



<b>Referencelaboratoriet for måling af emissioner til luften</b>	
<b>Titel</b>	<b>Hvordan skal forekomsten af outliers på lugtmålinger vurderes?</b>
<b>Undertitel</b>	-
<b>Forfatter(e)</b>	<b>Arne Oxbøl</b>
<b>Arbejdet udført, år</b>	<b>2015</b>
<b>Udgivelsesdato</b>	<b>10. oktober 2015</b>
<b>Revideret, dato</b>	-

## Indholdsfortegnelse

1	Baggrund .....	2
2	Vurdering .....	2
2.1	Grubbs test.....	4
2.2	Andre testmuligheder.....	6
2.3	Alternativ til outliertest.....	6

## 1 Baggrund

Flere myndigheder har tidligere indskrevet i vilkår om lugtmålinger, at spredningen mellem de fremkomne måleresultater skal være mindre end en vis størrelse. Hvis ikke den var det, skulle de højeste resultater anvendes (f.eks. to af tre resultater) uden hensyn til, om det evt. var den højeste værdi, der afveg fra de andre. Formålet har været at sikre en konservativ vurdering af lugtemissionen, så virksomheden ikke drog fordel af et evt. fejlagtigt lavt resultat.

Den anvendte model er formuleret således (bl.a. i spørgsmål til Referencelaboratoriets svartjeneste nr. 775, 2008):

"Hvis den geometriske spredning på et prøvesæt er <50% af den geometriske middelværdi, benyttes den geometriske middelværdi af enkeltprøverne ved OML-beregningerne. Hvis den geometriske spredning er >50 % af den geometriske middelværdi af enkeltprøverne, skal den geometriske middelværdi af de to højeste enkeltprøver benyttes ved OML-beregningerne."

## 2 Vurdering

I svaret til spørgsmålet i afsnit 1 anbefaler Referencelaboratoriet, at "lugtmålinger ikke behandles anderledes end andre målinger. Hvis der er meget stor forskel mellem resultater, er det for alle målinger god praksis at vurdere, om en afvigelse evt. kan forklares ved atypiske variationer i processen eller fejl i prøvetagning eller analyse".

Der er dog i dag et ønske hos Miljøstyrelsen både om at kunne identificere evt. stærkt afvigende resultater og om at erstatte 50 % reglen med en beregningsformel for outliers. Det er målet at finde en beregningsmetode, som på enkel vis kan bruges i vurderingen af, om evt. tilsyneladende afvigere kan skyldes analysevariation eller med stor sikkerhed skyldes f.eks. fejl i prøvetagningen eller utilsigtet procesvariation. Det er ikke defineret med hvor stor sikkerhed, vurderingen ønskes.

Det giver ikke mening at tale om en procentvis geometrisk spredning, og den kan ikke umiddelbart beregnes. Når resultater - som for lugt - er geometrisk normalfordelt, skal afvigelsesernes størrelse beskrives ved en faktor til hver side for middelværdien og ikke ved en procentværdi. Hvis man imidlertid tolker kravet således, at forskellen mellem den geometriske middelværdi (beregnet ved  $10^{\text{middel af logværdier}}$ ) og  $10^{(\text{middel af logværdier} + \text{geometrisk spredning})}$  ikke må overstige 50 %, giver det en vis mening.

Forskellen mellem den geometriske middelværdi (beregnet ved  $10^{\text{middel af logværdier}}$ ) og  $10^{(\text{middel af logværdier} + \text{geometrisk spredning})}$  betragtes i denne sammenhæng som et estimat af den "geometriske spredning" udtrykt i  $\text{OU}/\text{m}^3$ . Nedenstående eksempel illustrerer beregningen.

