

Näytteenoton virhelähteet, luotettavuuden estimointi ja näytteenottoketjun optimointi

Sisällysluettelo

1	Johdanto	17
2	Näytteenottosuunnitelma.....	17
3	Näytteenottokehteiden jaottelu	19
4	Näytteenoton virhelähteet ja virhekomponenttien ominaisuudet.....	20
4.1	Painotusvirhe.....	22
4.2	Näytteen materialisointivirheen, <i>ME</i> , komponentit: rajaamis- ja erottamisvirhe sekä preparointivirhe.....	22
4.3	Integrointivirhe, <i>IE</i>	25
4.3.1	Perusvirheen, <i>FE</i> , estimointi aineominaisuuksien avulla	26
4.3.1.1	Sovellusesimerkkejä.....	28
4.3.2	Ryhmittymis- ja lajittumisvirheen, <i>GSE</i> , estimointi	30
4.3.3	Pitkäaikaisen vaihtelun aiheuttamien virheiden, <i>IE₂</i> ja <i>IE₃</i> estimointi.....	31
4.3.3.2	Esimerkkejä variogrammin käytöstä	32
5	Näytteenotto- ja analyysiketjun optimointi (Sommerin [8] mukaan).....	36
5.1	Osajoukot eri suuria (Ositetun otannan teoriaa).....	36
5.1.1	Erän keskiarvon varianssin estimointi (yleisesti)	36
5.1.2	Näytteiden ositus, kun ainoastaan osien suhteelliset koot tunnetaan	37
5.1.3	Sekä osien suhteelliset koot että osien sisäiset varianssit, s^2 , tunnetaan	37
5.2	Näytteenotto- ja analyysiketjun optimointi, kun osat ovat samansuuruisia.....	38
5.2.1	Perusyhtälöt	39
5.2.2	Tarvittavien varianssikomponenttien estimointi	39
5.2.3	Ketjun optimointi	40
5.2.4	Joitakin käytännön ohjeita optimoinnin suorittamiseksi	41
6	Kirjallisuusviitteet	43

Esimerkit

1.	Vuotuisen rikki-taseen määrittäminen sellutehtaalle	45
2.	Esimerkki näytteenottovirheen (integroidivirheen) laskemisesta näytevälin funktiona käyttäen P. Gyn kehittämää variografista tekniikkaa	48
3.	Esimerkki puhdistetun jäteveden rikkipitoisuuden määrittämisen epävarmuuden estimoinnista	55
4.	Funktioita (Mathcad), joita voidaan käyttää P. Gyn teorian mukaisen perushajonnan laskemisessa	61
5.	Esimerkkejä ositetun otannan käytöstä	69
6.	Eri vaihtoehtojen tutkiminen haarautuvassa monivaiheisessa näytteenottoketjussa	74

1 Johdanto

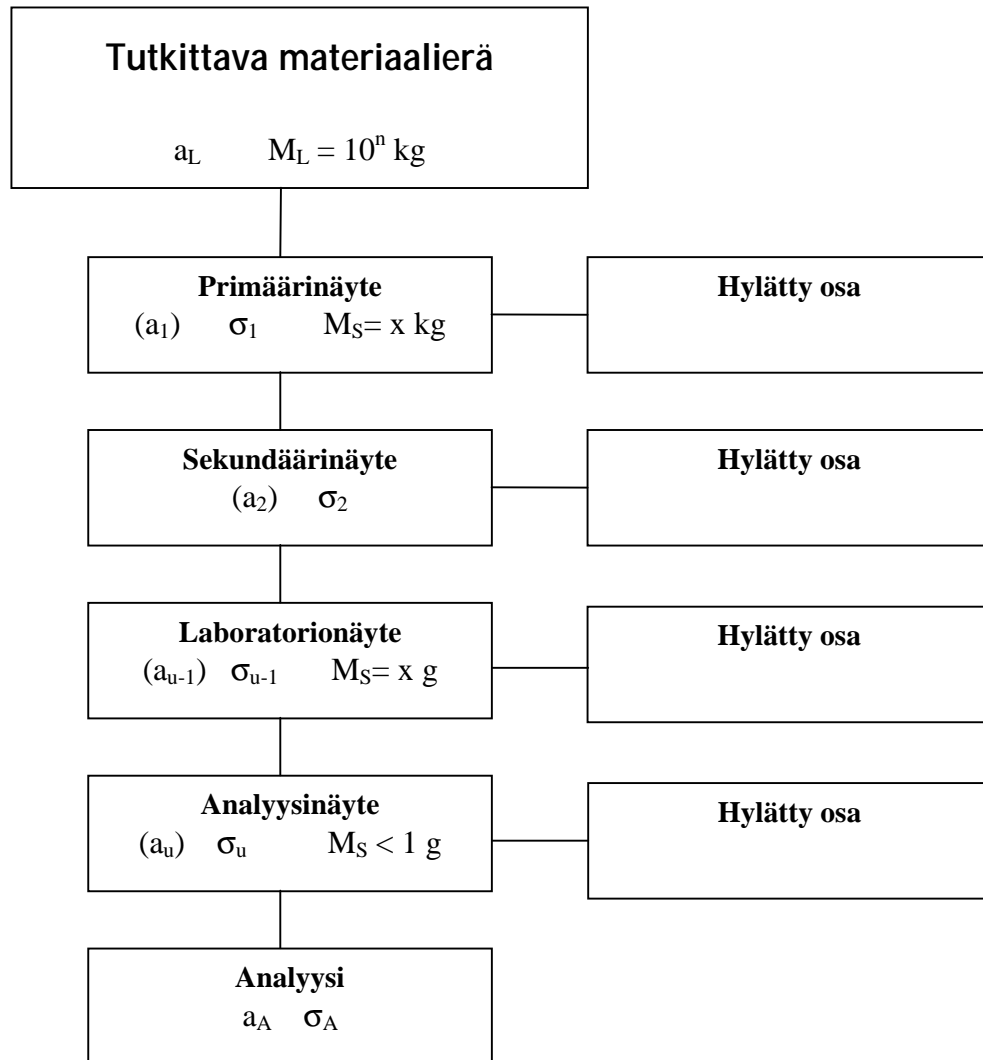
Prosessi- ja kemiantekniikassa tavallisin tapa saada kvantitatiivista tietoa tutkittavasta systeemistä on analysoida siitä otettuja näytteitä. Näytteenotto liittyy siten oleellisena ja välttämättömänä osana kaikkiin analyttisiin määrittämyksiin myös jatkuvatoimisiin prosessimittauksiin silloin, kun anturi "näkee" vain osan tutkittavasta virrasta. Analyttinen prosessi, johon usein liittyy monia näytteen-ottovaiheita (kuva 1), on luonteeltaan virhettä generoiva prosessi - jokainen vaihe vaikuttaa osaltaan lopputuloksen luotettavuuteen. Tutkittava materiaali, josta tietoa tarvitaan, voi olla kooltaan kymmeniä tonneja. Nykyisen instrumentaalianalytiikan vaatima näytekoko on tyypillisesti enintään gramman suuruusluokkaa. Tehtävä, jonka mukaan lopullisen analyysinäytteen tulisi olla mitattujen ominaisuuksien suhteen alkuperäisen erän pienoiskuva, ei olekaan triviaali.

Monivaiheisessa toimintaketjussa toisistaan riippumattomien näytteenotto- ja käsittelyvaiheiden sekä varsinaisen analyttisen määrittämyksen varianssit ovat additiivisia. Tästä seuraa, että käytännössä ketjun heikoin lenkki - työvaihe, jonka keskihajonta on selvästi muita vaiheita suurempi - määrää lopputuloksen luotettavuuden. Tätä valaisee seuraava esimerkki: Oletetaan, että näytteenoton suhteellinen keskihajonta on 5 % ja analyttisen määrittämyksen suhteellinen keskihajonta 2 %. Tällöin lopputuloksen suhteellinen keskihajonta on 5,4 %. Lopputuloksen luotettavuus ei parane oleellisesti määrittämenetelmää kehittämällä. Tuloksen luotettavuutta voidaan tässä tapauksessa parantaa vain kehittämällä näytteenotto- menetelmää.

Näytteenotossa on tilanteita, joihin ei löydy hyvää ratkaisua, aina kuitenkin ratkaisu, jolla päästään resursseihin nähden parhaaseen mahdolliseen tulokseen. Näytteenoton teoriaa on kehittänyt erityisesti Pierre Gy [1-4]. Gyn teoriaan perustuu myös valmisteilla oleva näytteenottoa käsittelevä ISO:n standardi [5].

2 Näytteenottosuunnitelma

Koska näytteenotto on ratkaiseva vaihe analyttisessä määrittämyketjussa, eikä siinä tehtyjä virheitä voi kompensoida myöhemmissä työvaiheissa, on se suunniteltava huolellisesti. Ryhmätyö, jossa on edustettuna sekä prosessitekniinen että analyttinen asiantuntemus takaa yleensä parhaan tuloksen. Näytteenottajien koulutusta tehtävänsä ei pidä unohtaa. Käytännössä ihan liian tavallinen tilanne, missä vähiten koulutusta saaneelle työryhmän jäsenelle annetaan lappi kouraan ja ohje käydä hakemassa näyte tutkittavaksi, voi aiheuttaa arvaamattomia ja hyvin kalliita seurauksia. Vain tehtävänsä merkityksen tuntevalta henkilöltä voi odottaa huolellisuutta, jota näytteenotossa tarvitaan.



Tavoite: $a_A = a_L$

Analyysituloksen kokonaiskeskihajonta:

$$\sigma_T(a_A) = \sqrt{\sum_{i=1}^u \sigma_i^2 + \sigma_A^2}$$

Kuva 1

Analyttinen prosessi on lähes aina monivaiheinen, virhettä generoiva prosessi, johon voi kuulua useita näytteenotto- ja käsittelyvaiheita.

Ennen yksityiskohtaisen suunnitelman laatimista tulisi saada vastaus moniin mitä?, miksi? milloin? -tyyppisiin kysymyksiin. Esimerkiksi:

- Mitä näytteistä halutaan tutkia?
- Millainen tieto halutaan estimoida?
 - Keskiarvo tietyltä ajanjaksolta (tunti, vuorokausi, viikko, ...), tai tietystä tilasta (työympäristö, varasto, raaka-ainelähetys, tuote-erä)
 - Pitoisuuden tai ominaisuuden jakautuma tutkittavassa tilassa
 - Korkein tai matalin pitoisuus.
- Onko aikaisemmista tutkimuksista käytävissä sellaista tietoa (esim. yksikkökustannukset, eri työvaiheiden varianssit), joita voidaan käyttää suunnittelun pohjana?
- Onko tarvittava välineistö saatavilla?
- Kuinka paljon näytteenotto saa maksaa?

Vasta kun yllä oleviin kysymyksiin on saatu vastaus, voidaan valita tekniikka (käsini. automaattisesti) ja optimaalinen strategia, jossa on ratkaistu mm. seuraavat seikat:

- Näytetiheys
- Näytteiden koko
- Näytteenottokohdat
- Erillis- vai koontanäytteet
- Otantatyypit
 - Satunnaisotanta
 - Systemaattinen ositettu otanta
 - Satunnainen ositettu otanta.

Korostettakoon vielä, että valittava näytteenoton strategia riippuu siitä, millaista tietoa tarvitaan. Esim. erillisnäytteiden yhdistäminen ja tutkiminen yhtenä näytteenä soveltuu keskiarvon määrittämiseen, mutta ei sen sijaan vaikkapa nopeahkojen prosessimuutosten vaikutuksen arviointiin.

