ( $p = 8,2 \cdot 10^{-4}$ ) med tre signifikanta variabler erhöles med ett  $Q^2 = 0,509$ . Detta betyder att variationer i normala utsläpp till vatten av  $^{60}\text{Co}$ , med de i figur 5 nedan visade omgivningsproverna, kan predikteras till ca 51 %.



**Figur 5.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{60}\text{Co}$  till vatten med 3 signifikanta variabler för modell M43, som är signifikant.

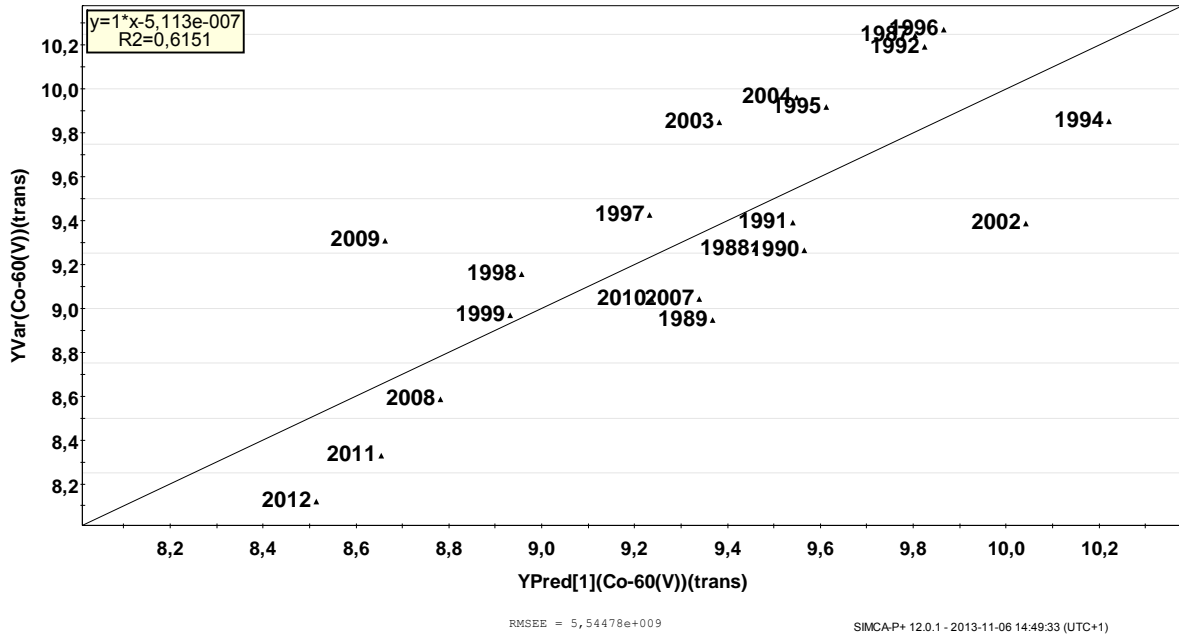
Nedan i figur 6, är observerat utsläpp av  $^{60}\text{Co}$  till vatten, avsatt mot predikterat utsläpp. För de flesta observationerna av utsläpp, som sträcker sig över två tiopotenser, är överensstämmelsen relativt bra. Observera att logaritmisk skala används i figuren.



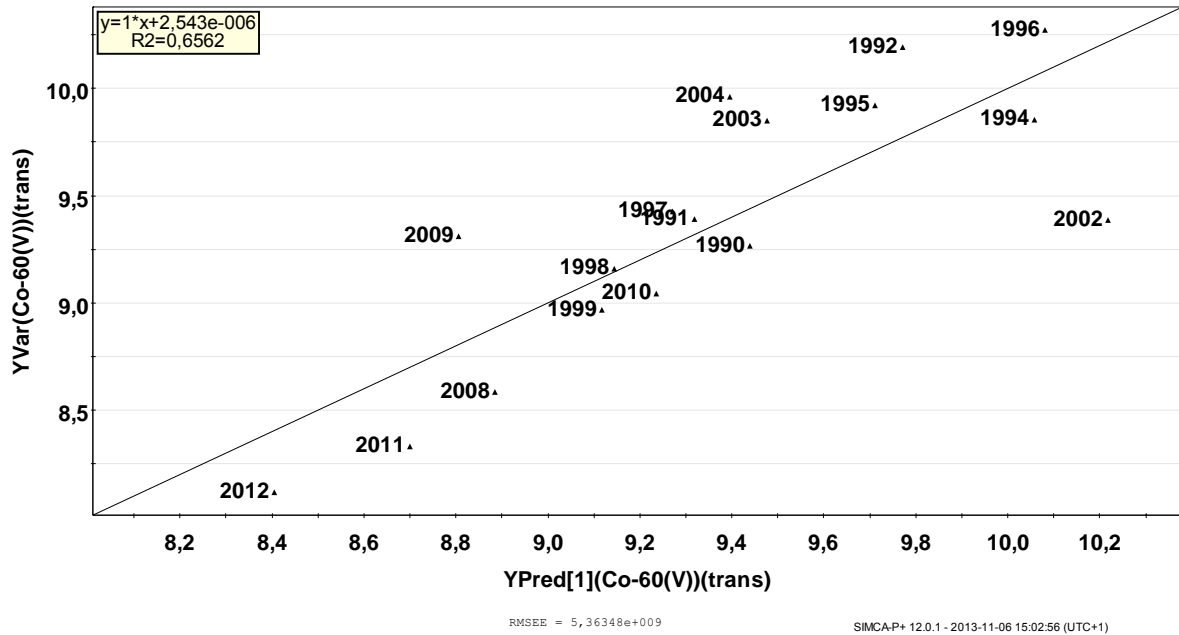
**Figur 6.** Observerat mot predikterat utsläpp av  $^{60}\text{Co}$  till vatten. Prediktionen är gjord med modell M43 med tre variabler. Observera log-skala.

En ytterligare förenkling av modellen, M43, kan göras till tre likvärdiga modeller (M44, M45 och M46) med en omgivningsvariabel. Det framgår av figur 5 att Blåstång vid de tre mätstationerna har likartade storlekar på regressionskoefficienterna och tillhörande osäkerhetsintervall. I figur 7 nedan kan inga större skillnader ses i observerad aktivitet mot predikterad för dessa tre en-variabelmodeller. Det skiljer något i prediktionsförmåga, Q2, som kan ses i tabell 2.

Studsvik. M44 (PLS-Class(1))  
Co-60(V) x=1(S3S-BlåST), p=4,5e-4

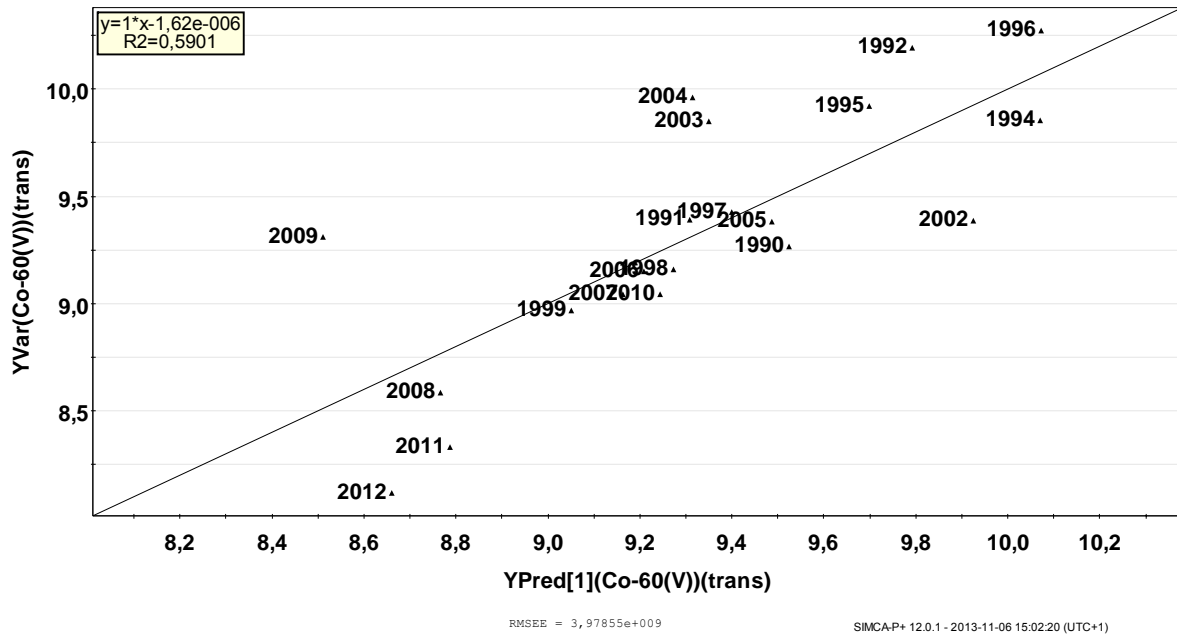


Studsvik, M45 (PLS-Class(1)),  
Co-60(V) x=1(S33-BlåST), p=1,2e-3





Studsвик, M46 (PLS-Class(1)),  
Co-60(V) x=1(S34-Blåst), p=6,8e-3



**Figur 7 a, b och c.** Observerat mot predikerat utsläpp av  $^{60}\text{Co}$  till vatten. Prediktionen är gjord med modell: a) M44 (med x-variabel  $^{60}\text{CO\_S3S-Blåstång}$ ), b) M45 (med x-variabel  $^{60}\text{CO\_S33-Blåstång}$ ) och c) M46 (med x-variabel  $^{60}\text{CO\_S34-Blåstång}$ ). Observera log-skala.

**Tabell 2.** Sammanställning av prediktionsförmåga, Q2, och modellsignifikans, p, för 3 en-variabelmodeller för utsläpp av  $^{60}\text{Co}$  till vatten.

