



Strål
säkerhets
myndigheten

Swedish Radiation Safety Authority

Författare: Rolf Bergman

Forskning

2013:12

Multivariatanalys av radioaktivitets-
data dels från utsläpp till luft och
vatten från OKG och Clab dels i
omgivningen av Oskarshamn

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 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 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 Zn-65. 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 en stor del av de använda radioaktivitetsvärdena hämtats från pappersrapporter och all data har inte stått att finna under den begränsade tid som fanns till förfogande.

Behov av ytterligare insatser

Denna studie analyserade data för sex specifika nuklider i utsläpp från OKG och Clab och i omgivningen runt dessa anläggningar. 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.

Projektinformation

Kontaktperson SSM: Charlotte Lager
Referens: SSM2012-4039



Strål
säkerhets
myndigheten

Swedish Radiation Safety Authority

Författare: Rolf Bergman, PhD, Assoc. Prof
Rolf Bergman Consulting, Uppsala

2013:12

Multivariatanalys av radioaktivitets-
data dels från utsläpp till luft och
vatten från OKG och Clab dels i
omgivningen av Oskarshamn

Datum: Februari 2013

Rapportnummer: 2013:12 ISSN:2000-0456

Tillgänglig på 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

SAMMANFATTNING	3
1. BAKGRUND	4
2. SYFTE	5
3. DATA	6
4. ANALYSVERKTYG	7
4.1. PCA.....	7
4.2. PLS.....	7
4.3. ANALYSGÅNG.....	9
5. RESULTAT OCH DISKUSSION	10
5.1. DATA	10
5.2. PCA – ALLA DATA	11
5.3. PCA – UTSLÄPPSDATA	13
5.4. PLS – ALLA DATA	14
5.5. PLS – UTSLÄPP TILL LUFT RESPEKTIVE VATTEN.....	16
5.6. PLS – PER NUKLID.....	17
5.6.1. ⁵⁸ Co	17
Utsläpp till luft.....	17
Utsläpp till vatten	18
5.6.2. ⁶⁰ Co	21
Utsläpp till luft.....	21
Utsläpp till vatten	21
5.6.3. ¹³⁷ Cs.....	23
Utsläpp till luft.....	23
Utsläpp till vatten	25
5.6.4. ⁵⁹ Fe.....	27
Utsläpp till luft.....	27
Utsläpp till vatten	27
5.6.5. ⁵⁴ Mn	28
Utsläpp till luft.....	28
Utsläpp till vatten	29
5.6.6. ⁶⁵ Zn.....	31
Utsläpp till luft.....	31
Utsläpp till vatten	34
5.7. PREDIKTIONSFÖRMÅGA OCH PREDIKTIONER	36
6. REKOMMENDATIONER	38
7. SLUTSATSER	39

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 för perioden 1984-2010. De studerade nukliderna är ^{58}Co , ^{60}Co , ^{137}Cs , ^{59}Fe , ^{54}Mn och ^{65}Zn . Utsläppen sker kontrollerat till luft och vatten från de kärntekniska anläggningarna i Oskarshamn; Oskarshamnsverket (OKG) och det centrala lagret för mellanlagring av använt kärnbränsle (CLAB).

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 OKG och CLAB genom mätvärden från ett fåtal olika provslag (t.ex. blåstång, abborre och gulål) vid någon eller några av de utvalda mätstationerna i omgivningarna av Oskarshamn.

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 319 stycken där varje enskild variabel definieras av nuklid, mätstation och provslag (ex. $^{58}\text{Co}_{S1_Gädda}$). Analysen visar att det för 5 av de 6 studerade nukliderna 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. Enda undantaget är ^{59}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 ^{65}Zn .

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 Strålskyddsinstitutet (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 1984, och det har sedan dess samlats ansevära mängder mätdata, från det stora antalet 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 källutsläppen 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 en observation) blir det i praktiken omöjligt att behandla värdena med traditionell en-variabelsstatistik, som vanligtvis kräver betydligt fler observationer. 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 Moduling of Chemical Analogies), från Umetrics i Umeå, 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 Oskarshamn (OKG och CLAB) 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 ^{58}Co , ^{60}Co , ^{137}Cs , ^{59}Fe , ^{54}Mn och ^{65}Zn , till luft (L) och vatten (V) från OKG och CLAB. Aktiviteten av respektive nuklid utgör medelvärden per varje år. Datamatriisen innehåller också årsvisa medelvärden av mätdata för varje nuklid i olika provslag, för luft respektive vatten vid olika mätstationer.

Mätstationer som finns representerade i datamatriisen är station 1, 2, 6, 7, 11, 11b, 12, 15, 17, 18, 19, 20, 23, A, B, C=Figeholm, D, E, Kristdala, vilka utgör en delmängd av de mätstationer som nämns i omgivningskontrollprogrammet.

Provslag är de som omnämns i omgivningskontrollprogrammet.

Datamatriisen är organiserad så att varje år utgör en observation, och varje variabel definieras av nuklid_mätstation_provslag (ex. $^{58}\text{Co}_{S1_Gädda}$). Det finns 27 observationer (1984-2010) och totalt 319 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 4 av 27 möjliga, dvs. mer än 85 % ”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. 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 för att avslöja detta.

En ”loading plot” 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 algoritmen 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 skall analyseras samtidigt, eller om endast ett Y i taget skall analyseras. I den följande analysen har båda sätten använts, dock för mer detaljerad information analyseras varje enskild Y -variabel.

R^2 visar hur väl framräknade modeller anpassas till data. Q^2 anger modellens prediktionsförmåga av nya Y med nya X -värden. Q^2 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.

Normalfördelningsplot visar om residualerna (differensen mellan observerat och predikerat värde) är normalfördelade. Grafen visar sannolikheten som funktion av standardavvikelsen (std) per varje observation (se t.ex. Figur 13). Observationer med mer än 4 std är sannolikt "outliers".

Framräknade regressionskoefficienter 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 skall 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-validated (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.

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" till högt Q^2 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 säkerheter 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 skall vara signifikant och med högt Q^2 -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 en omgivningsvariabel, kan modellen 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 Q^2

5. Resultat och diskussion

5.1. Data

Det allmänna intrycket, under sammanställandet av datamatriken, var att mätvärden i omgivningen i några fall var konstanta under flera år och i andra variabler var skillnaden i mätvärden förvånansvärt stor. I det första exempelfallet var aktiviteten i mjölk vid mätstation A (betecknad _SA i diagram och tabeller i förekommande fall) konstant 2 Bq vid varje mättillfälle under åren 1989-1992, för nukliderna ^{60}Co och ^{65}Zn . I det andra fallet var skillnaden, som mest, en dryg faktor 4000, för mätning utförd samma dag på samma provslag (ex. $^{58}\text{Co_S1_Röttslam}$ år 1996). Förvånande observanda som kanske har en enkel förklaring (i det senare fallet skulle eventuella "hot spots" kunna vara en möjlig förklaring), men som ger ett intryck av att mätvärdena i matrisen kan vara behäftade med stora mätosäkerheter, vilket i så fall begränsar möjligheterna att finna signifikanta korrelationssamband. Vidare innehåller matrisen, när det gäller omgivningsdata, ett stort antal "missing data" före år 1989 och efter 1999, vilket också medför reducerade möjligheter att få generella och robusta prediktionsmodeller. En anledning till att det är glest med data efter år 1999 är att omgivningsdata för åren 2000 och 2001 saknas helt då det inte fanns tid att få fram dessa i projektets inledning. Bedömningen gjordes dock att projektet skulle genomföras utan dessa data och det troliga är att omgivningsdata ligger på samma nivå som närliggande år eftersom utsläppen inte har förändrats radikalt. De luftrelaterade omgivningsvariablerna innehåller mer "missing data" än de för vatten, vilket kommer att återspeglas vid modellbyggandet.

Variabler med färre än 4 av totalt 27 observationer ("missing data" > 85 %) har direkt uteslutits ur analysen.

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 trovärdig 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 mycket "missing data". Detta bidrar också till svårigheter i 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. 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 kan vara en ”felkälla” som försvårar möjligheterna att finna signifikanta korrelationssamband för de kontrollerade utsläppen från OKG och CLAB. 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”. Nedan 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 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.

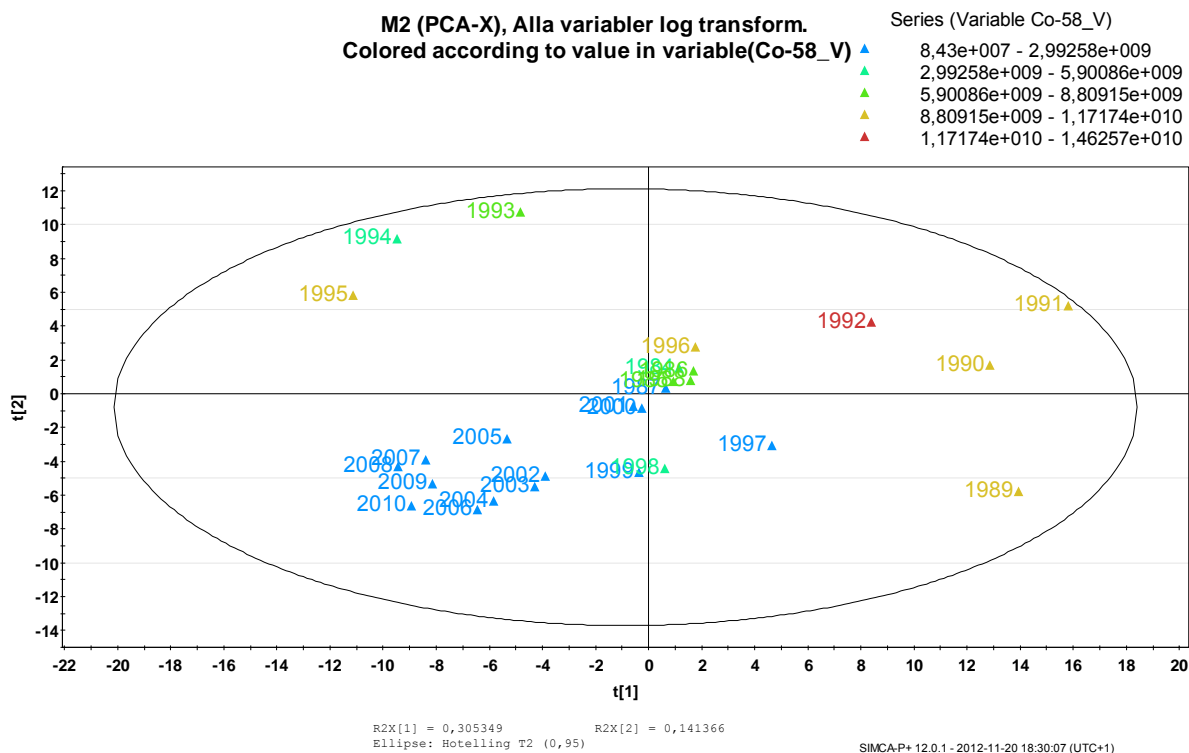
5.2. PCA – alla data

En PCA-modell på samtliga data (319 variabler och 27 observationer) utan någon förbehandling gav endast en signifikant principalkomponent (PC1). Modellens prediktionsförmåga blev $Q^2=0,07$. Således visar Q^2 att endast 7 % av variationen i data kan predikteras. Vid en kontroll av hur mätvärdena är fördelade i 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 $Q^2=0,167$. Den logaritmiska transformeringen resulterade således i en fördubbling av prediktionsförmågan, jämfört med icke transformerade data, till ungefär 17 %.

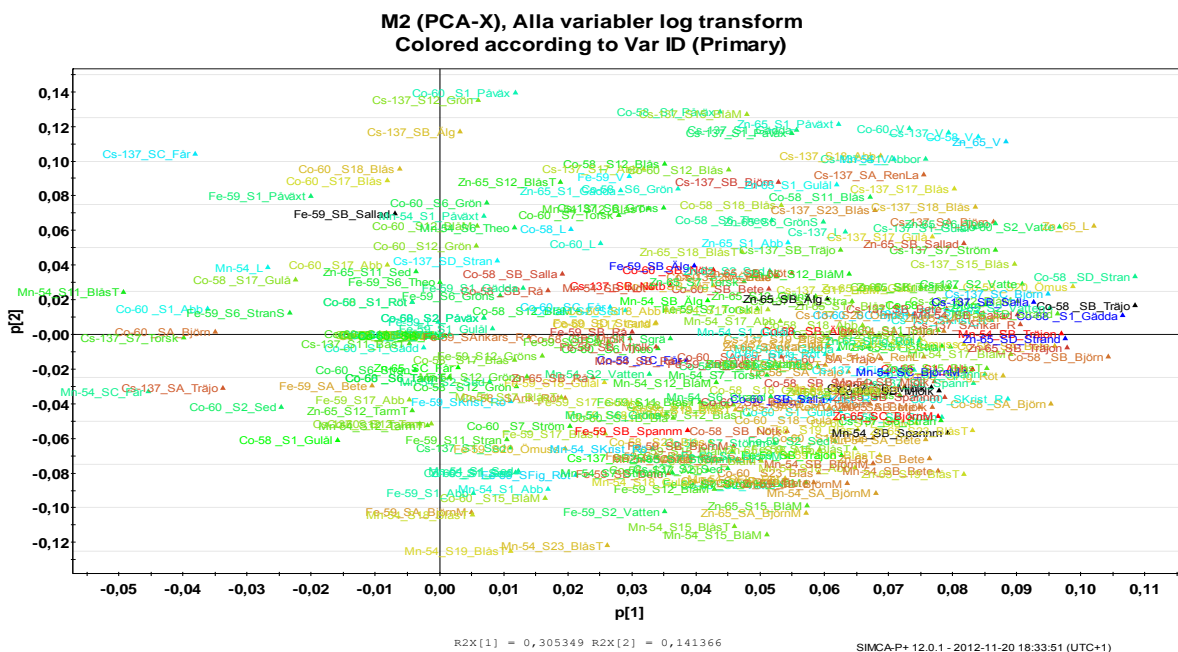
Alla följande analyser kommer att baseras på log-transformerade värden.