3 Näytteenottokehteiden jaottelu

Valittava näytteenoton tekniikka, myös näytteenoton vaikeus, riippuu kohteen laadusta, jotka voidaan jaotella esim. dimensioiden mukaan seuraavasti:

1. Kolmidimensioiset
Kasat, siilot, vaunut, suuret kappaleet, huone- tai ulkoilman epäpuhtaudet
2. Kaksidimensioiset
Paperiradat, metallilevyt, muovikalvot, yleensä sellaiset kohteet, joista voidaan saada näyte, joka ulottuu koko tutkittavan kerroksen läpi
3. Yksidimensioiset
Langat, tangot, kuljetinlinjoilla kulkeva materiaali, jos niistä voidaan leikata täydellinen poikkileikkaus näytteeksi

4. Nolladimensioiset

Tuotteet, joista kokonaiset yksiköt voidaan ottaa näytteeksi (esim. pakatut tuotteet, kuten säkit, lääkepurkit ym.)

Jos ajallinen vaihtelu on otettava huomioon, kasvaa jokaisen kohteen dimensio vastaavasti yhdellä.

Näytteenottokohteet voidaan jaotella myös materiaalin homogeenisuusasteen perusteella seuraavasti:

A. Homogeeniset: Ominaisuus on vakiotasolla koko kohteessa

- Täysin sekoittuneet fluidit
- Puhtaat metallit
- Puhtaat kiteet

B. Heterogeeniset: jakautuu kahteen alaryhmään

1. Muutokset diskreettejä

- Karkearakeiset materiaalit (malmipelletit, kiviäytteet ym.)
- Suspensiot
- Toisiinsa sekoittumattomien nesteiden seokset

2. Muutokset jatkuvia

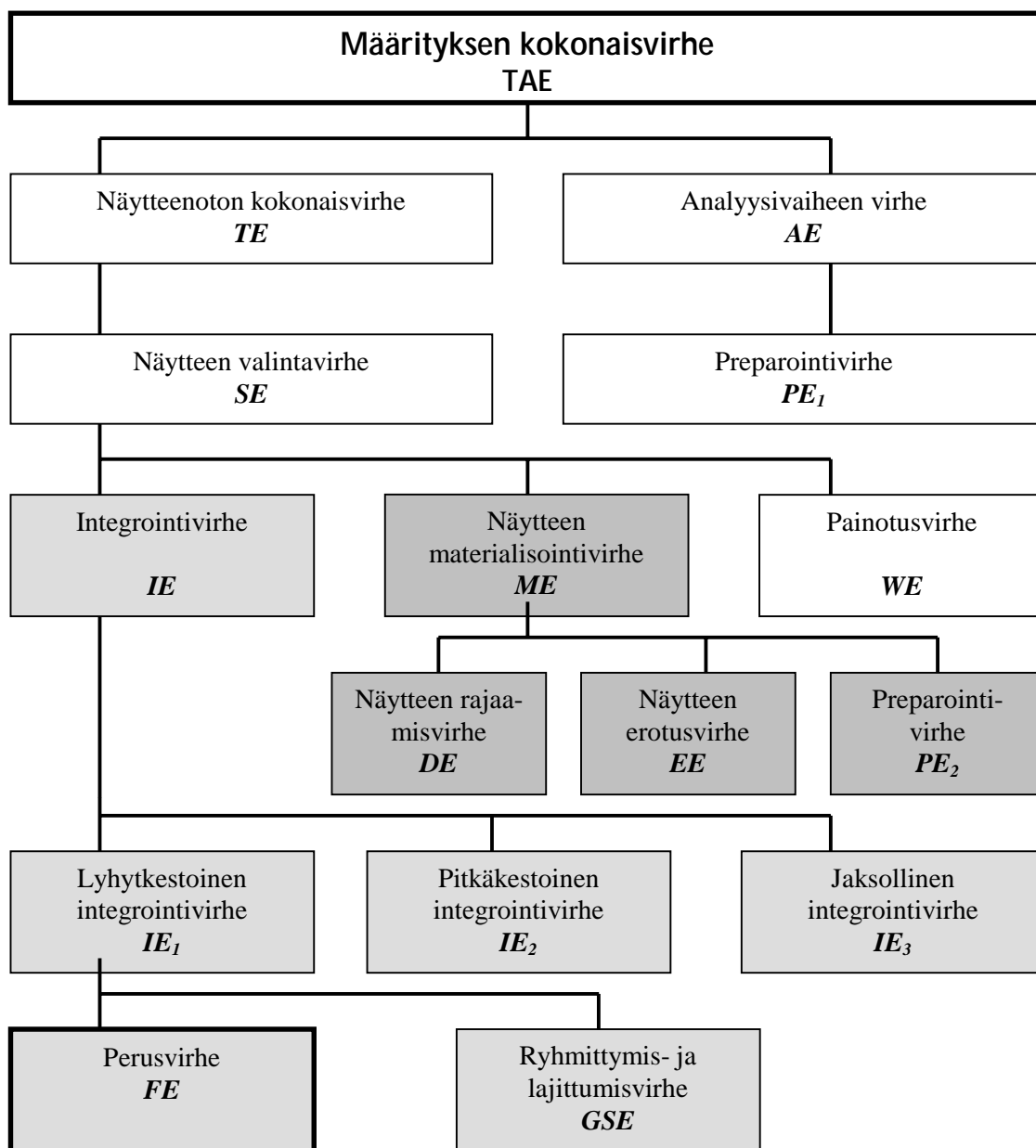
- Fluidit, joissa pitoisuus muuttuu tasaisesti
- Reaktioseokset
- Jauhemaiset materiaalit, kun partikkelikoko on hyvin paljon näytekokoaa pienempi

P. Gyn esittämä heterogeenisuuden määritelmä käsitellään myöhemmin.

4 Näytteenoton virhelähteet ja virhekomponenttien ominaisuudet

Pierre Gy [1,2] jakaa analyttisen määrittelyn virheet niiden syiden mukaan kolmeen pääluokkaan, jotka ovat: *Näytteenoton kokonaisvirhe (TE)*, varsinaisen *analyysivaiheen virheet (AE)* ja *painotusvirhe (WE)*.

Prosessinäytteenotossa eräs virhelähde on virtausmäärän vaihtelu tai epätarkka mittaus, joka aiheuttaa väärän painotuksen laskettaessa keskimääräistä massavirtausta tai tutkittavan komponentin kokonaismassaa. Tästä aiheutuvaa virhettä Gy kutsuu *painotusvirheeksi*. *Analyttinen virhe* liittyy varsinaiseen määrittelyvaiheeseen, eli niihin työvaiheisiin, jotka seuraavat lopullisen analyysinäytteen valmistumista sisältäen mm. kalibroinnin ja instrumentin kohinan tuloksiin aiheuttaman virheen. *Näytteenoton kokonaisvirhe (TE)* aiheutuu kahdesta päälähteestä, jotka ovat: *Näytteenottokohteen heterogeenisuus* (aiheuttaa *näytteen valintavirheen, SE*), sekä *preparointivirheet (PE)*, joiksi luetaan kaikki näytteen käsittelystä aiheutuvat virheet (kuten kontaminaatio, määritettävän aineen häviöt haihtumisen tai absorption vuoksi, ym.) ja joita voi esiintyä missä tahansa näytteen käsittelyvaiheessa.



$$TAE = TE + AE + PE + WE$$

$$= \underbrace{FE + GSE + IE_2 + IE_3}_{\text{Integroitivirhe, IE}} + \underbrace{DE + EE + PE_2}_{\text{Materialisointivirhe, ME}} + PE_1 + AE + WE$$

Kuva 2
 Analyttisen prosessin kokonaisvirheen varianssin komponenttien jaottelu Pierre Gyn mukaan, kun tavoitteena on prosessivirran keskipitoisuuden määrittäminen.

Kohteen heterogeenisuus puolestaan aiheuttaa sen, että diskreetti näytteenotto aiheuttaa väistämättä keskiarvoa laskettaessa tuloksiin virheen, jota voidaan pitää *integroitinvirheenä (IE)*, ja että näytteenoton tekniikalta ja siihen käytetyiltä laitteilta vaaditaan tiettyjä ominaisuuksia, joista poikkeaminen aiheuttaa näytteen *materialisointivirheen (ME)*, jonka komponentteja preparointivirheen ohella voivat olla näytteen *rajaamisvirhe (DE)* ja näytteen *erottamisvirhe (EE)*. Virheiden jaottelu ja näytteenoton *kokonaisvirheen, TE*, komponentit on esitetty kuvassa 2. Koska varianssit ovat additiivisia, on kokonaisvirheen varianssi osavirheiden varianssien summa.

4.1 Painotusvirhe

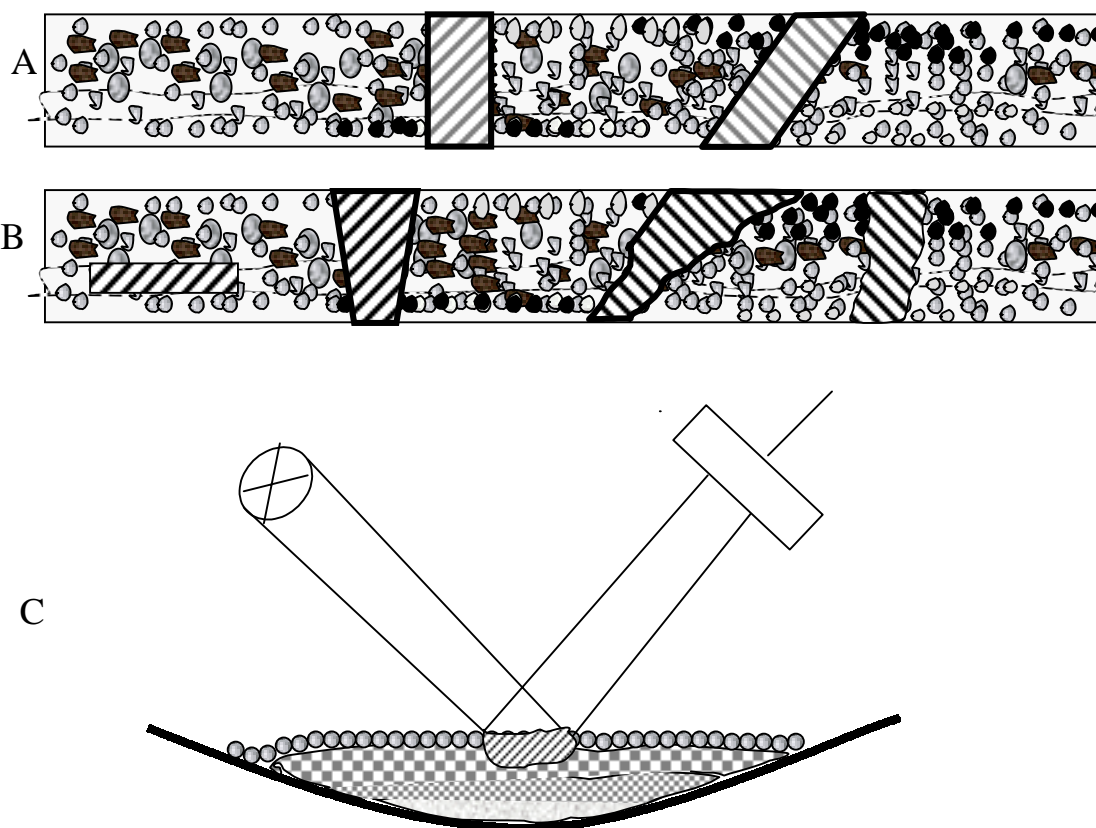
Jos näytteiden avulla on tarkoitus selvittää materiaalierän keskipitoisuus, tai määritettävän komponentin massavirtaus tai kokonaisuudessa, tulisi tulokset painottaa virtausmäärillä, erityisesti jos pitoisuustason vaihtelu riippuu virtauksen vaihtelusta, kuten on tavallista prosessiolosuhteissa. Ellei kokonaisvirtausta mitata, tai sitä ei pystytä mittaamaan tarkasti, aiheuttaa prosessivirtauksen vaihtelu tulokseen virheen, jota Gy [1] nimittää *painotusvirheeksi*. Virhettä voidaan pienentää tasoittamalla virtausta näytteenoton ajaksi ja lisäämällä näytteiden lukua. On huomattava, että tämä virhe liittyy vain keskiarvon määrityksiin. Jos tarkoituksena on estimoida pitoisuuden piste-estimaatti (suo., konsentraatio tietyllä ajanhetkellä tietyssä kohdassa), ei virhettä esiinny. Se tulee mukaan vasta, kun diskreetit tulokset halutaan yleistää koskemaan pidempää ajanjaksoa.