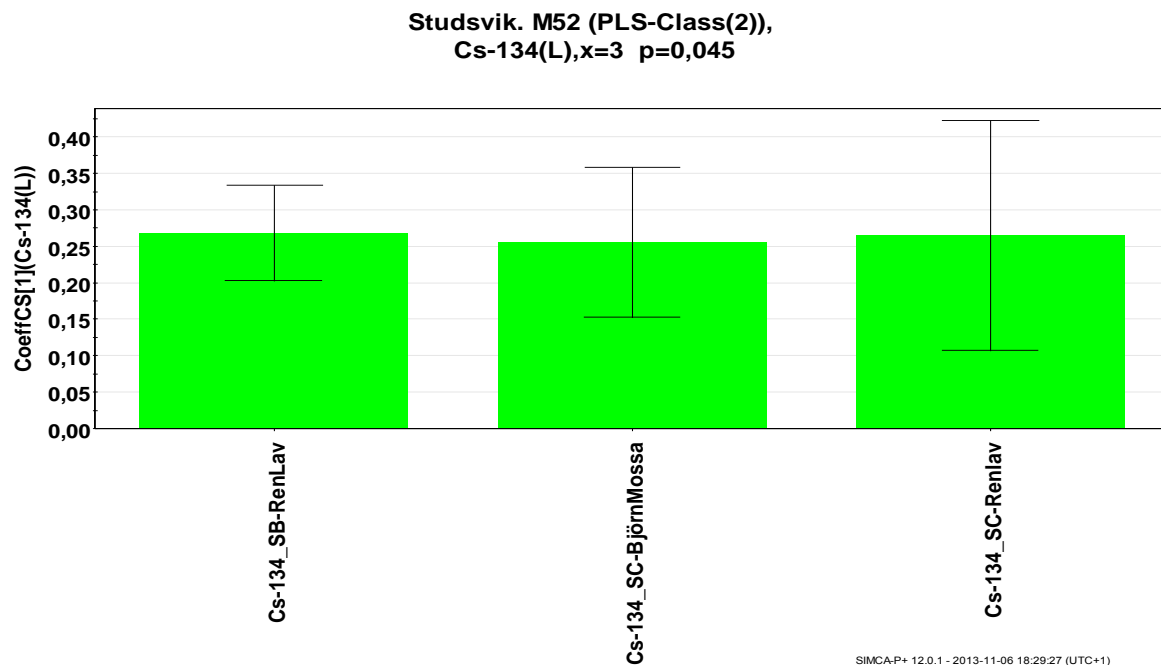
Modell	Mätstation	Provslag	Q2	p
M44	3S	Blåstång	0,575	$4,5 \cdot 10^{-4}$
M45	33	Blåstång	0,617	$1,2 \cdot 10^{-3}$
M46	34	Blåstång	0,444	$6,8 \cdot 10^{-3}$

Modell M45 har som synes den bästa prediktionsförmågan av dessa tre modeller. Alla möjliga en-variabelmodeller är inte testade, men de här redovisade modellerna var de som "föll ut" efter förenklingen som finns beskriven under punkt 4.3. Analysgång.

## 5.6.2. $^{134}\text{Cs}$

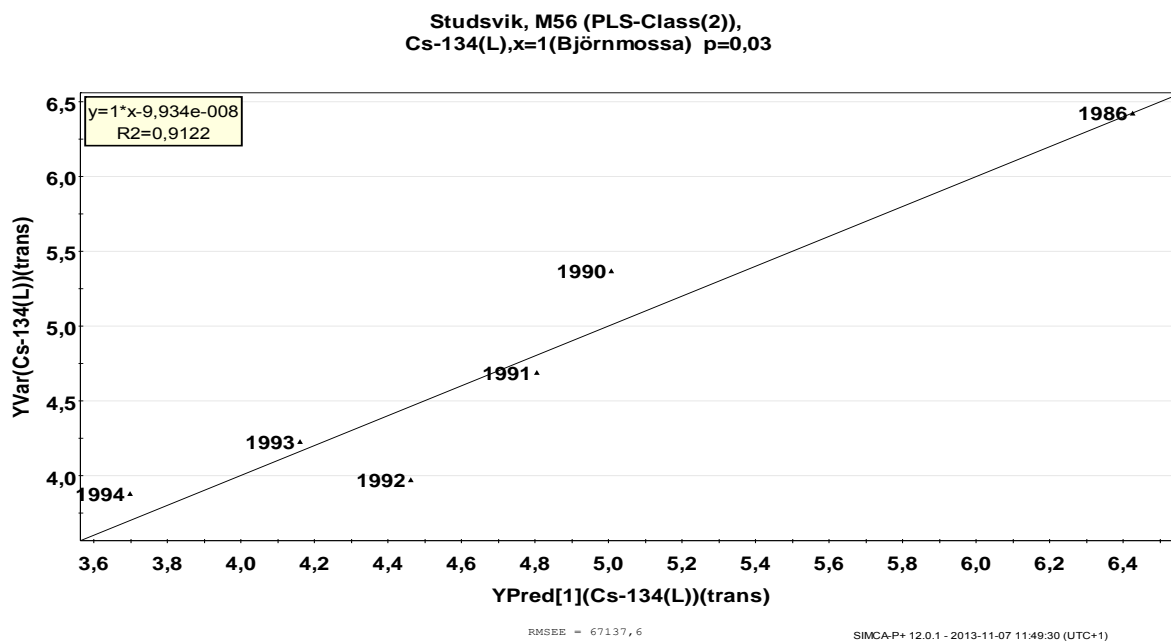
### Utsläpp till luft

En signifikant modell, M52, med 3 signifikanta variabler kan ses i figur 8. Prediktionsförmågan är  $Q^2=0,644$  och modellsignifikansen  $p=0,045$ .



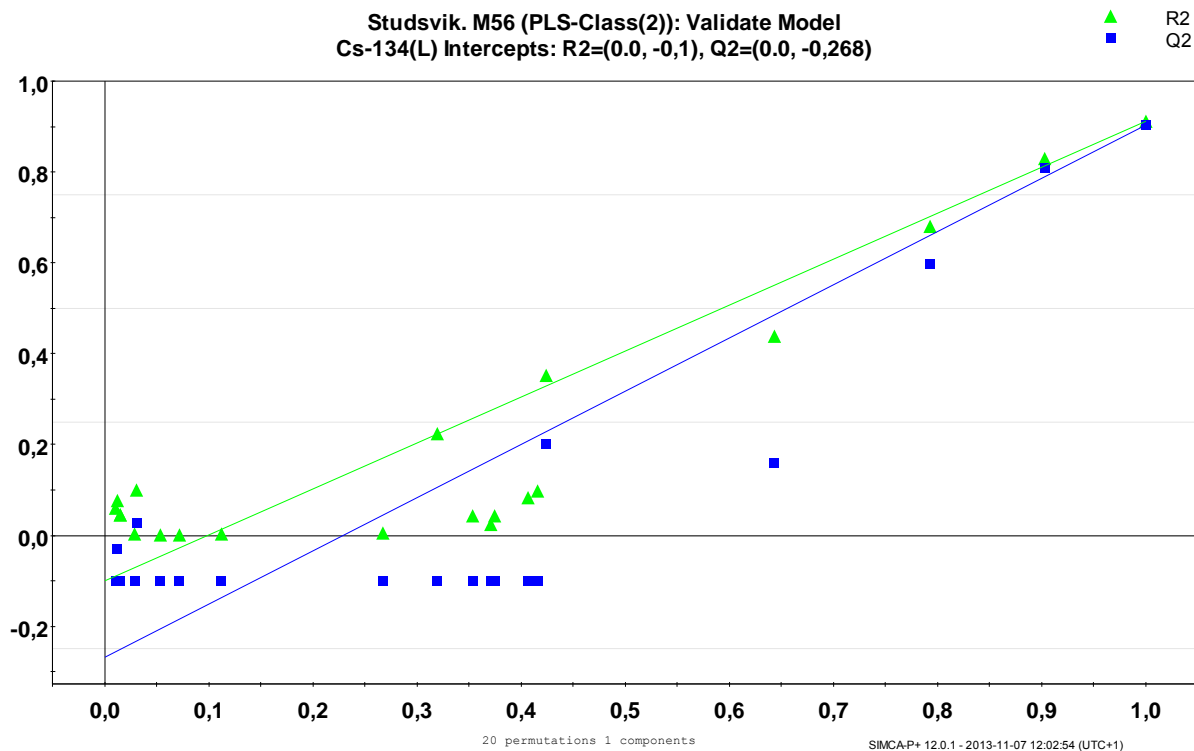
**Figur 8.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{134}\text{Cs}$  till luft med 3 signifikanta variabler för den signifikanta modellen M52.

En vidare förenkling av M52 leder fram till att det endast med variabeln  $^{134}\text{Cs\_SC-BjörnMossa}$  går att få en signifikant modell, M56. Detta trots att osäkerheten för  $^{134}\text{Cs\_SB-Renlav}$  verkar vara minst av dessa 3 variabler. Prediktionsförmågan för M56 är extremt hög,  $Q^2=0,904$  och  $p=0,030$ , vilket medför att drygt 90 % av utsläppen kan predikteras bara genom att mäta på BjörnMossa vid station C.



**Figur 9.** Observerat mot predikerat utsläpp av <sup>134</sup>Cs till luft. Prediktionen är gjord med modell M56 (<sup>134</sup>Cs\_SC-Björnmossa). Observera log-skala.

