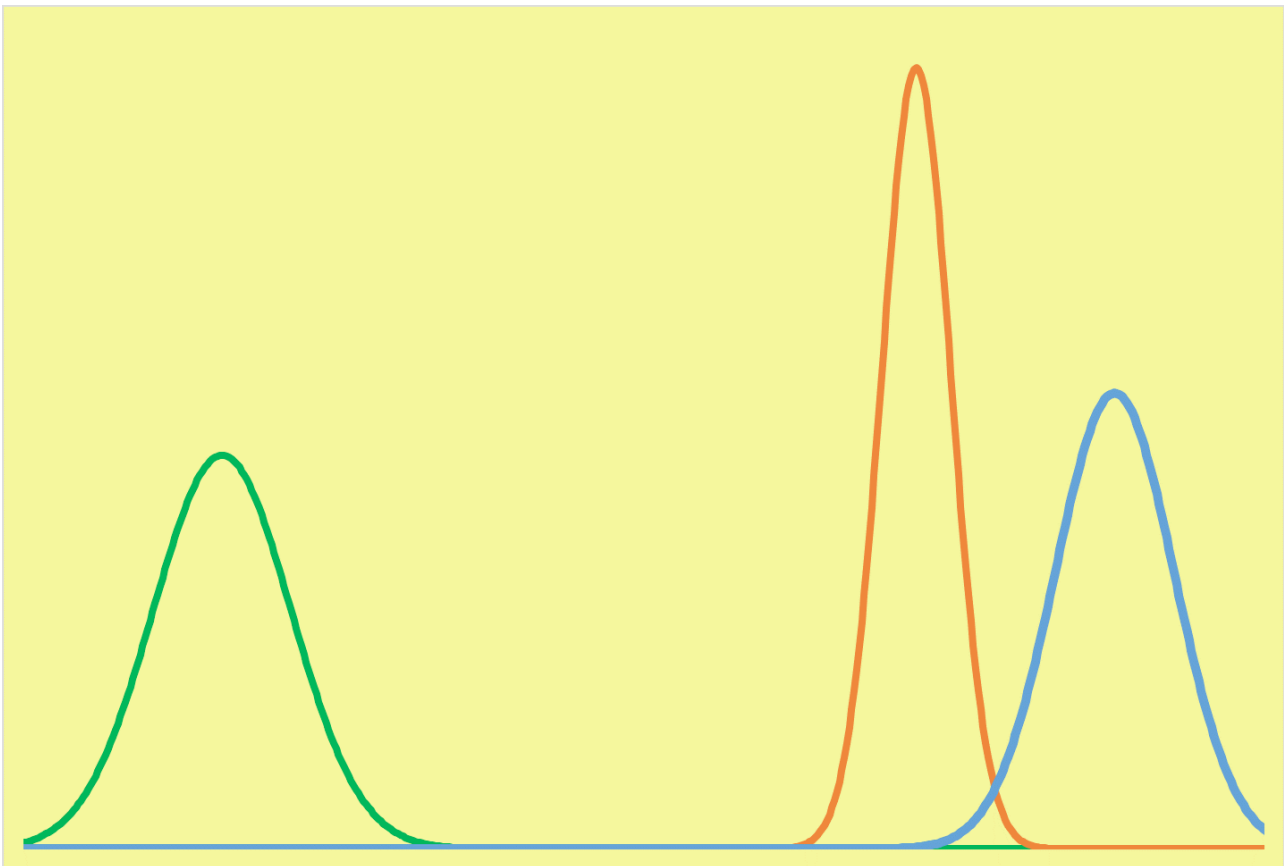


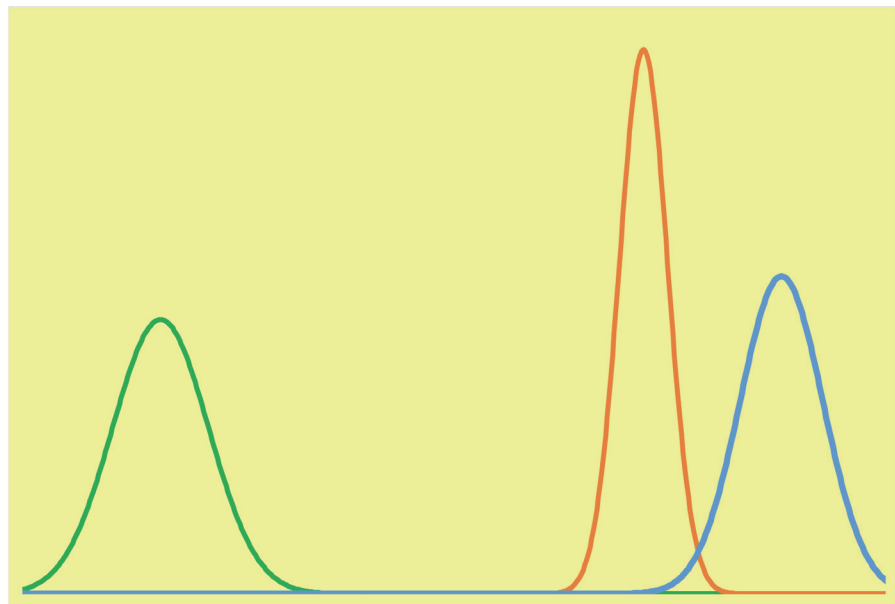
Sammanställning av resultat från deltagande i internationella kompetensprövningar

H. RAMEBÄCK, S. HOLMGREN RONDAHL, S. JONSSON,
P. LAGERKVIST, A. TOVEDAL, A. VESTERLUND



H. Ramebäck, S. Holmgren Rondahl, S. Jonsson,
P. Lagerkvist, A. Tovedal, A. Vesterlund

Sammanställning av resultat från deltagande i internationella kompetensprövningar



Titel	Sammanställning av resultat från deltagande i internationella kompetensprövningar
Title	Summary of results from participation in international proficiency tests
Rapportnr/Report no	FOI-R--4868--SE
Månad/Month	Juli
Utgivningsår/Year	2020
Antal sidor/Pages	14
ISSN	1650-1942
Kund/Customer	Fö
Forskningsområde	CBRN-frågor
FoT-område	Inget FoT-område
Projektnr/Project no	A405719
Godkänd av/Approved by	Niklas Brännström
Ansvarig avdelning	CBRN-skydd och säkerhet

Bild/Cover: Henrik Ramebäck, FOI

Detta verk är skyddat enligt lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk, vilket bl.a. innebär att citering är tillåten i enlighet med vad som anges i 22 § i nämnd lag. För att använda verket på ett sätt som inte medges direkt av svensk lag krävs särskild överenskommelse.

This work is protected by the Swedish Act on Copyright in Literary and Artistic Works (1960:729). Citation is permitted in accordance with article 22 in said act. Any form of use that goes beyond what is permitted by Swedish copyright law, requires the written permission of FOI.

Sammanfattning

I den här rapporten har resultat från deltagande i nio internationella kompetensprövningar anordnade av IAEA för mätning av radionuklider sammanställts. Vidare redovisar rapporten hur utvärdering av resultat kan göras, och hur den görs av IAEA.

