

# **MIRA O&O Project** **Statistische analyse van de meetresultaten** **van het fysisch-chemisch waterkwaliteitsmeet-** **net**

## **Eindverslag**

Prof. Dr. ir. Olivier Thas  
BioStat Onderzoeksgroep  
Vakgroep Toegepaste Wiskunde, Biometrie en Procesregeling,  
Universiteit Gent

**Studie uitgevoerd in opdracht van de Vlaamse**  
**Millieumaatschappij, MIRA**

MIRA/2007/02

juni 2007



Dit rapport verschijnt in de reeks MIRA Ondersteunend Onderzoek van de Vlaamse Milieumaatschappij. Deze reeks bevat resultaten van onderzoek gericht op de wetenschappelijke onderbouwing van het Milieurapport Vlaanderen.

Dit rapport is ook beschikbaar via [www.milieurapport.be](http://www.milieurapport.be)

Contactadres:  
Vlaamse Milieumaatschappij  
Milieurapportering (MIRA)  
Van Benedenlaan 34  
2800 Mechelen  
tel. 015 45 14 66  
[mira@vmm.be](mailto:mira@vmm.be)

Wijze van citeren:

Thas, O. (2007), Statistische analyse van de meetresultaten van het fysisch-chemisch waterkwaliteitsmeetnet, studie uitgevoerd in opdracht van de Vlaamse Milieumaatschappij, MIRA, MIRA/2007/02, Universiteit Gent.

# Inhoudsopgave

<b>1</b>	<b>Inleiding</b>	<b>5</b>
1.1	Doelstelling . . . . .	5
1.2	De Data . . . . .	6
1.3	Structuur van het Eindverslag . . . . .	7
1.4	Software . . . . .	8
1.5	Terminologie . . . . .	8
<b>2</b>	<b>Methodologie</b>	<b>10</b>
2.1	Inleiding . . . . .	10
2.2	Statistische Methoden voor Trendanalyse op Meetplaatsniveau . .	13
2.2.1	Notatie . . . . .	13
2.2.2	Het Schatten van de Afhankelijkheidsstructuur . . . . .	14
2.2.3	Test 1 . . . . .	15
2.2.4	Test 2 . . . . .	16
2.3	Statistische Methoden voor Trendanalyse op Aggregatieniveau . .	17
2.3.1	Inleiding . . . . .	17
2.3.2	Testen B1 en B2 . . . . .	18

2.3.3	Test B3 . . . . .	19
2.4	Regelmatige Steekproefname, Ontbrekende Waarnemingen en de Detectielimiet . . . . .	20
2.4.1	Regelmatige steekproefname en ontbrekende waarnemingen	20
2.4.2	De detectielimiet . . . . .	21
2.5	Multipliciteit . . . . .	22
2.6	Simulatiestudie . . . . .	24
2.6.1	Test T1 . . . . .	24
2.6.2	Test B3 . . . . .	25
<b>3</b>	<b>Resultaten voor de afzonderlijke meetplaatsen</b>	<b>26</b>
3.1	Inleiding . . . . .	26
3.2	Resultaten van de trendanalyse met 2005 als referentiejaar . . . .	27
3.2.1	Gedetailleerde resultaten . . . . .	27
3.2.2	Bondig overzicht van de resultaten per bekken . . . . .	31
<b>4</b>	<b>Resultaten op aggregatieniveau</b>	<b>34</b>
4.1	Inleiding . . . . .	34
4.2	Resultaten op de aggregatieniveaus van de trendanalyse met Test B2 en 2005 als referentiejaar . . . . .	35
<b>5</b>	<b>Resultaten voor heel Vlaanderen</b>	<b>39</b>
<b>6</b>	<b>Conclusies</b>	<b>56</b>

# Hoofdstuk 1

## Inleiding

### 1.1 Doelstelling

Volgens artikel 2.1.3 “Decreet houdende algemene bepalingen inzake milieubeleid” van 5 april 1995, BS 3 juni 1995, omhelst de opdracht voor de milieurapportering (MIRA) in Vlaanderen 3 elementen:

- een beschrijving, analyse en evaluatie van de bestaande toestand van het milieu;
- een beschrijving, analyse en evaluatie van het tot dan toe gevoerde milieubeleid;
- een beschrijving van de verwachte ontwikkeling van het milieu bij ongewijzigd beleid en bij gewijzigd beleid volgens een aantal relevant geachte scenario's.

Om deze opdracht te realiseren, worden verschillende rapporten uitgewerkt: jaarlijkse themarapporten (MIRA-T), vijfjaarlijkse scenariorapporten (MIRA-S) en tweejaarlijkse beleidsevaluatierapporten (MIRA-BE). Naast studie- en samenwerkingsovereenkomsten voor de invulling van de milieurapporten worden ook studies voorzien ter bredere ondersteuning van de milieurapportering. Het voorliggend verslag betreft het verslag van een dergelijke studie.

Doelstelling van deze studie is om op een objectieve en statistisch correcte wijze tot conclusies te komen over de evolutie (trendanalyse) van de waterkwaliteit in Vlaanderen. Daarvoor zal gebruik gemaakt worden van statistische methoden die rekening houden met de temporele en spatiale afhankelijkheden in de data. De onderzoeksvraag betreft trendanalyse op Vlaams en andere aggregatieniveaus (bv. bekkens) en wordt beperkt tot de basisparameters van het fysisch-chemisch meetnet (O<sub>2</sub>, NO<sub>3</sub>, NO<sub>2</sub>, NH<sub>4</sub>, BZV, CZV, oPO<sub>4</sub>).

Op de stuurgroepvergaderingen van 28 november 2006 en 21 februari 2007, werd in onderling overleg het volgende afgesproken:

- voor de trendanalyse wordt meetjaar 2005 vergeleken met alle individuele meetjaren uit de periode 1990-2004;
- De trends hebben betrekking op de gemiddelde concentraties;
- Meetpunten worden geaggregeerd volgens de volgende criteria: bekken, categorie, kwaliteit, saliniteit, stroming en waterloop.

De trendanalyse wordt op twee niveaus uitgevoerd:

1. op iedere individuele meetplaats, wordt ieder jaar vergeleken met het referentiejaar 2005 (de resultaten worden gegeven in Hoofdstuk 3.2 en Appendices 1 t.e.m. 11 en Appendix 13);
2. meetplaatsen worden vooraf geaggregeerd in bekkens of andere aggregaties. Voor ieder van deze aggregaties wordt eveneens een trendanalyse uitgevoerd waarbij ieder jaar met het referentiejaar 2005 wordt vergeleken. Dit betekent dat de conclusies hieruit betrekking hebben op de volledige aggregatie, i.p.v. op de individuele meetplaatsen (de resultaten worden gegeven in Hoofdstuk 4 en Appendix 14).

## 1.2 De Data

Voor deze studie worden enkel de basisparameters van het fysisch-chemisch meetnet (O<sub>2</sub>, NO<sub>3</sub>, NO<sub>2</sub>, NH<sub>4</sub>, BZV, CZV, oPO<sub>4</sub>) beschouwd. Hiervoor zijn de data ter beschikking gesteld van de jaren 1990-2006, afkomstig van 11 bekkens

(Beneden Schelde, Boven Schelde, Gentse kanalen, Brugse polders, Nete, Maas, Dender, Demer, Leie, Nete en Dijle-Zenne). In het totaal zijn er data van 3988 meetplaatsen.

In Appendix 12 worden voor ieder jaar, iedere meetplaats en iedere variabele een overzicht gegeven van de verdeling van de “teken”-verdeling van de data. Met “teken” wordt verwezen naar de codering in de VMM-meetdatabank. Er zijn drie codes:

- teken = +1  
observatie is boven de detectielimiet;
- teken=0  
observatie is kleiner dan de detectielimiet;
- teken=-1  
observatie is afwezig.

### 1.3 Structuur van het Eindverslag

Dit eindverslag bestaat uit verscheidene delen. Dit gedeelte bevat het tekstgedeelte met daarin de methodologie (Hoofdstuk 2), een bondig overzicht van de resultaten voor de afzonderlijke meetplaatsen (Hoofdstuk 3), een bondig overzicht van de resultaten op de aggregatieniveaus (Hoofdstuk 4), de resultaten en besluiten voor heel Vlaanderen (Hoofdstuk 5) en de conclusies (Hoofdstuk 6).

Aangezien de methodologie beschreven in Hoofdstuk 2 voor sommige lezers te technisch is, starten Hoofdstukken 3 en 4 met een korte niet-technische samenvatting van de methode. Deze hoofdstukken bevatten bovendien ook minstens één voorbeeld waarvan de interpretatie besproken wordt. Bijgevolg kan dit verslag ook gelezen worden door Hoofdstuk 2 over te slaan.

Naast dit tekstgedeelte, zijn er 15 appendices als afzonderlijke delen. In Appendices 1 t.e.m. 11 worden de figuren met daarin de resultaten van de afzonderlijke meetplaatsen weergegeven. Appendix 12 geeft informatie over de beschikbaarheid van de data gebruikt in deze studie. In het bijzonder wordt voor iedere variabele en voor iedere meetplaats aangegeven hoeveel observaties beschikbaar waren voor de analyses in deze studie, alsook het aantal observaties kleiner dan

de detectielimiet. in Appendix 13 wordt aan de hand van tabellen en figuren een samenvattend overzicht gegeven van de resultaten van de trenddetectie op individuele meetplaatsen. Appendix 14 bevat de figuren die de resultaten samenvatten van de trendanalyses op de verschillende aggregatieniveaus, en tenslotte wordt de programmatuur gegeven in Appendix 15.

## 1.4 Software

Alle berekeningen werden uitgevoerd met de R-software ([www.r-project.org](http://www.r-project.org)). Dit is een erg verspreide statistische programmeertaal. In Appendix 15 worden alle codes gegeven.

## 1.5 Terminologie

Enkele verduidelijkingen van de gebruikte terminologie:

- variabele:  
In VMM-terminologie is dit de parameter. In het bijzonder beschouwen we hier: O<sub>2</sub>, NO<sub>3</sub>, NO<sub>2</sub>, NH<sub>4</sub>, BZV, CZV en oPO<sub>4</sub>.  
We gebruiken heel bewust de term variabele om geen verwarring te krijgen met het begrip “parameter” in de conventionele statistiek-terminologie.
- regelmatige steekproefname:  
Wanneer observaties op gelijke tijdsintervallen van elkaar gemeten zijn, bv. maandelijks, dan spreken we over “regelmatige steekproefname”
- multipliciteit:  
Het probleem van multipliciteit treedt op wanneer de conclusies van meerdere statistische hypothesetesten simultaan geïnterpreteerd worden. Wanneer meerdere statistische testen worden uitgevoerd op het 5% significantieniveau, kan het volgende optreden. Stel dat we de zuurstofconcentratie op meetplaats 250000 analyseren. We hebben data van 11 jaren, zodat we 10 hypothesetesten kunnen uitvoeren. Iedere test vergelijkt één jaar met het referentiejaar 2005, en iedere test wordt uitgevoerd op het 5% significantieniveau. Dan is de globale kans op het maken van een type I fout groter



dan het nominale 5% significantieniveau. Er kan aangetoond worden dat de kans op het maken van een type I fout nu eventueel tot  $10 \times 5\% = 50\%$  kan oplopen. Dit betekent dat, wanneer er in werkelijkheid in geen enkel jaar een gemiddeld verschil is met het referentiejaar 2005, er een kans van maximaal 50% is om minimaal één significant testresultaat te bekomen, en dus een foute conclusie te formuleren.

- seizoenen:

In dit verslag wordt de term “seizoen” ruimer gebruikt dan de traditionele definitie van winter, lente, zomer en herfst. De exacte definitie hangt af van het aantal metingen per jaar. Indien er bv. 10 metingen zijn per jaar, dan spreken we over 10 seizoenen. Het correspondeert dus met de meetfrequentie van een tijdreeks.

- stationariteit:

De statistische methoden gebruikt in deze studie steunen op de veronderstelling van “stationariteit”. In de context van deze studie betekent dit:

(1) er wordt verondersteld dat de statistische afhankelijkheid tussen twee metingen op eenzelfde meetplaats, maar op een gegeven tijdsafstand van elkaar, niet afhangt van het tijdstip;

(2) er wordt verondersteld dat de statistische afhankelijkheid tussen twee metingen op eenzelfde tijdstip, maar op twee gegeven meetpunten, niet afhangt dan het tijdstip.

# Hoofdstuk 2

## Methodologie

### 2.1 Inleiding

In dit hoofdstuk worden de statistische methoden in detail beschreven. In Sectie 2.2 worden de methoden behandeld die gebruikt worden voor de trendanalyse op meetplaatsniveau, en in Sectie 2.3 bespreken we de methoden voor de trendanalyse met uitspraken op aggregatieniveau.