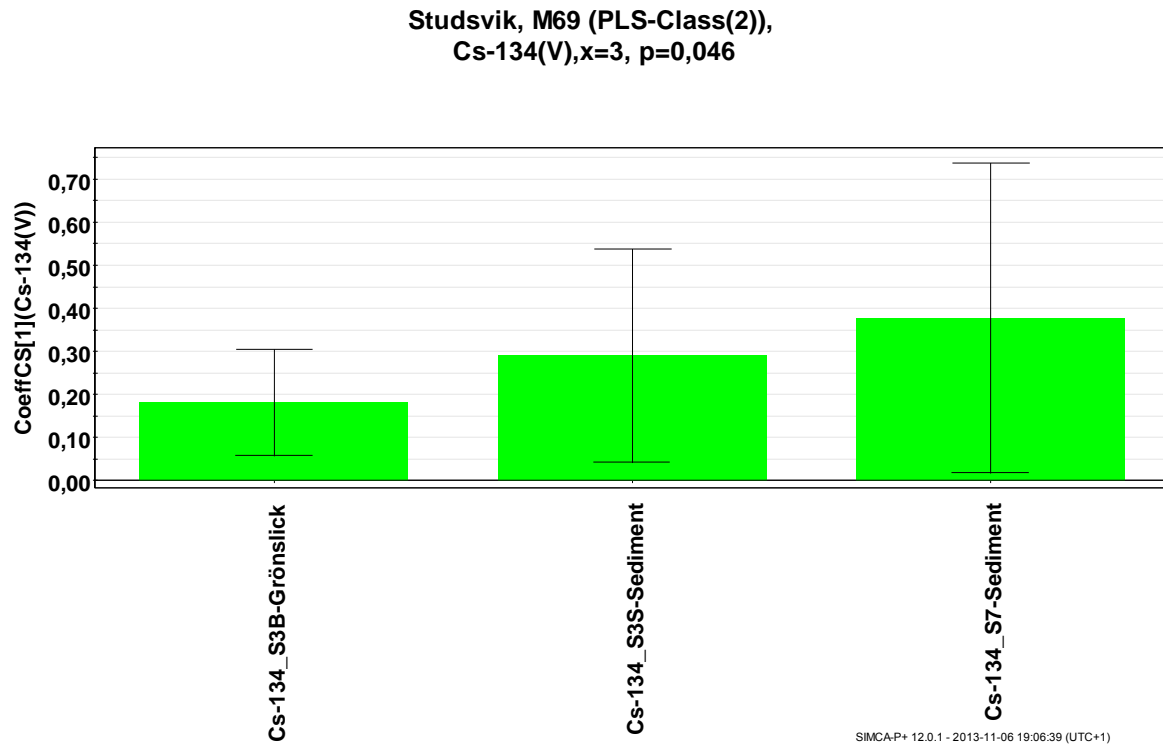
Observera att antalet observationer bara är 6 (inkluderat 1986), varför modell M56, skall användas med försiktighet. Se figur 10 nedan, "Validate Model", där 20 slumpvisa permutationer av Y-vektorn resulterar i modeller som har nästan lika höga R2 och Q2, som originalmodellen vid korrelationskoefficient=1, vilket minskar modellens validitet.



**Figur 10.** R2 och Q2 (vertikala axeln) avsatta för 20 slumpvisa permutationer av Y-värden mot korrelationskoefficienter (horisontella axeln) mellan ursprunglig Y och permuterat Y.

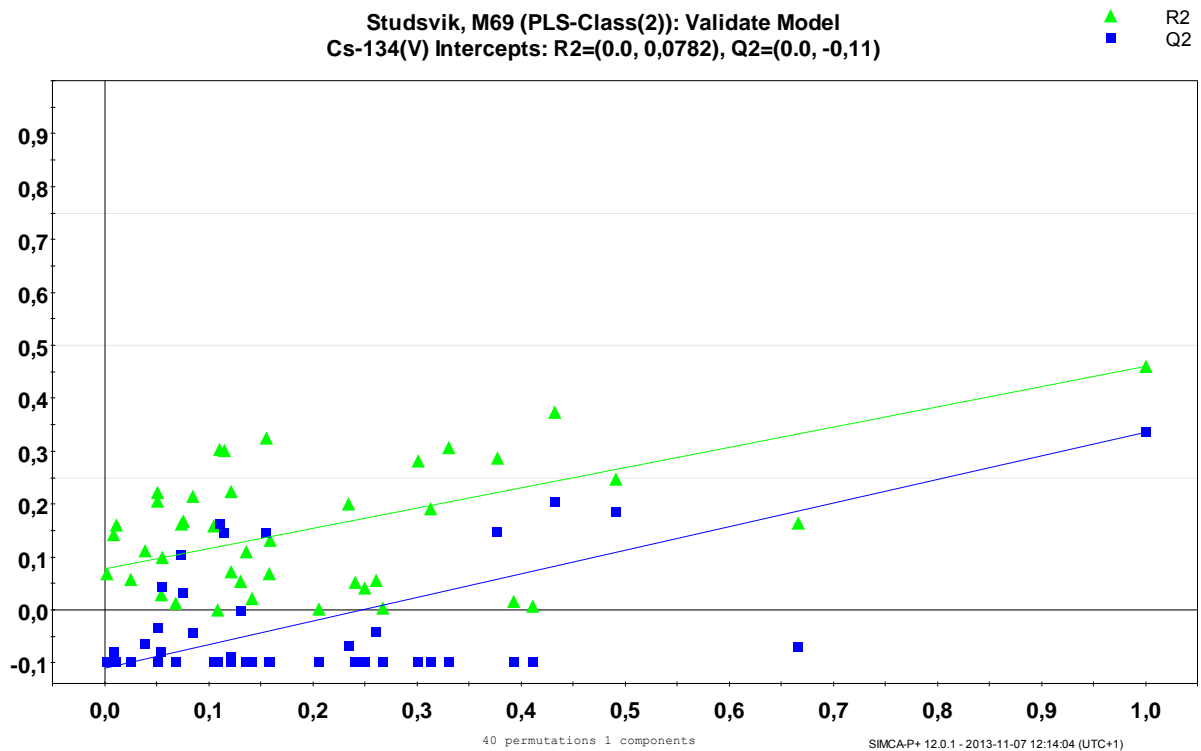
## Utsläpp till vatten

Mätningarna av utsläpp av  $^{134}\text{Cs}$  till vatten leder till att en signifikant modell, M69, kan beräknas och regressionskoefficienterna visas i figur 11. Modellen har relativt låg prediktionsförmåga,  $Q^2=0,336$  och modellsignifikansen  $p=0,046$ .



**Figur 11.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{134}\text{Cs}$  till vatten med 3 signifikanta variabler för den signifikanta modellen M69.

Att modell M69 inte är så robust kan anas dels av att p-värdet är nära signifikansgränsen  $p=0,05$  och dels av att osäkerhetsintervallen per variabel är relativt stora, men det kan också visas med hjälp av "Validate Model" i figur 12. I denna figur är  $R^2$  och  $Q^2$  (vertikala axeln) beräknade för 40 slumpvisa permutationer av Y-vektorn, avsatta mot korrelationskoefficienten. Korrelationskoefficienten=1 motsvarar originalmodellen.  $R^2$  och  $Q^2$  för några av permutationerna hamnar inte alltför långt ifrån originalet, vilket tolkas som att modellen, i detta fall visserligen är signifikant, men inte så robust.



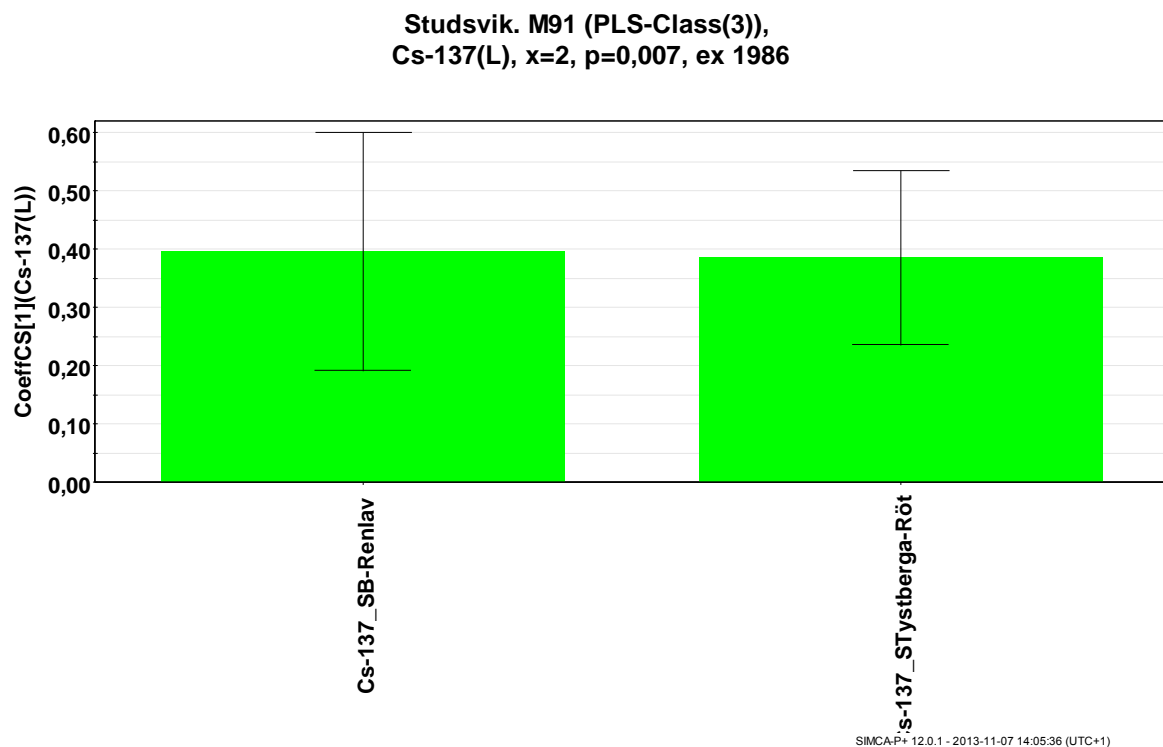
**Figur 12.** R2 och Q2 (vertikala axeln) avsatta för 40 slumpvisa permutationer av Y-värden mot korrelationskoefficienter (horisontella axeln) mellan ursprunglig Y och permuterat Y.

Vidare reducering av antalet variabler resulterar inte i någon signifikant modell för utsläppen av <sup>134</sup>Cs till vatten.

