

**Normen en richtlijnen voor
methodvalidatie in chemische laboratoria**

Wetenschappelijk ondersteuningsprogramma
voor de normalisatie

deel II

Eindverslag

Federale Diensten voor
WETENSCHAPPELIJKE, TECHNISCHE
EN CULTURELE AANGELEGENHEDEN

**Wetenschappelijk
ondersteuningsprogramma voor de
normalisatie**

**Onderzoeksovereenkomst Nr. NO/03/003
Normen en richtlijnen voor
methodvalidatie in chemische
laboratoria**

Eindverslag

Deelnemende laboratoria :

- **VUB : prof. D.L. Massart**
- **ULg : prof. J. Crommen**
- **KUL : prof. J. Hoogmartens en prof. E. Roets**

Wetenschappelijk ondersteuningsprogramma voor de normalisatie

Onderzoeksovereenkomst Nr. NO/03/003

Normen en richtlijnen voor methodenvalidatie in chemische laboratoria

Eindverslag

1. Inleiding

Overeenstemming tussen meetresultaten en hun onderlinge erkenning voor het nagaan van het naleven van de productspecificaties is een probleem dat belangrijk is voor iedereen die betrokken is in een consument-producent relatie. Dit is zowel een probleem voor de industrie als voor nationale en internationale beleidsorganen. Een dergelijke overeenkomst vereist dat de meetmethoden gevalideerd zijn, dit wil zeggen dat aangetoond moet worden dat hun precisie en andere performantiekarakteristieken voldoen aan de eisen van alle belanghebbende partijen. Er is een nood aan afspraken over de wijze waarop deze validatie moet uitgevoerd worden, of met andere woorden : men heeft standaarden en richtlijnen nodig voor de validatie van meetmethoden.

De technische en wetenschappelijke doelstelling van het project was om gestandaardiseerde benaderingen en richtlijnen te ontwikkelen voor de validatie van meetmethoden in chemische, farmaceutische, agro-voeding en aanverwante industrieën. Het is duidelijk dat standaardisatie niet enkel nodig is op nationaal vlak maar nog meer op internationaal vlak en dat de doelstelling daarom het definiëren van standaarden voor Europees gebruik of zelfs voor een wereldwijd gebruik is. Op organisatorisch vlak bracht dit project de bekwaamheid van een aantal Belgische onderzoekers en een aantal industrieën op het gebied van methodenvalidatie samen, teneinde de Belgische wetenschappers een prominente rol te laten spelen in de ontwikkeling van normen en richtlijnen.

Methodenvalidatie is dus een belangrijk regulatory probleem voor vele industrieën en is nodig om de validiteit aan te tonen van de methoden die aangewend worden om product specificaties na te gaan. De voorgestelde normen voor methodenvalidatie moeten kost-efficiënt zijn zeker wat de industrie betreft.

Bovendien is methodenvalidatie nodig voor de bescherming van de veiligheid van de consument en om de kwaliteit van consumptieproducten te garanderen. De Farmacopee, bijvoorbeeld, beschrijft standaardmethoden om de kwaliteits- en veiligheidskarakteristieken van geneesmiddelen te controleren. Er is echter nood aan een betere validatie van deze analysemethoden. Daarom werden in het kader van dit project een aantal case studies uitgevoerd die zich concentreerden op de veiligheid van geneesmiddelen. Hierbij handelt het

enerzijds over methodenvalidatie van farmakopee-analyses op bulkpreparaten en anderzijds van bioanalysemethodes van geneesmiddelen voor farmacokinetsiche doeleinden.

De ontwikkeling van snelle, multivariate technieken, zoals o.a. de nabij-infrarood analyse, voor het meten van de kwaliteit tijdens of in onmiddellijke relatie met een proces, is een belangrijke trend in de industrie. Pre-normatief onderzoek om guidelines te ontwikkelen voor de procesanalyse was gepland en werd uiteindelijk uitgevoerd behandeld.

Het hier uitgevoerde project sloot nauw aan bij een internationaal project genaamd "*Robust process analytical methods for industrial practice and their validation*" uit het Standards, Measurements and Testing programma van de Europese Unie. Dit project had als partners de Universiteiten van Brussel, Tarragona (Spanje) en Genua (Italië) en de volgende bedrijven Elf (Frankrijk), Rhone Poulenc (Frankrijk), Pfizer (Verenigd Koninkrijk), Novartis (Zwitserland) en Unilever (Nederland). Het was een sterk industrie-gericht project met de volgende doelstellingen :

- a) het voorstellen van werkmiddelen voor de methodenvalidatie in een industriële context, welke dan uitgetest werden door bovenvermelde belangrijke industriële partners, zodat bijgedragen wordt tot de Europese standardisatie in dit gebied,

- b) het vooruit helpen van methodenvalidatie in de procesanalyse.

De twee projecten zijn complementair aan elkaar, zowel wat de wetenschappelijke inhoud als de samenstelling van de deelnemers betreft. Terwijl het SMT project slechts de Vrije Universiteit Brussel als partner heeft, liet dit project toe een samenwerking op te zetten op nationaal niveau tussen de industrie, wetenschappelijke instituten (als leden van het begeleidingscomite) en de universiteiten van Leuven, Luik en Brussel. Wat de wetenschappelijke inhoud betreft legt het SMT project meer de nadruk op het ontwikkelen robuuste methodes, terwijl in dit project het accent lag op de kosten en de efficiëntie in het kader van methodenvalidatie. Alhoewel de projecten complementair waren aan elkaar werd gehoopt dat er tussen beide synergieën ontwikkeld konden worden, wat in de loop van beide projecten ook het geval gebleken is.

2. Methodologie

Binnen het project werden verschillende taken uitgevoerd en verschillende aspecten van de methodenvalidatie bestudeerd. De verschillende onderdelen waarvoor getracht werd richtlijnen te formuleren of bestaande richtlijnen te evalueren, zijn :

- a) de evaluatie van de robuustheid van analytische methodes,
- b) de validatie van chromatografische bioanalysemethodes, d.w.z. methodes die bepalingen uitvoeren in biologische milieus,
- c) bepaling van de detectielimiet,

- d) de vergelijking van alternatieve meetmethodes,
- e) methodenvalidatiedesigns die kostenefficiënt zijn.
- f) bepaling van de reproduceerbaarheid van een methode,
- g) robuuste procesanalyse met multivariate calibratie,

Voor de verschillende bovenvermelde onderdelen was de aanpak meestal gelijklopend. In eerste instantie werd de reeds bestaande situatie, d.w.z. de uit de literatuur voorhanden zijnde gegevens, dieper onderzocht. In een tweede stap werden enkele specifieke applicatiegebieden geselecteerd en werden hierin een aantal case studies uitgevoerd. Vervolgens werd tenslotte geprobeerd om een algemene benadering en gefinaliseerde standaardprocedures te definiëren.

2.1. Bestudering van de bestaande situatie

Verschillende standaardisatie-organisaties hebben richtlijnen voorgesteld voor de terminologie die gebruikt moet worden om de validatie-parameters (zoals repeteerbaarheid en bias) van een analysemethode te beschrijven. De belangrijkste wereldorganisaties die hierbij betrokken zijn, zijn o.a. ISO (International Organisation for Standardization), IUPAC (International Union of Pure and Applied Chemistry) en AOAC (Association of Official Analytical Chemists). Naast deze wereldwijde organisaties zijn er nog vele actoren op Europees (zoals Eurachem en CEN), nationaal (zoals de Duitse DIN, de Franse AFNOR en de Nederlandse Keuringsdienst voor Waren) en regionaal niveau. Er zijn vroeger enkele pogingen tot harmonisatie geweest die geleid hebben tot een paar geharmoniseerde protocols tussen ISO, AOAC en IUPAC. In sommige gevallen hebben bepaalde industriële sectoren, zoals die van de farmaceutische analyse, via het ICH comité (International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use) een overeenkomst bereikt over de terminologie. In dit specifieke geval werd er ook een overeenkomst bereikt over validatie-parameters die geanalyseerd moeten worden naargelang de aard van de toepassing. Men was daarbij echter nog ver van het bereiken van een overeenkomst over hoe deze parameters gemeten en statistisch behandeld moeten worden. In sommige sectoren heeft men getracht een stap verder te gaan. Op het gebied van de bioanalyse van geneesmiddelen voor farmacokinetische studies, specificeert de zogenaamde "Washington consensus guideline" hoe goed de repeteerbaarheid moet zijn en hoeveel replicaten er minimaal nodig zijn om dit te bepalen.

