

**Rapport nr. 714701012**

**Betrouwbaarheidsanalyse van de NH<sub>3</sub>-emissie  
berekningen voor Nederland in 1992**

**P.H.M. Janssen, A.A.M. Kusse,  
A.U.C.J. van Beurden, K.W. van der Hoek,  
P.M. van Egmond, N.J.P. Hoogervorst**

**September 1995**

Achtergrond-document bij rapport 714701011: Betrouwbaarheid  
van milieu-informatie: analyse van een stikstofketen

Dit onderzoek werd verricht in opdracht en ten laste van de Directie RIVM,  
in het kader van de projecten 714701 en 482532.

## Verzendlijst

1. Directie RIVM
  2. Drs. N. Joanknecht (DGM-DWL)
  3. Mr. V.G. Keizer (DGM-LE)
  4. Ir. S.M. Smeulders (DGM-LE)
  5. Drs. J. Dijk (LEI-DLO)
  6. Ing. D.A. Oudendag (LEI-DLO)
  7. Prof. dr. P.A. Burrough (RUU)
  8. Dr. ir. G.B.M. Heuvelink (UvA)
  9. Prof. dr. H.J. Scholten (VU)
  10. Drs. J.M.M. Aben
  11. Dr. L.C. Braat
  12. Dr. A.L.M. Dekkers
  13. Ir. H.S.M.A. Diederer
  14. Ir. G. van Drecht
  15. Ir. W. van Duijvenbooden
  16. Drs. J.C.H. van Eijkeren
  17. Dr. ing. J.W. Erisman
  18. Drs. A. van der Giessen
  19. Ir. B.J. de Haan
  20. Ir. G.J. Heij
  21. Dr. P.S.C. Heuberger
  22. Dr. J.A. Hoekstra
  23. Ing. J.A. van Jaarsveld
  24. Dr. A.S.M. Koeleman
  25. Drs. L.H.M. Kohsiek
  26. Mw. Dr. M.A.J. Kuijpers-Linde
  27. Dr. ir. J.W. de Kwaadsteniet
  28. Ir. F. Langeweg
  29. Dr. ir. D. van Lith
  30. Drs. R.J.M. Maas
  31. Dr. D. Onderdelinden
  32. Drs. E.J. Pebesma
  33. Drs. J.P.M. Ros
  34. Drs. A.A. van der Veen
  35. Drs. R.J. van de Velde
  36. Drs. J. van Woerden
  - 37-43. Auteurs
  - 44-45. Bibliotheek RIVM
    46. Bureau Project- en Rapportregistratie
    47. Hoofd Bureau Voorlichting en Public Relations
  - 48-58. reserve exemplaren t.b.v. Bureau Rapportenbeheer
- Mede ter informatie aan:
59. DGM-Strategische Planning
  60. Depot van Nederlandse Publicaties en Nederlandse Bibliografie



## Inhoudsopgave

Verzendlijst	ii
Inhoudsopgave	iii
Lijst van tabellen	iv
Lijst van figuren	iv
Nomenclatuur	v
Dankwoord	v
Voorwoord	vi
Summary	vii
Samenvatting	vii
1 Inleiding	1
2 De NH <sub>3</sub> -emissieberekeningen voor 5 x 5 km gridcellen	3
3 Afspraken en uitgangspunten	5
4 Onzekerheidsspecificaties voor foutenbronnen	11
5 Vereenvoudiging van de betrouwbaarheidsanalyse	20
6 Betrouwbaarheidsanalyse: werkwijze en resultaten	24
7 Conclusies en aanbevelingen	32
Appendix A. Bewijs van Stelling 1	34
Appendix B. Kwantificering van verrasterings- en classificatiefouten	36
Appendix C. Specificatie bandbreedte voor dierlijke NH <sub>3</sub> -emissiecijfers	39
Appendix D. Analytische onzekerheidsberekeningen voor NH <sub>3</sub> -emissies	48
Referenties	54

## Lijst van tabellen

1	NH <sub>3</sub> -emissie in Nederland (in kTon NH <sub>3</sub> ) voor de jaren 1990-1992 . . . . .	3
2	De relatieve bijdrage van de diersoorten tot de dierlijke NH <sub>3</sub> -emissie voor 1992 . . . . .	15
3	Experimenten voor de betrouwbaarheidsanalyse van de nominale NH <sub>3</sub> -emissieberekeningen . . . . .	24
4	Nummers en namen van verzuringsgebieden . . . . .	25
5	Correlaties van het optreden van onder/overschatting bij de berekening van de totale NH <sub>3</sub> -emissie, en de NH <sub>3</sub> -emissie uit dierlijke mest, kunstmest, huishoudens . . . . .	26
6	Correlaties tussen de relatieve afwijkingen van nominale berekening en gemiddeld verfijnde berekening voor alle 5 x 5 km gridcellen, betreffende totale NH <sub>3</sub> -emissie, en de NH <sub>3</sub> -emissies uit dierlijke mest, kunstmest en huishoudens . . . . .	27
7	Statistics van betrouwbaarheidsfactoren voor de 5 x 5 km gridcellen, per verzuringsgebied en Nederland totaal . . . . .	29
8	Informatie over de systematische fout en de random fout in de berekeningen voor NH <sub>3</sub> -emissietotalen per verzuringsgebied en Nederland totaal . . . . .	31
9	Specificaties voor berekening bandbreedte van NH <sub>3</sub> -emissie bijdrage rundvee . . . . .	40
10	Specificaties voor berekening bandbreedte van NH <sub>3</sub> -emissie bijdrage varkens . . . . .	41
11	Specificaties voor berekening bandbreedte van NH <sub>3</sub> -emissie bijdrage pluimvee . . . . .	42
12	Totaal aantal landbouwdieren volgens tellingen in mei 1992 . . . . .	47
13	De relatieve bijdrage van de mestsoorten tot de dierlijke NH <sub>3</sub> -emissie voor 1992 . . . . .	47

## Lijst van figuren

1	Drie situaties met dezelfde betrouwbaarheidsfactor (.6), maar met een verschillende rol voor de systematische en random afwijking . . . . .	55
2	Verzuringsgebieden in Nederland . . . . .	56
3	NH <sub>3</sub> -emissie (in Ton) in gridcellen, volgens nominale berekening (a), en volgens gemiddelde verfijnde berekening (b) . . . . .	56
4	Procentuele bijdrage van dierlijke mest (a), kunstmest (b), en huishoudens (c) aan de NH <sub>3</sub> -emissie in gridcellen, volgens gemiddelde verfijnde berekening . . . . .	57
5	Systematische afwijking tussen nominale berekening en gemiddelde verfijnde berekening voor NH <sub>3</sub> -emissie in gridcellen, uitgedrukt als percentage van de gemiddelde verfijnde berekening. Dit is weergegeven voor de totale NH <sub>3</sub> -emissie (a), en voor de emissiebijdrages van dierlijke mest (b), kunstmest (c), en huishoudens (d) . . . . .	58
6	Histogram van systematische afwijking (in %) tussen nominale berekening en gemiddelde verfijnde berekening voor NH <sub>3</sub> -emissie in gridcellen. Dit is weergegeven voor de totale NH <sub>3</sub> -emissie (a), en voor de emissiebijdrages van dierlijke mest (b), kunstmest (c), en huishoudens (d) . . . . .	59
7	Landgebruikstypering van Nederland volgens het LGN-bestand RS <sup>o</sup> (25 x 25 m) . . . . .	60
8	Landgebruikstypering van Nederland volgens de door het LBG samengestelde landgebruikskaart (500 x 500 m) . . . . .	61
9	Variatiecoëfficiënten (a) en gestandaardiseerde bias (b) voor NH <sub>3</sub> -emissie in gridcellen . . . . .	62
10	Betrouwbaarheidsfactoren voor NH <sub>3</sub> -emissie in gridcellen . . . . .	62
11	Histogram van betrouwbaarheidsfactoren, en bijbehorende percentielwaarden, voor NH <sub>3</sub> -emissie in gridcellen over Nederland totaal . . . . .	63
12	Histogram van betrouwbaarheidsfactoren voor NH <sub>3</sub> -emissie in gridcellen voor diverse verzuringsgebieden . . . . .	63
13	Histogram van de gestandaardiseerde bias voor alle gridcellen met een betrouwbaarheidsfactor < .7 . . . . .	64
14	Betrouwbaarheden voor de NH <sub>3</sub> -emissie per verzuringsgebied en Nederland totaal . . . . .	65
15	Histogram van verkleiningsfactor $\sum_{i \in G_j} [f_{d,*}^o(i; j)]^2$ voor de varianties van dierlijke NH <sub>3</sub> -emissies in de 5 x 5 km gridcellen over Nederland . . . . .	66

# Nomenclatuur

## Symbolen

- B: betrouwbaarheidsfactor (-)
- $N(\mu, \sigma^2)$ : normale verdeling met gemiddelde  $\mu$  en standaarddeviatie  $\sigma$
- $\mu$ : gemiddelde
- $\sigma$ : standaarddeviatie
- IE: verwachting; gemiddelde
- Med: mediaan
- IP: kans
- Var: variantie
- VC: variatiecoëfficiënt
- $X_{std}$ : standaard-normaal verdeelde grootheid ( $N(0, 1)$ )
- $y_{nom}$ : nominale waarde
- $y_{pos}$ : gepostuleerde waarde
- $E_d, E_h, E_k$ : NH<sub>3</sub>-emissie afkomstig van dierlijke mest (d), huishoudens (h), kunstmest (k)
- $G_{25}, G_{500}$ : gemeenteverrastering op 25x25m, resp. 500x500m gridcel nivo
- Inp: input-emissie data
- MS: modelspecificatie
- $N_h$ : gemiddelde jaarlijkse NH<sub>3</sub>-emissie per inwoner
- $N_{sam}$ : aantal getrokken samples
- RS: Remote Sensing
- $VS_h, VS_m$ : verdeelsleutel voor emissiebijdrage van huishoudens (h), mest (m)

## Afkortingen

- CBS** Centraal Bureau voor de Statistiek
- COROP** Coördinatie-commissie Regionaal Onderzoek Programma
- CWM** Centrum voor Wiskundige Methoden (RIVM)
- GIS** Geographical Information System
- IKC** Informatie en Kennis Centrum
- LAE** Laboratorium voor Afvalstoffen en Emissies (RIVM)
- LBG** Laboratorium voor Bodem en Grondwater (RIVM)
- LEI-DLO** Landbouw Economisch Instituut - Dienst Landbouwkundig Onderzoek
- LGN** Land Gebruik Nederland
- LLO** Laboratorium voor LuchtOnderzoek (RIVM)
- MB** Milieubalans (RIVM)
- MV** Milieuverkenning (RIVM)
- RIVM** Rijks Instituut voor Volksgezondheid en Milieuhygiëne
- RS** Remote Sensing
- SC-DLO** Staring Centrum - Dienst Landbouwkundig Onderzoek

## Dankwoord

De auteurs danken drs. J. Dijk (LEI-DLO), ing. D.A. Oudendag (LEI-DLO), drs. J.C.H. van Eijkeren (CWM) en dr. J.A. Hoekstra (CWM) voor commentaar op een eerdere versie van dit rapport. Dr. P.S.C. Heuberger (CWM) komt dank toe voor de geboden hulp bij het vervaardigen van de figuren.

## Voorwoord

Optimalisatie van een complex informatie-systeem begint bij een sterkte-zwakke analyse van het vigerende systeem. De betrouwbaarheid van de informatie uit de samenstellende delen wordt daarbij getoetst aan criteria voor de gewenste betrouwbaarheid. Vervolgens kan een kosten-baten analyse worden uitgevoerd.

Momenteel vindt een heroverweging plaats van het RIVM-milieumonitoring systeem, dat bestaat uit diverse programma's voor de verzameling van gegevens die tezamen met wiskundige modellen informatie geven over de milieukwaliteit. Een probleem daarbij is dat de betrouwbaarheid van de onderdelen vaak slechts in beperkte mate bekend is, of voor de onderdelen op verschillende wijze is gedefinieerd zodat vergelijking niet goed mogelijk is. Daarom is een onderzoek gestart onder de naam '*Betrouwbaarheid van het milieumonitoring systeem*'. Het onderzoek heeft tot doel een algemeen bruikbare methodiek voor betrouwbaarheidsanalyse te ontwikkelen. De methodiek omvat het definiëren van het begrip betrouwbaarheid, het ontwikkelen van een set criteria voor de gewenste betrouwbaarheid, en kwantificering van betrouwbaarheid in de informatieketen (inclusief foutenvoortplanting). Toetsing vindt plaats aan de hand van een concrete toepassing. Deze methodiek kan in een vervolgproject door medewerkers van de inhoudelijke laboratoria i.s.m. CWM-medewerkers worden toegepast bij modelberekeningen en meetnetanalyses, zodat tenslotte een compleet beeld van de betrouwbaarheid van het RIVM-milieumonitoring systeem ontstaat, ter onderbouwing van verdere optimalisatie.

Als concrete toepassing is gekozen voor de berekeningen en metingen in de keten van N-emissies tot N-in-ruw-water. In '*Betrouwbaarheid van milieu-informatie: analyse van een stikstofketen*' (rapportnr. 714701011) wordt een korte uitleg van de gebruikte methoden gegeven, tezamen met de resultaten van de toepassing.

Een nadere verantwoording en methodologische details zijn opgenomen in achtergrondsdocumenten, waarvan dit rapport er één is. Het betreft hier de analyse van de betrouwbaarheid van de atmosferische NH<sub>3</sub>-emissie berekeningen in 1992 voor de 5 x 5 km gridcellen die Nederland overdekken, en voor de verzuringsgebieden waarin Nederland is opgedeeld.

## Summary

By means of analytical uncertainty computations and model simulations the reliability is assessed for the  $\text{NH}_3$ -emission computations of 1992, on local (5 x 5 km grids), regional (acidification regions) as well as national scale (the Netherlands). Results are presented in terms of confidence intervals and reliability factors. It is studied how the various error sources affect the reliability.

## Samenvatting

Door middel van analytische onzekerheidsberekeningen en model simulaties wordt de betrouwbaarheid geëvalueerd van de atmosferische  $\text{NH}_3$ -emissieberekeningen voor 1992, die zowel op locale- (5 x 5 km gebieden), regionale- (verzuringgebieden), als ook op nationale schaal (Nederland) bepaald zijn. De resultaten van de uitgevoerde analyse worden vastgelegd in de vorm van betrouwbaarheidsintervallen en betrouwbaarheidsfactoren. De invloed van de diverse foutenbronnen op de betrouwbaarheid wordt onderzocht.



## 1 Inleiding

In het kader van het project ‘*Coördinatie, Optimalisatie en Integratie van Milieumonitoring*’ (projectnr. 714701) wordt binnen het deelproject ‘*Betrouwbaarheid van het milieu monitoring systeem*’ (projectnr. 714702) een betrouwbaarheidsanalyse verricht van de berekeningsketen die zich uitstrekt van ‘N-emissies’ tot en met het effect van ‘N-in-ruw-water’ (zie figuur 2.1-2.3 in [1]). Deze wordt hierna verder aangeduid als de N-keten of stikstofketen.

Het startpunt van deze keten wordt gevormd door de berekeningen die het LAE uitvoert om de NH<sub>3</sub>-emissies op 5 x 5 km gridcellen aan te leveren, ten behoeve van N-depositieberekeningen verderop in de keten. De berekende NH<sub>3</sub>-emissiecijfers kunnen afwijken van de werkelijkheid. Debet hieraan zijn onzekerheden (t.g.v. natuurlijke variabiliteit en/of onnauwkeurigheid c.q. onvolledigheid van data en kennis) in de basisgegevens voor deze berekeningen, alsmede de onvermijdelijke aannames en vereenvoudigingen bij dit rekenproces. De grootte van de aldus resulterende afwijkingen is onzeker (onbekend), en het is voor een adequate beoordeling en gebruik van de berekende resultaten essentieel om een indruk te krijgen van deze onzekerheid, en om zijn hoofdoorzaken in kaart te brengen. Er wordt in het kader van deze studie met name naar gestreefd om deze onzekerheid samen te vatten in één specifieke maat, te weten de *betrouwbaarheidsfactor* voorgesteld in [1]. Anderzijds zal, ten behoeve van berekeningen verderop in de N-keten, de onzekerheids-informatie ook in een meer gedetailleerde en basale vorm worden opgeslagen, bijv. in termen van afzonderlijke realisaties van NH<sub>3</sub>-emissie-velden. Informatie hierover wordt verschaft door het presenteren van figuren en tabellen met gemiddelden, percentielen, standaarddeviaties, betrouwbaarheidsintervallen etc.

Uitvoering van een dergelijke *betrouwbaarheidsanalyse* vereist allereerst een specificatie van de verschillende foutenbronnen die bij de berekeningen een rol kunnen spelen. Dit betreft o.a. bovengenoemde basisgegevens die betrekking hebben op de jaargemiddelde NH<sub>3</sub>-emissies uit dierlijke mest, kunstmest en huishoudens. Deze data worden deels door externe instituten (LEI-DLO, CBS) berekend, mede op basis van uitgangspunten van het RIVM, en worden beschikbaar gesteld als totaalcijfers per gemeente. Wegens beperkingen in tijd, menskracht en reken capaciteit was het niet mogelijk om hiervan een volledig beeld te geven; simplificaties en verwaarlozingen waren onvermijdelijk, en er is ernaar gestreefd om dit zo verantwoord en realistisch mogelijk te doen. De betrouwbaarheidsevaluatie van de NH<sub>3</sub>-emissie berekeningen zal hierdoor echter voornamelijk een *tentatief en indicatief* karakter hebben, en er moet voor gewaakt worden om absolute betekenis toe te kennen aan specifieke getalswaarden.

In hoofdstuk 4 van het hoofdrapport [1] worden de hoofdlijnen van de betrouwbaarheidsanalyse van de atmosferische NH<sub>3</sub>-emissie berekeningen beschreven. In dit rapport, dat als achtergronddocument bij [1] fungeert, wordt de uitgevoerde analyse in detail toegelicht. Nadat in hoofdstuk 2 is vermeld hoe deze berekeningen plaatsvinden, komen in hoofdstuk 3 enkele afspraken en uitgangspunten aan de orde die van belang zijn voor de latere hoofdstukken. De diverse foutenbronnen die bij de emissieberekeningen een rol spelen worden nader gespecificeerd in hoofdstuk 4. Deze specificatie geeft aanleiding tot aanzienlijke vereenvoudigingen van de betrouwbaarheidsberekeningen (hoofdstuk 5). In hoofdstuk 6 wordt vervolgens toegelicht en gemotiveerd welke specifieke betrouwbaarheidsanalyse uitgevoerd is, en wat de resultaten zijn. Het rapport wordt tenslotte afgesloten met de conclusies en aanbevelingen in hoofdstuk 7. Een aantal appendices met technische resultaten completeren het geheel. Om technische redenen zijn de figuren aan het eind geplaatst.

Het gepresenteerde materiaal is deels technisch wiskundig van aard, omdat het een volledige

beschrijving beoogt te geven van de uitgevoerde betrouwbaarheidsanalyse. Voor een goed begrip van de verkregen resultaten is het echter niet nodig om deze technische passages door te nemen. In de hoofdtekst staat expliciet vermeld wanneer dit het geval is (hoofdstuk 3, 5, en de appendices).

## 2 De NH<sub>3</sub>-emissieberekeningen voor 5 x 5 km gridcellen

NH<sub>3</sub>-emissies zijn afkomstig uit een viertal broncategorieën, te weten *dierlijke mest, kunstmest*, *industriële processen* en *huishoudens*. In tabel 1 zijn de absolute en procentuele emissiebijdrages van elk van deze bronnen vermeld voor de jaren 1990 tot en met 1992 ([2]). Verder zorgen buitenlandse NH<sub>3</sub>-emissies nog voor NH<sub>x</sub>-depositie in Nederland. Het buitenlandse aandeel bedroeg in 1989 19 % van de totale NH<sub>x</sub>-depositie ([3]).

Bron	1990	1991	1992
Dierlijke mest	188,9 (87.9 %)	195,3 (88.5 %)	145,8 (85.6 %)
<ul style="list-style-type: none"> <li>- stallen en opslag</li> <li>- beweiding</li> <li>- mestaanwending</li> </ul>	<ul style="list-style-type: none"> <li>{ 81.5(43.1%)</li> <li>{ 14.8( 7.8%)</li> <li>{ 92.6(49.0%)</li> </ul>	<ul style="list-style-type: none"> <li>{ 89.1(45.6%)</li> <li>{ 16.7( 8.6%)</li> <li>{ 89.5(45.8%)</li> </ul>	<ul style="list-style-type: none"> <li>{ 88.5(60.7%)</li> <li>{ 14.5( 9.6%)</li> <li>{ 42.9(29.4%)</li> </ul>
Kunstmest	10,6 ( 4.9 %)	10,3 ( 4.7 %)	9,4 ( 5.5 %)
Industrie	5,1 ( 2.4 %)	4,5 ( 2.0 %)	4,5 ( 2.7 %)
Huishoudens	10,4 ( 4.8 %)	10,5 ( 4.8 %)	10,6 ( 6.2 %)
Totaal	215,0 (100 %)	220,6 (100 %)	170,3 (100 %)

Tabel 1: NH<sub>3</sub>-emissie in Nederland (in kTon NH<sub>3</sub>) voor de jaren 1990-1992 (bron [2])

Ten behoeve van NH<sub>x</sub>-depositieberekeningen verderop in de N-keten bepaalt het LAE allereerst de NH<sub>3</sub>-emissie op 5 x 5 km resolutienivo. Dit betreft echter enkel de bijdragen voor dierlijke mest, kunstmest en huishoudens; de bijdragen van industriële bronnen en buitenland, die als puntbronnen respectievelijk oppervlaktebronnen worden aangemerkt, worden rechtstreeks verdisconteerd bij de depositieberekeningen en hoeven geen voorbereiding door het LAE te ondergaan. Dit aandeel blijft buiten beschouwing in dit rapport.

De *basisgegevens* die het LAE gebruikt voor de emissieberekening op 5 x 5 km gridcellen zijn gegevens op gemeentelijke schaal ([2]):

- Jaargemiddelde NH<sub>3</sub>-emissiecijfers uit *dierlijke mest* per gemeente worden door het LEI-DLO ter beschikking gesteld, en zijn bepaald aan de hand van uitgebreide modelberekeningen met gegevens uit de Landbouwtellingen, hierbij uitgangspunten van het RIVM hanterend (emissiefactoren, penetratie graden), die afgeleid zijn uit data van CBS, LEI-DLO, IKC Veehouderij e.a.
- De NH<sub>3</sub>-emissie uit *kunstmest* is verkregen door het jaargemiddelde stikstof kunstmestverbruik per gemeente (berekend door LEI-DLO) te vermenigvuldigen met een gemiddelde vervluchtigingscoëfficiënt van 2 %, en vervolgens om te rekenen in eenheden ammoniak (vermenigvuldigingsfactor  $\frac{17}{14}$ ).
- De NH<sub>3</sub>-emissiebijdrage van *huishoudens* wordt per gemeente vastgesteld door de gemeentelijke inwoneraantallen (bevolkingsstatistiek van het CBS) te vermenigvuldigen met een emissiefactor van 0.7 kg NH<sub>3</sub> per inwoner per jaar.

Deze gegevens worden door het LAE via GIS-bewerkingen omgerekend naar 5 x 5 km gridcellen die Nederland overdekken. Daartoe worden de gemeenten allereerst verrasterd naar 500 x 500 m arealen, omdat voor dit resolutienivo een landgebruikskaart beschikbaar is. Deze kaart is door het LBG samengesteld op basis van een herclassificatie (in 6 dominante landgebruiksklassen<sup>1</sup>) van informatie uit het LGN-bestand afkomstig van het Staring Centrum (SC-DLO; Wageningen). Dit bestand werd afgeleid uit een remote sensing studie uit

<sup>1</sup>De klassen die onderscheiden worden zijn: gras, maïs, akkerbouw, tuinbouw, bos en bebouwd. Tevens is er nog een rest-klasse die al het overige bevat.

1986-1988, en bevat informatie over het landgebruik in heel Nederland voor 25 x 25 m arealen (zie [4]). De 6 landgebruiksklassen die na herclassificatie resulteren, worden verder in 2 categorieën ingedeeld: *agraris*ch landgebruik en *niet-agraris*ch landgebruik.

De NH<sub>3</sub>-emissie uit dierlijke mest en kunstmest per gemeente wordt vervolgens homogeen verdeeld over de gemeentelijke 500 x 500 m arealen met agrarisch landgebruik (d.w.z. grasland, maïs- en bouwland). De emissie uit huishoudens wordt homogeen verdeeld over het *gehele* gemeente-oppervlakte, omdat de gehanteerde landgebruiksclassificatie van de 500 x 500 m gridcellen onvoldoende differentiatie biedt voor een adequate toekenning van deze emissie aan urbane gebieden. Al deze bijdragen gesommeerd per 5 x 5 km gridcel levert tenslotte de NH<sub>3</sub>-emissiecijfers voor deze gridcellen.

### 3 Afspraken en uitgangspunten

- In dit hoofdstuk worden enkele notationale afspraken en uitgangspunten gepresenteerd die bij de presentatie in de volgende hoofdstukken gebruikt worden. Tevens wordt de betrouwbaarheidsfactor geïntroduceerd, die in deze studie als maat gebruikt wordt voor het expliciet aanduiden van de betrouwbaarheid van de uitgevoerde modelberekeningen. Enkele belangrijke eigenschappen van deze grootheid worden besproken. De presentatie van deze eigenschappen is deels technisch van aard, en kan zonder al te grote bezwaren voor de leesbaarheid en het begrip van de vervolghoofdstukken worden overgeslagen.

Allereerst echter de notationale afspraken:

- $y(r \mid Inp; MS; G_{500}; LBG)$  duidt de berekende  $\text{NH}_3$ -emissie<sup>2</sup> in gridcel<sup>3</sup>  $r$  aan bij gegeven input-emissie data  $Inp$ , modelspecificatie  $MS$ , gemeente-verrastering  $G_{500}$  (d.w.z. op basis van het 500 x 500 m resolutie-nivo), en landgebruiksclassificatie via de LBG-kaart.

De input-emissie data  $Inp$  refereren naar de LEI-DLO/CBS-gegevens m.b.t. gemeentelijke  $\text{NH}_3$ -emissie uit mest en de inwoneraantallen.

De modelspecificatie  $MS$  betreft  $[VS_m; VS_h; N_h]$ , waarbij  $VS_m; VS_h$  de verdeelsleutels voor de bijdrage van mest resp. huishoudens zijn.  $N_h$  is de gemiddelde jaarlijkse  $\text{NH}_3$ -emissie per inwoner. De verdeelsleutels hebben betrekking op de vigerende verrastering en classificatie, d.w.z. op  $G_{500}$  en  $LBG$ , i.e. op 500 x 500 m gridcel-nivo.

- $y(r \mid Inp; MS; G_{25}; RS)$  duidt de berekende  $\text{NH}_3$ -emissie<sup>2</sup> in de gridcel  $r$  aan bij gegeven input-emissie data  $Inp$ , modelspecificatie  $MS$ , gemeente-verrastering  $G_{25}$  (i.e. op basis van 25 x 25 m resolutie-nivo), en landgebruiksclassificatie via RS-informatie (Remote Sensing), zoals deze is vastgelegd in het LGN-bestand.

Ook hier heeft de modelspecificatie  $MS$  (d.w.z.  $[VS_m, VS_h, N_h]$ ) betrekking op de vigerende verrastering en classificatie, d.w.z. op  $G_{25}$  en  $RS$ , i.e. op 25 x 25 m gridcel-nivo.

- De *nominale* waarden voor bovengenoemde grootheden (d.w.z. input-emissie data, modelspecificaties, verrasterings-, en landgebruiksclassificatie opties) worden aangeduid met de superfix <sup>o</sup>.

De nominale input-emissie data  $Inp^o$  betreffen de gemeentelijke  $\text{NH}_3$ -emissie gegevens uit mest en de inwoneraantallen zoals deze door het LEI-DLO respectievelijk het CBS aan het RIVM worden geleverd. De nominale situatie bij de modelspecificaties komt neer op het hanteren van een verdeelsleutel van 1:1:1 voor het verdelen van mest (dierlijke mest en kunstmest) over de categorieën grasland, maïs en bouwgrond ( $VS_m^o = 1:1:1$ ); de emissiebijdrage van de huishoudens wordt uniform verdeeld over het *totale* gemeentelijke oppervlak, met een verdeelsleutel van 1:1 voor de urbane en niet-urbane gebieden ( $VS_h^o = 1:1$ ), waarbij een gemiddelde jaarlijkse  $\text{NH}_3$ -emissie gehanteerd wordt van 0.7 kg  $\text{NH}_3$  per inwoner (d.w.z.  $N_h^o = 0.7$ ). De aldus resulterende nominale modelspecificatie  $MS^{(o)}$  wordt gegeven door  $MS^{(o)} = [VS_m^o, VS_h^o, N_h^o]$ .

Bovendien zal een *alternatieve* modelspecificatie worden beschouwd, die genoteerd wordt als  $MS^{(a)}$ , en die gegeven is door (zie hoofdstuk 4)  $MS^{(a)} := [VS_m^o, VS_h^{(a)}, N_h^o]$ , waarbij  $VS_h^{(a)} := 1 : 0$  de alternatieve verdeelsleutel voor de huishoudelijke emissiebijdrage betreft. Deze verdeelsleutel geeft aan dat de huishoudelijke emissies enkel worden toebedeeld aan urbane gebieden, en niet aan niet-urbane gebieden. De alternatieve modelspecificatie  $MS^{(a)}$  onderscheidt zich dus enkel van de nominale  $MS^{(o)}$  door het feit dat de huishoudelijke emissie anders gealloceerd wordt. Allocatie van kunstmest en dierlijke mest blijft daarentegen dezelfde.

- De  $\text{NH}_3$ -emissieberekeningen die momenteel binnen de N-keten worden uitgevoerd betreffen bovengenoemde nominale situatie. De bijbehorende  $\text{NH}_3$ -emissies worden verkort weergegeven

<sup>2</sup>De bijdrage van industrieën en buitenland is hierin niet verdisconteerd, omdat het LAE deze niet bewerkt!! Deze bijdragen, die puntbronnen resp. oppervlaktebronnen betreffen, worden rechtstreeks bij de depositieberekeningen gebruikt als invoerdata.

<sup>3</sup>Bijv. een 5 x 5 km gridcel.

als  $y_{nom}(r)$ , i.e.:

$$y_{nom}(r) := y(r \mid Inp^{\circ}; MS^{\circ}; G_{500}^{\circ}; LBG^{\circ}) \quad (1)$$

Om de betrouwbaarheid van de *huidige* emissieberekeningen  $y_{nom}(r)$  te onderzoeken, zouden deze berekeningen idealiter met hun corresponderende *werkelijke* waarden, zeg  $y_{real}(r)$ , vergeleken moeten worden. Deze werkelijke waarden zijn echter *niet beschikbaar*. Wel kunnen er emissieberekeningen worden uitgevoerd op een *verfijnder* gridnivo (op basis van de 25 x 25 m rastercellen van de remote sensing beelden), die vervolgens worden doorvertaald (geaggregeerd) naar emissiecijfers op het gewenste gridcelnivo (5 x 5 km): d.w.z.  $y(r \mid Inp; MS; G_{25}; RS)$  wordt berekend.

Deze berekeningen maken gebruik van meer gedetailleerde informatie dan de huidige N-keten berekeningen  $y_{nom}(r)$  die gebaseerd zijn op het 500 x 500 m resolutie-nivo (zie hoofdstuk 2), en worden daarom verondersteld de werkelijkheid nauwkeuriger te beschrijven dan  $y_{nom}(r)$ . Deze verfijnde modelberekeningen  $y(r \mid Inp; MS; G_{25}; RS)$  worden opgevat als de *gepostuleerde werkelijkheid*, en de aldus berekende NH<sub>3</sub>-emissie voor de r-de gridcel wordt kort genoteerd als  $y_{pos}(r)$ . Hij fungeert als substituuut voor de onbekende  $y_{real}(r)$ . Betrouwbaarheidsanalyse wordt nu gebaseerd op het evalueren van de afwijking  $\varepsilon(r) := y_{nom}(r) - y_{pos}(r)$  tussen nominale modeluitkomst en gepostuleerde werkelijkheid, die een benadering is van de (onbekende) werkelijke afwijking  $\varepsilon^*(r) := y_{nom}(r) - y_{real}(r)$ .