### 5.6.3. $^{137}\text{Cs}$

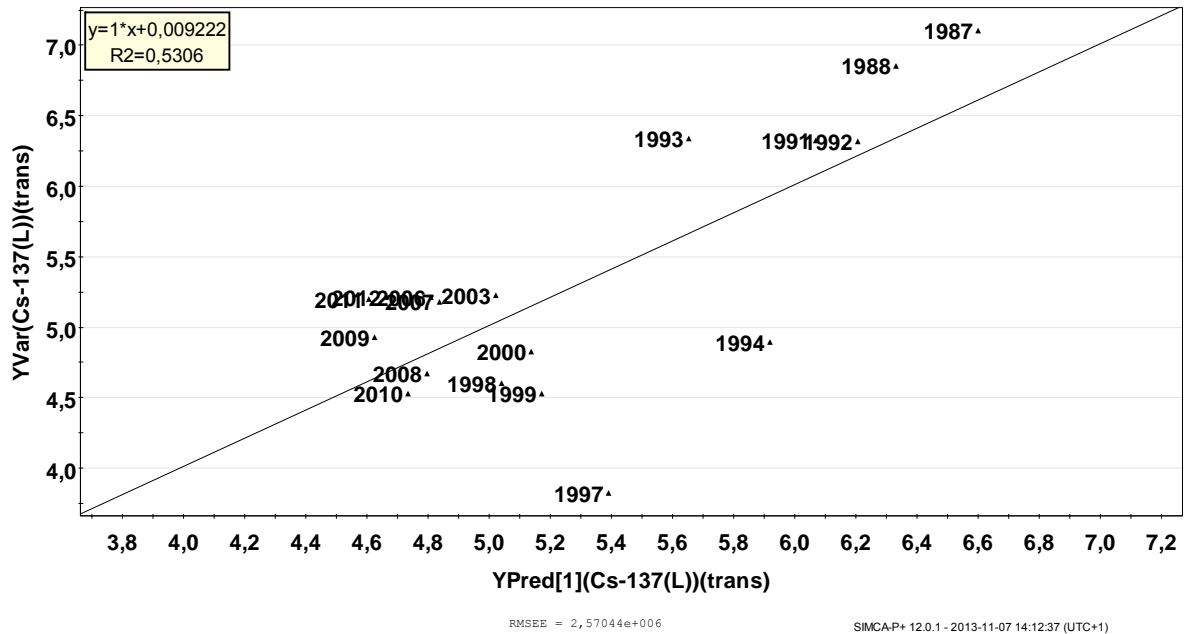
#### Utsläpp till luft

En signifikant modell, M91, med 2 signifikanta omgivningsvariabler,  $Q^2=0,484$  och modellsignifikans  $p=0,007$  kan erhållas om mätvärden under 1986 exkluderas (figur 13). Det finns en stark misstanke om att  $^{137}\text{Cs}$  från Tjernobyl, fördärvar regressionen mellan kontrollerade utsläpp till luft och uppmätta aktiviteter i luftrelaterade omgivningsvariablerna.



**Figur 13.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{137}\text{Cs}$  till luft med 2 signifikanta variabler för den signifikanta modellen M91.

Studsvik. M91 (PLS-Class(3)),  
Cs-137(L), x=2, p=0,007, ex 1986



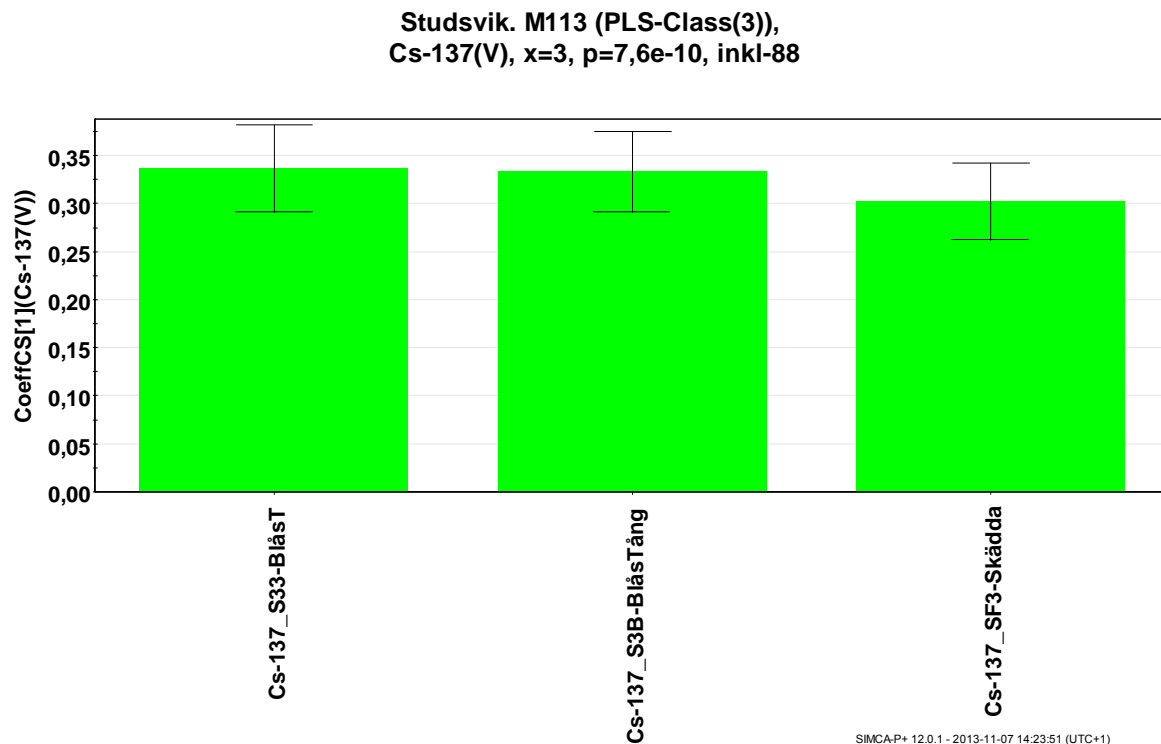
**Figur 14.** Observerat mot predikterat utsläpp av <sup>137</sup>Cs till luft. Prediktionen är gjord med modell M91 (<sup>137</sup>Cs\_SB-Renlav och <sup>137</sup>Cs\_STystberga-Röttslam). Observera log-skala.

Prediktionerna i figur 14, sträcker sig över drygt 2 tiopotenser och observationerna 1997 och 1994 ”sticker ut” lite på grund av att dessa år innehåller relativt mycket ”missing data”.

Ingen av dessa omgivningsvariabler resulterar i någon signifikant en-variabelmodell.

## Utsläpp till vatten

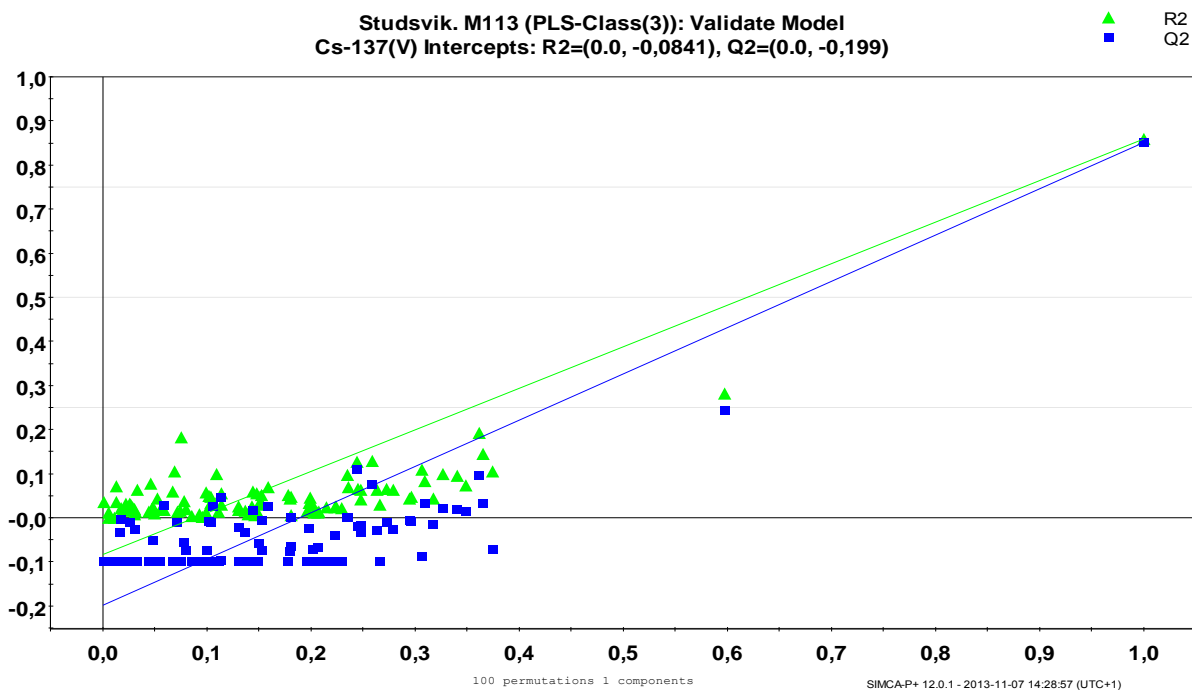
Den signifikanta modellen, M113, har 3 signifikanta omgivningsvariabler (figur 15). Prediktionsförmågan är  $Q^2=0,852$  och modellsignifikansen är  $p=7,6*10^{-10}$ .



**Figur 15.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{137}\text{Cs}$  till vatten med 3 signifikanta variabler för den signifikanta modellen M113.

De relativt snäva osäkerhetsintervallen som visas för regressionskoefficienterna indikerar att modellen är robust, vilket också modellsignifikansen  $p=7,6*10^{-10}$  vittnar om. Ytterligare belegg för att modell M113 är valid med väldigt hög prediktionsförmåga, ses i figur 16 ("validate model") där inga resultat från de 100 slumpvisa permutationerna av Y kan jämföras med det som fås ur originalmodellen.





**Figur 16.** R2 och Q2 (vertikala axeln) avsatta för 100 slumpvisa permutationer av Y-värden mot korrelationskoefficienter horisontella axeln) mellan ursprunglig Y och permuterat Y.