In de inleiding (Sectie 1.1) werd de doelstelling van dit project in detail besproken. Hieruit volgt dat de doelstelling voornamelijk herleid is tot het vergelijken van gemiddelden tussen verschillende jaren. Niettegenstaande het vergelijken van gemiddelden in de klassieke statistiek erg eenvoudig is (bv. t-testen), stellen er zich bij de VMM-meetdatabank toch enkele erg specifieke problemen die verhinderen om de klassieke methoden toe te passen. In het bijzonder:

- observaties op eenzelfde meetplaats vormen een tijdreeks, en observaties binnen een tijdreeks zijn niet onafhankelijk van elkaar. Meestal zijn observaties die dicht bij elkaar in de tijd geobserveerd zijn, sterker afhankelijk dan observaties verder in de tijd van elkaar gemeten.
- wanneer de observaties van meerdere meetpunten gebruikt moeten worden om uitspraken op bekkenniveau te doen, dan vinden we afhankelijkheid tussen observaties van meetplaatsen die dicht bij elkaar “in de ruimte” liggen. Zo zijn observaties van twee opeenvolgende meetpunten langs een-

zelfde waterloop sterker afhankelijk van elkaar dan observaties van meetpunten die verder uit elkaar liggen volgens de riviertopologie.

Aangezien de klassieke statistische methoden onafhankelijkheid veronderstellen, kunnen we deze hier niet correct toepassen. Hiermee bedoelen we het volgende. Indien we niet correct rekening houden met de afhankelijkheidsstructuur tussen de waarnemingen, dan

- hebben de berekende 95% betrouwbaarheidsintervallen niet de interpretatie van een 95% betrouwbaarheidsinterval: ze omvatten de werkelijke parameterwaarde (bv. het gemiddelde) niet noodzakelijk met de vooropgestelde kans van 95%. Wat dan wel de omvattingkans is van het berekende interval, is niet gekend. Zodoende is het berekende interval niet informatief en dus waardeloos!
- hebben de  $p$ -waarden van de statistische testen niet noodzakelijk de kansinterpretatie die ze zouden moeten hebben. Het gevolg is dat het niet gewaarborgd is dat de kans om foutief de nulhypothese te verwerpen (bv. het concluderen van een significante trend) niet groter is dan het vooropgestelde significantieniveau. Dus als het significantieniveau 5% is, dan zou het toch kunnen dat de werkelijke kans op het maken van een foutieve conclusie groter is dan 5%, en hoeveel groter de kans is, is niet gekend. Zodoende is de conclusie van de hypothesetest niet informatief en dus waardeloos!

Uit bovenstaande redeneringen volgt dat we statistische methoden nodig hebben die wel kunnen omgaan met afhankelijkheid tussen de observaties. Afhankelijkheid wordt in de statistiek typisch gekwantificeerd met de *covariantie* (of *correlatie*). Het zal daarom ook noodzakelijk zijn om de betrokken covarianties goed te kunnen schatten. Bij een foute schatting van deze covarianties, zullen de statistische methoden weerom niet betrouwbare resultaten geven. In Secties 2.2 en 2.3 zullen we daarom statistische methoden ontwikkelen die rekening houden met de afhankelijkheid, en zullen we behoefte hebben aan goede schattingen van de afhankelijkheidsstructuur.

Naast de problematiek van de afhankelijkheid tussen de observaties, vertonen de VMM-meetdata nog een andere moeilijkheid: ontbrekende waarnemingen en/of

onregelmatige steekproefname. Hiermee bedoelen we het volgende. De eenvoudigere methoden voor trendanalyse veronderstellen dat de observaties op gelijke tijdsintervallen van elkaar gemeten zijn, bv. maandelijks. Dit is zeker niet het geval voor de VMM-meetdata. Deze veronderstelling van “regelmatige steekproefname” vereenvoudigt de methode om de temporele afhankelijkheid te schatten onder bepaalde veronderstellingen. Indien de veronderstelling van regelmatige steekproefname niet gemaakt kan worden, dan moet de temporele afhankelijkheid modelmatig geschat worden. Dit is veel moeilijker, en de correctheid van deze methode hangt tevens af van de correctheid van het veronderstelde model! Bovendien zou het onderliggende model voor de temporele afhankelijkheid kunnen variëren van de ene meetplaats tot de andere. Dit modelleren vraagt erg veel tijd. Gezien het erg grote aantal meetplaatsen dat in deze studie geanalyseerd moet worden, wensen we deze modelmatige weg te vermijden, en moeten we toch trachten om te gaan met de veronderstelling van regelmatige steekproefname. We geven vervolgens enkele argumenten die ons laten vermoeden dat dit geen erg grote problemen zal veroorzaken. Op de meeste meetplaatsen wordt er beoogd om maandelijks te bemonsteren. Uit vorige studies (Thas en Ottoy, 1996) blijkt dat metingen op een afstand van één maand geen erg grote temporele afhankelijkheid meer vertonen. Dus als metingen niet exact op 30 of 31 dagen van elkaar gemeten zijn, zal dit geen erg grote fouten introduceren.

In de Secties 2.2 en 2.3 worden de statistische methoden beschreven. In deze beschrijvingen gaan we er steeds van uit dat de observaties voldoen aan de veronderstelling van regelmatige steekproefname. Later, in Sectie 2.4, gaan we dieper in op de problematiek van ontbrekende observaties en regelmatige steekproefname. We zullen daar de methode beschrijven hoe de data uit de VMM-meetdatabank aangepast zijn om zo weinig mogelijk problemen te hebben met de problematiek van ontbrekende observaties en de veronderstelling van regelmatige steekproefname.

In Sectie 2.4 vermelden we ook nog de problematiek van de observaties onder de detectielimiet. Ook hiermee wordt in de beschrijving van de methodes in de Secties 2.2 en 2.3 geen rekening gehouden. Ten slotte wordt in Sectie 2.5 een discussie gegeven over *multipliciteit*. Dit fenomeen duikt op omdat in deze studie heel veel statistische testen uitgevoerd worden en het simultaan interpreteren van de resultaten van meerdere statistische testen moet met de nodige voorzichtigheid gebeuren. Dit is het onderwerp van Sectie 2.5.

## 2.2 Statistische Methoden voor Trendanalyse op Meetplaatsniveau

### 2.2.1 Notatie

We beschouwen hier de data van één enkele meetplaats. Deze vormen dus een tijdreeks. Zoals aangehaald in Sectie 2.1, gaan we ervan uit dat de observaties voldoen aan de veronderstelling van regelmatige steekproefname. Niettegenstaande de meeste tijdreeksen in de VMM-meetdatabank maandelijkse metingen bevatten, zullen we hier een meer algemene terminologie hanteren: ieder jaar wordt ingedeeld in *seizoenen*. Bij maandelijkse metingen zijn de seizoenen gelijk aan maanden, bij wekelijkse metingen zijn de seizoenen gelijk aan weken, enz. Ieder observatie-tijdstip kan dus gerepresenteerd worden door een jaar en een seizoen.

Stel  $Y_{ijs}$  de observatie op meetplaats  $i$ , jaar  $j$  en seizoen  $s$ . We nummeren de meetplaatsen:  $i = 1, \dots, I$ . We nummeren de jaren:  $j = 1, \dots, J$ . We nummeren de seizoenen:  $s = 1, \dots, S$ . Verder definiëren we de vector

$$\mathbf{Y}_{ij}^t = (Y_{ij1}, \dots, Y_{ijS}),$$

welke de observaties bevat van meetplaats  $i$  en jaar  $j$ . De doelstellingen in dit project m.b.t. individuele meetplaatsen (bv. meetplaats  $i$ ) kunnen herleid worden tot het vergelijken van de gemiddelden van de observaties uit de jaren  $j$  en  $k$ , rekening houdend met een mogelijks seizoenseffect. Om met dit seizoenseffect rekening te houden, werken we verder met de paarsgewijze verschillen tussen de elementen van de vectoren  $\mathbf{Y}_{ij}$  en  $\mathbf{Y}_{ik}$  (i.e. verschillen tussen observaties van gelijke seizoenen). Stel

$$\mathbf{D}_{i[jk]}^t = \mathbf{Y}_{ij} - \mathbf{Y}_{ik},$$

welke deze paarsgewijze verschillen bevat. We stellen de individuele elementen van  $\mathbf{D}_{i[jk]}$  voor als  $D_{i[jk]s}$ ,  $s = 1, \dots, S$ , en we noemen zo'n individueel verschil de *verschilobservatie* van seizoen  $s$  (in meetplaats  $i$ ).

Tot hier toe hebben we nog geen distributionele veronderstellingen gemaakt. Voor de constructie van de statistische testen, hebben we enkele distributionele karakteristieken onder de nulhypothese nodig. De nulhypothese impliceert reeds dat het gemiddelde (verwachtingswaarde) van  $\mathbf{D}_{i[jk]}$  gelijk is nul. We hebben

verder ook nog de momenten van tweede orde nodig (i.e. varianties en covarianties). We stellen de variantie-covariantiematrix van  $\mathbf{D}_{i[jk]}$  voor als  $\Sigma_i$ , i.e.  $\text{Var}[\mathbf{D}_{i[jk]}] = \Sigma_i$ . Merk op dat  $\Sigma_i$  niet afhangt van de jaren  $j$  of  $k$ . We hebben hierbij immers de veronderstelling gemaakt dat de covariantiestructuur van  $\mathbf{D}_{i[jk]}$  niet wijzigt over de tijd. In technische bewoordingen wordt dit de *stationariteitsveronderstelling* genoemd. Deze veronderstelling moeten we noodzakelijk maken, omdat we anders niet in staat zijn de temporele afhankelijkheidsstructuur te schatten aan de hand van de data. Merk ook op dat de niet-diagonaalelementen van  $\Sigma_i$  de covarianties zijn tussen verschilobservaties uit verschillende seizoenen. Dit laat toe om op een niet-modelmatige wijze rekening te houden met de temporele afhankelijkheid tussen de observaties.

Aangezien  $\Sigma_i$  niet gekend is, moet deze matrix uit de data geschat worden. Dit wordt behandeld in Sectie 2.2.2. We stellen twee verschillende statistische testen voor. Deze worden in Secties 2.2.3 en 2.2.4 geïntroduceerd. Op de stuurgroepvergadering van 21 februari 2007 is besloten om enkel de eerste test toe te passen op de data.

## 2.2.2 Het Schatten van de Afhangelijkheidsstructuur

Zoals eerder vermeld, veronderstellen we dat de variantie-covariantiematrix niet varieert met het jaar. Dit laat toe data uit meerdere jaren te gebruiken om  $\Sigma_i$  te schatten. Aangezien we wel toelaten dat  $\Sigma_i$  niet gelijk is op alle meetplaatsen, gebruiken we voor het schatten van  $\Sigma_i$  enkel de data van meetplaats  $i$ .

Om  $\Sigma_i$  te kunnen schatten moeten we veronderstellen dat er minstens data van drie jaar aanwezig zijn (i.e.  $J \geq 3$ ).

Stel

$$\bar{\mathbf{D}}_i = \frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=i+1}^J \mathbf{D}_{i[jk]}.$$

Dit is de vector met de gemiddelde verschilobservaties, waarbij gemiddeld is over alle  $J(J-1)/2$  jaarcombinaties.

De schatter van  $\Sigma_i$  wordt nu gegeven door

$$\hat{\Sigma}_i = \frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=i+1}^J [\mathbf{D}_{i[jk]} - \bar{\mathbf{D}}_i] [\mathbf{D}_{i[jk]} - \bar{\mathbf{D}}_i]^t. \quad (2.1)$$

### 2.2.3 Test 1

Op de stuurgroepvergadering van 21 februari 2007 is besloten om enkel de test beschreven in deze sectie toe te passen op de data. De eerste test (Test 1 of T1) wordt gebruikt voor het testen dat de gemiddelde concentratie in de jaren  $j$  en  $k$  op meetplaats  $i$  gelijk is. Dit laat zich voorstellen als

$$H_0 : \sum_{s=1}^S E [D_{i[jk]s}] = 0, \quad (2.2)$$

waar  $E[\cdot]$  de verwachtingswaarde (= gemiddelde op populatieniveau) voorstelt. We zullen steeds tweezijdig testen. De alternatieve hypothese is dus

$$H_1 : \text{niet } H_0.$$

Het is belangrijk te begrijpen wat deze nulhypothese betekent. In termen van een voorbeeldje zegt  $H_0$ : de gemiddelde zuurstofconcentratie op meetplaats 250000 in 2004 is gelijk aan de gemiddelde zuurstofconcentratie op meetplaats 250000 in 2005. Dit laat dus toe dat misschien in de winter er een hogere gemiddelde concentratie is in 2004 dan in 2005, maar dat het omgekeerde geobserveerd wordt in de zomermaanden. Beide verschillen compenseren hier elkaar zodat over heel het jaar er geen gemiddeld verschil meer is vast te stellen.

Deze eerste test is een *liniare statistiek*. De rationale steunt op de observatie dat  $\sum_{s=1}^S D_{i[jk]s}$  een onvertekende schatter is van  $\sum_{s=1}^S E [D_{i[jk]s}]$ , en dit laatste is wat in de nulhypothese (Vergelijking (2.2)) expliciet voorkomt en gelijk aan nul verondersteld wordt.

We gebruiken verder de notatie  $\mathbf{1}$  voor een kolomvector met  $S$  elementen gelijk aan 1.

De teststatistiek is

$$T_{1;i[jk]} = \frac{\sum_{s=1}^S D_{i[jk]s}}{\hat{\sigma}_i}, \quad (2.3)$$

met

$$\hat{\sigma}_i^2 = \mathbf{1}^t \hat{\Sigma}_i \mathbf{1}. \quad (2.4)$$

(De uitdrukking in Vergelijking (2.4) is gelijk aan de som van alle elementen in de matrix  $\hat{\Sigma}_i$ .)