Indien de remote sensing informatie (*RS*) en de input-emissie data (*Inp*) *foutvrij* zijn<sup>4</sup>, dan ligt  $y_{pos}(r)$ , en daarmee  $\varepsilon(r)$ , eenduidig vast. In situaties met *onzekerheid* daarentegen, kan  $y_{pos}(r)$  in wezen diverse mogelijke waarden (realisaties) aannemen, zeg  $y_{pos}(r; j)$ ,  $j = 1, \dots, N_{sam}$ , resulterend in evenzovele realisaties van mogelijke afwijkingen  $\varepsilon(r; j) = y_{nom}(r) - y_{pos}(r; j)$ . Statistische analyse van deze afwijkingen (gemiddelde, variantie, percentielen) levert nuttige informatie over de betrouwbaarheid van de emissieberekening in de N-keten. Deze informatie kan bijv. worden samengevat in termen van een *betrouwbaarheidsfactor* (zie [1]), die een dimensieloze maat geeft voor de discrepantie tussen de huidige berekening  $y_{nom}(r)$  en de mogelijke ‘werkelijke’ waarden  $y_{pos}(r; j)$ . De geschiktheid van deze maat om de daadwerkelijke afwijking tussen model en werkelijkheid uit te drukken hangt direct samen met de mate waarin deze werkelijkheid adequaat wordt weergegeven door de gepostuleerde waarden  $y_{pos}(r; j)$  (*validatie-aspect*). Gebruik en interpretatie van de betrouwbaarheidsfactor zal dus altijd *conditioneel* zijn m.b.t. deze *gepostuleerde werkelijkheid*, en kan slechts na adequate validatie van  $y_{pos}(r; \cdot)$  gerelateerd worden aan de (‘echte’) werkelijkheid  $y_{real}(r)$ .

De *betrouwbaarheidsfactor* is gedefinieerd als ( $y_{nom}(r)$  en  $y_{pos}(r; j)$  worden  $> 0$  verondersteld):

$$B(y_{nom}(r)) := \exp \left( - \sqrt{\frac{1}{N_{sam}} \sum_{j=1}^{N_{sam}} [\log(y_{nom}(r)) - \log(y_{pos}(r; j))]^2} \right) \quad (2)$$

Deze grootheid neemt waarden aan tussen 0 en 1, waarbij waarden in de buurt van 0 resp. 1 wijzen op een zeer geringe resp. grote betrouwbaarheid. In wezen drukt hij de grootte van de relatieve<sup>5</sup> afwijking  $\frac{y_{nom}(r)}{y_{pos}(r; j)}$  tussen model en ‘werkelijkheid’ uit op de logaritmische schaal,

<sup>4</sup>D.w.z. gelijk zijn aan hun nominale waarden  $RS^{\circ}, Inp^{\circ}$ .

<sup>5</sup>Een dergelijke quantificering in termen van een relatieve fout is met name zinvol in situaties waarbij de verdeling van de afwijkingen tussen model en werkelijkheid scheef naar rechts of links is, bijv. t.g.v. de niet-negativiteit van de betrokken grootheden  $y_{nom}, y_{pos}$ .

als wortel van een kwadratensom (Root Mean Square). Gebruik van deze factor sluit aan bij foutendiscussies in de context van statistiek ([8], [9]) en modellen ([15], [11]).

De aldus gedefinieerde betrouwbaarheidsmaat is schaalinvariant, en beschrijft de totale (relatieve) fout, zonder expliciet onderscheid te maken tussen een ‘systematisch’ en een ‘random’ deel. In de volgende stelling staan een aantal eigenschappen genoemd, die de interpretatie en het gebruik van deze maat verduidelijken.

**Stelling 1** Voor de betrouwbaarheidsfactor gelden de volgende eigenschappen:

- (a) De betrouwbaarheidsfactor is samengesteld uit een *bias* en een *variantie* component:

$$B(y_{nom}(r)) = \exp\left(-\sqrt{\beta_*^2(r) + \sigma_*^2(r)}\right) \quad (3)$$

waarbij de bias  $\beta_*(r)$  (‘systematisch’ deel) en variantie  $\sigma_*^2(r)$  (‘random’ deel) gelijk zijn aan:

$$\beta_*(r) := \left[ \log(y_{nom}(r)) - \overline{\log(y_{pos}(r; \cdot))} \right] \quad (4)$$

$$\sigma_*^2(r) := \frac{1}{N_{sam}} \sum_{j=1}^{N_{sam}} [\log(y_{pos}(r; j)) - \overline{\log(y_{pos}(r; \cdot))}]^2 \quad (5)$$

Hierbij is  $\overline{\log(y_{pos}(r; \cdot))}$  het sample gemiddelde van  $\log(y_{pos}(r; j))$ ,  $j = 1, \dots, N_{sam}$ :

$$\overline{\log(y_{pos}(r; \cdot))} := \frac{1}{N_{sam}} \sum_{k=1}^{N_{sam}} \log(y_{pos}(r; k)) \quad (6)$$

- (b) Bovenstaande ‘bias’- en ‘variantie’-component kunnen rechtstreeks worden uitgedrukt in termen van het *geometrisch gemiddelde*

$$\overline{y_{pos}(r; \cdot)}^{geom} := \left[ \prod_{j=1}^{N_{sam}} y_{pos}(r; j) \right]^{\frac{1}{N_{sam}}} = \exp(\overline{\log(y_{pos}(r; \cdot))}) \quad (7)$$

als

$$\beta_*(r) = \log \left[ \frac{y_{nom}(r)}{\overline{y_{pos}(r; \cdot)}^{geom}} \right] \quad (8)$$

$$\sigma_*^2(r) = \frac{1}{N_{sam}} \sum_{j=1}^{N_{sam}} \left( \log \left[ \frac{y_{pos}(r; j)}{\overline{y_{pos}(r; \cdot)}^{geom}} \right] \right)^2 \quad (9)$$

- (c) Indien  $y_{pos}(r; \cdot)$  lognormaal is, dan is het *68 %-betrouwbaarheidsinterval*, waarin met 68 % kans de ‘werkelijke’ waarde  $y_{pos}(r; j)$  ligt, gelijk aan:

$$\left[ \exp(-\sigma_*(r) - \beta_*(r)) \cdot y_{nom}(r) \quad , \quad \exp(+\sigma_*(r) - \beta_*(r)) \cdot y_{nom}(r) \right] \quad (10)$$

Dit interval is geformuleerd in termen van de nominale berekening  $y_{nom}(r)$ , maar kan ook in termen van het geometrisch gemiddelde<sup>6</sup> geformuleerd worden als:

$$\left[ \exp(-\sigma_*(r)) \cdot \overline{y_{pos}(r; \cdot)}^{geom} \quad , \quad \exp(+\sigma_*(r)) \cdot \overline{y_{pos}(r; \cdot)}^{geom} \right] \quad (11)$$

Het *95 %-betrouwbaarheidsinterval* wordt verkregen indien  $\sigma_*(r)$  in bovenstaande uitdrukkingen (10) en (11) vervangen wordt door  $2\sigma_*(r)$ .

<sup>6</sup>Voor lognormaal verdeelde  $y_{pos}(r; \cdot)$  is het geometrisch gemiddelde een schatting voor de mediaan  $\text{Med}[y_{pos}(r)]$ .

- (d) Indien  $y_{pos}(r; \cdot)$  lognormaal is, dan kunnen de uitdrukkingen voor de bias  $\beta_*(r)$  en de variantie  $\sigma_*^2(r)$  bij benadering worden uitgedrukt in termen van de mediaan  $\text{Med} [y_{pos}(r)]$  en de variatiecoëfficiënt<sup>7</sup>  $\text{VC} [y_{pos}(r)]$ :

$$\beta_*(r) \approx \log \left( \frac{y_{nom}(r)}{\text{Med} [y_{pos}(r)]} \right) \quad (12)$$

$$\sigma_*^2(r) \approx \log \left( 1 + (\text{VC} [y_{pos}(r)])^2 \right) \quad (13)$$

Indien bovendien  $\beta_*(r)$  verwaarloosbaar is t.o.v.  $\sigma_*(r)$ , en indien de variatiecoëfficiënt  $\text{VC} [y_{pos}(r)]$  klein is (bijv. kleiner dan .25), dan geldt er dat de betrouwbaarheidsfactor bij benadering gelijk is aan:

$$B(y_{nom}(r)) \approx \exp(-\sigma_*(r)) \approx 1 - \sigma_*(r) \approx 1 - \text{VC} [y_{pos}(r)] \quad (14)$$

- (e) Indien er slechts één realisatie is, d.w.z.  $N_{sam} = 1$ , dan vereenvoudigt de uitdrukking voor de betrouwbaarheidsfactor tot:

$$\begin{aligned} B(y_{nom}(r)) &:= \exp \left( -\sqrt{[\log(y_{nom}(r)) - \log(y_{pos}(r; 1))]^2} \right) \\ &= \text{Min} \left[ \frac{y_{nom}(r)}{y_{pos}(r; 1)}, \frac{y_{pos}(r; 1)}{y_{nom}(r)} \right] \end{aligned} \quad (15)$$

Deze drukt enkel de systematische afwijking (bias) uit. Van variabiliteit is in dit geval geen sprake.

- (f) Indien  $y_{pos}(r; \cdot)$  lognormaal is, en indien  $y_{nom}(r)$  tussen de klasse-grenzen  $[L; H]$  ligt ( $0 < L < H$ ), dan is de kans op *mis-classificatie*<sup>8</sup> op zijn minst gelijk aan:

$$\mathbb{P}_{mis} \geq 2 \cdot \mathbb{P} \left( X_{stnd} \geq \frac{\log(H/L)}{2 \cdot \sigma_*(r)} \right) \quad (16)$$

waarbij  $X_{stnd}$  een standaard-normaal verdeelde grootte is (i.e.  $N(0, 1)$ ).

■

Het bewijs van deze stelling staat vermeld in appendix A. De eigenschappen die in de stelling genoemd zijn, worden in het hieronder nader besproken:

- (1) De ‘bias’  $\beta_*(r)$  geeft informatie over de *systematische* afwijking tussen nominaal model en (geometrisch) gemiddelde gepostuleerde werkelijkheid, gemeten in relatieve zin, op logaritmische schaal. De ‘variantie’  $\sigma_*^2(r)$  weerspiegelt de intrinsieke onzekerheid (*random* fout) in de gepostuleerde werkelijkheid. De betrouwbaarheidsfactor  $B(y_{nom}(r))$  bevat beider informatie in *geaggregeerde vorm*, en het is *niet mogelijk* om enkel uit de waarde van  $B(y_{nom}(r))$  af te leiden wat de grootte van de ‘systematische’ en de ‘random’ component is.

<sup>7</sup>De variatiecoëfficiënt is gelijk aan de ratio tussen de standaarddeviatie en het gemiddelde.

<sup>8</sup>Mis-classificatie betekent in dit geval dat de gepostuleerde ‘werkelijkheid’  $y_{pos}(r; \cdot)$  buiten die klasse-grenzen ligt.



- (2) Het ‘betrouwbaarheidsinterval’ (10) ligt niet noodzakelijkerwijze rond de nominale waarde  $y_{nom}(r)$ . Dit geldt enkel indien  $|\beta_*(r)| \leq \sigma_*(r)$ . Uit uitdrukking (11) blijkt daarentegen wel dat het interval (scheef-symmetrisch) gecentreerd is rond het geometrisch gemiddelde  $\overline{y_{pos}(r; \cdot)}^{geom}$  van de gepostuleerde werkelijkheid. Uit dit resultaat volgt tevens dat bovenstaand ‘betrouwbaarheidsinterval’ totaal *niet afhangt* van de waarde  $y_{nom}(r)$ , de nominale modeluitkomst!!! Dit is ook logisch, want de definitie van het betrouwbaarheidsinterval richt zich enkel op de samples  $y_{pos}(r; j)$ , en niet op het nominale model  $y_{nom}(r)$ .

Door nu de ligging van de nominale modeluitkomst  $y_{nom}(r)$  t.o.v. het geometrisch gemiddelde en het 68 %-betrouwbaarheidsinterval te bestuderen, krijgen we nader inzicht in de ‘systematische’ en ‘random’ afwijking tussen model en (gepostuleerde) ‘werkelijkheid’. Deze specifieke informatie is niet rechtstreeks voorhanden in de waarde van de betrouwbaarheidsfactor ( $B(y_{nom}(r))$ ) an sich, en kan nuttige suggesties opleveren ter verbetering van de modelvoorspellingen. Zie figuur 1 waarin drie verschillende situaties uitgebeeld staan met dezelfde betrouwbaarheidsfactor, maar met een duidelijk andere rol voor de systematische en random afwijkingen.

- (3) Indien de bias  $\beta_*(r)$  verwaarloosbaar<sup>9</sup> is t.o.v. de standaard deviatie  $\sigma_*(r)$ , dan vallen het geometrisch gemiddelde en de nominale uitkomst nagenoeg samen, en kan bovenstaand 68 %-betrouwbaarheidsinterval bij benadering vereenvoudigd worden tot:

$$[ B(y_{nom}(r)) \cdot y_{nom}(r) , [B(y_{nom}(r))]^{-1} \cdot y_{nom}(r) ] \quad (17)$$

en het 95 %-betrouwbaarheidsinterval tot:

$$[ [B(y_{nom}(r))]^2 \cdot y_{nom}(r) , [B(y_{nom}(r))]^{-2} \cdot y_{nom}(r) ] \quad (18)$$

Dit levert een rechtstreekse interpretatie van de betrouwbaarheidsfactor in termen van een betrouwbaarheidsinterval rond de nominale waarde. Zo geeft een betrouwbaarheidsfactor van bijv.  $B(\cdot) = .5$  aan dat de opgegeven (nominale) waarde met een kans van 32 % meer dan een factor 2 ( $= [B(\cdot)]^{-1}$ ) afwijkt van de (gepostuleerde) werkelijke waarde, en met kans 5 % meer dan een factor 4 ( $= [B(\cdot)]^{-2}$ ).

Deze interpretatie van de betrouwbaarheidsfactor als ‘de factor die men ernaast kan zitten’ is strikt genomen slechts geldig indien de bias verwaarloosbaar is!!

Indien dit niet het geval is, dan schiet de betrouwbaarheidsfactor te kort als duidelijke maatstaf voor de betrouwbaarheid, en kan alleen het betrouwbaarheidsinterval als voldoende indicatief gehanteerd worden; zie ook het voorgaande item (2).

- (4) Indien  $y_{pos}(r; \cdot)$  bij benadering normaal verdeeld is, dan kan  $[\mu - \sigma, \mu + \sigma]$  als ook 68 %-‘betrouwbaarheidsinterval’ fungeren, waarbij  $\mu$  en  $\sigma$  het geschatte gemiddelde en de standaarddeviatie van  $y_{pos}(r; j)$  zijn. Dit interval is een alternatief voor het interval uit (10) of (11) dat gebaseerd is op lognormaliteit van  $y_{pos}(r; \cdot)$ .

Indien de variatiecoëfficiënt  $VC[y_{pos}(r)]$  klein is (bijv. kleiner dan .25), dan kan  $y_{pos}(r; \cdot)$  bij benadering ook als lognormaal verdeeld worden opgevat, en vallen bovengenoemde intervallen nagenoeg samen. Immers het geometrisch gemiddelde is dan bij benadering gelijk aan de mediane waarde, en dus aan het gemiddelde  $\mu$ , terwijl volgens (13)

<sup>9</sup>Indien  $\beta_*(r) = 0$ , dan zijn relatieve fouten naar boven even waarschijnlijk als naar onder. Onder veronderstelling van lognormaliteit komt dit er in wezen op neer dat de werkelijkheid even vaak wordt onderschat als overschat met het nominale model.

$\exp(\pm\sigma_*(r)) \approx 1 \pm \sigma_*(r) \approx 1 \pm \text{VC}[y_{pos}(r)] \approx 1 \pm \frac{\sigma}{\mu}$ . Het interval (11) wordt dus bij benadering gelijk aan  $[\mu(1 - \frac{\sigma}{\mu}), \mu(1 + \frac{\sigma}{\mu})] = [\mu - \sigma, \mu + \sigma]$ .

- (5) Uit (16) volgt dat de kans op mis-classificatie minstens  $2 \cdot \mathbb{P}(X_{std} \geq f)$  is ( $f > 0$ ), indien de ratio  $\frac{H}{L}$  tussen de boven- en ondergrens van de klasse kleiner dan of gelijk aan  $[\exp(\sigma_*(r))]^{2 \cdot f}$  is. Evaluatie van deze uitdrukking voor  $f = \frac{1}{2}, 1, 2$  levert:

$$\frac{H}{L} \leq [\exp(\sigma_*(r))] \Rightarrow \mathbb{P}_{mis} \geq .6 \quad (19)$$

$$\frac{H}{L} \leq [\exp(\sigma_*(r))]^2 \Rightarrow \mathbb{P}_{mis} \geq .3 \quad (20)$$

$$\frac{H}{L} \leq [\exp(\sigma_*(r))]^4 \Rightarrow \mathbb{P}_{mis} \geq .05 \quad (21)$$

Omdat  $\exp(\sigma_*(r)) \approx \exp(\sqrt{\sigma_*^2(r) + \beta_*^2(r)}) = [B(y_{nom}(r))]^{-1}$ , indien de bias-invloed  $\beta_*^2$  verwaarloosbaar is t.o.v. de variantie  $\sigma_*^2$ , geeft dit rechtstreeks informatie over hoe ver de klasse-grenzen uit elkaar gekozen dienen te worden om te voorkomen dat er een grote kans op mis-classificatie bestaat:

$$\frac{H}{L} \leq [B(y_{nom}(r))]^{-1} \Rightarrow \mathbb{P}_{mis} \geq .6 \quad (22)$$

$$\frac{H}{L} \leq [B(y_{nom}(r))]^{-2} \Rightarrow \mathbb{P}_{mis} \geq .3 \quad (23)$$

$$\frac{H}{L} \leq [B(y_{nom}(r))]^{-4} \Rightarrow \mathbb{P}_{mis} \geq .05 \quad (24)$$

Dit resultaat kan als *ruwe richtlijn* fungeren voor de keuze van de klassebreedtes bij de presentatie van kaarten, indien de berekeningen met een bepaalde betrouwbaarheid zijn uitgevoerd. Bijv. de klassegrenzen dienen minstens met een factor  $[B(y_{nom}(r))]^{-2}$  op te lopen, om te voorkomen dat de kans op mis-classificatie groter is dan 30 %. Bij een betrouwbaarheidsfactor van  $\frac{1}{2}$  betekent dit dat de klassegrenzen minstens een factor 4 uit elkaar dienen te liggen. Bedenk dat dit strikt genomen slechts geldig is indien er voldaan wordt aan bovenvermelde voorwaarden, d.w.z.  $y_{pos}(r; \cdot)$  is lognormaal en de systematische afwijking is verwaarloosbaar is t.o.v. de random afwijking.

## 4 Onzekerheidsspecificaties voor foutenbronnen

Het evalueren van de betrouwbaarheid van de emissieberekeningen vereist allereerst dat de verschillende foutenbronnen die bij dit rekenproces een rol kunnen spelen, geïnventariseerd en gekwantificeerd worden. Dit is het onderwerp van dit hoofdstuk.

De fouten die bij de NH<sub>3</sub>-emissieberekeningen voor 5 x 5 km gridcellen kunnen optreden, zijn onder te verdelen in drie categorieën:

- (a) **Specificatiefouten in de basisgegevens** die bij de berekening gebruikt worden, i.e. de gemeentelijke NH<sub>3</sub>-emissiecijfers uit dierlijke mest en kunstmest (LEI-DLO), en de gemeentelijke NH<sub>3</sub>-emissies uit huishoudens (CBS).

De emissiebijdrage van dierlijke mest wordt bepaald aan de hand van simulaties met mest- en ammoniakmodellen van het LEI-DLO, waarvan de invoer bestaat uit gegevens van de jaarlijkse Landbouwtelling die op bedrijfsniveau beschikbaar zijn, gecombineerd met informatie betreffende landelijk gemiddelde excretiefactoren, emissiefactoren, penetratiegraden etc. Deze laatste informatie is gebaseerd op uitgangspunten van het RIVM, en is afgeleid uit gegevens van verschillende bronnen (LEI-DLO, CBS, IKC etc.).

De bijdrage van kunstmest is berekend door het stikstof kunstmestgebruik per gemeente (berekend door het LEI-DLO) te vermenigvuldigen met een vervluchtigingspercentage en een omrekeningsfactor.

De emissiebijdrage uit huishoudens tenslotte wordt ingeschat door een emissiefactor van 0.7 kg NH<sub>3</sub> per inwoner per jaar te hanteren, en door deze eenvoudigweg te vermenigvuldigen met de gemeentelijke inwoneraantallen uit de CBS bevolkingsstatistiek.

Door de gehanteerde berekeningsmethodiek (zie [2] voor nadere informatie), en door onnauwkeurigheden c.q. onvolkomenheden in de hiervoor benodigde gegevens, kunnen de opgegeven waarden voor de gemeentelijke emissiecijfers afwijken van de werkelijkheid (specificatiefouten).

- (b) **Verrasterings- en classificatiefouten** van 500 x 500 m arealen: Zoals beschreven in hoofdstuk 2, worden bij de berekeningen de gemeentekaarten verrasterd naar 500 x 500 m arealen, die aansluiten bij de resolutie van de door het LBG samengestelde landgebruikskaart. Elk van deze arealen wordt 'toegewezen' aan die gemeente die er, qua oppervlak, het meeste deel van uit maakt (d.w.z. de *dominante* gemeente). Hierdoor kunnen afwijkingen ontstaan in de oppervlakte toekenning aan gemeenten, en in de toewijzing van inwoners aan de 500 x 500 m arealen.

Aan de hand van de LBG-landgebruikskaart wordt per areaal het dominante landgebruikstype bepaald, en wordt vervolgens de ammoniakemissie berekend uit de gemeentelijke emissiecijfers op basis van de procedure beschreven in hoofdstuk 2. De fouten die bij deze 'verdeling' van de gemeentelijke emissies over de arealen ontstaan, hebben voor een groot deel te maken met een foutieve toewijzing van oppervlakten aan (dominante) landgebruikstypen en van inwoneraantallen aan urbane gebieden. De oorzaak hiervan is velerlei: enerzijds zitten er *classificatiefouten*<sup>10</sup> in de Remote Sensing

<sup>10</sup>Ten gevolge van eindige resolutie, sampling-, detectie-, interpolatie-, meet-, typeringsfouten etc. Bedenk dat de classificatie van een 25 x 25 m pixel vaak niet eenduidig en scherp is, enerzijds omdat het pixel niet totaal bedekt hoeft te zijn met land van één specifiek type, en anderzijds omdat de classificatie niet volledig strikt is (wanneer classificeert men land bijv. als type A, en wanneer als type B?).

informatie die de basis vormt van het LGN-bestand waaruit de landgebruikskaart van het LBG bepaald is. Anderzijds treden er *aggregatiefouten* op ten gevolge van opschaling/aggregatie van deze verfijnde 25 x 25 m informatie naar 500 x 500 m nivo. Ook worden bij herclassificatie van de LGN-informatie door het LBG mogelijkerwijze *interpretatie- en rubriceringsfouten* gemaakt, bijv. bij de update van deze informatie om veranderingen in het landgebruik te verdisconteren die sinds de uitvoering van de Remote Sensing studie (1986-1988) hebben plaatsgevonden.

- (c) **Allocatiefouten:** De fouten die bij de berekening van de NH<sub>3</sub>-emissie in 5 x 5 km gridcellen verder geïntroduceerd worden zijn het gevolg van de gehanteerde aannames en vereenvoudigingen bij allocatie van de gemeentelijke NH<sub>3</sub>-emissies aan landgebruikstypen. Hierbij wordt de NH<sub>3</sub>-emissie uit mest (dierlijk en kunstmest) gelijkmatig verdeeld over de gras-, maïs- en bouwland- arealen van de gemeente, zonder differentiatie naar mestsoort, vorm van mestaanwending, gewastype, grondsoort<sup>11</sup>. De huishoudelijke NH<sub>3</sub>-emissies worden homogeen over alle arealen van de gemeente verdeeld. De uiteindelijke sommatie van de emissiecijfers uit alle 500 x 500 m arealen per 5 x 5 km gridcel voegt echter geen extra fouten toe.

De aard van bovenstaande fouten is verschillend. Zo kunnen de specificatiefouten in de basisgegevens i.h.a. goed in een probabilistische context worden weergegeven. Bij verrasterings- en classificatiefouten is dat minder het geval, omdat verrastering en classificatie de toekenning van één discreet attribuut betreft (bijv. gemeente  $i_o$ ; landgebruikstype  $j_o$ ) aan de diverse 500 x 500 m arealen. De fout die bij zo'n toekenning gemaakt wordt, kan *in principe* gekwantificeerd worden door per areaal de fractie van de *werkelijke* oppervlakte van gemeente  $i_o$  of landgebruikstype  $j_o$  te vergelijken met de *toegekende* fractie (zie appendix B). Dit veronderstelt echter voorkennis over de werkelijke oppervlakte, die zelden voorhanden is. In de praktijk kan deze oppervlakte hoogstens worden ingeschat; de onzekerheidsaspecten van deze schatting kunnen in veel gevallen probabilistisch, d.w.z. via kansverdelingen, beschreven worden. Complicerende factor hierbij is echter dat de schattingsfouten onderling gecorreleerd zijn, omdat de totaaloppervlakte per areaal vast ligt. Bovendien kunnen er allerlei correlaties optreden tussen naburige arealen, hetgeen een goede probabilistische beschrijving bemoeilijkt.

Allocatiefouten zijn doorgaans nog moeilijker adequaat te specificeren, omdat ze rechtstreeks refereren naar veronderstellingen die ten grondslag liggen aan het berekeningsproces zelf. Veelal stelt men zich in de praktijk tevreden met het beschouwen van enkele alternatieve allocatie opties (bijv. andere verdeelsleutels), en met het analyseren van de gevolgen van deze alternatieve keuzen voor de berekeningen.

Een exacte bepaling van bovenstaande fouten is in de regel onmogelijk omdat de 'werkelijke' waarden van de betrokken grootheden (bijv. de gemeentelijke NH<sub>3</sub>-emissiedata) veelal onbekend is. Bovendien is de *onzekerheid* over de grootte van deze fouten in het algemeen slechts bij benadering te *kwantificeren*. Een nauwkeurige specificatie vereist bijv. immers een volledige onzekerheidsanalyse van alle berekeningen die verricht worden om de gemeentelijke NH<sub>3</sub>-emissiecijfers te bepalen, en is voorlopig onhaalbaar, gezien de beperkte tijd, menskracht en reken capaciteit die hiervoor beschikbaar was, en gezien het feit dat deze berekeningen grotendeels buiten het RIVM worden uitgevoerd. Daarom is in dit stadium volstaan met een

<sup>11</sup>De *gemeentelijke* emissiecijfers van het LEI-DLO zijn in eerste instantie wel bepaald op basis van zo'n differentiatie. Gezien het betrouwbaar karakter van de hiervoor benodigde gegevens, is deze informatie echter niet beschikbaar voor het LAE, en kunnen de gemeentelijke cijfers niet op adequate wijze 'gedisaggregeerd' worden bij de toedeling aan de 500 x 500 m arealen.

*ruwe schatting* van de onzekerheid, hierbij een aantal drastische aannames en vereenvoudigingen hanterend. Gebruik makend van de informatie uit appendix C, werd tot de volgende keuzen besloten:

**ad (a): Fouten in de basisgegevens:**

De emissiedata  $Inp$  die als input voor de berekeningen gebruikt worden, betreffen de gemeentelijke  $NH_3$ -emissiecijfers die afkomstig zijn van o.a. het LEI-DLO en het CBS. Deze zijn uitgesplitst naar de bijdrages van dierlijke mest ( $E_d$ ), kunstmest ( $E_k$ ) en huishoudens ( $E_h$ ) (d.w.z.  $Inp = [E_d, E_k, E_h]$ ). De onzekerheid in de emissiebijdragen van deze bronnen is gespecificeerd in termen van een *spreiding of relatieve 'bandbreedte'* rondom hun gerapporteerde, i.e. *nominale*, waarden. Deze keuze impliceert in wezen dat *systematische* fouten afwezig worden verondersteld in de gemeentelijke emissiedata, en dat fouten enkel een *willekeurig* (random) karakter hebben. De exacte specificatie van deze willekeurige fouten wordt in het hiernavolgende besproken:

*(1) Dierlijke mest per gemeente:*

Mogelijke afwijkingen tussen de door het LEI-DLO berekende emissiecijfers en hun echte (onbekende) waarden zijn het gevolg van de gehanteerde berekeningsmethodiek. Enerzijds kunnen de cijfers uit de Landbouwtelling die hierbij als invoer dienen van de mest- en ammoniakmodellen van het LEI-DLO onvolledig of onnauwkeurig zijn. Bovendien zijn de gehanteerde excretiefactoren, emissiefactoren etc. 'landelijk gemiddelde waarden' die sterk kunnen afwijken van waarden op lokaal (i.e. regionaal, gemeentelijk en bedrijfs) nivo. Immers bij de berekening van deze factoren worden allerlei (model)veronderstellingen gehanteerd betreffende voersamenstelling, N-gehalte in voer, mestproductie, mestopslag, die op landelijk nivo redelijk kloppen, maar op lokaal nivo soms forse verschillen laten zien tussen modelberekeningen en werkelijkheid. Ook de gegevens over mestaanwending, mesttransport kunnen grote fouten vertonen, omdat het ruimtelijk patroon gesimuleerd wordt met een transport model dat slechts een benadering van de werkelijkheid geeft.

De uiteindelijke spreiding in de  $NH_3$ -emissiecijfers uit dierlijke mest zal dus van vele factoren afhangen, bijv. mestsoort (diercategorie), mestproductie, mestopslag, mestaanwending (gewas/grondsoort combinatie; uitrijperiode), mesttransport etc. Met name de jaren 1992 tot en met 1994 kenmerken zich door een gewijzigd en gevarieerd beeld van mestaanwending ten gevolge van de invoering van een verscherpte mestregelgeving waarbij geleidelijk een overgang van emissierijke naar emissie-arme aanwending plaatsvindt. In deze overgangsperiode is de mestaanwending gedifferentieerd naar grondsoort en regio, en zijn de gemeenten daardoor niet altijd eenduidig te classificeren als emissie-rijk of emissie-arm. Er is bovendien onvoldoende extra informatie beschikbaar om adequate cijfers af te leiden voor de daadwerkelijke spreiding die hierdoor in de berekende emissiecijfers resulteert. Dit probleem zal echter grotendeels verdwenen zijn na volledige invoering van de nieuwe mestregelgeving. Daarom is besloten om de berekeningen ter vaststelling van de spreiding (bandbreedte) voor 1990 uit te voeren, een jaar waarin de mestaanwending emissie-rijk is (oppervlakkige aanwending). Bij gebrek aan voldoende adequate gegevens voor de andere jaren, zal deze bandbreedte ook voor de overige jaren worden toegepast, en dus ook voor het hier gerapporteerde jaar 1992.

Om tot een adequate schatting van deze bandbreedte te komen, is de dierlijke mest in drie ‘mestsoorten’ onderscheiden, te weten mest afkomstig van rundvee, varkens en pluimvee. Voor elk van deze soorten is een afzonderlijke ‘bandbreedte’ gehanteerd, die de (nationaal gemiddelde) spreiding in de jaarlijkse emissie per diersoort weerspiegelt. Deze bandbreedte is uitgedrukt als percentage van de gehanteerde nominale waarde (i.e. de berekende gemiddelde waarde), en geeft de marge aan waarbinnen een groot deel (d.w.z. 68 %, overeenkomend met de spreiding (standaarddeviatie) rond het gemiddelde) van alle emissievariaties valt. Hij is bepaald op basis van berekeningen waarbij de belangrijkste factoren die de  $\text{NH}_3$ -emissie beïnvloeden gevarieerd zijn, en blijkt gelijk te zijn aan respectievelijk 25 %, 20 % en 30 %, of te wel<sup>12</sup>  $\sigma_{d,r} = .25$ ;  $\sigma_{d,v} = .2$ ;  $\sigma_{d,p} = .3$ , voor rundvee (r), varkens (v) en pluimvee (p). Zie opmerking 5 in appendix C voor een uitgebreide toelichting van de totstandkoming van deze cijfers.

De bandbreedte  $\sigma_d(i)$  van de totale dierlijke  $\text{NH}_3$ -emissie per  $i$ -de gemeente wordt vervolgens uit deze afzonderlijke bandbreedtes  $\sigma_{d,r}, \sigma_{d,v}, \sigma_{d,p}$  afgeleid<sup>13</sup> op basis van de relatieve bijdragen van de afzonderlijke diersoorten aan de totale dierlijke  $\text{NH}_3$ -emissie per gemeente (zie appendix C):

$$\sigma_d(i) := \left( \sum_x [f_{d,x}^o(i) \cdot \sigma_{d,x}]^2 \right)^{\frac{1}{2}} \quad (25)$$

Hierbij loopt de som  $\sum_x$  over  $x = r, v, p$ , d.w.z. over de diersoorten rundvee, varkens en pluimvee. De weegfactoren  $f_{d,x}^o(i)$  geven aan welk deel van de dierlijke  $\text{NH}_3$ -emissie in gemeente  $i$  afkomstig is van diersoort  $x$ . De aan het RIVM geleverde gemeentelijke  $\text{NH}_3$ -emissiecijfers van het LEI-DLO bevatten echter geen directe informatie over deze fracties  $f_{d,x}^o(i)$ , omdat ze het resultaat zijn van een optelsom waarbij de differentiatie naar diersoort verloren<sup>14</sup> is gegaan. Daarom is besloten om de relatieve bijdragen op een andere wijze in te schatten, namelijk aan de hand van het *aantal dieren* per diersoort, rekenend met een nationaal gemiddelde  $\text{NH}_3$ -emissie  $P_x$  per dier uit diersoort  $x$ , d.w.z.:

$$f_{d,x}^o(i) := \frac{[N_{d,x}(i) \cdot P_x]}{\sum_x [N_{d,x}(i) \cdot P_x]} \quad (26)$$

waarbij  $N_{d,x}(i)$  het aantal dieren uit diersoort  $x$  is voor gemeente  $i$ . Deze werkwijze houdt o.a. in dat de invloed van getransporteerde mest niet geheel adequaat wordt meegenomen. Omdat deze landelijk gezien slechts zo’n 10 % tot de dierlijke  $\text{NH}_3$ -emissie bijdraagt wordt deze verwaarlozing acceptabel geacht<sup>15</sup>. Ook wordt de onzekerheid ten gevolge van mestaanwending niet expliciet meebeschoofd; deze informatie is moeilijk te traceren, en is bovendien reeds gedeeltelijk in de hiervoor

<sup>12</sup>  $\sigma_{d,x}$  duidt de spreiding (standaarddeviatie) aan van de emissie uit dierlijke mest, afkomstig van diersoort  $x$ .