Nyckelord: Kompetensprövning, mätkvalitet, radionuklider

Summary

This report summarizes results from nine international proficiency tests arranged by IAEA. Moreover, the report also shows how reported results can be evaluated and how evaluation is done by IAEA.

Keywords: proficiency tests, quality i measurements. radionuclides

Innehållsförteckning

1	Inledning	6
2	Teori	7
	2.1 IAEA:s utvärderingskriterier	8
3	Resultat och Diskussion	10
	3.1 Resultat utifrån IAEA:s bedömningskriterier	10
	3.2 Resultat utifrån zeta-scores	12
4	Slutsatser	13
5	Referenser	14

1 Inledning

Deltagande i kompetensprövningar (*proficiency tests*) är en central del för att laboratorium ska kunna påvisa kompetens inom en mätmetod. Det är dessutom ett krav för laboratorier som är, eller ämnar bli, ackrediterade för en specifik mätmetod [1]. I vissa sammanhang förekommer det s.k. jämförelsemätningar (*interlaboratory comparisons*), vilka inte ska förväxlas med kompetensprövningar. I en jämförelsemätning tilldelas den storhet som ska mätas sitt värde i efterhand, ofta som ett konsensusvärde som bygger på de deltagande laboratoriernas rapporterade resultat. Jämförelsemätningar mäter då överensstämmelse *mellan* laboratorier, men inte noggrannheten i enskilda rapporterade mätresultat eftersom systematiska effekter i rapporterade resultat kan inverka på konsensusvärdet [2]. I kompetensprövningar å andra sidan har man på förhand tilldelat värden på de storheter som ska mätas. Ett sätt att göra detta är via tillsats av en känd mängd till en matris, eller genom användande av certifierade referensmaterial [3].¹