Var och en av dessa 3 omgivningsvariabler ger signifikanta en-variabelmodeller, som har en prediktionsförmåga väl i paritet med den för tre-variabelmodellen, M113, se tabell 3.

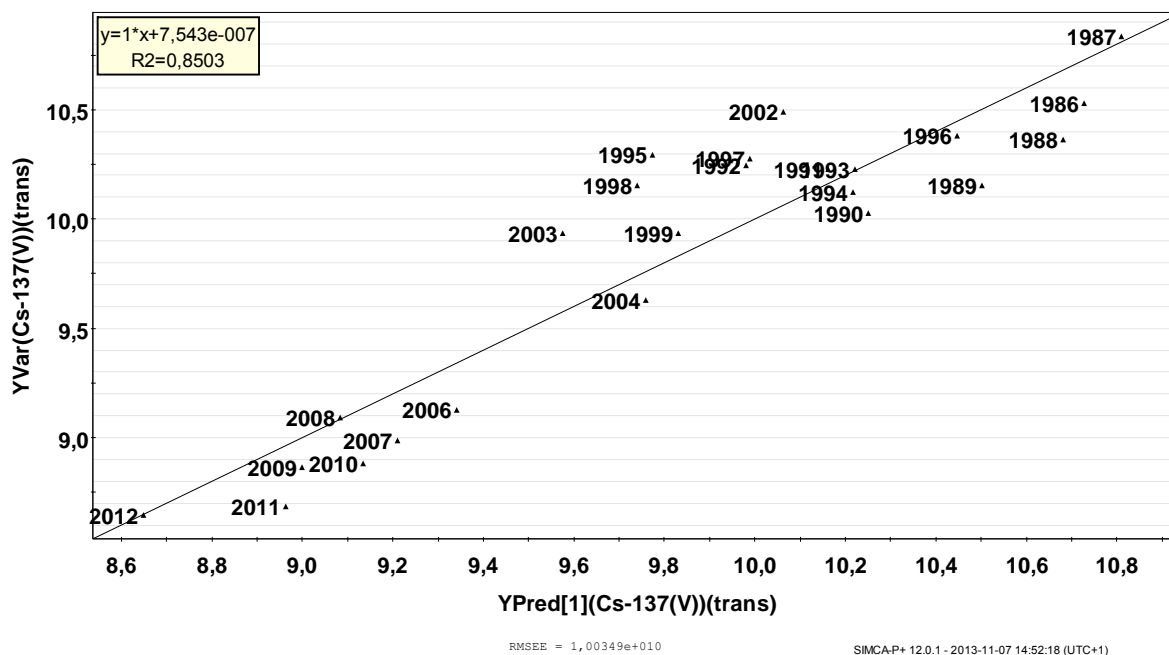
**Tabell 3.** Sammanställning av prediktionsförmåga, Q2, och modellsignifikans, p, för 3 en-variabelmodeller för utsläpp av <sup>137</sup>Cs till vatten.

Modell	Mätstation	Provslag	Q2	p
M114	33	Blåstång	0,870	$1,1 \cdot 10^{-8}$
M115	3B	Blåstång	0,845	$3,2 \cdot 10^{-9}$
M118	F3	Skrubbskädda	0,899	$3,4 \cdot 10^{-9}$

Nedan i figur 17 visas observerat mot predikerat värde för modell M115, som har den högsta modellsignifikansen av de i tabell 3. De övriga två modeller ger liknande goda överensstämmelse.

Den observante noterar att utsläppen till vatten från 1986 och efterföljande år, lite förvånande, passar väl in den aktuella modellen, M115, och "Tjernobyleffekten" verkar inte förstöra regressions sambandet för denna specifika omgivningsvariabel. Figur 17 visar också att utsläppen minskar med tiden, där senare delen av 2000-talet är 2 tiopotenser lägre än de i slutet av 1980-talet.

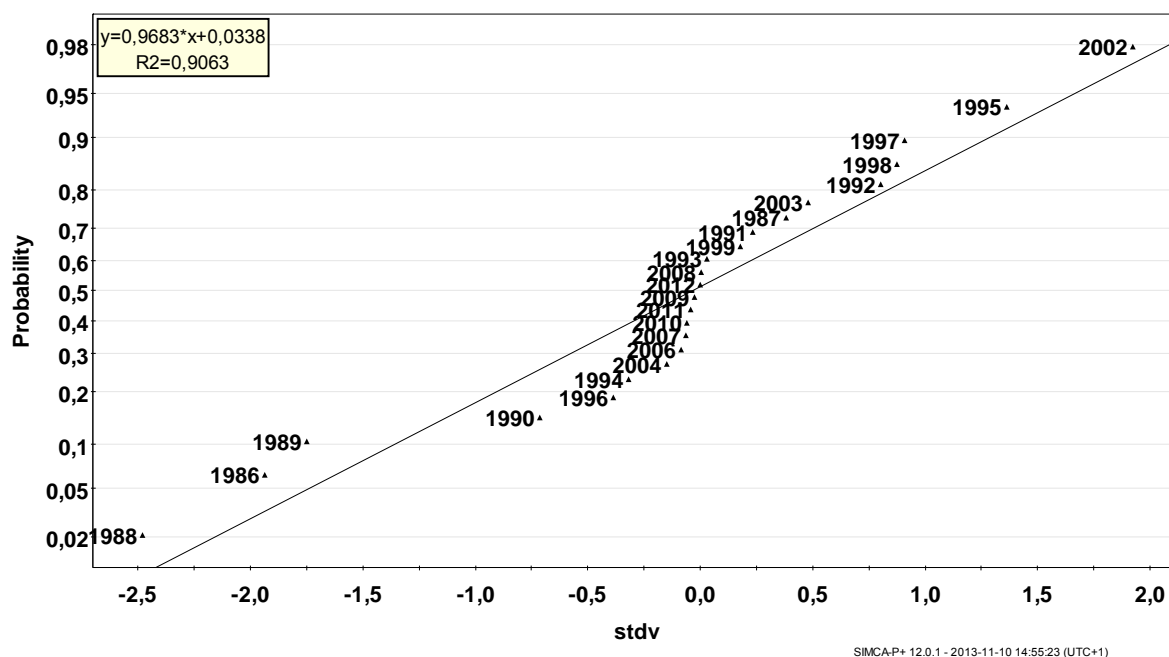
Studsvik, M115 (PLS-Class(3)),  
Cs-137(V), x=1(S3B-Blåst), p=3,2e-9, inkl-86-90



Figur 17. Observerat mot predikerat utsläpp av <sup>137</sup>Cs till vatten. Prediktionen är gjord med modell M115 (<sup>137</sup>Cs\_S3B-Blåstång). Observera log-skala.

I figur 18 nedan, normalfördelnings plot, ses att residuerna för modell M115, är normalfördelade då punkterna ansluter relativt väl till linjen.

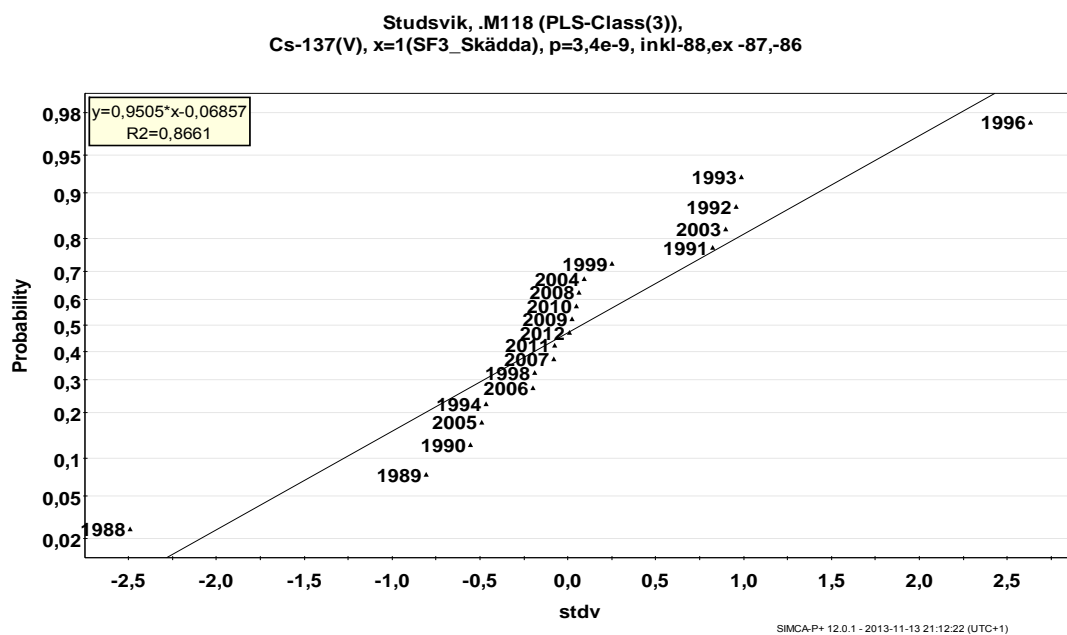
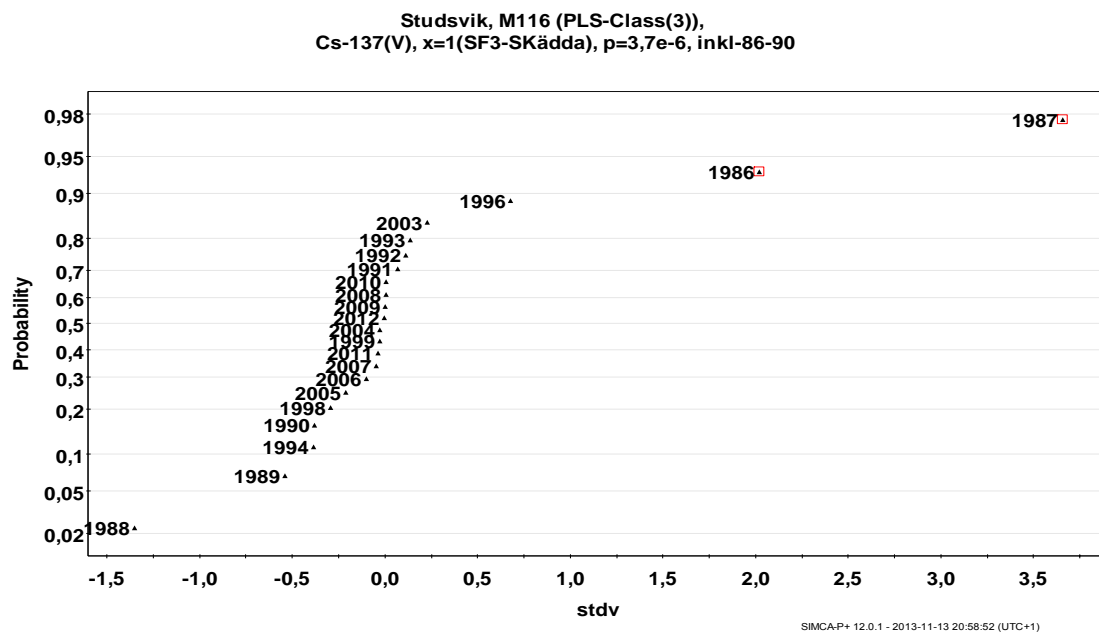
Studsvik, .M115 (PLS-Class(3)), Cs-137(V),  
x=1(S3B-Blåst), p=3,2e-9, inkl-86-90



