



Marie Lillehammer

Usikkerhetsanalyse for utslipp av farlige stoffer

Notater

Innhold

1. Bakgrunn og formål.....	2
2. Metode.....	2
2.1 Fastsettelse av usikkerhet på utslippsfaktorer	2
2.1.1 Usikkerhet på utslippsfaktorer fra kategori I.....	4
2.1.2 Usikkerhet på utslippsfaktorer fra kategori II	5
2.1.3 Usikkerheter på utslippsfaktorer fra kategori III	5
2.1.4 Usikkerhet på utslippsfaktorer fra kategori IV	6
2.2 Fastsettelse av usikkerhet på aktivitetsdata	6
2.3 Aggregering av usikkerhet	6
2.4 Usikkerhet på endring.....	9
2.5 Følsomhetsanalyse.....	9
3. Resultater av usikkerhetsanalyse.....	9
4. Kommentarer til resultatene	10
4.1 Utslipp av CMR-forbindelser	10
4.2 Utslipp av kronisk toksiske forbindelser	11
4.3 Resultater fra følsomhetsanalysen	11
5. Kommentarer til metoden	12
6. Referanser	12
7. Vedlegg	12

1. Bakgrunn og formål

Statistisk sentralbyrå har, på grunnlag av data fra produktregisteret, estimert utslipp av farlige stoffer for årene 2002-2006. Resultatene er publisert i Rapporten 2008/24 fra Statistisk sentralbyrå (Kittelsen og Hansen, 2008). Nå er det ønskelig å knytte usikkerhetsmål til disse estimatene. Det er ønskelig å kjenne usikkerheten til både nivåtall og endring, som er en differens mellom nivåtall. På grunn av endring i metoden for beregning av utslipp er det ikke gjort usikkerhetsanalyse lenger tilbake enn til 2005. I dette notatet presenteres resultatene for årene 2005-2007 sammen med program og dokumentasjon til senere bruk.

Det er flere årsaker til at man ønsker å kjenne usikkerheten på utslippene av farlige stoffer. For det første er det viktig å kjenne presisjonen til de publiserte tallene for å kunne vurdere disse på en riktig måte, spesielt når det gjelder endringen i utslipp, som man ønsker å fastslå om er positiv eller negativ. I tillegg er det ønskelig å fastslå hvilke kilder til usikkerhet som bidrar mye til usikkerhet på aggregert nivå. En presentasjon av presisjonen på de publiserte tallene oppnås ved å gjøre punkttestimatene om til konfidensintervall, noe som kan gjøres ved å bruke tabellene som er presentert i kapittel 3. En vurdering av hvilke forutsetninger som bidrar til høy usikkerhet, sammen med resultatene av en følsomhetsanalyse, er presentert i kapittel 4.

Hovedresultatene i rapport 2008/24 er utslipp innen fire fareklasser: kreftfremkallende, mutasjonsfremmende og reproduksjonsforstyrrende stoffer (CMR), miljøfarlige stoffer, allergifremkallende stoffer og kronisk toksiske stoffer. Usikkerhet er beregnet på disse fire aggregerte nivåene. Metoden og programmet som er brukt kan enkelt brukes til å også aggregere usikkerhet på andre nivåer, som for eksempel industrisektor.

Utslipp av farlige stoffer beregnes ved å multiplisere import/produksjon av et stoff (aktivitet) med andelen av stoffet man regner med slippes ut i naturen (utslippsfaktor). For å beregne usikkerhet på utslippet anslås først usikkerheten på henholdsvis aktivitet og utslippsfaktor. Begge disse usikkerhetene kan være asymmetriske. Herfra trengs det to trinn med aggregering for å finne fram til usikkerheten på det totale utslippet. Først beregnes usikkerheten i utslipp av et enkelt stoff, som er et produkt av aktivitet og utslippsfaktor for det aktuelle stoffet. Deretter skal usikkerheten på den totale mengden stoffer innen en fareklasse beregnes, hvor det totale utslippet er en sum av mange delutslipp.

2. Metode

Hvordan det totale utslippet beregnes er nærmere beskrevet i rapport 2008/24. Som nevnt ovenfor, beregnes utslipp som aktivitet*utslippsfaktor. Aktivitet defineres som produksjon+import-eksport av et stoff og hentes fra produktregisteret. Utslippsfaktorene er anslått med bakgrunn i ulike kilder, beskrevet nedenfor og i kapittel 5.2 i Rapporten 2008/24. For å beregne usikkerhet på utslipp må man ta hensyn til usikkerheten både i aktivitetsdataene og i utslippsfaktorene.

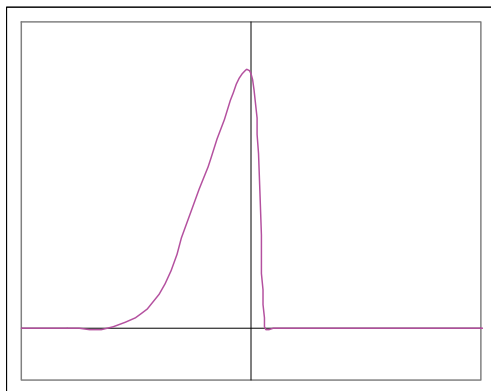
2.1 Fastsettelse av usikkerhet på utslippsfaktorer

Det er 4 hovedkategorier av utslippsfaktorer, sortert etter informasjonskilde (se kapittel 5.2 i rapport 2008/24). Kategoriene representerer også ulike nivå på sikkerhet, der kategori I har høyest sikkerhet på utslippsfaktoren, mens utslippsfaktor fra kategori IV er mest usikker.

Usikkerheten ble fastslått innen hver av disse gruppene på bakgrunn av tilgjengelig informasjon.