Voor elk van de onderdelen en aspecten van de methodenvalidatie waarvoor criteria en normen opgesteld moesten worden, werd in eerste instantie het uit de literatuur beschikbare materiaal verzameld en bestudeerd.

Systematische, algemene en overeengekomen benaderingen die preciese experimentele richtlijnen bevatten en advies geven over hoe de validatie-parameters statistisch berekend

moeten worden, waren oorspronkelijk nog niet beschikbaar en werden gedurende dit project gedefinieerd.

2.2 Specifieke applicatiegebieden

Traditionele methodenvalidatie

Twee van de specifiek te bestuderen applicatiegebieden werden gekozen omdat het proces voor de ontwikkeling van richtlijnen op Europees- en Wereldniveau bezig is. Deze applicatiegebieden werden bestudeerd teneinde ook Belgische wetenschappers de gelegenheid te geven een prominente rol op dit terrein te spelen. De twee applicatiegebieden zijn (i) richtlijnen voor methodenvalidatie in farmacopee-analyses en (ii) de bioanalyse van geneesmiddelen voor farmacokinetische doeleinden.

Het eerste wordt bestudeerd door de Belgische Pharmacopee-commissie op nationaal niveau. Validatie of voortgezette validatie van bestaande officiële methoden voor de analyse van geneesmiddelen, gepubliceerd in de farmacopees, wordt meer en meer belangrijk. De ontwikkeling van voldoende gevalideerde nieuwe of verbeterde methoden voor de analyse van geneesmiddelen is eveneens noodzakelijk. Ook binnen dit project werden in dit verband een aantal case studies uitgevoerd.

Bioanalyse van geneesmiddelen werd, voor de aanvang van dit project, bestudeerd door een informeel comité met leden van o.a. de FDA, de Canadese tegenhanger ervan en leden van de industrie. Dit resulteerde in de zogenaamde 'Washington consensus guidelines'. De VUB (dienst prof. Massart) publiceerde, eveneens voor de aanvang van dit project, een kritiek op deze richtlijn en was eveneens aanwezig op de follow-up conferentie in Munchen. Gedurende dit project werden richtlijnen opgesteld voor de validatie van chromatografische methodes voor de dosage van geneesmiddelen in biologische milieus. Dit gebeurde in samenwerking met een commissie van de Franse SFSTP (Société Française de Sciences et Techniques Pharmaceutiques).

Vooraleer algemene richtlijnen voor methodenvalidatie gedefinieerd kunnen worden, moesten eerst een aantal specifieke problemen in detail bestudeerd worden. Het betreft hier validatie-criteria die gekend waren als knelpunten in de ontwikkeling van een methodenvalidatiestrategie, zoals de bepaling van de robuustheid en de detectielimiet van een methode.

Het eerste criterium dat aan een grondige studie onderworpen werd, is de bepaling van de robuustheid van een methode. Volgens de meest gangbare definitie bestudeert een robuustheidstest het effect van kleine veranderingen in een analysemethode op het gevonden resultaat. De bedoeling van dergelijke test is om de lange termijn precisie of de reproduceerbaarheid van een methode te voorspellen en zonodig te verbeteren teneinde uitwisselbaarheid van methodes tussen laboratoria of instrumenten te verhogen. Binnen dit

project werden een aantal case studies uitgevoerd, werd een strategie vooropgesteld welke een statistische interpretatie van de resultaten toelaat evenals een strategie voor minimale evaluatie van de robuustheid, werd een softwareprogramma geschreven en werden richtlijnen opgesteld voor het opstellen en interpreteren van een robuustheidstest.

Het tweede validatiecriterium dat bestudeerd werd, is de detectielimiet. Twee belangrijke benaderingen zijn mogelijk : (i) de klassieke, gebaseerd op de meting van blancostalen, en (ii) deze die gebaseerd is op informatie van de calibratielijjn. De tweede is statistisch nogal complex, maar is diegene die IUPAC voorstelt in een (concept)richtlijn. Praktische ervaring ermee was echter beperkt en het was noodzakelijk ze in verschillende situaties uit te testen voordat ze in praktijk gebracht kan worden.

Een situatie die zich bij het ontwikkelen en valideren van een methode frequent voordoet is dat men de resultaten van deze methode wil vergelijken met deze van een bestaande referentiemethode. Richtlijnen werden voorgesteld voor de vergelijking van twee analysemethoden, nl. voor de bias, de verschillende precisiecomponenten en de detectielimieten.

Proces analyse

Proces analytische methoden brengen de meting dicht bij het proces of binnen het proces. De gebruikelijke benadering van het ontwikkelen van selectieve analysemethoden om de aanwezige substanties of de karakteristieken van een produkt te meten, levert dan vaak methoden op die te langzaam zijn. De chemisch selectieve methodes moeten dan vervangen worden door een aantal nieuwe methoden, die gebruik maken van minder selectieve informatie (een spectrum, een chromatogram waarin niet alle pieken zijn gescheiden of geïdentificeerd, een set van resultaten van sensoren) en essentieel een multivariaat karakter hebben. Deze methoden zijn minder tijdsrovend en laten toe om de meting dicht bij het proces te brengen. Het economisch voordeel is evident.

Er zijn twee typische problemen die gebruik kunnen maken van dergelijke methoden. In het eerste, dat mengselanalyse wordt genoemd, is men meer geïnteresseerd in de kennis van het aantal componenten dat aanwezig is en met welke concentratieprofielen. In het tweede, multivariate calibratie genaamd, heeft men referentiestalen voor de te analyseren bestanddelen, en is men geïnteresseerd in hun concentraties.

In de twee gevallen worden de metingen direct op een vloeistofstroom uitgevoerd of op het produkt als dusdanig met een minimale voorbehandeling. Dergelijke methoden worden het meest gebruikt in de agro-voedingsindustrie, maar toepassingen in de farmaceutische industrie (bijvoorbeeld, analyse van tabletten zonder deze op te lossen) worden frekwenter en in de petrochemische industrie meet men op deze manier bijvoorbeeld het octaangetal.

Om informatie te extraheren uit de multivariate gegevens vertrouwt men op mathematische (chemometrische) methoden zoals evolving factor analyse, partial least squares en vele andere. De introductie van deze methoden wordt echter bemoeilijkt omdat er problemen zijn met de acceptatie ervan als standaardmethode.

Dit komt omdat men pas recent begonnen is met het onderzoeken van de beperkingen van deze methodes. De grootste zorg is hun betrouwbaarheid. Bijvoorbeeld, in multivariate calibratie is het in essentie de robuustheid van de calibratiemodellen en soms hun transfer die aandacht vereist. Het was de bedoeling om binnen dit project pre-normatief onderzoek uit te voeren met betrekking tot de manier waarop dit type van methoden gevalideerd moet worden. Uiteindelijk werd de problematiek uitgebreid onderzocht en leidde dit tot de formulering van een aantal richtlijnen terzake.

3. Resultaten

De verschillende onderdelen waarvoor getracht werd richtlijnen te formuleren of bestaande richtlijnen te evalueren, zijn :

- a) de evaluatie van de robuustheid van analytische methodes,
- b) de validatie van chromatografische bioanalysemethodes, d.w.z. methodes die bepalingen uitvoeren in biologische milieus,
- c) bepaling van de detectielimiet,
- d) de vergelijking van alternatieve meetmethodes,
- e) methodenvalidatiedesigns die kostenefficiënt zijn.
- f) bepaling van de reproduceerbaarheid van een methode,
- g) robuuste procesanalyse met multivariate calibratie,

a) Robuustheid (robustness/ruggedness) van analytische methodes

I.v.m. de studie van de robuustheid werd een filosofie opgesteld wat betreft de keuze van de experimentele designs en de statistische interpretatie van de effecten. Deze filosofie is beschreven in het artikel " *Y. Vander Heyden, F. Questier and D.L. Massart, A ruggedness test strategy for procedure related factors : experimental set-up and interpretation; Journal of Pharmaceutical and Biomedical Analysis, 17 (1998) 153-168*". Ze werd eveneens geïmplementeerd in een computerprogramma dat beschreven is in het manuscript "*F. Questier, Y. Vander Heyden and D.L. Massart; RTS, a computer program for the*

experimental set-up and interpretation of ruggedness tests; Journal of Pharmaceutical and Biomedical Analysis 18 (1998) 287-303".