Figur 18. Normalfördelnings plot för modell M115(S3B\_Blåstång), för vattenutsläpp av <sup>137</sup>Cs.

För att modellerna M114 och M118 skall ge normalfördelade residuer och därmed förbättrad prediktionsförmåga och högre modellsignifikans måste observationerna 1986 och 1987 exkluderas. Multivariatanalysen kan inte förklara varför observationerna 1986 och 1987 stör regressionen i några omgivningsvariabler, men inte för andra.

Exempelvis visas en normalfördelningsplot, för modell M116, figur 19 a nedan. I denna figur är det uppenbart att framförallt observationen 1987 avviker starkt från övriga observationer (tänk en linje genom huvuddelen av observationerna) och som synes är också observationen 1986 avvikande. När dessa båda observationer exkluderats fås modell M118, för vilken en normalfördelningsplot visas i figur 19 b.



**Figur 19 a och b.** Normalfördelningsplot för vattenutsläpp av <sup>137</sup>Cs för a) modell M116 (<sup>137</sup>Cs\_SF3-Skädda) och b) modell M118(<sup>137</sup>Cs\_SF3-Skädda, exklusive åren 1986-1987).

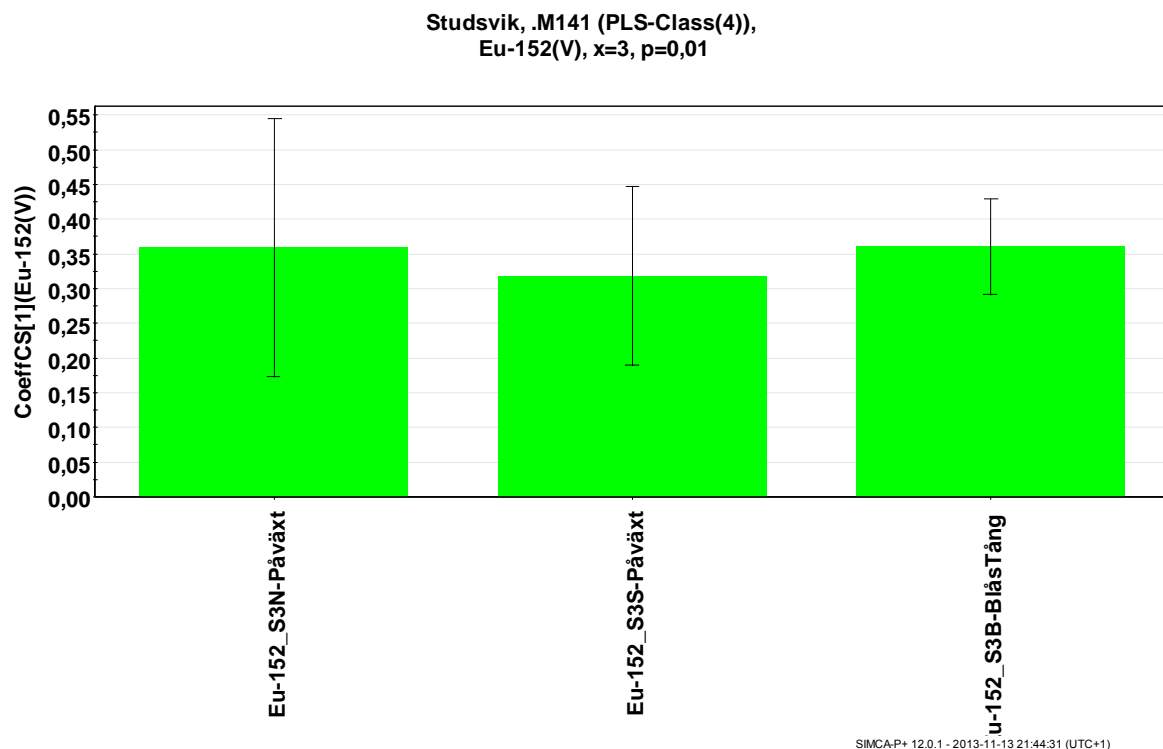
## 5.6.4. $^{152}\text{Eu}$

### Utsläpp till luft

Det finns endast 1 observation/årsmedelvärde registrerat under tiden 1977-2012 av utsläpp till luft för  $^{152}\text{Eu}$ , vilket medför att det svårligen går att finna någon korrelationsmodell mot de 3 luftrelaterade omgivningsvariablerna, som har "missing data" <80 %. Det går heller inte att se någon inbördes korrelation med PCA mellan de 3 omgivningsvariablerna.

### Utsläpp till vatten

En signifikant modell, M141, med 3 signifikanta omgivningsvariabler går att finna för utsläpp av  $^{152}\text{Eu}$  till vatten. Prediktionsförmågan är  $Q^2=0,778$  och modellsignifikansen är  $p=0,010$ .



**Figur 20.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{152}\text{Eu}$  till vatten med 3 signifikanta variabler för den signifikanta modellen M141.

Endast med omgivningsvariabeln  $^{152}\text{Eu}_\text{S3B-Blåstång}$ , som har det minsta osäkerhetsintervallet, gick det att finna 1 en-variabelmodell, M140. Prediktionsförmågan på  $Q^2=0,708$  och  $p=0,046$  är i paritet med tre-variabelmodellen M141, men är något mindre resurskrävande.

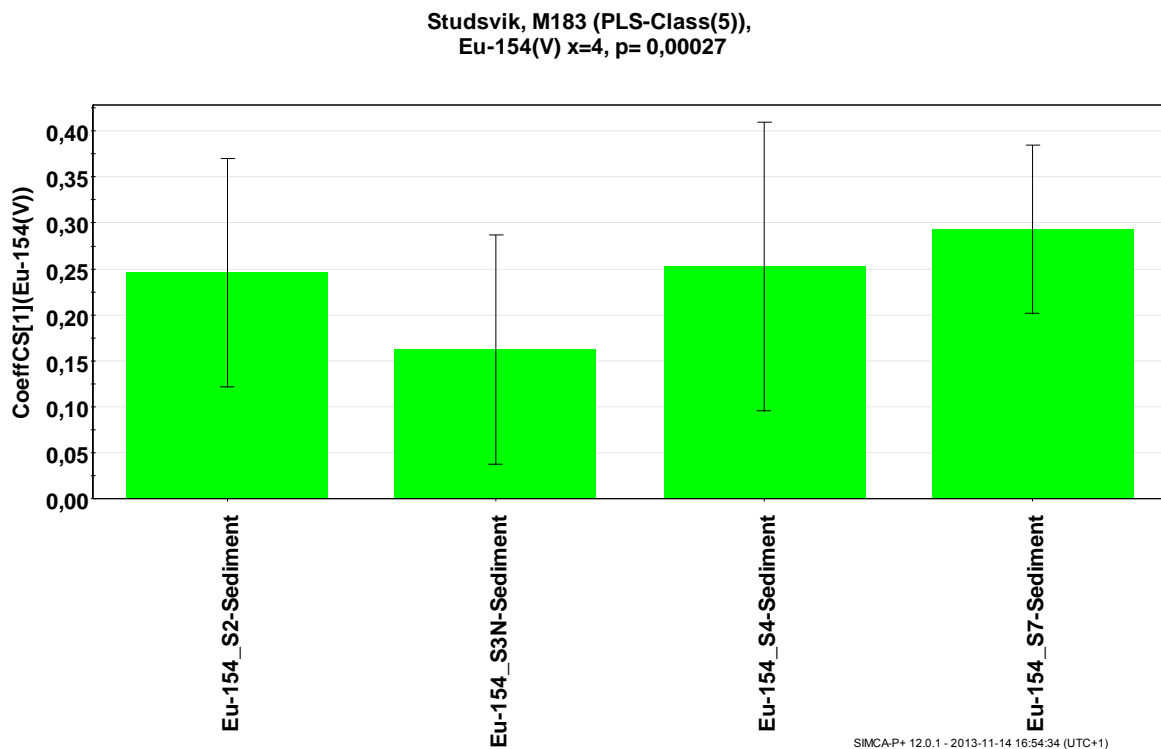
### 5.6.5. $^{154}\text{Eu}$

#### Utsläpp till luft

Det finns inga registrerade utsläpp av  $^{154}\text{Eu}$  till luft. Det finns heller inga registrerade mätvärden på luftrelaterade omgivningsvariabler.