Mens noen utslippsfaktorer er antatt forventningsrette, er andre utslippsfaktorer satt til, eller nær, en antatt maksimumsverdi. Dette, sammen med et avgrenset parameterområde, gjør at en del usikkerheter ikke er symmetriske om utslippsfaktoren. En lignende problemstilling oppstår ved inflasjonsframskrivninger (Blix og Sellin, 1998). Den svenske riksbanken publiserer inflasjonsframskrivninger som er aggregert fra flere kilder. Slike inflasjonsframskrivninger publiseres med usikkerhetsmål og er gjerne asymmetriske. De antas da å følge en two-piece eller two-part normalfordeling (TPN), som er en fordeling som beskrives av tre parametre: mode, σ_1 og σ_2 . Fordelingen antar på hver side av mode å være proporsjonal med en normalfordeling med varians henholdsvis σ_1^2 og σ_2^2 . Fordelingen har mange av normalfordelingens egenskaper, sentralgrense-teoremet kan brukes, men samtidig kan fordelingen være asymmetrisk. Figur 1 er en illustrasjon av hvordan en TPN-fordeling kan se ut.

Figur 1 – Illustrasjon av en TPN-fordeling



Tetthetsfunksjonen til fordelingen er gitt ved:

$$f(x; \mu, \sigma_1, \sigma_2) = \begin{cases} C e^{-\frac{1}{2\sigma_1^2}(x-\mu)^2} & x \leq \mu \\ C e^{-\frac{1}{2\sigma_2^2}(x-\mu)^2} & x > \mu \end{cases}$$

Der $C = k(\sigma_1 + \sigma_2)^{-1}$, $k = \sqrt{2/\pi}$ og μ er modalverdien (ikke forventningsverdien).

TPN-fordelingen kan sees som en blanding av to normalfordelinger med samme forventningsverdi (μ) men med ulik varians. Hvilken fordeling X er trukket fra avhenger av verdien på X (om den er større eller mindre enn μ). Blandingsandelene er tilnærmet lik:

$$\frac{\sigma_1}{\sigma_1 + \sigma_2} \text{ og } \frac{\sigma_2}{\sigma_1 + \sigma_2}$$

I utslippsdataene finnes to variable, utslippsfaktor (f) og aktivitet (A). Vi antar at estimatene av disse er TPN-fordelte, ved at:

$$\hat{f}_i = f_i + X_i \quad X_i \sim \text{TPN}(0, \sigma_{f1i}, \sigma_{f2i})$$

$$\hat{A}_i = A_i + Y_i \quad Y_i \sim \text{TPN}(0, \sigma_{a1i}, \sigma_{a2i}).$$

A_i og f_i er modalverdiene i fordelingene av $\hat{A}_i$ og $\hat{f}_i$. Estimatenes er dermed ikke forventningsrette estimater, men den mest sannsynlige verdien av estimatet er den sanne verdien.

Dette gir:

$$E(X_i) = \sqrt{\frac{2}{\pi}}(\sigma_{f2i} - \sigma_{f1i})$$

$$E(Y_i) = \sqrt{\frac{2}{\pi}}(\sigma_{a2i} - \sigma_{a1i})$$

$$\text{Var}(X_i) = \left(1 - \frac{2}{\pi}\right)(\sigma_{f2i} - \sigma_{f1i})^2 + \sigma_{f1i}\sigma_{f2i}$$

$$\text{Var}(Y_i) = \left(1 - \frac{2}{\pi}\right)(\sigma_{a2i} - \sigma_{a1i})^2 + \sigma_{a1i}\sigma_{a2i}$$

2.1.1 Usikkerhet på utslippsfaktorer fra kategori I

Stoffer i denne kategorien brukes i store mengder. Her har industrien blitt konsultert for å fremskaffe så nøyaktige utslippsfaktorer som mulig. Disse utslippsfaktorene er spesifisert for en kombinasjon av produkt og industrisektor. De har generelt lav usikkerhet. Usikkerhet på utslippsfaktorene i denne kategorien er presentert i tabellen under.

Utslippsfaktor	σ_1 / σ_2	Begrunnelse
≤ 0.001	Utslippsfaktor*0.5/0.0005	Disse utslippsfaktorene har blitt dokumentert til å være svært lave. Konfidensintervallet er trunkert for å ikke anta negative verdier.
”Avrundet” oppover	1.5*avrunding/0.5*avrunding	”Avrundingen” skal gjenspeile at utslippsfaktoren er unøyaktig ¹ . Antar at avrundingen er på ett standardavvik.
Intervall	(Estimat-nedre grense)*0.5 / (estimat-øvre grense)*0.5	Er utslippsfaktoren opprinnelig angitt som et intervall antas dette å være 95% konfidensintervall.
Oppgitt tall 0.06	0.0025/0.0025	Anslaget antas å være avrundet til nærmeste 0.01, altså å ligge mellom 0.055 og 0.065.
Oppgitt tall 0.1	0.025/0.025	Anslaget antas å være avrundet til nærmeste 0.1, altså å ligge mellom 0.05 og 0.15.
1	0.0005/0	Her er det veldokumentert at så godt som alt slippes ut. Antar likevel en liten negativ usikkerhet.

Når industrien har oppgitt en utslippsfaktor med flere desimaler enn det man tror gjenspeiler den virkelige usikkerheten er utslippsfaktoren ”rundet av” oppover før beregning av totalt utslipp. Etter avrundning skal antall desimaler samsvare bedre med usikkerhet på utslippsfaktoren.

2.1.2 Usikkerhet på utslippsfaktorer fra kategori II

Disse utslippsfaktorene er hentet fra den svenske SMED-rapporten (Skårman et al, 2006), hvor også usikkerhet på utslippsfaktorene er oppgitt.

Utslippsfaktor	σ_1 / σ_2	Begrunnelse
Alle	Utslippsfaktor*0.1/ Utslippsfaktor*0.1	Kilden for disse utslippsfaktorene (SMED-rapporten) angir en usikkerhet på disse på 20%. Usikkerheten er trunkert hvis estimat+usikkerhet ville overstige grenseverdien 1.