Gebaseerd op discussies met de leden van het begeleidingscomité kon een lijst met discussiepunten opgesteld worden die in de loop van het project bestudeerd dienden te worden. Het gaat om de volgende punten :

- Wat is de bij voorkeur te gebruiken definitie voor ruggedness/robustness? Worden in een robuustheidstest voornamelijk procedure-gebonden factoren onderzocht in screening designs of volgt men meer de United States Pharmacopeia methode waar voornamelijk niet-procedure gebonden factoren onderzocht worden in nested designs? Nog een andere mogelijkheid welke door bepaalde bedrijven toegepast wordt, is het testen van de robuustheid m.b.v. een interlaboratoriumtest.

Case studies op voornamelijk procedure-gebonden factoren werden uitgevoerd in volgende manuscripten : (i) "Y. Vander Heyden and D.L. Massart; Y. Zhu and J. Hoogmartens; J. De Beer; *Ruggedness tests on the HPLC assay of the United States Pharmacopeia XXIII for tetracycline hydrochloride : comparison of different columns in an interlaboratory approach; Journal of Pharmaceutical and Biomedical Analysis 14 (1996) 1313-1326*", (ii) "Y. Vander Heyden, A. Bourgeois and D.L. Massart, *Influence of the sequence of experiments in a ruggedness test when drift occurs; Analytica Chimica Acta 347 (1997) 369-384*". Nested designs werden bestudeerd in "Y. Vander Heyden, K. De Braekeleer, Y. Zhu, J. Hoogmartens, J. De Beer and D.L. Massart, *Nested designs in ruggedness testing, aanvaard voor publicatie in Journal of Pharmaceutical and Biomedical Analysis*"

- Wanneer voert men een ruggedness test uit? Alleen voor methodes die onderworpen worden aan een interlaboratoriumtest of ook voor methodes die een meer beperkt gebruik kennen, b.v. alleen in één lab. Past men voor deze laatste methodes ook steeds een ruggedness test toe of alleen wanneer de tussen-dagen precisie slecht is?

- In de huidige filosofie worden de designs zodanig gekozen dat een bepaalde statistische interpretatie mogelijk is. Soms zijn daarvoor relatief veel experimenten nodig, meer dan de analyst verkiest uit te voeren. Daarom wordt eraan gedacht een strategie in stappen uit te bouwen die de analyst moet toelaten toch bepaalde conclusies te trekken na het uitvoeren van een minder uitgebreid design (invoeren van minimale designs). Hierdoor zal bepaalde

informatie ingeleverd moeten worden en zal er bijvoorbeeld geen of slechts een beperkte statistische analyse mogelijk zijn.

Daarom werden minimale designs aan de strategie toegevoegd. Deze laten wel het berekenen van effecten toe maar niet de oorspronkelijke statistische interpretatie. Een aantal onder hen kunnen na een eerste evaluatie wel uitgebreid worden tot de designs uit de oorspronkelijke strategie. Voor deze minimale designs werd toch getracht een significantie aan te geven voor de berekende effecten m.b.v. randomisatietesten.

De randomisatietesten gaven problemen wanneer er meer dan één significant effect was. Een sterk significant effect kan namelijk verantwoordelijk zijn voor de onderschatting van de significantie van de andere effecten. Na herhaaldelijke correcties waarbij elk significant bevonden effect geëlimineerd wordt, bleek dat in het algemeen dezelfde effecten als significant aangeduid worden door de randomisatietesten als door de t-testen uitgevoerd bij de "recommended designs" en waarbij de standard error berekend wordt gebaseerd op interactie-effecten of dummy factor effecten.

Uiteindelijk werden in de minimale designs de significante effecten geschat, gebruik makend van de gegevens van de niet-significante (algorithme van Dong).

- Kunnen supersaturated designs gebruikt worden in een ruggedness test?

Het gebruik van supersaturated designs als minimale designs in een robuustheidstest werd onderzocht. De gebruikte supersaturated designs werden geconstrueerd uit Plackett-Burman designs en voor een aantal case studies werd de variantie van de supersaturated designs vergeleken met deze van de Plackett-Burman designs (de berekening van effecten is nl. niet meer mogelijk voor de supersaturated designs). De verkregen varianties voor de supersaturated designs waren over het algemeen niet verschillend van deze van de Plackett-Burman designs wat erop wijst dat de supersaturated designs toelaten de variantie veroorzaakt door de factoren op een geschikte manier te schatten. Er zijn echter nog meer case studies vereist om dit te kunnen veralgemenen.

In een volgende stap diende bepaald te worden of de variantie gemeten voor een supersaturated design, veroorzaakt wordt door een al dan niet robuuste methode, m.a.w. een statistische test is vereist. In deze context werden een aantal voorstellen gedaan waarbij een kritische waarde berekend wordt uitgaande zowel van een theoretische benadering als van

practische resultaten. De designvariantie moet nl. vergeleken worden met een variantie die een normale experimentele fout voorstelt. De praktische bruikbaarheid van de voorgestelde criteria dient nog nagegaan te worden. Vooraleer daar echter mee te starten werd er nagegaan of er statistisch aanvaardbare resultaten gevonden kunnen worden. Dit blijkt echter niet het geval te zijn door het geringe aantal vrijheidsgraden te wijten aan het lage aantal uitgevoerde experimenten. Afhankelijk van de gebruikte hypothese zal men ofwel een grote probabiliteit hebben om niet-robuste methodes als robuust te beschouwen, ofwel robuuste methodes te verwerpen als niet-robust. Daarom dienen er andere interpretatiemethodes voor dergelijke designs gezocht te worden.

- Moet een factor op 2 (extreme) of op 3 (2 extreme en 1 nominaal) niveaus onderzocht worden?

- Wenst de gemiddelde gebruiker in staat te zijn de effecten statistisch te interpreteren? Het antwoord lijkt te zijn dat nogal wat gebruikers dat niet wensen. Ook de Eurachem voorstellen lijken in die richting te gaan.

Vaak is een analyst meer geïnteresseerd in chemisch relevante effecten dan in statistisch significante. Hoe kunnen deze chemisch relevante effecten bepaald worden?

In het artikel "*Y. Vander Heyden, C. Hartmann and D.L. Massart; P. Nuyten, A.M. Hollands and P. Schoenmakers; Ruggedness testing of a size exclusion chromatographic assay for low molecular mass polymers; Journal of Chromatography A, 756 (1996) 89-106*" is getracht daar een antwoord op te geven.

- Kan men bij de bepaling van statistisch significante of chemisch relevante effecten er van uitgaan dat men reeds de repeteerbaarheid of de tijdsafhankelijke precisie kent?

- Hoe moet men bepaalde factoren includeren in de screening designs die uitgevoerd worden? B.v. (i) de componenten van een mengsel, (ii) de zure en basische component van een buffermengsel.

Het onderzoeken van deze factoren evenals van enkele andere in screening designs werd bestudeerd en is beschreven in een artikel "*Y. Vander Heyden, F. Questier and D.L. Massart; Ruggedness testing of chromatographic methods : selection of factors and levels; Journal of Pharmaceutical and Biomedical Analysis 18 (1998) 43-56*". In deze referentie

wordt er beschreven hoe voor een bepaalde set van proces- en mengselvariabelen men de meest economische experimentele set-up kan definiëren. Er wordt eveneens aangetoond dat afhankelijk van hun formulering, voor bepaalde factoren (buffercomponenten) men meer of minder fysisch relevante effecten kan berekenen.

In de tekst wordt tevens een algemene richtlijn gegeven voor de keuze van de extreme niveaus in een robuustheidstest, alsook worden enkele speciale gevallen besproken (keuze van asymmetrische intervallen rond het nominale niveau bijvoorbeeld).