Om de test te kunnen uitvoeren, moeten we de nulverdeling van de teststatistiek  $T_{1;i[jk]}$  kennen. Bij benadering is de nulverdeling een standaard normale verdeling. Stel  $z_{1-\alpha}$  gelijk aan het  $100(1 - \alpha)\%$  percentiel van een standaard normale verdeling, dan wordt de beslissingsregel van de tweezijdige test op het  $\alpha$  significantieniveau:

$$\begin{aligned} T_{1;i[jk]} &\leq z_{1-\alpha/2} && \text{aanvaard } H_0 \\ T_{1;i[jk]} &> z_{1-\alpha/2} && \text{verwerp } H_0. \end{aligned}$$

Voor de huidige significantieniveaus van  $\alpha = 0.01$ ,  $\alpha = 0.05$  en  $\alpha = 0.10$ , zijn de percentielen respectievelijk  $z_{0.995} = 2.576$ ,  $z_{0.975} = 1.960$  en  $z_{0.95} = 1.645$ .

De hier beschreven methode houdt geen rekening met de multipliciteit van hypothesetesten bij het gebruik van de methode voor het simultaan vergelijken van meerdere jaren. Meer details worden gegeven in Sectie 2.5.

## 2.2.4 Test 2

De tweede test (Test 2 of T2) wordt ook gebruikt voor het testen dat de gemiddelde concentratie in de jaren  $j$  en  $k$  op meetplaats  $i$  gelijk is, maar met de extra restrictie dat deze gelijkheid opgaat voor ieder seizoen afzonderlijk. Dit laat zich voorstellen als

$$H_0 : E [D_{i[jk]1}] = E [D_{i[jk]2}] = \dots = E [D_{i[jk]S}] = 0. \quad (2.5)$$

Voor het testen van deze nulhypothese stellen we een *kwadratische vorm* voor als teststatistiek. De teststatistiek is

$$T_{2;i[jk]} = \mathbf{D}_{i[jk]}^t \hat{\Sigma}_i^{-1} \mathbf{D}_{i[jk]}. \quad (2.6)$$

De nulverdeling van  $T_{2;i[jk]}$  is bij benadering een chi-kwadraat verdeling met  $S$  vrijheidsgraden (i.e.  $\chi_S^2$ ). In analogie met de standaard normale verdeling, stellen we het  $100(1 - \alpha)\%$  percentiel voor door  $\chi_{S;1-\alpha}^2$ . De beslissingsregel voor de test op het  $\alpha$  significantieniveau wordt nu:

$$\begin{aligned} T_{1;i[jk]} &\leq \chi_{S;1-\alpha}^2 && \text{aanvaard } H_0 \\ T_{1;i[jk]} &> \chi_{S;1-\alpha}^2 && \text{verwerp } H_0. \end{aligned}$$



Wanneer er 12 seizoenen zijn (maandelijke observaties) en voor de courante significantieniveaus van  $\alpha = 0.01$ ,  $\alpha = 0.05$  en  $\alpha = 0.10$ , zijn de percentielen respectievelijk  $\chi_{12;0.99}^2 = 26.22$ ,  $\chi_{12;0.95}^2 = 21.03$  en  $\chi_{12;0.95}^2 = 18.55$ .

Voor zowel Test 1 als voor Test 2, zullen we eerder gebruik maken van  $p$ -waarden. Als we  $p$  gebruiken als notatie voor de  $p$ -waarde van een test, dan wordt de beslissingsregel voor een test op het  $\alpha$  significantieniveau:

$$\begin{aligned} p &\geq \alpha && \text{aanvaard } H_0 \\ p &< \alpha && \text{verwerp } H_0. \end{aligned}$$

De hier beschreven methode houdt geen rekening met de multipliciteit van hypothestetesten bij het gebruik van de methode voor het simultaan vergelijken van meerdere jaren. Meer details worden gegeven in Sectie 2.5.

## 2.3 Statistische Methoden voor Trendanalyse op Aggregatieniveau

### 2.3.1 Inleiding

Ook de trendanalyse op aggregatieniveau beoogt het vergelijken van twee jaren. De nulhypothese die moet getest worden is dat de gemiddelde concentratie van een variabele gelijk is in de jaren  $j$  en  $k$ , en dit uitgemiddeld over alle meetplaatsen  $i = 1, \dots, I$  van dit aggregatieniveau. Voor een individuele meetplaats  $i$  kan de teststatistiek (2.3) beschouwd worden als een gestandaardiseerd verschil tussen de jaargemiddelden. De standardisatie zorgt er min of meer voor dat de teststatistieken  $T_{1;i[jk]}$  vergelijkbaar zijn over de meetplaatsen  $i = 1, \dots, I$  heen. De testen B1 t.e.m. B3 steunen op deze vaststelling. Aangezien alle testen in deze sectie gebaseerd zijn op teststatistiek 1 (i.e.  $T_{1;i[jk]}$  uit (2.3)) laten we in deze sectie de index 1 weg en schrijven we  $T_{i[jk]} = T_{1;i[jk]}$ .

Stel  $\mathbf{T}_{[jk]}^t = (T_{1[jk]}, \dots, T_{I[jk]})$ . Onder de nulhypothese is de verwachtingswaarde van  $\mathbf{T}_{[jk]}$  gelijk aan nul, i.e.  $E[\mathbf{T}_{[jk]}] = \mathbf{0}$ . We stellen de variantie-covariantiematrix van  $\mathbf{T}_{[jk]}$  voor als  $\Sigma_T$ . Deze kan geschat worden op analoge wijze als in

Sectie 2.2.2. Stel

$$\bar{\mathbf{T}} = \frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=i+1}^J \mathbf{T}_{[jk]},$$

dan is de schatter van  $\Sigma_T$  gegeven door

$$\hat{\Sigma}_T = \frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=i+1}^J [\mathbf{T}_{[jk]} - \bar{\mathbf{T}}] [\mathbf{T}_{[jk]} - \bar{\mathbf{T}}]^t.$$

De correctheid van deze procedure steunt op de veronderstelling dat de afhankelijkheidsstructuur tussen de meetplaatsen niet wijzigt over de tijd. Dit leidt tot de stationariteitsvereiste die deze procedure ondersteunt.

### 2.3.2 Testen B1 en B2

Het (gewogen) gemiddelde (of de som) van de teststatistieken  $T_{i[jk]}$  ( $i = 1, \dots, I$ ) vormt dus een goede basis voor een teststatiek voor het testen op aggregatieniveau. De testen in deze sectie testen de nulhypothese dat het algemene gemiddelde (over alle meetplaatsen binnen de aggregatie en over alle seizoenen) gelijk is in de jaren  $j$  en  $k$ , i.e.

$$H_0 : \sum_{i=1}^I \left\{ \sum_{s=1}^S \mathbb{E} [D_{i[jk]s}] \right\} = 0.$$

Testen B1 en B2 verschillen enkel in de wijze van standardisatie. In test B1 wordt pas gestandardiseerd na het berekenen van de som (of gemiddelde) over de meetplaatsen van de aggregatie, en in test B2 wordt iedere  $T_{i[jk]}$  eerst gestandardiseerd alvorens de som te berekenen. Na sommatie moet er uiteraard terug gestandardiseerd worden. Test B2 is dus eigenlijk een gewogen versie van B1, waarbij meetplaatsen met minder variabiliteit meer doorwegen. Volgens statistische theorie zou B2 een grotere *power* moeten hebben. Om deze reden werd geopteerd voor test B2. Voor de volledigheid geven we toch ook test B1

De teststatistiek B1 wordt gegeven door

$$B_{1;j}^k = \frac{\sum_{i=1}^I T_{i[jk]}}{\hat{\sigma}_B},$$

met

$$\hat{\sigma}_B^2 = \mathbf{1}^t \hat{\Sigma}_T \mathbf{1}$$

(i.e.  $\hat{\sigma}_B^2$  is de som van alle elementen uit de matrix  $\hat{\Sigma}_T$ ). Om de test te kunnen uitvoeren, moeten we de nulverdeling van de teststatistiek  $B_{1;j}^k$  kennen. Bij benadering is de nulverdeling een standaard normale verdeling. Stel  $z_{1-\alpha}$  gelijk aan het  $100(1-\alpha)\%$  percentiel van een standaard normale verdeling, dan wordt de beslissingsregel van de tweezijdige test op het  $\alpha$  significantieniveau:

$$\begin{aligned} B_{1;j}^k &\leq z_{1-\alpha/2} && \text{aanvaard } H_0 \\ B_{1;j}^k &> z_{1-\alpha/2} && \text{verwerp } H_0. \end{aligned}$$

Voor de courante significantieniveaus van  $\alpha = 0.01$ ,  $\alpha = 0.05$  en  $\alpha = 0.10$ , zijn de percentielen respectievelijk  $z_{0.995} = 2.576$ ,  $z_{0.975} = 1.960$  en  $z_{0.95} = 1.645$ .

De teststatistiek B2 wordt gegeven door

$$B_{2;j}^k = \left( \sum_{i=1}^I \frac{T_{i[jk]}}{\hat{\sigma}_{T_i}} \right) / \hat{\sigma}_T,$$

met  $\hat{\sigma}_{T_i}^2$  het  $i$ -de diagonaalelement uit de matrix  $\hat{\Sigma}_T$ , en

$$\hat{\sigma}_T = \mathbf{1}^t \mathbf{W}_T^{-1} \hat{\Sigma}_T \mathbf{W}_T^{-1} \mathbf{1},$$

waar  $\mathbf{W}_T$  de diagonaalmatrix voorstelt met elementen  $\hat{\sigma}_{T_i}^2$  ( $i = 1, \dots, I$ ). Om de test te kunnen uitvoeren, moeten we de nulverdeling van de teststatistiek  $B_{2;j}^k$  kennen. Bij benadering is de nulverdeling een standaard normale verdeling. De beslissingsregel voor de uitvoering van de B2 test is dus juist dezelfde als deze voor de B1 test.

### 2.3.3 Test B3

Test B3 verschilt van de testen B1 en B2 zoals testen 1 en 2 uit Secties 2.2.3 en 2.2.4 verschillen. De nulhypothese van test B3 luidt

$$H_0 : \sum_{s=1}^S \mathbb{E} [D_{1[jk]s}] = \sum_{s=1}^S \mathbb{E} [D_{2[jk]s}] = \dots = \sum_{s=1}^S \mathbb{E} [D_{I[jk]s}] = 0.$$

Dus, de nulhypothese stelt dat er op iedere meetplaats van de aggregatie geen trend is; compensatie is dus niet toegestaan. Deze nulhypothese kan getest worden aan de hand van de kwadratische vorm

$$B_{3;j}^k = \mathbf{T}_{[jk]}^t \hat{\Sigma}_T^{-1} \mathbf{T}_{[jk]}.$$

De nul distributie van  $B_{3;j}^k$  is bij benadering een chi-kwadraat distributie met  $I$  vrijheidsgraden (i.e.  $\chi_I^2$ ). In analogie met de standaard normale distributie, stellen we het  $100(1 - \alpha)\%$  percentiel voor door  $\chi_{I;1-\alpha}^2$ . De beslissingsregel voor de test op het  $\alpha$  significantieniveau wordt nu:

$$\begin{aligned} B_{3;j}^k &\leq \chi_{I;1-\alpha}^2 && \text{aanvaard } H_0 \\ B_{3;j}^k &> \chi_{I;1-\alpha}^2 && \text{verwerp } H_0. \end{aligned}$$

## 2.4 Regelmatige Steekproefname, Ontbrekende Waarnemingen en de Detectielimiet

### 2.4.1 Regelmatige steekproefname en ontbrekende waarnemingen

Zoals reeds aangehaald in Sectie 2.1, steunen onze methoden op de veronderstelling van regelmatige steekproefname. Dit houdt in dat de observaties worden ingedeeld in *seizoenen* zodat de covarianties tussen observaties uit seizoenen  $i$  en  $j$  gelijk zijn over alle jaren heen. Het is deze (stationariteits) veronderstelling die toelaat de covariantie te schatten. Aangezien de metingen in de VMM-meetdatabank niet op exact gelijke tijdsintervallen gemeten zijn, hanteren we hier een voorbereidingsalgorithme. We beschrijven het algorithme hier voor een tijdreeks die slechts twee jaren omvat, maar de veralgeming zal triviaal zijn.

Het algorithme:

Beschouw de data uit de jaren  $j_1$  en  $j_2$ , met respectievelijk  $n_1$  en  $n_2$  observaties.

1. Neem de data van het jaar met de meeste observaties. Zonder afbreuk te doen aan de algemeenheid, stel dat jaar  $j_1$  dit jaar is. Dan definiëren we het aantal *seizoenen* voor deze tijdreeks als  $S = n_1$ , i.e. het "grootste"

jaar bepaalt het aantal seizoenen voor de volledige tijdreeks. Noem deze seizoenen  $s = 1, 2, \dots, S$ . Conform de notatie uit Sectie 2.2.1 stellen we deze data voor als  $\mathbf{Y}_{ij_1}^t = (Y_{ij_1 1}, \dots, Y_{ij_1 S})$ .