<sup>13</sup> Verondersteld wordt dat de onzekerheid in de bijdragen van de diverse mestsoorten onderling onafhankelijk is.

<sup>14</sup> In de nieuwe versie van het ammoniakmodel, die gemaakt is ten behoeve van het RIVM, worden de emissiecijfers ook naar diersoort aangeleverd. In dat geval is het wel mogelijk om het relatieve aandeel per diersoort per gemeente rechtstreeks te bepalen.

<sup>15</sup> Bij regio’s in Zeeland en in de Flevopolder levert de emissie als gevolg van het aanwenden van geïmporteerde mest, de belangrijkste bijdrage. Voor deze regio’s is bovengenoemde verwaarlozing dus strikt genomen inacceptabel.

genoemde bepaling van de bandbreedtes  $\sigma_{d,x}$  voor de afzonderlijke diersoorten meegenomen. Door het gebruik van een nationaal gemiddelde  $NH_3$ -emissie  $P_x$  per diersoort  $x$  wordt er tevens impliciet vanuit gegaan dat de onderverdeling van diersoort  $x$  in subcategorieën<sup>16</sup> geen wezenlijk verschillend beeld laat zien per gemeente.

Bovengenoemde berekeningswijze leidt in principe tot bandbreedtes die per gemeente verschillen, omdat het aantal dieren  $N_{d,x}(i)$  per gemeente zal verschillen. Hij suggereert een mate van nauwkeurigheid, die niet gewettigd lijkt wegens de vele aannames en verwaarlozingen (zie appendix C). Daarom is uiteindelijk besloten om de bandbreedtes niet per gemeente te laten variëren, maar enkel *per provincie*, door uit te gaan van de landbouwdiertellingen per provincie, in plaats van de diertellingen per gemeente. Deze keuze is mede ingegeven door het feit dat deze cijfers eenvoudig beschikbaar waren (zie [7]). Dit resulteert in de bandbreedtes die in tabel 2 vermeld staan (zie appendix C). Deze waarden worden aan de diverse gemeenten uit de betreffende provincies toegekend. Het feit dat deze bandbreedte in een aantal gevallen kleiner is dan 20 % wordt veroorzaakt door de veronderstelde onafhankelijkheid van de emissiebijdragen van de afzonderlijke diersoorten, waardoor hun effecten elkaar deels opheffen.

Provincie	% bijdrage Rundvee	% bijdrage Varkens	% bijdrage Pluimvee	Bandbreedte in %
Groningen	80 %	11 %	9 %	20 %
Friesland	91 %	4 %	5 %	23 %
Drenthe	74 %	15 %	11 %	19 %
Overijssel	64 %	30 %	6 %	17 %
Gelderland	59 %	32 %	9 %	16 %
Utrecht	69 %	27 %	4 %	18 %
N-Holland	91 %	5 %	4 %	23 %
Z-Holland	81 %	15 %	4 %	21 %
Zeeland	69 %	18 %	13 %	18 %
N-Brabant	41 %	50 %	9 %	15 %
Limburg	33 %	49 %	18 %	14 %
Flevoland	84 %	8 %	8 %	21 %

Tabel 2: De relatieve bijdrage van de diersoorten tot de dierlijke  $NH_3$ -emissie, en de daaruit resulterende bandbreedte per provincie voor het jaar 1992

De *variatie*  $E_d(i)$  van de gemeentelijke dierlijke  $NH_3$ -emissie rond zijn door het LEI-DLO aangeleverde nominale waarde  $E_d^o(i)$  wordt normaal verdeeld verondersteld<sup>17</sup>, met een spreiding  $\sigma_d(i)$  die gegeven is door de bandbreedtes uit tabel 2:

$$E_d(i) := E_d^o(i) \cdot [1 + g_d(i) \cdot \sigma_d(i)] \quad (27)$$

<sup>16</sup> Bijv. de diersoort rundvee wordt onderverdeeld in melkvee en vleesvee, die eigenlijk een verschillende gemiddelde emissiefactor hebben.

<sup>17</sup> Deze berekeningswijze komt in feite neer op de veronderstelling dat zo'n 68 % van alle emissievariaties binnen de relatieve bandbreedte  $[(1 \pm \sigma_d(i)) \cdot E_d^o(i)]$  vallen, en dat ze normaal verdeeld zijn. De systematische fout wordt dus nul verondersteld, terwijl de random fout normaal verdeeld is met variantie  $[\sigma_d(i) \cdot E_d^o(i)]^2$ .

Hierbij is  $E_d^0(i)$  de nominale  $\text{NH}_3$ -emissie uit dierlijke mest voor gemeente  $i$ , en is  $E_d(i)$  de 'mogelijk' werkelijke waarde (realisatie).  $g_d(i)$  is een standaardnormaal verdeelde grootheid ( $N(0,1)$ ) die de variatie karakteriseert, welke plaatsvindt met een bandbreedte/spreiding  $\sigma_d(i)$ . Deze spreiding is volgens tabel 2 enkel afhankelijk van de provincie. Omdat  $g_d(i)$  in wezen alle waarden tussen  $-\infty$  en  $+\infty$  kan aannemen, laat uitdrukking (27) toe dat er negatieve emissies kunnen optreden. De kans hierop zal echter verwaarloosbaar zijn, omdat in de huidige toepassing  $\sigma_d(i)$  kleiner is dan .25 (zie tabel 2), terwijl de kans uiterst gering is dat  $g_d(i)$  waarden aanneemt die kleiner zijn dan -4.

Tenslotte moet nog beslist worden hoe bovengenoemde emissievariëaties uit (27) bij verschillende gemeenten met elkaar samenhangen. Gekozen is om alle gemeenten binnen *eenzelfde provincie* op *identieke* wijze te variëren: d.w.z. indien de emissie in gemeente A  $x$  % hoger is, dan zal ook de emissie in gemeente B  $x$  % hoger zijn ('volledige correlatie' tussen de fouten). De variaties bij gemeenten uit *verschillende provincies* worden daarentegen *ongecorreleerd* verondersteld<sup>18</sup>. Dus  $g_d(i) = g_d(j)$  indien gemeente  $i$  en  $j$  binnen dezelfde provincie liggen, terwijl  $g_d(i)$  en  $g_d(j)$  onafhankelijk zijn indien dit niet het geval is. Deze keuzes lijken tamelijk arbitrair, maar zijn ingegeven door het feit dat informatie over de daadwerkelijke samenhang tussen de gemeentelijke emissieonzekerheden nagenoeg totaal ontbreekt. Uiteraard zijn in deze omstandigheden veel alternatieven mogelijk, variërend van een volledige onafhankelijkheid tot een volledige afhankelijkheid van gemeentelijke emissievariëaties, maar gekozen is voor bovenstaande optie die zich ergens tussen deze twee uitersten in bevindt, in de hoop hiermee een enigszins acceptabel beeld van de realiteit te bereiken. Deze benadering kan pas echt op waarde worden ingeschat zodra er meer informatie voorhanden is over de werkelijke relatie tussen de onzekerheden in de gemeentelijke cijfers. In appendix D wordt op basis van theoretische beschouwingen een impressie gegeven van de invloed van de inter-gemeentelijke afhankelijkheidsstructuur op de onzekerheid in de  $\text{NH}_3$ -emissies. De resultaten duiden erop dat de emissie-variëatie/onzekerheid inderdaad sterk van deze facetten kan afhangen. Mogelijkerwijze kan een meer realistisch en gedifferentieerd beeld verkregen worden door de (gemeentelijke) cijfers op andere wijze te groeperen of te relateren dan puur administratief geografisch, zoals bij de huidige provincie-gebonden specificatie het geval is. Een indeling in bijv. ecodistricten, waarbij rekening gehouden kan worden met specifieke afhankelijkheden (m.b.t. grondsoort, landgebruik) zou bij een vervolgstudie de voorkeur verdienen.

(II) *Kunstmest per gemeente:*

De nominale  $\text{NH}_3$ -emissie ten gevolge van kunstmesttoepassing wordt berekend als het produkt van de hoeveelheid toegediende kunstmest (berekend met LEI-DLO modellen) en een vervluchtigingspercentage (zie [2]). Ook bij deze berekening spelen modelveronderstellingen een grote rol. Op landelijk nivo kloppen de cijfers redelijk, maar op lokaal nivo kunnen nog grote afwijkingen optreden tussen modelberekeningen en werkelijkheid.

<sup>18</sup>Idealiter zouden de emissievariëaties zodanig dienen te zijn, dat het totaal over Nederland nagenoeg hetzelfde is als in de nominale berekeningen. Deze extra beperking is echter bij de uitgevoerde analyses niet meegenomen. Afwijkingen treden op die kunnen variëren tussen -15% en +15%.



Analoog aan dierlijke mest, kan ook de gemeentelijke  $\text{NH}_3$ -bijdrage van kunstmest worden gevarieerd rond zijn nominale waarde:

$$E_k(i) := E_k^o(i) \cdot [1 + g_k(i) \cdot \sigma_k(i)] \quad (28)$$

Hierbij refereert het suffix  $k$  naar kunstmest. De keuze van de variatie  $g_k(i)$  en de 'bandbreedte'  $\sigma_k(i)$  kan bijv. weer afhankelijk gesteld worden van de provincies. Een bijkomend probleem is echter dat de kunstmestgift afhankelijk is van de dierlijke mestgift (d.w.z.  $g_k(i)$  en  $g_d(i)$  zullen samenhangen). Deze samenhang is een gevolg van de opgelegde mestregelgeving, en de preciese grootte daarvan wordt bepaald door de mate waarin deze normen en richtlijnen in de praktijk worden opgevolgd. Omdat dit in onvoldoende bekend is, is besloten om de specificatie van de *kunstmest-emissievariaties volledig buiten de onzekerheidsberekeningen te laten* (d.w.z.  $\sigma_k(i) = 0$ )!! Verwacht wordt dat het effect hiervan op de betrouwbaarheidsanalyse slechts gering is, omdat kunstmest op landelijk nivo slechts zo'n 5 % bijdraagt aan de  $\text{NH}_3$ -emissie, terwijl dierlijke mest 88 % voor haar rekening neemt (zie tabel 1). Op lokaal nivo kan deze bijdrage echter wel aanzienlijk zijn (zie figuur 4), en is deze verwaarlozing minder geoorloofd.

(III) *Huishoudens per gemeente:*

Net als bij de kunstmest bijdrage, zal ook de *huishoudelijke bijdrage niet bij de onzekerheidsbeschouwingen betrokken worden*, mede omdat de huishoudens landelijk<sup>19</sup> gezien ook slechts zo'n 5 á 6 % bijdragen tot de  $\text{NH}_3$ -emissie. Bovendien is het moeilijk om op een adequate wijze de onzekerheid te specificeren van de huishoudens, omdat hun invloed op een sterk geaggregeerde wijze beschreven is, namelijk in de vorm van een gemiddelde jaarlijkse  $\text{NH}_3$ -emissie  $N_h^o$  van 0.7 kg  $\text{NH}_3$  per inwoner.

Een aspect dat wel in de analyses wordt meegenomen, is de *wijze waarop* de bijdrage van de huishoudens wordt gealloceerd; zie ook het hiernavolgende item (c) betreffende allocatiefouten. Momenteel gebeurt dit door de huishoudelijke  $\text{NH}_3$ -emissie per gemeente te berekenen op basis van de inwoneraantallen, en vervolgens deze bijdrage homogeen over het *totale gemeenteoppervlak* te verdelen. Bestudeerd zal worden of een meer realistische aanpak, waarbij de huishoudelijke bijdrage enkel wordt toegekend aan de *urbane* (i.e. bewoonde) gebieden, essentieel andere resultaten zal opleveren.

**ad (b): Fouten in verrastering en classificatie ( $G_{500}^o$ ;  $LBG^o$ ):**

Omdat de effecten van verrastering en classificatie gecombineerd optreden<sup>20</sup> bij de  $\text{NH}_3$ -emissieberekening, zal ook de invloed van hun fouten op een gecombineerde wijze onderzocht worden.

Probleem is allereerst echter hoe de verrasterings- en classificatiefouten gespecificeerd moeten worden. Een probabilistische beschrijving lijkt niet volledig adequaat door het nominale, discrete karakter van de verrasterings- en classificatieprocedure. Het verdient veeleer de voorkeur om, indien mogelijk, deze fouten expliciet te beschrijven, bijv. door aan te geven welk deel van het oppervlak van een 500 x 500 m areaal daadwerkelijk

<sup>19</sup> Lokaal gezien kan de huishoudelijke bijdrage wel aanzienlijk zijn (zie figuur 4), en is deze verwaarlozing minder geoorloofd.

<sup>20</sup> Via verrastering wordt eenduidig bepaald bij welke gemeente een specifiek 500 x 500 m areaal behoort. Classificatie kent aan dit areaal een specifiek landgebruikstype toe. Beide gegevens bepalen vervolgens de hoeveelheid gemeentelijke  $\text{NH}_3$ -emissie die op dat areaal terecht komt.

tot een specifieke gemeente of tot een specifiek landgebruikstype behoort, en dit te vergelijken met de gemeente c.q. het landgebruikstype dat aan dat areaal is toegekend. Vervolgens kan de fout deels worden gekarakteriseerd via grootheden als de ‘fuzzyness- en de confusion graad’ (zie appendix B).

In de hier beschreven toepassing is deze werkwijze inderdaad mogelijk omdat Remote Sensing data beschikbaar zijn die op een fijn schaalnivo (voor 25 x 25 m rastercellen) informatie leveren over het landgebruik. Indien verondersteld wordt dat deze gegevens ( $RS^o$ ) zoals ze in het LGN bestand zijn vastgelegd nagenoeg *foutvrij*<sup>21</sup> zijn, dan kunnen hieruit rechtstreeks de *classificatiefouten* worden afgeleid die optreden bij de landgebruikstypering van de 500 x 500 m arealen volgens het LBG (i.e.  $LBG^o$ ). Indien het bovendien mogelijk is om deze 25 x 25 m rasters nauwkeurig en nagenoeg *foutvrij*<sup>22</sup> toe te wijzen aan specifieke gemeenten (d.w.z.  $G_{25}^o$ ), dan is ook directe informatie over de *verrasteringsfout* beschikbaar voor de 500 x 500 m arealen. Zie ook [5] en [6] voor nadere informatie over de fouten die optreden ten gevolge van ruimtelijke bewerkingen zoals vergridding en aggregatie.

Strikt genomen is het echter voor de uit te voeren betrouwbaarheidsevaluatie *niet nodig* om een *expliciete specificatie* van deze verrasterings- en classificatiefouten ‘an sich’ te hebben. De interesse gaat immers uiteindelijk uit naar de *invloed* van deze fouten op de  $NH_3$ -emissieberekeningen, en niet zo zeer naar hun eigenlijke grootte. Voor de betrouwbaarheidsanalyses kan dan ook volstaan worden met het berekenen van de  $NH_3$ -emissies aan de hand van de fjnschalige informatie uit het LGN bestand ( $RS^o$ ), gecombineerd met de 25 x 25 m verrastering ( $G_{25}^o$ ). Vergelijking van deze resultaten (i.e.  $y(r | Inp^o; MS^o; G_{25}^o; RS^o)$ ) met de nominale uitkomsten van de  $N$ -keten berekeningen (i.e.  $y_{nom}(r) = y(r | Inp^o; MS^o; G_{500}; LBG^o)$ ) levert de vereiste informatie over hoe verrasterings- en classificatie-effecten de betrouwbaarheid van de emissieberekening beïnvloeden.

Zoals gezegd is de bovengeschetste aanpak enkel bruikbaar indien de LGN-informatie ( $RS^o$ ) en de fijn-rastering ( $G_{25}^o$ ) nagenoeg *foutvrij* zijn, en geen noemenswaardige onzekerheden bevatten. Hoewel deze veronderstellingen discutabel zijn, zullen ze toch gehandhaafd worden omdat er vooralsnog onvoldoende informatie<sup>23</sup> is over de betreffende onzekerheden om een andere aanpak mogelijk te maken.

#### ad (c): Fouten in modelspecificaties ( $MS^o$ ) (m.n. allocatiefouten):

Fouten en onzekerheden in de gemiddelde jaarlijkse  $NH_3$ -emissie  $N_h^o$  van 0.7 kg  $NH_3$  per inwoner worden buiten beschouwing gelaten, omdat de onzekerheid in de emissiebijdrage van huishoudens niet meegenomen wordt. Ook wordt niet getornd aan de veronderstelling van homogeniteit bij de toewijzing van de emissie uit dierlijke mest en kunstmest aan de diverse landgebruikstypen.

<sup>21</sup>Strikt genomen is deze veronderstelling discutabel omdat op dit resolutienivo een grote kans op misclassificatie bestaat, met name indien veel klassen onderscheiden worden. Omdat uiteindelijk echter deze klassen weer verder geaggregeerd worden, tot bijv. (productie) ‘landbouw’ en ‘niet-landbouw’ of tot ‘urbaan’ en ‘niet-urbaan’, blijft de ernst van deze misclassificatie vermoedelijk beperkt.

<sup>22</sup>Dit vereist nauwkeurige gemeentekaarten op een fijne schaal, die na verrastering naar 25 x 25 m grids nog betrouwbare informatie verschaffen over de gemeentegrenzen. Omdat er kaarten gehanteerd worden met een schaal van 1:250.000, is er strikt genomen niet aan deze voorwaarde voldaan!!

<sup>23</sup>Voor de Remote Sensing data zou bijv. de kans bekend dienen te zijn op een geclassificeerde waarde van een 25 x 25 m pixel, op basis van zijn gelijkenis met een standaard reflectiepatroon. Deze informatie wordt echter niet standaard geleverd bij de satelliet-data.

Wel wordt gekeken naar de fouten die het gevolg zijn van de gehanteerde ruimtelijke verdeling (*allocatiefouten*) van de  $\text{NH}_3$ -emissie over de arealen behorende bij een gemeente. Deze ruimtelijke verdeling komt naar voren in de keuze van de ‘nominale verdeelsleutels’ voor mest (dierlijke mest en kunstmest) en voor huishoudens. Het idee is om voor een *alternatieve verdeelsleutel* de berekeningen uit te voeren, en de resultaten daarvan te vergelijken met de nominale modelresultaten  $y_{nom}(r)$ . Voor huishoudens ligt de keuze van een alternatief voor de hand, namelijk de huishoudelijke  $\text{NH}_3$ -emissie enkel homogeen te verdelen over bewoonde (urbane) gebieden (zowel op 500 x 500 m als ook op 25 x 25 m gridnivo), en niet over het gehele oppervlak. Voor mest is dat echter gecompliceerder: mocht het al mogelijk zijn om een realistische verdeelsleutel voor de *aanwending* van dierlijke mest en kunstmest over de diverse gewastypen te verkrijgen, dan nog blijft het bijzonder problematisch om de *andere* wijzen van dierlijke  $\text{NH}_3$ -emissie (stallen, opslag, beweiding) adequaat te alloceren aan de diverse landgebruikstypen. Voor mest zal dan ook geen alternatieve verdeelsleutel beschouwd worden.

De modelspecificatie die samenhangt met de zojuist beschreven alternatieve keuze wordt genoteerd als  $MS^{(a)} := [VS_m^o, VS_h^{(a)}, N_h^o]$ , waarbij  $VS_h^{(a)}$  de alternatieve verdeelsleutel voor de huishoudens is, d.w.z.  $VS_h^{(a)} := 1 : 0$  voor ‘urbaan’ versus ‘niet-urbaan’, terwijl de verdeelsleutel voor kunstmest en dierlijke mest hetzelfde blijft als in de nominale situatie (i.e.  $VS_m^o$ ). Zie ook de notationele afspraken in hoofdstuk 3.

## 5 Vereenvoudiging van de betrouwbaarheidsanalyse

Om een indruk te krijgen van de betrouwbaarheid van de  $\text{NH}_3$ -emissieberekeningen die het LAE voor de 5 x 5 km gridcellen uitvoert, dient geanalyseerd te worden wat de invloed is van de diverse foutenbronnen die bij deze berekeningen een rol spelen. Omdat de fouten/onzekerheden in de basisgegevens op adequate wijze statistisch beschreven kunnen worden middels kansverdelingen (zie hoofdstuk 4), is een Monte Carlo georiënteerde aanpak een gangbare manier om de betrouwbaarheid te analyseren. Hierbij worden voor de betreffende basisgegevens (i.e. gemeentelijke  $\text{NH}_3$ -emissiecijfers van dierlijke mest) diverse realisaties getrokken uit de gespecificeerde verdelingen, d.w.z. diverse waarden worden bepaald voor de standaardnormaal verdeelde grootheid  $g_d(i)$  in vergelijking (27). Vervolgens wordt met elk van deze realisaties de emissie berekend voor de 5 x 5 km gridcellen. Dit vereist een veelheid van berekeningen en dito opslag van gegevens, die vervolgens nader geanalyseerd dienen te worden op hun statistische eigenschappen (gemiddelden, varianties, correlaties, histogrammen, betrouwbaarheidsfactoren, 68 % betrouwbaarheidsintervallen).

Echter, ten gevolge van de specifieke keuzes van de onzekerheden in deze basisgegevens kunnen deze berekeningen aanzienlijk vereenvoudigd worden. Ze kunnen zelfs grotendeels op een *analytische* wijze worden uitgevoerd, in plaats van gebruik te maken van uitgebreide Monte Carlo berekeningen. Dit wordt in het hiernavolgende gedetailleerd uitgewerkt. Omdat bij de bespreking van de resultaten in hoofdstuk 6 en 7 geen rechtstreeks gebruik wordt gemaakt van dit materiaal, kan dit technische gedeelte bij globale lezing zonder bezwaar worden overgeslagen.

**Opmerking 1:** Het trekken van realisaties uit de onzekerheidsspecificaties in hoofdstuk 4 komt in wezen neer op het trekken van  $N$  samples voor de dierlijke  $\text{NH}_3$ -emissies  $E_d(\cdot)$  volgens formule (27). Omdat emissievariatië voor gemeenten uit dezelfde provincie identiek zijn, maar voor gemeenten uit verschillende provincies onafhankelijk zijn, betekent dit dat er uit 12 (i.e. aantal provincies) normaal verdeelde onafhankelijke parameters  $g_d(\cdot)$  getrokken moet worden. Indien hiervoor Latin Hypercube Sampling gebruikt wordt (zie [14]), dan kan voor een adequaat beeld van de onzekerheid volstaan worden met  $N = 60 (= 5 \cdot 12)$  trekkingen.

■

De totale emissie  $y(r | \cdot)$  per 5 x 5 km gridcel kan opgebouwd worden gedacht uit de bijdrages van dierlijke mest  $y_d(r | \cdot)$ , kunstmest  $y_k(r | \cdot)$ , en huishoudens  $y_h(r | \cdot)$ . Door deze opsplitsing te hanteren wordt de rekenprocedure vereenvoudigd, en kunnen de emissiecijfers rechtstreeks uit de nominale cijfers ( $Inp^o$ ) worden afgeleid.

Inmers, uit het feit dat de gemeentelijke  $\text{NH}_3$ -emissie uit dierlijke mest varieert volgens (27) volgt er (zie appendix D) dat de bijbehorende dierlijke  $\text{NH}_3$ -emissie voor de  $r$ -de gridcel direct gerelateerd kan worden aan de corresponderende nominale waarde (zie de notatie afspraken uit hoofdstuk 3):

$$y_d(r | Inp; *) = y_d(r | Inp^o; *) \cdot [1 + \sigma_d^{(r)} \cdot g_d^{(r)}] \quad (29)$$

Hierbij duidt  $Inp$  de input-emissie data aan, d.w.z. de gemeentelijke  $\text{NH}_3$ -emissiecijfers afkomstig van dierlijke mest, kunstmest en huishoudens  $E_d(i)$ ,  $E_k(i)$ ,  $E_h(i)$ , terwijl  $Inp^o$  betrekking heeft op hun corresponderende nominale waarden, d.w.z.  $E_d^o(i)$ ,  $E_k^o(i)$ ,  $E_h^o(i)$ .  $g_d^{(r)}$  is een standaardnormaal verdeelde grootheid die refereert naar de emissie-variëte van de gemeenten die bijdragen tot de  $r$ -de gridcel. De grootheid  $\sigma_d^{(r)}$  duidt de bandbreedte voor de  $r$ -de gridcel aan.

Strikt genomen is de uitdrukking (29) alleen geldig indien de emissievariatië  $E_d(i)$  (zie (27)) van de gemeenten  $i$  die aan de  $r$ -de gridcel bijdragen volledig afhankelijk (!) zijn, d.w.z.

indien alle  $g_d(i)$  uit (27) zijn gelijk aan  $g_d^{(r)}$ . Dit is het geval voor gridcellen die volledig binnen één provincie vallen (zie hoofdstuk 4). De bandbreedte  $\sigma_d^{(r)}$  uit (29) is dan gelijk aan de bandbreedte die volgens tabel 2 bij deze provincie behoort; bovendien zal  $g_d^{(r)}$  ten gevolge van de afspraken uit hoofdstuk 4 dezelfde zijn voor alle gridcellen  $r$  binnen die provincie.

Voor gridcellen die op provinciegrenzen liggen, is de formule (29) strikt genomen enkel bij benadering geldig, omdat de bijbehorende gemeentelijke bijdragen niet volledig afhankelijk zijn. Omdat deze gridcellen slechts een klein deel uitmaken van alle gridcellen die Nederland overdekken, wordt de vereenvoudiging uit formule (29) toch gehandhaafd, waarbij de gridcellen op de provinciegrenzen worden 'toegekend' aan de provincie waarin hun centrum ligt.

Analoge uitdrukkingen kunnen worden geformuleerd voor  $y_k(r | Inp; *)$  en  $y_h(r | Inp; *)$ . Immers onder de veronderstelling dat de gemeentelijke  $\text{NH}_3$ -emissie afkomstig van kunstmest en huishoudens varieert volgens

$$E_k(i) = E_k^o(i) \cdot [1 + \sigma_k(i) \cdot g_k(i)] \quad (30)$$

$$E_h(i) = E_h^o(i) \cdot [1 + \sigma_h(i) \cdot g_h(i)] \quad (31)$$

is de bijbehorende  $\text{NH}_3$ -emissie voor de  $r$ -de gridcel gelijk aan

$$y_k(r | Inp; *) = y_k(r | Inp^o; *) \cdot [1 + \sigma_k^{(r)} \cdot g_k^{(r)}] \quad (32)$$

$$y_h(r | Inp; *) = y_h(r | Inp^o; *) \cdot [1 + \sigma_h^{(r)} \cdot g_h^{(r)}] \quad (33)$$

waarbij  $\sigma_k^{(r)}, \sigma_h^{(r)}, g_k^{(r)}, g_h^{(r)}$  rechtstreeks verwijzen naar de provincie waarvan de  $r$ -de gridcel deel uitmaakt.

De totale emissie  $y(r | Inp; *)$  kan nu samengesteld worden als de som van de afzonderlijke bijdrages:

$$y(r | Inp; *) = y_d(r | Inp; *) + y_k(r | Inp; *) + y_h(r | Inp; *) \quad (34)$$

Dit alles samenvattend, levert dat de totale  $\text{NH}_3$ -emissie van gridcel  $r$  varieert rond zijn nominale waarde als:

$$y(r | Inp; *) = y(r | Inp^o; *) \cdot [1 + \sigma_{d,tot}^{(r)} \cdot g_d^{(r)} + \sigma_{k,tot}^{(r)} \cdot g_k^{(r)} + \sigma_{h,tot}^{(r)} \cdot g_h^{(r)}] \quad (35)$$

waarbij de bijbehorende 'bandbreedte'  $\sigma_{x,tot}^{(r)}$  voor  $x = d, k, h$  gelijk is aan:

$$\sigma_{x,tot}^{(r)} := \frac{y_x(r | Inp^o; *)}{y(r | Inp^o; *)} \cdot \sigma_x^{(r)} \quad (36)$$

De fractie  $\frac{y_x(r | Inp^o; *)}{y(r | Inp^o; *)}$  waarmee  $\sigma_x^{(r)}$  vermenigvuldigd wordt is de relatieve bijdrage van de bron  $x$  (i.e. dierlijke mest, kunstmest, huishoudens) tot de nominale totale  $\text{NH}_3$ -emissie van gridcel  $r$ . Merk op dat 'bandbreedte'  $\sigma_{x,tot}^{(r)}$  afhangt van de gridcel  $r$ , en kleiner dan of gelijk aan  $\sigma_x^{(r)}$  is. De variatie  $g_x^{(r)}$  in deze bandbreedte rond de nominale waarde zal echter voor alle gridcellen *binnen* eenzelfde provincie identiek zijn, d.w.z. niet afhangen van  $r$ .

Op bovengeschetste wijze is de variatie in de totale  $\text{NH}_3$ -emissie  $y(r | Inp; *)$  een gewogen som van de variaties in de afzonderlijke emissies  $y_x(r | Inp; *)$ . Door vergelijking (35) te gebruiken kunnen uit de nominale  $\text{NH}_3$ -emissiecijfers  $y(r | Inp^o; *)$  voor gridcel  $r$  rechtstreeks de actuele emissies  $y(r | Inp; *)$  berekend worden, zonder alle afzonderlijke emissieberekeningen expliciet uit te voeren voor de diverse realisaties van  $Inp$ . Dit levert een *aanzienlijke*

*reductie van rekentijd, en ook van opslagcapaciteit!!* Wel is bij deze aanpak vereist dat, bij het doorrekenen van de nominale situatie, de *afzonderlijke bijdrage* van de diverse emissiebronnen  $y_x(r | Inp^o; *)$  (d.w.z. dierlijke mest, kunstmest en huishoudens) expliciet wordt bepaald en opgeslagen. Dit gebeurt bij de huidige berekeningen al standaard om de bijdrage van de diverse emissiebronnen te verkrijgen.

Uit relatie (35) kunnen eenvoudig analytische uitdrukkingen worden afgeleid voor het *gemiddelde, de mediaan, de percentielen, de (co)varianties en de variatiecoëfficiënt (VC)*. Omdat volgens de specificaties uit hoofdstuk 4, de onzekerheden in de bijdrage van kunstmest en huishoudens buiten beschouwing blijven (d.w.z.  $E_k(i) = E_k^o(i)$ ;  $E_h(i) = E_h^o(i)$ , i.e.  $\sigma_k(i) = \sigma_h(i) = 0$ ), geldt er dat<sup>24</sup>:

$$\mathbb{E}[y(r | Inp; *)] = y(r | Inp^o; *) \quad (37)$$

$$\text{Med} [y(r | Inp; *)] = y(r | Inp^o; *) \quad (38)$$

$$\text{Perc}_\alpha [y(r | Inp; *)] = y(r | Inp^o; *) \cdot [1 + \sigma_{d,tot}^{(r)} \cdot \text{Perc}_\alpha [g_d^{(r)}]] \quad (39)$$

$$\begin{aligned} \text{Var} [y(r | Inp; *)] &= [y(r | Inp^o; *)]^2 \cdot [\sigma_{d,tot}^{(r)}]^2 \\ &= [y_d(r | Inp^o; *)]^2 \cdot [\sigma_d^{(r)}]^2 \end{aligned} \quad (40)$$

$$\begin{aligned} \text{VC} [y(r | Inp; *)] &= \sigma_{d,tot}^{(r)} \\ &= \frac{y_d(r | Inp^o; *)}{y(r | Inp^o; *)} \cdot \sigma_d^{(r)} \end{aligned} \quad (41)$$

$$\begin{aligned} \text{Cov} [y(r | Inp; *) ; y(s | Inp; *)] &= [y(r | Inp^o; *) \cdot y(s | Inp^o; *)] \cdot \sigma_{d,tot}^{(r)} \cdot \sigma_{d,tot}^{(s)} \\ &= [y_d(r | Inp^o; *) \cdot y_d(s | Inp^o; *)] \cdot \sigma_d^{(r)} \cdot \sigma_d^{(s)} \end{aligned} \quad (42)$$

waarbij de uitdrukking voor de covariantie enkel geldt voor gridcellen  $r$  en  $s$  die binnen dezelfde provincie liggen. Voor gridcellen binnen verschillende provincies is de covariantie daarentegen nul.