- Een robuustheidstest moet aangeven in welke mate fouten/afwijkingen in de factoren aanleiding geven tot wijzigingen in de respons. Het doel van een robuustheidstest moet zijn dat de gevonden resultaten en de getrokken conclusies leiden tot een verbetering van de resultaten van een interlaboratoriumtest d.w.z. de reproduceerbaarheid verbetert. Daarom zouden de resultaten van een robuustheidstest moeten leiden tot het creëren van system suitability criteria voor de (kritieke) factoren. Een voorstel terzake is gedaan in "*Y. Vander Heyden, M. Jimidar, E. Hund, N. Niemeijer, R. Peeters, J. Smeyers-Verbeke, D.L. Massart and J. Hoogmartens; Determination of system suitability limits with a robustness test, aanvaard voor publicatie in J. of Chromatography A*".

- Stel dat men een methode niet zelf ontwikkeld heeft, maar men er toch een robuustheidstest op uitvoert. De nominale waarden voor de onderzochte factoren worden in dit geval overgenomen van een bepaalde bron. Wanneer nu bepaalde effecten groot zijn d.w.z. de methode is niet robuust, hoe kan men gebaseerd op de ruggedness test resultaten, niveaus definiëren voor deze factoren om het responsoppervlak te bepalen en tot een nieuwe optimalisatie te komen?

Uit de opgedane kennis werden er tenslotte richtlijnen gedefinieerd voor het opstellen en interpreteren van een robuustheidstest welke de definitie voor robuustheid van de ICH volgen. Deze richtlijnen bevatten eveneens de mogelijkheid om system suitability limieten te bepalen uit de resultaten van een robuustheidstest. Dit laatste wordt eveneens aanbevolen door de ICH. De verschillende stappen in de set-up en interpretatie van een robuustheidstest worden in de richtlijnen gegeven. De richtlijnen werden opgesteld in samenwerking met Unilever (Vlaardingen, Nederland), terwijl er een coöperatie bestond met Novartis (Basel,

Zwitserland), Bayer (Dormagen, Duitsland) en Shell (Amsterdam, Nederland) tijdens het uitwerken van een aantal case studies die het uiteindelijk opstellen van de richtlijnen toelieten. De richtlijn bevat twee delen, namelijk de selectie van de experimentele set-up (definitie van factoren en hun niveaus, en selectie van de designs) en de behandeling van de resultaten van een uitgevoerd design. Ze bevat bijvoorbeeld voorstellen voor het uitvoeren design (Plackett-Burman of fractional factorial design) in functie van het aantal te bestuderen factoren. De behandeling van de resultaten bestaat in de berekening van effecten en hun grafische en statistische interpretatie teneinde hun significantie te bepalen. Wanneer een methode dan als robuust beschouwd wordt wat zijn kwantitatief aspect betreft, kunnen voor een aantal responsen, zoals resolutie, analysetijd, asymmetriefactoren in chromatografie, system suitability limieten gedefinieerd worden. Deze richtlijn (*Y. Vander Heyden, A.Nijhuis, B.G.M. Vandeginste and D.L. Massart; Robustness/Ruggedness Tests in Method Validation*) werd toegevoegd in bijlage.

Daarnaast werken een aantal leden van dit onderzoeksprogramma, nl. P. Chiap (ULg), Ph. Hubert (ULg) en Y. Vander Heyden (VUB), en van het begeleidingscomité (B. Boulanger (E. Lilly) samen binnen een commissie "Robustesse" van de Franse SFSTP aan een artikel dat eveneens een aantal richtlijnen zal voorstellen voor het opstellen en het interpreteren van een robuustheidstest en dus gedeeltelijk overlapt met de hogervermelde richtlijnen. Het artikel zal twee grote delen omvatten, nl. (i) het implementeren van robuustheid en robuustheidscriteria in methodeoptimalisatie, en (ii) het uitvoeren van robuustheidstesten in methodenvalidatie. Het gedeelte over de robuustheid in methodenvalidatie zal grotendeels gebaseerd zijn op de resultaten en richtlijnen beschreven in de doctoraatsthesis van Y. Vander Heyden. Hieruit blijkt duidelijk de invloed, binnen deze commissie, van de Belgische leden van dit normalisatieprogramma op het Franse niveau.

Een onderwerp dat naar het einde van het project toe onderzocht werd, is het gebruik van zogenaamde asymmetrische designs in robuustheidstesten. Deze designs laten toe sommige factoren op meer dan twee niveaus te onderzoeken. Dit laatste is voor kwalitatieve factoren in feite een vereiste om een min of meer representatief idee te krijgen over hun invloed op de robuustheid van een analyseresultaat. De resultaten van dit onderzoek zijn beschreven in de volgende manuscripten "*E. Hund, Y. Vander Heyden, M. Haustein, D.L. Massart and J. Smeyers-Verbeke, Comparison of several criteria to decide on the significance of effects in*

a robustness test with an asymmetrical factorial design, document in voorbereiding” en “*E. Hund, Y. Vander Heyden, M. Haustein, D.L. Massart and J. Smeyers-Verbeke, Application of fractional factorial and asymmetrical factorial designs in the robustness testing of a RP-HPLC method, document in voorbereiding*”.

b) Validatie van chromatografische bioanalysemethodes, d.w.z. methodes die bepalingen uitvoeren in biologische milieus

In samenwerking met de SFSTP werd een strategie besproken voor de validatie van chromatografische methodes voor bepalingen in biologische milieus. Deze besprekingen leidden enerzijds tot de publicatie van een aantal richtlijnen i.v.m. validatie in biologische milieus (*S. Bervoas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, P. Hubert, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, E. Chapuzet et N. Mercier; Méthodes chromatographiques de dosage dans les milieus biologiques : stratégie de validation; S.T.P. Pharma Pratiques 7, 3 (1997) 169-194*) en tot een aantal case studies ter evaluatie van de voorgestelde richtlijnen ((i) *S. Bervoas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, P. Hubert, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, E. Chapuzet et N. Mercier; Méthodes chromatographiques de dosage dans les milieus biologiques. Exemple d'application de la stratégie de validation; S.T.P. Pharma Pratiques 8, 2 (1998) 81-107*, (ii) *Ph. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie et J.C. Nivet; The SFSTP Guide on the validation of chromatographic methods for drug bioanalysis: from the Washington Conference to the laboratory, ingediend voor publicatie*, (iii) *P. Chiap, Ph. Hubert, B. Boulanger, J. Crommen, Validation of an automated method for the liquid chromatographic determination of atenolol in plasma: application of a new validation protocol, ingediend voor publicatie*, en (iv) *B. Boulanger, Ph. Hubert, P. Chiap, J. Crommen; An analysis of the SFSTP guide on validation of chromatographic bioanalytical methods: progress and limitations, ingediend voor publicatie*) en anderzijds tot de ontwikkeling van een softwaresysteem voor methodevalidatie in deze biologische milieus.

c) Bepaling van de detectielimiet

Een aanvang werd genomen met het vergelijken van twee benaderingen voor de evaluatie van de detectielimiet : *i)* gebaseerd op de meting van blancostalen; *ii)* gebaseerd op de informatie van de calibratielijn. Ze werden toegepast op NIR metingen gebaseerd op univariate calibratie. Verdere evaluatie is echter noodzakelijk.

d) De vergelijking van alternatieve meetmethodes

Een aantal bestaande standaarden en richtlijnen, zoals ISO 5725, AFNOR T90-120 en Eurachem werden verzameld en bestudeerd met het oog op hun bruikbaarheid voor het vergelijken van analysemethoden. Het doel van deze studie is richtlijnen te hebben die toelaten aan te tonen dat (i) een nieuwe methode even goed of beter presteert dan een bestaande referentiemethode, of (ii) een bepaalde methode zonder problemen getransfereerd kan worden van een researchcentrum naar een industrieel labo of van het ene naar het andere industrieel labo. De ISO-normen zijn beperkt tot het gebruik van een methode in een bepaalde matrix en voor een bepaalde concentratie en zijn bedoeld voor interlaboratoriumstudies. Ze moesten geherdefinieerd worden voor het bestuderen van meerdere matrices en meerdere concentratieniveaus en voor intralaboratoriumstudies. Het voordeel van de ISO-norm tegenover andere validatiestandaarden is dat zowel de β -fout als interval hypothesetesten geïnccludeerd zijn. In verband met het vergelijken van methodes werd dus een protocol opgesteld uitgaande van de bestaande standaarden. De resultaten zijn beschreven in het artikel "*S. Kuttatharmakul, D.L. Massart and J. Smeyers-Verbeke, Comparison of alternative measurement methods, accepted for publication in Analytica Chimica Acta*".