#### Utsläpp till vatten

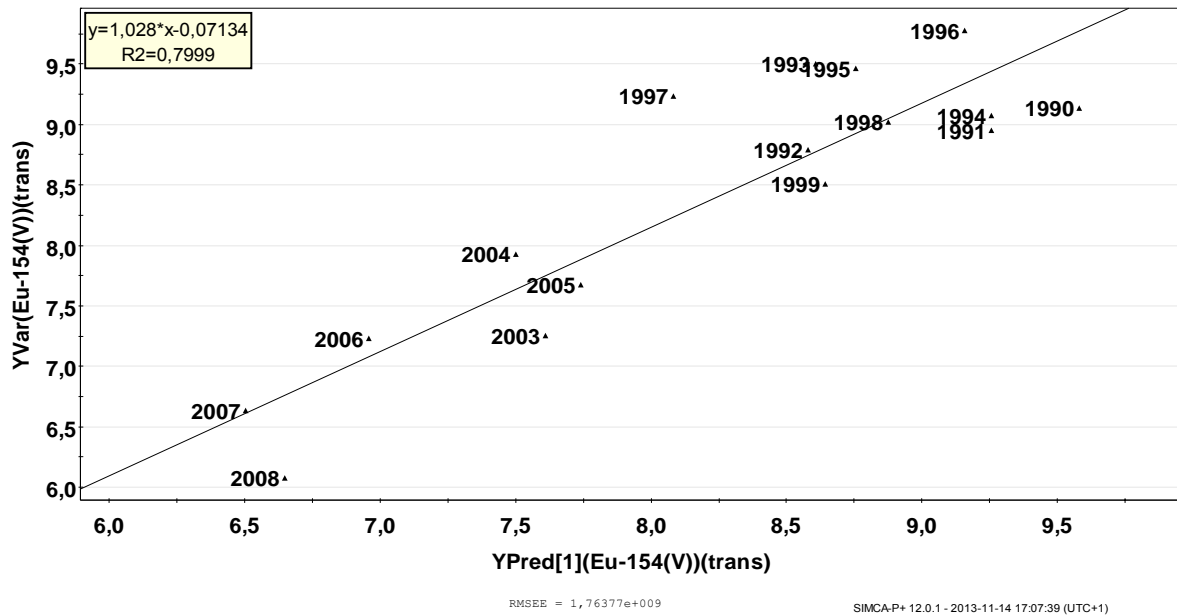
Trots att bara 9 omgivningsvariabler, med "missing data" < 80 %, finns registrerade går det att hitta 2 modeller bestående av 4 delvis olika signifikanta omgivningsvariabler. Modell M183, figur 21, har bäst prediktionsförmåga av dessa,  $Q^2=0,718$ , och modellsignifikansen är  $p=0,00027$ .



**Figur 21.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av  $^{154}\text{Eu}$  till vatten med 4 signifikanta variabler för den signifikanta modellen M183.

I figur 22 nedan, har observerat utsläpp avsatts mot predikerat för modell M183, och det bör observeras att grafen spänner över drygt 3 tiopotenser med relativt hög precision.

Studsvik, M183 (PLS-Class(5)),  
Eu-154(V) x=4, p= 0,00027



**Figur 22.** Observerat mot predikerat utsläpp av <sup>154</sup>Eu till vatten. Prediktionen är gjord med modell M183 med 4 omgivningsvariabler. Observera log-skala.

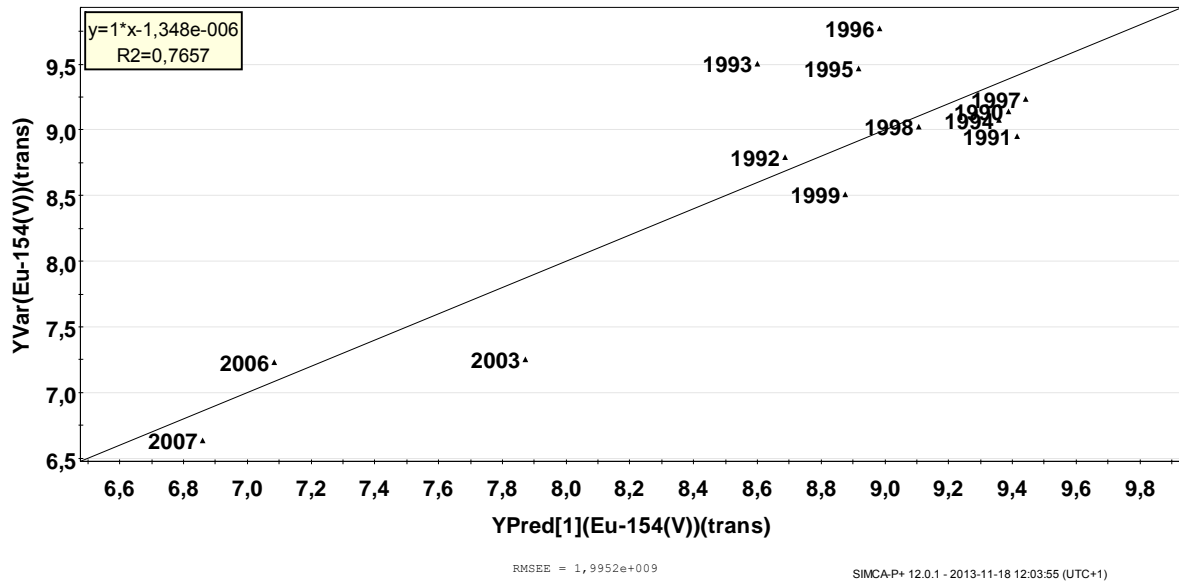
**Tabell 4.** Sammanställning av prediktionsförmåga, Q2, och modellsignifikans, p, för 4 en-variabelmodeller för utsläpp av <sup>154</sup>Eu till vatten.

Modell	Mätstation	Provslag	Q2	p
M150	2	Sediment	0,470	0,016
M152	4	Sediment	0,421	0,085 (NS)
M153	7	Sediment	0,736	0,0013
M184	3N	Sediment	0,175	0,35 (NS)

Notera att en-variabelmodellen M153 har något högre prediktionsförmåga, Q2=0,736, än det från modell M183, med 4 variabler, där Q2=0,718. Modell M152 och M184 i tabell 4, är inte strikt signifikanta då p-värdet är större än 0,05.

Nedan i figur 23 visas observerat mot predikerat värde för modell M153 med omgivningsvariabel (<sup>154</sup>Eu\_S7-Sediment).

Studsвик, M153 (PLS-Class(5)),  
Eu-154(V) x=1(S7-Sediment), p=0,0013



**Figur 23.** Observerat mot predikterat utsläpp av <sup>154</sup>Eu till vatten. Prediktionen är gjord med modell M153 med omgivningsvariabeln (<sup>154</sup>Eu\_S7-Sediment). Observera log-skala.

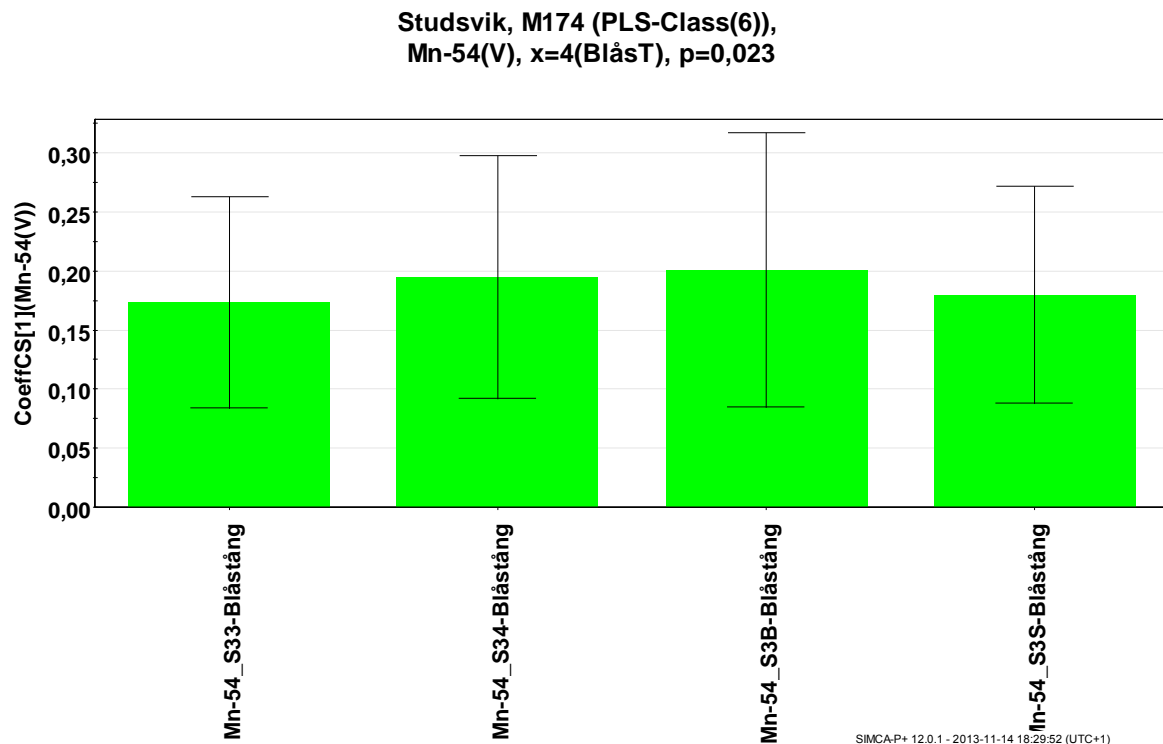
### 5.6.6. <sup>54</sup>Mn

#### Utsläpp till luft

Det finns endast 4 observationer av utsläpp av <sup>54</sup>Mn till luft mellan 1981-2012, och bara 2 omgivningsvariabler med "missing data" <80 %. Detta medför att det inte går att finna någon korrelationsmodell för luftrelaterade variabler.

#### Utsläpp till vatten

Totalt finns 15 omgivningsvariabler med "missing data" <80 %, men bara 2 av dessa har <50 % "missing data". Det går att finna 1 signifikant modell M174, figur 24, med 4 signifikanta omgivningsvariabler med Q<sup>2</sup>=0,468 och p=0,023.