Omdat de aldus berekende  $y(r | Inp; *)$  normaal verdeeld is ( $g_d^{(r)}$  in vergelijking (35) is normaal verdeeld), kan op basis van het voorgaande eenvoudig het *68 %-betrouwbaarheidsinterval*  $[\mu - \sigma, \mu + \sigma]$  worden afgeleid, waarbij  $\mu$  en  $\sigma$  het gemiddelde en de standaarddeviatie van de normaal verdeelde grootte zijn. Het is echter niet mogelijk om een gesloten-vorm uitdrukking voor de *betrouwbaarheidsfactor* af te leiden. Reden hiervan is dat er geen expliciete uitdrukking bestaat voor het gemiddelde en de variantie van  $\log(y(r | Inp; *))$  voor normaal verdeelde  $y(r | Inp; *)$ . Wel kunnen de relaties uit stelling 1, sub (d), in hoofdstuk 3 gebruikt worden om de betrouwbaarheidsfactor bij *benadering* te evalueren. Deze benadering is adequaat indien de variatiecoëfficiënt van  $y(r | Inp; *)$  (i.e.  $\sigma_{d,tot}^{(r)}$ ) klein is (bijv.  $\leq .25$ ), en geeft aanleiding tot de volgende uitdrukking voor de betrouwbaarheidsfactor:

$$B(y_{nom}(r)) = \exp \left( -\sqrt{\beta_*^2(r) + \sigma_*^2(r)} \right) \quad (43)$$

waarbij de bias  $\beta_*(r)$  en de variantie  $\sigma_*^2(r)$  benaderd worden door

$$\beta_*(r) \approx \log \left( \frac{y_{nom}(r)}{y_{pos}^o(r)} \right) \quad (44)$$

$$\sigma_*^2(r) \approx \log \left( 1 + (\text{VC} [y_{pos}(r)])^2 \right) = \log \left( 1 + \left[ \frac{y_{d,pos}^o(r)}{y_{pos}^o(r)} \cdot \sigma_d^{(r)} \right]^2 \right) \quad (45)$$

<sup>24</sup>Er wordt hierbij verondersteld dat er enkel stochastiek/onzekerheid is in  $Inp$ , d.w.z. in  $E_d(i)$  c.q.  $g_d^{(r)}$

Hierbij is  $y_{pos}(r) = y(r | Inp; *)$  de gepostuleerde werkelijkheid, terwijl  $y_{pos}^o(r) = y(r | Inp^o; *)$  diens ‘nominale’ waarde (referentie) aanduidt.

**Opmerking 2:** De uitdrukking (35) kan ook worden gebruikt om de in hoofdstuk 3 genoemde discrepantie tussen het nominale model  $y_{nom}(r)$  en de gepostuleerde werkelijkheid  $y_{pos}(r)$  nader uit te werken.

Laat daartoe het nominale model gegeven zijn door vergelijking (1), d.w.z.

$$y_{nom}(r) = y(r | Inp^o; R^o) \quad (46)$$

waarbij  $R^o := [MS^o; G_{500}^o; LBG^o]$  de rest-specificatie is van het nominale model.

Laat de gepostuleerde werkelijkheid gegeven zijn door:

$$y_{pos}(r) := y(r | Inp; R) \quad (47)$$

waarbij rest-specificatie  $R$  een nog nader in te vullen vorm kan aannemen, bijv.  $R := [MS^o; G_{25}^o; RS^o]$ . De discrepantie tussen deze grootheden is nu:

$$\begin{aligned} \varepsilon(r) &:= y_{nom}(r) - y_{pos}(r) \\ &= y(r | Inp^o; R^o) - y(r | Inp; R) \\ &= y(r | Inp^o; R^o) - y(r | Inp^o; R) \cdot \left[1 + \sum_{x=d,k,h} \sigma_{x,tot}^{(r)} \cdot g_x^{(r)}\right] \\ &= y(r | Inp^o; R^o) - y(r | Inp^o; R) - y(r | Inp^o; R) \cdot \sum_{x=d,k,h} \sigma_{x,tot}^{(r)} \cdot g_x^{(r)} \\ &\stackrel{zie(36)}{=} [y_{nom}(r) - y_{pos}^o(r)] - \sum_{x=d,k,h} y_{x,pos}^o(r) \cdot \sigma_x^{(r)} \cdot g_x^{(r)} \\ &= \varepsilon^o(r) - \sum_{x=d,k,h} y_{x,pos}^o(r) \cdot \sigma_x^{(r)} \cdot g_x^{(r)} \end{aligned} \quad (48)$$

Hierbij zijn

$$y_{pos}^o(r) := y(r | Inp^o; R) \quad (49)$$

$$y_{x,pos}^o(r) := y_x(r | Inp^o; R) \quad (50)$$

de  $\text{NH}_3$ -emissies voor de ‘nominale’<sup>25</sup> gepostuleerde werkelijkheid, en is

$$\varepsilon^o(r) := y_{nom}(r) - y_{pos}^o(r) \quad (51)$$

de ‘nominale’ discrepantie.

De relatie (48) drukt de discrepantie  $\varepsilon(r)$  uit in termen van de *systematische fout*  $\varepsilon^o(r)$  (nominale discrepantie) en een *toevallige fout* die afhangt van de ‘nominale’ gepostuleerde werkelijkheid  $y_{x,pos}^o(r)$ .

■

<sup>25</sup>Nominaal in de zin dat de input-emissie data hun nominale waarden  $Inp^o$  hebben.

## 6 Betrouwbaarheidsanalyse: werkwijze en resultaten

In hoofdstuk 4 werden de foutenbronnen gespecificeerd die een rol spelen bij de  $\text{NH}_3$ -emissie berekeningen uit de N-keten (zie hoofdstuk 2), en die bij een betrouwbaarheidsevaluatie van deze berekeningen onderzocht kunnen worden. Hiermee ligt de uit te voeren betrouwbaarheidsanalyse echter nog geenszins volledig vast. Immers de diverse foutenbronnen kunnen hierbij afzonderlijk of gezamenlijk meegenomen worden.

Tabel 3 geeft een indruk van de diverse combinaties die mogelijk zijn. Bij elk van deze combinaties staat het referentiepunt  $y_{pos}(r)$  (i.e. de gepostuleerde werkelijkheid) vermeld, waarmee de nominale emissieberekening  $y_{nom}(r)$  uit de N-keten vergeleken wordt. Bovendien staat weergegeven hoeveel realisaties ( $N_{sam}$ ) van de gepostuleerde werkelijkheid  $y_{pos}(r)$  beschouwd worden, of de betrouwbaarheidsfactor samengesteld is uit een ‘bias’ en/of ‘variantie’ component, en of de betrouwbaarheidsanalyse voornamelijk een *intern* dan wel een *extern* karakter<sup>26</sup> heeft (zie [1]). Merk op dat mogelijke fouten/onzekerheden in de Remote Sensing gegevens en de 25 x 25 m verrastering buiten beschouwing blijven bij de voorgestelde analyses.  $RS^\circ$  en  $G_{25}^\circ$  worden dus verondersteld foutvrij te zijn.

Exp.	Fout in Input Emissie	Fout in Raster/ Classific.	Fout in Verdeel-sleutel	Gepostuleerde Werkelijkheid $y_{pos}(r)$	Aantal Realis. $y_{pos}(r)$	Betrouwbaarheids-factor	Aard Betrouwbaarheid
I	X	-	-	$y(r   Inp; MS^\circ; G_{500}^\circ; LBG^\circ)$	$N_{sam}$	bias/var.	Intern
II	-	X	-	$y(r   Inp^\circ; MS^\circ; G_{25}^\circ; RS^\circ)$	1	bias	‘Extern’
III	-	-	X	$y(r   Inp^\circ; MS^{(a)}; G_{500}^\circ; LBG^\circ)$	1	bias	Intern
IV	X	-	X	$y(r   Inp; MS^{(a)}; G_{500}^\circ; LBG^\circ)$	$N_{sam}$	bias/var.	Intern
V	X	X	-	$y(r   Inp; MS^\circ; G_{25}^\circ; RS^\circ)$	$N_{sam}$	bias/var.	‘Extern’
VI	-	X	X	$y(r   Inp^\circ; MS^{(a)}; G_{25}^\circ; RS^\circ)$	1	bias	‘Extern’
VII	X	X	X	$y(r   Inp; MS^{(a)}; G_{25}^\circ; RS^\circ)$	$N_{sam}$	bias/var.	‘Extern’

Tabel 3: Experimenten voor de betrouwbaarheidsanalyse van de nominale  $\text{NH}_3$ -emissieberekeningen [ $y_{nom}(r) = y(r | Inp^\circ; MS^\circ; G_{500}^\circ; LBG^\circ)$ ].

Met het oog op de beperkt beschikbare tijd, is uit de veelheid van experimenten slechts één experiment onderzocht, en wel het meest complete waarbij alle foutenbronnen gezamenlijk zijn meegenomen (experiment VII in tabel 3). Dit heeft als consequentie dat de effecten van de afzonderlijke foutenbronnen op de betrouwbaarheid van de emissiecijfers niet altijd onderscheiden kunnen worden. Zo zal bijv. de allocatiefout voor de bijdrage van de huishoudens altijd gecombineerd optreden met de fouten ten gevolge van verrastering en classificatie.

<sup>26</sup>Bij een ‘interne’ betrouwbaarheidsanalyse worden de nominale modeluitkomsten vergeleken met uitkomsten uit hetzelfde (nominale) model, maar met andere settings. Bij een ‘externe’ betrouwbaarheidsanalyse vindt daarentegen de vergelijking plaats met werkelijke waarden (metingen), of, indien deze niet rechtstreeks beschikbaar zijn, met resultaten uit verfijndere modelberekeningen (bijv. op basis van  $G_{25}^\circ, RS^\circ$ ) die verondersteld worden de werkelijkheid adequaat weer te geven. Uiteraard is in dit laatste geval het ‘extern’ karakter van de analyse van een minder uitgesproken orde dan in de situatie met echte metingen.



nr.	Verzuringengebied	nr.	Verzuringengebied	nr.	Verzuringengebied
1	Groningen	8	Z-Gelderland	15	W-Noord Brabant
2	Friesland	9	Utrecht	16	M-Noord Brabant
3	Drenthe	10	N-Noord Holland	17	NO-Noord Brabant
4	W/NO Overijssel + NO polder	11	Z-Noord Holland + Flevopolder	18	ZO-Noord Brabant
5	ZO Overijssel	12	N-Zuid Holland	19	N-Limburg
6	NW Gelderland	13	Z-Zuid Holland	20	Z/M Limburg
7	NO Gelderland	14	Zeeland		

Tabel 4: Nummers en namen van verzuringengebieden

Bij het voorgestelde experiment wordt de standaardberekening zoals deze in de N-keten gebruikt is (d.w.z.  $y_{nom}(r) = y(r | Inp^o; MS^o; G_{500}^o; LBG^o)$ ; zie hoofdstuk 2), vergeleken met de gepostuleerde werkelijkheid ( $y_{pos}(r) = y(r | Inp; MS^{(a)}; G_{25}^{(a)}; RS^o)$ ), die verkregen is via analoge berekeningen op basis van het verfijnde schaalnivo (25 x 25 m), uitgevoerd met de onzekerheidsspecificaties uit hoofdstuk 4. Deze verfijnde berekeningen worden verondersteld representatief te zijn voor de (onzekere) werkelijkheid.

De benodigde rekeninspanning is aanzienlijk ingeperkt door het gebruik van de vereenvoudigingen uit hoofdstuk 5, die het mogelijk maken om de gewenste informatie op theoretische wijze uit de nominale berekeningen af te leiden, zonder de toevlucht te nemen tot een rekenintensieve en meer empirisch georiënteerde Monte Carlo benadering.

**Opmerking 3:** De discrepantie tussen de nominale berekening  $y_{nom}(r)$  en gepostuleerde werkelijkheid  $y_{pos}(r)$  wordt bij dit experiment gegeven door:

$$\varepsilon(r) := y_{nom}(r) - y_{pos}(r) = y(r | Inp^o; R_{500}^o) - y(r | Inp; R_{25}^{(a)}) \quad (52)$$

waarbij  $R_{500}^o := [MS^o; G_{500}^o; LBG^o]$  naar de nominale LBG situatie verwijst, en  $R_{25}^{(a)} := [MS^{(a)}; G_{25}^{(a)}; RS^o]$  naar de Remote Sensing situatie met alternatieve verdeelsleutel  $MS^{(a)}$ . Deze discrepantie kan worden uitgesplitst als (zie ook opmerking (2)):

$$\begin{aligned} \varepsilon(r) &= [y(r | Inp^o; R_{500}^o) - y(r | Inp^o; R_{25}^{(a)})] + [y(r | Inp^o; R_{25}^{(a)}) - y(r | Inp; R_{25}^{(a)})] \\ &= [y_{nom}(r) - y_{pos}^o(r)] + [y_{pos}^o(r) - y_{pos}(r)] \end{aligned} \quad (53)$$

De term  $[y_{nom}(r) - y_{pos}^o(r)]$  geeft de invloed weer van de verrastering/classificatiefout in combinatie met de allocatiefout, en komt exact overeen met de discrepantie bij experiment VI uit tabel 3. In wezen geeft deze term de *systematische* afwijking weer tussen nominaal model en gepostuleerde werkelijkheid.

De tweede term  $[y_{pos}^o(r) - y_{pos}(r)]$  beschrijft de invloed van de specificatiefout in de basisgegevens  $Inp^o$ , maar dan wel geëvalueerd voor de ‘Remote-Sensing situatie’  $R_{25}^{(a)}$ . Deze term drukt de *random* fout uit, die bij de gehanteerde onzekerheidsspecificaties uit hoofdstuk 4 enkel zijn oorsprong vindt in de specificatiefouten in de basisgegevens. Merk op dat bovengenoemde grootte  $[y_{pos}^o(r) - y_{pos}(r)]$  enigszins verschilt van de discrepantie die in experiment I uit tabel 3 onderzocht wordt, en die betrekking heeft op de ‘LBG situatie’  $R_{500}^o$ .

■

De betrouwbaarheidsevaluatie is er allereerst op gericht om een beeld te krijgen van de betrouwbaarheid van de berekende  $NH_3$ -emissiecijfers voor de 1657 *gridcellen* (5 x 5 km) die Nederland bedekken. Daarnaast gaat de interesse uit naar de emissiecijfers op *hoger aggregatienivo*, namelijk voor *Nederland als totaal* en voor de 20 *verzuringengebieden* waarin Nederland is opgedeeld (zie figuur 2 en tabel 4; deze opdeling is bepaald op basis van zogenaamde COROP gebieden). De resultaten zijn als volgt:

**(1) Betrouwbaarheid op 5 x 5 km grid-niveau**

De standaardberekening (nominale berekening) (zie figuur 3, sub (a)) wordt vergeleken met de 'verfijnde' modelberekeningen (gepostuleerde werkelijkheid), waarvan de gemiddelde waarden staan weergegeven in figuur 3, sub (b). De procentuele bijdrage van dierlijke mest, kunstmest en huishoudens aan deze cijfers is vermeld in figuur 4. Deze resultaten illustreren dat de bijdrages van huishoudens en kunstmest, hoewel zij op landelijk niveau ieder slechts zo'n 5 % bedragen (zie tabel 1), op lokaal niveau aanzienlijk kunnen variëren. Met name in dichtbevolkte gebieden (randstad) is de bijdrage van de huishoudens substantieel, terwijl kunstmest bijv. in Zeeland, Groningen en de Flevopolder relatief gezien veel bijdraagt aan de NH<sub>3</sub>-emissie.

De mate waarin de standaardberekening (i.e. nominale berekening) uit figuur 3, sub (a), afwijkt van de gemiddelde gepostuleerde waarden uit figuur 3, sub (b), wordt belicht in de figuren 5 en 6:

- (i) De totale NH<sub>3</sub>-emissie wordt nagenoeg even vaak onder- of overschat. De grootte van deze onder- c.q. overschatting is echter scheef verdeeld, hetgeen door figuur 6, sub (a), wordt geïllustreerd, evenals door het feit dat bijv. in zo'n 5 % van de gridcellen een overschatting met meer dan 25 % optreedt, terwijl in 10 % van alle cellen in een onderschatting van dezelfde orde optreedt.

Analoge eigenschappen gelden ook voor de emissiebijdragen uit dierlijke mest en kunstmest (zie figuur 6, sub (b,c)). Bovendien blijkt uit tabel 5 dat het optreden van onder- c.q. overschatting voor emissies uit kunstmest en dierlijke mest sterk gecorreleerd is (correlatie .92), terwijl de correlatie met de totale NH<sub>3</sub>-emissie geringer is (ongeveer .75).

	Totaal	Dierlijke mest	Kunstmest	Huishoudens
Totaal	1.	.76	.74	.36
Dierlijke mest	.76	1	.92	.12
Kunstmest	.74	.92	1	.11
Huishoudens	.36	.12	.11	1

Tabel 5: Correlaties van het optreden van onder/overschatting bij de berekening van de totale NH<sub>3</sub>-emissie, en de NH<sub>3</sub>-emissie uit dierlijke mest, kunstmest, huishoudens, voor alle 5 x 5 km gridcellen

- (ii) De huishoudelijke emissiebijdrage wordt daarentegen vaker overschat (in 61 % van de gridcellen) door de standaardberekeningen. Reden daarvan is het feit dat de huishoudelijke emissie in de standaardberekeningen gelijkelijk over het totale gemeente oppervlak verdeeld worden, waardoor de niet-urbane gebieden meer emissie krijgen toebedeeld. Bovendien is de onder- of overschatting bij huishoudens aanzienlijk groter dan bij kunstmest en dierlijke mest: uit berekeningen bleek dat in slechts 7 % van de gridcellen de nominale emissieberekening voor dierlijke mest en kunstmest met meer dan 50 % afwijkt van zijn gepostuleerde waarde (i.h.a. onderschatting), terwijl dit voor huishoudens bij 40 % van alle gridcellen het geval is (i.h.a. overschatting; zie ook de figuren 5, sub (b,c,d), en de histogrammen uit figuur 6, sub (b,c,d)).

Belangrijkste oorzaak van dit afwijkend gedrag is dat bij de huishoudelijke bijdrage niet enkel het effect van verrastering en classificatie optreedt, zoals bij kunstmest

en dierlijke mest, maar ook het effect van een alternatieve verdeelsleutel die ervoor zorgt dat bij de verfijnde berekeningen de huishoudelijke emissie enkel gealloceerd wordt aan bewoonde gebieden. Overigens blijken deze effecten voor de totale  $\text{NH}_3$ -emissie weer enigszins uitgemiddeld te worden: in slechts 7 % van alle gridcellen wijkt de totale  $\text{NH}_3$ -emissie meer dan 50 % af van zijn gepostuleerde waarde (zie ook figuur 6, sub (a)).

- (iii) Onderlinge vergelijking van de subplots in figuur 5, evenals in figuur 6, lijkt een sterke verwantschap te suggereren tussen de relatieve afwijking in de totale  $\text{NH}_3$ -emissies, en in de  $\text{NH}_3$ -emissies uit dierlijke mest en kunstmest. De resultaten uit tabel 6 maken echter duidelijk dat deze verwantschap voor een deel slechts schijn is: de relatieve afwijking bij dierlijke mest en kunstmest zijn weliswaar in hoge mate gecorreleerd, hun correlatie met de relatieve afwijking in de totale  $\text{NH}_3$ -emissie is daarentegen zwak. Een verklaring hiervoor is te vinden in het feit dat enerzijds dierlijke mest en kunstmest op vergelijkbare wijze gealloceerd worden, terwijl anderzijds de totale emissie ook nog de bijdrage van de huishoudens bevat. Deze laatste leidt blijkens het voorgaande item (ii) tot grote afwijkingen tussen standaardberekening en gepostuleerde waarden, waardoor de correlatie van dierlijke mest c.q. kunstmest met de totale  $\text{NH}_3$ -emissie zwak blijft.

	Totaal	Dierlijke mest	Kunstmest	Huishoudens
Totaal	1.	.10	.10	.68
Dierlijke mest	.10	1	.99	-.17
Kunstmest	.10	.99	1	-.17
Huishoudens	.68	-.17	-.17	1

Tabel 6: Correlaties tussen de relatieve afwijkingen van nominale berekening en gemiddeld verfijnde berekening voor alle 5 x 5 km gridcellen, betreffende totale  $\text{NH}_3$ -emissie, en de  $\text{NH}_3$ -emissies uit dierlijke mest, kunstmest en huishoudens

- (iv) Uit figuur 5, sub (a), kunnen de gridcellen worden geïdentificeerd waarvoor de gepostuleerde  $\text{NH}_3$ -emissies met meer dan 25 % worden onder- of overschat. Dit betreft respectievelijk zo'n 10 % (onderschatting) en 5 % (overschatting) van alle cellen (vergelijk ook figuur 6, sub (a)). Deze zijn met name gelegen op de Waddeneilanden, langs de kustlijn, in Zeeland, in de Flevopolder, op de Veluwe, op de grens van Limburg en Brabant met België en Duitsland. De reden hiervan is niet meteen duidelijk: de 'verfijnde' landgebruikskaart uit figuur 7 op basis van het LGN-bestand ( $RS^\circ$ ), laat voor die gebieden weinig duidelijke verschillen zien met de door het LBG samengestelde landgebruikskaart (zie figuur 8) die in de standaardberekening gehanteerd is. Wel is het voor de hand liggend dat de berekening van de huishoudelijke emissiebijdrage substantieel andere resultaten oplevert in de voornoemde gebieden omdat de bebouwing een sterk heterogeen karakter heeft. Dit wordt onderstreept door figuur 5, sub (d), waarin de gridcellen aangegeven zijn waarvoor de gepostuleerde huishoudelijke bijdrage met meer dan 75 % overschat wordt. Dit zijn ruim 25 % van alle cellen<sup>27</sup>, die met name gelegen zijn in dun c.q. ongelijk bevolkte gebieden (Veluwe, Flevopolder, kustgebieden, grensgebieden<sup>28</sup>

<sup>27</sup>Een onderschatting van dezelfde grootte vindt in nog geen 1 % van de cellen plaats.

<sup>28</sup>De grote afwijking in de grensgebieden is te wijten aan het feit dat bij rasterverfijning een deel van de (verfijnde) cellen volledig buiten Nederland ligt, en dus geen emissie krijgt toegewezen.

etc.).

In feite leveren bovenstaande resultaten over de afwijking  $[y_{nom}(r) - y_{pos}^o(r)]$  tussen de standaardberekeningen ( $y_{nom}(r)$ ) en de gemiddelde gepostuleerde emissies ( $y_{pos}^o(r)$ ) nuttige informatie over de *systematische fout* die bij de berekening gemaakt wordt. Deze fout is te wijten aan de invloed van verrastering, classificatie, en het hanteren van een alternatieve verdeelsleutel, en vormt een onderdeel van de totale fout die de betrouwbaarheid van de berekeningen bepaalt (zie opmerking (3)).

Daarnaast is er ook nog sprake van een *random fout*, die bij de hier gehanteerde onzekerheidsspecificatie volledig veroorzaakt wordt door specificatiefouten in de basisgegevens (zie opmerking (3)). De grootte van deze random fout kan bijv. uitgedrukt worden in de *variatioecoëfficiënt* die de relatieve spreiding van de gepostuleerde waarden ( $y_{pos}(r)$ ) rond hun gemiddelde ( $y_{pos}^o(r)$ ) weergeeft (zie figuur 9, sub (a)). De ruimtelijke verdeling in deze figuur is nauw verwant met die van de relatieve bijdrage van dierlijke mest aan de totale  $\text{NH}_3$ -emissie (vergelijk figuur 4, sub (a)); zo is bijv. de onzekerheid (spreiding) gering in gridcellen waar deze bijdrage laag is. De reden van deze verwantschap is duidelijk wanneer we de uitdrukking voor de variatioecoëfficiënt uit formule (37) e.v. beschouwen: de variatioecoëfficiënt is gelijk aan het produkt van deze relatieve bijdrage met de bandbreedte  $\sigma_d^{(r)}$  behorend bij de provincie.

Door nu bovenstaande systematische fout  $[y_{nom}(r) - y_{pos}^o(r)]$  te delen door de spreiding (i.e. de standaarddeviatie  $\sigma(y_{pos}(r))$ ) in de random fout, verkrijgen we de *gestandaardiseerde bias*  $\frac{[y_{nom}(r) - y_{pos}^o(r)]}{\sigma(y_{pos}(r))}$ , die het *relatieve belang* van de systematische fout t.o.v. de random fout weergeeft (zie figuur 9, sub (b)). Met name voor normaal verdeelde grootheden is dit een goede maat. Waarden in de buurt van 0 (bijv. tussen  $-0.25$  en  $+0.25$ ) weerspiegelen een geringe invloed van de systematische fout t.o.v. de random fout, terwijl grotere waarden (bijv.  $> 1$  of  $< -1$ ) aanduiden dat de systematische fout van meer belang is. Uit figuur 9, sub (b), blijkt bijv. dat met name in kustgebieden, grensgebieden, de Flevopolder, de randstad en de Veluwe de absolute waarde van de gestandaardiseerde bias groot is ( $> 2$ ).

De gecombineerde invloed van zowel systematische als ook random fouten komt tot uitdrukking in de *betrouwbaarheidsfactor* (zie figuur 10), die de totale afwijking tussen standaardberekening en gepostuleerde werkelijkheid in één getalswaarde uitdrukt. Deze factor blijkt aanzienlijk te kunnen variëren over Nederland. Dit wordt nader onderstreept door figuur 11, 12 en tabel 7 waarin zowel voor Nederland als totaal, als voor (diverse van) de 20 verzuringsgebieden waarin Nederland is opgedeeld, enkele belangrijke statistische karakteristieken vermeld staan. Deze informatie verduidelijkt dat de verdeling van de betrouwbaarheidsfactoren scheef is, en dat hun getalswaarden ver uiteen kunnen lopen (bijv. tussen 0.01 en 0.96 voor Nederland als totaal). De variabiliteit kan per verzuringsgebied veel verschillen: zo hebben volgens tabel 7 de betrouwbaarheidsfactoren in de verzuringsgebieden 2, 6, 10, en 20 een grote variatioecoëfficiënt, terwijl deze voor de gebieden 3, 4, 8 en 17 veel minder groot is, omdat daar de minima en maxima niet zo ver uit elkaar liggen. Deze grote variatie, evenals het voorkomen van lage betrouwbaarheidsfactoren, wordt *grotendeels veroorzaakt door de systematische fout* tussen de standaardberekening en de gepostuleerde werkelijkheid. De invloed van de *random fout* is hierbij *minder van belang*, immers berekeningen tonen aan dat de betrouwbaarheidsfactoren voor de gridcellen allen tussen .8 en 1. zullen liggen, indien er enkel sprake zou zijn van een random fout (d.w.z. indien de bias-component

Gebied	Med	Mean	SDev	MAD	Min	5 %	10 %	90 %	95 %	Max	VC	rMAD
V1	0.84	0.81	0.15	0.07	0.01	0.52	0.82	0.86	0.87	0.88	0.19	0.08
V2	0.81	0.75	0.20	0.11	0.01	0.04	0.64	0.82	0.82	0.83	0.27	0.15
V3	0.84	0.84	0.04	0.02	0.57	0.80	0.82	0.87	0.87	0.88	0.05	0.02
V4	0.85	0.83	0.07	0.03	0.35	0.70	0.82	0.86	0.86	0.87	0.08	0.04
V5	0.85	0.82	0.08	0.05	0.30	0.65	0.73	0.86	0.86	0.89	0.10	0.06
V6	0.82	0.71	0.21	0.16	0.15	0.21	0.34	0.86	0.87	0.87	0.29	0.23
V7	0.85	0.82	0.09	0.04	0.17	0.68	0.79	0.86	0.86	0.87	0.11	0.05
V8	0.86	0.83	0.07	0.05	0.54	0.64	0.74	0.87	0.87	0.96	0.08	0.05
V9	0.84	0.81	0.10	0.05	0.27	0.60	0.74	0.85	0.87	0.90	0.12	0.07
V10	0.81	0.71	0.21	0.16	0.02	0.17	0.30	0.85	0.86	0.89	0.30	0.22
V11	0.81	0.76	0.16	0.10	0.01	0.38	0.59	0.86	0.87	0.93	0.21	0.14
V12	0.83	0.80	0.12	0.07	0.22	0.50	0.67	0.89	0.92	0.95	0.14	0.08
V13	0.84	0.82	0.10	0.06	0.24	0.61	0.69	0.88	0.90	0.94	0.12	0.07
V14	0.86	0.80	0.16	0.10	0.06	0.41	0.58	0.89	0.89	0.91	0.19	0.12
V15	0.86	0.82	0.09	0.06	0.36	0.57	0.69	0.88	0.89	0.90	0.11	0.07
V16	0.86	0.82	0.14	0.07	0.03	0.47	0.75	0.88	0.88	0.88	0.18	0.09
V17	0.86	0.84	0.05	0.03	0.58	0.67	0.80	0.87	0.87	0.87	0.07	0.04
V18	0.84	0.82	0.07	0.05	0.51	0.64	0.70	0.87	0.87	0.90	0.09	0.06
V19	0.83	0.78	0.17	0.10	0.07	0.31	0.70	0.88	0.89	0.91	0.21	0.12
V20	0.85	0.77	0.19	0.13	0.05	0.20	0.50	0.88	0.90	0.93	0.25	0.16
Nld	0.84	0.79	0.15	0.08	0.01	0.48	0.67	0.87	0.88	0.96	0.18	0.10

Tabel 7: Statistics van betrouwbaarheidsfactoren voor de 5 x 5 km gridcellen, per verzuingsgebied en Nederland totaal<sup>a</sup>

<sup>a</sup>De rubrieken Med, Mean en SDev duiden de mediaan, het gemiddelde en de standaarddeviatie aan. MAD refereert naar de 'Mean Absolute Deviation', die een meer robuuste maat voor de spreiding in een dataset is dan de standaarddeviatie (zie [12]). De overige rubrieken refereren naar het minimum, de 5, 10, 90 en 95 percentiel punten, en het maximum. VC en rMAD tenslotte duiden de variatiecoëfficiënt en de relatieve MAD aan, die verkregen worden door SDev, respectievelijk MAD te delen door het gemiddelde. Op deze wijze wordt de data-scatter uitgedrukt als fractie van het gemiddelde.

in formule (3) nul wordt verondersteld). Als daarentegen alleen de systematische fout zijn invloed zou doen gelden (d.w.z. indien de variantie-component in formule (3) gelijk wordt gesteld aan nul), dan variëren de betrouwbaarheidsfactoren tussen .01 en 1., terwijl dan in ruim 15 % van alle gridcellen de betrouwbaarheidsfactor kleiner is dan .8. De belangrijke rol die de systematische fout dus blijkt te spelen wordt nogmaals bevestigd in figuur 13. Deze figuur illustreert dat de gestandaardiseerde bias groot is voor gridcellen met een kleine betrouwbaarheid.

De voorgaande feiten verduidelijken de invloed van de diverse foutcategorieën op de betrouwbaarheid van de emissiecijfers:

- (a) De *systematische fout* (zie o.a. figuur 5) is immers het rechtstreekse gevolg van verrasterings- en classificatiefouten, gecombineerd met allocatiefouten (wat betreft de huishoudelijke bijdrage). Gridcellen met een geringe betrouwbaarheid (bijv. betrouwbaarheidsfactor  $< 0.7$ ; zie figuur 10) verwijzen dientengevolge naar plaatsen waar de gecombineerde effecten van verrastering, classificatie en een alternatieve verdeelsleutel aanzienlijke fouten veroorzaken. De  $\text{NH}_3$ -emissie in bovengenoemde gridcellen (zie figuur 3) is echter in het algemeen laag, waardoor de ernst van de lage betrouwbaarheden beperkt blijft.
- (b) De *random fout* hangt direct samen met de specificatiefouten in de basisgegevens. Hun invloed is zichtbaar in de variatiecoëfficiënten, die de onzekerheid karakteriseren van de emissiecijfers rond hun gemiddelde gepostuleerde waarde (figuur 9, sub (a)).

## (2) Betrouwbaarheid voor verzuringsgebieden en Nederland totaal

De  $\text{NH}_3$ -emissiecijfers voor de verzuringsgebieden en Nederland als totaal worden bepaald door de emissiebijdrage van alle gridcellen die van deze gebieden deel uit maken, te sommeren. Deze cijfers, alsmede de bijbehorende betrouwbaarheidsfactoren, -intervallen en geometrisch gemiddelden, staan in figuur 14, sub (a), vermeld, waarbij gerekend is met onafhankelijke emissievariaties per provincie. Het geometrisch gemiddelde, dat een schatting is voor de mediane waarde, is gelijk gekozen aan het gemiddelde  $\mu$ , terwijl het betrouwbaarheidsinterval het  $(\mu \pm \sigma)$ -interval is, dat met 68 % kans de 'ware' (gepostuleerde) waarde bevat. Deze benaderingen zijn geoorloofd, omdat de gepostuleerde emissiecijfers normaal verdeeld zijn op grond van de gehanteerde onzekerheidsspecificaties uit hoofdstuk 4. De informatie in figuur 14, sub (a), is genormaliseerd rond de uitkomsten van de standaard-berekeningen  $y_{nom}(r)$ , en levert daarom rechtstreeks visuele duidelijkheid over de invloed van de systematische en de random fout. Meer gedetailleerde getalsmatige informatie over de grootte van deze fouten staat vermeld in tabel 8, waarbij de (procentuele) systematische fout ook apart vermeld staat voor de diverse emissiebron-categorieën.