I Figur 1, ”score plot”, nedan 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 2, 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”.



Figur 1. "Score plot" av samtliga observationer (år) med hänsyn till samtliga variabler utan uppdelning för luft och vatten. Storleken av utsläppen (Bq/år) av, i detta exempel, ^{58}Co har kodats med färg.

Den blå färgen, i framför allt den tredje kvadranten i Figur 1, antyder att utsläppen av ^{58}Co har minskat med åren (se färgkodningen i figuren). Det syns också att utsläppen var något högre under början av 1990-talet där det högsta mätvärdet av ^{58}Co noterades 1992.



Figur 2. "Loading plot" av samtliga variabler.

En "loading plot" visas i Figur 2 för att ge en första överblick av hur variablerna är fördelade och förhåller sig till varandra. Den 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. Generellt gäller att observationerna i t.ex. 1:a kvadranten i Figur 1 förknippas, i positiv mening, med variablerna i 1:a kvadranten i Figur 2.

Notera att den underliggande modellen, M2, bara har en prediktionsförmåga på c:a 17 %, samt att data inte är uppdelade på luft- respektive vattenutsläpp, så några långtgående slutsatser kan ännu inte dras.

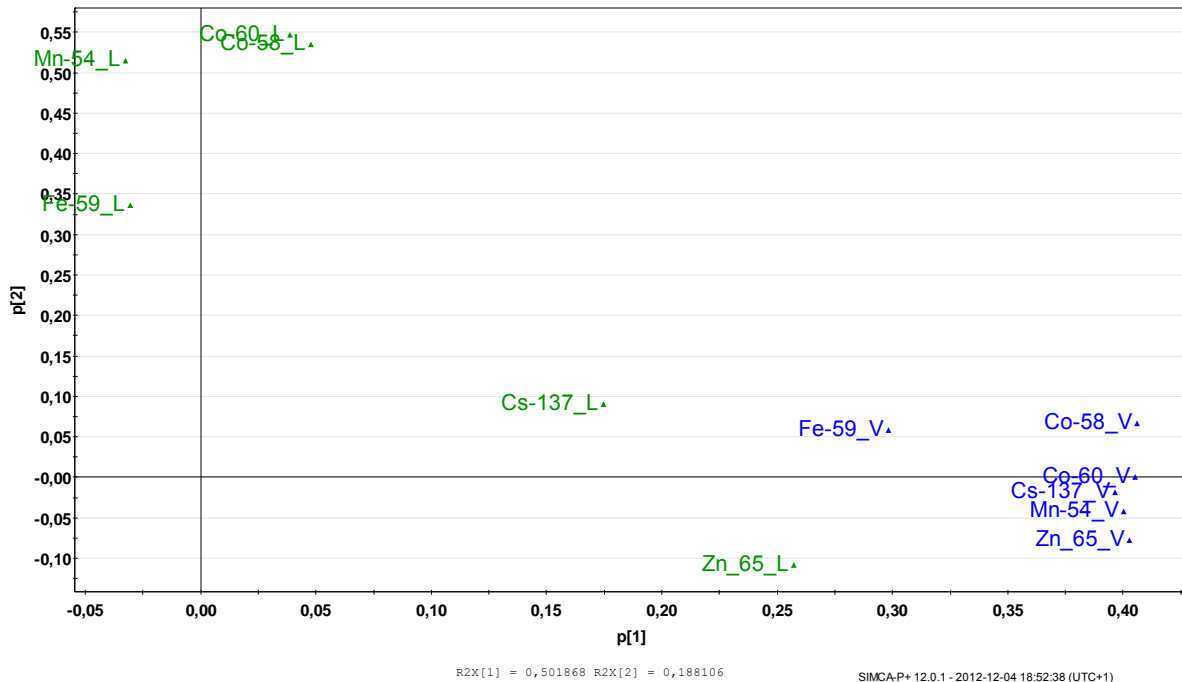
5.3. PCA – utsläppsdata

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 vatten också medför hög aktivitet i luftutsläppet? 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.

En indikation om korrelationsmönstret ses i Figur 3 nedan ("score plot"). Det är en PCA som är gjord på enbart Y-variablerna, där det nästan inte finns några "missing data", undantaget ⁵⁹Fe till luft (L) med ungefär 50 % "missing data". Det är viktigt att komma ihåg att korrelationsmönstret, i Figur 3 för modell M260, 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.

Det framgår att nästan alla Y(V) (blå), är starkt förknippade med varandra, utom för $^{59}\text{Fe}(\text{V})$ som avviker något från mönstret. Det framgår också att de flesta Y(L) (gröna) är mer utspridda (mycket svaga korrelationer) i ”score plot”, men där $^{65}\text{Zn}(\text{L})$ och i någon mån $^{137}\text{Cs}(\text{L})$ kan ha någon korrelation till utsläppen till vatten. Övriga Y(L) ligger nära noll längs PC1 och har minimal betydelse för modellens Q2, som har en total prediktionsförmåga på 44,9 % av variationen, varav 43,3 % förklaras av PC1 och endast 1,6 % av PC2. Analysen visar att det inte tycks finnas någon entydigt korrelation mellan luft- och vattenutsläpp, men det framgår i alla fall att Y(V) är starkt korrelerade.

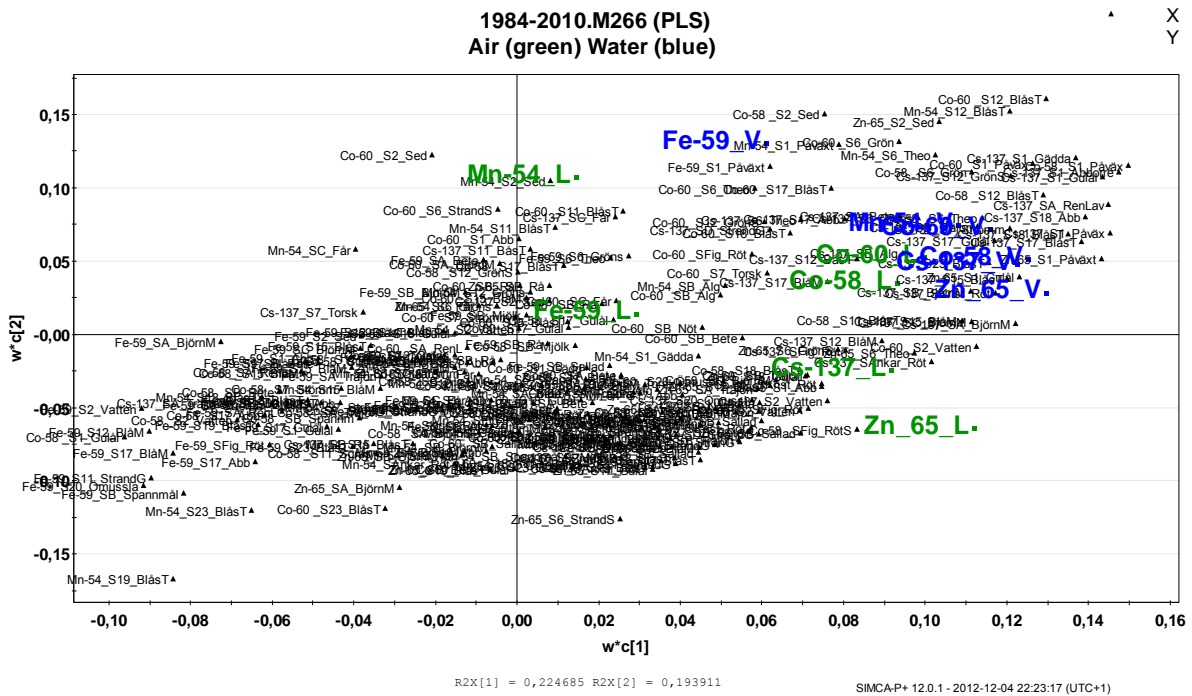
1984-2010.M260 (PCA-X&Y), L-V korrelation exkl1991
Air green, Water blue



Figur 3. PCA på Y-data (inga omgivningsvariabler ingår) där utsläppen till luft (L) är gröna och utsläppen till vatten (V) är blå.

5.4. PLS – alla data

PLS på alla data, där utsläpp till luft (L) och vatten (V) representeras av Y och omgivningsvariablerna av X. Figur 4, modell M266, visar i vissa avseenden ett likartat mönster som det i Figur 3 (PCA), men med vissa skillnader. De blå Y(V) är väl samlade och de gröna Y(L) är mer spridda och nu observeras de båda cobolt-isotoperna nära Y(V)-grupperingen. Osäkerheten i (L)-variablerna är avsevärt mycket större än motsvarande (V)-variabler, då generellt (L) består av mycket ”missing data”, speciellt före 1989 och efter 1999. Den glesa (L)-matrisen gör att det är svårt att hitta något tillförlitligt korrelations samband mellan Y(L) och Y(V) ur dessa data, i den mån det skulle finnas något samband. Frågan är då, finns det någon fysikalisk förklaring till varför bara vissa nuklider, från luftutsläppen, tycks ha en svag korrelation med nukliderna från vattenutsläppen, och andra inte? Resultatet från denna analys kan inte helt utesluta en korrelation mellan Y(L) och Y(V), men det verkar inte finnas någon trovärdig robust modell för denna koppling.

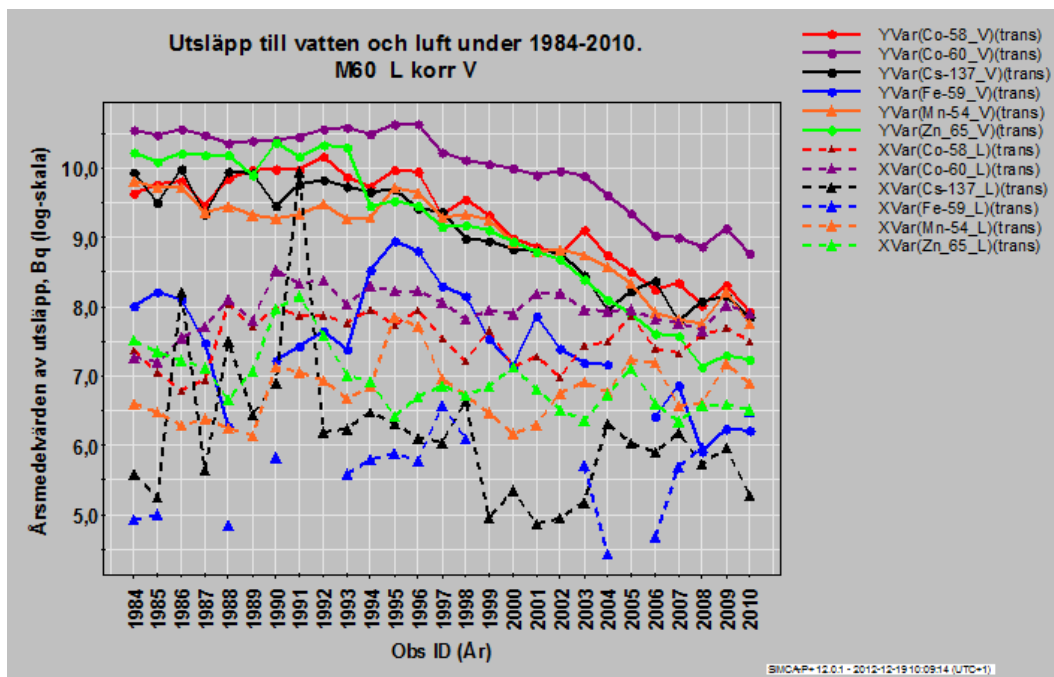


Figur 4. "Loading plot" med alla variabler, både X och Y. Där utsläppen till luft är gröna och till vatten är blå.

Den fortsatta analysen kommer inte att samanalysera (L)- och (V)-variabler. En eventuell svag korrelation, under rådande omständigheter med få variabler och stora luckor i datamattisen för (L)-variablerna, kommer att ha liten betydelse för det slutliga resultatet att finna eventuella kors samband mellan utsläpp och aktiviteter i omgivningen.

En del av det som antytts ovan, om eventuella samband, kan faktiskt observeras i Figur 5 som visar årsmedelvärden av utsläppen till vatten respektive luft per år. Till exempel är utsläppen till vatten (prickar med heldragna linjer) betydligt högre och "jämnare" än utsläppen till luft (trianglar med streckade linjer). Utsläppen av ⁵⁹Fe(V) (blå) är mer i nivå med utsläppen till luft, och ⁵⁹Fe(L) saknar c:a 50 % av data. För ¹³⁷Cs(L) ses ett "taggigt" mönster och ett extremt högt värde 1991. Det senare kan förklaras med att Oskarshamn fick en bränsleskada 1991, som sedan togs ut under 1992.

Generellt verkar utsläppen till vatten ha ett inbördes likartat utseende, de minskar långsamt med tiden. Luftutsläppen varierar mer och det är inte lika lätt att identifiera trender i tiden. Dessa senaste observationer är med stor sannolikhet avgörande för att det inte finns några tydliga korrelationsmönster mellan luft- och vattenutsläpp i Figur 3 och Figur 4, samt att det ur här analyserade data bara går att finna ett signifikant regressions samband mellan utsläpp till luft och luftrelaterade omgivningsvariabler för ⁶⁵Zn(L) (se den nuklidspecifika analysen nedan).



Figur 5. Graf visande årsmedelvärden av utsläpp till vatten (V) och luft (L) i Bq i logaritmisk skala. Varje nuklid har samma färg för (V) och (L). Heldragna linjer gäller för (V).

5.5. PLS – utsläpp till luft respektive vatten

Data har, förutom separation av (L)- och (V)-variabler, klassindelats så att varje nuklid analyseras separat för att ge ett mer lättolokat resultat.