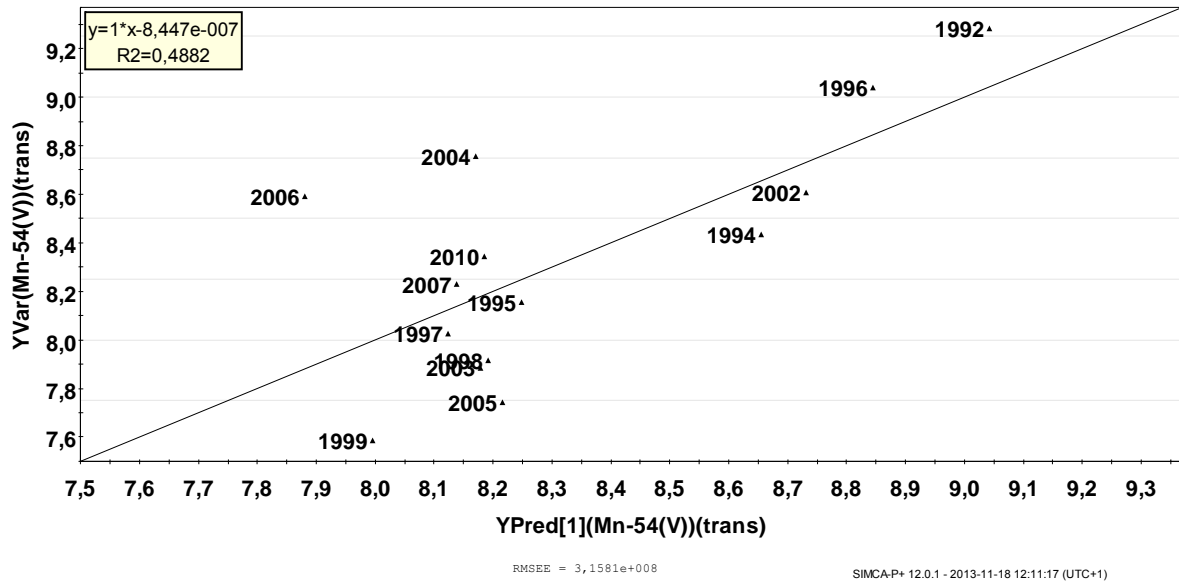
**Figur 24.** Plot av regressionskoefficienter (scaled & centered) för utsläpp av <sup>54</sup>Mn till vatten med 4 signifikanta variabler, för den signifikanta modellen M174 (Q<sup>2</sup>=0,468 och p=0,023).

Vidare förenkling mot en-variabelmodeller resulterar inte i någon strikt signifikant modell. Modell M178 med Q<sup>2</sup>=0,402 kommer närmast, men är inte strikt signifikant då p=0,059.

Observerat mot predikerat utsläpp av <sup>54</sup>Mn till vatten visas, trots icke strikt modellsignifikans, i figur 25 nedan. Prediktionsfelen verkar i alla fall maximalt vara under en 10-potens.



Studsвик, M178 (PLS-Class(6)),  
Mn-54(V), x=(S34-BlåsT), p=0,059



**Figur 25.** Observerat mot predikterat utsläpp av <sup>154</sup>Eu till vatten. Prediktionen är gjord med modell M153 med omgivningsvariabeln (<sup>54</sup>Mn\_S7-Sediment). Observera log-skala.

## 5.7. Prediktionsförmåga och prediktioner

Nedan följer en sammanställning av prediktionsförmåga och prediktioner för framförallt signifikanta PLS-modeller, men också några icke strikt signifikanta 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. Observera att det är endast för <sup>134</sup>Cs(L) som det finns en signifikant modell, med 1 variabel, för utsläpp till luft och luftrelaterade omgivningsvariabler.

Nuklid	Modell	Antal variabler (n)	Q2_n (%)	Modell	1 variabel	Q2_1 (%)
<sup>60</sup> Co(V)	M43	3	50,9	M45	S33_Blåstång	61,7
<sup>134</sup> Cs(V)	M69	3	33,6	-	-	-
<sup>137</sup> Cs(V)	M113	3	85,2	M115	S3B_Blåstång	84,5
<sup>152</sup> Eu(V)	M141	3	77,8	M140	S3B_Blåstång	70,8
<sup>154</sup> Eu(V)	M183	4	71,8	M153	S7_Sediment	73,6
<sup>54</sup> Mn(V)	M174	4	46,8	M178 <sup>1</sup>	S34_Blåstång	40,2
<sup>60</sup> Co(L)	-	-	-	-	-	-
<sup>134</sup> Cs(L)	M52	3	64,4	M56	SC_Björnmossa	90,4
<sup>137</sup> Cs(L)	M91	2	48,4	-	-	-
<sup>152</sup> Eu(L)	-	-	-	-	-	-
<sup>154</sup> Eu(L)	-	-	-	-	-	-
<sup>54</sup> Mn(L)	-	-	-	-	-	-

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

Modell	M45	M115	M140	M153	M178 <sup>1</sup>	M56
Nuklid	<sup>60</sup> Co(V)	<sup>137</sup> Cs(V)	<sup>152</sup> Eu(V)	<sup>154</sup> Eu(V)	<sup>54</sup> Mn(V)	<sup>134</sup> Cs(L)
Konstant (B)	7,98604	6,12787	6,64665	4,27595	7,78472	2,23371
Koefficient (k)	0,91761	2,01553	2,08446	3,05682	0,893114	1,53565

Regressions sambandet 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 1998. Detta år valdes på grund av att det inte saknas några värden för de kontrollerade utsläppen, men också för att omgivningsvariablerna

<sup>1</sup> M178 är inte strikt signifikant, p=0,059.

generellt inte har så mycket "missing data" detta år. För  $^{134}\text{Cs(L)}$  fanns dock ingen mätning på  $^{134}\text{Cs\_SC-Björnmossa}$  1998, därför användes ett mätvärde från 1994 för att kunna göra prediktionen.

**Tabell 7.** Mätta och predikterade värden för utsläpp till vatten och vattenrelaterade omgivningsaktivitet och för ett utsläpp till luft och luftrelaterade omgivningsvariabler, under år 1998. 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) 1998	Predikterat utsläpp	Mätt på en omgivningsvariabel (Bq) 1998	Predikterat för en omgivningsvariabel	Variabel
$^{60}\text{Co(V)}$	1,44E+09	1,39E+09	18,3	19	S33_BlåsT
$^{137}\text{Cs(V)}$	1,42E+10	5,48E+09	61,9	99	S3B_BlåsT
$^{152}\text{Eu(V)}$	1,00E+09	5,00E+08	9,65	13	S3B_BlåsT
$^{154}\text{Eu(V)}$	1,06E+09	1,27E+09	38	36	S7_Sediment
$^{54}\text{Mn(V)}$	8,25E+07	1,55E+08	2,85	1	S34_BlåsT
$^{134}\text{Cs(L)}$	1,10E+04	5,00E+03	9,0 <sup>2</sup>	15	SC_BjörnM

Prediktioner av utsläppen i tabell 7 visar god överensstämmelse med faktiskt mätta värden. Prediktionsfelen är klart mindre än en tiopotens.

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".

<sup>2</sup> Mätning 1994

## 6. Rekommendationer

- Denna analys har visat på nödvändigheten av att ha välorganiserade data och kontinuitet i registrering av mätvärden 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 regressions samband 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 av omgivningsvariabler, under detektionsgräns noteras, 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 regressions sambanden är t.ex. blåstång och sediment.
- Luftrelaterade omgivningsvariabler visar i denna studie generellt mycket låg mätfrekvens. Endast för  $^{134}\text{Cs(L)}$  erhöles en signifikant modell med 1 variabel ( $^{134}\text{Cs\_SC-Björnmossa}$ ).
- För att få en uppfattning om spridningskinetiken från utsläpp till omgivningen bör ett kontrollerat utsläpp följas upp med mätning av aktiviteten i några olika (nuklid\_mätstationer- provslag) 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 av viss nuklid till olika mätstationer och provslag.

## 7. Slutsatser

- Data som analyserats med multivariata verktyg (PCA och PLS) visar att rapporteringen av de kontrollerade utsläppen av de radioaktiva nukliderna  $^{60}\text{Co}$ ,  $^{134}\text{Cs}$ ,  $^{137}\text{Cs}$ ,  $^{152}\text{Eu}$ ,  $^{154}\text{Eu}$  och  $^{54}\text{Mn}$  till luft (L) och vatten (V) från de kärntekniska anläggningarna i Studsvik har hög frekvens i tidsserierna 1977-2012. Undantaget  $^{152}\text{Eu}$ ,  $^{154}\text{Eu}$  och  $^{54}\text{Mn}$ , där endast utsläpp till vatten har en, för multivariatanalys, användbar frekvens av mätvärden.
- De kontrollerade utsläppen av radioaktiva ämnen till vatten tenderar att minska med tiden.
- Aktiviteten av utsläppen till luft är oftast i medeltal ungefär 2 tiopotenser lägre än aktiviteten av utsläpp till vatten.
- Insamlade mätvärden på omgivningsvariabler, olika provslag och mätstationer, har dessvärre, med några undantag, mycket låg frekvens.
- Multivariatanalysen visar att det finns kvantifierbara samband mellan utsläpp till vatten och till vattenrelaterade omgivningsvariabler för de 6 studerade nukliderna, men endast ett samband för luftrelaterat utsläpp,  $^{134}\text{Cs}$  (L).
- För nukliderna  $^{60}\text{Co}$ ,  $^{137}\text{Cs}$ ,  $^{152}\text{Eu}$ ,  $^{154}\text{Eu}$  och  $^{54}\text{Mn}$  finns signifikanta vattenrelaterade en-variabelsamband.
- PLS-modeller med 1 signifikant variabel har i denna studie nästan lika bra och i några 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 bör vara signifikanta. Observera att inte alla omgivningsvariabler i denna rapport kan användas.
- 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.







2014:15

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)