2. Voor  $s = 1, \dots, S$ , doe het volgende:
  - stel  $Y_{ij_2 s}$  gelijk aan de observatie uit jaar  $j_2$  waarvan het metingstijdstip het dichtst bij dat van  $Y_{ij_1 s}$  ligt. Indien deze tijdsafstand meer dan 60 dagen bedraagt, dan wordt  $Y_{ij_2 s}$  op “ontbrekend” gezet.

Bovenstaand algoritme kan dus aanleiding geven tot ontbrekende observaties (binnen de rigide seizoentaliteitsstructuur). De beschreven statistische methoden kunnen op de volgende manier aangepast worden:

- De variantie-covariantie matrix  $\Sigma$  uit Vergelijking (2.1) kan elementsgewijs geschat worden. Merk op dat ieder element correspondeert met één (diagonaal element) of twee seizoenen (niet-diagonaal element). Voor iedere element worden dan enkel de jaren genomen waarvoor er geen ontbrekende waarnemingen zijn voor de betrokken seizoenen. Deze methode heeft als enige nadeel dat  $\hat{\Sigma}$  nu niet meer noodzakelijk positief-definiet is. Dit kan in de R-code opgevangen worden.
- De teststatistieken uit Secties 2.2.3 en 2.2.4 kunnen eenvoudig aangepast worden door in de vectoren  $\mathbf{D}_{i[jk]}$  enkel de elementen  $D_{i[jk]s}$  op te nemen die geen betrekking hebben op ontbrekende observaties. Voor zowel Test1 als Test2, moeten hiervoor dan ook enkel de overeenkomstige kolommen en rijen uit  $\hat{\Sigma}$  gebruikt worden. Dus, als bv. voor de vergelijking van 2003 met 2005 er een ontbrekende observatie is voor oktober, dan wordt de overeenkomstige rij en kolom uit  $\hat{\Sigma}$  weggenomen alsvorens de teststatistieken te berekenen.

## 2.4.2 De detectielimiet

In deze studie werd opzettelijk geopteerd voor de analyse van de basisparameters van het fysisch-chemisch meetnet (O<sub>2</sub>, NO<sub>3</sub>, NO<sub>2</sub>, NH<sub>4</sub>, BZV, CZV, oPO<sub>4</sub>) omdat er voor deze variabelen weinig observaties onder de detectielimiet liggen. Gezien het geringe aantal (zie Appendix 12 voor de juiste aantallen) voorzien we weinig problemen. De hier gebruikte statistische methoden leiden bovendien

niet erg onder deze geringe problemen met de detectielimiet, o.a. omdat alle variantie-covariantiematrices op een niet-modelgebaseerde wijze geschat worden.

Indien de hier voorgestelde methodologie zou moeten toegepast worden op bv. de zware metalen, waarvoor wel veel observaties kleiner zijn dan de detectielimiet, dan kan deze methodologie tot fouten leiden. Een mogelijk aanpassing bestaat erin om de statistische methoden te baseren op rank-getransformeerde data. Dit maakt de technieken minder gevoelig voor de problematiek rond de detectielimiet, maar het zal ook leiden tot een vermindering van de power van de statistische testen (i.e. de kans op het detecteren van trends wordt kleiner). De ontwikkeling van deze methoden valt echter buiten het kader van dit project.

## 2.5 Multipliciteit

Het probleem van multipliciteit treedt op wanneer de conclusies van meerdere statistische hypothesetesten simultaan geïnterpreteerd worden. De kern van het probleem zit in de controle van de type I fout. Wanneer één individuele test wordt uitgevoerd op het  $\alpha = 5\%$  significantieniveau, dan garandeert de statistische test dat de kans op het maken van een type I fout gelijk is aan  $\alpha$ . In de huidige context betekent dit dat er slechts een kans is van  $\alpha = 5\%$  om te besluiten dat bv. de gemiddelde zuurstofconcentratie op meetplaats 250000 in het jaar 2004 verschilt van deze van het referentiejaar 2005, wanneer er in werkelijkheid geen verschil is.

Wanneer echter meerdere statistische testen worden uitgevoerd op het  $5\%$  significantieniveau, kan het volgende optreden. Stel dat we de zuurstofconcentratie op meetplaats 250000 analyseren. We hebben data van 11 jaren, zodat we 10 hypothesetesten kunnen uitvoeren. Iedere test vergelijkt één jaar met het referentiejaar 2005, en iedere test wordt uitgevoerd op het  $5\%$  significantieniveau. Dan is de globale kans op het maken van een type I fout groter van het nominale  $5\%$  significantieniveau. Er kan aangetoond worden dat de kans op het maken van een type I fout nu eventueel tot  $10 \times 5\% = 50\%$  kan oplopen. Dit betekent dat, wanneer er in werkelijkheid in geen enkel jaar een gemiddeld verschil is met het referentiejaar 2005, er een kans van maximaal  $50\%$  is om minimaal één significant testresultaat te bekommen, en dus een foute conclusie te formuleren.

Er bestaan vele correctiemethoden, maar we beschrijven hier enkel de meest een-

voudige. De Bonferroni correctiemethode bestaat erin om het nominale significantieniveau van iedere individuele test te delen door het totaal aantal simultaan te interpreteren testen. In het bovenstaand voorbeeld zouden we dus iedere test op het  $5\%/10 = 0.5\%$  significantieniveau uitvoeren. Bovenstaande redenering herhalend, zien we dat nu inderdaad de globale kans gecontroleerd wordt op  $10 \times 5\%/10 = 5\%$ , zoals gewenst.

De Bonferroni methode is heel eenvoudig te implementeren omdat de bovenstaande aanpassing aan het significantieniveau van de individuele testen eigenlijk ook rechtstreeks op de individuele  $p$ -waarden kan gebeuren. Bijvoorbeeld: indien er 10 simultane testen te interpreteren zijn, dan moeten alle individuele  $p$ -waarden met 10 vermenigvuldigd worden. Deze gecorrigeerde  $p$ -waarden kunnen nu weer met het traditionele  $\alpha = 5\%$  significantieniveau vergeleken worden.

Het enige nadeel van de Bonferroni methode is dat ze *conservatief* is. Dat wil zeggen dat de methode TE veilig is. De methode garandeert dat de eigenlijke globale kans op het maken van een type I fout zeker niet groter is dan 5%, maar misschien is de kans wel kleiner dan 5%. Door te-veilig te zijn, is er een risico op verlies van kracht (=power) van de statistische testen. We hebben er hier echter voor geopteerd om een correcte, maar eenvoudige methode voor te stellen, zodat

1. de resultaten uit deze studie voor meerdere doeleinden kunnen gebruikt worden. Bijvoorbeeld: (1) als de resultaten gebruikt moeten worden om enkel 2004 met het referentiejaar 2005 te vergelijken, moeten er geen correcties uitgevoerd worden; (2) als de resultaten gebruikt worden om de jongste 5 jaren met 2005 te vergelijken, moeten alle  $p$ -waarden met 5 vermenigvuldigd worden; (3) als de resultaten gebruikt worden om de jongste 10 jaren met 2005 te vergelijken, moeten alle  $p$ -waarden met 10 vermenigvuldigd worden; enz.

Andere correctiemethoden vragen dikwijls naar een volledige herrekening.

2. de Bonferroni methode is eenvoudig. Andere methoden zijn vaak computationeel veel zwaarder (bv. via resampling technieken), wat niet haalbaar is gezien de erg grote omvang van de VMM-meetdata.

Op de stuurgroepvergadering van 21 februari 2007 is echter besloten om alle resultaten in deze studie te presenteren zonder toepassing van enige correctiemethode. Op deze wijze kan iedere lezer zelf beslissen hoe om te gaan met de multipliciteit, afhankelijk van de context.

## 2.6 Simulatiestudie

Alle testen ontwikkeld in dit project hebben een asymptotische standaard normale nulverdeling. Dit is een theoretische eigenschap die eigenlijk pas gegarandeerd is in heel erg grote datasets. In deze sectie gaan we na of het gebruik van de standaard normale verdeling voor de berekening van de  $p$ -waarden verantwoord is in kleine steekproeven.

De correctheid (*onvertekendheid*) van een statistische test wordt nagegaan door de werkelijke kans op het maken van een type I fout te schatten. Als een test wordt uitgevoerd op het  $\alpha$ -significantieniveau, dan zou de kans op het maken van een type I fout gelijk moeten zijn aan  $\alpha$ . Een type I fout betekent: "het onterecht verwerpen van de nulhypothese". Dit is dus het verwerpen van de nulhypothese wanneer in werkelijk de nulhypothese juist is. De kans op het maken van deze fout kan geschat worden in een simulatiestudie door meermaals (bv. 10000 maal) een dataset te simuleren onder de conditie van de nulhypothese, en op iedere dataset de test toe te passen op het  $\alpha$  significantieniveau. Een teller houdt bij hoeveel maal de nulhypothese toch verworpen wordt. Dit aantal gedeeld door het aantal simulaties levert een goede schatting van de werkelijke kans op het maken van de type I fout. Indien er gesimuleerd wordt onder de condities waarop de theorie van de test gebaseerd is (bv. normaliteit en erg grote steekproeven), dan moet de werkelijke kans op een type I fout exact gelijk zijn aan  $\alpha$ , want dit is wat de theorie garandeert. Hier zullen we simuleren onder condities die we afleiden van de data uit de VMM databank (dus niet-normaal, geen hele grote steekproeven).

Aangezien enkel testen T1 (Sectie 2.2.3) en B3 (Sectie 2.3.3) verder gebruikt worden, worden enkel deze testen bestudeerd in deze sectie.

### 2.6.1 Test T1

De simulatiestudie die we hier presenteren is gebaseerd op de data van ammonium uit het bekken van de Benedenschelde. Hiervoor hebben we in de periode 1990-2005 ieder jaar tussen de 49 en de 114 bruikbare meetplaatsen. Als voorbereiding op de simulatiestudie, wordt van iedere observatie het gemiddelde afgetrokken van de betreffende meetplaats/jaar. Op deze wijze voldoen de data zeker aan de nulhypothese. We noemen dit de *residuele data*. In iedere *run* van de simulatiestudie, worden voor de jaren 1990-2005 at random 7 tot 12 residu-



elen geselecteerd voor ieder van deze 16 jaren. Om de seriële correlatie zo goed mogelijk te behouden, worden deze 16 observaties geconstrueerd als 4 blokjes van 4 opeenvolgende residuen uit de residuele dataset. Dit is analoog aan de *moving block bootstrap* methode die gebruikt om met serieel gecorreleerde data te werken. We hebben 10000 simulatie runs gedaan. Op deze gesimuleerde data wordt test T1 toegepast op het  $\alpha = 5\%$  significantie niveau. De geschatte kans op een type I fout is 0.061. Dit is een klein beetje groter dan de vooropgestelde 0.05, maar binnen de huidige context is dit zeker aanvaardbaar. We besluiten dus dat Test 1 betrouwbare resultaten levert.

### 2.6.2 Test B3

Voor test B3 wordt een gelijkaardige simulatiestudie opgezet. We hebben nu gekozen voor CZV in de Basiskwaliteitsklasse van de kwaliteitsaggregatie. Hiervoor hebben we voor de jaren 1990-2005 tussen de 289 en 556 meetplaatsen. Ook deze observaties worden eerst getransformeerd naar de residuele data om aan de nulhypothese te voldoen. Een analoge simulatiestudie wordt opgezet, terug volgens het *moving block* principe om seriële correlatie te behouden. In de simulatiestudie hebben we voor ieder jaar slechts 100 meetplaatsen beschouwd. In iedere simulatierun worden deze 100 meetplaatsen at random geselecteerd. Dit laat toe ook de afhankelijkheid tussen de meetplaatsen te incorporeren in de gesimuleerde data om zodoende erg realistische datasets te simuleren. Doordat we slechts 100 meetplaatsen beschouwen i.p.v. 200 of 500, dagen we onze B3 test eigenlijk nog meer uit. Ook hier hebben we 10000 simulaties uitgevoerd, en werd de B3 test steeds uitgevoerd op het 5% significantieniveau. De geschatte kans op een type I fout is 0.060. Dit is een klein beetje groter dan de vooropgestelde 0.05, maar binnen de huidige context is dit zeker aanvaardbaar. We besluiten dus dat ook Test B3 betrouwbare resultaten levert.

# Hoofdstuk 3

## Resultaten voor de afzonderlijke meetplaatsen

### 3.1 Inleiding

In dit hoofdstuk worden de resultaten gegeven van de statistische trenddetectie op individuele meetplaatsen en voor iedere variabele afzonderlijk. Voor een gegeven meetplaats en een gegeven variabele wordt voor ieder jaar de gemiddelde concentratie vergeleken met de gemiddelde concentratie in het referentiejaar 2005.

De statistische test (Test 1 uit Sectie 2.2.3) houdt rekening met een eventuele aanwezigheid van seizoenale variabiliteit. De ontwikkelde test houdt eveneens rekening met de zogenaamde *seriële correlatie* van de observaties over de tijd (i.e. observaties gemeten over de tijd zijn niet statistisch onafhankelijk). De test steunt niet op de veronderstelling van normaliteit.

Alle testen zijn tweezijdig uitgevoerd op het  $\alpha = 5\%$  significantieniveau. Dit betekent dat zowel neerwaartse als opwaartse trends gedetecteerd kunnen worden.