## Rapport nr. 77

### Hvordan skal forekomsten af outliers på lugtmålinger vurderes?

	Resultat	log(resultat)
	OU/m <sup>3</sup>	
Prøve 1	10.000	4,000
Prøve 2	11.000	4,041
Prøve 3	5.300	3,724
Geometrisk spredning		0,172
Geometrisk middel	8.354	3,922
Middel + spredning	12.424	4,094
Estimat af geometrisk spredning	4.070	-
Estimat i % af middel	49	-
(middel + spredning)/middel	1,5	-
Middel - spredning	5.617	3,750
Estimat af geometrisk spredning	2.737	-
Estimat i % af middel	33	-

\* spredningen på 0,172 er den maksimalt tilladelige spredning på n-butanol målinger, jf. DS/EN 13725

Problemet i at estimere en geometrisk spredning udtrykt i OU/m<sup>3</sup> ses af, at estimatet for spredningen afhænger af, om man regner på middel + spredning eller middel – spredning.

I standarden for lugtanalyser (DS/EN 13725) kræves, at et lugtlaboratorium kan præstere gentagne målinger af referencestoffet n-butanol med en geometrisk spredning på mindre end 0,172. Det betyder netop, at den geometriske gennemsnitsværdi af alle værdier plus én gange spredningen omregnet til LE/m<sup>3</sup> er 1,5 gange gennemsnitsværdien<sup>1</sup> (se eksempel ovenfor, hvor den geometriske spredning netop er 0,172 ved en afvigelse på 49%). Hvis den største værdi er mindre end gennemsnitsværdien gange 1,5 (eller den mindste værdi er større end gennemsnitsværdien divideret med 1,5) overholder målingen derfor kravene i DS/EN 13725.

Der er således en god regnemæssig overensstemmelse mellem kravet og standarden, hvis standardens 0,172 er et udtryk for en accepteret usikkerhed på lugtanalyser. Spredningen på <0,172 dokumenteres ved analyser af prøver af n-butanol. Der er imidlertid intet belæg for, at spredningen på n-butanolanalyser giver et retvisende billede af spredningen på almindelige prøver. Lugten af n-butanol bliver over tid kendt for panelisterne, hvilket måske fører til en større ensartethed i deres svar og dermed en mindre spredning. Erfaringsmæssigt har de danske laboratorier en spredning, der er væsentligt lavere end 0,172. Der er således intet grundlag for at betragte det enkelte laboratoriums spredning på n-butanolprøver som retvisende for dets arbejde med andre prøver. Det vurderes derfor, at standardens 0,172 indtil videre er det bedste bud på en beskrivelse af usikkerheden i analysen.

Det skitserede krav vurderes at være et "skrap" krav, idet 95% af en række gentagelser af lugtanalyse af den samme prøve vil ligge i et interval, der går fra gennemsnitsværdien - 1,96 \* spredningen til gennemsnitsværdien + 1,96 \* spredningen. Med spredningen fra standarden (0,172) svarer det til, at intervallet målt i OU/m<sup>3</sup> går fra gennemsnitsværdien/2,17 til gennemsnitsværdien \* 2,17<sup>2</sup>. Forholdet mellem højeste og laveste intervalgrænse er således 2,17 \* 2,17 = 4,6. Hvis man begrænser sig til én gange spredningen (eller en faktor 1,5), ligger kun 67% af resultaterne i intervallet.

<sup>1</sup>  $10^{0,172} = 1,49$

<sup>2</sup>  $10^{(1,96*0,172)} = 2,17$

## Rapport nr. 77

### Hvordan skal forekomsten af outliers på lugtmålinger vurderes?

---

På grund af den ikke korrekte håndtering af procentvise afvigelser for den logaritmiske normalfordeling af lugtresultater ønskes en mere anerkendt testmetode.

#### 2.1 Grubbs test

Grubbs test er en velkendt test for outliers. Grubb's teststørrelse for en suspekt værdi beregnes af

$$G_p = \frac{|X_{\text{suspekt}} - X_{\text{middel}}|}{s}$$

hvor  $X_{\text{suspekt}}$  er den formodede outlier  
 $X_{\text{middel}}$  er middelværdien af samtlige resultater  
 $s$  er spredningen

Grubbs test er beskrevet i ISO 5725-2<sup>3</sup>, og i denne publikations tabel 5 findes de kritiske værdier for hhv. 5% konfidensniveau og 1% konfidensniveau.