Op basis van de ISO 5725 (deel 6-sectie 8) norm werd een voorstel van richtlijnen uitgewerkt voor het vergelijken van analysemethoden. Daartoe werd de ISO norm, die gebaseerd is op interlaboratoriumstudies, aangepast voor het vergelijken van twee analysemethoden binnen eenzelfde organisatie (intralaboratoriumvergelijking). Dit betekent b.v. dat de precisie niet bestudeerd wordt onder reproduceerheidscondities maar dat intermediaire precisie condities (v.b. verschillende dagen en operatoren) moeten beschouwd worden. De voorgestelde richtlijnen houden nu rekening met de "time-different intermediate" precisie en dienen uitgebreid te worden tot de "[time+operator]-different intermediate" precisie.

Een eerste voorstel van richtlijn start zoals de ISO norm met een bepaling van het aantal metingen dat nodig is om de β -fout te controleren (de β -fout in deze context is de probabiliteit dat een relevante bias niet gedetecteerd wordt of dat een aanvaardbare precisie niet gehaald wordt). Deze benadering kan echter aanleiding geven tot een aantal uit te voeren metingen dat in praktijk niet of moeilijk haalbaar is. Daarom werd een tweede richtlijn ontwikkeld die start met een door de gebruiker bepaald aantal metingen en die de β -fout evalueert. Op die manier krijgt de analyst ten minste een idee van de waarschijnlijkheid dat een methode die eigenlijk niet aan de voorwaarden voldoet, aanvaard wordt.

De eerder voorgestelde richtlijnen voor het vergelijken van analysemethoden houden enkel rekening met de "time-different intermediate" precisie ($s_{I(T)}^2$). Nagegaan werd hoe een

uitbreiding naar de "[time+instrument+operator]-different intermediate" precisie ($s_{I(OIT)}^2$) mogelijk is. Het probleem bij een intralaboratoriumvergelijking van methoden is namelijk dat meestal het aantal instrumenten gebruikt voor de analyse alsook het aantal analisten dat de analyses uitvoert beperkt is. Dit leidt tot een slechte schatting van de variantiecomponenten voor instrumenten en operatoren. Daarom wordt voorgesteld $s_{I(OIT)}^2$ niet als de som van de individuele variantiecomponenten te berekenen maar deze af te leiden uit de kwadraatsommen van de verschillen tussen de daggemiddelden en het algemeen gemiddelde. Alhoewel de op deze manier bekomen $s_{I(OIT)}^2$, die afhankelijk kan zijn van het aantal beschouwde instrumenten en operatoren, misschien geen goede schatter is voor de "[time+instrument+operator]-intermediate" precisie is het bruikbaar in de vergelijking van methodes op voorwaarde dat eenzelfde aantal operatoren en instrumenten voor beide methodes beschouwd worden.

De twee hoger voorgestelde richtlijnen voor het vergelijken van analysemethoden (de eerste gebaseerd op een bepaling van het aantal metingen nodig om de β -fout tot een gespecificeerde waarde te beperken, de tweede vertrekkend van een door de gebruiker gespecificeerd aantal metingen en een evaluatie van de β -fout) werden uitgewerkt en aangepast. Daarbij werd rekening gehouden met opmerkingen gemaakt door o.a. Elf en Novartis (industriële partners in het hoger vermelde Europese SMT project gecoördineerd door de VUB).

Het was daarbij noodzakelijk om in de eerste richtlijn een beslissing te nemen i.v.m. de aanvaardbare β -fout (hier de probabiliteit dat een relevante bias niet gedetecteerd wordt of een gespecificeerde precisie niet gehaald wordt) voor het berekenen van het aantal uit te voeren metingen (n). Om praktische redenen (n neemt toe naarmate men meer zekerheid wil dat b.v. een relevante bias inderdaad gedetecteerd zal worden, dit betekent naarmate men een kleinere β -fout wil) werd de voorkeur gegeven om $\beta=20\%$ te specificeren in plaats van 5% zoals in de eerste versie voorgesteld. Vermits de verschillende statistische testen op het klassieke 5% significantieniveau ($\alpha=5\%$) uitgevoerd worden, betekent dit dat nog altijd meer belang gehecht wordt aan de α - dan aan de β -fout. Met andere woorden men beschouwt het, louter om praktische redenen, belangrijker om een methode die eigenlijk aanvaardbaar is niet te verwerpen dan om een methode die eigenlijk niet geschikt is te aanvaarden.

Een betere balans tussen de α - en de β -fout zou verkregen kunnen worden door $\alpha=10\%$ en $\beta=10\%$ te specificeren. Het aantal uit te voeren metingen met deze waarden voor α en β is

vergelijkbaar met deze voor $\alpha=5\%$ en $\beta=20\%$. Alhoewel dit logischer is, werd het in de richtlijn niet voorgesteld omdat dit zou betekenen dat de statistische testen uitgevoerd worden op het 10% in plaats van op het klassieke 5% significantieniveau. Dit laatste is zo standaard dat afwijken hiervan momenteel waarschijnlijk moeilijk aanvaard zou worden.

Vermits voorgesteld werd om de richtlijnen in te dienen bij de CEN werden ze samengevoegd in één document dat verspreid werd onder de partners in dit onderzoeksprogramma en in het SMT project.

In de hoger vermelde richtlijnen voor het vergelijken van analysemethodes wordt enkel rekening gehouden met de "time-different intermediate" precisie. Zoals beschreven in het vorig verslag werd nagegaan hoe een uitbreiding naar de "[time+instrument+operator]-different intermediate" precisie mogelijk is. Een artikel in dat verband is aanvaard voor publicatie in *Analytica Chimica Acta*, nl. "S. Kuttatharmakul, D.L. Massart and J. Smeyers-Verbeke, *Comparison of alternative measurement methods, accepted for publication in Analytica Chimica Acta*"

Vergelijking van analysemethodes : de voorgestelde richtlijnen voor de vergelijking van twee analysemethodes, beperken zich tot een vergelijking van de bias en van verschillende precisiecomponenten. De voorgestelde richtlijnen zijn ofwel gebaseerd op een bepaling van het aantal metingen nodig om de β -fout tot een gespecificeerde waarde te beperken, ofwel vertrekkend van een door de gebruiker gespecificeerd aantal metingen en een evaluatie van de β -fout. Nagegaan werd ook hoe de detectielimieten van beide methodes vergeleken konden worden. Dit is uiteraard belangrijk wanneer het methoden voor sporenanalyse betreft.

e) Methodvalidatiedesigns die kostenefficiënt zijn : lineariteit ijklijn

Wat betreft de studie van methodevalidatie designs die kosten-efficiënt zijn, werden verschillende designs vergeleken voor de evaluatie van de lineariteit van de ijklijn. Een evaluatie van lineariteitstesten (Anova lack-of-fit test en testen van de significantie van de kwadratische coëfficiënt b_2) werd uitgevoerd aan de hand van gesimuleerde niet-lineaire calibratielijnen met een verschillende graad van afbuiging die gebaseerd zijn op een beperkt aantal metingen ($6 < n \leq 10$). Verschillende designs waarin de n metingen verdeeld zijn over 3, 4 of 5 concentratieniveaus werden daarbij bestudeerd. De waarschijnlijkheid dat een lack-of-

fit niet gedetecteerd wordt (= de β -fout), is het kleinst wanneer slechts 3 concentratieniveaus beschouwd worden. Heteroscedastiteit (niet constante variantie over het beschouwde concentratiegebied), die gezien het gering aantal metingen, niet aan de hand van de experimentele gegevens kan onderzocht worden, geeft aanleiding tot een aanzienlijke toename van de β -fout. Vermits de β -fout ook afhankelijk is van de precisie werd de invloed van deze laatste ook nagegaan. Dit is belangrijk omdat dit toelaat te bepalen welke precisie noodzakelijk is om een relevante afwijking van de lineariteit te detecteren. In praktijk is het echter moeilijk te specificeren welke afwijking van lineariteit als belangrijk beschouwd wordt. Vermits een lack-of-fit echter aanleiding geeft tot fouten in de voorspelde concentraties (en dus tot een bias) wordt nu de β -fout van de lineariteitstesten in functie van de predictiefout bestudeerd. Bovenvermelde resultaten zijn beschreven in het manuscript "S. Kuttatharmakul, D.L. Massart and J. Smeyers-Verbeke, Influence of precision, sample size and design on the beta error of linearity tests; *Journal of Analytical Atomic Spectroscopy (JAAS)* 13 (1998) 109-118"

f) Bepaling van de reproduceerbaarheid van een methode en methodenvalidatie in farmacopee-analyses

Een van de applicatiegebieden waarbinnen het uittesten van richtlijnen en normen gepland waren, is de methodenvalidatie in farmacopee-analyses.