4.2 Näytteen materialisointivirheen, *ME*, komponentit: rajaamis- ja erottamisvirhe sekä preparointivirhe

Näytteen materialisointivirhe (P. Gy: *Increment Materialization Error, ME*) liittyy näytteen erottamiseen tutkittavasta kohteesta ja se voi olla kolmen erilaisen virheen, *DE, EE ja PE₂*, summa. Näistä näytteen *rajaamisvirhe, (DE, increment delimitation error)* liittyy näytteen muotoon ja on puhtaasti geometrinen operaatio. Otetaan esimerkki näytteenotosta kuljetinhihnalta. Kuljetuksen aikana pyrkii materiaali lajittumaan, joten vain siinä tapauksessa, että näyte muodostaa tasapaksun viipaleen koko materiaalivirran poikkileikkauksesta, voi näyte olla edustava (ja *DE = 0*). Tätä valottaa kuva 3. Esim. jatkuvatoimisten prosessianalysointien ongelma on, että ne useimmiten saavat informaatiota vain prosessivirran pintaosasta. Näytteen *erottamisvirhe (EE, increment extraction error)* syntyy, vaikka näytteen profiili olisi periaatteessa oikea, jos näytteenotin syrjii tai suosii jotakin materiaalin partikkelilajia. *Näytteen preparointivirhe (PE)* voi liittyä näytteenkäsittelyketjun jokaiseen vaiheeseen ja sisältää näytteen muuttumisen esim. kontaminaation tai haihtumisen vaikutuksesta. Tähän kohtaan liittyvää preparointivirhettä on kuvassa 2 merkitty alaviitteellä 2.

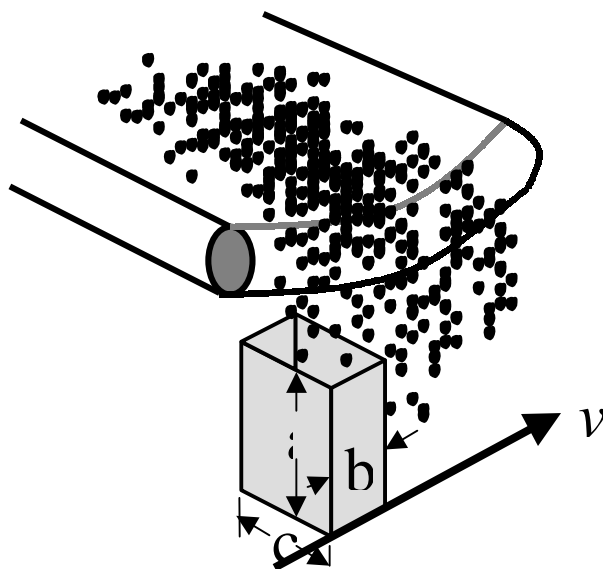
Helpoimmin materiaalivirroista saadaan oikeanmuotoinen näyte sellaisissa kohdissa prosessia, missä materiaali on vapaassa putoamisliikkeessä. Paikallaan olevasta materiaalista, kuten kasoista, siiloista, junanvaunuista ja autokuormista on oikean näytteen rajaaminen käytännössä mahdotonta eikä niistä otettuja näytteitä voi pitää tilastomatematisissa mielessä edustavina. Tällaista näytteenottoa tulisi välttää, erityisesti silloin, kun tulosten perusteella määrätään tuotteen hinta. Vaikka näytteen profiili olisi oikea, täytyy näyte saada erotetuksi siten, että kaikilla osalajeilla on niiden

lukumäärää vastaava todennäköisyys joutua näytteeseen. Tämä varmistetaan näytteenottimen oikealla mitoituksella. Kuva 4 esittää niin sanottua traversoivaa näytteenotinta, jota voidaan nimittää myös lineaariottimeksi. Vain oikeanmuotoinen poikkileikkaus materiaalivirrasta varmistaa sen, että $DE \approx EE \approx 0$ ja että materiaalivirtaa voidaan käsitellä 1-dimensioisena näytteenoton kohteena. Jotta näytteen profiili olisi oikea, tulee keräimen leikkuaukon reunojen olla yhdensuuntaiset ja keräimen nopeuden v vakio koko näytteen keruun ajan. Jos b_0 on kapein hyväksyttävä suuaukon leveys ja d suurimpien partikkeleiden halkaisija voidaan näytteenotin mitoittaa seuraavasti:



Kuva 3

Jos materiaalivirrassa on lajittumista, niin ainoastaan näytteessä, joka muodostaa täydellisen tasapaksun viipaleen prosessivirrasta (A ja B), voivat materiaalivirran kaikki osat olla tasapuolisesti edustettuina. Näytteen rajaamisvirhe on siten potentiaalinen virhelähde esim. käytettäessä 'on-line'-prosessianalysaattoreita, jotka useimmiten saavat informaatiota vain suhteellisen ohuesta materiaalikerroksesta (C).



Kuva 4
Traversoiva näytteenotin I. lineaariotin

- a) $v \leq 0.6$ m/s
jos $d > 3$ mm niin $b \geq 3d = b_0$
jos $d < 3$ mm niin $b \geq 10$ mm = b_0 (aukko tukkeutuu helposti, jos se on tätä kapeampi)
- b) Jos käytetään suurempaa aukkoa kuin b_0 , niin näytteen leikkausnopeutta voidaan nostaa, ts. jos $b = n b_0$ ja
 $n > 1$ niin
 $v \leq v_{max} = (1+n) 0.3$ m/s

Jos käytetään pienempää aukkoa tai suurempaa leikkausnopeutta, kuin edellä olevissa yhtälöissä on esitetty, aiheuttaa reunaefekti sen, että isoimmat partikkelit alkavat syrjäytyä näytteestä, ts.. näytteen *erotusvirhe*, *EE*, alkaa kasvaa, vaikka näytteen profiili olisikin oikea.

Näytteenkeruulaitteiden toimintaa arvioitaessa tulisi aina ensin selvittää, millainen on materiaalivirrasta leikatun näytteen profiili ja toiseksi noudattaako laite yllä olevia mitoitusyhtälöitä. Ottimen pituuden, c , tulee olla selvästi suurempi kuin materiaalivirran paksuus ja syvyyden a niin suuri, ettei astia täyty keräysjakson aikana.

Näytteenoton kannalta vaikean ongelman muodostaa pölynäytteen otto virtauskanavista. Tällöin on yleensä mahdotonta saada täydellistä poikkileikkausta virtauksesta näytteeksi. Koska kaasun ja kiinteän aineen tiheysero on suuri, pyrkii kiintoainne gravitaation ja keskipakovoimien ansiosta muodostamaan epätasaisen jakautuman. Jos näyte joudutaan imemään yhdestä tai muutamasta kohdasta kanavaa, eikä pitoisuusjakautumaa saada esim. staattisilla sekoittimilla tasatuksi ennen näytteenottoa, aiheutuu tästä vaikeasti arvioitava näytteen *rajaamisvirhe*. Näytteen *erottamisvirheenä* voidaan tässä yhteydessä pitää virhettä, joka aiheutuu siitä, jos näytteen imunopeus poikkeaa isokineettisestä. Jos imunopeus on suurempi kuin virtausnopeus kanavassa, imee sondi kaasua ja pölyä myös suuaukkonsa ulkopuolelta. Hienoin aines tulee tällöin ylijedustetuksi näytteeseen, koska se pystyy seuraamaan kaasuvirtausta helpommin kuin

raskaammat partikkelit. Toisaalta taas, jos immunopeus on pienempi kuin virtausnopeus kanavassa, karkaa osa hienommasta aineksesta sondin suuaukon ohi ja tällöinkin näyte on virheellinen. Pölypäästöjen mittauksista antaa ohjeita kotimainen standardi [8].

Näytteen materialisointivirheen komponenttien suuruutta ei voida normaalisti estimoida teoreettisesti eikä niitä kannata myöskään yrittää estimoida kokeellisesti - eikä se ole tarpeenkaan, koska ne voidaan käytännössä minimoida, ja teoriassa eliminoida, suunnittelemalla näytteenottoon käytetyt laitteet oikein.

4.3 Integrintivirhe, IE

Näytteenottokohteen *heterogeenisuus* l. mitattavan ominaisuuden vaihtelu ajan tai paikan funktiona näytteenottoon perustuvassa pitoisuuden arvioinnissa aiheuttaa keskiarvoa laskettaessa virheen, jota voidaan pitää integrintivirheenä, koska virhe on luonteeltaan samanlainen, mikä aiheutuu integroitaessa jatkuvaa funktiota numeerisesti. Näytteenoton kohteen heterogeenisuudella voi olla kaksi eri syytä: se voi johtua materiaalin luonteesta, kuten silloin, kun materiaali on koostunut erilaisista partikkeleista, tai sitten mitattavan ominaisuuden jakautuman heterogeenisuudesta, jonka syynä voi olla esim. raaka-aineen tai prosessin vaihtelut materiaalityypistä riippumatta. Vastaavasti integrintivirhe voi olla kolmen erityyppisen virheen summa: $IE = IE_1 + IE_2 + IE_3$.

IE_1 johtuu lyhytkestoisesta vaihtelusta (kun näytteen kokoa tai näytteen keruu-aikaa käytetään mittana) ja sillä on kaksi komponenttia: *Perusvirhe*, *FE* ja *ryhmittymis- ja lajittumisvirhe* (tai *ryhmittymis- ja segregoitumisvirhe*), *GSE*. Perusvirhe on materiaalin partikkeliluonteesta johtuva puhdas tilastollinen virhe, ts.. se virhe, joka jää jäljelle, jos kaikki muut virhelähteet on saatu eliminoiduksi. Sen suuruus riippuu näytteen koosta ja materiaalin ominaisuuksista ja se on ainoa näytteenoton virheistä, jota ei edes teoreettisesti voida koskaan kokonaan eliminoida. Muut virheet voidaan ainakin teoriassa, joskaan ei käytännössä, eliminoida täysin, jos näytteenoton kohde voidaan homogenisoida (l. sekoittaa ideaalisti) ennen näytteenottoa ja käytetään moitteetonta näytteenoton tekniikkaa (jolloin $ME \equiv 0$). Perusvirhe määrää siten sen teoreettisen minimivirheen, johon tiettyä toimenpideketjua käyttäen voidaan päästä. Sen arviointi on perustana, kun tutkitaan erilaisia näytteenkäsittelyvaihtoehtoja. Lisäksi perusvirhe, toisin kuin muut näytteenoton virheet, voidaan estimoida teoreettisesti materiaalin ominaisuuksista, jos ne tunnetaan riittävän tarkasti. *Lajittumisvirhe* puolestaan tarkoittaa materiaalin lajittumisen aiheuttamaa virhettä; esim. kuljetinhihnalla jakautuman epätasaisuutta poikkileikkauksessa ja sen näytteenottoon aiheuttamaa virhettä.