Deltagande i kompetensprövningar är inte enbart centralt för att påvisa kompetens. En annan aspekt är att man genom avvikande resultat kan identifiera problem i en mätmetod och, efter identifiering av orsaken till avvikelsen, vidta korrigerande åtgärder. Detta är då en del i ett kontinuerligt förbättringsarbete. Det är viktigt att inte se avvikande resultat som rena misslyckanden, utan en möjlighet till lärdom, under förutsättning att orsaken identifieras och åtgärder införs i metodbeskrivningar, och som en del i ett ständigt förbättringsarbete.

I den här rapporten har resultat från FOI:s deltagande i nio internationella kompetensprövningar anordnade av IAEA för mätning av radionuklider sammanställts. I rapporten redovisas även hur utvärdering av rapporterade mätresultat genomförts. Dessutom nämns några specifika problem som förekommit och som efter identifiering av orsak åtgärdats.

¹ I jämförelsemätningar kan också certifierade referensmaterial givetvis användas.

2 Teori

Nedan redovisas några vanliga utvärderingskriterier [4]. Ett mått på avvikelse, eller mätfel är

$$D_i = x_i - x_{pt} \quad (1)$$

där x_i är det rapporterade värdet och x_{pt} det tilldelade värdet på mätstorheten. Ekvation (1) kan också uttryckas i relativa mått

$$D_i\% = 100(x_i - x_{pt})/x_{pt} \quad (2)$$

Det är ett enkelt utvärderingsmått, men brister främst i att det saknar koppling till vad som kan anses vara *fit-for-purpose* för en mätmetod². En relativt stor avvikelse kan erhållas om analytkoncentrationen är låg, men det kan fortfarande vara ett användbart mätresultat. Det kopplar inte heller till mätosäkerheten. Ett mått som adresserar mätosäkerhet är zeta-score

$$\zeta = \frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}} \quad (3)$$

där $u^2(x_i)$ är mätosäkerheten i det rapporterade värdet och $u^2(x_{pt})$ är mätosäkerheten i det tilldelade värdet, båda med $k=1$. Utvärderingskriterier brukar vara:

- Överensstämmelse gäller om $|\zeta| \leq 2$.
- Avvikelse om $|\zeta| > 3$. Ska resultera i att man utreder orsak till avvikelsen och inför en korrigerande åtgärd.
- Värden mellan 2 och 3 anses eventuellt kunna ifrågasättas, men kan också bero på slumpmässiga fel. Då ett flertal värden, speciellt i rad, mellan 2 och 3 förekommit ska orsaken till de upprepade avvikelserna utredas.

Ett problem här är att en deltagare i en kompetensprövning kan manipulera sitt zeta-score genom att medvetet överskatta mätosäkerheten $u(x_i)$, och på så sätt sänka $|\zeta|$ [5]. Zeta-scores >2 eller <-2 orsakas troligtvis antingen av signifikanta systematiska fel eller av underskattad mätosäkerhet, eller en kombination av båda. Ett mått som adresserar *fitness-for-purpose* är z-score

$$Z = \frac{x_i - x_{pt}}{\sigma_{pt}} \quad (4)$$

där σ_{pt} är standardavvikelsen för kompetensprövningen. Utvärderingskriterier är samma som för zeta-score. Nackdelen här är återigen att måttet inte tar hänsyn till rapporterad mätosäkerhet. Å andra sidan kan måttet inte manipuleras genom att man medvetet överskattar mätosäkerheten. För en kompetensprövning bestäms σ_{pt} ofta utifrån ett kriterium baserat på *fitness-for-purpose*. Kombinerar z-score och zeta-score kan man dra följande slutsatser:

² *Fit-for-purpose* är ett koncept som anger om en metod är ändamålsenlig, d.v.s. om man kan använda ett mätresultat för det syfte man hade tänkt.