De toepassing van standaarden en richtlijnen in verband met meetmethoden uit de farmacopee voor de analyse van geneesmiddelen maakte dus een belangrijk onderdeel uit van het project. De bestaande voorschriften en richtlijnen, moeten daarom eerst toegepast worden op enkele case studies. Op punt stellen van nieuwe of verbeterde methoden blijft eveneens noodzakelijk.

Na toetsing werd gekozen om de analysemethodes voor enkele belangrijke antibiotica als model te nemen. De biotechnologie van de industriële antibiotica is op dit ogenblik trouwens in volle expansie, wat ongetwijfeld een diepgaande invloed zal hebben op de aard en samenstelling van deze producten.

Onze keuze kan als volgt toegelicht worden. Spiramycine en tylosine, beide behorend tot de groep van macrolide antibiotica, worden niet alleen aangewend respectievelijk in de humane

en veterinaire geneeskunde, maar hebben ook toepassing gevonden in de agro-voeding als groeibevorderende stoffen.

Bacitracine, een polypeptide antibioticum, wordt veelal aangewend als zinkcomplex ter bevordering van de voedselopname efficiëntie bij sommige diersoorten en ziekte controle. Het is in deze sector het economisch belangrijkste produkt.

Amoxicilline, een belangrijke vertegenwoordiger van de β -lactam groep werd eveneens als model gekozen. Het is een semi-synthetische penicilline met activiteit tegen zowel gram-positieve als gram-negatieve bacteria. Het wordt gebruikt in verschillende farmaceutische vormen zoals capsules, orale suspensies als injectievloeistoffen.

De 3 eerst vermelde antibiotica kunnen alleen als fermentatieprodukt verkregen worden door klassieke biosynthese. Dit houdt in dat het isoleren gebeurt uitgaande van complexe voedingsmilieus ontstaan tijdens de gecontroleerde groei van veelal gemodificeerde micro-organismen.

De produktievoorwaarden en isoleringsschema's, bekomen na een uitgebreide reeks van experimenten, leiden tot economisch rendabele opbrengsten en behoren tot de confidentiële kennis van de producenten. De bekomen actieve grondstoffen bezitten meestal een complexe samenstelling met een groep actieve componenten naast een wisselend aantal inactieve bestanddelen. Hun chemische aard is de oorzaak dat gemakkelijk ongewenste nevenprodukten kunnen ontstaan door chemische omzetting tijdens het isoleren of de opzuivering. Deze onzuiverheden bezitten meestal geen activiteit en sommige kunnen zelfs toxisch zijn.

De oorspronkelijke microbiologische procedure die tot op heden voor heel wat antibiotica wordt toegepast, is gebaseerd op de activiteit ten opzichte van gevoelige micro-organismen. Deze methoden zijn zeer bewerkelijk, langdurig en onderhevig aan tal van interferenties. De wisselende samenstelling van de monsters ten opzichte van de gebruikte standaard geeft aanleiding tot problemen bij deze activiteitsbepalingen, wat ook zal blijken bij de bespreking van de bekomen resultaten! Het gebruik van alternatieve methoden, zoals de scheidingsmethoden HPLC en CE, dringt zich dan ook op.

Het gebruik van fysico-chemische methoden om de kwaliteitscontrole van antibiotica en analoge farmaceutische produkten uit te voeren, is reeds enkele jaren in opgang. Er moet

naar gestreefd worden om de vereiste selectiviteit te bereiken. De signalen die gemeten worden moeten selectief zijn voor de te bepalen stoffen. De methode die tegenwoordig meest gebruikt wordt, is een combinatie van een chromatografische scheiding gekoppeld aan een spectrometrische bepaling. Voor de onderzochte modellen is vloeistofchromatografie of capillaire elektroforese met ultraviolet detectie de voor de hand liggende keuze.

Interlaboratoriumstudies werden opgesteld op twee analysemethodes (farmacopee methodes) voor de bepaling van het antibioticum tylosine. Deze studies werden opgesteld volgens de ISO-normen terzake (ISO 5725-2). De eigenlijke interlaboratoriumexperimenten werden voorafgegaan door een zogenaamde trainingsronde. Dergelijke trainingsronde laat toe een aantal problemen die zich voordoen in de deelnemende labo's op te lossen zodanig dat men in de uiteindelijke studies deze problemen kan vermijden en tot een maximaal aantal bruikbare resultaten komt.

De behandeling van de resultaten van de interlaboratoriumstudies gebeurde volgens de ISO 5725-2 richtlijn. De gegevens werden eerst getest op outliers (zowel outliers op het niveau van de binnen-laboratoriumvariantie als van de tussen-laboratoriumvariantie). Na verwijderen van de outliers wordt de repeteerbaarheid en de reproduceerbaarheid van de methode berekend en wordt er nagegaan of er een verband bestaat (lineair of logaritmisch) tussen de variantie en de concentratie te bepalen bestanddeel (homo/heteroscedasticiteit).

Aangezien de studie uitgevoerd werd m.b.v. twee bepalingsmethodes op dezelfde stalen, konden de resultaten van beide methodes vergeleken worden. Er werden zowel kwantitatieve parameters (varianties, gemiddelde resultaten) als kwalitatieve (resolutie, analysetijd, enz.) beschouwd.

De resultaten van dit onderzoek werden opgenomen in een tutorial artikel "*Y. Vander Heyden, J. Saevels, E. Roets, J. Hoogmartens, D. Decolin, M.G. Quaglia, W. Van den Bossche, R. Leemans, O. Smeets, F. Van de Vaart, B. Mason, G.C. Taylor, W. Underberg, A. Bult, P. Chiap, J. Crommen, J. De Beer, S.H. Hansen and D.L. Massart, Interlaboratory studies on two HPLC assays for tylosin (tartrate), Journal of Chromatography A, 830 (1999) 3-28*".

Wat de andere substanties (spiramycine, bacitracine, amoxicilline) betreft, werd zowel methodenoptimalisatie als methodenvalidatie uitgevoerd. De resultaten ervan zijn in meer detail beschreven in (i) *Y.M. Li, A. Van Schepdael, Y. Zhu, E. Roets, J. Hoogmartens; Development and validation of amoxicillin determination by micellar electrokinetic capillary chromatography, Journal of Chromatography A, 812 (1998) 227-236*, en (ii) *L. Liu, J. Saevels, P. Louis, H. Nelis, S. Rico, K. Dierick, S. Guyomard, E. Roets, J. Hoogmartens; Interlaboratory study comparing the microbiological potency of spiramycins I, II and III; in druk in Journal of Pharmaceutical and Biomedical Analysis.*

g) Robuuste procesanalyse met multivariate calibratie

Wat de ontwikkeling van een filosofie voor procesanalyse betreft, werden twee problemen bestudeerd. Een eerste was het nagaan of de gegevens afwijkende metingen bevatten. Als oplossing van dit probleem werd een strategie vooropgesteld, voor NIR metingen, gebaseerd op het meten van leverage punten en op een evolutieprogramma, lijkend op een genetisch algoritme.

Het tweede probleem was het bepalen van de factoren welke een invloed hebben op de precisie. Hiervoor werd gevonden dat een aantal duplicate of replicate metingen van een beperkt aantal standaarden een even goed model en een even goede precisie oplevert als een groot aantal standaarden die slechts éénmaal gemeten worden.