2.1.3 Usikkerheter på utslippsfaktorer fra kategori III

Utslippsfaktorer i denne kategorien, med unntak av utslippsfaktor 0.0025, stammer fra en kategorisering av utslipp i høyt, middels og lavt, fra KemI-rapporten (Fischer et al, 2005). Etter hvert er også en fjerde kategori inkludert, svært lavt, som brukes om råvarer, der stoffet inngår i et nytt produkt og eneste utslipp er mulige "lekkasjer". Usikkerheten er her fastsatt på en slik måte at kvantifiseringen av de ulike kategoriene av utslipp til sammen skal dekke hele området 0 til 1, også når kategorien svært lavt holdes utenfor. Utslippsfaktoren 0.0025 for brenslere er inkludert i kategori III, til tross for at den er beregnet og ikke resultat av en kategorisering, siden KemI-rapporten er kilde også for denne utslippsfaktoren.

Utslippsfaktor	σ_1 / σ_2	Begrunnelse
0.0001	0.00005/0.0005	Denne kategorien brukes om antatte råvarer, der produktet inngår i et ferdig produkt og svært små mengder slippes ut. Her brukes samme usikkerhet som ved veldig lavt utslipp i kategori I.
0.0025	0.00125/0.00125	Denne utslippsfaktoren brukes på brensel, og er beregnet som et gjennomsnitt av 0.000005 og 0.005. Usikkerheten er satt slik at hele intervallet dekkes. For detaljer: se kapittel 4.3.
0.1	0.05/0.05	Denne faktoren er en kvantifisering av "lavt utslipp" og laveste mulige faktor på en del produkter. Derfor er det rimelig å la konfidensintervallet gå ned til 0 ved å sette usikkerheten lik utslippsfaktoren.
0.5	0.15/0.15	Denne utslippsfaktoren brukes som en kvantifisering av "middels" utslipp. Den antatte usikkerheten gir et 95% konfidensintervall på [0.2,0.8], dersom en antar at verdiene stammer fra en normalfordeling. Det er også tilnærmet usikkerheten en får om en antar en uniform fordeling mellom 0.2 og 0.8.
1	0.07/0	Dette er en kvantifisering av "høyt utslipp", og dataene antas å stamme fra en blandet fordeling. En andel av produktene har sant utslipp på 1, mens de resterende har et sant utslipp på over 0.8. En andel har da usikkerhet ≈ 0 , mens en andel har usikkerhet ≈ 0.2 (n). En videre antagelse om at ca 1/3 har utslipp akkurat 1, gir en usikkerhet på 0.14 ($\sigma_1=0.07$)

2.1.4 Usikkerhet på utslippsfaktorer fra kategori IV

Dersom det er registrert aktivitet der industrisektor eller produkt er angitt som ”ukjent” blir utslippsfaktoren satt til 1. Dette er tilfeller der det mangler informasjon om utslippsfaktor, så usikkerheten er svært stor.

Utslippsfaktor	σ_1 / σ_2	Begrunnelse
1	0.5/0	Denne utslippsfaktoren brukes der man ikke har noen informasjon om utslippet. Den antas å i realiteten kunne anta en hvilken som helst verdi, så den har en negativ usikkerhet på 1 ($\sigma_1=0.5$).

2.2 Fastsettelse av usikkerhet på aktivitetsdata

Det er ikke tatt hensyn til eventuell underdekning i produktregisteret. Skårman et al (2006) fant at svenske aktivitetsdata hadde en usikkerhet på omtrent 15%. Det norske produktregisteret antas å være sammenlignbart med det svenske, så usikkerheten i norske aktivitetsdata antas å være 15%.

For noen produkter er det mulighet for å foreta forenklet deklarasjon, der mengde blir deklart som ”fra – til” tall. I disse tilfellene blir max-mengde brukt til å beregne totalt utslipp. Den positive usikkerheten er satt til 15% som ved andre aktivitetsdata, mens den negative usikkerheten antas å være bredden på intervallet. Minste mengde blir satt til 0 dersom den er negativ.

2.3 Aggregering av usikkerhet

De antatte parametrene i fordelingen av komponentene $\hat{A}_i$ og $\hat{f}_i$ brukes til å beregne forventning og varians på utslipp ($\hat{U}_i$) som videre kan brukes til å estimere usikkerhet på et hvilket som helst aggregert nivå.

U_i – utslipp av stoff i, $U_i = A_i f_i$.

$$\hat{U}_i = \hat{A}_i \hat{f}_i = (A_i + Y_i)(f_i + X_i) = A_i f_i + A_i X_i + f_i Y_i + X_i Y_i$$

$$\hat{U}_i - U_i = A_i X_i + f_i Y_i + X_i Y_i$$

Vi kan da finne forventning og varians til $\hat{U}_i - U_i$.

$$E(\hat{U}_i - U_i) = A \sqrt{\frac{2}{\pi}} (\sigma_{f2i} - \sigma_{f1i}) + f \sqrt{\frac{2}{\pi}} (\sigma_{a2i} - \sigma_{a1i}) + \frac{2}{\pi} (\sigma_{f2i} - \sigma_{f1i}) (\sigma_{a2i} - \sigma_{a1i})$$

$$\begin{aligned} Var(\hat{U}_i) &= EVar(\hat{A}_i \hat{f}_i | \hat{f}_i) + VarE(\hat{A}_i \hat{f}_i | \hat{f}_i) = E(\hat{f}_i^2) Var(\hat{A}_i) + (E\hat{A}_i)^2 Var(\hat{f}_i) \\ &= Var(\hat{f}_i) Var(\hat{A}_i) + (E\hat{f}_i)^2 Var(\hat{A}_i) + (E\hat{A}_i)^2 Var(\hat{f}_i) \\ &= Var(X_i) Var(Y_i) + (f_i + EX_i)^2 Var(Y_i) + (A_i + EY_i)^2 Var(X_i) \end{aligned}$$