Integrintivirheen komponentti IE_2 aiheutuu prosessin pitkäaikaisesta, ei-jaksollisesta, vaihtelusta ja IE_3 jaksollisesta vaihtelusta ja niiden vaikutus näytteenoton virheeseen voidaan estimoida kokeellisesti. Näiden virheiden suuruus riippuu lisäksi otantatyyppistä ja on erilainen satunnaisotannalle, ositetulle satunnaisotannalle ja systemaattiselle l. tasaväliselle otannalle. Jos näytteenoton tarkoituksena on saada ominaisuuden piste-estimaatti, tarvitsee ottaa huomioon vain lyhytaikainen vaihtelu.

4.3.1 Perusvirheen, FE, estimointi aineominaisuuksien avulla

Näytteenoton perusvirheen suhteellinen keskihajonta, σ_r , voidaan estimoida seuraavien kaavojen avulla:

$$\sigma_r^2 = HI \left(\frac{1}{M_S} - \frac{1}{M_L} \right) \quad (1a)$$

$$\approx \frac{HI}{M_S}, \text{ jos } M_S \ll M_L \quad (1b)$$

Kaavoissa M_S = näytteen koko, M_L = sen materiaalierän koko, josta näyte otetaan, HI (=Heterogeneity Invariant) = materiaalista riippuva näytevakio, josta Gy vanhemmissa julkaisuissaan on käyttä merkintää Z .

$\sigma_r = \sigma/a_L$, missä σ on pitoisuuden keskihajonta ja a_L keskipitoisuus tutkittavassa materiaalierässä.

Yhtälöitä (1) voidaan käyttää yleisesti materiaalityypistä riippumatta, kun halutaan estimoida puhtaan satunnaisvirheen riippuvuutta näytekoosta.

Kiinteän, partikkeleista koostuvan materiaalin näytteenoton suhteellinen keskihajonta saadaan arvioiduksi aineominaisuuksista seuraavasti:

$$\sigma_r^2 = Cd^3 \left(\frac{1}{M_S} - \frac{1}{M_L} \right) \quad (2a)$$

$$\approx \frac{Cd^3}{M_S}, \text{ jos } M_S \ll M_L \quad (2b)$$

C = materiaalista riippuva näytevakio ($Z = C d^3$) ja d on partikkelikokojakautuman 95 % yläraja (yleensä d :nä käytetään sen seulan aukon läpimittaa, jonka materiaalista n. 95% läpäisee). Tavallisesti voidaan käyttää yhtälöä (2b), koska lähes aina - jotkut näytteenjako-operaatiot ovat tästä poikkeuksena - paljon pienempi kuin materiaalierä, josta se otetaan.

Näytevakio C on neljän aineominaisuuksia kuvaavan parametrin tulo:

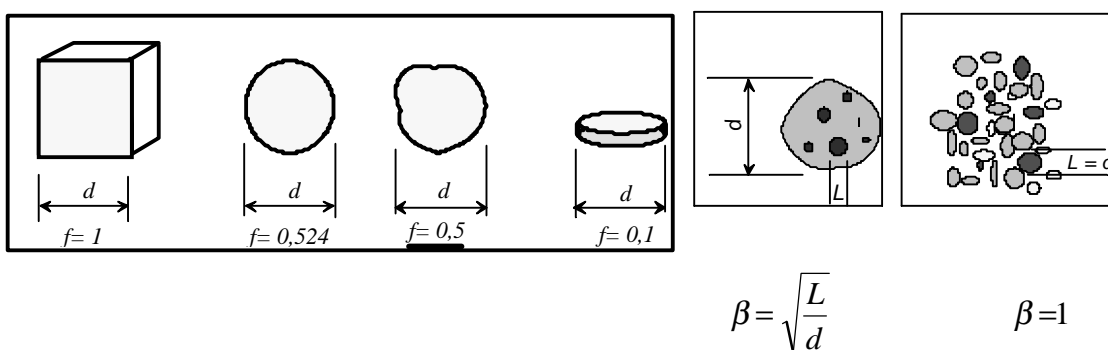
$$C = fg\beta c \quad (3)$$

Parametreistä f on raemuototekijä, joka kuutiolle on 1; jauhatuksessa useimmat materiaalit pyöristyvät niin, että niille voidaan käyttää keskimääräistä arvoa $f=0,5$. g on raekokojakautumatekijä, joka ottaa huomioon raekokojakauman laajuuden vaikutuksen näytteenoton virheeseen. Tasarakeiselle materiaalille se on 1 ja jos jakautuman 95 %:n ylärajan ja 5 %:n alarajan suhde $d/d_{0,05}$ on yli 4, voidaan käyttää oletusarvona $g = 0,25$. Normaalissa jauhatuksessa ja murskauksessa saadaan tavallisesti tämä arvo. Ääriarvojen väli voidaan estimoida seuraavien arvojen avulla:

Laaja raekokojakautuma ($d/d_{0,05} > 4$) tavallisin arvo $g = 0,25$
 Keskikokoinen raekokojakautuma ($d/d_{0,05} = 4...2$) $g = 0,50$

Kapea raekokojakautuma ($1 < d/d_{0,05} < 2$) $g = 0,75$
 Tasarakeinen materiaali ($d/d_{0,05} = 1$) $g = 1,00$

Puhtaaksijauhautumistekijä $\beta = 1$, jos matriisi ja määritettävän aineen sisältävä partikkelilaji ovat materiaalissa erillisinä partikkeleina. Jos taas partikkelit, jotka sisältävät pääosan määritettävästä aineesta, ovat sulkeumina matriisipartikkeleissa, arvioidaan β seuraavasti: $\beta = \sqrt{L/d}$, missä L on mielenkiintoisen partikkelilajin koko. β :n maksimiarvo on kuitenkin aina 1. Kuva 5 havainnollistaa f :n ja β :n estimointia.



Kuva 5
 Partikkelimuototekijän, f , ja puhtaaksijauhaantumistekijän, β , estimointi

Koostumustekijä c arvioidaan seuraavan kaavan avulla:

$$c = \frac{(1 - a_L / \alpha)^2}{a_L / \alpha} \rho_c + \left(1 - \frac{a_L}{\alpha}\right) \rho_m \quad (4)$$

missä a_L on keskimääräinen pitoisuus, α sen partikkelilajin pitoisuus, joka sisältää määritettävän aineen, ρ_c ja ρ_m ovat mielenkiintoisen partikkelilajin ja matriisin tiheydet. Joillekin materiaaleille C :n laskemiseksi tarvittavat materiaaliominaisuudet ovat suhteellisen helposti arvioitavissa. Monimutkaisille seoksille C ja HI voidaan estimoida kokeellisesti analysoimalla riittävän suuri määrä rinnakkaisia näytteitä. Tuloksista lasketaan suhteellinen keskihajonta, josta on vähennettävä puhtaan analyttisen virheen osuus, jos se on näytteenottovirheen rinnalla merkitsevä:

$$s_r^2 = s_t^2 - s_a^2 \quad (5)$$

s_t on näytteistä laskettu suhteellinen kokonaiskeskihajonta, s_r näytteenoton ja s_a analyysin keskihajonta. Kun s_r on saatu määritetyksi, voidaan näytevakio C estimoida yhtälön (2b) avulla. Jos näytteet eivät ole samansuuruisia, on kokonaiskeskihajonta laskettava näytekoolla painotetuista tuloksista.

Gy:n yhtälöitä voidaan käyttää seuraavantyyppisten ongelmien käsittelyyn:

- Mikä on näytteenotosta johtuva epävarmuus tietynsuuruista näytekoko, M_s , käytettäessä? (Yhtälöt 1 ja 2).
- Kuinka suuri näyte on otettava, jos näytteenotolta vaaditaan tietty luotettavuustaso, σ_r ?

$$M_s = \frac{Cd^3}{\sigma_r^2}$$

- Kuinka hienoksi materiaali on jauhettava, jos sekä näytekoko että luotettavuustaso ovat määrättyt?

$$d = \left(\frac{M_s \sigma_r^2}{fgc\sqrt{L}} \right)^{0,4}, \text{ tämä arvo hyväksytään jos } d \geq L$$

tai

$$d = \left(\frac{M_s \sigma_r^2}{fgc} \right)^{\frac{1}{3}}, \text{ jos } d < L.$$

- Monivaiheisten näytteenotto- ja käsittelyketjujen analysointi ja optimointi.

Saatavilla on myös Gy:n teoriaan perustuva tietokoneohjelma, SAMPEX [6,7], joka on kätevä apuväline näytteenotto-ongelmien ratkaisussa.

4.3.1.1 Sovellusesimerkkejä

Esimerkki 1

Testattaessa kuparipitoisen lannoitteen rakeistusprosessia sekoitettiin ammoniumfosfaattilannoitteeseen, tiheys 1,7 kg/dm³, kuparisulfaattia niin paljon, että lannoitteen Cu-pitoisuus oli keskimäärin 0,020 %. Seostamiseen käytetyn kuparisulfaatin raekoko (=puhtaaksijauhaantumiskoko, L) oli 0,2 mm, tiheys 2,28 kg/dm³ ja kuparipitoisuus 25 %. Lannoite on rakeistettiin 5...8 mm:n suuruisiksi rakeiksi. Sekoitusprosessin tehokkuuden testaamiseksi otettiin valmistuserästä 20 näytettä, jotka olivat kooltaan 5 kg. Näytteiden väliseksi suhteelliseksi keskihajonnaksi saatiin 22 %. Toimiiko sekoitusprosessi ideaalisti?

Ratkaisu:

Gy:n yhtälöön sijoitettavat lähtötiedot ovat tehtävän mukaan seuraavat: $a=25\%$, $a_L=0,02\%$, $\rho_c=2,28\text{ kg/dm}^3$, $\rho_m=1,7\text{ kg/dm}^3$, $f=0,5$ (keskim. arvo), $g=0,9$ (arvio), $L=0,2\text{ mm}$ ja $d=8\text{ mm}$. Näiden avulla voidaan laskea näytevaktion C tekijät: $\beta=0,16$, $c=2850\text{ kg/dm}^3$ ja edelleen $C=203\text{ kg/dm}^3$. Sijoittamalla C , d sekä $MS=5\text{ kg}$ yhtälöön (2b) saadaan 5 kg:n näytteen perushajonta, $\sigma_r=14,4\%$. Jos ainearvoja voidaan pitää riittävän tarkkoina voidaan nyt testata F-testin avulla, poikkeako havaittu hajonta 22 % merkitsevästi perushajonnasta:

$$F = (s_{hav} / \sigma_r)^2 = (22 / 14,4)^2 = 2,33$$

F-jakautuman 1 %:n taulukkoarvo vapausasteilla (19,∞) on 1,90, joten on aihetta olettaa, että tuotteessa on jonkin verran lajittumista, eikä sekoitus siten ilmeisestikään ole toiminut täysin ideaalisti. Ainearvojen avulla voidaan myös edelleen laskea, että 50 kg:n suuruisten säkkien välinen perushajonta olisi 4,6 %, mikä 95 %:n luottamustasolla antaisi säkkien väliseksi pitoisuusvaihteluksi 0,02 % ± 0,0018 % kuparia.