Uit dit alles blijkt dat de emissies voor verzuringsgebieden en Nederland als totaal met een beduidend grotere betrouwbaarheid kunnen worden bepaald dan de emissies voor de 5 x 5 km gridcellen, en dat de invloed van de systematische fout in het niet valt bij de invloed van de random fout. De verklaring hiervoor is voor de hand liggend: verrastering, classificatie en allocatie, die de systematische fout veroorzaken, hebben enkel invloed op de specifieke emissieverdeling *binnen* een gemeente, en niet zozeer op de *totale*  $\text{NH}_3$ -emissie per gemeente. De emissiecijfers per verzuringsgebied ondervinden dientengevolge nauwelijks invloed van eventuele systematische fouten. Enkel random fouten

Gebied	Systematische Afwijking (in procent)				VarCoef	Gestand. Bias
	NH <sub>3</sub> -tot	NH <sub>3</sub> -dier	NH <sub>3</sub> -kunst	NH <sub>3</sub> -huis		
V1	0.87	0.33	0.34	9.65	0.16	0.05
V2	0.04	0.01	-0.05	1.26	0.21	0.00
V3	0.35	0.36	0.06	0.83	0.16	0.02
V4	0.05	0.08	0.15	-1.09	0.16	0.00
V5	-0.03	-0.02	-0.01	-0.28	0.16	-0.00
V6	-0.33	-0.68	-0.35	5.38	0.15	-0.02
V7	-0.40	-0.44	-0.45	1.38	0.15	-0.03
V8	0.56	0.75	0.44	-1.93	0.14	0.04
V9	0.96	1.23	0.60	-1.68	0.16	0.06
V10	0.10	0.01	0.11	0.97	0.18	0.01
V11	-0.15	-0.21	-0.08	-0.04	0.15	-0.01
V12	0.63	-0.30	-0.32	4.94	0.16	0.04
V13	-1.30	-0.31	-0.24	-5.49	0.15	-0.09
V14	0.32	0.31	0.18	0.68	0.12	0.03
V15	-0.56	-0.35	-0.32	-3.07	0.13	-0.04
V16	1.33	1.17	1.12	4.78	0.14	0.10
V17	0.08	0.34	-0.29	-8.47	0.14	0.01
V18	-0.47	-0.62	-1.02	3.38	0.14	-0.03
V19	-1.32	-1.35	-0.37	-1.31	0.13	-0.10
V20	0.18	0.30	0.20	-0.79	0.12	0.02
Nld	0.05	0.02	0.02	0.37	0.16	0.00

Tabel 8: Informatie over de systematische fout en de random fout in de berekeningen voor NH<sub>3</sub>-emissietotalen per verzuringsgebied en Nederland totaal; de emissievariaties van verschillende provincies zijn onderling onafhankelijk verondersteld.

spelen een rol die het gevolg zijn van specificatiefouten in de gemeentelijke basisgegevens. Deze fouten zullen nauwelijks uitmiddelen omdat de meeste verzuringsgebieden deel uitmaken van slechts één provincie, waarbij alle gemeenten binnen de provincie relatief gezien dezelfde specificatiefout hebben.

Uitmiddelingseffecten treden echter wel op voor Nederland als totaal, omdat de emissievariaties voor gemeenten in verschillende provincies onafhankelijk zijn (zie de onzekerheidsspecificaties in hoofdstuk 4). Hierdoor treedt er dus een verdere betrouwbaarheidsverbetering op en is de betrouwbaarheidsfactor voor Nederland totaal dan ook hoger (0.95) dan wanneer alle gemeentelijke emissievariaties over heel Nederland volledig afhankelijk zouden zijn geweest. Uit figuur 14, sub (b), blijkt immers dat in een situatie met volledige afhankelijkheid de betrouwbaarheid voor de totale NH<sub>3</sub>-emissie in Nederland gelijk zou zijn aan 0.86. Dit illustreert het grote effect van de inter-gemeentelijke afhankelijkheidsstructuur op de betrouwbaarheid. Merk tot slot op dat in figuur 14, sub (b), de betrouwbaarheidsfactor voor de verzuringsgebieden echter nauwelijks verschilt van die in figuur 14, sub (a). Oorzaak hiervan is het feit dat de verzuringsgebieden grotendeels binnen één provincie liggen, waardoor de bijbehorende emissievariaties sowieso al volledig afhankelijk zijn, los van het feit of deze variaties in verschillende provincies wel of niet onderling onafhankelijk zijn.

## 7 Conclusies en aanbevelingen

De resultaten van de uitgevoerde analyse laten zien dat de betrouwbaarheid van de emissiecijfers sterk afhangt van het schaalniveau waarop ze betrekking hebben. De betrouwbaarheid van de  $\text{NH}_3$ -emissiecijfers op 5 x 5 km schaal varieert tamelijk sterk (het 10 percentiel ligt op 0.67, en het 90 percentiel op 0.87), waarbij lage betrouwbaarheden grotendeels veroorzaakt worden door fouten ten gevolge van ruimtelijke operaties zoals verrastering, classificatie en emissieverdeling. Schaalvergroting heeft hierop een sterk vereffenend effect: de betrouwbaarheidsfactor van de totale  $\text{NH}_3$ -emissies per verzuringsgebied varieert tussen 0.81 en 0.89, en de rol van bovengenoemde fouten is hierbij verwaarloosbaar. De betrouwbaarheid wordt voor dit schaalniveau grotendeels bepaald door specificatiefouten in de basisgegevens betreffende gemeentelijke ammoniakemissies. De uiteindelijke schaalvergroting naar Nederland als totaal levert een nog verdere verbetering van de betrouwbaarheidsfactor op (0.95) omdat de specificatiefouten elkaar wegens hun veronderstelde onafhankelijkheid (per provincie) deels opheffen.

Deze bevindingen werpen enerzijds een kritisch licht op de huidige emissieberekeningen voor de 5 x 5 km gridcellen. Immers, het gebruik van fijnschalige (25 x 25 m) informatie betreffende geografie, landgebruik en emissieverdeling leidt tot resultaten die soms aanzienlijk kunnen verschillen van de huidige berekeningen, dit alles nog afgezien van het feit of de gebruikte basisgegevens betreffende gemeentelijke emissies wel juist zijn. Kwaliteitsverbetering van de  $\text{NH}_3$ -emissiecijfers op 5 x 5 km schaal moet dan ook o.a. gezocht worden in het verzamelen van nauwkeuriger gegevens betreffende deze aspecten, en een verantwoorde verwerking daarvan in de berekeningsmethodiek. Zo zou een meer gedifferentieerde datatoelevering door het LEI-DLO, die rechtstreeks gekoppeld is aan arealen en bedrijven op het gewenste schaalniveau, voorkomen dat er grote fouten geïntroduceerd worden door de ruimtelijke operaties die momenteel nog vereist zijn omdat de basisgegevens enkel op gemeentelijk niveau worden aangeleverd. Met de huidige cijfers en ammoniak- en emissiemodellen behoort deze optie inderdaad tot de verwezenlijkbare mogelijkheden.

Anderzijds suggereert de verrichte studie dat bovengenoemde effecten van verrastering, classificatie en emissieverdeling op een grover schaalniveau (verzuringgebieden; Nederland als totaal) een veel minder grote rol spelen, en dat de betrouwbaarheid daar voornamelijk bepaald wordt door de kwaliteit van de basisgegevens.

Omdat de onzekerheidsspecificatie van deze gegevens wegens tijdgebrek noodgedwongen een tentatief en onvolledig karakter had, dienen de berekende betrouwbaarheidsfactoren, -intervallen etc. met gepaste zorg geïnterpreteerd en gebruikt te worden. Strikt genomen bieden ze enkel informatie over de situatie waarvoor ze berekend zijn. Deze kenmerkt zich enerzijds door verwaarlozing van allerlei foutenbronnen (eventuele fouten in emissiebijdrages uit kunstmest en huishoudens, landgebruiks- en geografische informatie op 25 x 25 m schaal, emissieverdeling over landgebruikstypen etc. worden niet meebeschouwd), waardoor de fijnschalige berekeningen als gepostuleerde werkelijkheid kunnen worden opgevat waarmee de nominale berekening vergeleken wordt. Anderzijds wordt de uitgevoerde studie gekarakteriseerd door het feit dat de door het LEI-DLO aangeleverde gemeentelijke  $\text{NH}_3$ -emissiecijfers als uitgangspunt genomen zijn, waaromheen additieve onzekerheid gezet is met een specifieke bandbreedte en afhankelijkheidsstructuur die enkel per provincie verschilt; d.w.z. mogelijke systematische afwijkingen in de gemeentelijke cijfers zijn buiten beschouwing gebleven, evenals alternatieve afhankelijkheidsstructuren, die bijv. gekoppeld zijn aan andere geografische indelingen (ecodistricten). Bovendien is de betrouwbaarheid voor de gemeentelijke emissiecijfers tamelijk groot verondersteld, namelijk ongeveer 0.8.



Tegen deze achtergrond is het duidelijk dat de uitgevoerde betrouwbaarheidsanalyse vooral een indicatief karakter heeft, in de zin dat veeleer bepaalde tendensen/trends worden aangeduid (bijv. de invloed van de verschillende foutcategorieën op lokale en regionale schaal), dan dat de 'werkelijkheid' exact weerspiegeld wordt. De analyse geeft een te simplistische en optimistische kijk op de betrouwbaarheid van de emissiecijfers, gezien de vele (soms strikte) veronderstellingen die gehanteerd zijn (vergelijk [10] waarin een beeld geschetst wordt van de mogelijke grootte van de onzekerheden in de  $\text{NH}_3$ -emissiecijfers in Groot Brittanië). Resultaten met een groter realiteitsgehalte kunnen slechts na een meer omvangrijke studie verkregen worden, waarbij gebruik wordt gemaakt van uitgebreidere en realistischere onzekerheidsspecificaties m.b.t. dierlijke mest, kunstmest en huishoudens en de intergemeentelijke afhankelijkheidsstructuur. Dit zal o.a. vereisen dat de LEI-DLO berekeningen van de gemeentelijke  $\text{NH}_3$ -emissiecijfers aan een uitgebreide onzekerheidsanalyse worden onderworpen. Omdat er nauwelijks directe emissiemetingen beschikbaar zijn waarmee de berekende cijfers kunnen worden vergeleken, zal zo'n analyse nog steeds een overwegend 'intern' karakter hebben. Een rechtstreekse validatie, c.q. 'externe' betrouwbaarheidsanalyse is voorlopig nog niet mogelijk. Wel is het mogelijk en wenselijk om de berekende emissiecijfers op een 'indirecte' wijze te toetsen, bijv. door concentratiemetingen te vergelijken met concentratieberekeningen op basis van transportmodellen die gebruik maken van emissiecijfers die recht doen aan de lokale situatie.

De hier gerapporteerde betrouwbaarheidsfactoren dienen om voorgaande redenen dan ook als bovengrenzen beschouwd te worden. Een ruwe alternatieve schatting van de betrouwbaarheid op basis van eenvoudige berekeningen, waarbij tevens de fouten in de mest-aanwending, Landbouwtellingen, hoeveelheden veevoer en N-gehalte van veevoer zijn meegenomen, leidt tot een betrouwbaarheidsfactor van 0.6 voor de emissies per 5 x 5 km grid (mediaan). Dit cijfer dient als een conservatieve, maar wellicht meer realistische schatting van de betrouwbaarheid beschouwd te worden.

## Appendix A: Bewijs van Stelling 1

Het resultaat uit stelling 1(a) volgt eenvoudig via uitwerking van de kwadratensom in vergelijking (2).

Vervolgens kunnen de betrekkingen uit stelling 1(b) rechtstreeks worden afgeleid uit het feit dat

$$\begin{aligned} \overline{\log(y_{pos}(r; \cdot))} &= \frac{1}{N_{sam}} \sum_{j=1}^{N_{sam}} \log(y_{pos}(r; j)) \\ &= \log \left[ \prod_{j=1}^{N_{sam}} y_{pos}(r; j) \right]^{\frac{1}{N_{sam}}} = \log \left[ \overline{y_{pos}(r; \cdot)}^{geom} \right] \end{aligned} \quad (54)$$

Lognormaliteit van  $y_{pos}(r; \cdot)$  impliceert dat  $\log(\frac{y_{pos}(r; \cdot)}{y_{nom}(r)})$  normaal verdeeld is met gemiddelde<sup>30</sup>  $\mathbb{E}[\log(y_{pos}(r; \cdot)) - \log(y_{nom}(r))]$  en variantie  $\text{Var}(\log(y_{pos}(r; \cdot)))$ . Hieruit volgt onmiddellijk de uitdrukking (10) voor het 68 %-betrouwbaarheidsinterval. Gebruikmaking van de uitdrukking van de bias uit stelling 1(b) leidt vervolgens tot de alternatieve formulering (11).

De uitdrukkingen in stelling 1(d) volgen uit het feit dat  $\overline{\log(y_{pos}(r; \cdot))}$  en  $\sigma_*^2(r)$  schattingen zijn voor  $\mathbb{E}[\log(y_{pos}(r))]$  en  $\text{Var}[\log(y_{pos}(r))]$ , en dat deze laatste grootheden voor lognormaal verdeelde  $y_{pos}(r)$  rechtstreeks kunnen worden uitgedrukt in termen van de mediaan  $\text{Med}[y_{pos}(r)]$  en de variatiecoëfficiënt  $\text{VC}[y_{pos}(r)]$  volgens:

$$\mathbb{E}[\log(y_{pos}(r))] = \log(\text{Med}[y_{pos}(r)]) \quad (55)$$

$$\text{Var}[\log(y_{pos}(r))] = \log\left(1 + (\text{VC}[y_{pos}(r)])^2\right) \quad (56)$$

De uitspraak onder stelling 1(e) is evident.

Om de bewering over de kans op mis-classificatie uit stelling 1(f) af te leiden merken we op dat lognormaliteit van  $y_{pos}(r; \cdot)$  impliceert dat  $\log(\frac{y_{pos}(r; \cdot)}{y_{nom}(r)})$  normaal verdeeld is met gemiddelde  $\mu(r)$  en variantie  $\sigma^2(r)$ . Merk op dat  $\sigma_*^2(r)$  uit (5) een schatting is voor  $\sigma^2(r)$ , terwijl  $\mu(r) \approx [\beta_*(r)]^{-1}$ . Uitwerking van de kans op mis-classificatie levert vervolgens:

$$\begin{aligned} \mathbb{P}_{mis} &= \mathbb{P}(y_{pos}(r; \cdot) > H \vee y_{pos}(r; \cdot) < L) \\ &= \mathbb{P}\left(\left[\log\left(\frac{y_{pos}(r; \cdot)}{y_{nom}(r)}\right) - \mu(r)\right] > \left[\log\left(\frac{H}{y_{nom}(r)}\right) - \mu(r)\right] \vee \right. \\ &\quad \left. \left[\log\left(\frac{y_{pos}(r; \cdot)}{y_{nom}(r)}\right) - \mu(r)\right] < \left[\log\left(\frac{L}{y_{nom}(r)}\right) - \mu(r)\right]\right) \\ &= \mathbb{P}\left(X_{stnd} > \frac{\left[\log\left(\frac{H}{y_{nom}(r)}\right) - \mu(r)\right]}{\sigma(r)} \vee X_{stnd} < \frac{\left[\log\left(\frac{L}{y_{nom}(r)}\right) - \mu(r)\right]}{\sigma(r)}\right) \\ &\geq 2 \cdot \mathbb{P}\left(X_{stnd} > \frac{\left[\log\left(\frac{H}{y_{nom}(r)}\right) - \mu(r)\right]}{2 \cdot \sigma(r)} - \frac{\left[\log\left(\frac{L}{y_{nom}(r)}\right) - \mu(r)\right]}{2 \cdot \sigma(r)}\right) \\ &= 2 \cdot \mathbb{P}\left(X_{stnd} > \frac{\log(H/L)}{2 \cdot \sigma(r)}\right) \end{aligned} \quad (57)$$

Hierbij is  $X_{stnd}$  gedefinieerd als:

$$X_{stnd} = \frac{\left[\log\left(\frac{y_{pos}(r; \cdot)}{y_{nom}(r)}\right) - \mu(r)\right]}{\sigma(r)} \quad (58)$$

<sup>30</sup>  $\mathbb{E}$  duidt de verwachtingswaarde (het gemiddelde) van een stochastische grootheid aan.

$X_{std}$  heeft een standaard normale verdeling ( $N(0, 1)$ ) en heeft daarom de eigenschap ( $b > a$ ):

$$\mathbb{P}(X_{std} > b \vee X_{std} < a) \geq 2 \cdot \mathbb{P}\left(X_{std} > \frac{(b-a)}{2}\right) \quad (59)$$

Deze eigenschap is gebruikt bij de ongelijkheid in bovenstaande afleiding.

## Appendix B: Kwantificering van verrasterings- en classificatiefouten

Om bij een betrouwbaarheidsanalyse de effecten van ruimtelijke bewerkingen zoals verrastering en aggregatie nader te bestuderen, dienen de fouten gekwantificeerd te worden die kunnen ontstaan bij verrastering van gemeentekaarten en bij classificatie van het landgebruik voor 500 x 500 m arealen. Omdat verrastering en classificatie betrekking hebben op de toekenning van één discreet attribuut (bijv. gemeente  $i_o$ ; landgebruikstype  $j_o$ ) aan elk van de 500 x 500 m arealen, lijkt een probabilistisch getinte benadering voor deze kwantificering in eerste instantie minder geschikt. Het verdient veeleer de voorkeur om, indien mogelijk, de fout die bij zo'n toekenning gemaakt wordt te kwantificeren door per areaal de fractie van de *werkelijke* oppervlakte van gemeente  $i_o$  of landgebruikstype  $j_o$  te vergelijken met de *toegekende* fractie:

Veronderstel dat de oppervlakte van een specifiek 500 x 500 m rastercel bestaat uit diverse landgebruikstypen  $j \in T$ , die elk in werkelijkheid een fractie  $p^*(j)$  van de rastercel bedekken. Hierbij zijn  $p^*(j)$  getallen tussen 0 en 1 die tot 1 sommeren. Indien een classificatie gehanteerd wordt die slechts één type, zeg  $j_o \in T$ , toekent aan de specifieke rastercel, dan zullen de toegekende fracties, zeg  $p(j), j \in T$ , op één na alle nul zijn. Deze ene heeft waarde 1 toegekend gekregen (d.w.z.  $p(j) = 0; j \neq j_o$ ).

De fout bij deze classificatie wordt nu volledig gekarakteriseerd door de getallen reeks  $\{p(j) - p^*(j)\}_{j \in T}$ . De grootheid

$$\|p - p^*\|_1 := \sum_{j \in T} |p(j) - p^*(j)| \quad (60)$$

levert een maat voor de *totale classificatiefout* bij de specifieke rastercel. Natuurlijk kunnen ook andere maten gehanteerd worden, bijv.

$$\|p - p^*\|_\infty := \max_{j \in T} |p(j) - p^*(j)| \quad (61)$$

Deze laatste maat levert echter geen directe informatie over de bijdrage van de niet maximale classificatiefouten.

### Opmerking 4:

- (i) Beide voorgaande maten zijn sterk aan elkaar gerelateerd. Immers, als  $j_o$  het landgebruikstype is dat eenduidig aan de rastercel wordt toegekend (i.e.  $p(j_o) = 1$ ), dan volgt er (wegens  $\sum_{j \in T} p^*(j) = 1$ ) dat:

$$\|p - p^*\|_1 = 1 - p^*(j_o) + \sum_{j \neq j_o} p^*(j) = 2 \cdot [1 - p^*(j_o)] \quad (62)$$

Bovendien geldt er wegens  $p(j_o) - p^*(j_o) = 1 - p^*(j_o) = \sum_{j \neq j_o} p^*(j) \geq p^*(j) = |p(j) - p^*(j)|$  voor  $j \neq j_o$ , dat:

$$\|p - p^*\|_\infty = 1 - p^*(j_o) \quad (63)$$

Hieruit volgt dus dat:

$$\|p - p^*\|_1 = 2 \cdot \|p - p^*\|_\infty \quad (64)$$

- (ii) Uit de classificatiefout van de afzonderlijke 500 x 500 m rastercellen kan eenvoudig de 'classificatiefout' worden afgeleid voor grotere gridcellen. Beschouw bijv. de oppervlakte  $O^j(r)$ , die op basis van een classificatie is 'toegekend' aan een specifiek landgebruikstype  $j$  binnen een gridcel

$r$  (bijv. een 5 x 5 km LLO grid). Deze oppervlakte kan worden uitgedrukt in termen van de afzonderlijke bijdragen van de 500 x 500 m rastercellen binnen deze gridcel:

$$O^j(r) := \sum_{k \in R(r)} p(j; k) \cdot A_o \quad (65)$$

Hierbij duidt  $p(j; k)$  de ‘toegekende’ fractie van grondtype  $j$  binnen de  $k$ -de 500 x 500 m rastercel aan.  $R(r)$  is de index-verzameling van de 500 x 500 m rastercellen die binnen de  $r$ -de gridcel vallen.  $A_o$  tenslotte duidt de oppervlakte van een 500 x 500 m rastercel aan.

De ‘werkelijke’ oppervlakte van landgebruikstype  $j$  binnen de  $r$ -de gridcel is:

$$O_{\#}^j(r) := \sum_{k \in R(r)} p^*(j; k) \cdot A_o \quad (66)$$

waarbij  $p^*(j; k)$  de ‘werkelijke’ fractie van grondtype  $j$  binnen de  $k$ -de 500 x 500 m rastercel aanduidt.

De fout in de oppervlakte specificatie van gridcel  $r$ , is nu een functie van de classificatiefouten in alle 500 x 500 m rastercellen die deel uitmaken van deze gridcel:

$$O^j(r) - O_{\#}^j(r) = \sum_{k \in R(r)} [p(j; k) - p^*(j; k)] \cdot A_o \quad (67)$$

■

Een alternatieve aanpak wordt gegeven in [13], waar uitgaande van een *possibilistische* context, de ‘fuzzyness’- en de ‘confusion’-graad gehanteerd wordt om de ‘onzekerheidsaspecten’ van classificatie c.q. verrastering te beschrijven. Deze maten zijn sterk gerelateerd aan de bovengenoemde classificatiefouten  $[p(j) - p^*(j)]_{j \in T}$ :

- Immers, indien het meest waarschijnlijke (d.w.z. dominante) type ( $j_o := \arg \max_{j \in T} [p^*(j)]$ ) toegekend wordt aan de onderhavige rastercel, dan is de classificatiefout (i.e.  $1 - p^*(j_o)$ ) voor dit specifieke type gelijk aan de ‘fuzzyness-graad’. Merk op dat deze classificatiefout bovendien de maximale is, i.e.  $\|p - p^*\|_{\infty}$  (zie de voorgaande opmerking).
- De ‘confusion-graad’ is gelijk aan de som van de (in *absolute* waarde) *twee grootste* classificatiefouten<sup>31</sup>, en geeft informatie over de *eenduidigheid* van de classificatie. Indien deze fouten dicht bij elkaar liggen, dan is de toekenning van de dominante klasse tamelijk arbitrair en kan aanleiding geven tot veel verwarring (‘confusion’). Als maat voor de betrouwbaarheid van de classificatie, i.e. voor de classificatiefout, lijkt de ‘confusion-graad’ minder geschikt, omdat hij slechts de twee grootste classificatiefouten beschouwt. Het lijkt logischer om hiervoor de absolute som van *alle* classificatiefouten, d.w.z.  $\|p - p^*\|_1$ , te nemen. Zoals in de vorige opmerking werd aangetoond komt dit in wezen neer op het hanteren van de norm  $\|p - p^*\|_{\infty}$ , en dus van de fuzzyness-graad.

De kwantificering van classificatie- en verrasteringsfouten die hierboven werd voorgesteld, vereist *voorkennis over de werkelijke oppervlakte* die vaak niet voorhanden is. In de praktijk kan deze oppervlakte hoogstens worden ingeschat. Indien deze schatting nauwkeurig is, dan kunnen de onzekerheidsaspecten achterwege blijven. In alle andere gevallen dient echter bij de kwantificering rekening gehouden te worden met de inherente onzekerheid van de schatting.

<sup>31</sup>Er geldt dat de grootste classificatie-fout gelijk is aan  $p(j_o) - p^*(j_o) = 1 - p^*(j_o)$ , terwijl alle overige classificatie fouten gelijk zijn aan  $p(j) - p^*(j) = -p^*(j)$ .

Veelal gebeurt dit in een probabilistische context, middels het aangeven van kansverdelingen c.q. 'likelihoods' die de schattingsfouten karakteriseren. Hierbij dient rekening gehouden te worden met de aanwezigheid van allerlei correlaties, zowel binnen rastercellen<sup>32</sup>, alsook tussen rastercellen onderling. Naast deze onzekerheid die debet is aan gebrekkige gegevens (onnauwkeurigheid), is er vaak ook nog sprake van een niet-eenduidigheid van de classificatie: bijv. wanneer is een stuk grond van het type A, en wanneer van het type B?? Dit alles heeft tot gevolg dat een adequate kwantificering van classificatie- en verrasteringsfouten vaak moeilijk kan zijn indien geschikte voorkennis ontbreekt. In dit rapport wordt de fout bij classificatie of verrastering van 500 x 500 m rastercellen rechtstreeks ingeschat door uit te gaan van de landgebruiks-informatie die op 25 x 25 m pixels beschikbaar is uit de Remote Sensing data  $RS^o$ , of van de gemeentelijke gebiedsinformatie op 25 x 25 m grids van geografische kaarten  $G_{25}$ . Aannemende dat deze informatie op dat nivo voldoende nauwkeurig/betrouwbaar is, kan hiermee vervolgens voor de 500 x 500 m rastercellen bepaald worden, welke fractie van het oppervlak van deze rastercellen tot een specifiek landgebruikstype hoort of tot een specifieke gemeente.

Echter, de klasse toekenning bij Remote Sensing data zal in de praktijk zeker niet foutloos verlopen, en met name als het aantal klassen in het satellietbeeld groot is, is bovenstaande veronderstelling discutabel. Bovendien is, gezien de beperkte resolutie van de gemeentekaarten, een nauwkeurige verrastering naar 25 x 25 m grids in het algemeen niet goed mogelijk. Dientengevolge zijn de hier gehanteerde aannames van foutvrije Remote Sensing informatie ( $RS^o$ ), en fijn-rastering ( $G_{25}^o$ ) discutabel. Omdat er echter weinig tot geen specifieke informatie beschikbaar is over de betreffende fouten zullen deze aannames toch gehandhaafd blijven.

---

<sup>32</sup>Omdat het totaaloppervlak per 500 x 500 m rastercel vastligt, zullen schattingsfouten binnen een rastercel onderling gecorreleerd zijn.

## Appendix C: Specificatie bandbreedte voor dierlijke NH<sub>3</sub>-emissiecijfers

Het LEI-DLO bepaalt aan de hand van simulaties met mest- en ammoniakmodellen de jaargemiddelde gemeentelijke NH<sub>3</sub>-emissiecijfers uit dierlijke mest. Bij deze berekeningen wordt gebruik gemaakt van gegevens uit de Landbouwtellingen (CBS) die op bedrijfsnivo beschikbaar zijn, en van *landelijk gemiddelde* waarden voor excretiefactoren, emissiefactoren, penetratiegraden etc., die door het RIVM als uitgangspunten zijn vastgesteld op basis van informatie van het CBS, LEI-DLO, IKC Veehouderij etc. ([2]). De 'werkelijke' waarden van deze factoren op regionaal, gemeentelijk en bedrijfsnivo kunnen forse afwijkingen vertonen van deze nominale cijfers. Deze afwijkingen worden veroorzaakt door het feit dat deze factoren berekend worden op basis van allerlei veronderstellingen en vereenvoudigingen over voersamenstelling, N-gehalte in voer, mestproductie, mestopslag, die weliswaar op landelijk nivo redelijk zijn, maar op lokaal nivo soms forse verschillen laten zien tussen modelberekening en werkelijkheid. Ook de gegevens over mestaanwending, mesttransport kunnen grote fouten vertonen, omdat het ruimtelijk patroon gesimuleerd wordt met een transport model dat slechts een benadering van de werkelijkheid geeft. Daarnaast bevatten de Landbouwtellingen en de opgegeven gehalten natuurlijk ook nog fouten. Het zal in de praktijk nagenoeg onmogelijk zijn om te specificeren wat de exacte fouten in de berekende cijfers zijn, tenzij na een nauwkeurige en tijdrovende onzekerheidsanalyse van deze berekeningen. Wegens de beperkte tijd en menskracht die ter beschikking stond voor de betrouwbaarheidsanalyse van de N-keten, was zo'n volledige analyse onbereikbaar. Daarom werd volstaan met een *ruwe schatting* van de gemaakte fouten, en werd de onzekerheid in de cijfers gespecificeerd in de vorm van een *relatieve 'bandbreedte'* rondom hun nominale waarde (d.w.z. het berekende gemiddelde). In deze appendix wordt weergegeven hoe deze bandbreedte bepaald is voor de NH<sub>3</sub>-emissie bijdrage uit dierlijke mest.

De volgende procedure is daarbij gevolgd: Allereerst is de dierlijke mest in een *drietal hoofdcategorieën* gesplitst (rundvee, varkens, pluimvee), en is voor elke categorie een spreidingsfactor/bandbreedte gespecificeerd. Vervolgens is uit deze informatie een bandbreedte bepaald voor de *totale* dierlijke NH<sub>3</sub>-emissie. In het navolgende wordt in detail beschreven hoe dit tot stand komt. Aan het eind staan de belangrijkste aspecten kort samengevat.

### [I] Specificatie bandbreedtes van NH<sub>3</sub>-emissie per afzonderlijke diercategorie:

Om een indruk te krijgen van de afwijkingen in de nominale gemeentelijke emissiecijfers t.o.v. hun 'werkelijke' waarden, werden enkele spreadsheet berekeningen uitgevoerd, waarbij voor de belangrijkste factoren die de NH<sub>3</sub>-emissie beïnvloeden waarden worden gehanteerd die rondom de referentie (nominale) waarden liggen die bij de LEI-DLO berekeningen gehanteerd worden. Zo werd bijv. de N-toevoer (hoeveelheid voer en N-gehalte) op  $100 \pm 10\%$  van de nominale (gemiddelde) waarde gezet voor varkens en pluimvee, terwijl de melkproductie van melkvee op  $100 \pm 25\%$  gezet is. Voor de andere belangrijke factoren die de NH<sub>3</sub>-emissie beïnvloeden, zoals het rantsoen en het beweidingssysteem voor melkvee, en de huisvesting bij vleesvarkens en leghennen, werden telkens een tweetal alternatieve opties gekozen 'rond' de nominale optie die bij de LEI-DLO berekeningen gehanteerd wordt. Op deze manier wordt geprobeerd om een indruk te krijgen van de range in de NH<sub>3</sub>-emissie(factoren) ten gevolge van (realistische) afwijkingen in de landelijk gemiddelde excretiefactoren etc. Zie opmerking 5.

'Variabiliteit' ten gevolge van verschillende mestaanwendingstechnieken werd niet mee-beschouwd, evenmin als 'variabiliteit' in de NH<sub>3</sub>-vervluchtigingsfactor per techniek die

bij de berekeningen van NH<sub>3</sub>-emissie t.g.v. mestaanwending gebruikt wordt. Deze factor werd eenduidig vastgelegd op 25 % voor melkvee, en op 18 % voor vleesvarkens en leghennen. Deze waarden weerspiegelen de mestregelgeving in 1990, en komen overeen met proefveldmetingen tijdens uitrij-periodes.

Doorrekenend met de diverse alternatieve instelwaarden werd per diercategorie (rundvee, varkens, pluimvee) een 'bandbreedte' rond de nominale NH<sub>3</sub>-emissiecijfers bepaald. Dit resulteerde bij benadering in een bandbreedte van 100±25% voor rundvee, 100±20% voor varkens en 100 ± 30% voor pluimvee (zie opmerking 5). Verondersteld werd dat ongeveer 70 % van de gevallen binnen deze bandbreedtes valt (dit komt overeen met 1 maal de standaarddeviatie). Verder werd aangenomen dat de verdeling binnen die bandbreedte bij benadering *normaal* is.