Her er A og f de sanne verdiene for aktivitet og utslippsfaktor. Et forventningsrett estimat for A er $\hat{A} = E(Y)$, tilsvarende for f . Dette gir en estimert varians på:

$$\text{Var}(\hat{U}_i) = \text{Var}(X_i)\text{Var}(Y_i) + \hat{f}_i^2 \text{Var}(Y_i) + \hat{A}_i^2 \text{Var}(X_i).$$

$\text{Var}(\Sigma \hat{U}_i)$ finnes ved $\Sigma \text{var}(\hat{U}_i)$, forutsatt at alle $\hat{U}_i$ er uavhengige. Tilsvarende er $E(\Sigma \hat{U}_i) = \Sigma E(\hat{U}_i)$. I tillegg til den aggregerte variansen, trenger vi den aggregerte skjevheten for å kunne angi riktig asymmetri i konfidensintervallet for $\Sigma \hat{U}_i$.

Dersom man antar at $\Sigma \hat{U}_i$ er TPN-fordelt med parametre $(\Sigma U_i, \sigma_1, \sigma_2)$ så kan σ_1 og σ_2 bestemmes ved:

$$\text{Var}(\Sigma \hat{U}_i) = \left(1 - \frac{2}{\pi}\right)(\sigma_2 - \sigma_1)^2 + \sigma_1 \sigma_2$$

$$E(\Sigma \hat{U}_i) = \sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1)$$

(ligning 12 og 13 i Blix og Sellin (1998)).

Løsningen på denne blir en annengradslikning, der vanligvis bare en av løsningene gir positive σ_1 og σ_2 .

La z_α være øvre α -punkt i $N(0,1)$ -fordelingen. Et konfidensintervall kan deretter utledes ved å bruke at $P(\Sigma \hat{U}_i > \Sigma U_i + z_\alpha \sigma_2) = \frac{2\sigma_2}{\sigma_1 + \sigma_2} \alpha$.

Deretter må z_α bestemmes slik at $\frac{2\sigma_2}{\sigma_1 + \sigma_2} \alpha = 0.025$. Tilsvarende for den nedre grensen.

Denne metoden for å beregne skjevhet på resultatet er brukt i inflasjonsfremskrivninger (Blix og Sellin, 1998), men som Wallis (2005) påpeker er det ikke bevist at summen av TPN-fordelte variable er TPN-fordelt.

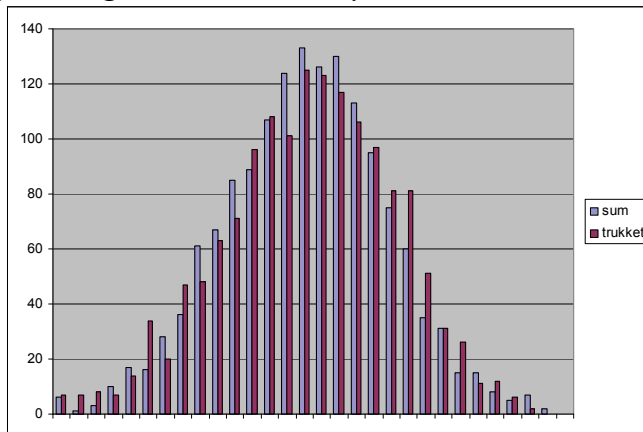
En annen mulighet er å anta at summen av de TPN-fordelte variablene er normalfordelt, siden summen av uavhengige variable vanligvis nærmer seg normalfordeling. Skjevheten i konfidensintervallet skyldes i tilfelle ikke en skjev fordeling, men at estimatet ikke er forventningsrett slik at konfidensintervallet er forskjøvet. For å beregne et slikt konfidensintervall beregnes varians som summen av varians og forventning som summen av forventningene etter samme formel som ovenfor. Et konfidensintervall beregnes symmetrisk om forventningen under forutsetning av at summen av de TPN-fordelte variablene er normalfordelt.

For å teste hvilken (om noen) av de foreslåtte metodene gir et riktig konfidensintervall trakk vi 15 TPN-fordelte variable og summerte disse. Dette ble gjentatt 1500 ganger for å få et datasett med 1500 verdier, fordelt etter samme fordeling som summen av 15 TPN-fordelte variable. Fordelingen av verdiene ble sammenlignet med en TPN-fordeling med σ_1 og σ_2 beregnet etter metoden fra Blix og Sellin (1998) og empirisk konfidensintervall ble sammenlignet med konfidensintervall dersom man forutsatte at summen var TPN-fordelt og med konfidensintervall dersom man forutsatte at summen var normalfordelt. Simuleringen ble gjentatt 10 ganger så man hadde 10 empiriske konfidensintervall.

Dersom skjevhetene blant de 15 opprinnelige TPN-fordelte variablene utjevnet hverandre slik at summen blir tilnærmet forventningsrett gir begge de to foreslåtte metodene tilfredsstillende

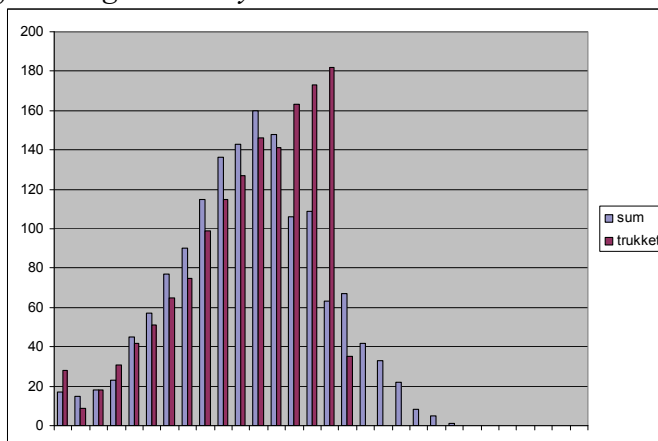
resultat. TPN-fordelingen nærmer seg en normalfordeling når forskjellen mellom σ_1 og σ_2 er liten. Forskjellen mellom observasjoner trukket fra en TPN-fordeling og observasjonene som framkommer ved summen av 15 TPN-fordelte variable er illustrert i figur 2.