Esimerkki 2

Edellisen tehtävän 5 kg:n näytteestä halutaan valmistaa analyysinäyte kolmessa vaiheessa: aluksi näyte murskataan ja siitä otetaan 0,5 kg:n suuruinen näyte. Tämän partikkelikokoa pienennetään edelleen ja siitä otetaan 50 g:n laboratorionäyte. Laboratoriojauhatuksen jälkeen kuparipitoisuus määritetään 1 g:n näytteistä. Mihin hienouteen on jauhatus kussakin vaiheessa suoritettava, jos analyysin suhteellinen keskihajonta on 1 % ja määrittelyn kokonaishajonnaksi, jossa ei huomioida 5 kg:n primäärinäytteen hajontaa (16 %), halutaan 4 %?

Ratkaisu:

Edellisen tehtävän ainearvoista muuttuu kokojakautumatekijän g arvo, koska murskauksessa ja jauhatuksessa muodostuu laaja kokojakautuma. g :lle käytetään keskimääräistä arvoa 0,25. Tehtävän ratkaisu on alla kuvassa 6 esitetty SAMPEX-ohjelman tulostuksena. Ohjelma on laskenut jauhatusasteet siten, että kunkin näytteenottovaiheen keskihajonta on sama (2,24 %).

```

** SAMPEX ** FUNCTION 1 = INITIALIZATION, ENTER INITIAL DATA **
** COMPUTE SAMPLING CONSTANT C **
CONC. OF THE COMPONENT OF INTEREST IN PURE COMPOUND (alpha) = 25
AVERAGE CONCENTRATION OF THE COMPONENT OF INTEREST (a) = .02
DENSITY OF THE PURE COMPONENT (g/cm^3) = 2.28
DENSITY OF SAMPLE MATRIX (g/cm^3) = 1.7
PARTICLE SHAPE FACTOR f = .5
PARTICLE SIZE RANGE FACTOR g = .25
** SAMPEX ** FUNCTION 5 = MULTI-STAGE SAMPLING *****
                        = sampling in several steps *****
3. Particle sizes for predetermined sample sizes
ENTER THE LIBERATION SIZE L OF THE MATERIAL (mm) = .2
ENTER NUMBER OF SAMPLING STEPS N (1...8) = 3
ENTER SIZE OF THE LOT 'M' (g) = 5000

+----- THE OPTIMUM PARTICLE SIZES OF MULTI-STAGE SAMPLING -----+
|STEP 1: 1.25 mm          Sample size = 500 g          |
|STEP 2: 0.497 mm        Sample size = 50 g           |
|STEP 3: 0.113 mm        Sample size = 1 g            |
+-----+-----+-----+-----+-----+-----+-----+
RELATIVE STANDARD DEVIATION OF EACH SAMPLING STEP = 2.24 %
REL. STAND. DEV. OF ANALYSIS (No. of parallel determinations= 1) = 1 %
TOTAL RELATIVE STANDARD DEVIATION = 4 %
Do You wish to run again with different estimates for std:s (Y/N) = ?N
    
```

Kuva 6
 Esimerkin 2 ratkaisu SAMPEX-tietokoneohjelman tulostuksena

4.3.2 Ryhmittymis- ja lajittumisvirheen, GSE, estimointi

Gy määrittelee ryhmittymis- ja lajittumisvirheen seuraavasti:

$$GSE = IE_1 - FE \approx YZ \cdot FE \quad (6)$$

Y on ryhmittymistekijä ja se on verrannollinen näytteeksi otetun ryhmän sisältämien partikkelien lukuun, jos partikkeleita on näytteessä runsaasti, niin sitä voidaan approksimoida suoraan partikkeleiden luvulla. Z puolestaan on lajittumistekijä, jonka minimiarvo on 0 (täysin sekoitettu materiaali) ja maksimiarvo 1, kun näytteen komponentit ovat täysin erkaantuneet omiksi ryhmikseen. Käytännössä tämän virheen suuruutta ja vakioita Y ja Z ei yritetä estimoida, vaan se otetaan tarvittaessa huomioon varmuuskertoimena: $GSE_{\max} = 2 \cdot FE$.

4.3.3 Pitkäaikaisen vaihtelun aiheuttamien virheiden, IE_2 ja IE_3 estimointi

Kemiallisissa prosesseissa, joissa siirretään suuria ainemääriä, ovat prosessin aikavakiot yleensä suhteellisen pitkiä. Tästä johtuen eivät perättäin otetut näytteet ole toisistaan riippumattomia muuta kuin siinä tapauksessa, että näyteväli on erittäin lyhyt, jolloin prosessivaihtelu on pientä perushajontaan verrattuna, tai että näyteväli on hyvin pitkä prosessin aikavakioon verrattuna. Arvioitaessa prosessivirran keskiarvoa tietyltä ajanjaksolta, voidaan periaatteessa käyttää kolmea erilaista otantamenetelmää:

- 1) satunnaisotantaa
- 2) ositettua otantaa ja
- 3) systemaattista l. tasavälistä otantaa.

Jos perättäisten näytteiden välillä on korrelaatiota, niin keskiarvon luotettavuus riippuu otantamenetelmästä, ellei korrelaatiota ole, antavat kaikki kolme menetelmää saman luotettavuuden.

Satunnaisotannassa haluttu näytemäärä kerätään siten, että näytteenottoajat arvotaan esim. satunnaislukujen avulla. Ositetussa otannassa kohde aluksi jaetaan tasasuuriin osiin, joista kustakin otetaan satunnaisnäyte. Systemaattisessa otannassa puolestaan voidaan aloituskohta valita satunnaisesti, jonka jälkeen kohteesta kerätään tarvittava näytemäärä tasavälein. Jos prosessin keskihajonta on s_p , kohteesta otetaan n näytettä ja keskiarvon keskihajonta on $s_{\bar{x}}$, niin seuraava vertailu otantamenetelmien kesken on yleisesti voimassa:

- 1) Satunnaisotanta: $s_{\bar{x}} = \frac{s_p}{\sqrt{n}}$ (yleensä epäluotettavin tulos)
- 2) Ositettu otanta: $s_{\bar{x}} \leq \frac{s_p}{\sqrt{n}}$
- 3) Systemaattinen otanta: $s_{\bar{x}} \leq \frac{s_p}{\sqrt{n}}$ (yleensä luotettavin tulos, paitsi jaksollisissa prosesseissa, jos näytteenottoväli on jokin prosessin jakson harmoninen monikerta, jolloin tämä menetelmä on epäluotettavin ja voi aiheuttaa tuloksiin systemaattiseen virheeseen)

Jos prosessin jaksollisuudesta ei ole varmistuttu, niin menetelmä 2) on aina varmin vaihtoehto. Se on myös jaksollisissa prosesseissa turvallinen menetelmä, koska siinä näytteenottovirheen pitkäaikainen keskiarvo lähenee nollaa. P. Gy on osoittanut, että näytteenoton pitkäaikaisen ja jaksollisen vaihtelun aiheuttama virhe ($IE_2 + IE_3$) voidaan estimoida ns. variografisen kokeen avulla. Variografisessa kokeessa riittävä määrä näytteitä (yleensä vähintään 30, mieluummin huomattavasti enemmän) kerätään tasavälein. Gy määrittelee prosessin tai näytteenoton kohteen *heterogeenisuuden* mitattavan ominaisuuden suhteelliseksi vaihteluksi tutkittavan materiaalierän (tai prosessijakson) keskiarvon ympärillä. Tutkittujen variografisen kokeen näytteiden avulla *heterogeenisuus*, h_i , estimoidaan seuraavasti:

$$h_i = \frac{a_i - a_L}{a_L} \frac{M_i}{M} \quad i=1, 2, \dots, N \quad (7)$$

i on näytteen järjestysluku, a_i näytteen i analyysituloksena, a_L tutkittavan materiaalierän

(prosessijakson) keskiarvo, M_i näytteen i koko ja \bar{M} näytteiden keskikoko. Heterogeenisuuden arvojen avulla voidaan nyt laskea prosessin variogrammi, V_j näytevälillä j funktiona:

$$V_j = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (h_{i+j} - h_i)^2, \quad j=1,2,\dots,\frac{N}{2} \quad (8)$$

Jotta variogrammista voitaisiin arvioida näytteenoton varianssi, joudutaan variogrammin yhtälö integroimaan:

$$W(j) = \frac{1}{j} \int_0^j V(j) dj = \frac{1}{j} S(j) \quad (9)$$

Ositetun otannan varianssi estimoidaan nyt seuraavasti:

$$s_{os}^2 = \frac{2}{j^2} \int_0^j S(j) dj \quad (10)$$

ja systemaattisen otannan varianssi:

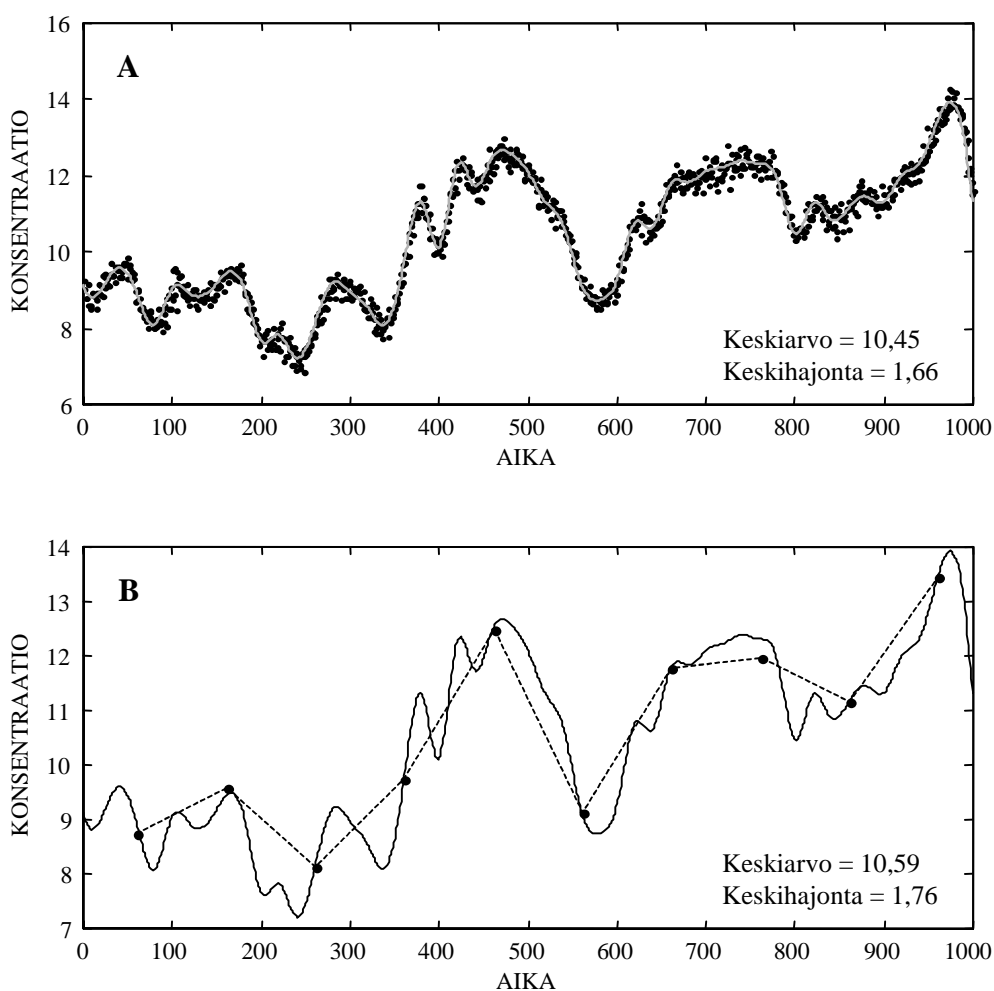
$$s_{sy}^2 = 2W\left(\frac{j}{2}\right) - s_{os}^2 \quad (11)$$

Gy suosittelee kokeellisen variogrammin analysoimiseksi suoraa numeerista integrointia.