- Ett resultat som passerar zeta- men inte z-testet har en stor avvikelse och en rapporterad mätosäkerhet som är för stor.
- Ett resultat som passerar z- men inte zeta-testet har en liten avvikelse och en rapporterad mätosäkerhet som är för liten.

Alltså kommer en kombination av båda testen att adressera både avvikelser och rapporterad mätosäkerhet. I kompetensprövningar arrangerade av IAEA utvärderas båda dessa, även om utvärderingskriterierna baseras på andra mått vilka redovisas nedan.

2.1 IAEA:s utvärderingskriterier

IAEA utvärderade rapporterade mätresultat fram t.o.m. 2012 enligt följande kriterier:

Noggrannheten i mätresultatet får "A"³ om⁴

$$A1 \leq A2 \quad (5)$$

där

$$A1 = |x_{pt} - x_i| \quad (6)$$

$$A2 = 2.58 \cdot \sqrt{u_{pt}^2 + u_i^2} \quad (7)$$

Precisionen utvärderas med

$$P = 100 \cdot \sqrt{\left(\frac{u_{pt}}{x_{pt}}\right)^2 + \left(\frac{u_i}{x_i}\right)^2} [\%] \quad (7)$$

Värdet på P är beroende av laboratoriets rapporterade mätosäkerhet och jämförs med en förutbestämd gräns, LAP , (*acceptance limit for precision*) som kan variera beroende på analyt och analytkoncentration. Precisionen får "A" om $P \leq LAP$. I den slutgiltiga bedömningen erhåller det deltagande laboratoriet ett "A" om man fått "A" både för noggrannhet och precision. Får man "N" för båda kriterierna blir det slutgiltiga omdömet "N". Om ett av kriterierna erhåller "N" men det andra "A" görs en fortsatt analys. Då jämför man laboratoriets relativa avvikelse, Ekvation (2), med en maximalt tillåten avvikelse, MAB (*maximum acceptable bias*), som likt LAP är satt på förhand. Om $D_i \% \leq MAB$ erhåller laboratoriet "W". Om $D_i \% > MAB$ erhåller laboratoriet dock "N" som slutlig bedömning. Här speglar ett "W" två möjliga situationer: a) ett avvikande resultat med liten mätosäkerhet, dock inom MAB , b) ett mätresultat med liten avvikelse men med stor mätosäkerhet.

³ IAEA ger "A" för överensstämmelse (*Accepted*), "N" för avvikelse (*Not accepted*) och "W" för ett tveksamt resultat (*Warning*). För riktighet och precision ges antingen "A" eller "N", medan i den totala utvärderingen kan även "W" ges, se texten.

⁴ Här har samma nomenklatur som ovan använts. I det här fallet avser index pt IAEA, som är anordnare av kompetensprövningen.

Fr.o.m. 2012 införde IAEA nya utvärderingskriterier. För ett mätresultats noggrannhet erhålls "A" om

$$|D_i \%| \leq MAB \quad (8)$$

Vidare ska den utvidgade relativa sammanlagda osäkerheten överlappa den relativa avvikelserna $D_i\%$

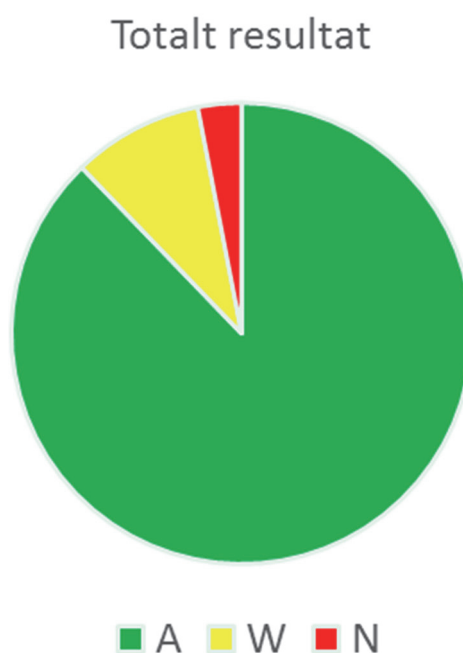
$$|D_i \%| \leq k \cdot P \quad (9)$$

där P ges av ekvation (7), och k är en täckningsfaktor som för 99 % konfidensintervall är 2,58. Om $P \leq MAB$ och $|D_i \%| \leq k \cdot P$ får det rapporterade värdet ett "A" för precision, om ett av villkoren inte uppfylls blir det "N" för precision. I den slutgiltiga bedömningen erhålls ett "A" om både ekvation (8) och (9) är uppfyllda, ett "N" om ekvation (9) är uppfylld men inte ekvation (8). Om ekvation (8) är uppfylld men inte ekvation (9) erhålls ett "W" som slutlig bedömning. Även denna utvärdering ger information om den rapporterade mätosäkerheten.