Hvis en teststørrelse er mindre end eller lig med 5% værdien, kan resultatet accepteres. Hvis teststørrelsen er større end 5% værdien og mindre end eller lig med 1% værdien, kaldes resultatet en "straggler", som bør underkastes nærmere undersøgelse. Hvis teststørrelsen er større end 1% værdien, er resultatet en outlier og bør forkastes.

Testen anvender den spredning, der kan beregnes mellem de aktuelle resultater. Spredningen skyldes bl.a. variationer i både materialet/produktionen, prøveudtagningen og analysen. Det betyder, at tilstedeværelsen af en outlier øger spredningen. Da outlieren således påvirker både tæller og nævner i ovenstående formel, bliver  $G_p$  aldrig særlig stor, og konsekvensen er, at en outlier skal ligge meget langt fra de øvrige værdier for at blive betragtet som outlier. Testen kan karakteriseres som svag – især for få prøver. Konkret gælder det for tre prøver, at det tilsyneladende ikke kan lade sig gøre at opnå en testværdi højere end den kritiske værdi, hvilket illustreres nedenfor for både en lav og en høj afvigende værdi. Der er foretaget flere beregningseksempler uden at opnå en testværdi højere end den kritiske værdi.

---

<sup>3</sup> Accuracy (trueness and precision) of measurement methods and results – Basic methods for the determination of repeatability and reproducibility of a standard measurement method

## Rapport nr. 77

### Hvordan skal forekomsten af outliers på lugtmålinger vurderes?

Måling	Resultat	Log(resultat)		Måling	Resultat	Log(resultat)
1	10.000	4,000		1	10.000	4,000
2	11.000	4,041		2	11.000	4,041
3	100	2,000		3	200.000	5,301
4				4		
Antal		3		Antal		3
Spredning, s		1,167		Spredning, s		0,739
Middel	2.224	3,347		Middel	28.020	4,447
	Høj værdi	Lav værdi			Høj værdi	Lav værdi
Beregnet teststørrelse	0,595	1,1545		Beregnet teststørrelse	1,1542	0,605
Straggler				Straggler		
Outlier				Outlier		
5% kritisk værdi	1,155			5% kritisk værdi	1,155	
1% kritisk værdi	1,155			1% kritisk værdi	1,155	

Teststørrelse nærmer sig, men når aldrig, den kritiske værdi, når det laveste resultat bliver lavere hhv. når det høje resultat bliver højere.

Når det gælder en myndigheds ønske om at teste tre lugtresultater, er det oftest en præmis for undersøgelsen, at prøverne er udtaget under konstant produktion, dvs. lugten bør ikke ændres som følge af ændret produktion. Myndigheden ønsker derfor at sikre sig med en vis sandsynlighed, at prøverne er taget under de ønskede, konstante forhold og, at evt. variationer kan tilskrives variationen i analysen. Det kunne derfor foreslås at benytte Grubbs test modificeret således, at standardens krævede, maksimale spredning på 0,172 anvendes i beregningen af teststørrelsen. Beregninger med denne spredning illustreres nedenfor.

Måling	Resultat	Log(resultat)		Måling	Resultat	Log(resultat)
1	10.000	4,000		1	10.000	4,000
2	11.000	4,041		2	11.000	4,041
3	5.200	3,716		3	20.900	4,320
4				4		
Antal		3		Antal		3
Spredning, s		0,172		Spredning, s		0,172
Middel	8.301	3,919		Middel	13.198	4,121
	Høj værdi	Lav værdi			Høj værdi	Lav værdi
Beregnet teststørrelse	0,71	1,181		Beregnet teststørrelse	1,16	0,701
Straggler				Straggler		
Outlier		Ja		Outlier	Ja	
5% kritisk værdi	1,155			5% kritisk værdi	1,155	
1% kritisk værdi	1,155			1% kritisk værdi	1,155	

Det ses af eksemplet, at en værdi, der er ca. halvt så stor som den højeste værdi, bliver dømt som outlier. Set i lyset af, at 95% af en serie analyseresultater af den samme prøve vil ligge i et interval fra gennemsnittet/2,17 til gennemsnittet \* 2,17 (se tidligere i teksten) giver det ingen mening at forkaste 5.200 som outlier i det viste eksempel.