In het laatste deel van het project werkte de VUB partner eveneens mee aan het opstellen van validatierichtlijnen voor Nabij Infrarood (NIR) spectrofotometrie. Deze richtlijnen worden opgesteld op Europees niveau en zijn bedoeld voor toepassing in de industrie. Andere partners in dit project zijn o.a. de multinationals Pfizer, Shell en Elf.

Veel methoden voor multivariate calibratie zijn in de literatuur voorgesteld. Het bleek nu dat meerdere van deze methoden een analoge performantie hebben. Om verwarring te vermijden, stellen we voor dat slecht de volgende methodes beschouwd zullen worden

- Multiple linear regression (MLR) : multi-pele lineaire regressie
- Principal component regression (PCR) : principale componentenregressie
- Partial least squares (PLS)

- Neural networks (NN) : neurale netwerken
- Locally weighted regression (LWR) : lokaal gewogen regressie
- Radial basis functions combined with PLS (RBF-PLS) : radial basis functies gecombineerd met PLS

Ze werden opgenomen op basis van theoretische overwegingen, die bevestigd werden door onderlinge vergelijking van hun performantie uitgevoerd op NIR data sets met verschillende data structuren.

Tijdens bovenvermelde vergelijkingen werden nog vele andere methodes uitgetest, zoals niet-lineaire varianten van PLS, PCR en MLR, nearest-neighbour methodes, methodes waarin de regressie is uitgevoerd op Fourier coëfficiënten, en wavelets. Deze werden tenslotte niet weerhouden omdat sommige geen goede resultaten opleverden, terwijl andere vergelijkbare resultaten met de bovenvermelde methodes gaven maar veel moeilijker te begrijpen of minder algemeen zijn. Om deze redenen is hun gebruik niet aangewezen door gebruikers die niet vertrouwd zijn met chemometrie.

- MLR is vanuit statistisch oogpunt gezien de best gekende methode. De methode is gebaseerd op de selectie van variabelen. Ze verdient de voorkeur wanneer de selectie van variabelen eenvoudig is, d.w.z. wanneer bepaalde variabelen relatief selectief zijn voor de componenten of eigenschappen die bepaald worden.
- PCR is eveneens een gekende methode die, van de beschreven methodes, het dichtst aanleunt bij PCR. Het is een tweestapsprocedure. In de eerste stap worden principale componenten bepaald. Dit zijn lineaire combinaties van de originele variabelen. Ze kunnen als nieuwe variabelen beschouwd worden die op een optimale manier de variantie bundelen welke in de spectra aanwezig is. Dergelijke variabelen die de originele variabelen combineren op een dusdanige manier dat deze combinaties een bepaald criterium optimaliseren (hier de verklaarde variantie maximaliseren) worden ook latente variabelen genoemd. In de tweede stap wordt dan MLR toegepast op de nieuw verkregen latente variabelen.
- PLS is eveneens een methode die vaak gebruikt wordt in multivariate calibratie. Het lijkt op PCR maar werkt in één stap. De methode bepaalt eveneens latente variabelen,

maar het gebruikte criterium hier is de maximale covariantie tussen de y-waarden en de spectrale variabelen.

- Neurale netwerken (NN) : slechts backpropagation netwerken werden gebruikt. Neurale netwerken starten met het maken van combinaties van de originele variabelen. Deze combinaties worden dan non-lineair gemaakt door het gebruik van transferfuncties.
- Lokaal gewogen regressie (LWR) gebruikt PCR of PLS, in combinatie met een wegingssysteem zodat de calibratiestandaarden die het dichtst staan tot het te voorspellen staal, ook het grootste gewicht krijgen.
- RBF-PLS : dit is een lokale methode die zeer goed blijkt te werken voor moeilijke datastructuren

De publicaties gemaakt in het kader van dit onderzoek zijn opgenomen in de bijlage.

Op basis van de opgedane kennis kon volgend beslissingsschema opgesteld worden.

Data set acquisition and pre-processing

Data inspection :
- PC score plots
- Plot of most correlated PCs vs y
- Histogram of y-values

Obvious outliers detected

yes

Remove outliers

no

Calibration space covers expected prediction space

no

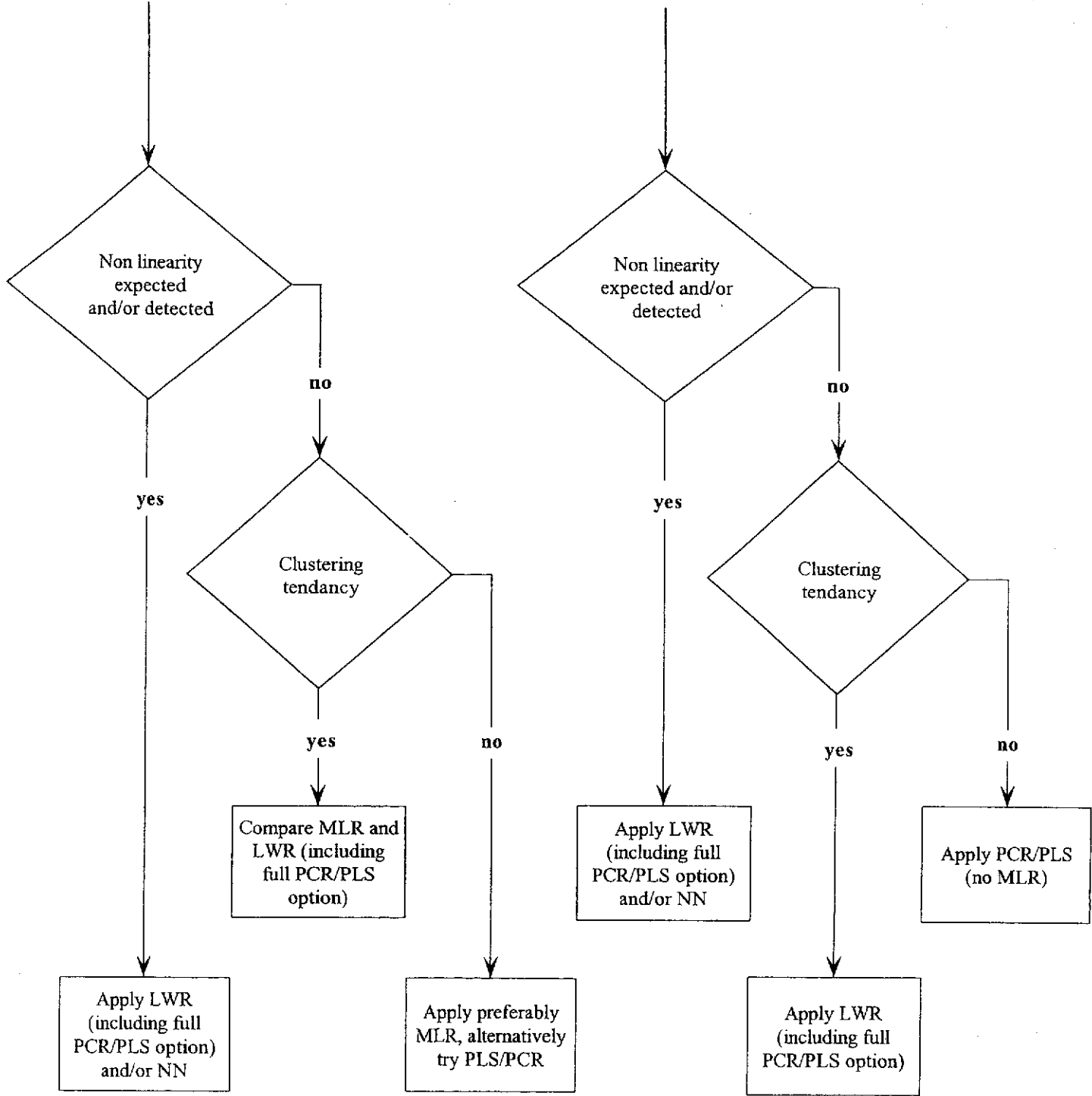
Accurate prediction required outside calibration space

yes

yes

no

Predicted y-values outside the calibration space will indicate that a different model must be applied



4. Besluiten en aanbevelingen

De verschillende onderdelen waarvoor in dit project richtlijnen geformuleerd werden of bestaande richtlijnen geëvalueerd, waren :

- a) de evaluatie van de robuustheid van analytische methodes,
- b) de validatie van chromatografische bioanalysemethodes, d.w.z. methodes die bepalingen uitvoeren in biologische milieus,
- c) bepaling van de detectielimiet,
- d) de vergelijking van alternatieve meetmethodes,
- e) methodenvalidatiedesigns die kostenefficiënt zijn, met name voor de lineariteit van een ijklijn,
- f) bepaling van de reproduceerbaarheid van een methode,
- g) robuuste procesanalyse met multivariate calibratie,

Methodenvalidatie is noodzakelijk voor de industrie, maar ook duur. Om die reden werd er bij het opstellen van bepaalde normen en richtlijnen, eveneens getracht om volgende zaken te definiëren : (i) methodenvalidatiedesigns die kostenefficiënt zijn, en (ii) minimale vereisten voor methodenvalidatie.