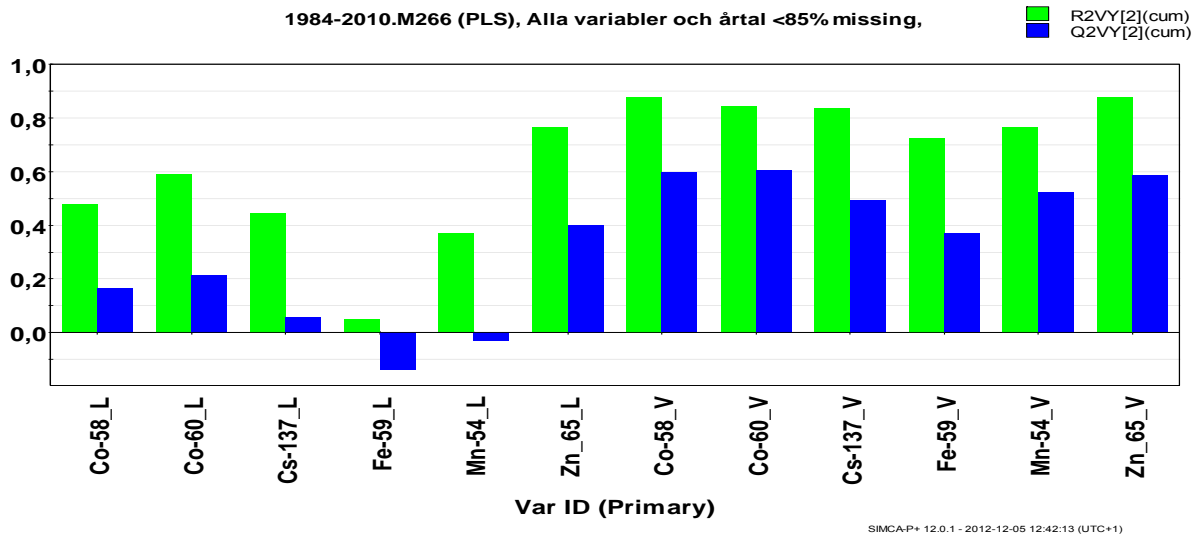
För att kvantifiera korrelationerna mellan utsläpp och aktiviteten i olika provslag vid olika mätstationer har PLS använts. Vanligtvis betraktas utsläppen som grunden till att aktivitet kan mätas i de olika provslagen i omgivningen och skulle därmed betraktas som X. I den kommande analysen har i stället aktiviteten i provslagen vid de olika mätstationerna hanterats som X 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.

Figur 6 nedan visar en översikt av sambanden från PLS-modell M266 för de olika nukliderna vad det gäller anpassning till data, R2 (grön), och prediktionsförmåga, Q2 (blå). Denna modell ger en indikation om sannolikheten att, efter ”trimning”, finna signifikanta regressionsmodeller. Modellen är baserad på alla variabler med mindre än 85 % ”missing data” och där några årtal (observationer) på grund av samma anledning utesluts.

Negativa Q2 (blå staplar) i figuren, visar att respektive utsläpp inte kan predikteras, utifrån gjorda aktivitetsmätningar. Det är bara brus som modelleras när icke signifikanta variabler ingår i beräkningarna, trots att det i de flesta fall tycks vara en rela-

ti tvt bra anpassning till data (R2). På motsvarande sätt ses att samtliga utsläpp till vatten (V), utom ⁵⁹Fe, relativt väl, verkar kunna beskrivas av mätta aktiviteter (höga Q2). Observera att utsläppet till luft av ⁶⁵Zn(L) verkar ha en relativt bra prediktionsförmåga. Den fortsatta mer detaljerade analysen får utvisa de nuklidspecifika modellernas validitet.



Figur 6. Anpassning till data, R2 (grön), och prediktionsförmågan, Q2 (blå), för varje studerad nuklid, för modell M266.

5.6. PLS – per nuklid

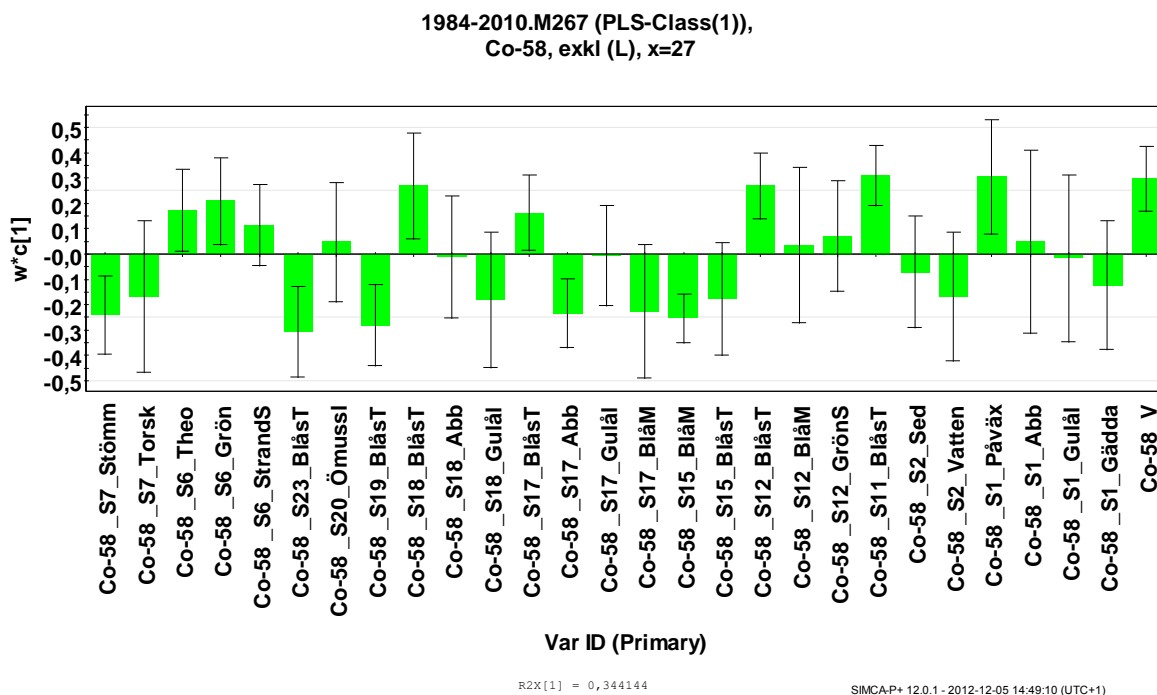
PLS på klassindelade data per varje nuklid med relaterade variabler för luft respektive vatten.

5.6.1. ⁵⁸Co

Utsläpp till luft

Nuvarande data ger ingen signifikant korrelationsmodell mellan utsläppen till luft och luftrelaterade omgivningsvariabler. Mjukvaran föreslog några modeller där bland annat aktiviteten i ett antal provslag minskade med ökade utsläpp till luft, och det förfaller inte rimligt. Förmodligen är det för mycket ”missing data”.

Utsläpp till vatten

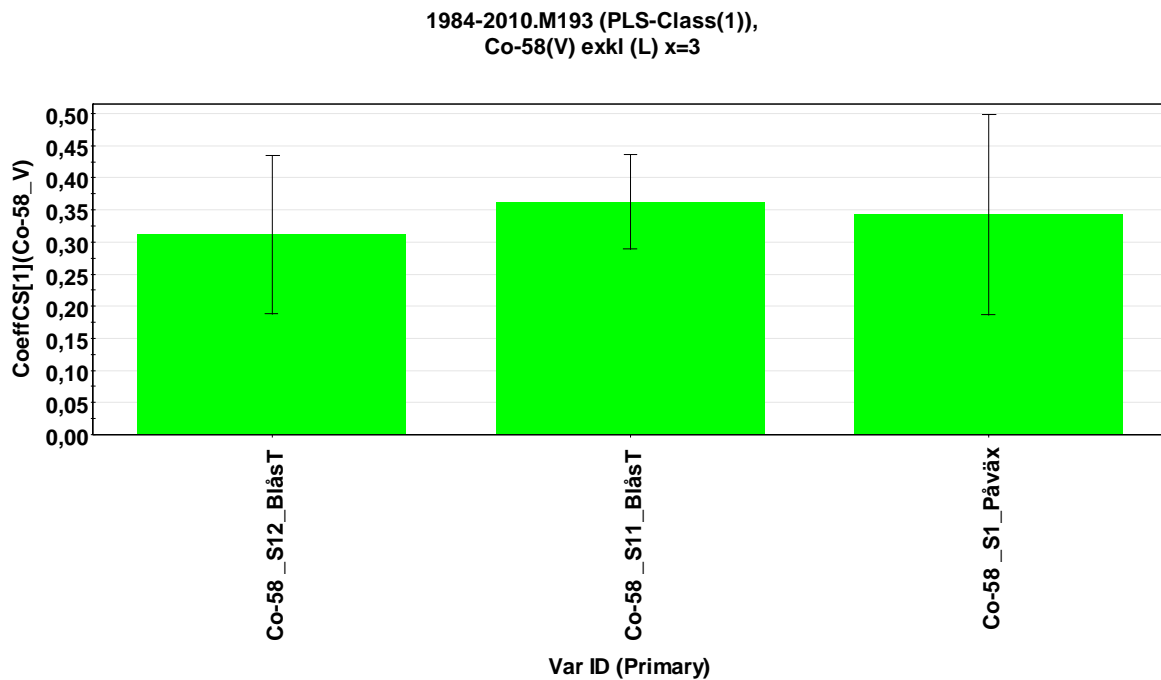


Figur 7. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ^{58}Co till vatten med relaterade variabler. Modell M267 har 27 variabler.

I Figur 7 visas regressionskoefficienterna med sina respektive osäkerhetsintervall för samtliga omgivningsvariabler. Modellen är inte signifikant, men illustrerar hur det kan se ut med en blandning av signifikanta (osäkerhetsintervallet omfattar inte noll) och icke signifikanta (osäkerhetsintervallet omfattar noll) koefficienter.

För att om möjligt finna en signifikant modell exkluderas succesivt icke signifikanta variabler som har stora osäkerheter och dessutom små numeriska värden. Det är fortfarande ett stort antal variabler, så det måste beaktas att slumpmässig korrelation kan förekomma. Stegvis eliminering av variabler, ”trimmning”, är att rekommendera. Vidare kommer negativa korrelationer att uteslutas, då ökande utsläpp till vatten, av i detta fall ^{58}Co , rimligtvis inte kan leda till minskad aktivitet i provslagen. PLS-algoritmerna söker en ”bästa möjliga” anpassning av X till Y och därför kan det också dyka upp negativa korrelationer.

Ett exempel på en förenklad signifikant modell (M193) med 3 signifikanta variabler visas i Figur 8.



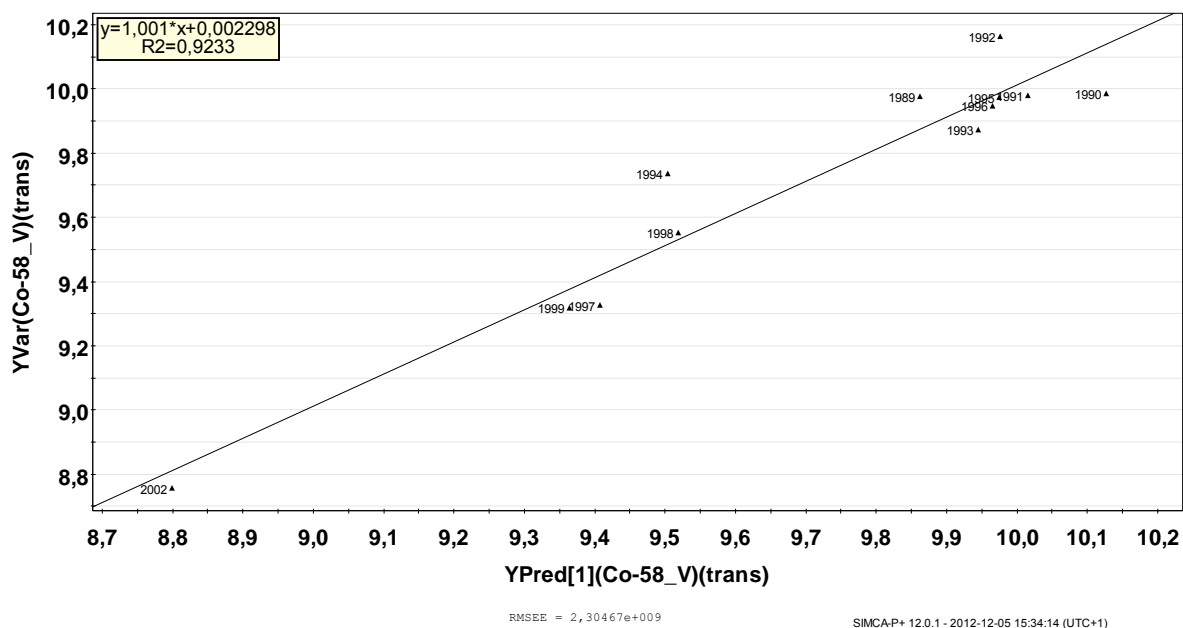
Figur 8. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ⁵⁸Co till vatten med 3 signifikanta variabler för modell M193.

Modellen M193 har en prediktionsförmåga på nästan 92 % ($Q^2=0,917$), vilket är mycket högt. Modell M267 i Figur 7, baserad på 27 variabler, har ”bara” ett $Q^2=0,617$. Dessutom är inte den modellen signifikant. Den enklare modellen ger en avsevärt säkrare prediktion av utsläppet, dessutom signifikant, och endast tre omgivningsvariabler behöver mätas. Således en betydande resursbesparing med bättre precision.