## 3.2 Resultaten van de trendanalyse met 2005 als referentiejaar

### 3.2.1 Gedetailleerde resultaten

Gezien de erg grote hoeveelheid meetplaatsen, zijn alle resultaten grafische weergegeven en opgenomen in afzonderlijke appendices (Appendices 1 t.e.m. 11). In deze sectie geven we slechts één figuur en één tabel als voorbeeld en ter verduidelijking van de interpretatie.

Figuur 3.1 toont de resultaten voor CZV voor meetplaats 871100. Voor ieder jaar wordt met een horizontale lijn het gemiddelde van de metingen aangegeven, alsook een foutenvlag die bepaald is door éénmaal de standaardafwijking van de observaties uit dat jaar. Dit geeft dus een idee over de variabiliteit van de observaties binnen het betreffende jaar. Merk evenwel op dat ook de seizoentaliteit hierin vervat zit. Bovenaan de figuur, wordt voor ieder jaar het aantal observaties vermeld waarop de statistische analyse betrekking heeft. Indien er voor een bepaald jaar een significant verschil met 2005 wordt gedetecteerd op het  $\alpha = 0.05$  significantieniveau, wordt dit onderaan met een \* weergegeven. Voor dit voorbeeld vinden we enkel voor 1995 en 1997 een significant verschil met 2005 m.b.t. de gemiddelde CZV concentratie. De  $p$ -waarden worden weergegeven in Tabel 3.1. Het is belangrijk om hier op te merken dat we hier nog NIET gecorrigeerd hebben voor multipliciteit. In de volgende paragraaf gaan we iets meer in op de interpretatie m.b.t. de multipliciteit.

De interpretatie van de “significante” resultaten voor 1995 en 1997 is afhankelijk van het standpunt dat ingenomen wordt m.b.t. de multipliciteit. We stellen drie scenario's voor:

1. het naïeve foutieve scenario:  
Uit de resultaten besluiten we op het 5% significantieniveau dat zowel in 1995 als in 1997 de gemiddelde CZV concentratie significant groter was dan in 2005.  
Het probleem met deze interpretatie is dat de kans dat minstens één van beide gedetecteerde trends (1995 en 1997) vals-positief is, groter is dan het vooropgeste significantieniveau van 5%.
2. een correct scenario waarbij multipliciteit geen rol speelt:

Stel dat we vooraf (i.e. voor het bekijken van de statistische analysere-sultaten) als doelstelling gesteld hadden dat we enkel één jaar wensen te vergelijken met het referentiejaar 2005. Stel bv. dat dit het jaar 1995 was. Dan mag je op basis van de resultaten correct besluiten op het 5% significantieniveau dat er een significante grotere CZV concentratie was in 1995 t.o.v. 2005. De  $p$ -waarde voor deze conclusie mag uit Tabel 3.1 overgenomen worden:  $p = 0.0278$ .

3. een correct scenario met correctie voor multipliciteit:

Stel dat we vooraf enkel vooropstellen dat we alle jaren sinds 1995 wensen te vergelijken met het referentiejaar 2005. Voor dit voorbeeld betekent dit dat er data van 8 jaren (voor 2005) beschikbaar zijn, en dat we dus 8 individuele statistische hypothesetesten dienen uit te voeren alvorens een algemeen besluit te vormen. Indien we dit correct wensen te doen, moeten we corrigeren voor multipliciteit. Een eenvoudige methode is de Bonferroni methode. Deze bestaat erin de individuele  $p$ -waarden te vermenigvuldigen met het aantal hypothese testen (8). Uiteraard volstaat het om deze vermenigvuldiging te doen op de individuele niet-gecorrigeerde  $p$ -waarden kleiner dan 5%. Hier dus voor 1995 en 1997: we vinden  $p = 8 \times 0.0278 = 0.2224$  en  $p = 8 \times 0.0071 = 0.0568$  voor respectievelijk 1995 en 1997. De gecorrigeerde  $p$ -waarden zijn alle groter dan 5% en dus moeten we besluiten dat voor geen enkel jaar er voldoende informatie in de data aanwezig is om te concluderen dat er een gemiddelde verschillende CZV concentratie is t.o.v. 2005.

De conclusie wordt hier dus duidelijk mede bepaald door het aantal hypothesetesten en het is best om het aantal zo laag mogelijk te houden. Indien er bv. gewenst is om enkel de tweede helft van de jaren '90 te bestuderen (1995-1999), dan zouden de gecorrigeerde  $p$ -waarden worden:  $p = 5 \times 0.0278 = 0.139$  en  $p = 5 \times 0.0071 = 0.0355$  voor respectievelijk 1995 en 1997, en zouden we toch een significant resultaat voor 1997 gevonden hebben.

Het is van uitzonderlijk groot belang om vooraf te bepalen welke jaren bestudeerd zullen worden.

Uit bovenstaande discussie volgt dat de rol van multipliciteit bij de interpretatie afhangt van de vraagstelling. Deze vraagstelling moet steeds vooraf duidelijk geformuleerd te worden (in het bijzonder: van hoeveel jaren willen we de resultaten simultaan interpreteren).

Tabel 3.1: Resultaten ( $p$ -waarden) voor meetplaats 871100 voor CZV. Referentiejaar is 2005.

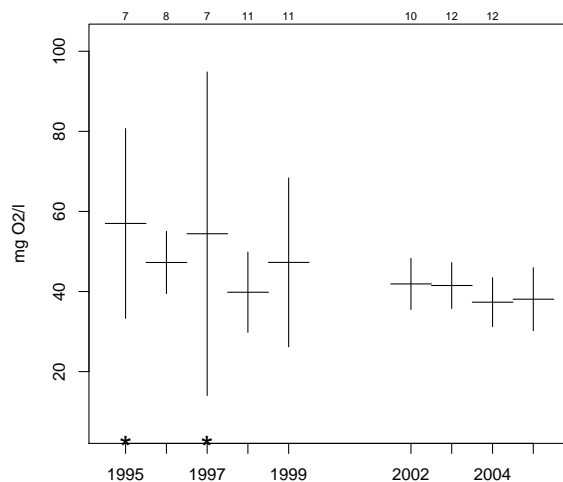
1995	1996	1997	1998	1999	2002	2003	2004
0.0278	0.1123	0.0071	0.8656	0.1486	0.5407	0.5757	0.9331

In het voorbeeld vinden we op het 5% significantieniveau voor de jaren 1996 en 1998-2004 geen significante verschillen met het referentiejaar 2005. Dit betekent dat de individuele  $p$ -waarden hiervoor groter zijn dan 5% (zie Tabel 3.1). Op deze niet-significante verschillen heeft het multipliciteitsprobleem geen betrekking. Immers, een correctie voor multipliciteit zou de individuele  $p$ -waarden nog vergroten, maar dan blijven ze uiteraard groter dan 5% en dus niet-significant.

Uit bovenstaande discussie kan misschien foutief geconcludeerd worden dat de correctiemethode tot tegengestelde conclusies kan leiden: soms wel en soms niet significant, afhankelijk van het aantal jaren dat bestudeerd wordt. Dit is evenwel zeker geen inconsistentie. Om dit te begrijpen, moeten we ons realiseren dat de conclusie van “niet significant” en van “wel significant” niet even sterk zijn. Besluiten dat gemiddelden significant zijn, is een “sterke conclusie” omdat de statistische methodologie van hypothesetesten een bescherming biedt tegen vals-positieve conclusies, maar het besluiten dat er geen significant verschil is, is een “zwakke conclusie” omdat de statistische theorie geen bescherming biedt tegen vals-negatieve conclusies. Dus door het aantal hypothesetesten te vergroten, moeten volgens de Bonferroni methode de  $p$ -waarden met een grotere waarde vermenigvuldigd worden, en zullen dus waarschijnlijk minder significante verschillen kunnen geconcludeerd worden. Voor de jaren waarvoor de  $p$ -waarde door de vermenigvuldiging plots groter dan 5% geworden zijn, wordt nu een niet-significantie besloten, maar aangezien dit een zwakke conclusie is, betekent dit niet dat er bewezen is dat er geen verschil in gemiddelde CZV concentratie is. Het enige dat je hiervoor kan concluderen is dat er onvoldoende informatie in de data is om het tegengestelde te bewijzen.

Zoals eerder vermeld, is er op de stuurgroepvergadering van 21 februari 2007 besloten om alle resultaten te presenteren zonder multipliciteitscorrectie. Op deze wijze kunnen de resultaten op meerdere manieren aangewend worden.

Het is tevens belangrijk om steeds naar de schaal/eenheden op de verticale as te kijken. Figuur 3.2 geeft twee voorbeelden. Het significante verschil in gemiddelde ammonium concentratie tussen 1990 en 2005 op meetplaats 689000 betreft

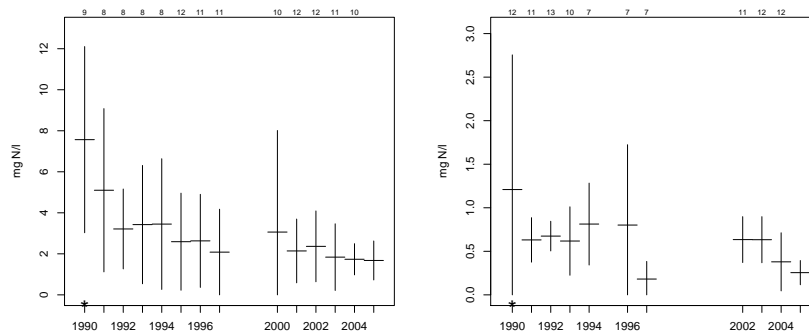


Figuur 3.1: Resultaten voor meetplaats 871100 voor CZV. Referentiejaar is 2005.

minder dan 1 mg N/l, daar waar voor meetplaats 7400 het significante verschil tussen 1990 en 2005 ongeveer 6 mg N/l betreft. Daarom is het belangrijk om steeds naar de eenheden op de verticale as te kijken. Bovendien illustreert dit dat een significant verschil niet steeds een relevant verschil impliceert. Een statistische methode houdt immers geen rekening met de biologische kennis. Als er maar voldoende observaties zijn, dan zal de test waarschijnlijk een significant resultaat geven, maar dit is daarom geen relevant resultaat!

De statistische methoden voorgesteld in dit project zouden aangepast kunnen worden zodat er slechts tot significantie besloten wordt indien het geobserveerde verschil in gemiddelde groter is dan een voorgesteld kleinste relevante verschil. Een dergelijke aanpak zou bovendien leiden tot een verhoging van de power. Dit valt echter buiten de doelstellingen van dit project.

Ten slotte nog een opmerking over de foutenvlaggen (+/- standaardafwijking van de observaties) die op de figuren weergegeven zijn. Deze geven een beeld van de variabiliteit van de observaties binnen één jaar, inclusief de variabiliteit ten gevolge van de seizoensaliteit. De statistische hypothesetesten houden nog rekening met de seriële afhankelijkheid. Om deze reden zou het gevoelsmatig soms

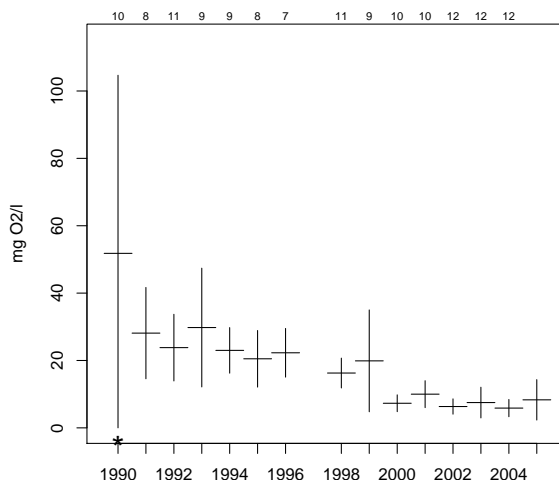


Figuur 3.2: Resultaten voor meetplaatsen 7400 (links) en 689000 (rechts) voor NH<sub>4</sub>. Referentiejaar is 2005.

tot tegenstellingen tussen hypothesetesten en foutenvlaggen kunnen komen. In dergelijke gevallen zijn de  $p$ -waarden formeel correct te interpreteren. De foutenvlaggen worden dus enkel weergegeven als een ondersteunen bij de exploratie van de data. Een voorbeeld hiervan wordt geïllustreerd in Figuur 3.3. De data uit 1990 vertonen een heel erg grote variabiliteit; dit is duidelijk te zien aan de hand van de foutenvlag. En toch is dit het enige jaar waarvoor de gemiddelde BZV verschilt met deze uit 2005. Het is de statistische test dat de juiste afweging maakt tussen variabiliteit en verschil in gemiddelde concentratie.