Figur 2 – sammenligning av den empiriske fordelingen til en variabel som er en sum av 15 TPN-fordelte variable og til en variabel som er TPN-fordelt med σ_1 og σ_2 beregnet etter metoden fra Blix og Sellin (1998). Fordelingene er valgt slik at den estimerte TPN-fordelingen blir tilnærmet symmetrisk.



Dersom mange av de opprinnelige TPN-fordelte variablene hadde skjevhet med samme fortegn, slik at de ulike skjevhetene ikke utjevnet hverandre viser figur 3 en tydelig forskjell på fordelingen av summen av 15 TPN-fordelte variable og fordelingen av en variabel trukket fra en beregnet TPN-fordeling (beregnet som beskrevet i Blix og Sellin, 1998). Det beregnede konfidensintervallet var også noe forskjøvet, i forhold til gjennomsnittet av de empiriske. Et konfidensintervall beregnet under forutsetning av normalfordeling ble likt gjennomsnittet av de empiriske konfidensintervallene. Vi antar derfor at summen av TPN-fordelte variable antar en normalfordeling, symmetrisk fordelt om forventningsverdien. Konfidensintervallet er likevel ikke symmetrisk rundt punkttestimatet, siden punkttestimatet er en sum av modalverdier og ikke forventningsrette verdier.

Figur 3 – sammenligning av den empiriske fordelingen til en variabel som er en sum av 15 TPN-fordelte variable og til en variabel som er TPN-fordelt med σ_1 og σ_2 beregnet etter metoden fra Blix og Sellin (1998). Fordelingene er valgt slik at den estimerte TPN-fordelingen blir asymmetrisk.



Usikkerheten på utslipp ble beregnet både under forutsetning av at summen var TPN-fordelt og under forutsetning om normalfordeling for også å kunne vurdere hvordan modellantagelsene påvirker resultatene.

2.4 Usikkerhet på endring

Det er liten forskjell på utslippsfaktorene fra år til år. Fra og med 2005 til og med 2007 er det ingen endring i disse så variasjon i utslipp skyldes variasjon i aktivitet. Usikkerheten på differansen mellom år vil da knyttes til usikkerhet i variasjon i aktivitet, dvs. vi bruker den betingede varians for differansen gitt estimert utslippsfaktor som mål på usikkerheten i endring.

$$Var(\hat{f}_i(\hat{A}_{i2} - \hat{A}_{i1}) | \hat{f}_i) = \hat{f}_i^2(Var(\hat{A}_{i2})) + \hat{f}_i^2(Var(\hat{A}_{i1})),$$

hvor $\hat{A}_{it}$ er aktivitet for stoff i ved tidspunkt t .

Variansen på endring beregnes for hvert stoff og summeres, på samme måte som variansen på nivå. Konfidensintervall for endring antas å være symmetrisk rundt forventet endring, som avviker fra estimert endring med forventet avvik ved $t=2$ minus forventet avvik ved $t=1$.

2.5 Følsomhetsanalyse

Enkelte kombinasjoner ble identifisert som store bidragsytere til høy usikkerhet. Her ble det testet hvor mye de aggregerte usikkerhetene ville endre seg dersom usikkerheten i utslippsfaktor eller aktivitet for disse kombinasjonene ble redusert, altså dersom man kunne fremskaffe sikrere data.

3. Resultater av usikkerhetsanalyse

Usikkerhet på nivå på utslipp av farlige stoffer. Negativ og positiv usikkerhet angir avstanden fra estimatet til henholdsvis nedre og øvre grense på 95% konfidensintervallet (tonn), estimert under forutsetning om at summen av TPN-fordelte variable også er TPN-fordelt.

	2005		2006		2007	
	Negativ	Positiv	Negativ	Positiv	Negativ	Positiv
CMR	3 903	4 656	3 369	4 279	4 975	6 348
Miljøfarlig	475 ^a	0 ^a	587	45	604	56
Allergifremkallende	167 ^a	0 ^a	168 ^a	0 ^a	189	16
Kronisk toksisk	254	183	283	210	306	192

^a Den estimerte σ_2 var negativ. σ_2 ble derfor satt til 0 og et ensidig konfidensintervall som ga riktig varians ble estimert.

Usikkerhet på nivå på utslipp av farlige stoffer, estimert under forutsetning om at summen av TPN-fordelte variable er normalfordelt.

	2005		2006		2007	
	Negativ	Positiv	Negativ	Positiv	Negativ	Positiv
CMR	4 118	4 622	3 603	4 211	5 326	6 244
Miljøfarlig	617	79	526	156	543	169
Allergifremkallende	236	8	239	9	169	51
Kronisk toksisk	248	200	276	228	293	217

Siden de samme variansberegningene ligger til grunn for begge beregningene av usikkerhet er størrelsesorden på usikkerheten den samme uavhengig av hva slags fordeling summen av utslippene antas å følge. Det viser seg at antagelsen om at summen er TPN-fordelt gir større skjevhet i usikkerhetene enn antagelsen om at summen er normalfordelt, noe som bekrefter resultatene fra simuleringen, der forutsetning om TPN-fordeling overestimerte skjevheten i usikkerhet. Ved noen kombinasjoner av år og fareklasse førte forutsetningen om TPN-fordelt sum til negativt estimat for σ_2 , noe som også bekrefter at denne forutsetningen ikke er oppfylt. Metoden som i de tilfellene ble brukt til å beregne et ensidig konfidensintervall gjorde at man beholdt riktig varians men fikk feil forventningsverdi på summen. Det var ikke mulig å finne en TPN-fordeling i disse tilfellene som ga riktig forventning og varians.