3 Resultat och Diskussion

3.1 Resultat utifrån IAEA:s bedömningskriterier

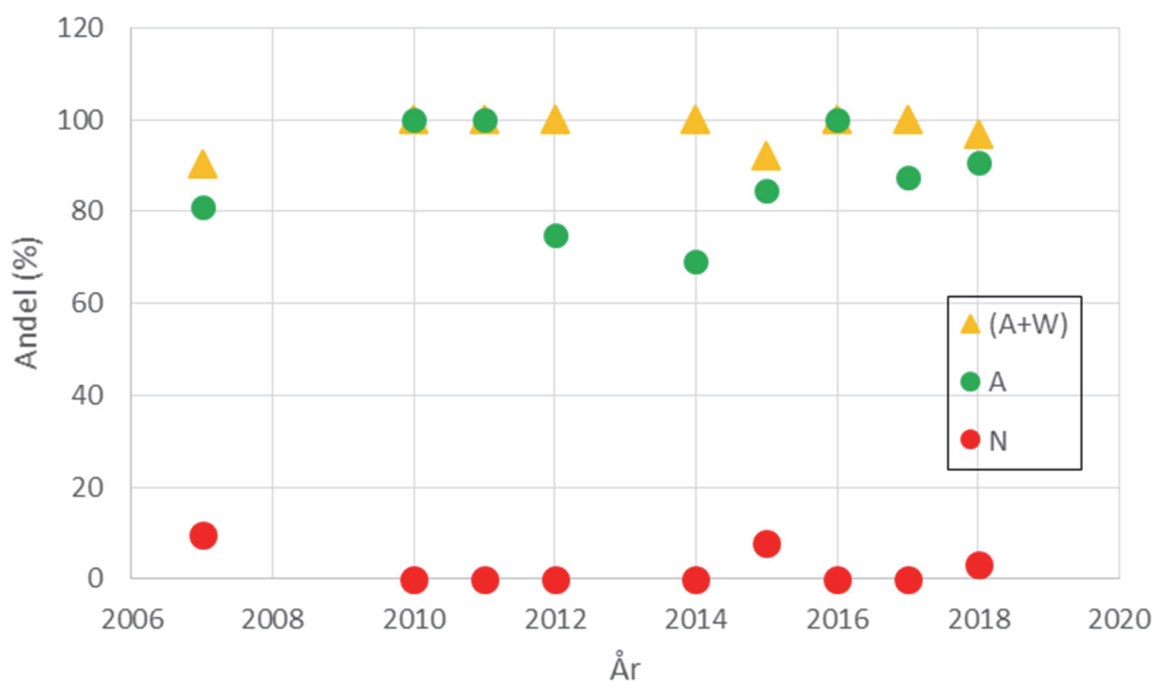
Analysen i denna rapport avser alla rapporterade värden oavsett analyt och mätmetod. Sammanställningen gäller för de år som FOI deltagit i IAEA:s kompetensprövningar (2007, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017 och 2018). I figur 1 visas den slutliga bedömningen över alla år 2007-2018 (för alla analyter och alla mätmetoder).



Figur 1. Slutlig bedömning för alla analyter och alla mätmetoder.