### 3.2.2 Bondig overzicht van de resultaten per bekken

De resultaten zijn samengevat in figuren. Voor iedere combinatie van variabele en bekken worden de resultaten in één figuur samengevat. Deze worden gegeven in Appendix 13. Een voorbeeld wordt gegeven in Figuur 3.4. Deze figuur toont voor ieder jaar het percentage meetplaatsen uit dit bekken dat een verschil vertoont met het referentiejaar 2005 (licht grijze balkjes). Er wordt een onderscheid gemaakt tussen opwaartse en neerwaartse trends. Opwaarts betekent dat de NH<sub>4</sub> concentratie in het betreffende jaar kleiner is dan in 2005, m.a.w. er is een chronologische opwaartse trend. Binnen de licht grijze balkjes worden a.d.h.v. donker grijze balkjes aangegeven wat het percentage meetplaatsen is dat een significant verschil vertoont met het referentiejaar 2005 (significantieniveau van 5%,



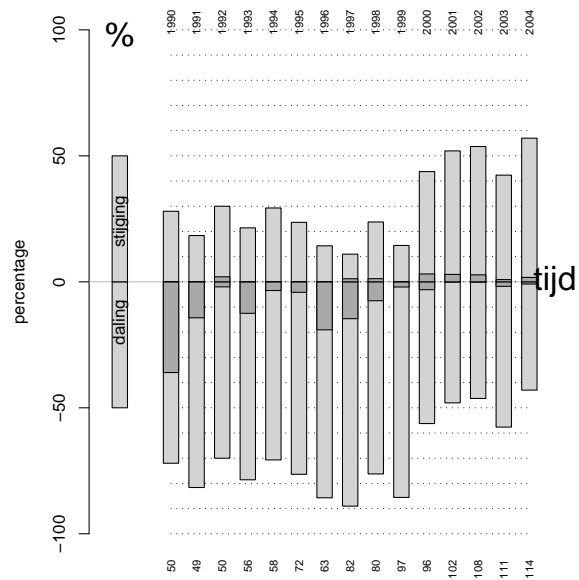
Figuur 3.3: Resultaten voor meetplaats 28000 voor BZV. Referentiejaar is 2005.

zonder correctie voor Multipliciteit). Onderaan de figuur wordt het totaal aantal meetplaatsen weergegeven. De figuur toont duidelijk dat er meer significante neerwaartse trends gedetecteerd zijn dan opwaartse trends. Voornamelijk in de beginjaren '90 werden veel neerwaartse trends gevonden. Bijvoorbeeld, in 1990 werd in vele van de 50 (72% van 50) meetplaatsen van het Beneden Schelde bekken een daling vastgesteld, en in 36% van deze 50 meetplaatsen was de daling bovendien significant op het 5% niveau (zonder multipliciteitscorrectie). Met "daling" wordt bedoeld dat de gemiddelde NH4 concentratie in 1990 groter was dan in 2005. In 2004, daarentegen, werd nog in slechts 43% van de 114 meetplaatsen van het Beneden Schelde bekken een daling van de gemiddelde NH4 concentratie geobserveerd, en slechts in ongeveer 1% van de meetplaatsen bleek dit significant te zijn. Van de 57% stijgende trends, was er ook maar een hele kleine fractie significant op het 5% niveau.

Merk op dat niet in ieder jaar er evenveel meetplaatsen gebruikt zijn in de statistische analyses. Dit heeft uiteraard te maken met de beschikbaarheid van de data die wijzigt over de jaren heen.

De resultaten werden niet gecorrigeerd voor multipliciteit.





Figuur 3.4: Samenvattende resultaten voor NH4 in het bekken van de Beneden Schelde. Referentiejaar is 2005.

In Hoofdstuk 5 worden deze figuren gegeven die de resultaten samenvatten voor alle Vlaamse meetplaatsen.

# Hoofdstuk 4

## Resultaten op aggregatieniveau

### 4.1 Inleiding

In dit hoofdstuk worden de resultaten gegeven van de statistische trenddetectie voor iedere variabele afzonderlijk, maar op verschillende aggregatieniveaus. We beschouwen de volgende aggregaties: bekken, saliniteit, categorie, kwaliteit, waterloop en stroming. In het bijzonder worden voor deze aggregaties de volgende aggregatieniveaus beschouwd:

- bekken:  
Beneden Schelde, Boven Schelde, Gentse kanalen, Brugse polders, Nete, Maas, Dender, Demer, Leie, Nete en Dijle-Zenne
- categorie:  
onbekend, waterloop eerste categorie, waterloop tweede categorie, waterloop derde categorie, en bevaarbare waterloop
- saliniteit:  
onbekend, zoet en brak
- kwaliteit:  
onbekend, viswater, basiskwaliteit, drink- en viswater, en zwemwater
- waterloop:

onbekend, natuurlijke waterloop, kunstmatige waterloop, en geïsoleerd water

- stroming:  
stromend, en stilstaand

Voor een gegeven aggregatieniveau en een gegeven variabele wordt voor ieder jaar de gemiddelde concentratie vergeleken met de gemiddelde concentratie in het referentiejaar 2005. Er wordt dus uitgemiddeld over meetplaatsen, en binnen de jaren ook over de seizoenen. De statistische methode houdt evenwel rekening met seizoenaliteit.

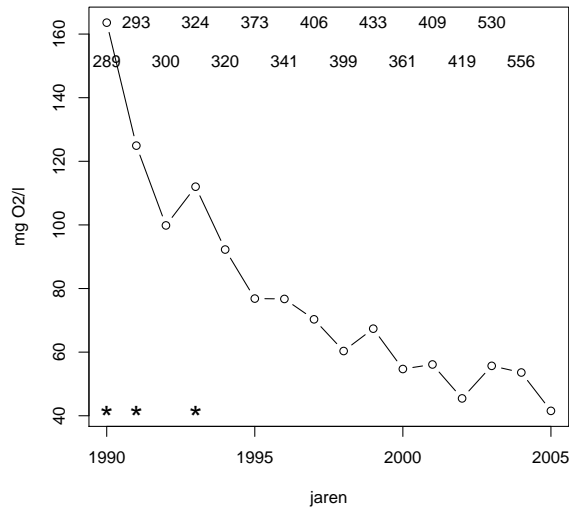
De statistische test (Test B2 uit Sectie 2.3.2) houdt rekening met een eventuele aanwezigheid van seizoenale variabiliteit. De ontwikkelde test houdt eveneens rekening met de zogenaamde *seriële correlatie* van de observaties over de tijd (i.e. observaties gemeten over de tijd zijn niet statistisch onafhankelijk), en met de *spatiale correlatie* (i.e. correlaties tussen de verschillende meetplaatsen binnen een aggregatieniveau). De test steunt niet op de veronderstelling van normaliteit.

Alle testen zijn tweezijdig uitgevoerd op het  $\alpha = 5\%$  significantieniveau. Dit betekent dat zowel neerwaartse als opwaartse trends gedetecteerd kunnen worden.

De meeste van de opmerkingen uit Sectie 3.2.1 zijn hier eveneens van toepassing.

## 4.2 Resultaten op de aggregatieniveaus van de trendanalyse met Test B2 en 2005 als referentiejaar

De volgende aggregatieniveaus worden beschouwd: bekken, categorie, kwaliteit, saliniteit, stroming en waterloop. Voor iedere combinatie van variabele en aggregatieniveau worden de resultaten samengevat in een figuur en een tabel. De volledige verzameling van figuren en tabellen is opgenomen in Appendix 14. Als voorbeeld tonen we hier in Figuur 4.1 de resultaten van CZV volgens de kwaliteit-saggregatie “basiskwaliteit”. Voor ieder jaar toont deze figuur de gemiddelde concentratie (uitgemiddeld over de meetplaatsen binnen de kwaliteitsaggregatie



Figuur 4.1: Samenvattende resultaten voor CZV in de Basiskwaliteit kwaliteitsaggregatie. Referentiejaar is 2005.

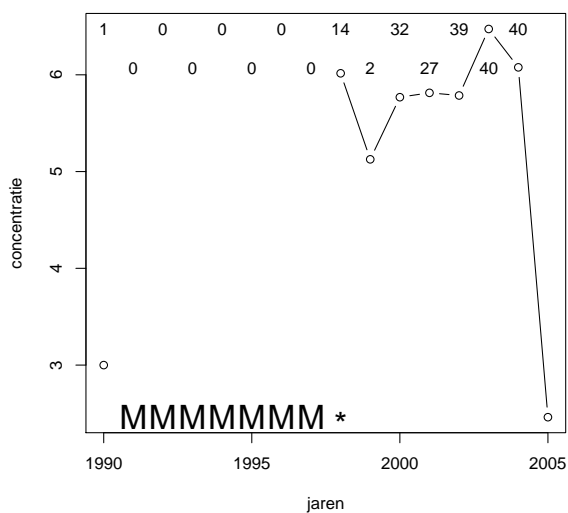
“basiskwaliteit”). Bovenaan de figuur wordt het aantal meetplaatsen vermeld dat gebruikt is voor de statistische analyse. Indien er op het  $\alpha = 5\%$  significantieniveau een significant verschil gedetecteerd is met test B2 t.o.v. het referentiejaar 2005, dan wordt dit onderaan met een \* aangeduid. In dit voorbeeld kunnen we dus besluiten voor de meetplaatsen binnen de kwaliteitsaggregatie “basiskwaliteit”, dat er in de jaren 1990, 1991 en 1993 een significant hogere gemiddelde CZV concentratie was t.o.v. 2005. De  $p$ -waarden worden getoond in Tabel 4.1. De resultaten werden niet gecorrigeerd voor multipliciteit. Het omgaan met multipliciteit is volledig analoog als voor bij de interpretatie van de trenddetectie op individuele meetplaatsen (zie Sectie 3.2.1 voor meer details).

Wanneer er onvoldoende data waren om de statistische test uit te voeren, wordt dit in de figuur met de letter “M” aangeduid (zie Figuur 4.2). In de tabel met de  $p$ -waarden wordt dit met “NA” aangeduid (zie Tabel 4.2).

In tegenstelling tot de presentatie van de resultaten op individueel meetplaatsniveau, is er hier geopteerd om geen foutenvlaggen te tonen op de figuur. De reden is dat de variabiliteit hier op verschillende niveaus plaatsvindt: variabiliteit

Tabel 4.1: Resultaten ( $p$ -waarden) van CZV voor Basiskwaliteit van aggregatie Kwaliteit. Referentiejaar is 2005.

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
0.0317	0.028	0.3099	0.009	0.0512	0.2547	0.0598	0.6099	0.431	0.3525
		2000	2001	2002	2003	2004	2005		
		0.7535	0.7018	0.8962	0.722	0.5673	1		



Figuur 4.2: Samenvattende resultaten voor BZV in de vijvers categorie-aggregatie. Referentiejaar is 2005.

Tabel 4.2: Resultaten ( $p$ -waarden) van BZV voor Vijver van aggregatie Categorie. Referentiejaar is 2005.

1990	1991	1992	1993	1994	1995	1996	1997	1998	
0.5491	NA	NA	NA	NA	NA	NA	NA	0.03	
		2000	2001	2002	2003	2004	2005		
		0.7705	0.4053	0.9112	0.5521	0.9193	0.8859	1	

ten gevolge van de seizoenen en verschillen tussen de meetplaatsen. De standaardafwijkingen zouden te veel gedomineerd worden door deze twee bronnen van variabiliteit, zonder evenwel een beeld te geven over het relatieve belang van de verschillende bronnen van variabiliteit. Uit ervaring weten we bovendien dat veel lezers de foutenvlaggen foutief interpreteren en ze het gevoel hebben dat ze in tegenspraak zijn met de resultaten van de statistische testen. Het is daarom veel correcter en informatiever om naar de  $p$ -waarden te kijken. Ten slotte merken we nog op dat deze foutenvlaggen ook geen rekening houden met seriële en spatiale correlaties, dewelke wel in rekening genomen worden bij de uitvoering van de statistische testen.

We hebben ook geopteerd om geen foutenvlaggen weer te geven in termen van de standaardafwijking op het geschatte gemiddelde over de meetplaatsen binnen een aggregatie/jaar. De reden is dezelfde als voorheen: vele ongeoeffende lezers zouden deze foutief kunnen interpreteren. Om dit te vermijden, werd gekozen om geen foutenvlaggen aan te brengen.

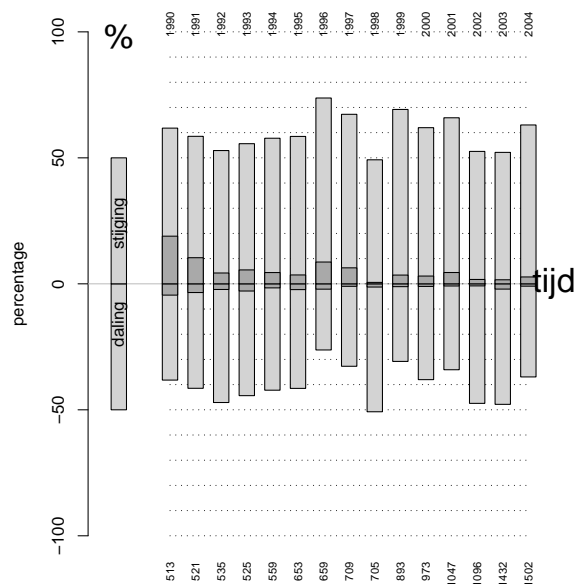
# Hoofdstuk 5

## Resultaten voor heel Vlaanderen

In dit hoofdstuk worden de overzichtsfiguren gegeven van de individuele trendanalyses op meetplaatsniveau voor iedere variabele, maar samengevat voor heel Vlaanderen. De figuren 5.1 t.e.m. 5.13 moeten op dezelfde wijze geïnterpreteerd worden als deze uit Sectie 3.2.2. Ook de resultaten van test B2 toegepast op de aggregatie van alle meetplaatsen van Vlaanderen worden hier weergegeven. Figuren 5.2 t.e.m. 5.14 en Tabellen 5.1 t.e.m. 5.7 tonen deze resultaten van de analyses voor de aggregatie over heel Vlaanderen. We geven en bespreken de resultaten hier voor iedere variabele afzonderlijk. Bij de interpretatie passen we op de resultaten van de B2 test een Bonferroni multipliciteitscorrectie toe. Dit houdt in dat de  $p$ -waarden uit de tabellen 5.1 t.e.m. 5.7 met 15 vermenigvuldigd worden omdat er 15 jaren vergeleken worden met 2005. Dit is dus een hele strenge analyse!