Aggregert usikkerhet på endring i utslipp innen hver fareklasse (tonn):

	2005-2006		2006-2007		2005-2007	
	Negativ	Positiv	Negativ	Positiv	Negativ	Positiv
CMR	752	856	749	1 059	723	1 137
Miljøfarlig	194	362	240	236	187	351
Allergifremkallende	103	101	36	148	35	145
Kronisk toksisk	58	58	81	53	79	51

Usikkerhetene ble beregnet under forutsetning om at summen av usikkerhetene er normalfordelt, symmetrisk om forventet endring. 95% konfidensintervallene blir skjeve dersom forventningsskjevheten er forskjellig mellom de aktuelle årene. For at en endring skal være statistisk signifikant må endringen være større enn usikkerheten. Dersom det er en negativ endring (en nedgang) må den overstige den positive usikkerheten for å være signifikant. En positiv endring må overstige den negative usikkerheten for å være signifikant.

4. Kommentarer til resultatene

Det er lav usikkerhet på utslippene av miljøfarlige og allergifremkallende stoffer, disse ligger på rundt 4-7 % av utslippsnivået. Utslippt av kronisk toksiske stoffer og CMR har høyere relativ usikkerhet (20-35%). For disse fareklassene presenteres derfor noen stoffer/forutsetninger som bidrar mye til den høye usikkerheten. Ved å endre disse forutsetningene, ved for eksempel å fremskaffe kategori I utslippsfaktorer for stoffer som brukes i store kvanta, kan usikkerheten reduseres betraktelig.

4.1 Utslipp av CMR-forbindelser

B55100 og B55300 brukes i store mengder, og er en kilde til store usikkerhetstall (spesielt B55100). Usikkerheten kan reduseres mye ved å fremskaffe sikrere data for disse varene. Både sikrere aktivitetsdata og sikrere utslippsfaktorer kan redusere usikkerheten. Effekten av sikrere utslippsfaktorer er beskrevet i følsomhetsanalysen i kapittel 4.3.

B15315 har forenklet deklarasjon, og det har utslippsfaktor 0.5 fra kemi-rapporten. Det medfører høy usikkerhet både på aktivitet og utslippsfaktor. Her kan det være noe å hente på å fremskaffe sikrere data.

4.2 Utslipp av kronisk toksiske forbindelser

CAS nr 108-88-3 (vare L10202, M10300, M05243, O15100) viser seg å være en stor bidragsyter til usikkerheten på utslipp av kronisk toksiske forbindelser. Disse har utslippsfaktorer fra SMED-rapporten som gir høy usikkerhet ved høy utslippsfaktor. Andre betydelige kilder til usikkerhet på utslipp av denne typen forbindelser er brensl (B55100 og B55300) og en del forbindelser som brukes i store kvanta og som har fått utslippsfaktor 0.5 fra kemi-rapporten.

4.3 Resultater fra følsomhetsanalysen

Effekten av å redusere usikkerheten ble testet for tre kilder til høy usikkerhet: forenklet deklarasjon, høye utslippsfaktorer fra SMED-rapporten og utslippsfaktoren for brensel (B55100).

Forenklet deklarasjon: Dersom midtpunktet i intervallet hadde vært brukt som estimat ved forenklet deklarasjon, og intervallet hadde vært betraktet som et 95 % konfidensintervall, ville dette omtrent halvert usikkerheten i aktivitet for disse stoffene. Det må likevel vurderes om dette er hensiktsmessig, for det kan føre til underestimering hvis den sanne mengden ligger nærmere max-mengde enn min-mengde. En slik endring kan føre til litt høyere positiv usikkerhet, men vil gi lavere negativ usikkerhet. Følsomhetsanalysen viste at en slik endring kunne redusere usikkerheten med 4-9 % for fareklassene miljøfarlige stoffer og allergifremkallende stoffer. Det ville ikke ha noen effekt på usikkerheten i de andre fareklassene. Tiltaket vil dermed ha liten effekt på usikkerheten, siden det først og fremst påvirker usikkerheter som allerede er lave. Tiltaket vil imidlertid bidra til å forskyve konfidensintervallene slik at de blir mer symmetriske om estimatet, da det vil redusere forskjellen mellom estimert og forventet utslipp (reducere bias), dette under forutsetning av at midtpunktet i intervallet oppgitt ved forenklet deklarasjon er et forventningsrett estimat for aktiviteten.

Utslippsfaktorer fra SMED-rapporten: Ved å sette usikkerheten lik 20 % av utslippsfaktoren får man høy usikkerhet på de høye utslippsfaktorene. På utslippsfaktorer fra andre kilder er derimot usikkerheten høyest på utslippsfaktorer rundt 0.5 og lavest når utslippsfaktoren nærmer seg 0 eller 1. Dette virker intuitivt mer logisk, siden man kan fastslå utslippet med stor sikkerhet hvis tilnærmet alt eller ingenting slippes ut. Dersom usikkerheten settes til 20 % av utslippsfaktoren dersom utslippsfaktoren var mindre eller lik 0.5, og 20 % av (1-utslippsfaktor) der utslippsfaktoren er større reduseres usikkerheten med 2-3 % for allergifremkallende stoffer, men det har ingen effekt på usikkerheten i andre fareklasser. Dette tiltaket vil altså ha liten effekt.

Brensel (B55100): I Kemi-rapporten ble det framskaffet utslippsstatistikk på bensin/diesel (utslippsfaktor=0.005) og på fyringsolje (utslippsfaktor=0.000005). Andre brensl ble antatt å ha utslippsfaktor som lå midt mellom disse (utslippsfaktor=0.0025). I den norske utslippsstatistikken er utslippsfaktoren 0.0025 for andre brensl mye brukt. Nå er usikkerheten antatt å være 0.0025, med bakgrunn i at de såkalte "andre brensl" kan ligge nærmere enten fyringsolje eller bensin/diesel. Dette er den aller største kilden til usikkerhet. Å fremskaffe egen statistikk på dette kan redusere usikkerheten på utslipp av CMR-forbindelser kraftig. Dersom usikkerheten på denne faktoren hadde vært redusert til 0.0005 ville usikkerheten på nivået av CMR-utslipp vært redusert med ca 72-75 %. Å fremskaffe utslippsstatistikk for B55100 vil dermed ha stor effekt på usikkerheten, først og fremst for CMR-forbindelser.