**Opmerking 5:** De bandbreedte voor de ammoniak emissie van de diverse diersoorten wordt bepaald door eenvoudige spreadsheet berekeningen uit te voeren waarbij de belangrijkste factoren die de emissie beïnvloeden op waarden gezet worden rond hun referentiewaarden. Het cijfermateriaal waarmee de berekeningen zijn uitgevoerd heeft betrekking op het jaar 1990, en leverde voor de diverse mestsoorten<sup>33</sup> de volgende specificaties en resultaten op:

- (a) **Rundvee:** Melkvee levert 84 % van deze emissie, en vleesvee 16 %. De voornaamste factoren die de NH<sub>3</sub>-emissie bij melkvee beïnvloeden zijn:
- *Melkproductie:* De gemiddelde productie per melkkoe bedraagt 6050 liter melk per jaar; gerekend is met een bandbreedte van 25 %.
  - *Rantsoen:* Het gemiddelde rantsoen is gebaseerd op een evenredige verdeling van alle ruw- en krachtvoer over alle melkkoeien. Voor het *lage* melkproductie nivo heeft de koe minder voer nodig: het gemiddelde rantsoen bevat geen snijmaïs meer. Voor het *hoge* melkproductie nivo heeft de koe meer voer nodig: het gemiddelde rantsoen is aangevuld met 40 % meer snijmaïs en 25 % meer krachtvoer.
  - *Beweidingssysteem:* Bij gemiddelde beweiding komt 30 % van de in de zomer geproduceerde mest in de mestkelder terecht, waardoor de emissie in de weide afneemt en de emissie uit de stal en bij mestaanwenden toeneemt. Er zijn twee varianten door-gerekend, namelijk (i) onbeperkt beweiden [O] in de zomer waarbij 10 % van de mest in de kelder terechtkomt, en (ii) beperkt beweiden [B] in de zomer waarbij 50 % van de mest in de kelder terecht komt.

In tabel 9 staan deze keuzen vermeld, en is weergegeven wat de bijbehorende totale NH<sub>3</sub>-emissie is. Hierbij is verondersteld dat bij mestaanwending 25 % van de stikstof vervluchtigt. Uit deze resultaten wordt de bandbreedte voor melkvee, en voor rundvee vastgesteld op 25 %.

Melkvee	Basis	Laag		Hoog	
Melkproductie	6050 liter	75 %		125 %	
Excretie	136 kg N	90 %		113 %	
Beweidingssysteem	Gemiddeld	O	B	O	B
Totale NH <sub>3</sub> -emissie	37.5 kg NH <sub>3</sub>	78 %	102 %	98 %	128 %

Tabel 9: Specificaties voor berekening bandbreedte van NH<sub>3</sub>-emissie bijdrage rundvee.  
Bron: K.W. van der Hoek (1995), persoonlijke communicatie.

- (b) **Varkens:** Vleesvarkens leveren 67 % van deze emissie, en zeugen met biggen 33 %. De voornaamste factoren die de NH<sub>3</sub>-emissie bij vleesvarkens beïnvloeden zijn:

<sup>33</sup> Landelijk gezien levert rundvee ongeveer 60% van de dierlijke NH<sub>3</sub>-emissie, varkens 32% en pluimvee 8%.



- *Voerinput*: Een vleesvarken verbruikt per jaar 755 kg voer met 25,8 gram N per kg. Voor de N-toevoer (i.e. hoeveelheid voer en N-gehalte) is gerekend met een bandbreedte van 10 %.
- *Huisvesting*: De gemiddelde stal-emissie is gebaseerd op een gewogen mix (40:60 %) van half- en volledig roostervloerstallen. Verder zijn twee extreme situaties bekeken, namelijk (i) alle stallen zijn halfroostervloerstallen (Half) met een stal-emissie van 2,5 kg NH<sub>3</sub> per dierplaats per jaar, en (ii) alle stallen zijn volledige roostervloerstallen (Volledig) met een emissie van 3,0 kg NH<sub>3</sub>.

Bij zeugen met biggen is de *voerinput* de belangrijkste factor die de NH<sub>3</sub>-emissie beïnvloed. Wederom is hierbij voor de N-toevoer gerekend met een bandbreedte van 10 %.

In tabel 10 staan bovengenoemde keuzen vermeld, en is weergegeven wat de bijbehorende totale NH<sub>3</sub>-emissie is. Hierbij is verondersteld dat bij mestaanwending 18 % van de stikstof vervluchtigt. Uit deze resultaten wordt de bandbreedte voor (vlees)varkens vastgesteld op 20 %.

Vleesvarkens	Basis	Laag		Hoog	
Voerinput	755 kg	90 %		110 %	
Excretie	13.2 kg N	85 %		115 %	
Huisvestingssysteem	Gemiddeld	Half	Volledig	Half	Volledig
Totale NH <sub>3</sub> -emissie	5.24 kg NH <sub>3</sub>	81 %	87 %	109 %	118 %
Zeugen met biggen	Basis	Laag		Hoog	
Voerinput	1713 kg	90 %		110 %	
Excretie	31.7 kg N	86 %		114 %	

Tabel 10: Specificaties voor berekening bandbreedte van NH<sub>3</sub>-emissie bijdrage varkens.  
Bron: K.W. van der Hoek (1995), persoonlijke communicatie.

- (c) **Pluimvee**: Leghennen leveren 67 % van deze emissie, en vleeskuikens 33 %. De voornaamste factoren die de NH<sub>3</sub>-emissie bij leghennen beïnvloeden zijn:
- *Voerinput*: Voor de N-toevoer (i.e. hoeveelheid voer en N-gehalte) is gerekend met een bandbreedte van 10 %.
  - *Huisvesting*: De gemiddelde stal-emissie is gebaseerd op een gewogen mix van alle staltypen voor leghennen. Verder is er de situatie bekeken waarbij het aantal deep-pitstallen en stallen met grondhuisvesting gehalveerd wordt. Dit resulteert in een instelling die 35 % minder stal-emissie geeft dan de basisinstelling. Ook is de situatie beschouwd waarbij dit aantal stallen verdubbeld is, hetgeen leidt tot een instelling die 50 % hoger ligt dan de referentie instelling.

Bij vleeskuikens is wederom de *voerinput* de belangrijkste factor die de NH<sub>3</sub>-emissie beïnvloed. Ook hierbij is voor de N-toevoer gerekend met een bandbreedte van 10 %.

In tabel 5: staan bovengenoemde keuzen vermeld, en is de bijbehorende totale NH<sub>3</sub>-emissie weergegeven. Hierbij is verondersteld dat bij mestaanwending 18 % van de stikstof vervluchtigt. Uit deze resultaten wordt de bandbreedte voor pluimvee gelijk gesteld aan 30 %.



**Opmerking 6:** De bandbreedte in de emissie bijdrages van de diverse diercategorieën is volgens voorgaande opmerking berekend door het model te runnen met diverse combinaties instelwaarden. De volgende beschouwing dient ter motivatie/onderbouwing van deze aanpak: Stel dat

Leghennen	Basis	Laag		Hoog	
Voerinput	42 kg	90 %		110 %	
Excretie	0.68 kg N	85 %		115 %	
Huisvestingssysteem	Gemiddeld	- 35 %	+ 50 %	- 35 %	+ 50 %
Totale NH <sub>3</sub> -emissie	0.22 kg NH <sub>3</sub>	75 %	100 %	102 %	136 %
Vleeskuikens (jaar 1991)	Basis	Laag		Hoog	
Voerinput	30 kg	90 %		110 %	
Excretie	0.60 kg N	82 %		118 %	

Tabel 11: Specificaties voor berekening bandbreedte van NH<sub>3</sub>-emissie bijdrage pluimvee.  
Bron: K.W. van der Hoek (1995), persoonlijke communicatie.

parameter  $a$  met kans  $P_a$  tussen  $a_1$  en  $a_2$  ligt, en dat parameter  $b$  met kans  $P_b$  tussen  $b_1$  en  $b_2$  ligt. Laat de modeluitkomst  $y$  een functie van  $a$  en  $b$ , zijn, zeg  $y = F(a, b)$ . Indien  $F(\cdot, \cdot)$  *monotoon* (stijgend of dalend) is in  $a, b$ , en indien  $a$  en  $b$  *onafhankelijk* zijn van elkaar, dan zal de modeluitkomst  $y$  met een kans  $P$ , die  $\geq P_a \cdot P_b$  is, liggen tussen  $[\text{Min}_{i,j=1,2} F(a_i, b_j), \text{Max}_{i,j=1,2} F(a_i, b_j)]$ . Zo wordt een (ruwe) schatting verkregen van een 'bandbreedte' op basis van het doorrekenen van 'extreme' varianten. Merk op dat de verkregen bandbreedte in wezen *conservatief* is (worst-case gebaseerd): de *werkelijke* kans  $P$  dat de modeluitkomst binnen de berekende grenzen valt kan groter zijn dan  $P_a \cdot P_b$ .

Bedenk wel dat bovenstaand resultaat *strikt genomen slechts geldig* is voor 'monotone' modellen met onafhankelijke parameters. Indien niet aan deze voorwaarden voldaan is, dan kan de aldus bepaalde 'bandbreedte' een inadequaat beeld geven. ■

Tot slot wordt in formulevorm weergegeven hoe nu, in principe althans<sup>34</sup>, de NH<sub>3</sub>-emissie per diercategorie bepaald kan worden uit de nominale emissiecijfers.

Laat daartoe  $E_{d,x}^o(i; *)$  de *nominale*<sup>35</sup> dierlijke NH<sub>3</sub>-emissie zijn voor diercategorie  $x$  in gemeente  $i$ . Hierbij kan  $x$  de waarden  $r, v, p$  aannemen, refererend naar 'rundvee', 'varkens' en 'pluimvee'. De '*werkelijke*' NH<sub>3</sub>-emissie  $E_{d,x}(i; *)$  voor diercategorie  $x$  in gemeente  $i$  varieert nu binnen een bandbreedte rondom deze nominale waarde, en wordt beschreven als:

$$E_{d,x}(i; *) := E_{d,x}^o(i; *) \cdot [1 + g_{d,x}(i) \cdot \sigma_{d,x}(i)] \quad \text{voor } x = r, v, p \quad (68)$$

waarbij  $g_{d,x}(i)$  een standaardnormale verdeling  $N(0, 1)$  heeft, en waarbij  $\sigma_{d,x}(i)$  de 'relatieve' bandbreedte karakteriseert voor de NH<sub>3</sub>-bijdrage van diercategorie  $x$ , in gemeente  $i$ .  $g_{d,x}(i)$  geeft in feite weer wat de daadwerkelijke relatieve afwijking is t.o.v. de nominale waarde, gemeten in termen van de bandbreedte  $\sigma_{d,x}(i)$ . Deze berekeningswijze komt neer op de veronderstelling dat zo'n 68 % van alle emissievariatiën binnen de relatieve bandbreedte ( $E_{d,x}^o(i; *) \cdot [1 \pm \sigma_{d,x}(i)]$ ) vallen, dat ze een normale verdeling hebben, en dat ze geen systematische afwijking met de nominale situatie vertonen. De 'bandbreedtes'  $\sigma_{d,x}(i)$  die hierbij voor de diverse mest-categorieën gehanteerd worden, zijn op basis van de voorgaande berekeningen gelijk genomen aan (zie opmerking (5)):

$$\sigma_{d,r}(i) = .25, \sigma_{d,v}(i) = .2, \sigma_{d,p}(i) = .3 \quad (69)$$

Merk op dat deze bandbreedtes voor *alle* gemeenten dezelfde zijn, i.e. onafhankelijk van  $i$ .

<sup>34</sup> Met de oude versie van het ammoniak-model werd enkel de *totale* dierlijke NH<sub>3</sub>-emissie bepaald. Inmiddels kan met de nieuwe versie ook de bijdrage per diercategorie berekend worden.

<sup>35</sup> De subscript <sup>o</sup> verwijst naar de nominale situatie, d.w.z. gebaseerd op landelijk gemiddelde waarden voor excretiefactoren, emissiefactoren etc.

[II] Specificatie van bandbreedte voor totale dierlijke NH<sub>3</sub> emissie:

De bandbreedte  $\sigma_{d,*}(i)$  van de totale dierlijke NH<sub>3</sub>-emissie, per gemeente, wordt bepaald uit de afzonderlijke bandbreedtes  $\sigma_{d,x}(i)$ , op basis van de relatieve bijdrage  $f_{d,x}^o(i;*)$

$$f_{d,x}^o(i;*) := \frac{E_{d,x}^o(i;*)}{E_{d,*}^o(i;*)} \quad \text{voor } x = r, v, p \quad (70)$$

van de afzonderlijke diersoorten aan de totale dierlijke NH<sub>3</sub>-emissie per gemeente. Hierbij is  $E_{d,*}^o(i;*)$  de *nominale totale* dierlijke NH<sub>3</sub>-emissie voor gemeente  $i$ , terwijl  $E_{d,x}^o(i;*)$  de bijdrage van diercategorie  $x$  behelst. Omdat

$$E_{d,*}^o(i;*) = \sum_x E_{d,x}^o(i;*) \quad (71)$$

geldt er dat de fracties  $f_{d,x}^o(i;*) \geq 0$  zijn voor  $x = r, v, p$ , en tot 1 sommeren.

Op basis van uitdrukking (68) voor de 'werkelijke' dierlijke NH<sub>3</sub>-emissie van diercategorie  $x$ , volgt dan dat de 'werkelijke' totale dierlijke NH<sub>3</sub>-emissie voor gemeente  $i$  (d.w.z.  $E_{d,*}(i;*)$ ) gelijk is aan:

$$\begin{aligned} E_{d,*}(i;*) &:= \sum_x E_{d,x}(i;*) = \sum_x E_{d,x}^o(i;*) \cdot [1 + g_{d,x}(i) \cdot \sigma_{d,x}(i)] \\ &= E_{d,*}^o(i;*) \cdot \left( \sum_x f_{d,x}^o(i;*) \cdot [1 + g_{d,x}(i) \cdot \sigma_{d,x}(i)] \right) \\ &= E_{d,*}^o(i;*) \cdot \left( 1 + \sum_x [g_{d,x}(i) \cdot f_{d,x}^o(i;*) \cdot \sigma_{d,x}(i)] \right) \end{aligned} \quad (72)$$

Dit resultaat suggereert hoe de bandbreedte  $\sigma_{d,*}(i)$  voor de totale dierlijke NH<sub>3</sub>-emissie gedefinieerd kan worden in termen van de bandbreedtes  $\sigma_{d,x}(i)$  voor de afzonderlijke diercategorieën. Twee manieren worden onderscheiden:

- (a) *Volledig afhankelijke bijdragen (versterking)*: Hierbij wordt verondersteld dat als de NH<sub>3</sub>-bijdrage van diercategorie  $x$  10 % meer bedraagt dan nominaal, dat dit dan voor diercategorie  $y$  ook zo zal zijn, d.w.z. in formule (72) is  $g_{d,x}(i) \equiv g_{d,y}(i)$  voor  $x \neq y$ . Uit (7) volgt dan onmiddellijk dat de bandbreedte (zeg  $\tilde{\sigma}_{d,*}(i)$ ) van de totale dierlijke NH<sub>3</sub>-emissie gelijk is aan:

$$\tilde{\sigma}_{d,*}(i) := \sum_x f_{d,x}^o(i;*) \cdot \sigma_{d,x}(i) \quad (73)$$

De 'werkelijke' totale dierlijke NH<sub>3</sub>-emissie  $E_{d,*}(i;*)$  in gemeente  $i$  varieert binnen deze bandbreedte rondom de nominale waarde  $E_{d,*}^o(i;*)$ , en wordt beschreven als:

$$E_{d,*}(i;*) := E_{d,*}^o(i;*) \cdot [1 + g_{d,*}(i) \cdot \tilde{\sigma}_{d,*}(i)] \quad (74)$$

waarbij  $g_{d,*}(i)$  een standaardnormaal ( $N(0, 1)$ ) verdeelde grootheid is. Merk op dat de variantie van de aldus bepaalde dierlijke NH<sub>3</sub>-emissie  $E_{d,*}(i;*)$  gelijk is aan:

$$\sigma_{E_{d,*}(i;*)}^2 = [E_{d,*}^o(i;*)]^2 \cdot \tilde{\sigma}_{d,*}^2(i) \quad (75)$$

- (b) *Volledig onafhankelijke bijdragen (uitmiddeling)*: Hierbij wordt verondersteld dat de NH<sub>3</sub>-bijdragen van de diverse diercategorieën volledig onafhankelijk van elkaar zijn, d.w.z. in formule (72) is  $g_{d,x}(i)$  onderling onafhankelijk van  $g_{d,y}(i)$  voor  $x \neq y$ . Omdat de variantie van de som van onafhankelijke grootheden gelijk is aan de som van de varianties, geldt er dat de resulterende bandbreedte (zeg  $\sigma_{d,*}(i)$ ) van de totale dierlijke NH<sub>3</sub>-emissie gelijk is aan:

$$\sigma_{d,*}(i) := \left( \sum_x [f_{d,x}^o(i;*) \cdot \sigma_{d,x}(i)]^2 \right)^{\frac{1}{2}} \quad (76)$$

De ‘werkelijke’ totale dierlijke  $\text{NH}_3$ -emissie  $E_{d,*}(i;*)$  in gemeente  $i$  varieert binnen deze bandbreedte rondom de nominale waarde  $E_{d,*}^o(i;*)$ , en wordt beschreven als:

$$E_{d,*}(i;*) := E_{d,*}^o(i;*) \cdot [1 + g_{d,*}(i) \cdot \sigma_{d,*}(i)] \quad (77)$$

waarbij  $g_{d,*}(i)$  weer een standaardnormaal ( $N(0, 1)$ ) verdeelde grootheid is. Merk op dat de variantie van de aldus berekende dierlijke  $\text{NH}_3$ -emissie  $E_{d,*}(i;*)$  gelijk is aan:

$$\sigma_{E_{d,*}(i;*)}^2 = [E_{d,*}^o(i;*)]^2 \cdot \sigma_{d,*}^2(i) \quad (78)$$

■

Deze twee aanpakken geven dus aanleiding tot verschillende uitdrukkingen voor de bandbreedte van de totale dierlijke  $\text{NH}_3$ -emissies. Eenvoudig volgt er dat de bandbreedte onder aanpak (b) kleiner is dan onder aanpak (a):

$$\sigma_{d,*}(i) \leq \tilde{\sigma}_{d,*}(i) \quad (79)$$

Dit hangt samen met het feit dat de aanpak onder (a) in feite een ‘conservatieve’ (‘worst case’) aanpak is, waarbij verondersteld wordt dat de emissie-varianties in de afzonderlijke diercategorieën allen precies dezelfde kant opwerken, en elkaar dus *versterken*. In situatie (b) daarentegen zijn de varianties onafhankelijk en treedt er een *uitmiddeling* op. Omdat het aannemelijk is dat de onzekerheden in de afzonderlijke emissies  $E_{d,x}(i;*)$  in werkelijkheid inderdaad in hoge mate onafhankelijk zijn, krijgt de definitie van de bandbreedte uit situatie (b), d.w.z. uit vgl. (76), de voorkeur. Merk op dat, ook al verschilt de bandbreedte  $\sigma_{d,x}(i)$  per diercategorie niet per gemeente  $i$ , de bandbreedte voor de *totale* dierlijke  $\text{NH}_3$ -emissie  $\sigma_{d,*}(i)$  toch *per gemeente kan verschillen*, omdat hij afhangt van de verdeelsleutels  $f_{d,x}^o(i;*)$  die per gemeente anders kunnen zijn; zie formules (73) en (76). De afschattingen uit de volgende stelling geven enige informatie over de mogelijke varianties in de bandbreedtes,  $\sigma_{d,*}(i)$ , zowel als  $\tilde{\sigma}_{d,*}(i)$ , van de diverse gemeenten ten gevolge van deze verdeelsleutels:

**Stelling 2** Voor de bandbreedtes gedefinieerd in (73) en (76) geldt:

$$\underline{\sigma}_{d,*}(i) \leq \tilde{\sigma}_{d,*}(i) \leq \bar{\sigma}_{d,*}(i) \quad (80)$$

$$\left( \sum_x [f_{d,x}^o(i;*)]^2 \right)^{\frac{1}{2}} \cdot \underline{\sigma}_{d,*}(i) \leq \sigma_{d,*}(i) \leq \left( \sum_x [f_{d,x}^o(i;*)]^2 \right)^{\frac{1}{2}} \cdot \bar{\sigma}_{d,*}(i) \quad (81)$$

$$\left( \sum_x \sigma_{d,x}^{-2}(i) \right)^{-\frac{1}{2}} \leq \sigma_{d,*}(i) \leq \bar{\sigma}_{d,*}(i) \quad (82)$$

Hierbij is

$$\underline{\sigma}_{d,*}(i) := \min_x [\sigma_{d,x}(i)] \quad (83)$$

$$\bar{\sigma}_{d,*}(i) := \max_x [\sigma_{d,x}(i)] \quad (84)$$

■

De eerste twee ongelijkheden volgen eenvoudig uit de definities van  $\sigma_{d,*}(i)$  en  $\tilde{\sigma}_{d,*}(i)$ . De laatste ongelijkheid volgt uit het feit dat het maximum en het minimum van  $\sum_k (f_k^2 \sigma_k^2)$  over  $f_k \geq 0$  met restrictie  $\sum_k f_k = 1$ , gelijk zijn aan  $\max_k [\sigma_k^2]$ , en  $(\sum_l \sigma_l^{-2})^{-1}$  respectievelijk<sup>36</sup>.

<sup>36</sup>Dit kan worden afgeleid via toepassing van de Lagrange multiplier methode, waarbij minimalisatie van  $\sum_k (f_k^2 \sigma_k^2)$  onder de restrictie  $\sum_k f_k = 1$  wordt vertaald in minimalisatie zonder restrictie van de uitdrukking  $[\sum_k (f_k^2 \sigma_k^2)] - \lambda \cdot [1 - \sum_k f_k]$ . Het minimum wordt aangenomen voor  $f_k = \sigma_k^{-2} \cdot [\sum_l \sigma_l^{-2}]^{-1}$ .

Deze stelling geeft aan dat de bandbreedte  $\tilde{\sigma}_{d,*}(i)$  kan variëren tussen de kleinste en grootste bandbreedte van de afzonderlijke diersoorten  $\sigma_{d,x}(i)$ . Voor  $\sigma_{d,*}(i)$  ligt dit ingewikkelder, en speelt ook nog de verdeelsleutel  $f_{d,x}^o(i,*)$  een rol. De laatste afchatting geeft aan tussen welke grenzen  $\sigma_{d,*}(i)$  maximaal kan liggen voor alle mogelijk denkbare verdeelsleutels.

Deze stelling geeft aanleiding tot het volgende specifieke resultaat:

### Stelling 3

- (a) Indien de bandbreedtes  $\sigma_{d,x}(i)$  voor de afzonderlijke diersoorten gelijk zijn aan elkaar, zeg  $\sigma^o$ , dan geldt er dat:

$$\tilde{\sigma}_{d,*}(i) = \sigma^o \quad (85)$$

$$\sigma_{d,*}(i) = \left( \sum_x [f_{d,x}^o(i;*)]^2 \right)^{\frac{1}{2}} \cdot \sigma^o \quad (86)$$

Bovendien geldt:

$$\sqrt{\frac{1}{3}} \cdot \sigma^o \leq \sigma_{d,*}(i) \leq \sigma^o \quad (87)$$

- (b) Indien de afzonderlijke bandbreedtes gelijk zijn aan  $\sigma_{d,r}(i) = .25$ ,  $\sigma_{d,v}(i) = .2$  en  $\sigma_{d,p}(i) = .3$ , dan geldt er dat

$$.2 \leq \tilde{\sigma}_{d,*}(i) \leq .3 \quad (88)$$

$$.2 \times \left( \sum_x [f_{d,x}^o(i;*)]^2 \right)^{\frac{1}{2}} \leq \sigma_{d,*}(i) \leq .3 \times \left( \sum_x [f_{d,x}^o(i;*)]^2 \right)^{\frac{1}{2}} \quad (89)$$

$$.1385 \leq \sigma_{d,*}(i) \leq .3 \quad (90)$$

■

Dit resultaat verduidelijkt tussen welke waarden de bandbreedtes voor de diverse gemeenten maximaal kunnen variëren in de specifieke situaties die beschouwd worden.

Tot slot wordt, op basis van voorgaande discussie, samengevat hoe de variaties in de gemeentelijke NH<sub>3</sub>-emissiecijfers voor dierlijke mest bepaald worden:

### Resumé:

De procedure ter bepaling van de variaties in de dierlijke NH<sub>3</sub>-emissies verloopt als volgt:

- (i) Bepaal allereerst de verdeelsleutels  $f_{d,x}^o(i;*)$  voor  $x = r, v, p$ , die weergeven welk deel van de totale nominale dierlijke NH<sub>3</sub>-emissie in gemeente  $i$  afkomstig is van diersoort  $x$  ( $x = r, v, p$  refereert naar de soorten 'rundvee', 'varkens' en 'pluimvee'). Omdat het LEI-DLO per gemeente enkel de *totale* nominale dierlijke NH<sub>3</sub>-emissie aanlevert, niet gedifferentieerd naar diersoort, kunnen deze verdeelsleutels niet rechtstreeks volgens formule (70) uit de gemeentelijke NH<sub>3</sub>-emissiegegevens bepaald worden. Daarom is besloten om de verdeelsleutels te baseren op de landbouwdiertellingen die door het CBS worden geregistreerd. Dit levert de volgende uitdrukking voor de fractie

$f_{d,x}^o(i)^{37}$ :

$$f_{d,x}^o(i) := \frac{[N_{d,x}(i) \cdot P_x]}{\sum_x [N_{d,x}(i) \cdot P_x]} \quad (91)$$

waarbij  $N_{d,x}(i)$  het aantal dieren uit categorie  $x$  is voor gemeente  $i$ .  $P_x$  is de nationaal gemiddelde  $\text{NH}_3$ -emissie per dier uit categorie  $x$ .

- (ii) Bereken de bandbreedte voor de totale dierlijke  $\text{NH}_3$ -emissie in gemeente  $i$  uit de bandbreedtes  $\sigma_{d,x}(i)$  van de afzonderlijke diercategorieën, volgens:

$$\sigma_{d,*}(i) := \sqrt{\sum_x [f_{d,x}^o(i; *) \cdot \sigma_{d,x}(i)]^2} \quad (92)$$

waarbij de afzonderlijke bandbreedtes gelijk zijn aan  $\sigma_{d,r}(i) = .25$ ,  $\sigma_{d,v}(i) = .2$  en  $\sigma_{d,p}(i) = .3$ .

- (iii) Gewapend met deze informatie wordt vervolgens de gemeentelijke dierlijke  $\text{NH}_3$ -emissie rond zijn nominale waarde gevarieerd met bandbreedte  $\sigma_{d,*}(i)$ , volgens<sup>38</sup>:

$$E_{d,*}(i; *) := E_{d,*}^o(i; *) \cdot [1 + g_{d,*}(i) \cdot \sigma_{d,*}(i)] \quad (93)$$

waarbij  $E_{d,*}^o(i; *)$  de totale *nominale* dierlijke  $\text{NH}_3$ -emissie is voor gemeente  $i$ , terwijl  $E_{d,*}(i; *)$  de mogelijk '*werkelijke*' waarde is (realisatie).  $g_{d,*}(i)$  heeft een standaardnormale verdeling  $N(0, 1)$ , en geeft in feite weer wat de daadwerkelijke (relatieve) emissievariatie is t.o.v. de nominale waarde  $E_{d,*}^o(i; *)$ , gemeten in termen van de bandbreedte  $\sigma_{d,*}(i)$ . ■

Toepassing van bovenstaande procedure leidt tot bandbreedtes die per gemeente verschillen omdat het aantal dieren  $N_{d,x}(i)$  per gemeente zal verschillen. De verkregen cijfers suggereren een nauwkeurigheid, die niet gewettigd is wegens de diverse aannames en verwaarlozingen die met de berekeningswijze samenhangen. Daarom is uiteindelijk besloten om de bandbreedtes niet per gemeente te laten variëren, maar enkel per provincie, door uit te gaan van de landbouwdiertellingen per provincie, in plaats van de diertellingen per gemeente. De specifieke bandbreedtes die in deze studie voor het jaar 1992 gehanteerd worden, kunnen dan berekend worden aan de hand van de landbouwdiertellingen uit mei 1992, die vermeld staan in [7]. De gegevens betreffende de aantallen landbouwdieren per provincie staan samengevat in tabel 12. Volgens tabel 18 uit [2] dragen rundvee, varkens en pluimvee in 1992 totaal 84975, 48088 en 12776 Ton  $\text{NH}_3$  bij aan de totale ammoniakemissie uit dierlijke mest op landelijk niveau. Gebruik makend van de getallen uit de laatste rij in tabel 12 volgt dan dat de nationaal gemiddelde  $\text{NH}_3$ -emissie (in kg) per dier gelijk is aan  $P_r = 17.270$ ;  $P_v = 3.396$ ;  $P_p = 0.125$ , voor rundvee, varkens en pluimvee respectievelijk. Hierbij zijn de subcategorieën (vb. melkvee, vleesvee, jongvee) binnen een hoofdcategorie (rundvee) geaggregeerd.

Combinatie van deze informatie met de informatie uit tabel 12 levert dat de relatieve bijdragen  $f_{d,x}^o(\cdot)$  uit vergelijking (91) de waarden (in procenten) hebben die in tabel 13 vermeldt staan. Merk op dat bij deze berekening in wezen impliciet is aangenomen dat in

<sup>37</sup>Het schatten van de  $\text{NH}_3$ -bijdrage van de diverse diersoorten uit de diertellingen per gemeente is strikt genomen niet volledig adequaat. Zo is bijv. de gemiddelde  $\text{NH}_3$ -emissie per dier niet in elke gemeente gelijk, omdat er verschillen zijn in voer, staltype, veebezetting etc. Ook ontstaan er fouten doordat een boer zijn dieren kan beweiden op percelen die in een andere gemeente liggen. Anderzijds treden onvolkomenheden op doordat in werkelijkheid ook de getransporteerde mest een bijdrage tot de  $\text{NH}_3$ -emissie levert. Omdat er onvoldoende informatie is over de samenstelling van de getransporteerde mest op gemeentelijk nivo. en omdat de getransporteerde mest slechts een geringe bijdrage heeft (landelijk gemiddeld ongeveer 10 % van de totale dierlijke mest), wordt ook deze fout buiten beschouwing gelaten.

<sup>38</sup>Deze berekeningswijze komt in feite neer op de veronderstelling dat zo'n 68 % van alle emissievariaties binnen de relatieve bandbreedte ( $E_{d,*}^o(i; *) \cdot [1 \pm \sigma_{d,*}(i)]$ ) vallen, en dat de emissievariaties een normale verdeling zullen hebben.

elke provincie het aandeel van de voornoemde subcategorieën in de betreffende hoofdcategorie dezelfde is. Deze aanname is echter discutabel.

Met de aldus afgeleide cijfers kan tenslotte uit formule (92) bepaald worden wat de bandbreedte  $\sigma_{d,*}(\cdot)$  voor de totale  $\text{NH}_3$ -emissie is. Deze staat in de voorlaatste kolom van tabel 13 vermeld. Deze waarden worden vervolgens aan de diverse gemeenten uit de betreffende provincies toegekend. Bovendien is de betrouwbaarheidsfactor geëvalueerd die bij deze cijfers hoort. Dit is gedaan op basis van de formule  $B \approx \exp(-\sigma_*) = \exp(-\sqrt{\log(1 + [VC]^2)})$  die uit stelling 1 (d) in hoofdstuk 3 volgt. Hierbij is de variatiecoëfficiënt  $[VC]$  gelijk aan bandbreedte  $\sigma_{d,*}(\cdot)$ .

Provincie	Totaal Rundvee	Totaal Varkens	Totaal Pluimvee
Groningen	201348	142405	3208799
Friesland	599218	140889	4870356
Drenthe	278081	274610	5843399
Overijssel	755212	1845964	9537090
Gelderland	1085533	3009681	22865756
Utrecht	228617	451811	1913858
N-Holland	199550	54086	1145958
Z-Holland	251138	243755	1458553
Zeeland	54562	74844	1369557
N-Brabant	961839	6037887	30725830
Limburg	241337	1852074	18410698
Flevoland	63854	32762	852615
Totaal	4920289	14160768	102202469

Tabel 12: Totaal aantal landbouwdieren volgens tellingen in mei 1992 (bron [7]).