### Zuurstof

Voor zuurstof concluderen we het volgende uit Figuur 5.1. We zien voor ieder jaar relatief meer meetplaatsen waarvoor een significante stijging van de zuurstofconcentratie gedetecteerd is (steeds naar 2005 toe). Het aantal meetplaatsen is echter eerder beperkt (steeds minder dan 20% van de meetplaatsen). Onafhankelijk van de significantie (dus op basis van de licht grijze balkjes), moeten we ook vaststellen dat er ieder jaar iets meer meetplaatsen zijn waarvoor een stijging van de zuurstofconcentratie waargenomen werd, dan dat er meetplaat-

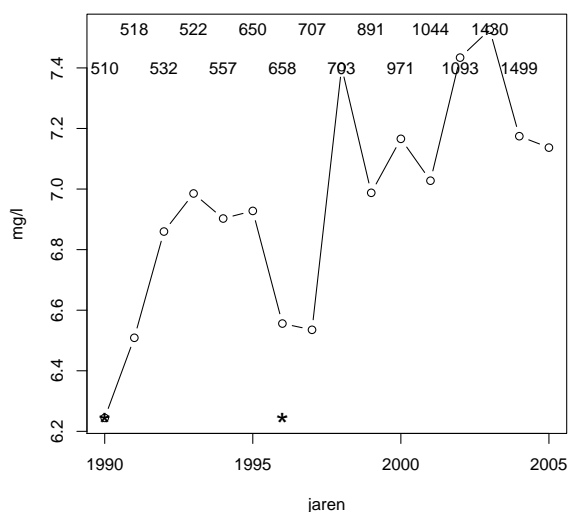


Figuur 5.1: Stijgende en dalende trends (% meetplaatsen) voor O<sub>2</sub> in heel Vlaanderen. Referentiejaar is 2005. Donkere grijze balken geven het percentage meetplaatsen waarvoor een significante trend (alpha=5%) gedetecteerd is. Onderaan wordt het totaal aantal meetplaatsen gegeven.

sen zijn waarvoor een daling geobserveerd werd. Het verschil is echter niet erg groot. Samengevat, lijken deze analyses de indiceren dat er in ieder jaar tussen 1990-2004 relatief meer meetplaatsen zijn waarvoor de gemiddelde zuurstofconcentratie lager was dan in 2005.

Deze conclusie wordt bevestigd door Figuur 5.2 en Tabel 5.1: niettegenstaande er enkel in 1990 en 1996 een significante lagere gemiddelde zuurstofconcentratie t.o.v. 2005 gedetecteerd is, toont de figuur een overwegend stijgende trend tot midden jaren 1990. Na correctie voor multipliciteit, blijft allen het significante resultaat van 1990 bestaan ( $15 \times 0.0023 = 0.0345 < 0.05$ ). De jongste jaren lijkt de gemiddelde zuurstofconcentratie gestagneerd.





Figuur 5.2: Resultaten van O2 voor heel Vlaanderen. Referentiejaar is 2005 (\*: significant verschil t.o.v. 2005 op het 5% significantieniveau). Bovenaan staat het aantal meetplaatsen vermeld waarop de statistische analyse betrekking heeft.

Tabel 5.1: Resultaten ( $p$ -waarden) van O2 voor heel Vlaanderen. Referentiejaar is 2005 .

1990	1991	1992	1993	1994	1995	1996	1997	1998
0.0023	0.1569	0.8559	0.1698	0.0589	0.2633	0.0162	0.1582	0.9457
	1999	2000	2001	2002	2003	2004	2005	
	0.8045	0.305	0.1512	0.6565	0.7634	0.3601	1	

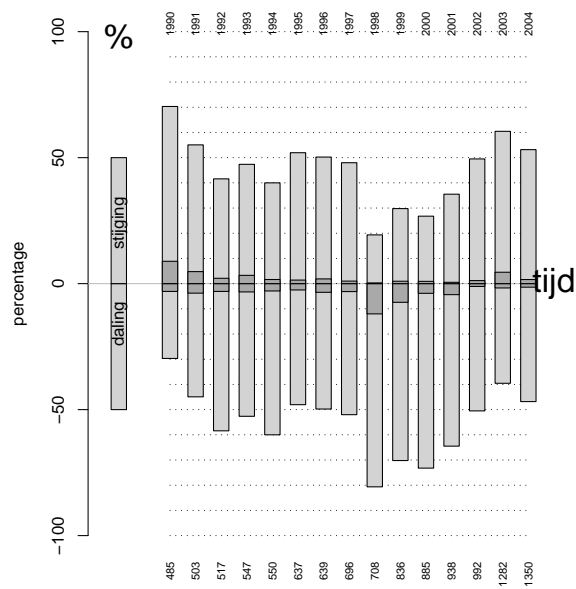
Tabel 5.2: Resultaten ( $p$ -waarden) van NO<sub>3</sub> voor heel Vlaanderen. Referentiejaar is 2005 .

1990	1991	1992	1993	1994	1995	1996	1997	1998
0.1078	0.6729	0.8034	0.9251	0.6967	0.7813	0.8501	0.4987	0.1067
	1999	2000	2001	2002	2003	2004	2005	
	0.1994	0.0706	0.341	0.9381	0.4945	0.8568	1	

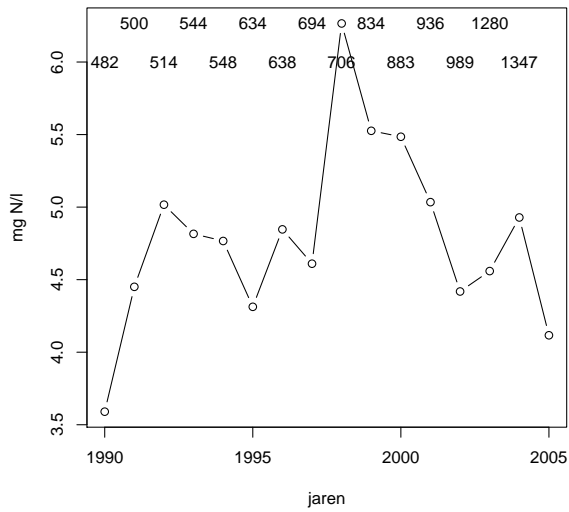
### Nitraat

Voor nitraat komen we tot de volgende conclusies uit Figuur 5.3. In 1990 waren er nog meer meetplaatsen waarvoor de gemiddelde nitraatconcentratie lager was dan in 2005. Van 1991 t.e.m. 1997 zijn er ongeveer evenveel meetplaatsen met een hogere als een lagere gemiddelde nitraatconcentratie in vergelijking met 2005. In 1998 waren er duidelijk meer meetplaatsen met een hogere gemiddelde nitraatconcentratie dan in 2005. Dezelfde vaststelling, maar dan minder uitgesproken, gelden ook voor 1999, 2000 en 2001. De volgende jaren lijken eerder stabiel te zijn t.o.v. 2005, i.e. ongeveer evenveel meetplaatsen met een kleinere als een grotere gemiddelde nitraatconcentratie dan in 2005. Het aantal meetplaatsen waarvoor significante verschillen met 2005 vastgesteld werden, is eerder klein (bijna steeds kleiner dan 10%).

De resultaten van de statistische testen over heel Vlaanderen (Figuur 5.4 en Tabel 5.2) geven voor geen enkel jaar een significant verschil t.o.v. 2005. De figuur toont verder dat er weinig verandering zit in de gemiddelde nitraatconcentratie over heel Vlaanderen. Tussen 1990 en 1998 lijkt er een stijging te zijn, en nadien terug een daling, maar deze vaststellingen zijn niet significant aangetoond. Niet-significantie kan o.a. een gevolg zijn van de grote variabiliteit tussen de meetplaatsen (uit Figuur 5.3 lezen we dat er ongeveer evenveel meetplaatsen zijn met stijgende als met dalende gemiddelde nitraatconcentraties).



Figuur 5.3: Stijgende en dalende trends (% meetplaatsen) voor NO3 in heel Vlaanderen. Referentiejaar is 2005 . Donkere grijze balken geven het percentage meetplaatsen waarvoor een significante trend (alpha=5%) gedetecteerd is. Onderaan wordt het totaal aantal meetplaatsen gegeven.



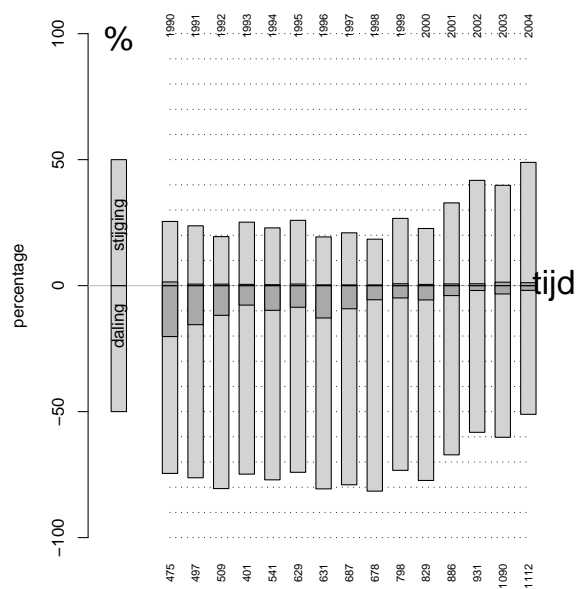
Figuur 5.4: Resultaten van NO<sub>3</sub> voor heel Vlaanderen. Referentiejaar is 2005 (\*: significant verschil t.o.v. 2005 op het 5% significantieniveau). Bovenaan staat het aantal meetplaatsen vermeld waarop de statistische analyse betrekking heeft.

Tabel 5.3: Resultaten ( $p$ -waarden) van NO<sub>2</sub> voor heel Vlaanderen. Referentiejaar is 2005 .

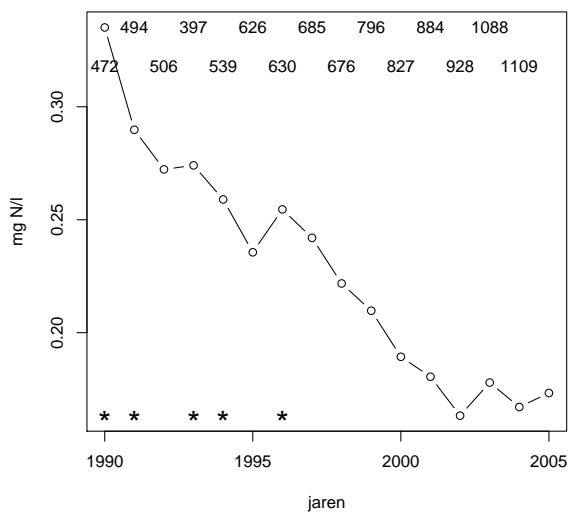
1990	1991	1992	1993	1994	1995	1996	1997	1998
0.0185	0.0112	0.2508	0.046	0.0283	0.1625	0.0144	0.2248	0.251
	1999	2000	2001	2002	2003	2004	2005	
	0.3132	0.0743	0.667	0.6958	0.6337	0.6319	1	

### Nitriet

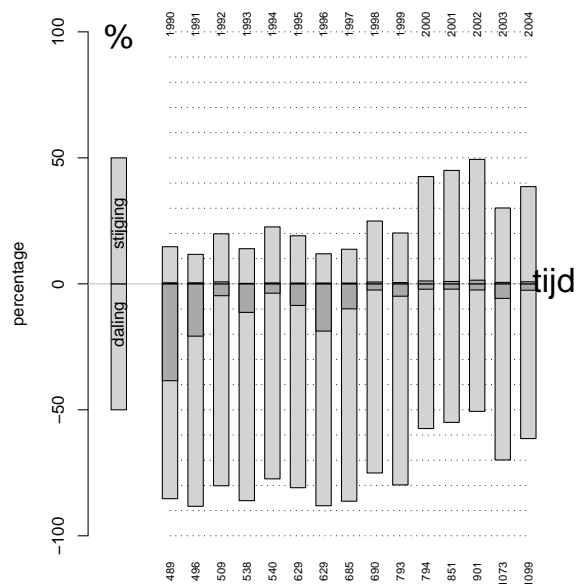
Op basis van Figuur 5.5 stellen we voor nitriet over de gehele periode vast dat er ieder jaar duidelijk meer meetplaatsen zijn die een hogere gemiddelde nitrietconcentratie hebben dan in 2005. Vooral in begin jaren 1990 was dit sterk aanwezig met tussen 10% – 20% van de meetplaatsen met een significant verschil. Maar ook onafhankelijk van significanties, moeten we ieder jaar vaststellen dat er meer meetplaatsen zijn met een daling van de gemiddelde nitrietconcentratie naar 2005 toe (70% – 80% van de meetplaatsen). Vanaf 2000 begint er een verschuiving op te treden, zodat in 2004 er ongeveer evenveel meetplaatsen met een lagere als een hogere gemiddelde nitrietconcentratie dan in 2005 geobserveerd werden. Deze conclusies worden bevestigd door de resultaten van de statistische test B2 over de volledige Vlaamse aggregatie. Figuur 5.6 toont duidelijk een daling gedurende de volledige jaren 1990, vanaf 2000 lijkt de nitrietconcentratie gestabiliseerd te zijn. De gemiddelde nitrietconcentratie in 2005 is significant lager dan deze uit een aantal jaren vóór 1997, maar na zware correctie voor de multiplicitéit van 15 jaren, verdwijnt deze significantie (zie  $p$ -waarden in Tabel 5.3). Samengevat toont deze analyse dat de gemiddelde nitrietconcentratie voornamelijk in de jaren 1990 een duidelijke daling vertoonde, maar dat deze daling gestabiliseerd is in de eerste helft van de jaren 2000.