5. Kommentarer til metoden

Simuleringsforsøket tok utgangspunkt i at hver av de summerte variablene var TPN-fordelt, men usikkerheten på $\hat{U}_i$, som er et produkt av $\hat{A}_i$ og $\hat{f}_i$, er ikke vist å være TPN-fordelt. Dette er imidlertid mindre problematisk dersom man forutsetter at summen skal være normalfordelt. Summen av mange uavhengige variable vil nærme seg normalfordelingen også om hver av variablene har en annen fordeling enn antatt.

Forutsetningen som er gjort og som kan påvirke resultatet, i tillegg til at summen er normalfordelt, er at de opprinnelige faktorene $\hat{A}_i$ og $\hat{f}_i$ er TPN-fordelt. Dette gjør at vi på bakgrunn av et asymmetrisk konfidensintervall kan beregne varians på disse komponentene. Det er ikke undersøkt hvordan det vil påvirke resultatet dersom de skjevt fordelte komponentene egentlig stammer fra en annen fordeling.

6. Referanser

Blix, M og P Sellin. 1998. Uncertainty bands for inflation forecasts. Economics Department, Sveriges Riksbank

Fischer, S, Å Almkvist, E Karlsson og M Åkerblom. 2005. Sluttrapport från projektet: Framtagande av produktregisterbaserat ExponeringsIndex. *Kemikalieinspektionen*.

Kittelsen, MO og KL Hansen. 2008. Use and emissions of hazardous substances in Norway, 2002-2006. *Rapporter*, Statistisk Sentralbyrå

Skårman, T, H Danielsson, E Henningsson, M Östman. 2006. Revised method for estimating emissions of NMVOC from solvent and other product use in Sweden. *Swedish Methodology for Environmental Data*.

Wallis, KF. 2005. Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics*, 67, Supplement 0305-9049

7. Vedlegg

Ett vedlegg: SAS-program som gjør usikkerhetsanalysen.

Vedlegg 1 – SAS-program brukt til usikkerhetsanalysen.

```
/*
Dette programmet beregner usikkerhet på utslippsdata og tar utgangspunkt i
en datafil som
inneholder kolonne for aktivitet (mengde stoff importert eller produsert)
og utslipp (mengde
stoff antatt sluppet ut). I tillegg inneholder fila kilde for
utslippsfaktoren som er angitt,
noe som blir brukt til å anslå usikkerhet på utslippsfaktoren.
*/

/*
Data importeres fra Excel. Bane og navn på datafil kan erstattes dersom
usikkerhets-
analysen skal gjentas med andre data. Bør bruke samme datafil til å
estimere utslipp og
usikkerhet på utslipp.
*/

PROC IMPORT OUT= WORK.kjem_full
            DATAFILE= "H:\Marie\usikkerhetsanalyse utslipp\Kjem_fareklasse
_til_Marie_2.xls"
            DBMS=EXCEL REPLACE;
            SHEET="Ark1$";
            GETNAMES=YES;
            MIXED=NO;
            SCANTEXT=YES;
            USEDATE=YES;
            SCANTIME=YES;
RUN;

/*
Data revideres. Mesteparten av revisjonen er gjort tidligere, men f.eks.
min_mengde er en
variabel som ikke brukes andre steder enn i usikkerhetsanalysen, så
opplagte feil rettes her.
*/

data alle_aar;
set kjem_full;

if max_mengde=0 then delete; /*Fjerner alle observasjoner som ikke bidrar
til utslipp*/
if min_mengde<0 then min_mengde=0; /*fjerner negative mengder, se notat*/
if min_mengde>max_mengde then min_mengde=max_mengde;

/*Fjerner tilfeller som er utelatt fra analysen
Tror de fleste eller alle disse tilfellene er fjernet tidligere, men
likevel...*/
if NACE>=11 then if NACE<=11.999 then delete;
if vare>='005000' then if vare<='005999' then delete; /* Usikker på om
denne virker uansett*/
if CAS_nr=127184 then if NACE>=61 then if NACE<=61.999 then if
vare='R10100' then delete;

/* BESTEMME KATEGORI - variabelen kategori som beskriver graden av
sikkerhet på utslippsfaktoren
(se notat), først og fremst på bakgrunn av kilde, inkluderes.*/
```

```

if nivaa=1 then kategori=1;
else if Kilde_for_data='SMED (2006)' then kategori=2;
else if Kilde_for_data='KemI (2005)' then kategori=3;
else if Kilde_for_data='Info fra PR' then kategori=1;
else if Kilde_for_data='Info fra DMU og PR' then kategori=1;
else if Kilde_for_data='Dokumentasjonsrapport kjemikalier 2007' then
kategori=3;
else if nivaa=10 then kategori=4;
else kategori=1; /*Data fra andre kilder får usikkerhet bestemt
individuelt*/

/* DEL C - BESTEMME USIKKERHETER */

Utslippsfaktor=utslipp/max_mengde;

/*Bestemmer usikkerhet (usp og usn) og varians (vp og vn), kategori for
kategori. Usikkerhet er 2 standardfeil*/

if kategori=4 then usn=1;
if kategori=4 then usp=0;

if kategori=3 then usp=utslippsfaktor;
if kategori=3 then usn=utslippsfaktor; /*Generelt (eg. hvis uf=0.1)
Unntakene kommer nedenfor*/
if kategori=3 then if utslippsfaktor=1 then usn=0.14;
if kategori=3 then if utslippsfaktor=1 then usp=0;
if kategori=3 then if utslippsfaktor=0.5 then usn=0.3;
if kategori=3 then if utslippsfaktor=0.5 then usp=0.3;
if kategori=3 then if utslippsfaktor<=0.0001 then usn=utslippsfaktor;
if kategori=3 then if utslippsfaktor<=0.0001 then usp=0.001;

if kategori=2 then usn=utslippsfaktor*0.2;
if kategori=2 then usp=utslippsfaktor*0.2;

/*I kategori 1 settes først noen faste regler*/
if kategori=1 then if utslippsfaktor=1 then usn=0.001;
if kategori=1 then if utslippsfaktor=1 then usp=0;
if kategori=1 then if utslippsfaktor<=0.001 then usp=0.001;
if kategori=1 then if utslippsfaktor<=0.001 then usn=utslippsfaktor;

/*Deretter bestemmes usikkerhet individuelt der vi har mer informasjon*/
if kategori=1 then if CAS_nr=15096523 then usp=0.0000674;
if kategori=1 then if CAS_nr=15096523 then usn=0.0002022; /*Antar at tallet
er avrundet ett standardavvik*/

if kategori=1 then if CAS_nr=7775099 then if NACE>=21.000 then if
NACE<=21.999 then usn=0.005;
if kategori=1 then if CAS_nr=7775099 then if NACE>=21.000 then if
NACE<=21.999 then usp=0.005; /*Antas etter nøyaktigheten dette tallet er
angitt med*/

if kategori=1 then if CAS_nr=7439921 then usn=0.0000024;
if kategori=1 then if CAS_nr=7439921 then usp=0.0000008; /*Antar at tallet
er avrundet ett standardavvik*/

if kategori=1 then if CAS_nr=1317391 then usn=0.1;