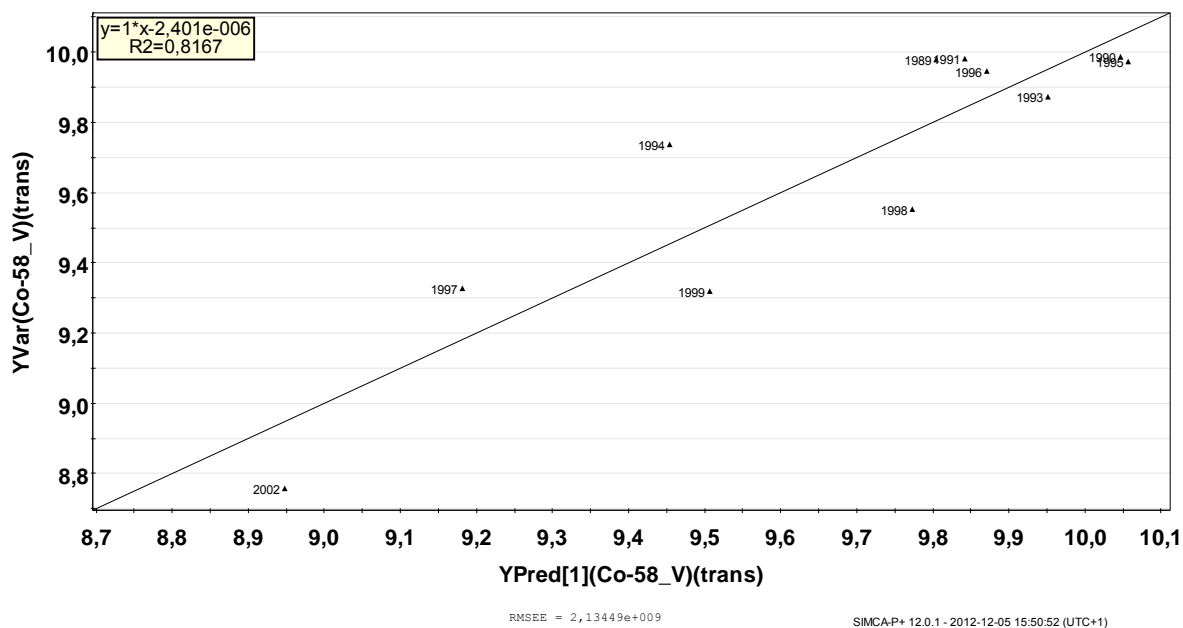
Det går naturligtvis att ta ett steg till mot en ytterligare förenklad modell genom att välja att mäta på bara ett provslag. I detta fall skulle Blåstång (*Fucus*) vid station 11 varit det naturliga valet, eftersom osäkerheten som visas i Figur 8 är den minsta. Dock fanns det inte tillräckligt med observationer, efter år 2000, vid den mätstationen för att ge en signifikant en-variabelmodell. Blåstång vid station 12 däremot gav en strikt signifikant modell med denna enda variabel och med $Q^2=0,767$. Uppenbarligen kan nästan 77 % av variationen i utsläppet av ⁵⁸Co till vatten predikteras i normalfallet.

Nedan i Figur 9 visas en jämförelse av den praktiska prediktionsförmågan i två grafer av observerade mot predikterade värden.

1984-2010.M193 (PLS-Class(1)), Co-58(V) exkl (L), x=3



1984-2010.M269 (PLS-Class(1)), Co-58(V), exkl (L), x=1



Figur 9 a och b. Observerat mot predikerat utsläpp (logaritmisk skala) av ⁵⁸Co till vatten i övre grafen M193, med 3 variabler, och den undre grafen, M269 med en variabel.

Naturligtvis observeras en skillnad i prediktionsförmågan (se R2 i graferna) mellan dessa två alternativa modeller, men det kanske räcker med en-variabelsfallet för att konstatera om utsläppen följer ”det vanliga mönstret”. Observera den logaritmiska skalan.

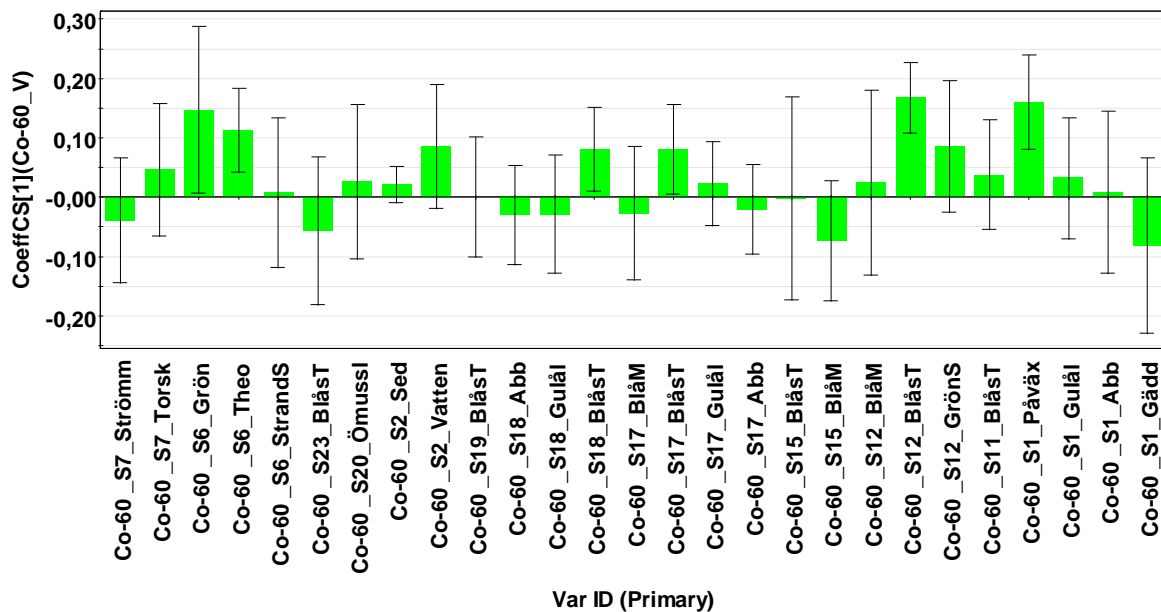
5.6.2. ^{60}Co

Utsläpp till luft

Trots tre signifikanta variabler och en prediktionsförmåga på 28 % är modellen (M148) icke signifikant enligt CV_ANOVA-testet. Således hittas ingen signifikant korrelation mellan utsläpp till luft av ^{60}Co och till relaterade omgivningsvariabler.

Utsläpp till vatten

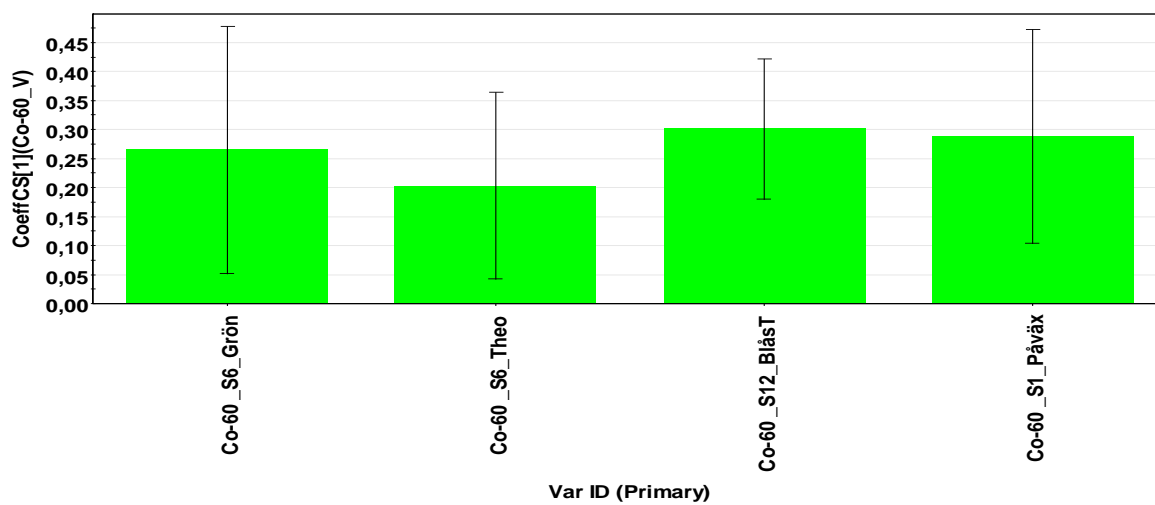
1984-2010.M288 (PLS-Class(2)), Co-60(V),exkl luft, x=27



Figur 10. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ^{60}Co till vatten med relaterade variabler. Modell M288 har 27 variabler.

Modell M288 visar ett antal icke signifikanta variabler och dessutom några variabler som är negativt korrelerade till utsläppet, och som behöver exkluderas. Successiv eliminering av dessa variabler visas i modell M164, som har 4 signifikanta variabler och är signifikant med $p=1,0e-8$. p -värdet indikerar sannolikhetsnivån att en modell är signifikant och för p -värden lägre än 0,05 anses en modell vara signifikant. Prediktionsförmågan för denna modell är $Q^2=0,90$. Därmed kan 90 % av normal variation i utsläppet av ^{60}Co till vatten predikteras.

1984-2010.M164 (PLS-Class(2)), Co-60(V),exkl luft x=4



SIMCAP+ 12.0.1 - 2012-12-06 12:07:23 (UTC+1)

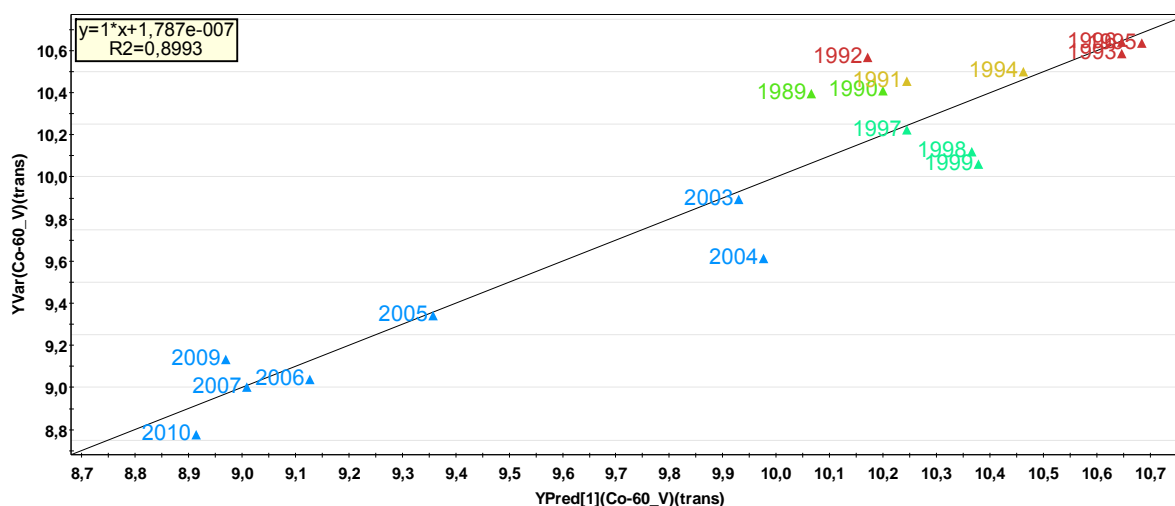
Figur 11. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ⁶⁰Co till vatten med relaterade variabler. Modell M164 har 4 variabler.

En ytterligare förenklad signifikant modell ($p=4,5e-8$) med endast en variabel (S12_Blåstång), modell M270, har obetydligt lägre $Q^2=0,895$ än för 4-variabelsmodellen (M164). Fortfarande en hög prediktionsförmåga med bara mätning på en omgivningsvariabel. Det förtjänar att påpekas att Blåstång vid station 12 bara har enstaka ”missing data”.

1984-2010.M270 (PLS-Class(2)),
Co-60(V), exkl (L), x=1(S12_Blåst)
Colored according to value in variable (Co-60_V)

Series (Variable Co-60_V)

- ▲ 5,943e+008 - 9,20416e+009
- ▲ 9,20416e+009 - 1,7814e+010
- ▲ 1,7814e+010 - 2,64239e+010
- ▲ 2,64239e+010 - 3,50337e+010
- ▲ 3,50337e+010 - 4,36436e+010



RMSEE = 8,77924e+009

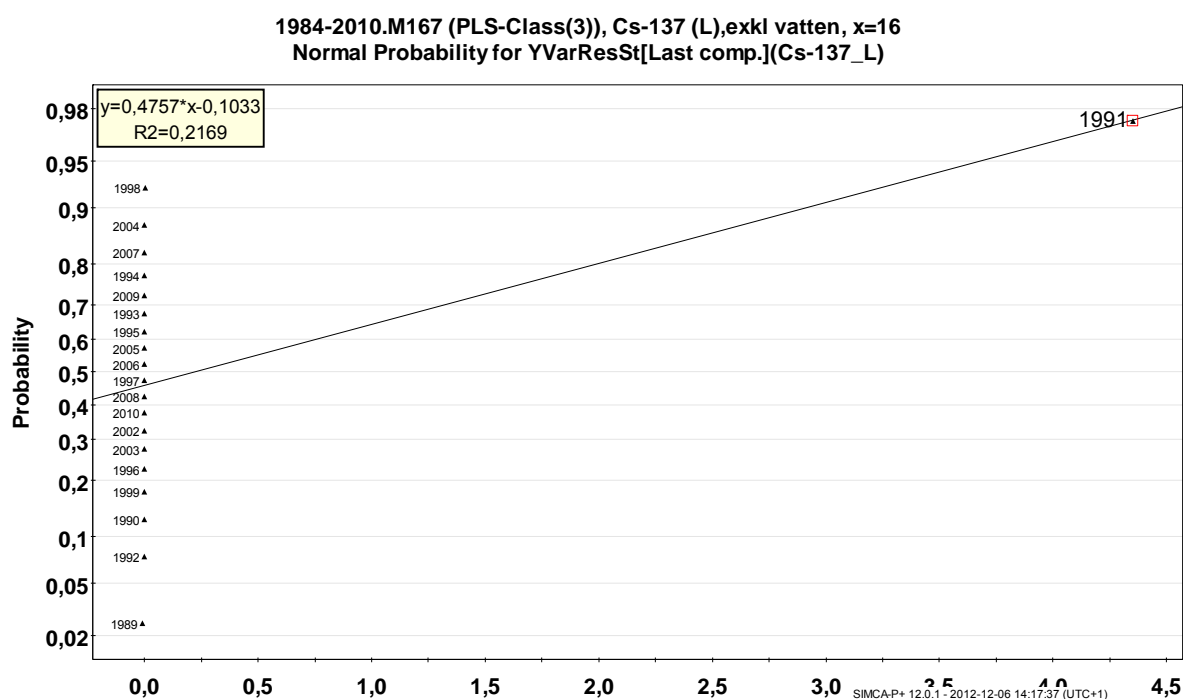
SIMCAP+ 12.0.1 - 2012-12-06 12:38:53 (UTC+1)

Figur 12. Observerat mot predikerat utsläpp av ⁶⁰Co baserat på aktivitetsmätning av endast blåstång vid station 12. Observera logaritmisk skala.