## Rapport nr. 77

### Hvordan skal forekomsten af outliers på lugtmålinger vurderes?

---

Grubbs test forekommer derfor ikke anvendelig, når det gælder få prøver (typisk tre ved lugtprøvninger), hverken i sin oprindelige form eller modificeret med værdien for spredning fra EN 13.725.

## 2.2 Andre testmuligheder

Nair's test har tidligere været diskuteret og illustreret. Den er i sin grundform identisk med Grubbs test, men bruger nogle andre kritiske værdier. Således er den kritiske 5% værdi for tre prøver lig med 1,74. Det betyder, at hvis en prøve afviger mere end 1,74 gange spredningen fra middelværdien dømmes den som outlier. En prøve skal derfor i denne test afvige mere fra middelværdien for at blive dømt som outlier end i Grubbs test. Testen er mindre kendt og vil ikke blive diskuteret yderligere.

Der findes en relativt simpel test for outliers, der er kendt som Dixon's test eller Q-test. Ved test af, om den mindste værdi i en serie er en outlier, beregnes teststørrelsen,  $Q$ , som

$$Q = \frac{X_{næstmindst} - X_{mindst}}{X_{størst} - X_{mindst}} \text{ (og tilsvarende for den højeste værdi).}$$

Teststørrelsen beregnes således, hvor stor forskellen mellem den lille, formodede outlier og den nærmest liggende værdi er i forhold til hele måleseriens interval. Denne test lider af den samme svaghed som Grubbs test, at både tæller og nævner i formlen påvirkes i samme retning, når den mindste værdi bliver mindre. Stor spredning mellem værdierne gør det vanskeligt at udpege en outlier. Særligt i forhold til ønsket om at teste, om en tilsyneladende afvigende værdi skyldes noget andet end analyseusikkerhed, er testens anvendelighed tvivlsom.

## 2.3 Alternativ til outliertest

Referencelaboratoriet vurderer, at vi ikke kan finde en tilstrækkeligt god outliertest, men må ty til en anden måde at vurdere en måleserie på. Som tidligere nævnt ligger 95% af resultaterne af gentagne målinger af den samme prøve inden for et interval, hvor øvre intervalgrænse/nedre intervalgrænse = 4,6 under forudsætning af, at spredningen er mindre end eller lig med 0,172. Hvis forholdet mellem højeste og laveste værdi i en måleserie er større end 4,6, er der stor sandsynlighed for, at resultaterne ikke tilhører samme (analyse)fordeling (ved antaget konstant produktion og "ens" prøver). Risikoen for fejlagtigt at drage denne konklusion er 5%.

I Holland beskriver den hollandske lugtvejledning, NTA9065<sup>4</sup>, at en måleserie kan betragtes som valid, hvis forholdet mellem højeste og laveste resultat i serien er mindre end 5. Det er ikke bekræftet, om denne faktor er et resultat af samme overvejelser som ovenfor, der fører til en faktor 4,55. Det er imidlertid sandsynligt.

Når en måleserie skal vurderes, er der to situationer:

Hensigten med målingerne er at dokumentere emissionen i en konstant driftssituation: Hvis forholdet mellem højeste og laveste værdi er større end 4,55, kan myndigheden med rimelig sikkerhed sige, at enten er der lavet en prøvetagningsfejl, eller produktion har ændret sig væsentligt. Det vil herefter være et skøn, om målingerne skal tages om, eller om én af værdierne skal bortkastes. Det gælder dog altid, at laboratoriet

---

<sup>4</sup> NTA9065:2012, Air quality - Odour measurements – Odour measurement and calculation, december 2013

**Rapport nr. 77****Hvordan skal forekomsten af outliers på lugtmålinger vurderes?**

---

skal vurdere den afvigende prøve for evt. at finde de omstændigheder, der forårsager afvigelsen (fortyndingsfejl, defekt prøveemballage, regnefejl, evt. oplysninger om produktionen).

Hvis hensigten med målingerne er at dokumentere en gennemsnitlig emission i en periode med varierende produktion, gælder betragtningerne om evt. afvigende værdier ikke.