4.3.3.2 Esimerkkejä variogrammin käytöstä

Kuva 7A esittää simuloitua ns. *random walk* -mallilla kehitettyä prosessia, jossa edelliseen lukuun on aina lisätty uusi positiivinen tai negatiivinen satunnaisluku, jonka keskiarvo = 0. Peräkkäiset havainnot eivät siten ole toisistaan riippumattomia. Kuvaa varten on generoitu 1000 lukua. Yhtenäinen viiva edustaa jatkuvaa prosessia, jonka keskiarvo on 10,45 ja keskihajonta 1,66. Pisteet viivan ympärillä vastaavat perushajontaa. Kuvassa 7B on otettu tästä funktiosta kymmenen näytettä tasaväliotoksena. Näiden avulla voidaan funktion kulku nähdä pääpiirteittäin ja otoksen keskiarvo, 10,59, onkin varsin lähellä oikeaa ja näytteiden välinen keskihajonta 1,76 on myös varsin hyvä estimaatti prosessin keskihajonnalle. Taulukossa 1 puolestaan on kymmenen otoksen keskiarvo ja keskihajonta. Kussakin otoksessa on kymmenen tasavälein otettua havaintoa, ainoastaan aloituskohta on ollut joka otoksessa eri kohdassa. Jos luvut olisivat normaalijakaantuneita satunnaislukuja, olisi kymmenen näytteen otoksen keskiarvojen keskihajonnan odotusarvo $1,66/\sqrt{10}=0,525$ (suhteellinen keskihajonta 5,02 %) ja 95 %:n luottamusväli keskiarvon virheelle 1,05. Taulukon keskiarvojen vaihteluväli oikean tuloksen ympärillä on kuitenkin huomattavasti pienempi, suurin virhe on -0,44 (suhteellinen virhe -4,13%), samoin kuin keskiarvojen virheiden keskihajontakin, joka on vain 0,19; vastaavasti suhteellisten virheiden keskihajonta on 1,79 %.

Kuvassa 8 A on esitetty kuvan 7 A näytepisteiden avulla laadittu variogrammi ja kuvassa 8 B systemaattisen ja ositetun otannan keskihajonta näytevälin funktiona. Jos näyteväli (=ositus) on 100, niin tällöin tämän pituisen jakson suhteelliset keskihajonnat ovat 5,7 % systemaattiselle otannalle ja 8,0 % ositetulle otannalle. Jos erästä otetaan 10 näytettä käyttäen samaa näyteväliä niin niiden keskiarvojen suhteelliset keskihajonnat ovat tämän mukaan $5,7\% / \sqrt{10} = 1,8\%$ systemaattiselle otannalle ja $8,0 / \sqrt{10} = 2,5\%$ ositetulle otannalle.



Kuva 7

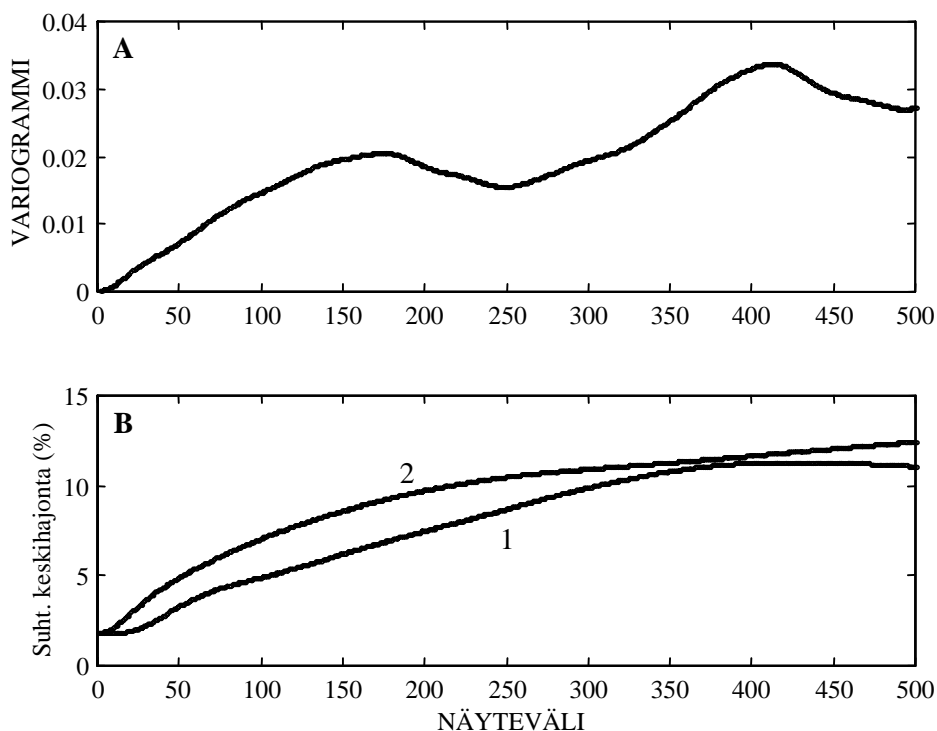
- A. Simuloitu 'random walk' -mallilla kehitetty jatkuva prosessi. Yhtenäinen viiva kuvaa prosessia ja pisteet sen ympärillä näytteenoton ja analyysimenetelmän lyhytkestoista hajontaa l. perusvirheen hajontaa.
- B. Funktiosta tasavälein otetun 10 näytteen avulla voidaan sen yleinen kulku estimoida jo suhteellisen luotettavasti (katkoviiva).

Taulukko 1

Tulokset, jotka saatiin, kun kuvan 7 A funktiosta otettiin kymmenen kappaletta näytteitä tasavälein. Otosten koko oli 10 ja näyteväli 100 ja aloituskohta määritettiin kussakin tapauksessa satunnaisesti väliltä 1...100.

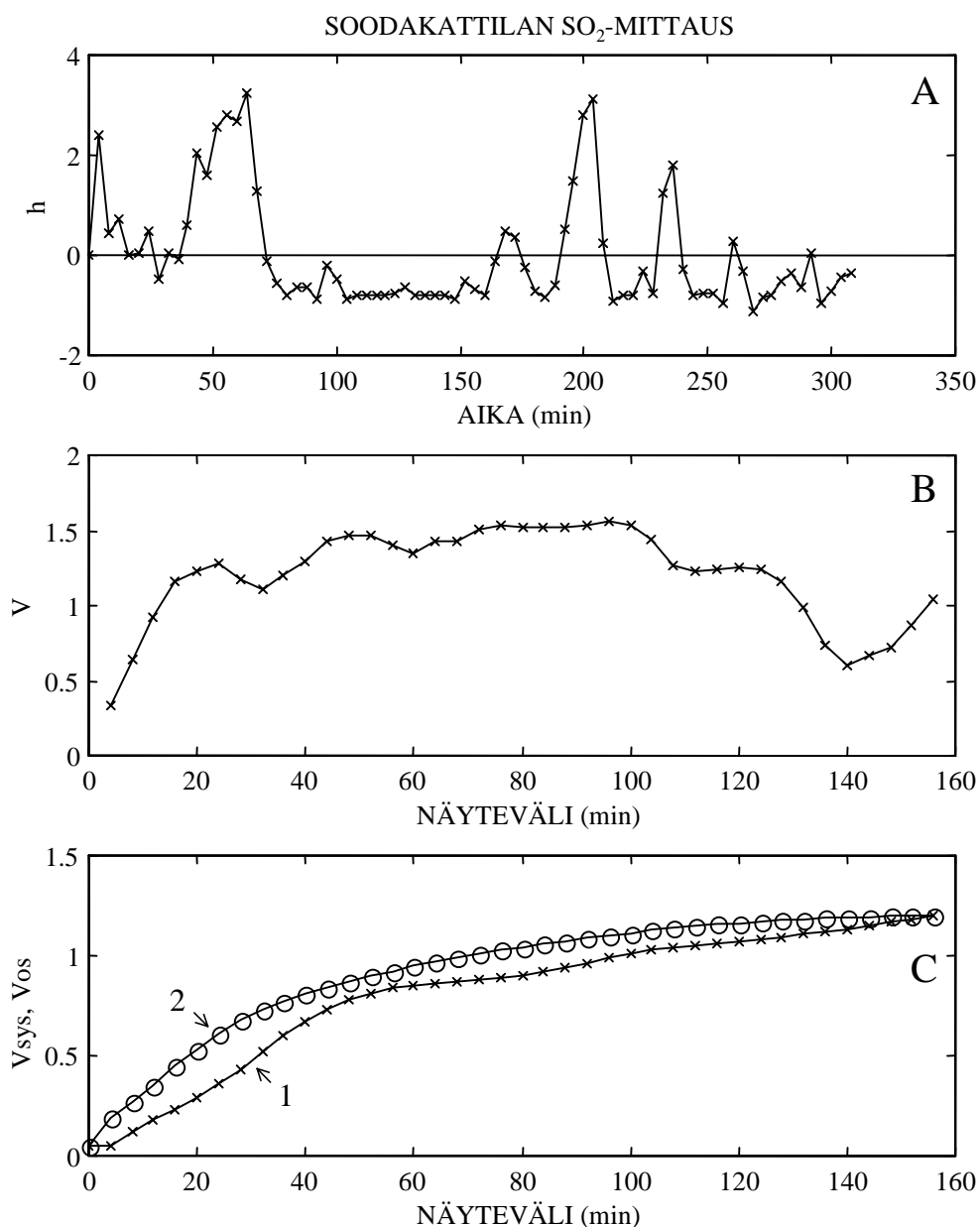
Keskiarvot		Otosten keskiarvojen virheet		Otosten keskihajonnat
Otos	Oikea	Absoluuttinen	Suhteellinen (%)	
10,21	10,55	-0,34	-3,26	2,11
10,43	10,72	-0,29	-2,71	1,56
10,64	10,76	-0,11	-1,03	1,60
10,66	10,61	0,05	0,45	1,92
10,29	10,67	-0,38	-3,55	1,57
10,71	10,73	-0,03	-0,23	1,56
10,23	10,68	-0,44	-4,13	1,56
10,63	10,59	0,04	0,37	1,87
10,66	10,61	0,05	0,45	1,92
10,50	10,64	-0,13	-1,25	1,65
Otosten yhdistetty keskihajonta				1,74
Virheiden keskihajonnat		0,19	1,79	

Jos 10 näytettä olisi otettu puhtaasti satunnaisotantana, niin keskiarvojen välisen keskihajonnan odotusarvo olisi $= 1,74/\sqrt{10} = 0,55 > 0,19$



Kuva 8

- A. Kuvan 7 A pisteiden avulla laskettu näytteenottovirheen variogrammi näytevälin funktiona.
 B. Variogrammin avulla estimoitu näytteenoton suhteellinen keskihajonta näytevälin funktiona systemaattiselle otannalle (1) ja ositetulle otannalle (2)



Kuva 9

A. Soodakattilan SO₂-emissiomittausten tulokset esitettynä heterogeenisuutena.

Näyteväli konsentraatiomittauksissa oli n. 4. min.