5.6.3. ¹³⁷Cs

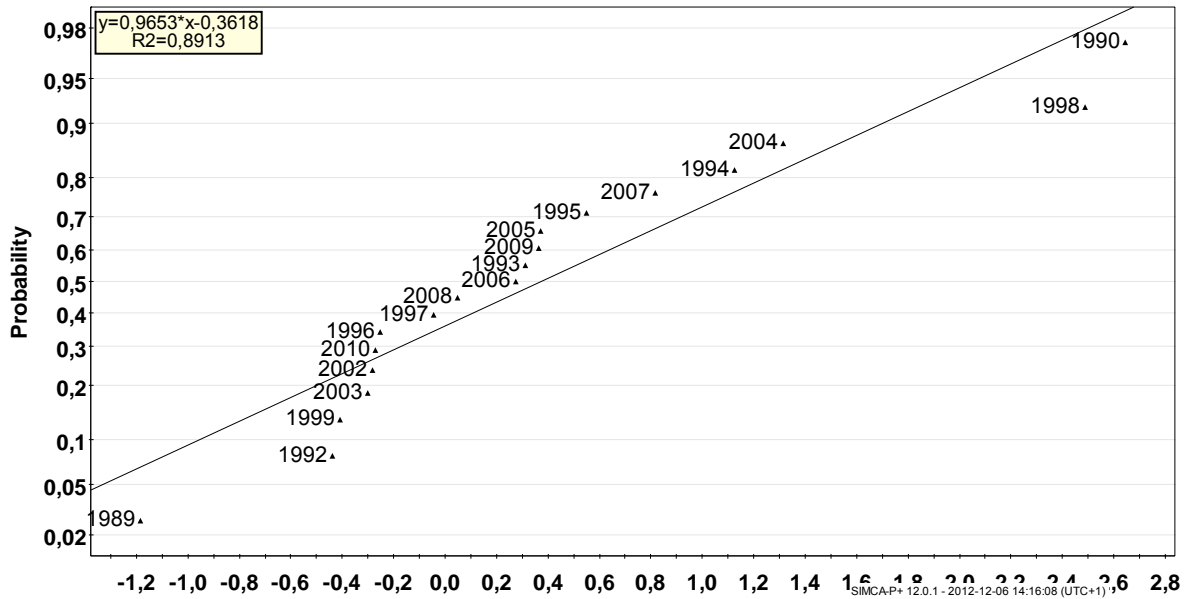
Utsläpp till luft

I Figur 13 nedan, normalfördelningsplotten, framgår det tydligt att observationen 1991 är klart avvikande från övriga observationer. Ändå har modell M167 ett $Q^2=0,171$, eller kanske snarare på grund av observationen 1991. När den observationen exkluderas ur data och en ny modell (M168) beräknas sjunker $Q^2=0,053$, vilket är en faktor 3 lägre än föregående. Det är således den borttagna observationen som ger någon form av riktning i data, men den är så pass annorlunda att det måste ha skett något speciellt detta år. Årsmedelvärdet på utsläppet av ¹³⁷Cs till luft år 1991 är 3000 gånger högre än jämfört med omkringliggande år, vilket kan förklaras med att block 2 hade en bränsleskada det året.



Figur 13. Normalfördelningsplot till modell M167 för ¹³⁷Cs(L).

1984-2010.M168 (PLS-Class(3)),
 Cs-137 (L),exkl vatten, x=16, exkl 1991
 Normal Probability for YVarResSt[Last comp.](Cs-137_L)

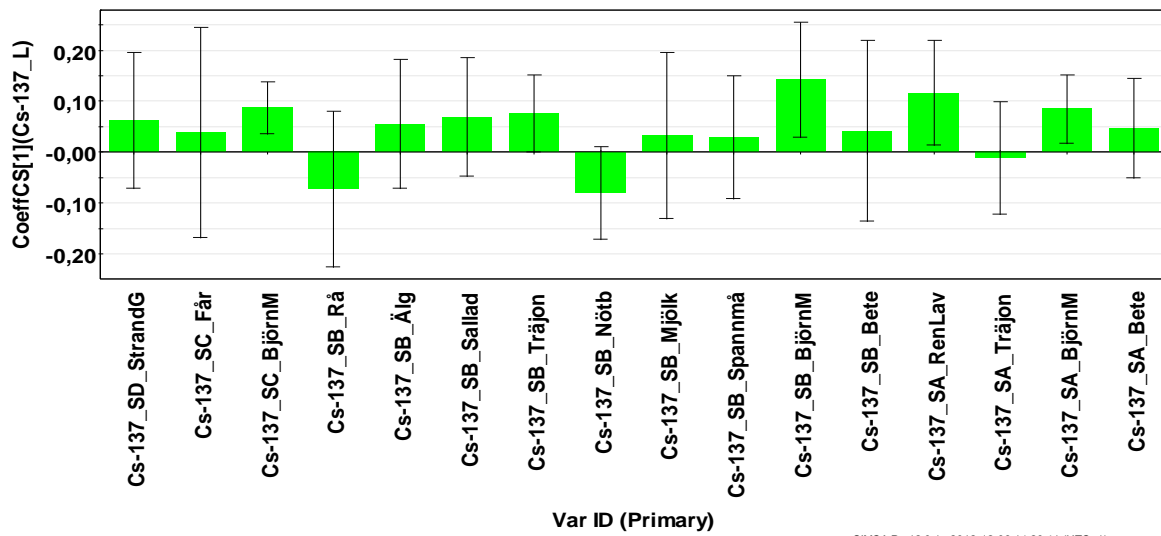


Figur 14. Normalfördelningsplot till modell M168 för $^{137}\text{Cs}(\text{L})$ där observationen 1991 exkluderas.

Normalfördelningsplotten för modell M168, Figur 14, visar inte på några ytterligare extrema observationer.

Även om observationen 1991 exkluderas går det inte att trimma de succesivt förenklade modellerna till signifikans, även om det finns signifikanta variabler i modell M168. Se t.ex. björnmossa och renlav i Figur 15, som båda har relativt många observationer. Således hittas ingen signifikant regression mellan utsläpp av ^{137}Cs till luft och relaterade luftvariabler.

1984-2010.M168 (PLS-Class(3)),
Cs-137 (L),exkl vatten, x=16, exkl 1991

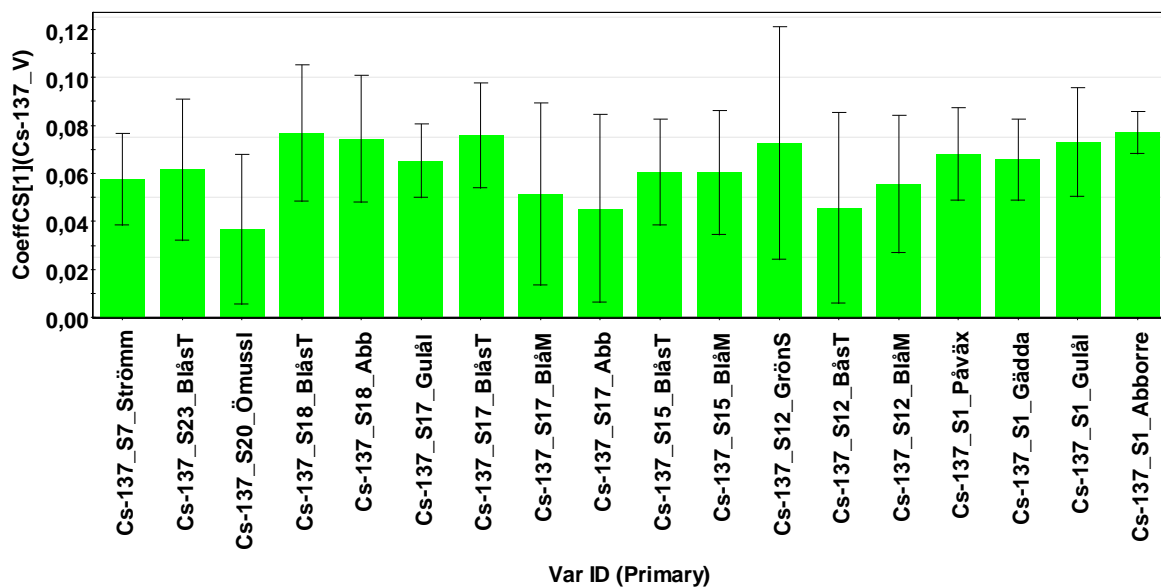


Figur 15. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ¹³⁷Cs till luft med relaterade variabler. Modell M168 har 16 variabler.

Utsläpp till vatten

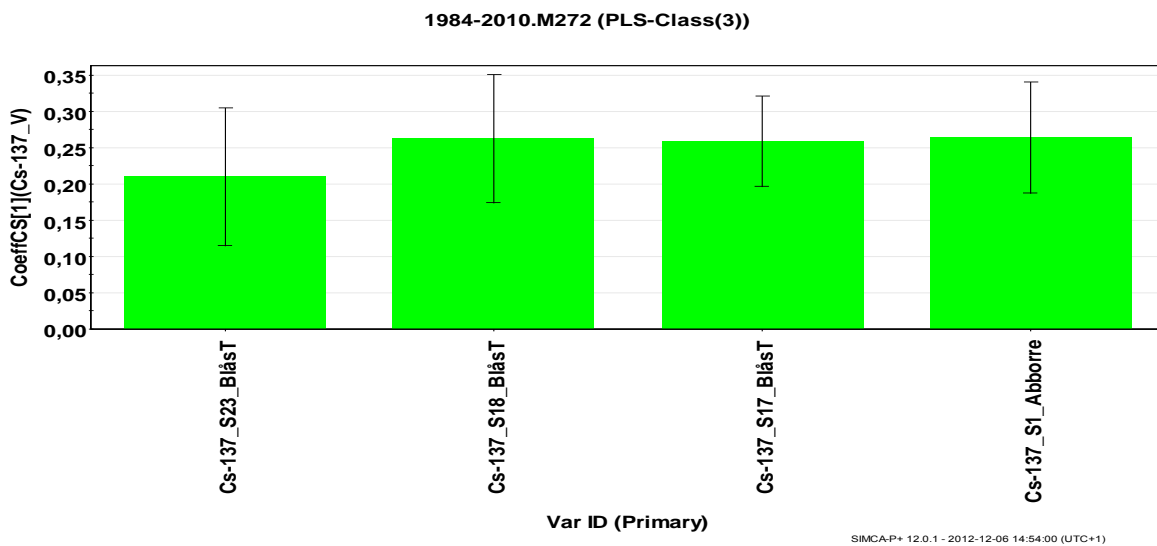
Modell M195, är signifikant med $p=1,5e-6$ och med 18 signifikanta omgivningsvariabler. Modellens $Q^2=0,794$, visar på god prediktionsförmåga.

1984-2010.M195 (PLS-Class(3)),
Cs-137(V),ekl luft, x=18



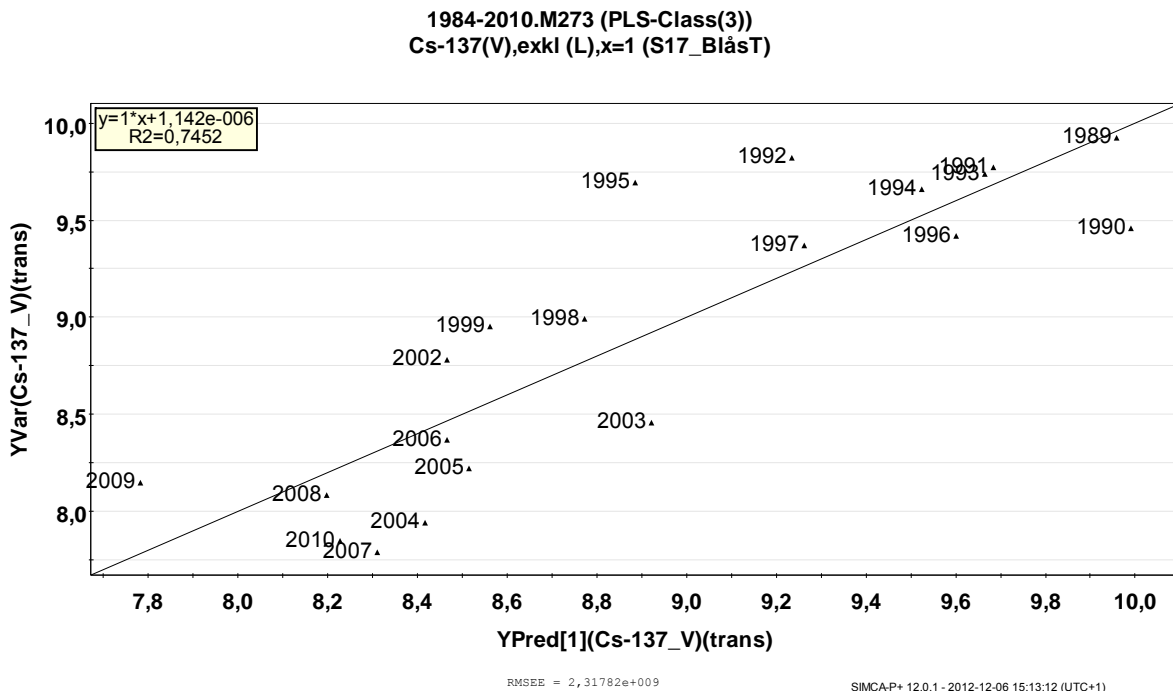
Figur 16. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ¹³⁷Cs till vatten med relaterade variabler. Modell M195 har 18 signifikanta omgivningsvariabler.