- (i) Een van de meest belangrijke vragen in methodenvalidatie is welke en hoeveel experimenten uit te voeren. Hoe meer experimenten er uitgevoerd worden, hoe beter de prestatiekenmerken bepaald kunnen worden, maar des te duider ook de validatie. Goede experimentele designs laten toe het aantal experimenten te beperken en toch genoeg informatie te verkrijgen. Het experimentele design hangt af van de situatie. Het is bijvoorbeeld verschillend voor een methode toegepast voor het testen van sommige productspecificaties, waar de concentratierange die wordt getest beperkt is, en voor situaties waar de range groot kan zijn, zoals in de voedingsmiddelen- of bioanalyse.
- (ii) Methoden kunnen in verschillende mate gevalideerd worden. In veel praktische situaties is een uitgebreide validatie niet haalbaar, noch noodzakelijk. De vraag is dan of overeenstemming kan worden gevonden over at het minimum is, afhankelijk van de context. Een benadering voor het definiëren van dergelijke minimum vereisten moet ook diagnostica of regels bevatten die toelaten te beslissen wanneer het minimum niet genoeg is en er meer gedaan moet worden.

Er dient opgemerkt te worden dat in dit project voor een aantal methodenvalidatieonderdelen wel rekening gehouden werd met het opstellen van kostenefficiënte designs (b.v. het nagaan van de lineariteit van een ijklijn) en met minimale vereisten (b.v. gebruik van minimale designs in robuustheidstesten), maar dat dit niet op een veralgemeende basis het geval is. In

de toekomst is er dus nog zeker werk weggelegd voor het ontwikkelen van verdere minimale en kostenefficiënte methodenvalidatieonderdelen, zodat men uiteindelijk tot een situatie komt waarbij men validatierichtlijnen heeft zowel voor een volledige als voor een minimale validatie.

Alhoewel de gebruikelijke statistische pakketten kunnen helpen in methodenvalidatie, is er geen enkel beschikbaar dat voldoende algemeen is en expliciet gericht op methodenvalidatie. De mogelijkheden van de bestaande methodenvalidatieprogramma's daarentegen zijn echter sterk gelimiteerd in vergelijking tot het hier onderzochte aantal onderwerpen. Inderdaad moet men rekening houden met de chemische context, die testen en tools vereisen die niet beschikbaar zijn in de meeste klassieke statistische pakketten. Een eenvoudig voorbeeld is het vergelijken van de helling van twee calibratielijnen (is er een statistisch significant verschil tussen beide?), hetgeen vaak noodzakelijk is voor het detecteren van interferenties met de zogenaamde standaardadditiemethode. Omdat statistische conclusies vaak beperkt zijn tot fouten van de eerste soort is het niet mogelijk te besluiten hoeveel experimenten men moet uitvoeren om twee methoden of twee laboratoria te vergelijken, zoals vaak vereist is in methodenvalidatie. Bovendien zijn de meeste chemici niet voldoende getraind in statistiek en in de terminologie, zodat een klassiek geconstrueerd pakket niet geschikt is.

Om die reden is het nodig specifieke software te produceren. Naast de voorgestelde richtlijnen is software dus één van de eindproducten van het project. Er werden programma's ontwikkeld voor de validatie van bioanalytische methodes, voor het opstellen en interpreteren van robuustheidstesten, voor het vergelijken van methodes en voor multivariate calibratie. Omdat het maken van professionele, gebruiksvriendelijke en gevalideerde software moeilijk en kostbaar is, werd geopteerd voor het ontwikkelen van demonstratiesoftware.

Het is de bedoeling om in de toekomst enerzijds de door ons opgestelde richtlijnen te laten officialiseren en anderzijds om ze te promoten via verspreiding langs het Internet. De richtlijnen opgesteld i.v.m. robuustheidstesten zullen b.v. voorgelegd worden aan de ICH teneinde ze te laten officialiseren terwijl andere aan ISO zullen voorgesteld worden. De opgestelde richtlijnen zullen eveneens beschikbaar gesteld worden door ze op het Internet te publiceren, in combinatie met tutorials, praktische voorbeelden en de eventueel reeds beschikbare software. Dit moet de verspreiding van de in het project opgestelde richtlijnen bevorderen.

5. Synthese van het onderzoek

5.1. Inleiding

Overeenstemming tussen meetresultaten en hun onderlinge erkenning voor het nagaan van het naleven van de productspecificaties is een probleem dat belangrijk is voor iedereen die betrokken is in een consument-producent relatie. Dit is zowel een probleem voor de industrie als voor nationale en internationale beleidsorganen. Een dergelijke overeenkomst vereist dat de meetmethoden gevalideerd zijn, dit wil zeggen dat aangetoond moet worden dat hun precisie en andere performantiekarakteristieken voldoen aan de eisen van alle belanghebbende partijen. Er is een nood aan afspraken over de wijze waarop deze validatie moet uitgevoerd worden, of met andere woorden : men heeft standaarden en richtlijnen nodig voor de validatie van meetmethoden.

De technische en wetenschappelijke doelstelling van het project was om gestandaardiseerde benaderingen en richtlijnen te ontwikkelen voor de validatie van meetmethoden in chemische, farmaceutische, agro-voeding en aanverwante industrieën. Op organisatorisch vlak bracht dit project de bekwaamheid van een aantal Belgische onderzoekers en een aantal industrieën op het gebied van methodenvalidatie samen, teneinde de Belgische wetenschappers een prominente rol te laten spelen in de ontwikkeling van normen en richtlijnen.

Methodenvalidatie is dus een belangrijk regulatory probleem voor vele industrieën en is nodig om de validiteit aan te tonen van de methoden die aangewend worden om product specificaties na te gaan. Bovendien is methodenvalidatie nodig voor de bescherming van de veiligheid van de consument en om de kwaliteit van consumptieproducten te garanderen.

De ontwikkeling van snelle, multivariate technieken, zoals o.a. de nabij-infrarood analyse, voor het meten van de kwaliteit tijdens of in onmiddellijke relatie met een proces, is een belangrijke trend in de industrie. Pre-normatief onderzoek om guidelines te ontwikkelen voor de procesanalyse was gepland en werd uiteindelijk uitgevoerd behandeld.

5.2. Methodologie

Binnen het project werden verschillende taken uitgevoerd en verschillende aspecten van de methodenvalidatie bestudeerd. De verschillende onderdelen waarvoor getracht werd richtlijnen te formuleren of bestaande richtlijnen te evalueren, zijn : a) de evaluatie van de robuustheid van analytische methodes, b) de validatie van chromatografische bioanalysemethodes, d.w.z. methodes die bepalingen uitvoeren in biologische milieus, c) bepaling van de detectielimiet, d) de vergelijking van alternatieve meetmethodes, e) methodenvalidatiedesigns die kostenefficiënt zijn, met name wat betreft de lineariteit van ijklijnen, f) bepaling van de reproduceerbaarheid van een methode, g) robuuste procesanalyse met multivariate calibratie.

Voor de verschillende bovenvermelde onderdelen was de aanpak meestal gelijklopend. In eerste instantie werd de reeds bestaande situatie, d.w.z. de uit de literatuur voorhanden zijnde gegevens, dieper onderzocht. In een tweede stap werden enkele specifieke applicatiegebieden geselecteerd en werden hierin een aantal case studies uitgevoerd. Vervolgens werd tenslotte geprobeerd om een algemene benadering en gefinaliseerde standaardprocedures te definiëren.