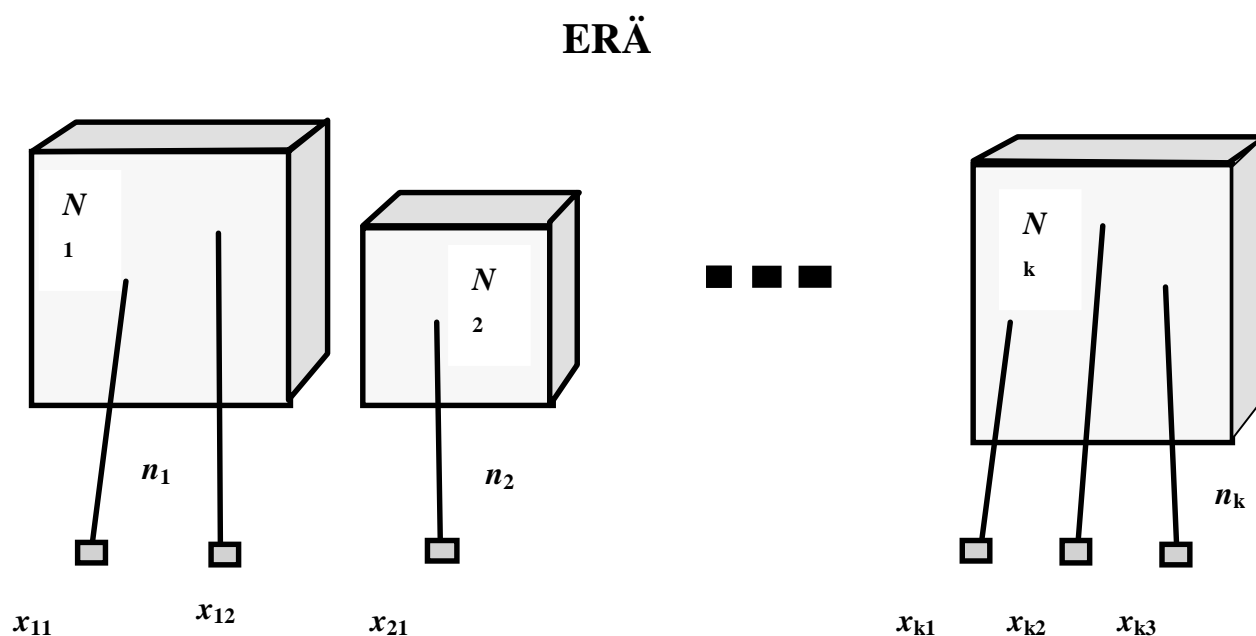
B. Heterogeenisuuden variogrammi

C. Variogrammin avulla estimoitu näytteenoton suhteellinen varianssi näytevälin funktiona systemaattiselle otannalle (1) ja ositetulle otannalle (2)

Kuvassa 9 on puolestaan esitetty erään soodakattilan SO₂-mittausten perusteella laadittu variogrammi sekä sen perusteella laaditut systemaattisen ja ositetun otannan suhteellisten varianssien riippuvuudet näytteiden välisestä etäisyydestä. Käyrien avulla voidaan helposti laatia mittausstrategia, jolla päästään haluttuun keskiarvon luotettavuustasoon.

5 Näytteenotto- ja analyysiketjun optimointi (Sommerin [8] mukaan)

5.1 Osajoukot eri suuria (Ositetun otannan teoriaa)



Kuva 10

Ositettuun otantaan liittyvät suureet, kun osat ovat eri suuret

5.1.1 Erän keskiarvon varianssin estimointi (yleisesti)

Kuvan 10 mukaisen, eri suuruisista osista koostuvan erän, keskiarvon keskihajonta lasketaan seuraavasti:

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^k \left(\frac{N_i}{N} \right)^2 \frac{N_i - n_i}{N_i - 1} \frac{\sigma_i^2}{n_i} \quad (12 \text{ a})$$

$$\approx \sum_{i=1}^k \left(\frac{N_i}{N} \right)^2 \frac{\sigma_i^2}{n_i}, \text{ jos } n_i \ll N_i \quad (12 \text{ b})$$

missä

$$\sigma_x^2 = \text{erän keskiarvon varianssi, } (\bar{x} = \frac{1}{N} \sum_{i=1}^k N_i \bar{x}_i \text{ ja } \bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i})$$

N_i = osan i suhteellinen koko (= potentiaalisten näytteiden luku)

$$N = \sum_{i=1}^k N_i = \text{tutkittavan erän koko}$$

σ_i^2 = osan i sisäinen varianssi

n_i = osasta i otettujen näytteiden luku

k = osien luku

5.1.2 Näytteiden ositus, kun ainoastaan osien suhteelliset koot tunnetaan

Jos erästä otetaan enintään n näytettä, eikä osien sisäisiä variansseja tunneta, niin paras tulos saavutetaan, kun kokonaisnäyttemäärä jaetaan osien kesken siten, että kustakin osajoukosta otettujen näytteiden luku on verrannollinen osan kokoon:

$$n_i = \frac{N_i}{N} n \tag{13}$$

missä

$$n = \text{analysoitavien näytteiden kokonaismäärä} = \sum n_i$$

Tulos on luotettavampi, kuin mitä saataisiin, jos kokonaisnäyttemäärä jaetaan satunnaisesti osien kesken

n_i :t on pyöristettävä kokonaisluvuiksi siten, ettei niiden summa, n , ylitä suurinta sallittua arvoa

5.1.3 Sekä osien suhteelliset koot että osien sisäiset varianssit, s_i^2 , tunnetaan

Tässä tapauksessa optimoitu näytteenottosuunnitelma voidaan tehdä kahdella eri tavalla:

- a) Minimoidaan keskiarvon varianssi, kun kokonaiskustannusten maksimi (l., näytteiden maksimimäärä, n , joka voidaan tutkia) on etukäteen määrätty:

$$n_i = \frac{N_i s_i}{\sum_{i=1}^k N_i s_i} n \tag{14}$$

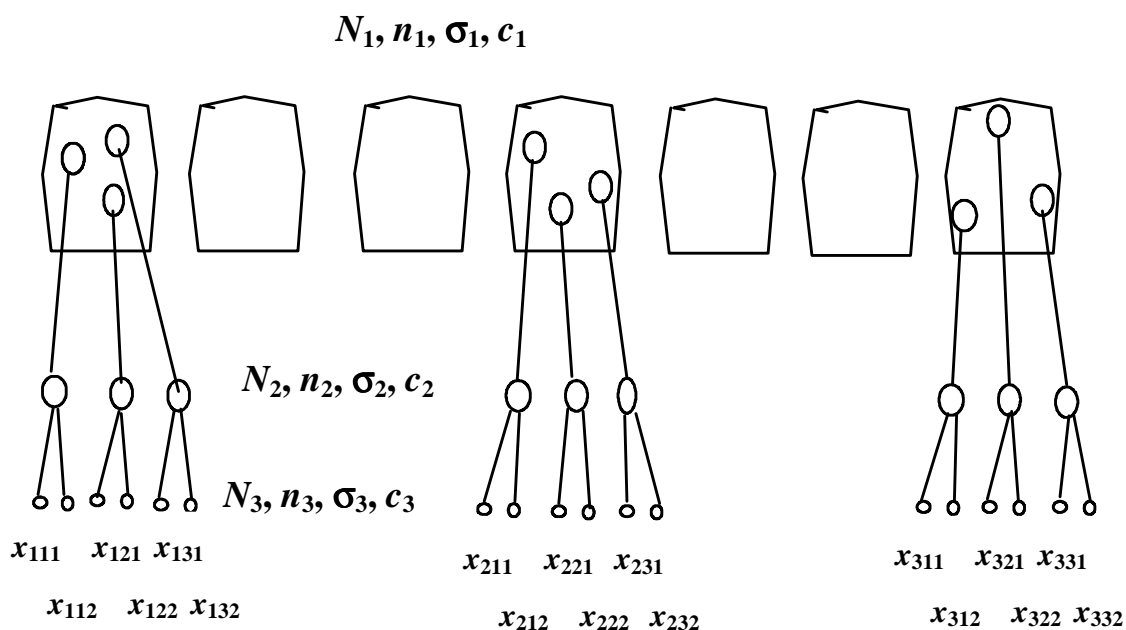
- b) Minimoidaan kustannukset, kun keskiarvolta vaaditaan tietty luotettavuustaso, ts.. keskiarvon tavoitevarienssi, s_T^2 , on etukäteen määrätty (ts., näytteiden kokonaismäärä, n , on minimoitava):

$$n_i = \frac{N_i s_i}{N^2 \sigma_T^2} \sum_{i=1}^k N_i s_i \quad (15)$$

Oletukset:

Näytteen tutkimuskulut ovat samat riippumatta siitä, mistä osajoukosta ne on otettu.

5.2 Näytteenotto- ja analyysiketjun optimointi, kun osat ovat samansuuruisia



Kuva 11

Ositettuun otantaan liittyvät suureet, kun osat ovat saman suuruiset ja näytteenotto toteutetaan useassa vaiheessa

Erä:

Koostuu N_l :stä osajoukosta, joista n_l :stä otetaan näyte tai näytteitä, s_l = osien keskiarvojen välinen keskihajonta ja c_l = näytteenoton kohteen, l. osan valinnan, yksikkökustannus (hyvin usein $c_1 \approx 0$).

Primäärinäytteet:

N_2 = osajoukkojen suhteellinen koko (tai potentiaalisten primäärinäytteiden luku yhdessä osajoukossa), n_2 = yhdestä osajoukosta otettujen näytteiden luku, s_2 = osajoukon sisäinen varianssi, c_2 = yhden primäärinäytteen oton ja käsittelyn aiheuttamat kustannukset.

Sekundäärinäytteet (tai analyysivaihe, jos tämä on optimoitavan ketjun viimeinen vaihe):

N_3 = primäärinäytteen suhteellinen koko (tai potentiaalinen sekundäärinäytteiden luku primäärinäytteessä), n_3 = primäärinäytteestä tehtyjen rinnakkaisnäytteiden luku, s_3 =

rinnakkaismäärittysten varianssi, $c_3 =$ yhden sekundäärinäytteen valmistus- ja tutkimuskustannukset (tai yhden analyysin yksikkökustannus).

5.2.1 Perusyhtälöt

Seuraavissa yhtälöissä i, j, k ovat määrittysketjun peräkkäisten tasojen indeksit summauksissa (osajoukko, primäärinäyte ja sekundääri- tai analyysinäyte kuvan 11 tapauksessa).