Modell M272, Figur 17, trimmad till 4 signifikanta variabler med $Q^2=0,806$.



Figur 17. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ^{137}Cs till vatten med relaterade variabler. Modell M272 har 4 signifikanta variabler.

Det går att trimma regressionsmodellen till en signifikant variabel (S17_Blåstång). Modell M273 har fortfarande ett väldigt högt $Q^2=0,72$ och är mindre resurskrävande än båda föregående modeller, utan att förlora mycket av prediktionsförmågan. Modellen är signifikant med $p=8,7e-7$.



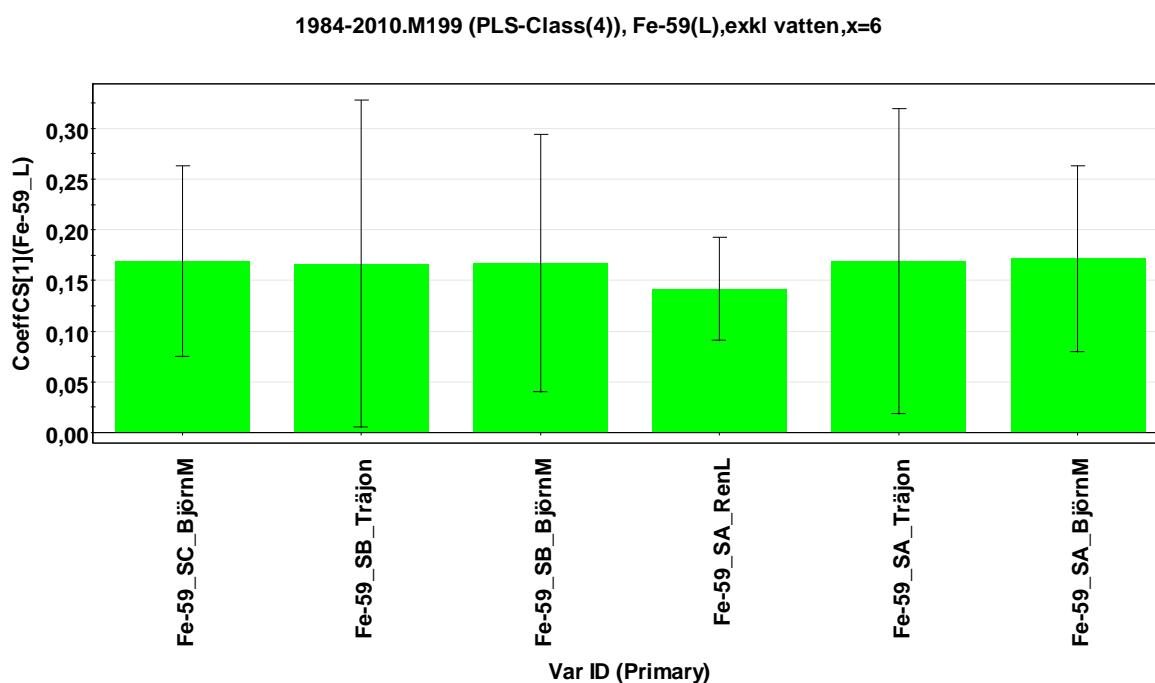
Figur 18. Observerat mot predikerat utsläpp av ^{137}Cs till vatten för modell M273, med en variabel S17_Blåstång. Observera logaritmisk skala.

Det finns alternativa en-variabellösningar för vattenutsläpp av ^{137}Cs , t.ex. abborre, som ger ett jämförbart, väldigt bra resultat, som det för blåstång.

5.6.4. ^{59}Fe

Utsläpp till luft

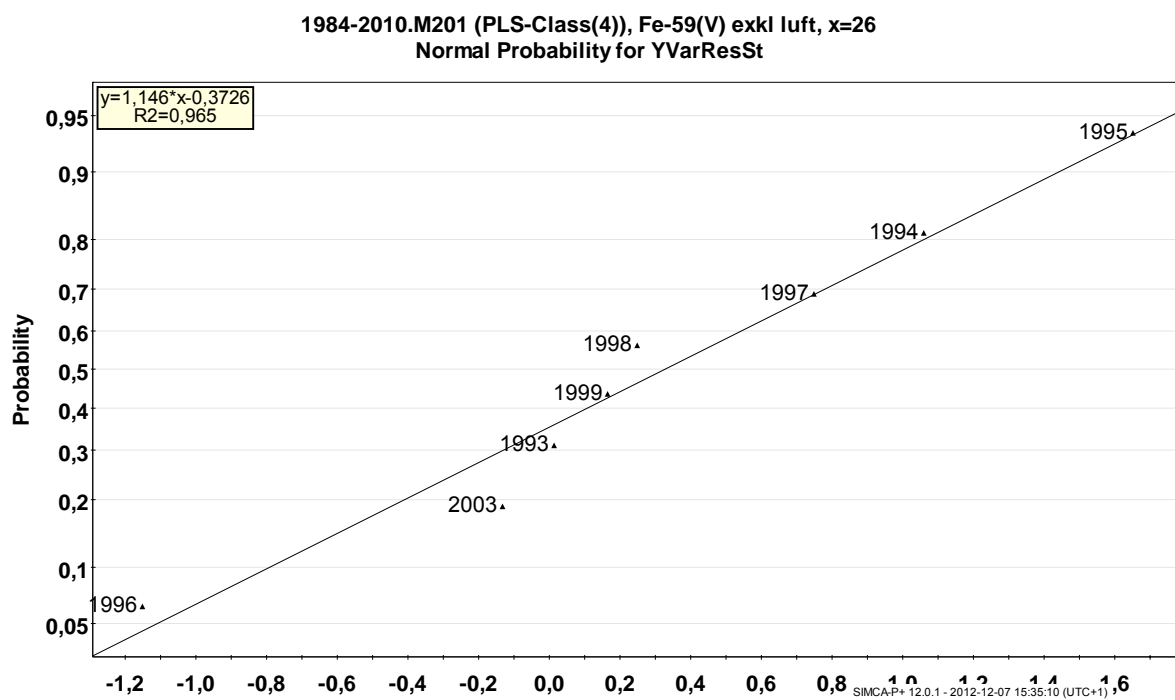
Ingen signifikant korrelationsmodell kunde beräknas mellan utsläpp av ^{59}Fe till luft och luftrelaterade omgivningsvariabler. Modell M199 ($Q^2=0,692$) baserad på 6 signifikanta omgivningsvariabler har $p=0,226$ och är således icke signifikant. Ytterligare trimning gav heller inga signifikanta modeller. Det kan vara värt att notera att luftutsläppen av ^{59}Fe har 50 % ”missing data” och relaterade omgivningsvariabler är ännu glesare, vilket naturligtvis påverkar resultatet. Mätosäkerheten har naturligtvis också betydelse, men bristen på data förefaller vara viktigast.



Figur 19. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ^{59}Fe till luft med relaterade variabler. Modell M199 har 6 signifikanta variabler.

Utsläpp till vatten

Inte heller utsläppen av ^{59}Fe till vatten, gav några signifikanta regressionsmodeller med tillhörande omgivningsvariabler. Det framgår av normalfördelningsploten i Figur 20 att det till antalet är relativt få observationer som är väl samlade runt den inlagda regressionslinjen. Det ser lovande ut, men det gick inte att finna någon signifikant modell.



Figur 20. Normalfördelningsplot för modell M201, för vattenutsläpp av ⁵⁹Fe.

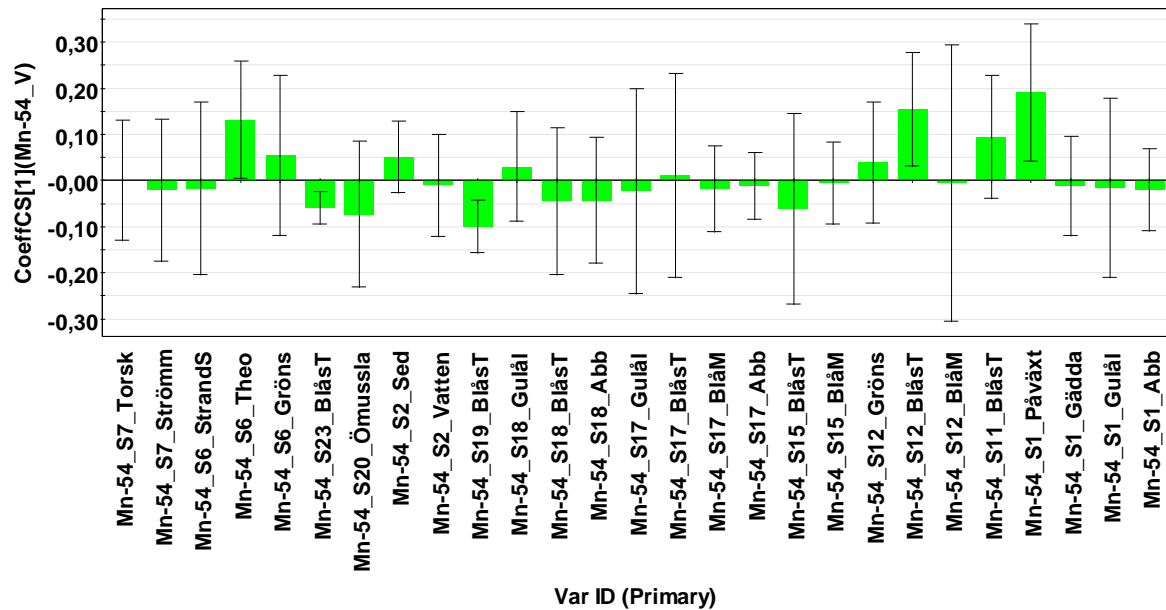
5.6.5. ⁵⁴Mn

Utsläpp till luft

Inte heller för utsläpp av ⁵⁴Mn till luft, gick det att finna någon signifikant modell.

Utsläpp till vatten

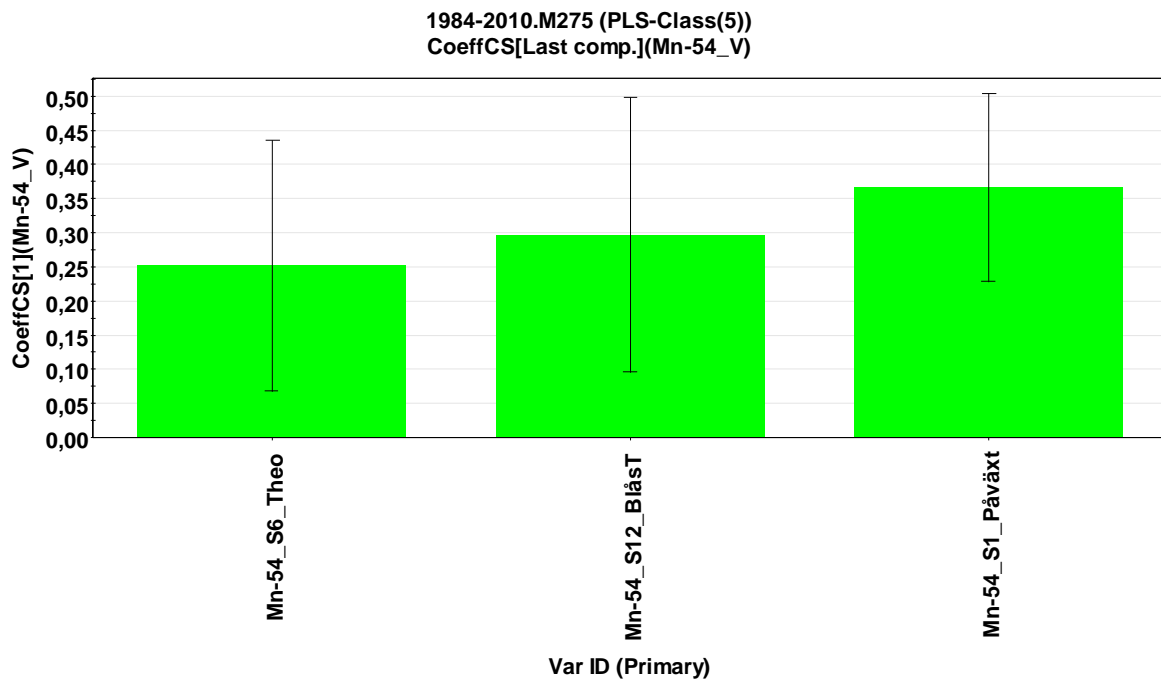
1984-2010.M230 (PLS-Class(5)), Mn-54(V)exkl luft, x=27



SIMCA-P+ 12.0.1 - 2012-12-06 16:22:33 (UTC+1)

Figur 21. Plot av regressionskoefficienter (scaled & centered) för utsläpp av ⁵⁴Mn till vatten med relaterade variabler.

Stora osäkerhetsintervall ses i Figur 21 för de flesta variablerna, men det finns ett fåtal variabler med positiv korrelation som är signifikanta. En trimmad modell baserad på dessa visas nedan.



Figur 22. Plot av regressionskoefficienter, modell M275, (scaled & centered) för utsläpp av ⁵⁴Mn till vatten, med 3 signifikanta variabler.