```

```

if kategori=1 then if CAS_nr=1317391 then usp=0; /*Fra-til tall angitt på
miljostatus.no*/

if kategori=1 then if kilde_for_data='Gjetning, basert på info fra Magne Mo
UMB og Otter Kjus i Addcon Nordic' then usp=0.05 ;
if kategori=1 then if kilde_for_data='Gjetning, basert på info fra Magne Mo
UMB og Otter Kjus i Addcon Nordic' then usn=0.05 ;

/* NB!!!!!!
Spesialreglene i kategori 1 er satt slik at det ikke er overlapp pr de som
i dag har kategori 1.
Dersom flere kombinasjoner får kategori 1 må ALLE spesialreglene gjennomgås
på nytt, for å være sikker på at det ikke
oppstår overlapp, f.eks. hvis samme kilde er kilde for flere
utslippsfaktorer
NB!!!!!!!!!! */

if (utslippsfaktor-usn<0) then usn=utslippsfaktor; /*Forsikrer at
konfidensintervallene ligger mellom 0 og 1*/
if (utslippsfaktor+usp>1) then usp=1-utslippsfaktor;

if max_mengde=min_mengde then A_n=max_mengde*0.15;
else A_n=(max_mengde-min_mengde);
A_p=max_mengde*0.15;

run;
/*Tidligere angitt usikkerhet er det dobbelte av hva vi vil bruke som
sigma1 og sigma2.
Her beregnes sigma1, sigma2 for både aktivitet og utslippsfaktor og
forventninger og
varianser som beskrevet i notatet.*/

data tpn;
set alle_aar;

sigf1=usn/2;
sigf2=usp/2;
sigal=A_n/2;
siga2=A_p/2;

forv_a=max_mengde+sqrt(2/(CONSTANT('pi')))*(siga2-sigal);
forv_f=utslippsfaktor+sqrt(2/(CONSTANT('pi')))*(sigf2-sigf1);

varx=(1-2/(constant('pi')))*(sigf2-sigf1)*(sigf2-sigf1)+sigf1*sigf2;
vary=(1-2/(constant('pi')))*(siga2-sigal)*(siga2-sigal)+sigal*siga2;

forventning=forv_a*sqrt(2/(CONSTANT('pi')))*(sigf2-
sigf1)+forv_f*sqrt(2/(CONSTANT('pi')))*(siga2-
sigal)+2/(CONSTANT('pi'))*(sigf2-sigf1)*(siga2-sigal);
varians=varx*vary+utslippsfaktor*utslippsfaktor*vary+max_mengde*max_mengde*
varx;

delsum_endring=utslippsfaktor*utslippsfaktor*vary;

run;

/*
Til slutt aggregeres dataene for å få usikkerhet på aggregert nivå.

```

```

For å beregne usikkerhet på andre aggregerte nivå lages andre
datasett/klasser.
*/

data K2005; /*Lager et datasett som bare inneholder observasjoner fra
2005.*/
set tpn;
if aar=2005;
run;

proc summary sum print data=K2005;
class fareklasse; /*Sorterer på fareklasse*/
var forventning varians delsum_endring utslipp; /*Summerer forventning,
variens, delsum og utslipp*/
run;
data K2006; /*Gjentar for 2006*/
set tpn;
if aar=2006;
run;

proc summary sum print data=K2006;
class fareklasse;
var forventning varians delsum_endring utslipp;
run;

data K2007; /*Gjentar for 2007*/
set tpn;
if aar=2007;
run;

proc summary sum print data=K2007;
class fareklasse;
var forventning varians delsum_endring utslipp;
run;

/*
Sum av forventning og varians brukes til å finne forventning og varians til
fordelingen av
summen innen det aktuelle år og fareklasse. Forventning er forventningen
til feilen, slik at
den beskriver avstand fra estimat til forventning.

Sum av delsum_endring brukes til å beregne varians på endring. For å finne
variens på endring
mellom to år summeres delsummen fra de respektive to årene.

Sum av utslipp brukes til å kontrollere at det estimerte utslippet stemmer
med det som er
estimert fra før (i.e. at man har brukt samme eller tilsvarende datafil til
alle analyser).
*/

```