Analysoitavien näytteiden kokonaismäärä:

$$n = n_1 n_2 n_3 \quad (16)$$

Erän keskiarvo:

$$\bar{x} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} x_{ijk} \quad (17)$$

Keskiarvon varianssi:

$$\sigma_x^2 = \frac{N_1 - n_1}{N_1 - 1} \frac{\sigma_1^2}{n_1} + \frac{N_2 - n_2}{N_2 - 1} \frac{\sigma_2^2}{n_1 n_2} + \frac{N_3 - n_3}{N_3 - 1} \frac{\sigma_3^2}{n_1 n_2 n_3} \quad (18a)$$

$$\approx \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} + \frac{\sigma_3^2}{n_1 n_2 n_3}, \text{ jos jokainen } n_i \ll N_i \text{ ja } N_i \gg 1 \quad (18b)$$

Tutkimuksen kokonaiskustannukset:

$$c = n_1 c_1 + n_1 n_2 c_2 + n_1 n_2 n_3 c_3 \quad (19)$$

5.2.2 Tarvittavien varianssikomponenttien estimointi

Tähän on kolme mahdollisuutta:

- A priori* (ts., käytetään jostakin aikaisemmasta tutkimuksesta saatuja tuloksia hyväksi)
- Teoreettisesti (esim. estimoimalla Gyn teorian mukainen perushajonta ja käyttämällä sopivaa varmuuskerrointa)
- Kokeellisesti joko Gyn esittämän variogarafisen kokeen avulla, tai varianssianalyysin (ANOVA) avulla. Jos tulokset organisoidaan siten, kuin kuvassa 11 on tehty, niin ANOVA voidaan suorittaa yhtälöiden 20-25 mukaisesti. Tällöin joka tasolla on tehtävä vähintään 2 rinnakkaisnäytettä.

$$\sigma_3^2 \approx s_3^2 = \frac{1}{n_1 n_2 (n_3 - 1)} \sum_{i=1}^{n_3} \sum_{j=1}^{n_2} \sum_{k=1}^{n_1} (x_{ijk} - \bar{x}_{ij})^2 \quad (20)$$

$$\bar{x}_{ij} = \frac{1}{n_3} \sum_{k=1}^{n_3} x_{ijk} \quad (21)$$

$$\sigma_2^2 \approx s_2^2 = \frac{1}{n_1 (n_2 - 1)} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\bar{x}_{ij} - \bar{x}_i)^2 - \frac{s_3^2}{n_3} \quad (22)$$

$$\bar{x}_i = \frac{1}{n_2} \sum_{j=1}^{n_2} \bar{x}_{ij} \quad (23)$$

$$\sigma_1^2 \approx s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 - \frac{s_2^2}{n_2} - \frac{s_3^2}{n_2 n_3} \quad (24)$$

$$\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} \bar{x}_i = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} x_{ijk} \quad (25)$$

5.2.3 Ketjun optimointi

Olettamalla kaikki n_i :t jatkuviksi funktioiksi voidaan johtaa analyttinen ratkaisu optimin löytämiseksi differentioimalla yhtälöt 18b ja 19. Jälleen optimointi voidaan suorittaa kahdella tavalla.

- a) Määritetään minimivarianssi, johon voidaan päästä, kun tutkimuksen maksimikustannukset, c_{max} , on määrätty:

$$n_3 = \frac{\sigma_3}{\sigma_2} \sqrt{\frac{c_2}{c_3}} \quad (26)$$

$$n_2 = \frac{\sigma_2}{\sigma_1} \sqrt{\frac{c_1}{c_2}} \quad (27)$$

$$n_1 = \frac{c_{max}}{\sigma_1 \sqrt{c_1} + \sigma_2 \sqrt{c_2} + \sigma_3 \sqrt{c_3}} \frac{\sigma_1}{\sqrt{c_1}} \quad (28)$$

- b) Optimoidaan ketju siten, että keskiarvon tavoitevarianssi, s_T^2 , saavutetaan minimikustannuksin

$$n_3 = \frac{\sigma_3}{\sigma_2} \sqrt{\frac{c_2}{c_3}} \quad (29)$$

$$n_2 = \frac{\sigma_2}{\sigma_1} \sqrt{\frac{c_1}{c_2}} \quad (30)$$

$$n_1 = \frac{\sigma_1}{\sqrt{c_1}} \frac{\sigma_1 \sqrt{c_1} + \sigma_2 \sqrt{c_2} + \sigma_3 \sqrt{c_3}}{\sigma_T^2} \quad (31)$$

Tulokset on pyöristettävä kokonaisluvuiksi siten, että c_{max} a) -tapauksessa ja s_T^2 b) -tapauksessa eivät ylitä. Pyöristyksessä on lisäksi otettava huomioon, että hyväksyttävä ratkaisu ei millään tasolla saa ylittää reunaehtoja: $N \geq n \geq 1$

5.2.4 Joitakin käytännön ohjeita optimoinnin suorittamiseksi

Koska yhtälöitä 26-31 johdettaessa n_i :t on oletettu jatkuviksi funktioiksi, ja näytteiden luku voi olla ainoastaan kokonaisluku, niin yhtälöiden 26-31 avulla saadut tulokset ovat vain likimääräisiä. Aika usein käytännössä näytteiden preparointi ja homogenisointi pienentää ketjun myöhempien vaiheiden variansseja ja suurentaa yksikkökustannuksia, jolloin $\sigma_1 > \sigma_2 > \sigma_3$ ja $c_1 < c_2 < c_3$. Tässä tapauksessa yhtälöiden 14 ja 15 avulla saadaan tulos $n_2 < 1$ ja $n_3 < 1$, jolloin on valittava näille arvot $n_2 = n_3 = 1$, eikä yhtälöitä 28 ja 31 voi enää käyttää n_i :n estimointiin. n_i voidaan nyt estimoida joko yhtälön 18 avulla (tapaus a: varianssin minimointi) tai yhtälön 19 avulla (tapaus b: kustannusten minimointi):

Jos on valittu arvot $n_2 = n_3 = \dots = 1$ niin yhtälöstä 19 ratkaistaan

$$n_1 = \frac{c_{max}}{c_1 + c_2 + c_3} \quad (32)$$

kun tutkimuksen maksimikustannukset, c_{max} , on määrätty,

Jos taas s_T on määrätty, niin yhtälön 18b avulla saadaan tulos:

$$n_1 = \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}{\sigma_T^2} \quad (33)$$

Jos ehto $n_i \ll N_i$ ei ole voimassa, mikä oli yhtälössä 18b tehty approksimaatio, niin n_1 voidaan myös ratkaista yhtälöstä 18a (kun $n_2 = n_3 = 1$):

$$n_1 = \frac{\frac{N_1}{N_1 - 1} \sigma_1^2 + \sigma_2^2 + \sigma_3^2}{\sigma_T^2 + \frac{\sigma_1^2}{N_1 - 1}} \quad (34)$$

Aina, kun n_2 :n ja n_3 :n suhteellinen pyöristysvirhe on suurehko (esim. 1,5 pyöristetään 2:ksi), n_1 on varminta ratkaista lopullisesti yhtälöistä 18 tai 19 yhtälöiden 28 ja 31 asemasta. Yhtälö 18a antaa tulokseksi, kun n_2 ja n_3 on määrätty:

$$n_1 = \frac{\frac{N_1}{N_1-1} \sigma_1^2 + \frac{N_2 - n_2}{N_2 - 1} \frac{\sigma_2^2}{n_2} + \frac{N_3 - n_3}{N_3 - 1} \frac{\sigma_3^2}{n_2 n_3}}{\sigma_T^2 + \frac{\sigma_1^2}{N_1 - 1}} \quad (35)$$

Optimointi voidaan käytännössä suorittaa myös kokeilemalla esim. tietokoneen avulla. Koska n on kokonaisluku, jolloin kokeiltavien lukujen määrä on rajoitettu, ei tällaisen optimointialgoritmin laatiminen ole kovin monimutkainen tehtävä.

Esimerkki

Tehtaalle tuleva erä raaka-ainetta on pakattu 50 kg:n säkkeihin ja erän koko on 10000 kg. Alustavien kokeiden perusteella on saatu vastaanottotarkastuksessa määritettävälle epäpuhtaudelle seuraavat keskihajonnan estimaatit:

Säkkien välinen hajonta	= 10 %
Säkkien sisäinen hajonta	= 5 % (kun näyteko on 100 g)
Analyysin hajonta	= 2 % (analyysinäytteen koko 1 g)

Eri työvaiheiden yksikkökustannukset on arvioitu seuraaviksi:

Säkin valinta	= ei kustannuksia
Näytteenotto säkistä	= 20 mk
Analyysi	= 40 mk

Erän keskipitoisuus halutaan määrittää siten, että keskiarvon keskihajonta saa olla enintään 2,0 %. Optimoit näytteenotto- ja analyysiketju siten, että tavoitteeseen päästään minimikustannuksin.

Ratkaisu:

Tarkastusketjussa on kolme virhettä generoivaa vaihetta, jotka ovat 1) säkin valinta, 2) primäärinäytteen otto säkistä, ja 3) analyysin suoritus. Näille vaiheille saadaan tehtävästä seuraavat lähtötiedot:

$$N_1 = \frac{10000 \text{ kg}}{50 \text{ kg}} = 200, \quad s_1 = 10 \%, \quad c_1 = 0 \text{ mk}, \quad n_3 = ?$$

$$N_2 = \frac{50 \text{ kg}}{0,1 \text{ kg}} = 500, \quad s_2 = 5 \%, \quad c_2 = 20 \text{ mk}, \quad n_2 = ?$$

$$N_3 = \frac{100 \text{ g}}{1 \text{ g}} = 100, \quad s_3 = 2 \%, \quad c_3 = 40 \text{ mk}, \quad n_3 = ?$$

Tavoite: $s_x \leq s_T = 2\%$

Kaavojen 26 ja 27 avulla saadaan tulokset:

$$n_3 = \frac{s_3}{s_2} \sqrt{\frac{c_2}{c_3}} = \frac{2\%}{5\%} \sqrt{\frac{20mk}{40mk}} = 0,28, \text{ ja valitaan } n_3 = 1$$

$$n_2 = \frac{s_2}{s_1} \sqrt{\frac{c_1}{c_2}} = \frac{5\%}{10\%} \sqrt{\frac{0mk}{20mk}} = 0, \text{ jolloin valitaan } n_2 = 1$$

n_1 on ratkaistava nyt kaavasta 33 tai 34. Kaava 33 antaa tulokseksi:

$$n_1 = \frac{s_1^2 + s_2^2 + s_3^2}{s_T^2} = \frac{(10\%)^2 + (5\%)^2 + (2\%)^2}{(2\%)^2} = 32,2 \approx 32$$

Koska tulos 32 on yli 10 % 200:sta, niin lasketaan ratkaisu vielä myös kaavan 34 avulla:

$$n_1 = \frac{\frac{N_1}{N_1-1} s_1^2 + s_2^2 + s_3^2}{s_T^2 + \frac{s_1^2}{N_1-1}} = \frac{\frac{200}{200-1} (10\%)^2 + (5\%)^2 + (2\%)^2}{(2\%)^2 + \frac{(10\%)^2}{200-1}} = 28,8 \approx 29$$

Kuten havaitaan, ei likimääräisen ja tarkan ratkaisun välillä ole kovin suurta eroa. Minimikustannuksin tavoitteeseen päästään, kun toimituserän 200 säkistä valitaan (mieluiten nekin ositetun otannan periaatteella) 29 säkkiä, joista kustakin otetaan yksi primäärinäyte, joista edelleen kustakin suoritetaan yksi analyysi.

6 Kirjallisuusviitteet

1. Gy, P.M., Sampling of Particulate Materials, Theory and Practice, Elsevier, Amsterdam, 1982.
2. Gy, P.M. Sampling of Heterogeneous and Dynamic Material Systems, Elsevier, Amsterdam, 1992.
3. Pitard, F.F., Pierre Gy's Sampling Theory and Sampling Practice, Vols. I and II, CRC Press Boca Raton, 1989.
4. Gy, P. M., Sampling for Analytical Purposes, John Wiley & Sons Ltd, Chichester, 1998.
5. Statistical Aspects of Sampling from bulk materials - Part 2: Sampling of particulate materials, Draft International standard ISO/DIS 11648-2, 1997.
6. Minkkinen, P., SAMPEX - A computer program for solving sampling problems, Chemometrics and Intelligent Laboratory Systems, 7 (1989) 189-194.

7. SAMPEX - A computer program for solving sampling problems, User's Manual, Prosessilaskenta Oy, Kauniainen, puh. 09-5051585.
8. Sommer, K., Probenahme von Pulvern und körnigen Massengütern, Springer, Berlin, 1979.