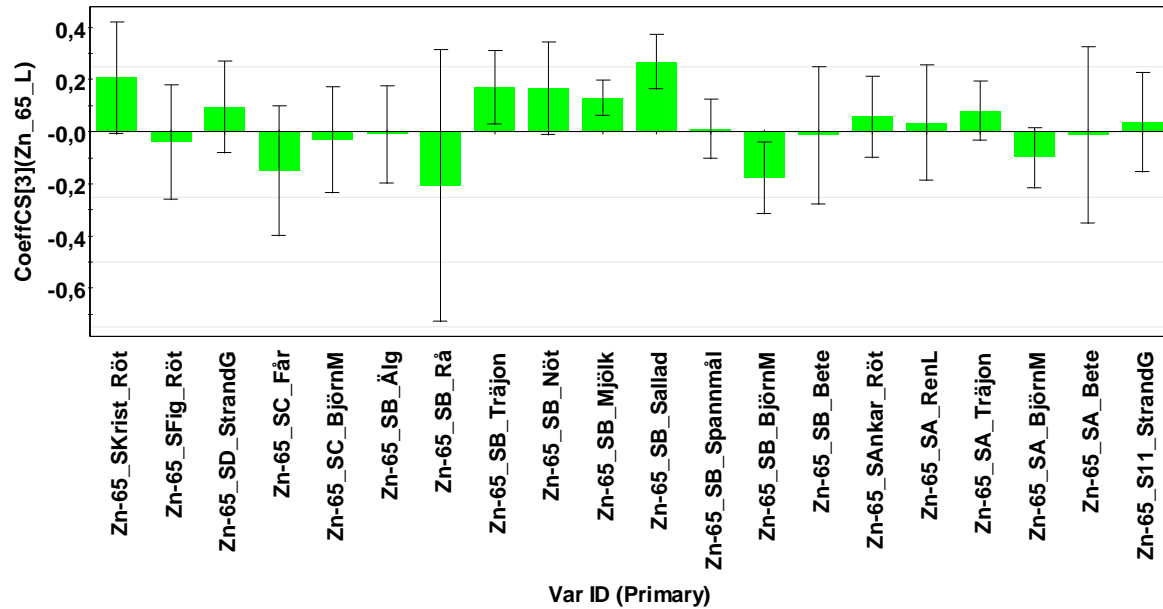
Modell M275 är signifikant ($p=4,0e-4$) och som synes är de tre variablerna också signifikanta. M275 har ett $Q^2=0,70$.

Ytterligare trimning till endast en signifikant variabel (S12_Blåstång) gav modell M276, som också är signifikant ($p=0,0012$), med $Q^2=0,642$. Även med S1_påväxt, som visar något mindre osäkerhet i Figur 22, fås en signifikant modell med $Q^2=0,618$, vilket också är väldigt bra. Relativt lite av prediktionsförmågan går förlo-rad när något av de signifikanta en-variabelsalternativen väljs. Det verkar således vara resursbesparande, vilket inte är helt betydelselöst.

5.6.6. ⁶⁵Zn

Utsläpp till luft

1984-2010.M236 (PLS-Class(6)), Zn-65(L) exkl vatten,x=20

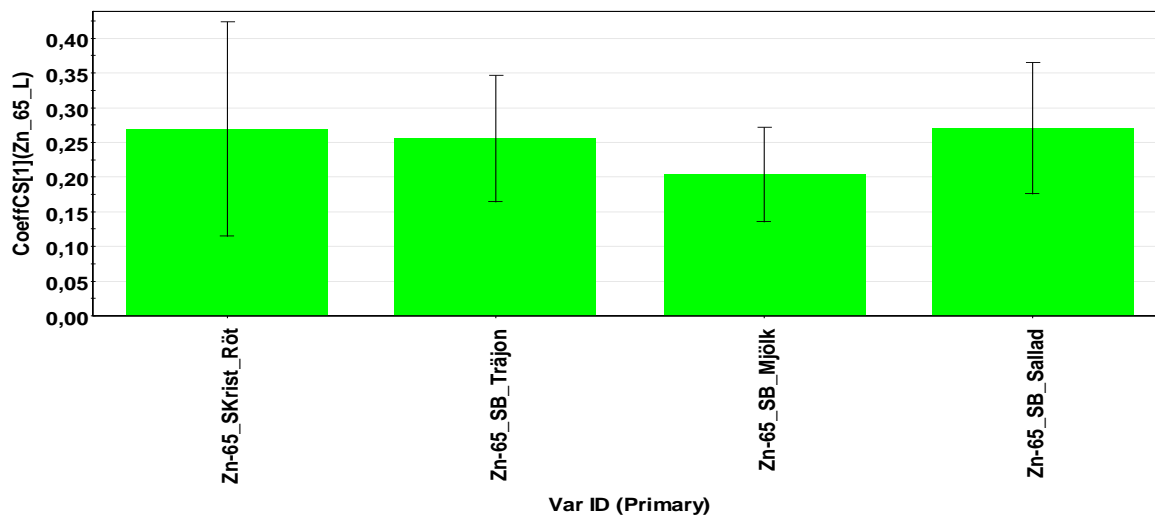


SIMCAP+ 12.0.1 - 2012-12-06 16:59:01 (UTC+1)

Figur 23. Plot av regressionskoefficienter, modell M236, (scaled & centered) för utsläpp av ⁶⁵Zn till luft.

I Figur 23 visas luftrelaterade omgivningsvariabler, där ett fåtal är signifikanta och positivt korrelerade till utsläppet av ⁶⁵Zn till luft. I Figur 24 visas en trimmad modell, M247, som är signifikant med $p=0,003$ och $Q2=0,763$. De fyra omgivningsvariablerna är också signifikanta.

1984-2010.M247 (PLS-Class(6)), Zn-65(L) exkl vatten,x=4



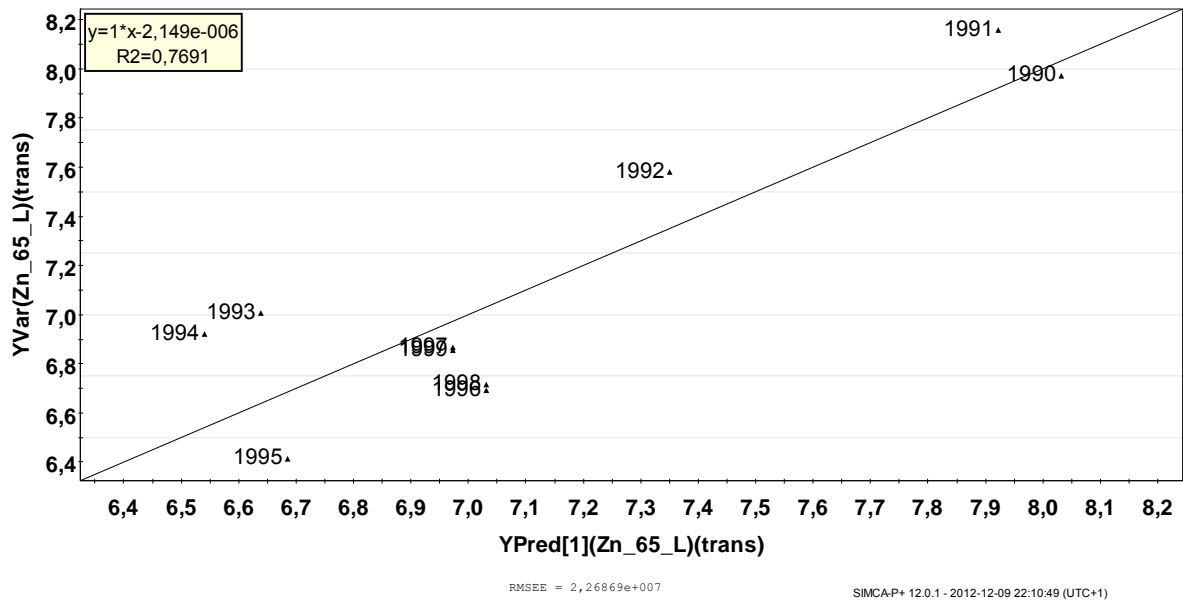
SIMCA-P+ 12.0.1 - 2012-12-06 17:07:11 (UTC+1)

Figur 24. Plot av regressionskoefficienter, modell M247, (scaled & centered) för utsläpp av ^{65}Zn till luft, med 4 signifikanta variabler.

Vidare trimning resulterade i modell M280, baserad på en signifikant variabel (SB_Träjon). M280 är signifikant med $p=0,012$ och $Q^2=0,714$. Detta är den första luftrelaterade modellen i den här studien som visas vara signifikant, trots stora mätosäkerheter för t.ex. mjölk och generellt mycket ”missing data”.

Tidigare kunde det observeras i Figur 3, att $^{65}\text{Zn(L)}$ har det högsta värdet längs PC1 av luftvariablerna/utsläppen längs PC1 och korrelerar i någon grad till vattenutsläppen. En fråga blir då, ”varför blir bara en av de 6 studerade fallen med utsläpp till luft signifikant korrelerad med de till luft relaterade omgivningsvariablerna?”. Det kan undersökas vidare om det t.ex. finns någon mätteknisk förklaring, eller om fysikalisk/biologisk halveringstid eller brist på kontinuitet i mätningarna påverkar. Vid en visuell översyn av samtliga de luftrelaterade variablerna har ^{137}Cs och de båda Co-isotoperna få men ändå något fler observationer än de för ^{65}Zn , men det har ändå inte gått att finna signifikanta korrelationsmodeller för dessa.

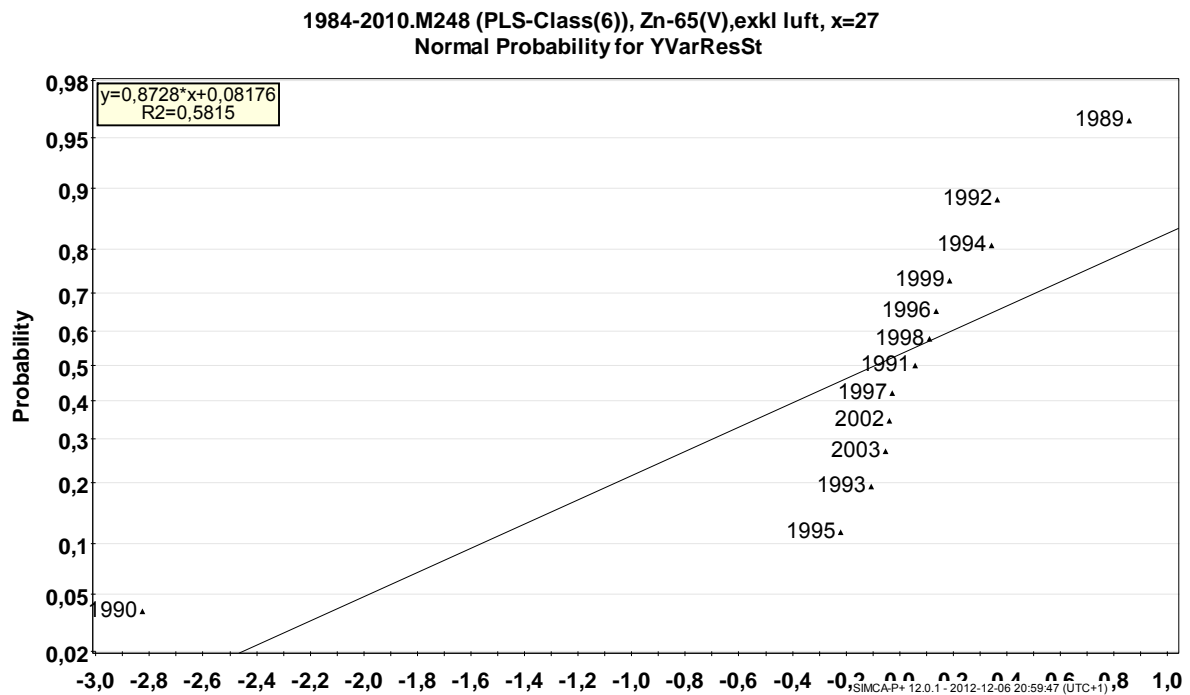
1984-2010.M280 (PLS-Class(6)),
Zn-65(L),x=1(SB_Träjon)



Figur 25. Observerat mot predikerat utsläpp av $^{65}\text{Zn}(\text{L})$ till luft, modell M280, med en omgivningsvariabel SB_Träjon. Obs logaritmisk skala.

I Figur 25 visas hur väl utsläppen av $^{65}\text{Zn}(\text{L})$ till luft kan predikteras. Notera att det inte finns några observationer före 1989 eller efter 1999, på grund av allt för få mätdata.

Utsläpp till vatten

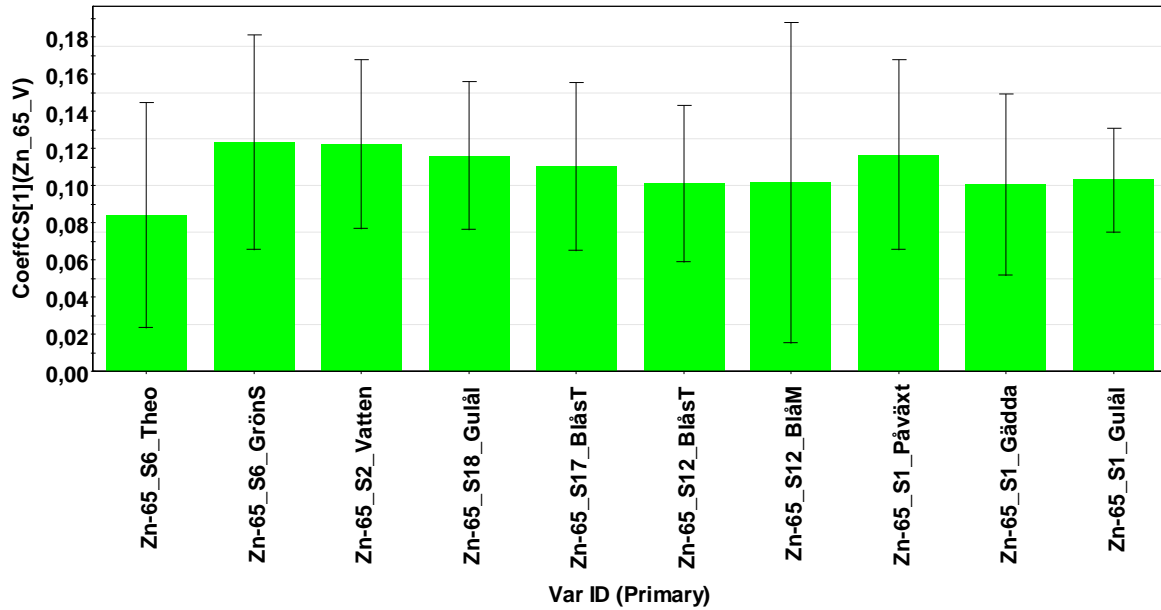


Figur 26. Normalfördelnings plot för modell M248, med 27 omgivningsvariabler, för utsläpp av ^{65}Zn till vatten

I Figur 26 verkar observationen 1990 vara något avvikande i något avseende (t.ex. extremt hög medelaktivitet i påväxtprovet vid station 1). Modellen M248, är inte strikt signifikant, men har ett $Q^2=0,74$.