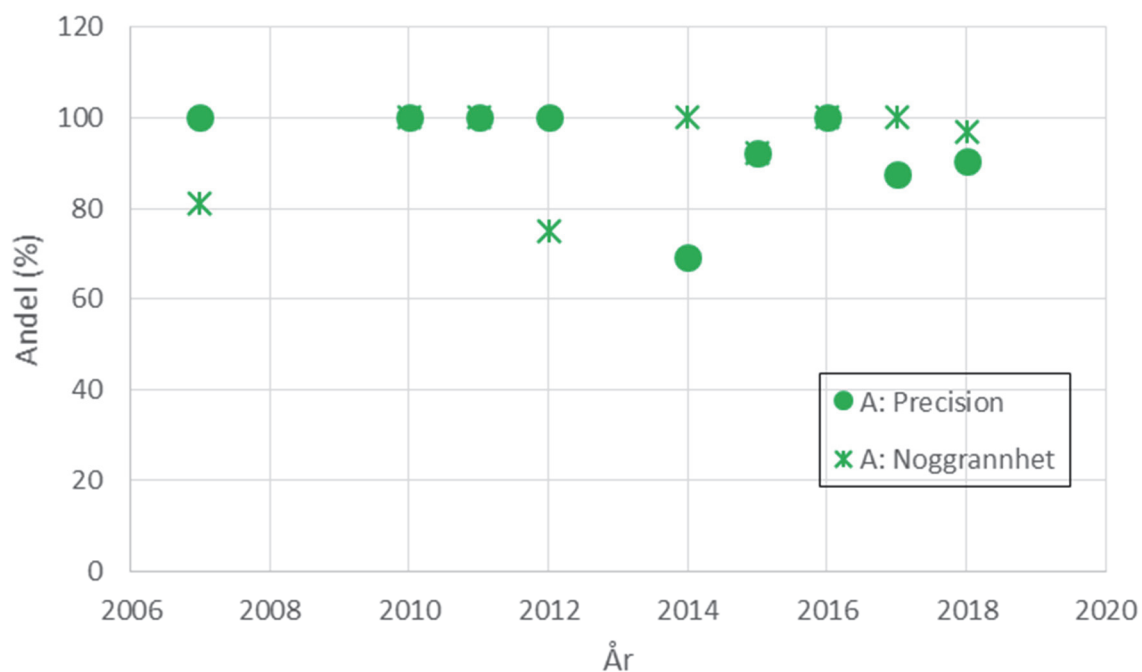
Över alla år har 88 % av alla rapporterade värden fått slutgiltigt ”A”, 9 % ”W” och 3 % ”N”⁵. Bilden över tid visas i figur 2. Vissa år har enbart ett fåtal analyter rapporterats, varför ett avvikande resultat ger en stor andel ”W” eller ”N”.

⁵ Av totalt 131 rapporterade mätresultat var 115 ”A”, 12 ”W” och 4 ”N”.



Figur 2. Andel "A", "A+W" samt "N" över tid sedan 2007. Data avser alla analyser och alla mätmetoder.

I figur 3 visas andelen "A" rörande precision och noggrannhet över alla år och för alla analyser/mätmetoder.



Figur 3. Andel "A", för precision och noggrannhet över tid sedan 2007. Data avser alla analyser och alla mätmetoder.

Det är svårt att uttyda några trender över tid, men man kan konstatera att resultaten får anses som goda (97 % A+W). Bilden som ges i

sammanfattningar av IAEA är att generellt erhåller ca. 60-70 % av alla rapporterade resultat ett slutgiltigt "A".

I kompetensprövningen 2007 erhöles nästan 20 % "N" för noggrannhet (4 av 21). I tre av fallen berodde detta på att FOI Umeå vid den tiden inte hade metodik för korrektion av koincidenssummersförluster i gammaspektrometri, bortsett från ^{134}Cs som korrigerades med en empirisk korrektionsfaktor⁶. Korrektion för koincidenssummersförluster började implementeras 2010-2011. Det fjärde fallet rörde en interferens av bly-röntgen vid mätning av ^{109}Cd (ett prov med ^{109}Cd fick "A" men det var ett prov med lägre totalaktivitet än det som fick "N"). I det senare fallet var åtgärden att införa en kopparplåt i blyskyddet, vilket reducerar inducerad bly-röntgen. I kompetensprövningen 2012 rapporterades enbart fyra analyter, varav en fick "N" för noggrannhet, vilket då slår mycket på andelen. I kompetensprövningen 2014 fick knappt 70 % av resultaten "A" för precision. Dock fick 100 % av rapporterade resultat "A" för noggrannhet, vilket ledde till 31 % "W" i det slutliga omdömet. Vid ett par tillfällen har problematik vid utvärdering av låga aktiviteter av gammastrålande radionuklider identifierats. Grundorsaken har då varit hur programvaran för utvärdering anpassar bakgrunden för en topp. Genom att granska denna anpassning kan detta problem undvikas genom att genomföra anpassningen med andra parametrar. Ett annat alternativ är att öka mätosäkerheten.