Provincie	% bijdrage Rundvee	% bijdrage Varkens	% bijdrage Pluimvee	Bandbreedte	$B(\cdot)$
Groningen	79.72 %	11.09 %	9.19 %	20.24 %	$(\sigma_{d,*}(\cdot) = .20)$ 0.82
Friesland	91.50 %	4.18 %	5.32 %	22.69 %	$(\sigma_{d,*}(\cdot) = .23)$ 0.80
Drenthe	74.28 %	14.42 %	11.30 %	19.10 %	$(\sigma_{d,*}(\cdot) = .19)$ 0.83
Overijssel	63.61 %	30.57 %	5.82 %	17.13 %	$(\sigma_{d,*}(\cdot) = .17)$ 0.84
Gelderland	59.91 %	32.11 %	8.98 %	16.29 %	$(\sigma_{d,*}(\cdot) = .16)$ 0.85
Utrecht	69.00 %	26.82 %	4.18 %	18.11 %	$(\sigma_{d,*}(\cdot) = .18)$ 0.84
N-Holland	91.34 %	4.87 %	3.79 %	22.88 %	$(\sigma_{d,*}(\cdot) = .23)$ 0.80
Z-Holland	81.11 %	15.48 %	3.41 %	20.54 %	$(\sigma_{d,*}(\cdot) = .21)$ 0.82
Zeeland	68.90 %	18.58 %	12.52 %	18.02 %	$(\sigma_{d,*}(\cdot) = .18)$ 0.84
N-Brabant	40.56 %	50.06 %	9.38 %	14.53 %	$(\sigma_{d,*}(\cdot) = .15)$ 0.87
Limburg	33.67 %	49.29 %	18.04 %	13.90 %	$(\sigma_{d,*}(\cdot) = .14)$ 0.87
Flevoland	83.51 %	8.42 %	8.07 %	21.08 %	$(\sigma_{d,*}(\cdot) = .21)$ 0.81

Tabel 13: De relatieve bijdrage van de mestsoorten tot de dierlijke  $\text{NH}_3$ -emissie, en de daaruit resulterende bandbreedte, standaarddeviatie en betrouwbaarheidsfactor  $B(\cdot)$  per provincie voor het jaar 1992

## Appendix D: Analytische onzekerheidsberekeningen voor NH<sub>3</sub>-emissies

Het LAE berekent aan de hand van GIS operaties uit gemeentelijke basisgegevens betreffende NH<sub>3</sub>-emissies, wat er uiteindelijk aan NH<sub>3</sub> geëmitteerd wordt in de specifieke 5 x 5 km gridcellen die Nederland overdekken (zie [2]). Ten gevolge van de onzekerheden in de basisgegevens zijn de berekende emissiecijfers voor de specifieke gridcellen (bijv. 5 x 5 km-grids) ook onzeker. In deze appendix wordt op theoretische wijze nagegaan hoe groot deze onzekerheden zijn. Hierbij wordt gebruik gemaakt van het feit dat de NH<sub>3</sub>-emissie afkomstig is van diverse soorten bronnen (dierlijke mest, kunstmest, huishoudens, industrieën, buitenland), en dat voor elk van deze bronnen de onzekerheidsspecificaties verondersteld worden in principe beschikbaar te zijn.

Allereerst wordt de contributie geëvalueerd van *dierlijke mest*, die landelijk gezien het meeste bijdraagt. Hierbij worden de volgende notaties en veronderstellingen gehanteerd:

- $G_j$  is de indexverzameling van alle gemeenten  $i$  die bijdragen tot de  $j$ -de gridcel.
- $E_{d,*}^o(i; j)$  is de *nominale* dierlijke NH<sub>3</sub>-emissiebijdrage van gemeente  $i$  aan gridcel  $j$ .  $E_{d,*}(i; j)$  is de *'werkelijke'* dierlijke NH<sub>3</sub>-emissiebijdrage van gemeente  $i$  aan gridcel  $j$ .
- $E_{d,*}^o(*; j)$  is de totale *nominale* dierlijke NH<sub>3</sub>-emissie van de  $j$ -de gridcel.  $E_{d,*}(*; j)$  is de totale *'werkelijke'* dierlijke NH<sub>3</sub>-emissie van de  $j$ -de gridcel.
- $f_{d,*}^o(i; j)$  duidt de relatieve (fractionele) bijdrage aan van gemeente  $i$  aan de nominale dierlijke NH<sub>3</sub>-emissie in gridcel  $j$ , en is gelijk aan:

$$f_{d,*}^o(i; j) := \frac{E_{d,*}^o(i; j)}{E_{d,*}^o(*; j)} \quad \text{voor } (i \in G_j) \quad (94)$$

Deze grootheden kunnen in principe bepaald worden uit de GIS-berekeningen die het LAE voor de nominale situatie uitvoert. Dit vereist een nauwkeurige boekhouding.

- Verondersteld wordt dat de *'werkelijke'* bijdrages van de gemeenten  $i$  op een analoge wijze over de gridcellen verdeeld worden als de nominale bijdrages, d.w.z.

$$\frac{E_{d,*}(i; j)}{E_{d,*}(i; *)} = \frac{E_{d,*}^o(i; j)}{E_{d,*}^o(i; *)} \quad (95)$$

Omdat bij de GIS-berekeningen van het LAE deze verdeling verricht wordt op basis van de oppervlaktes van de diverse landgebruikstypen van gemeenten  $i$  in gridcel  $j$ , impliceert deze veronderstelling dat de oppervlakteverdeling voor de nominale situatie dezelfde is als voor de *'werkelijke'* situatie, en dus niet als onzekere factor wordt meebeschoofd!!

■

Op basis van het materiaal uit appendix C kan nu eenvoudig de *'werkelijke'* dierlijke NH<sub>3</sub>-emissie  $E_{d,*}(*; j)$  in de  $j$ -de gridcel worden uitgedrukt als som van de *'werkelijke'* gemeentelijke bijdrages  $E_{d,*}(i; j)$ . Dit resulteert in:

$$\begin{aligned} E_{d,*}(*; j) &:= \sum_{i \in G_j} E_{d,*}(i; j) \stackrel{zie(95)}{=} \sum_{i \in G_j} \frac{E_{d,*}^o(i; j)}{E_{d,*}^o(i; *)} \cdot E_{d,*}(i; *) \\ &\stackrel{zie(77)}{=} \sum_{i \in G_j} E_{d,*}^o(i; j) \cdot [1 + g_{d,*}(i) \cdot \sigma_{d,*}(i)] \\ &\stackrel{zie(94)}{=} \sum_{i \in G_j} E_{d,*}^o(*; j) \cdot f_{d,*}^o(i; j) \cdot [1 + g_{d,*}(i) \cdot \sigma_{d,*}(i)] \end{aligned}$$



$$\begin{aligned}
&= E_{d,*}^{\circ}(*; j) \cdot \left[ \sum_{i \in G_j} f_{d,*}^{\circ}(i; j) \cdot [1 + g_{d,*}(i) \cdot \sigma_{d,*}(i)] \right] \\
&= E_{d,*}^{\circ}(*; j) \cdot \left[ 1 + \sum_{i \in G_j} f_{d,*}^{\circ}(i; j) \cdot \sigma_{d,*}(i) \cdot g_{d,*}(i) \right] \quad (96)
\end{aligned}$$

De laatste overgang is gebaseerd op het feit dat  $\sum_{i \in G_j} f_{d,*}^{\circ}(i; j) \equiv 1$ . De verkregen uitdrukking geeft weer hoe de werkelijke emissie  $E_{d,*}(*; j)$  gerelateerd is aan de nominale  $E_{d,*}^{\circ}(*; j)$  via adequate weging van de emissievariatiaties  $g_{d,*}(i)$  in de afzonderlijke gemeenten.

Merk op dat de verwachtingswaarde (het gemiddelde) van  $E_{d,*}(*; j)$  gelijk is aan de nominale waarde  $E_{d,*}^{\circ}(*; j)$ , hetgeen betekent dat er geen systematische afwijking optreedt (bias is nul). Verder kan uit bovenstaande formule een *analytische uitdrukking voor de variantie* van  $E_{d,*}(*; j)$  afgeleid worden:

$$\text{Var} [E_{d,*}(*; j)] = [E_{d,*}^{\circ}(*; j)]^2 \cdot \sum_{i \in G_j} \sum_{l \in G_j} [f_{d,*}^{\circ}(i; j) \sigma_{d,*}(i)] \cdot [f_{d,*}^{\circ}(l; j) \sigma_{d,*}(l)] \cdot C_{g_d}(i, l) \quad (97)$$

waarbij

$$C_{g_d}(i, l) := \mathbb{E} [g_{d,*}(i) \cdot g_{d,*}(l)] \quad (98)$$

de covariantie is van  $g_{d,*}(i)$  en  $g_{d,*}(l)$  (deze hebben verwachtingswaarde (gemiddelde) 0). Ook kan eenvoudig de *covariantie* van de dierlijke NH<sub>3</sub>-emissie in gridcel  $j$  en gridcel  $m$  worden bepaald:

$$\begin{aligned} \text{Cov} [E_{d,*}(*; j); E_{d,*}(*; m)] &= [E_{d,*}^{\circ}(*; j) \cdot E_{d,*}^{\circ}(*; m)] \times \\ &\quad \sum_{i \in G_j} \sum_{l \in G_m} [f_{d,*}^{\circ}(i; j) \sigma_{d,*}(i)] \cdot [f_{d,*}^{\circ}(l; m) \sigma_{d,*}(l)] \cdot C_{g_d}(i, l) \quad (99) \end{aligned}$$

Deze uitdrukkingen hangen o.a. af van de covariantie structuur  $C_{g_d}(i; l)$  van de stochasten  $g_{d,*}(i)$  die de variaties rond de nominale waarden karakteriseren voor de diverse gemeenten.

**Opmerking 7:** Uitwerking van deze uitdrukkingen voor een tweetal extreme situaties levert de volgende resultaten:

- (A) De emissievariatiaties in alle gemeenten die bijdragen tot de  $j$ -de en  $m$ -de gridcel zijn volledig *afhankelijk*: In dit geval is  $g_{d,*}(i) = g_{d,*}(l) (= g_{d,*}^{(j)}$ , zeg) voor  $i \neq l$ . Dan is  $C_{g_d}(i, l) = 1$  voor alle  $i, l$ , en vereenvoudigt de uitdrukking (96) tot:

$$E_{d,*}(*; j) = E_{d,*}^{\circ}(*; j) \cdot \left[ 1 + \sigma_{d,*}^{(j)} \cdot g_{d,*}^{(j)} \right] \quad (100)$$

waarbij

$$\sigma_{d,*}^{(j)} := \sum_{i \in G_j} f_{d,*}^{\circ}(i; j) \sigma_{d,*}(i) \quad (101)$$

de bandbreedte representeert voor gridcel  $j$ , verkregen door gewogen middeling van de bandbreedtes van de afzonderlijke gemeenten die deel uitmaken van de gridcel.  $\sigma_{d,*}^{(j)}$  kan verder vereenvoudigd worden tot  $\sigma_{d,*}^{\circ}$  indien de bandbreedtes  $\sigma_{d,*}(i)$  voor alle gemeenten in die gridcel hetzelfde zijn, zeg  $\sigma_{d,*}^{\circ}$ . In dit geval is kennis over de specifieke waarden van  $f_{d,*}^{\circ}(i; j)$  *niet vereist* om de bandbreedte voor de gridcel te bepalen!!

De uitdrukkingen (97) en (99) voor de variantie en de covarianties van de dierlijke NH<sub>3</sub>-emissie in de gridcellen  $j$  en  $h$  kunnen vereenvoudigd worden tot:

$$\text{Var} [E_{d,*}(*; j)] = [E_{d,*}^{\circ}(*; j)]^2 \cdot [\sigma_{d,*}^{(j)}]^2 \quad (102)$$

$$\text{Cov} [E_{d,*}(*; j); E_{d,*}(*; m)] = [E_{d,*}^{\circ}(*; j) \cdot E_{d,*}^{\circ}(*; m)] \cdot \sigma_{d,*}^{(j)} \cdot \sigma_{d,*}^{(m)} \quad (103)$$

Via bovenstaande uitdrukkingen kan eenvoudig een expliciete onzekerheidsevaluatie en een gerealiseerd emissieveld verkregen worden uit de *nominale* emissiecijfers per gridcel.

Bovendien kunnen, in analogie met stelling 2 in appendix C, de volgende afschattingen worden afgeleid voor de onzekerheden:

$$(\underline{\sigma}_d[G_j])^2 \leq \frac{\text{Var}[E_{d,*}(*;j)]}{[E_{d,*}^o(*;j)]^2} \leq (\bar{\sigma}_d[G_j])^2 \quad (104)$$

$$(\underline{\sigma}_d[G_j]) \cdot (\underline{\sigma}_d[G_m]) \leq \frac{\text{Cov}[E_{d,*}(*;j); E_{d,*}(*;m)]}{[E_{d,*}^o(*;j) \cdot E_{d,*}^o(*;m)]} \leq (\bar{\sigma}_d[G_j]) \cdot (\bar{\sigma}_d[G_m]) \quad (105)$$

waarbij

$$\underline{\sigma}_d[K] := \min_{i \in K} [\sigma_{d,*}(i)] \quad (106)$$

$$\bar{\sigma}_d[K] := \max_{i \in K} [\sigma_{d,*}(i)] \quad (107)$$

voor een willekeurige indexverzameling  $K$ .

- (B) De emissievariatiaties in alle gemeenten die bijdragen tot de  $j$ -de en  $m$ -de gridcel zijn volledig *onafhankelijk*: In dit geval zijn  $g_{d,*}(i)$  en  $g_{d,*}(l)$  voor  $i \neq l$  onderling onafhankelijk. Dat betekent dat  $C_{g_d}(i, l) = \delta_{il}$ , waarbij  $\delta_{il}$  de Kronecker delta is (d.w.z.  $\delta_{il} = 1$  als  $i = l$ , en 0 als  $i \neq l$ ). De uitdrukkingen (97) en (99) voor de variantie en de covarianties van de dierlijke  $\text{NH}_3$ -emissie in de gridcellen  $j$  en  $m$  kunnen vereenvoudigd worden tot:

$$\text{Var}[E_{d,*}(*;j)] = [E_{d,*}^o(*;j)]^2 \cdot \sum_{i \in G_j} [f_{d,*}^o(i; j) \sigma_{d,*}(i)]^2 \quad (108)$$

$$\begin{aligned} \text{Cov}[E_{d,*}(*;j); E_{d,*}(*;m)] &= [E_{d,*}^o(*;j) \cdot E_{d,*}^o(*;m)] \times \\ &\sum_{i \in G_j \cap G_m} f_{d,*}^o(i; j) \cdot f_{d,*}^o(i; m) \cdot [\sigma_{d,*}(i)]^2 \end{aligned} \quad (109)$$

In analogie met stelling 2 in appendix C kunnen de volgende afschattingen worden afgeleid:

$$\left( \sum_{i \in G_j} [f_{d,*}^o(i; j)]^2 \right) \cdot (\underline{\sigma}_d[G_j])^2 \leq \frac{\text{Var}[E_{d,*}(*;j)]}{[E_{d,*}^o(*;j)]^2} \leq \left( \sum_{i \in G_j} [f_{d,*}^o(i; j)]^2 \right) \cdot (\bar{\sigma}_d[G_j])^2 \quad (110)$$

$$\left( \sum_{i \in G_j} \sigma_{d,*}^{-2}(i) \right)^{-\frac{1}{2}} \leq \frac{\text{Var}[E_{d,*}(*;j)]}{[E_{d,*}^o(*;j)]^2} \leq (\bar{\sigma}_d[G_j])^2 \quad (111)$$

$$\frac{\text{Cov}[E_{d,*}(*;j); E_{d,*}(*;m)]}{[E_{d,*}^o(*;j) \cdot E_{d,*}^o(*;m)]} \geq \left( \sum_{i \in G_j \cap G_m} f_{d,*}^o(i; j) \cdot f_{d,*}^o(i; m) \right) \cdot (\underline{\sigma}_d[G_j \cap G_m])^2 \quad (112)$$

$$\frac{\text{Cov}[E_{d,*}(*;j); E_{d,*}(*;m)]}{[E_{d,*}^o(*;j) \cdot E_{d,*}^o(*;m)]} \leq \left( \sum_{i \in G_j \cap G_m} f_{d,*}^o(i; j) \cdot f_{d,*}^o(i; m) \right) \cdot (\bar{\sigma}_d[G_j \cap G_m])^2 \quad (113)$$

Indien er algemene onder- en bovengrenzen beschikbaar zijn voor de gemeentelijke bandbreedtes  $\sigma_{d,*}(i)$  dan kunnen voor bovenstaande situaties (A) en (B) uit de nominale emissiecijfers, ruwe onder- en overschattingen bepaald worden voor de onzekerheid in de dierlijke  $\text{NH}_3$ -emissiecijfers per gridcel.

Merk op dat variantie en covarianties en hun geschatte bovengrenzen in situatie (A) groter zullen zijn dan in situatie (B). Bij situatie (A) zijn de variaties immers volledig gecorreleerd, en zullen elkaar dus versterken, terwijl ze in situatie (B), wegens hun ongecorrleerdheid, een uitmiddeling tot stand brengen. Situatie (A) geeft dus een 'conservatief' en pessimistisch beeld van de onzekerheid (te grote bandbreedte). Situatie (B) leidt daarentegen vermoedelijk tot een te optimistische inschatting (te smalle bandbreedte). De 'werkelijkheid' zal waarschijnlijk ergens tussen deze twee 'extremen'

in liggen, omdat de emissievariatiën van de afzonderlijke gemeenten correlaties zullen vertonen, die elkaars effecten zowel kunnen versterken als opheffen. Foutenbronnen zijn hierbij bijv. andere emissiepercentages en excretiefactoren dan verondersteld, verkeerde aannames over waar welke mest wordt uitgereden etc. Het zal in de praktijk echter nagenoeg onmogelijk zijn om hiervan, en van hun onderlinge correlaties<sup>39</sup> adequate specificaties te geven. Om te voorkomen dat de onzekerheid te laag wordt ingeschat, zal voor de uit te voeren betrouwbaarheidsanalyse veeléer naar de worst-case situatie (A) gekeken worden, dan naar situatie (B). *Binnen de gehanteerde aannames* zullen de resultaten dientengevolge vermoedelijk een *te pessimistisch beeld* schetsen van de onzekerheid. Vergelijking van formule (102) met (108), of (104) met (110), geeft een indruk hoe veel deze worst-case situatie (A) kan afwijken van de situatie (B) met onafhankelijke gemeenten. Indien bijv. alle  $\sigma_{d,*}(i)$ 's dezelfde zouden zijn (zeg  $\sigma_{d,*}^o$ ), dan blijkt de variantie/onzekerheid in situatie (B)  $\sum_{i \in G_j} [f_{d,*}^o(i; j)]^2$  keer zo klein te zijn als de variantie in de worst-case situatie (A). Uit de nominale NH<sub>3</sub>-emissieberekeningen die het LAE uitvoert voor het jaar 1992 blijkt dat deze verkleiningsfactor  $\sum_{i \in G_j} [f_{d,*}^o(i; j)]^2$  voor de diverse gridcellen ( $j$ ) waarden kan aannemen die variëren tussen .19 en 1. De verdeling van deze waarden is zichtbaar in het histogram uit figuur 15.

■

Bovenstaande berekeningen werden uitgevoerd voor de dierlijke NH<sub>3</sub>-emissie. Analoge beschouwingen kunnen worden opgezet voor de bijdrage van *kunstmest en huishoudens*<sup>40</sup>. Dit resulteert in verwante uitdrukkingen voor hun emissiebijdrages (vergelijk (96)):

$$E_k(*; j) := E_k^o(*; j) \cdot \left[ 1 + \sum_{i \in G_j} f_k^o(i; j) \cdot \sigma_k(i) \cdot g_k(i) \right] \quad (114)$$

$$E_h(*; j) := E_h^o(*; j) \cdot \left[ 1 + \sum_{i \in G_j} f_h^o(i; j) \cdot \sigma_h(i) \cdot g_h(i) \right] \quad (115)$$

Hierbij refereren de suffixen  $k$  en  $h$  naar 'kunstmest' en 'huishoudens' respectievelijk. Omdat de invloed van huishoudens bepaald wordt aan de hand van een nominale NH<sub>3</sub>-emissiefactor per jaar, per hoofd van de bevolking, die voor *heel Nederland* gehanteerd wordt, zullen  $\sigma_h(i)$  en  $g_h(i)$  voor alle gemeenten dezelfde zijn, bij analoge variaties in de bevolkingsaantallen, d.w.z.

$$\sigma_h(i) =: \sigma_h^o; \quad g_h(i) =: g_h^o \quad \text{voor alle } i; \quad (116)$$

Dan kan bovenstaande uitdrukking voor  $E_h$  vereenvoudigd worden tot:

$$E_h(*; j) = E_h^o(*; j) \cdot \left[ 1 + \sigma_h^{(j)} \cdot g_h^o \right] \quad (117)$$

waarbij

$$\sigma_h^{(j)} := \sigma_h^o \quad (118)$$

Indien de emissievariatiën in de kunstmestbijdrage *voor alle gemeenten in gridcel  $j$  hetzelfde* zijn<sup>41</sup>, d.w.z.  $g_k(i) = g_k(l) (=: g_k^o, \text{ zeg})$ , dan vereenvoudigt de uitdrukking voor  $E_k$  tot:

$$E_k(*; j) = E_k^o(*; j) \cdot \left[ 1 + \sigma_k^{(j)} \cdot g_k^o \right] \quad (119)$$

<sup>39</sup>Veelal valt te verwachten dat de onderlinge correlaties van deze foutenbronnen gering zullen zijn. Indien daarentegen de fouten ook een systematisch karakter zullen hebben, dan zal de onderlinge samenhang groot zijn.

<sup>40</sup>De emissiebijdrage van industrieën en buitenland blijven buiten beschouwing.

<sup>41</sup>Deze keuze komt overeen met bovengeschetste 'worst-case' situatie (A).

waarbij

$$\sigma_k^{(j)} := \sum_{i \in G_j} f_k^o(i; j) \sigma_k(i) \quad (120)$$

de bandbreedte representeert van de kunstmestemissie voor gridcel  $j$ , verkregen door gewogen middeling van de bandbreedtes  $\sigma_k(i)$  van de afzonderlijke gemeenten  $i$  die deel uitmaken van de  $j$ -de gridcel. Omdat de kunstmestgift in de praktijk deels gerelateerd zal zijn met de dierlijke mestgift, zal in principe ook de bandbreedte  $\sigma_k(i)$  voor kunstmest, net als de bandbreedte  $\sigma_{d,*}(i)$  voor dierlijke mest, per gemeente kunnen variëren.

Indien tenslotte de hiervoor genoemde emissiebijdrages samengevoegd worden tot:

$$E_t(*; j) = E_{d,*}(*; j) + E_k(*; j) + E_h(*; j) \quad (121)$$

dan is het duidelijk dat het gemiddelde ( $\mathbb{E}$ ) en de (co)variantie van deze ‘totale’ emissie gelijk zijn aan:

$$\mathbb{E}[E_t(*; j)] = E_{d,*}^o(*; j) + E_k^o(*; j) + E_h^o(*; j) \quad (122)$$

$$\begin{aligned} \text{Var}[E_t(*; j)] &= \text{Var}[E_{d,*}(*; j)] + \text{Var}[E_k(*; j)] + \text{Var}[E_h(*; j)] + \\ &\quad \sum_{\substack{v \neq w \\ v, w \in \{d, k, h\}}} \text{Cov}[E_v(*; j); E_w(*; j)] \end{aligned} \quad (123)$$

$$\begin{aligned} \text{Cov}[E_t(*; j); E_t(*; m)] &= \text{Cov}[E_{d,*}(*; j); E_{d,*}(*; m)] + \text{Cov}[E_k(*; j); E_k(*; m)] + \\ &\quad \text{Cov}[E_h(*; j); E_h(*; m)] + \\ &\quad \sum_{\substack{v \neq w \\ v, w \in \{d, k, h\}}} \text{Cov}[E_v(*; j); E_w(*; m)] \end{aligned} \quad (124)$$

Evaluatie hiervan vereist dat naast de bepaling van de (co)varianties van de afzonderlijke emissiebronnen (dierlijke mest; kunstmest; huishoudens), ook nog de covarianties

$\text{Cov}[E_v(*; j); E_w(*; m)]$ ,  $v \neq w$ , tussen deze bronnen onderling bepaald dienen te worden. Om hiervan een goed beeld te krijgen dienen dus de *onderlinge correlaties* tussen de variaties  $g_{d,*}(i)$ ;  $g_k(i)$ ;  $g_h(i)$  gespecificeerd te worden. Met name de correlaties tussen  $g_{d,*}$  en  $g_k$  (i.e. dierlijke mest en kunstmest) kunnen van belang zijn. Specificatie daarvan is echter moeilijk, en vergt veel werk. Om die reden worden de correlaties vaak niet meegenomen. Met het oog op de beperkte tijd en menskracht die voor de huidige betrouwbaarheidsberekeningen beschikbaar was, zijn niet alleen de correlaties, maar ook de *onzekerheden zelf in de bijdragen van kunstmest en huishoudens helemaal buiten beschouwing gelaten*. Dit wordt mede gemotiveerd vanuit het feit dat deze emissiebronnen, in verhouding tot dierlijke mest, slechts weinig bijdragen tot de totale  $\text{NH}_3$ -emissie. Hoewel dit op landelijk nivo inderdaad het geval is, hoeft dit echter op lokale schaal niet te gelden.

**Opmerking 8:** Een alternatieve aanpak (worst-case) bestaat daarentegen uit het meenemen van *alle* onzekerheden, waarbij de onderlinge correlaties maximaal verondersteld worden, d.w.z.  $g_{d,*}(i) = g_k(i) = g_h(i) (= g^o \text{ zeg})$ . In dat geval gaat (121) over in

$$E_t(*; j) = E_t^o(*; j) \cdot [1 + g^o \cdot \sigma_t^{(j)}] \quad (125)$$

waarbij

$$E_t^o(*; j) := E_{d,*}^o(*; j) + E_k^o(*; j) + E_h^o(*; j) \quad (126)$$

$$\sigma_t^{(j)} := \left[ \frac{E_{d,*}^o(*; j)}{E_t^o(*; j)} \cdot \sigma_{d,*}^{(j)} + \frac{E_k^o(*; j)}{E_t^o(*; j)} \cdot \sigma_k^{(j)} + \frac{E_h^o(*; j)}{E_t^o(*; j)} \cdot \sigma_h^{(j)} \right] \quad (127)$$

Hierbij is de uitdrukking  $\sigma_t^{(j)}$  (i.e. de standaarddeviatie in de  $j$ -de gridcel) een gewogen som van de afzonderlijke standaarddeviaties. De weging is gebaseerd op de relatieve bijdrage van de diverse emissiebronnen aan de totale nominale emissie.

De (co)varianties voor deze situatie zijn gelijk aan:

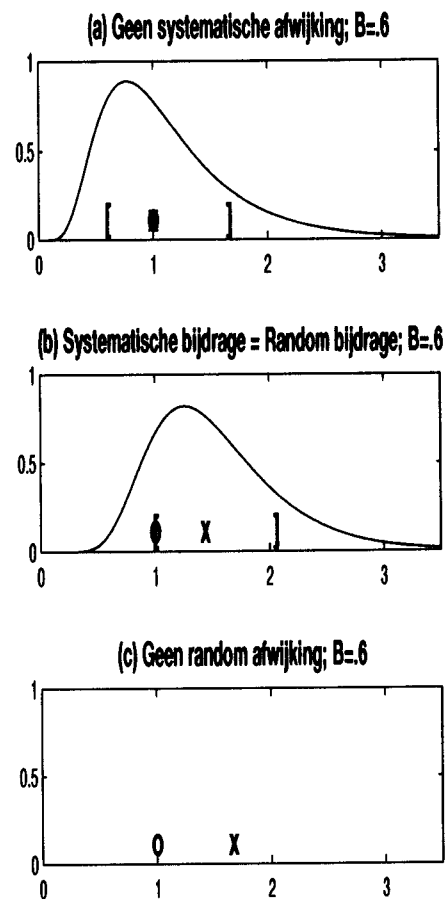
$$\text{Var} [E_t(*; j)] = [E_t^o(*; j)]^2 \cdot [\sigma_t^{(j)}]^2 \quad (128)$$

$$\text{Cov} [E_t(*; j); E_t(*; m)] = E_t^o(*; j) \cdot E_t^o(*; m) \cdot \sigma_t^{(j)} \cdot \sigma_t^{(m)} \quad (129)$$

■

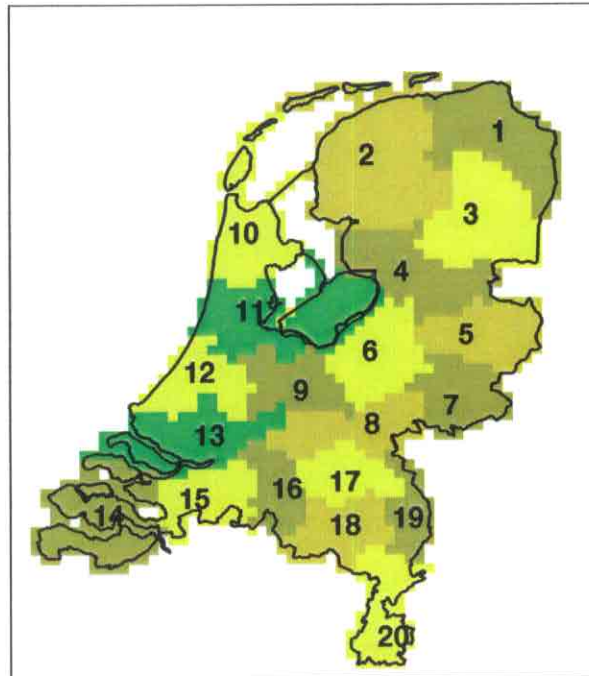
## Referenties

- [1] Hoekstra, J.A. en Heuberger, P.S.C. (eds.) (1995), *Betrouwbaarheid van milieu-informatie: analyse van een stikstofketen*. RIVM rapport nr. 714701011.
- [2] Hoek, K.W. van der (1994), *Berekeningsmethodiek ammoniakemissie in Nederland voor de jaren 1990, 1991 en 1992*. RIVM rapport nr. 773004003.
- [3] Heij, G.J., Erisman, J.W., Vries, W. de, (1992), De effecten van atmosferische depositie op het Nederlandse bos. *Milieu*, no. 4, pp. 101-106.
- [4] Thunissen, H., Olthof, R., Getz, P., Vels L. (1992), *Grondgebruiksdatabank van Nederland vervaardigd met behulp van Landsat Thematic Mapper opnamen*. Staring Centrum, SC-DLO, rapport nummer 168, Wageningen.
- [5] Beurden, A.U.C.J. van, Kusse, A.A.M., Janssen, P.H.M., (1995), Reliability assessment in GIS with an example of ammonia emissions. Proceedings of the *First Joint European Conference and Exhibition on Geographical Information*. The Hague, The Netherlands, 1995, pp. 285-291.
- [6] Kusse, A.A.M., Beurden, A.U.C.J. van, Janssen, P.H.M., (1995), Where policy and quality meet: a reliability assessment in GIS of ammonia emissions. Artikel in voorbereiding.
- [7] *CBS-Maandstatistiek van de Landbouw*. Februari 1993; jaargang 41.
- [8] Kirkwood, B.R. (1979), Geometric means and measures of dispersion. *Biometrics*, Vol. 35, pp. 905-909.
- [9] Kirkwood, B.R. (1988), *Essentials of medical statistics*. Blackwell Scientific Publications, Oxford.
- [10] Lee, D.S., Dollard, G.J. (1994), Uncertainties in current estimates of emissions of ammonia in the United Kingdom. *Environmental Pollution*, Vol. 86, pp. 267-277.
- [11] Leggett, R., Williams, L.R. (1981), A reliability index for models. *Ecological Modelling*. Vol. 13, pp. 303-312.
- [12] Press, W.H., Flannery B.P., Teukolsky, S.A., Vetterling W.T., (1988) *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge.
- [13] Hootsmans, R. en Wel, F. van der, (1993) Detection and visualization of ambiguity and fuzziness in composite spatial datasets. In *Proceedings Fourth European Conference on GIS*, 29 maart- 1 april, Genua, Italië. pp. 1035-1046.
- [14] Janssen, P.H.M., Heuberger, P.S.C. en Sanders R. (1992), *UNCSAM 1.1: a Software Package for Sensitivity and Uncertainty Analysis: Manual*. RIVM rapport nr. 959101004.
- [15] Wallach, D., Goffinet, B. (1989), Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modelling*. Vol. 44, pp. 299-306.