Exkludering av observationen 1990 och några ytterligare observationer som innehåller extrema mätvärden, samt variabeltrimning leder fram till en signifikant modell, M254 ($p=0,003$), med 10 variabler och $Q^2=0,758$.

1984-2010.M254 (PLS-Class(6)), Zn-65(V),
exkl luft, -90,-93,-00,-01,-04, x=10, CoeffCS

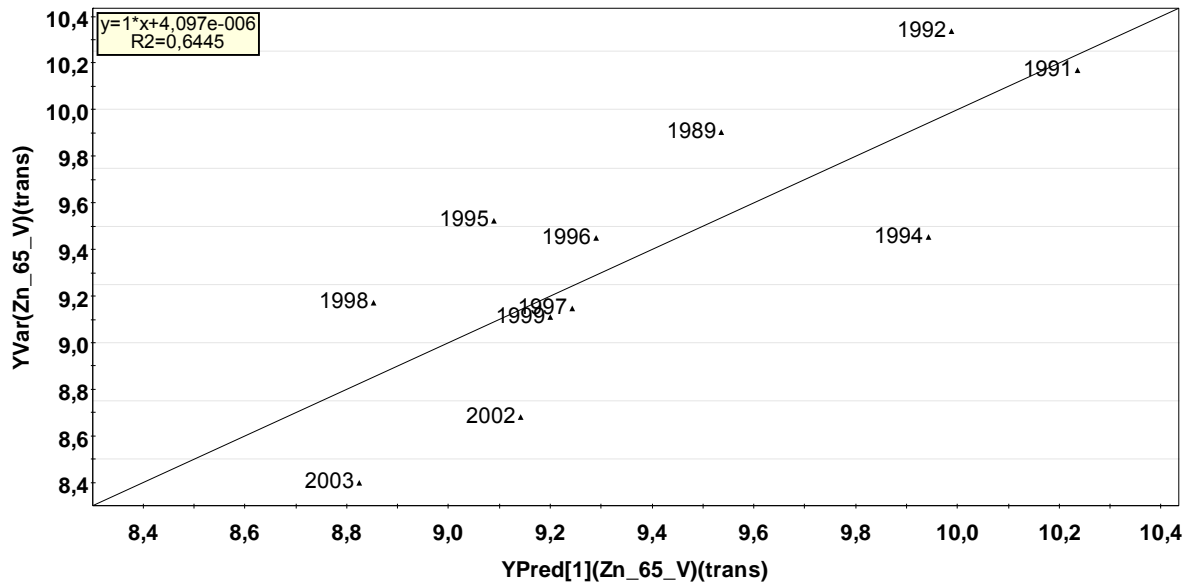


SIMCA-P+ 12.0.1 - 2012-12-06 21:20:42 (UTC+1)

Figur 27. Plot av regressionskoefficienter, modell M254, (scaled & centered) för utsläpp av ⁶⁵Zn till vatten, med 10 signifikanta variabler.

Ytterligare trimning till en signifikant omgivningsvariabel, Gulål vid station 1, ger modell M287 (p=0,026) med Q2=0,596.

1984-2010.M287 (PLS-Class(6)), Zn-65 (V), exkl luft, x=1 (S1_Gulål)



RMSEB = 4,87953e+009

SIMCA-P+ 12.0.1 - 2012-12-06 21:55:30 (UTC+1)

Figur 28. Observerat mot predikerat utsläpp av ⁶⁵Zn till vatten, modell M287, med en omgivningsvariabel S1_Gulål. Obs logaritmisk skala.

I Figur 28 visas att det går att prediktera utsläppet av ^{65}Zn till vatten, med relativt gott resultat, med en modell baserad på en omgivningsvariabel. Det verkar dock inte vara möjligt att finna någon annan en-variabellösning än den med Gulål vid station 1 med aktuella data.

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 1. Sammanställning av prediktionsförmåga för utsläppen till vatten av respektive nuklid för modeller med flera (n) variabler mot modeller med en variabel. För ^{59}Fe finns ingen signifikant modell till vare sig luft eller vatten. Observera att det är endast för $^{65}\text{Zn(L)}$ det finns signifikanta modeller för utsläpp till luft och luftrelaterade omgivningsvariabler.

Nuklid	Antal variabler (n)	Q2_n (%)	En variabel	Q2_1 (%)
$^{58}\text{Co(V)}$	3	91,7	S12 BlåsT	76,7
$^{60}\text{Co(V)}$	4	90,0	S12 BlåsT	89,5
$^{137}\text{Cs(V)}$	18	79,4	S17 BlåsT	72,0
$^{59}\text{Fe(V&L)}$	-	-	-	-
$^{54}\text{Mn(V)}$	3	70,0	S12 BlåsT	64,2
$^{65}\text{Zn(V)}$	10	75,8	S1 Gulål	59,6
$^{65}\text{Zn(L)}$	4	76,3	SB_Träjon	71,4

Tabell 2. Regressionskoefficienter (logaritmskala) för modeller med en signifikant variabel (se Tabell 1).

Modell	M269	M270	M273	-	M276	M287	M280
	^{58}Co	^{60}Co	^{137}Cs	^{59}Fe	^{54}Mn	$^{65}\text{Zn(V)}$	$^{65}\text{Zn(L)}$
Konstant (B)	8,848	8,269	3,895	-	8,015	8,549	6,222
Koefficient (k)	0,6796	1,125	3,159	-	1,106	0,9149	1,81

I Tabell 2 avviker regressionskoefficienterna för ^{137}Cs något, från mönstret av de övriga nukliderna. Modeller för ^{137}Cs ger inte lika bra resultat som för övriga signifikanta nuklid-modeller. Det kan inte uteslutas att det i den mätta aktiviteten för ^{137}Cs , i de olika provslagen, också inkluderas ”utsläpp” från andra sidan av Östersjön, och inte enbart aktivitet från OKG.

För ^{59}Fe saknas signifikanta regressionsmodeller både för utsläpp till luft och till vatten. Observera att det bara är för $^{65}\text{Zn(L)}$ som det går att finna en signifikant luftrelaterad modell.

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

Tabell 3. Mätta och predikterade värden för utsläpp till vatten och vattenrelaterade omgivningsaktivitet under år 1995 samt ett luftrelaterat utsläpp av $^{65}\text{Zn(L)}$. För ^{59}Fe finns ingen signifikant modell. Predikerat utsläpp baseras på mätt aktivitet i respektive variabel i Tabell 1. Predikerad aktivitet för en omgivningsvariabel baseras på mätta utsläpp till vatten eller luft.

Nuklid	Mätt utsläpp (Bq), 1995	Predikerat utsläpp (Bq), 1995	Mätt aktivitet (x=1)	Predikerad aktivitet (x=1)
^{58}Co	9,41E+09	1,14E+10	60,0	45
^{60}Co	4,30E+10	4,82E+10	140,0	126
^{137}Cs	4,95E+09	7,68E+08	38,0	69
^{59}Fe	-	-	-	-
^{54}Mn	5,25E+09	6,97E+09	45,0	35
$^{65}\text{Zn(V)}$	3,34E+09	1,23E+09	3,9	12
$^{65}\text{Zn(L)}$	2,58E+06	4,83E+06	1,8	1

Prediktioner av utsläppen i Tabell 3 visar relativt bra överensstämmelse med faktiskt mätta värden. Prediktionsfelen är klart mindre än en tiopotens. Bäst överensstämmelse ses för ^{60}Co , i linje med vad som indikeras av respektive modells Q2-värde. Den något större underskattningen av predikerat utsläpp av ^{137}Cs till vatten, kan indikera att den mätta aktiviteten inte enbart kan relateras till utsläpp från OKG. Den enda signifikanta modellen för luftutsläpp ses för $^{65}\text{Zn(L)}$. Uppenbarligen ger modellen bra prediktioner både när det gäller utsläpp och aktivitet i provslaget (SB_Träjon).

Det bör påpekas att de signifikanta modellerna i denna studie bör ge relativt bra prediktioner för utsläpp som ligger inom det utsläppsintervall som legat till grund för ”modellbygget”. Därmed skall 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 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.
- Några av de studerade vattenrelaterade variablerna visar att det går att uppskatta storleken på gjorda utsläpp, åtminstone för nukliderna ^{58}Co , ^{60}Co , ^{137}Cs , ^{54}Mn och ^{65}Zn till vatten. Mätningar på provslag som har resulterat i signifikanta samband är framför allt blåstång vid några mätstationer. Även några fiskarter bör kunna användas, vid bättre kontinuitet i dataserierna. Således krävs bättre kontinuitet i mätningarna på utvalda omgivningsvariabler.
- För luftrelaterade utsläpp är det inte lika klart att peka på provslag som kan användas i signifikanta regressionsmodeller. Någon/några av arterna träjon, rötslam, mjölk, björnmossa, renlav och i viss mån sallad och mjölk, skulle kunna vara ”referensarter”. Bedömning av vilken art som ger ”bästa” regressionsresultat kan göras när fler mätvärden finns tillgängliga för dessa. Det kan vara en fördel att, ur ett mätvärdesperspektiv, välja provslag som om möjligt är biologiskt aktiva under större delen av året och som är lätta att hantera.
- Ur ytterligare ett mätperspektiv är det inte nödvändigt att kontrollera omgivningen, gällande vattenrelaterade aktiviteter, i den omfattning som beskrivs i ”omgivningskontrollprogrammet” för att ha god kunskap om storleken av gjorda utsläpp av nukliderna som studerats i denna rapport. För luftrelaterade utsläpp behövs, som nämnts, mer kontinuitet i mätserierna, och kanske med fokus på några av ovan nämnda potentiella provslag.
- 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_mätstationer med regelbunden frekvens. Kanske veckovis eller varje månad. Viktigt dock att prover tas ”samtidigt”. Dessa mätdata kan även kombineras med meteorologiska data (t.ex. förhärskande vindriktning och vindhastighet, vattentemperatur och strömningshastighet). Därefter skulle en multivariat tidsserieanalys kunna ge en uppfattning om kinetik av upptag/deponering till olika provslag_provstationer.
- Regressionsstudier för ”svärmätta”/speciella nuklider och specifika upptag av vissa nuklider i något provslag, skulle kunna göras i laboratorieförsök tillsammans med ”lättmätta” nuklider med känd upptagshastighet i ett visst provslag. Eventuella modeller kan då användas för att uppskatta aktiviteten av dessa vid normal kontroll av övriga variabler i omgivningskontrollen. Resultat i rapporten indikerar att blåstång för vatten och kanske träjon för luft kunde vara intressanta provslag att starta med.

7. Slutsatser

- Data, som analyserats med multivariata verktyg (PCA och PLS), visar att rapporteringen av de kontrollerade utsläppen av radioaktiva nuklider (^{58}Co , ^{60}Co , ^{137}Cs , ^{59}Fe , ^{54}Mn och ^{65}Zn), från OKG och CLAB, till luft och vatten, har hög kontinuitet i tidsserierna 1984-2010.
- De kontrollerade utsläppen av radioaktiva ämnen till vatten tenderar att minska med tiden. För motsvarande utsläpp till luft är aktiviteten, i medeltal, oftast lägre än till vatten men den avtagande trenden är inte lika tydlig.
- Insamlade mätvärden på omgivningsvariabler, olika provslag och mätstationer, för samma nuklider är dessvärre, med några undantag, mycket dåligt representerade före år 1989 och efter 1999. Det visuella intrycket av dessa värden är också att det existerar stora mätosäkerheter, då stora skillnader i mätning av ett provslag samma dag observeras, men också konstanta värden under flera år för något provslag.
- Multivariatanalysen visar att det finns kvantifierbara samband mellan källutsläpp till vatten och till vattenrelaterade omgivningsvariabler, för 5 av de 6 studerade nukliderna, men endast ett samband för luftrelaterat utsläpp, $^{65}\text{Zn(L)}$.
- För ^{59}Fe finns inga signifikanta korrelationer till vare sig vatten eller luft.
- Signifikanta PLS-modeller med fler ingående variabler har något bättre prediktionsförmåga, Q^2 , än modeller med bara en variabel.
- Storleken på ett utsläpp (\mathbf{Y}) kan predikteras genom mätning av aktiviteten i en eller flera omgivningsvariabler. De använda PLS-modellerna (med en eller flera variabler) såväl som omgivningsvariablerna måste vara signifikanta. Observera att inte alla omgivningsvariabler, i denna rapport, kan användas.
- Aktiviteten i en omgivningsvariabel (\mathbf{X}) kan beräknas med PLS-modeller med en signifikant variabel vid känt källutsläpp. Efter att regressionsuttrycket för källutsläppet omformats till $\log(\mathbf{X}) = (\log(\mathbf{Y}) - \mathbf{B})/\mathbf{k}$, och \mathbf{Y} =mätt utsläpp/nuklid, och där \mathbf{k} och \mathbf{B} är konstanter för varje nuklid, provslag och mätstation. För signifikanta modeller fås relativt god överensstämmelse mellan observerat och predikerat värde. Prediktionsfelet är 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 en signifikant variabel mindre resurser och arbetsinsatser än modeller med fler variabler. Analytiskt/praktiskt förloras emellertid oftast bara en mindre del av prediktionsförmågan, Q^2 , vilket bör tas i beaktning när kostnad mot 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 aktiviteter i omgivningsvariabler utifrån uppmätta utsläpp. 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.



2013:12

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
Web: stralsakerhetsmyndigheten.se