3.2 Resultat utifrån zeta-scores

Ett annat sätt att utvärdera resultat från kompetensprövningar är utifrån zeta-score (ζ) [4], se ovan. Den analyt med flest rapporterade värden är ^{137}Cs (20 stycken). Av dessa 20 rapporterade resultat var två stycken utanför gränserna ± 2 och inget utanför ± 3 . Näst flest rapporterade resultat är för ^{134}Cs (14 stycken). Även här var två resultat utanför ± 2 . Totalt sex stycken rapporterade resultat av 131 har varit utanför ± 3 : ^{210}Po , ^{235}U , ^{234}U , ^{40}K , ^{226}Ra och ^{54}Mn (en gång per analyt).⁷ Orsaken till avvikande resultat för ^{40}K och ^{54}Mn , vilka båda härrör sig från 2007, är att kalibreringen för HPGe-detektorn då var behäftad med systematiska fel i energiregionen högre än ~ 700 keV, se diskussionen ovan samt fotnot 6. För ^{210}Po , zeta-score=-36, berodde avvikelsen på felaktig korrektion till referenstidpunkt. Halveringstidskorrektion genomfördes med halveringstiden för ^{210}Po i stället för halveringstiden för ^{210}Pb . För alla rapporterade analyter var 82 % innanför ± 2 och 95 % innanför ± 3 .

⁶ I själva verket berodde avvikelserna på att kalibreringen var behäftad med systematiska fel p.g.a. närvaro av radionuklider i kalibreringsmaterialet som sönderföll via kaskadsönderfall (^{60}Co och ^{88}Y). Detta resulterade i att ^{40}K och ^{54}Mn , vilka själva inte ger upphov till koincidenssummersförluster, fick avvikande resultat.

⁷ För de här analyterna fick ^{210}Po och ^{54}Mn "N" utifrån IAEA:s bedömningskriterier, de andra "W".

4 Slutsatser

Deltagande i kompetensprövningar är en central del för ett laboratorium att påvisa kompetens inom en mätmetod. Vidare är utvärdering av resultat, främst gällande avvikelser från deltagande, en viktig del i laboratoriets förbättringsarbete. Under åren har några avvikelser resulterat i identifiering av möjliga problem, vilka åtgärdats. Resultaten från deltagande i dessa kompetensprövningar anses som goda. Totalt sett har, utifrån IAEA:s bedömningskriterier, 88 % fått "A" i slutligt omdöme, och 97 % ("A"+"W").

5 Referenser

- 1 SS-EN ISO/IEC 17025:2018: *Allmänna kompetenskrav för provnings och kalibreringslaboratorier.*
- 2 A. Shakhashiro, A. Fajgelj, U. Sansone: *Comparison of different approaches to evaluate proficiency test data*, International Workshop Combining and Reporting Analytical Results: The Role of Traceability and Uncertainty for Comparing Analytical Results, Volume: The Royal Society of Chemistry, Special Publication No. 307, Cambridge, pp. 220-228, 2007.
- 3 Eurachem: *Selection, use and interpretation of proficiency testing (PT) schemes*, 2nd Edition, 2011.
- 4 ISO 13528:2015: *Statistical methods for use in proficiency testing by interlaboratory comparisons.*
- 5 Analytical methods committee, AMCTB No. 74: *z-scores and other scores in chemical proficiency testing--their meanings, and some common misconceptions*, Analytical Methods, 8, 5553, 2016.

FOI är en huvudsakligen uppdragsfinansierad myndighet under Försvarsdepartementet. Kärnverksamheten är forskning, metod- och teknikutveckling till nytta för försvar och säkerhet. Organisationen har cirka 1000 anställda varav ungefär 800 är forskare. Detta gör organisationen till Sveriges största forskningsinstitut. FOI ger kunderna tillgång till ledande expertis inom ett stort antal tillämpningsområden såsom säkerhetspolitiska studier och analyser inom försvar och säkerhet, bedömning av olika typer av hot, system för ledning och hantering av kriser, skydd mot och hantering av farliga ämnen, IT-säkerhet och nya sensorers möjligheter.



FOI
Totalförsvarets forskningsinstitut
164 90 Stockholm

Tel: 08-55 50 30 00
Fax: 08-55 50 31 00

www.foi.se