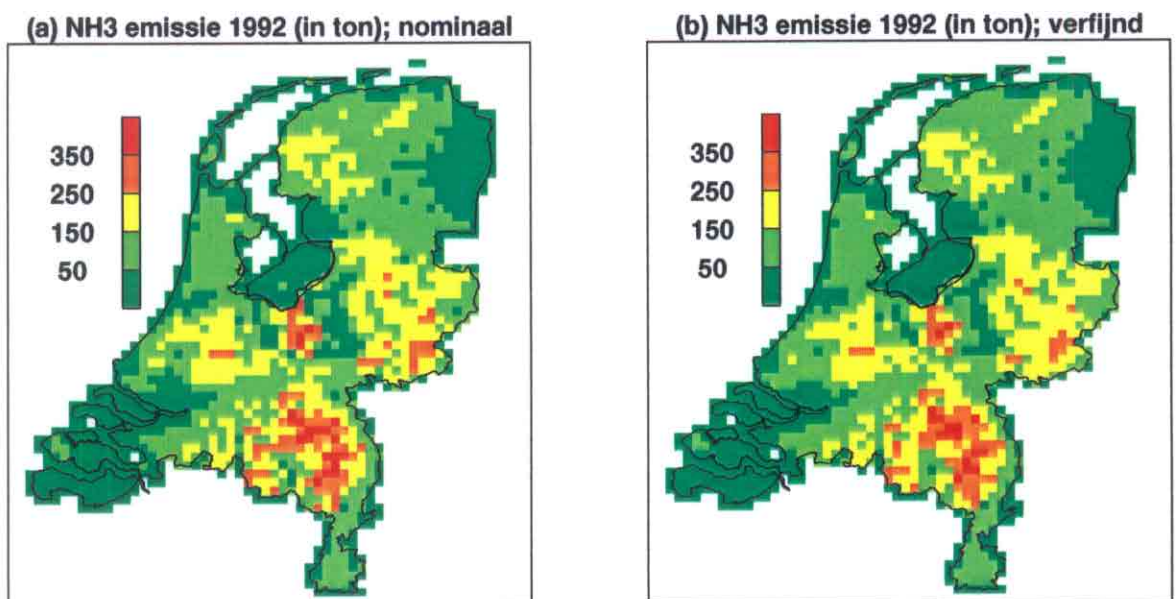


Het 68 % betrouwbaarheidsinterval wordt aangeduid door het interval  $[\cdot \cdot \cdot]$ .  
De nominale waarde wordt door (o) aangeduid en het geometrisch gemiddelde door (x).

Figuur 1: Drie situaties met dezelfde betrouwbaarheidsfactor (.6), maar met een verschillende rol voor de systematische en random afwijking

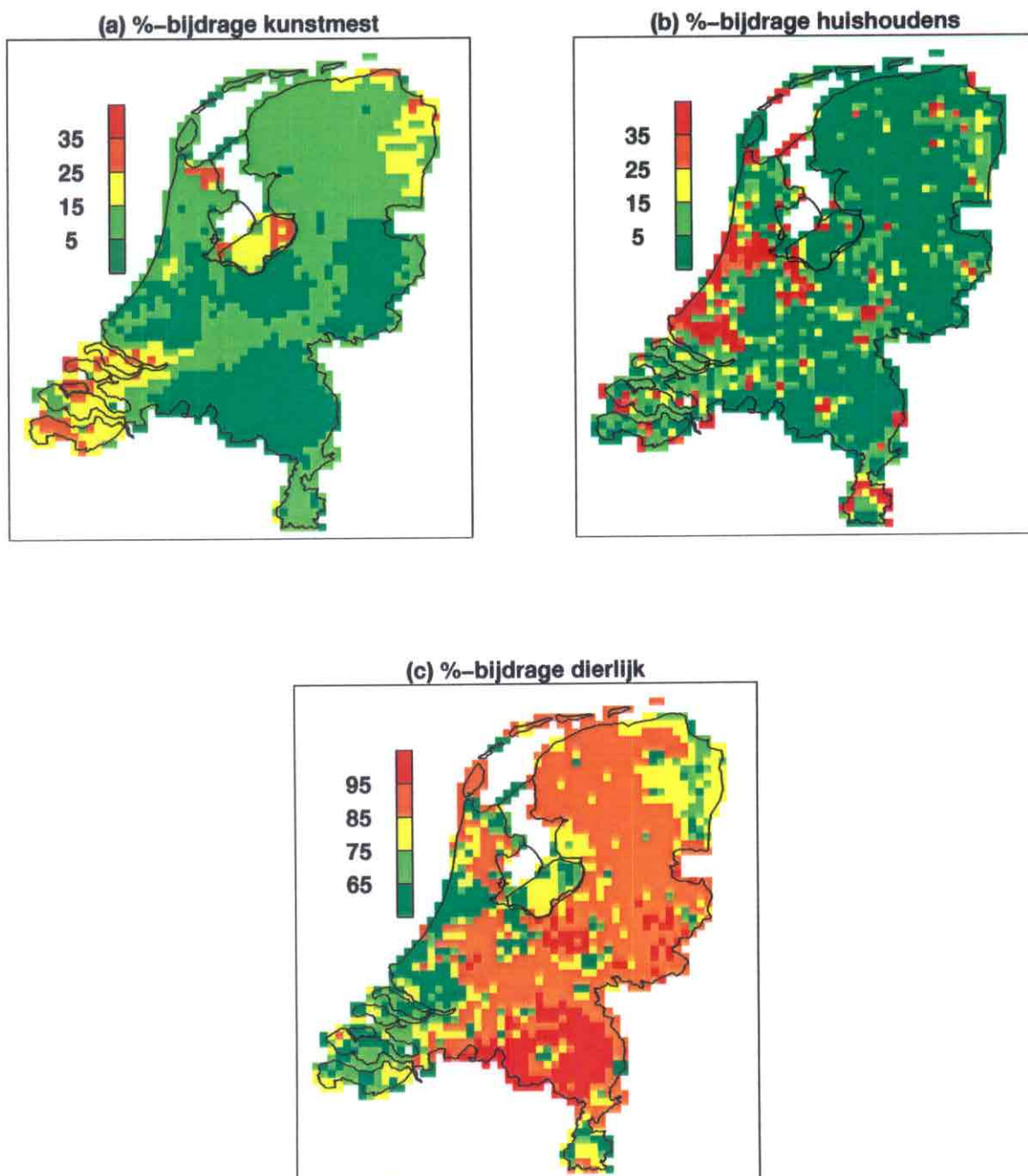


Figuur 2: Verzuringgebieden in Nederland

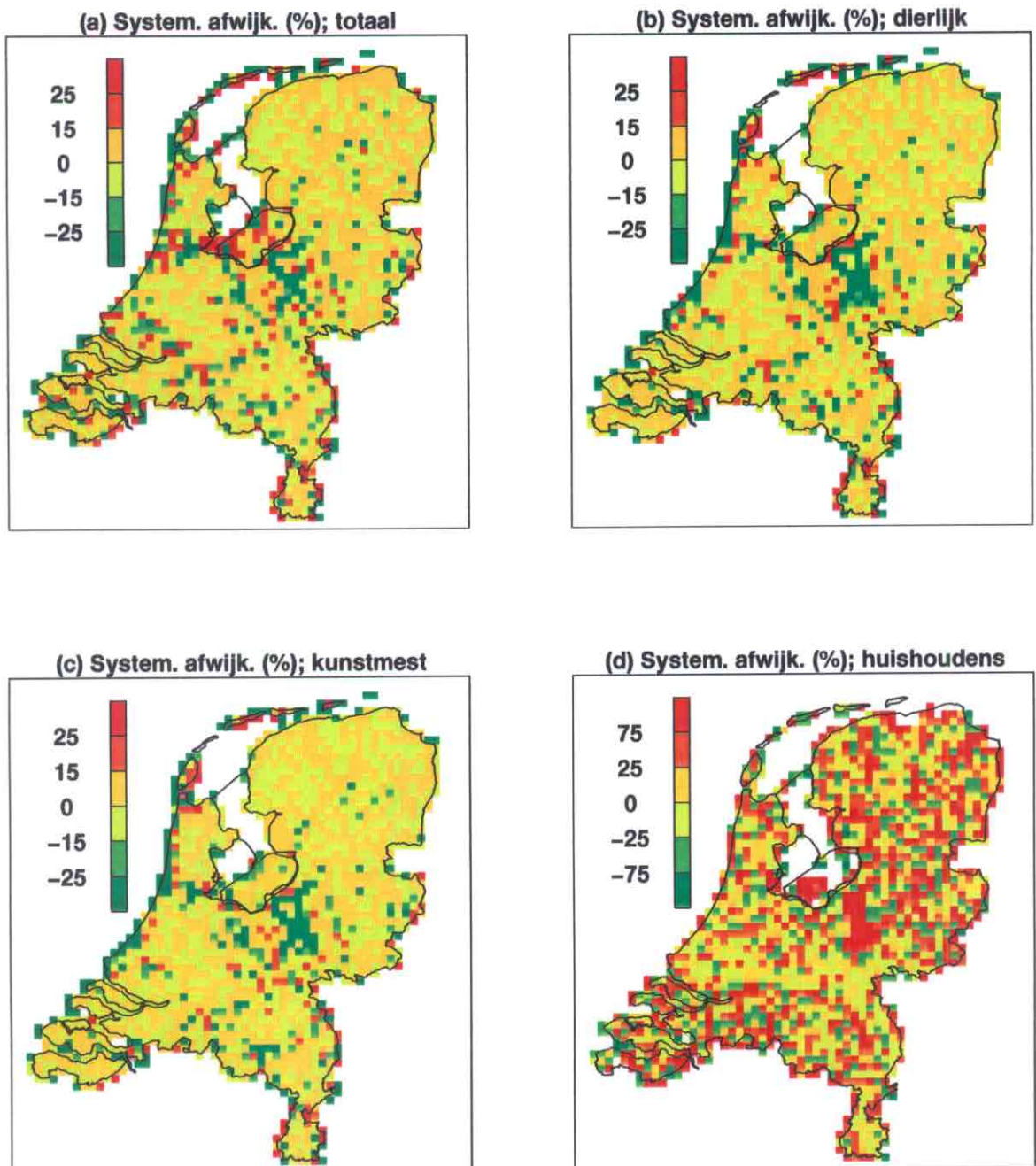


Figuur 3: NH<sub>3</sub>-emissie (in Ton) in gridcellen, volgens nominale berekening (a), en volgens gemiddelde verfijnde berekening (b)

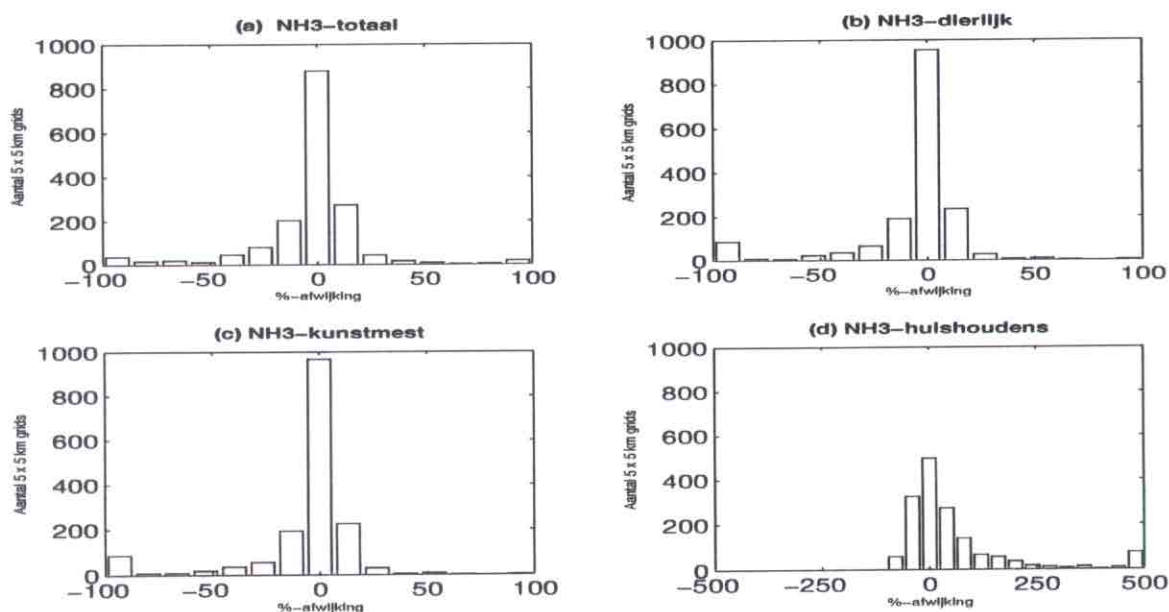




Figuur 4: Procentuele bijdrage van dierlijke mest (a), kunstmest (b), en huishoudens (c) aan de  $\text{NH}_3$ -emissie in gridcellen, volgens gemiddelde verfijnde berekening



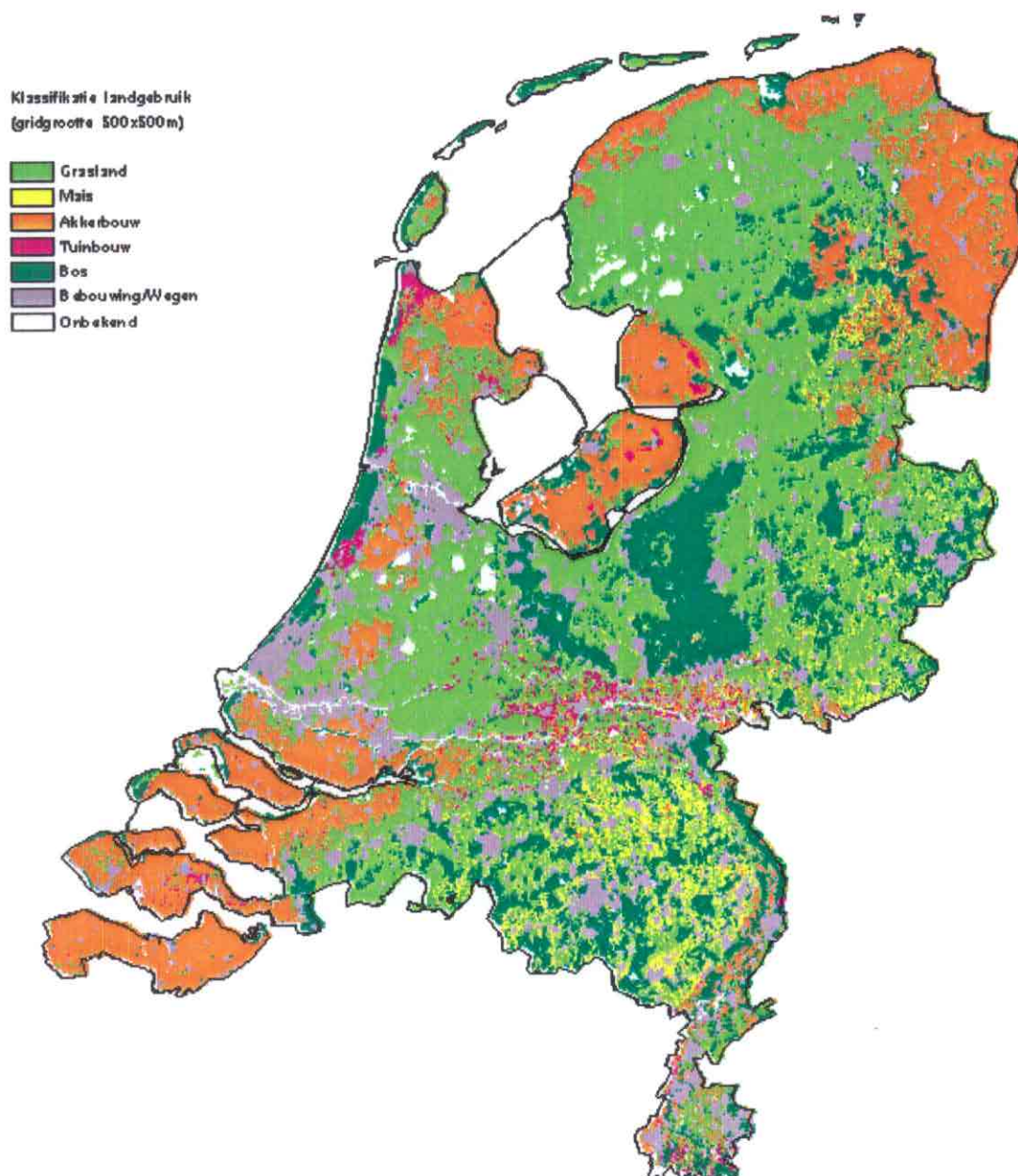
Figuur 5: Systematische afwijking tussen nominale berekening en gemiddelde verfijnde berekening voor  $\text{NH}_3$ -emissie in gridcellen, uitgedrukt als percentage van de gemiddelde verfijnde berekening. Dit is weergegeven voor de totale  $\text{NH}_3$ -emissie (a), en voor de emissiebijdrages van dierlijke mest (b), kunstmest (c), en huishoudens (d)



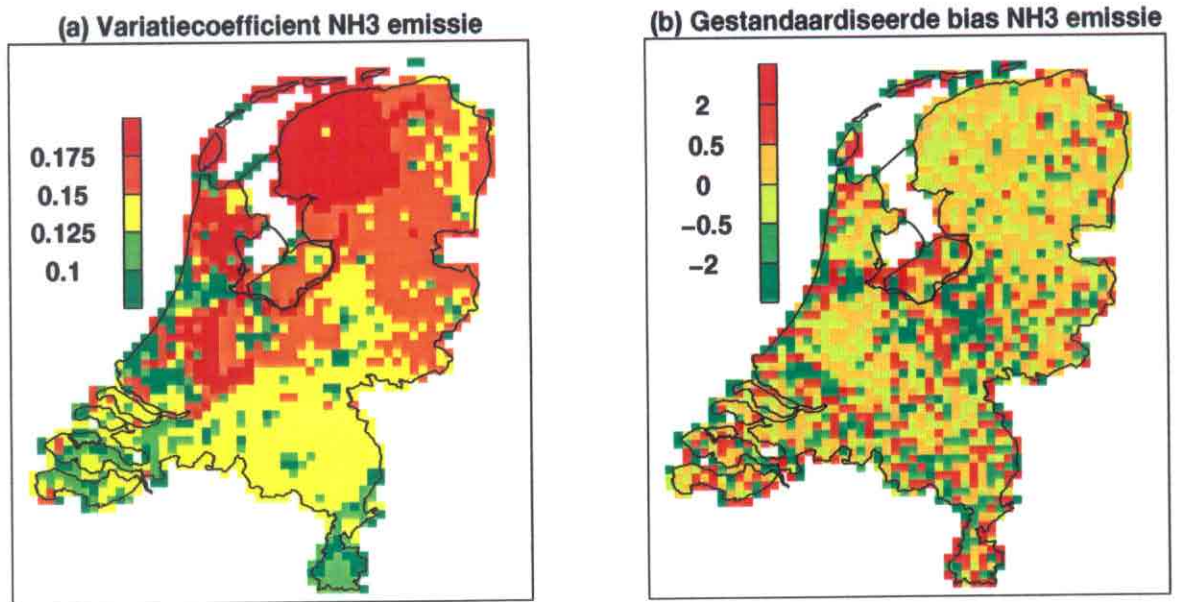
Figuur 6: Histogram van systematische afwijking (in %) tussen nominale berekening en gemiddelde verfijnde berekening voor NH<sub>3</sub>-emissie in gridcellen. Dit is weergegeven voor de totale NH<sub>3</sub>-emissie (a), en voor de emissiebijdrages van dierlijke mest (b), kunstmest (c), en huishoudens (d). Totaal aantal grids is 1657.



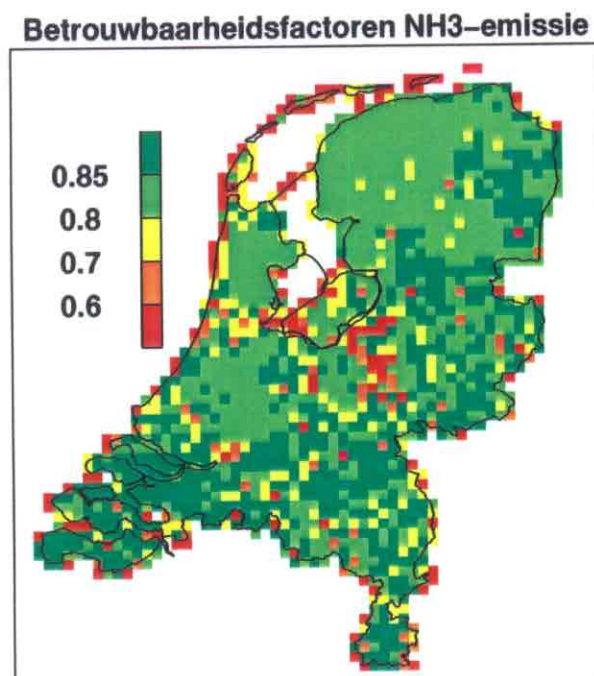
Figuur 7: Landgebruikstypering van Nederland volgens het LGN-bestand  $RS^0$  (25 x 25 m)



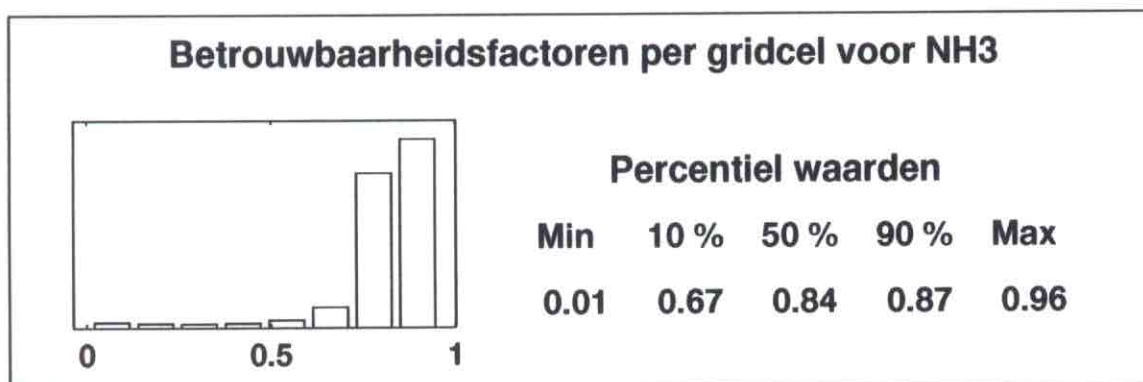
Figuur 8: Landgebruikstypering van Nederland volgens de door het LBG samengestelde landgebruikkaart (500 x 500 m)



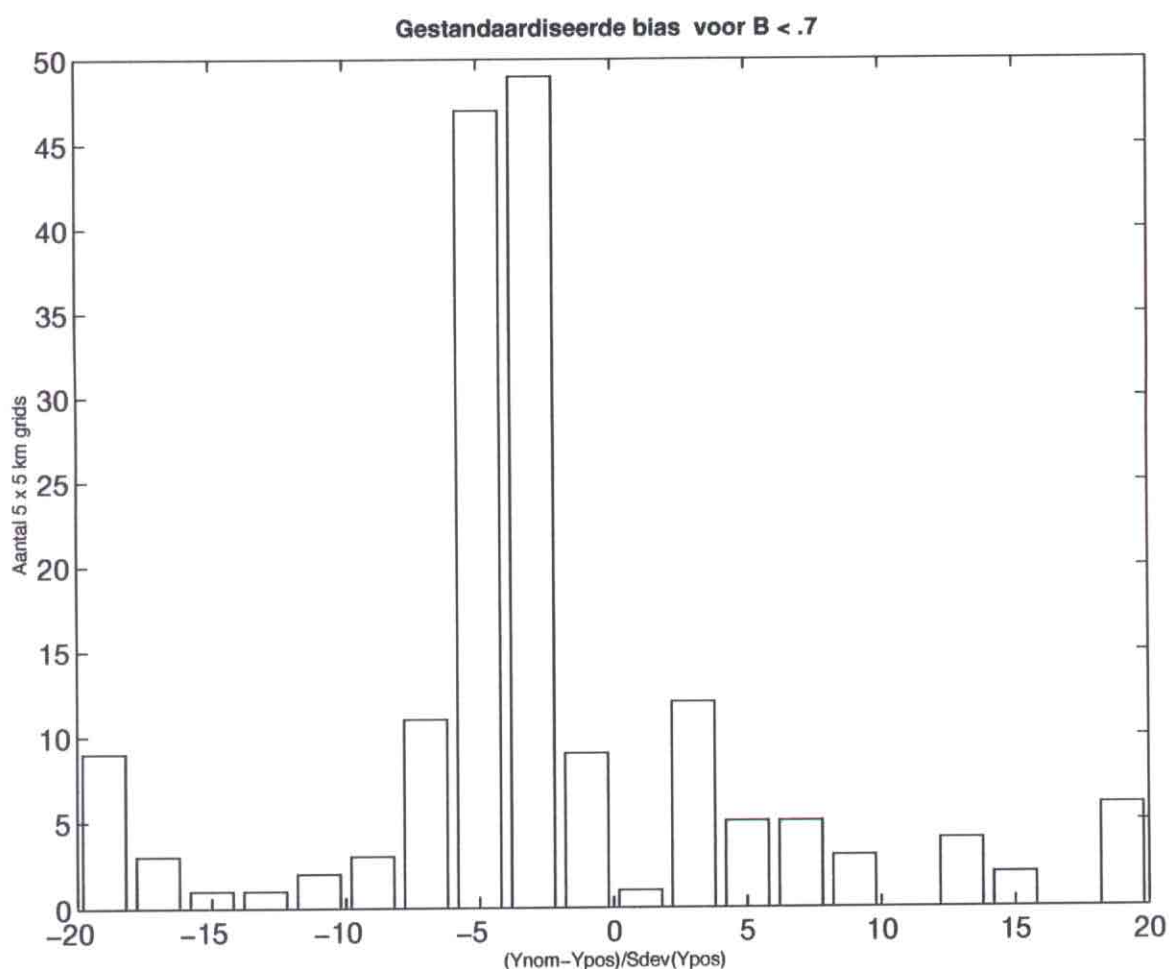
Figuur 9: Variatiecoëfficiënten (a) en gestandaardiseerde bias (b) voor NH<sub>3</sub>-emissie in gridcellen



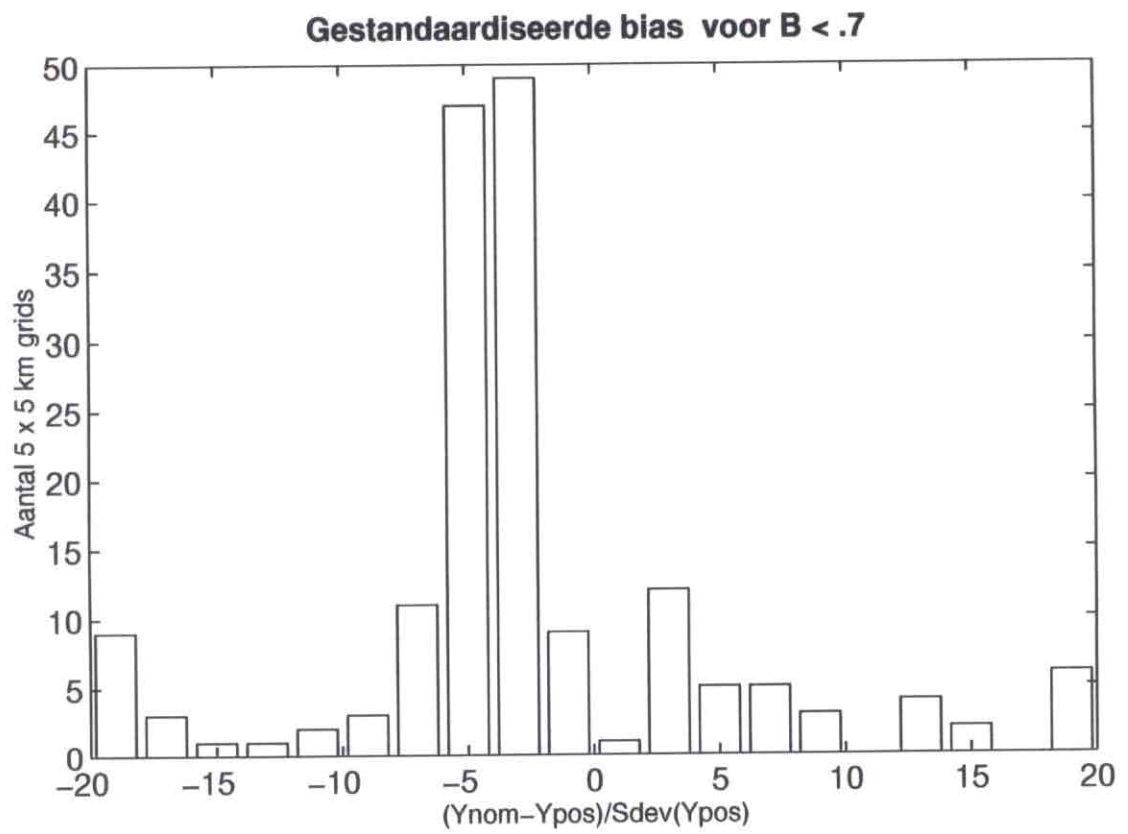
Figuur 10: Betrouwbaarheidsfactoren voor NH<sub>3</sub>-emissie in gridcellen



Figuur 11: Histogram van betrouwbaarheidsfactoren, en bijbehorende percentielwaarden, voor NH<sub>3</sub>-emissie in gridcellen over Nederland totaal

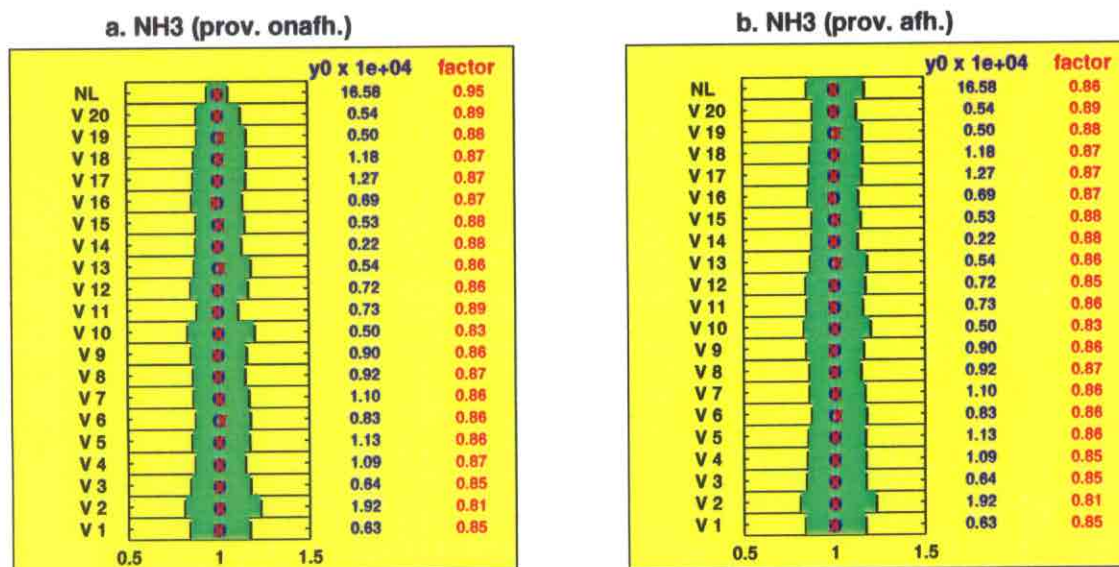


Figuur 12: Histogram van betrouwbaarheidsfactoren voor NH<sub>3</sub>-emissie in gridcellen voor diverse verzuringsgebieden



Figuur 13: Histogram van de gestandaardiseerde bias voor alle gridcellen met een betrouwbaarheidsfactor < .7

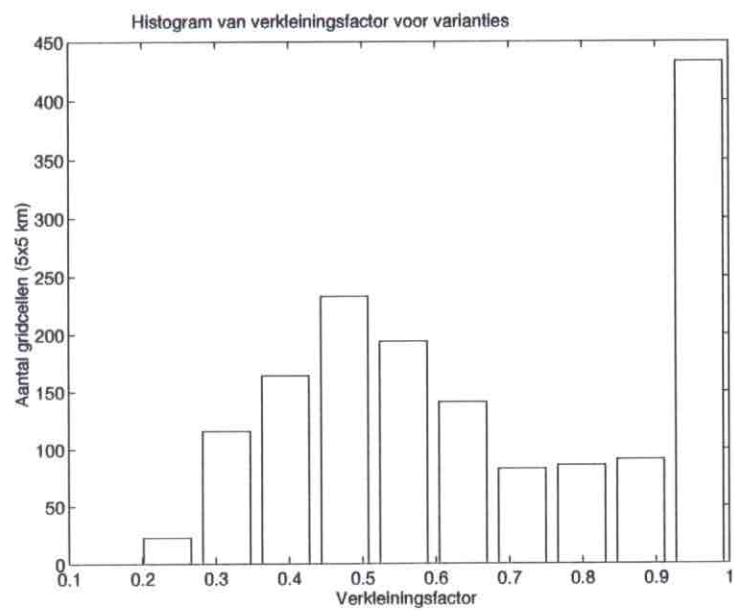




Geschaald t.o.v.  $y^o$ . 68 % betrouwbaarheidsinterval in groen  
 Nominale waarden  $y^o$  (o) in blauw en geometrisch gemiddelden ( $x$ ) in rood.  
 In cijfers: nominale waarden en betrouwbaarheidsfactoren

Figuur 14: Betrouwbaarheden voor de NH<sub>3</sub>-emissie per verzorgingsgebied en Nederland totaal.

Resultaten zijn weergegeven voor de situatie waarbij de gemeentelijke emissievariatiaties in verschillende provincies onafhankelijk (zie (a)) c.q. volledig afhankelijk (zie (b)) zijn .



Figuur 15: Histogram van verkleiningsfactor  $\sum_{i \in G} [f_{d,*}^o(i; j)]^2$  voor de varianties van dierlijke NH<sub>3</sub>-emissies in de 5 x 5 km gridcellen over Nederland