Figuur 5.5: Stijgende en dalende trends (% meetplaatsen) voor NO2 in heel Vlaanderen. Referentiejaar is 2005 . Donkere grijze balken geven het percentage meetplaatsen waarvoor een significante trend (alpha=5%) gedetecteerd is. Onderaan wordt het totaal aantal meetplaatsen gegeven.



Figuur 5.6: Resultaten van NO<sub>2</sub> voor heel Vlaanderen. Referentiejaar is 2005 (\*: significant verschil t.o.v. 2005 op het 5% significantieniveau). Bovenaan staat het aantal meetplaatsen vermeld waarop de statistische analyse betrekking heeft.



Figuur 5.7: Stijgende en dalende trends (% meetplaatsen) voor NH4 in heel Vlaanderen. Referentiejaar is 2005. Donkere grijze balken geven het percentage meetplaatsen waarvoor een significante trend (alpha=5%) gedetecteerd is. Onderaan wordt het totaal aantal meetplaatsen gegeven.

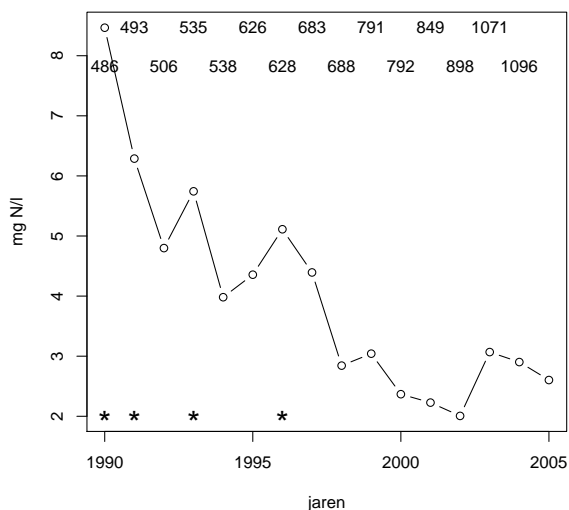
### Ammonium

Voor ammonium zijn de conclusies heel gelijkaardig aan deze van nitriet. Vooral de vergelijking van 1990 met 2005 is heel overtuigend: Figuur 5.7 toont dat in meer dan 80% van de meetplaatsen werd in 1990 een hogere gemiddelde ammoniumconcentratie geobserveerd dan in 2005, en in bijna 40% van de meetplaatsen was de trend significant.

Ook de resultaten van de B2 test (Figuur 5.8 en Tabel 5.4) tonen een daling de gemiddelde ammoniumconcentratie gedurende de jaren 1990, gevolgd door een nivellering. Tussen 1990 en 1996 zijn er enkele jaren met een significant hogere ammoniumconcentratie dan in 2005, maar na correctie van de multipliciteit verdwijnen de significanties.

Samengevat toont deze analyse dat de gemiddelde ammoniumconcentratie voornamelijk in de jaren 1990 een duidelijke daling vertoonde, maar dat deze daling gestabiliseerd is sinds 2000.

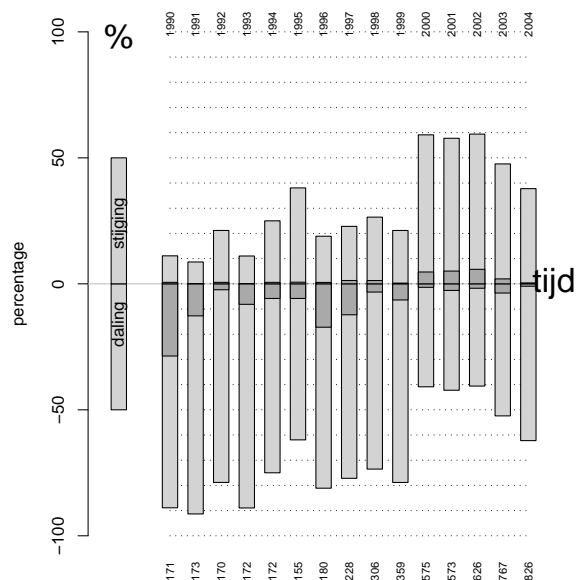




Figuur 5.8: Resultaten van NH4 voor heel Vlaanderen. Referentiejaar is 2005 (\*: significant verschil t.o.v. 2005 op het 5% significantieniveau). Bovenaan staat het aantal meetplaatsen vermeld waarop de statistische analyse betrekking heeft.

Tabel 5.4: Resultaten ( $p$ -waarden) van NH4 voor heel Vlaanderen. Referentiejaar is 2005 .

1990	1991	1992	1993	1994	1995	1996	1997	1998
0.0235	0.0223	0.6051	0.0372	0.2171	0.3816	0.0184	0.3813	0.5141
1999	2000	2001	2002	2003	2004	2005		
0.5032	0.7808	0.9563	0.8703	0.5175	0.8292	1		



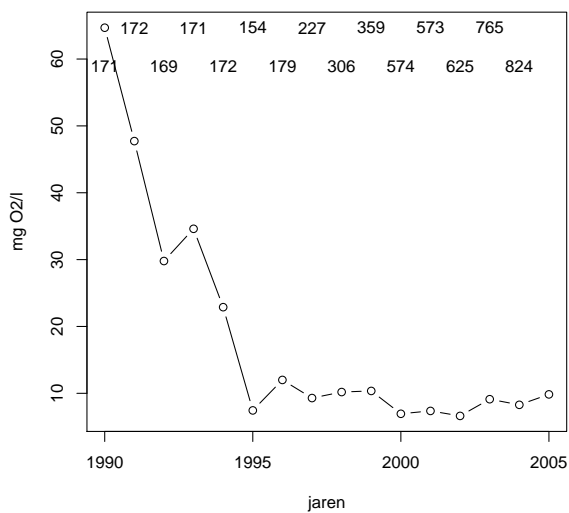
Figuur 5.9: Stijgende en dalende trends (% meetplaatsen) voor BZV in heel Vlaanderen. Referentiejaar is 2005. Donkere grijze balken geven het percentage meetplaatsen waarvoor een significante trend (alpha=5%) gedetecteerd is. Onderaan wordt het totaal aantal meetplaatsen gegeven.

### BZV

Van 1990 t.e.m. 1999 zijn de conclusies uit Figuur 5.9 voor BZV bijna indientiek aan deze voor nitriet en ammonium, maar vanaf 2000 moeten we vaststellen dat er terug meer meetplaatsen zijn waarvoor de BZV groter is dan in 2005. In 2004 waren er nagenoeg geen significante wijzigingen meer.

Figuur 5.10 toont een hele scherpe daling van de gemiddelde BZV gedurende de eerste helft van de jaren 1990, onmiddellijk gevolgd door een duidelijke stagnatie. Er werden echter geen significante verschillen op Vlaams niveau vastgesteld.

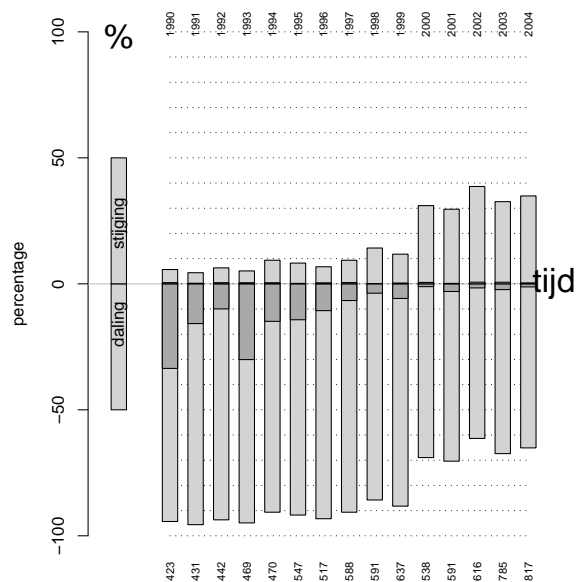
Samengevat kunnen we stellen dat vooral in de eerste helft van de jaren 1990 er een daling van de BZV plaatsgevonden heeft in vele meetplaatsen, maar de jongste jaren is er een evenwicht tussen neerwaartse en opwaartse trends zodat het Vlaams gemiddelde constant blijft.



Figuur 5.10: Resultaten van BZV voor heel Vlaanderen. Referentiejaar is 2005 (\*: significant verschil t.o.v. 2005 op het 5% significantieniveau). Bovenaan staat het aantal meetplaatsen vermeld waarop de statistische analyse betrekking heeft.

Tabel 5.5: Resultaten ( $p$ -waarden) van BZV voor heel Vlaanderen. Referentiejaar is 2005 .

1990	1991	1992	1993	1994	1995	1996	1997	1998
0.4154	0.5672	0.6618	0.2883	0.2702	0.9877	0.3806	0.6742	0.6703
1999	2000	2001	2002	2003	2004	2005		
0.3185	0.2773	0.8212	0.4386	0.8746	0.3151	1		



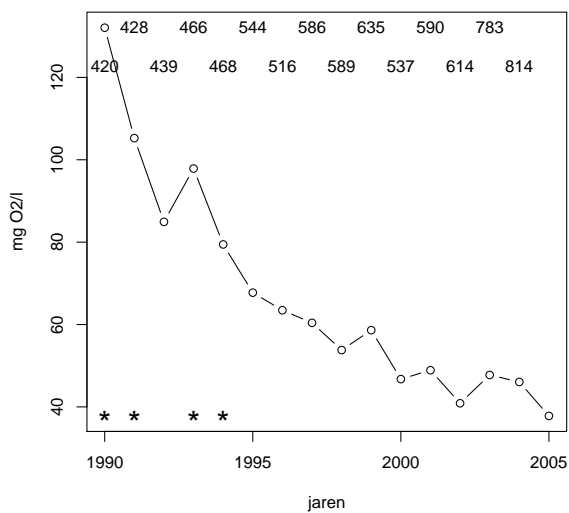
Figuur 5.11: Stijgende en dalende trends (% meetplaatsen) voor CZV in heel Vlaanderen. Referentiejaar is 2005. Donkere grijze balken geven het percentage meetplaatsen waarvoor een significante trend ( $\alpha=5\%$ ) gedetecteerd is. Onderaan wordt het totaal aantal meetplaatsen gegeven.

### CZV

De conclusies voor CZV zijn analoog aan deze voor nitriet. Figuur 5.11 toont dat er in het verleden steeds veel meetplaatsen waren waar een hogere gemiddelde CZV vastgesteld werd in vergelijking met 2005. Er zijn nagenoeg geen meetplaatsen waar in het verleden de gemiddelde CZV lager was dan in 2005. Onafhankelijk van de resultaten van de statistische testen, leert de figuur ons dat er sinds 2000 toch ook terug meer (ongeveer 30% van de meetplaatsen) meetplaatsen zijn waar de gemiddelde CZV gestegen is naar 2005 toe.

Uit Figuur 5.12 leren we dat er een duidelijke daling van de gemiddelde CZV geobserveerd wordt. In de eerste helft van de jaren 1990 waren de gemiddelde CZV concentratie significant hoger dan in 2005 (Tabel 5.6), maar de significantie verdwijnt na correct voor multiplicititeit.

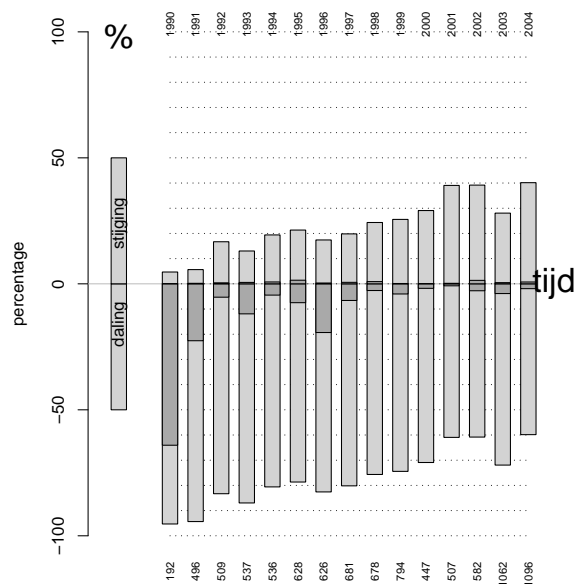
Samengevat zeggen we dat de gemiddelde CZV concentratie in vele meetplaatsen gedaald is.



Figuur 5.12: Resultaten van CZV voor heel Vlaanderen. Referentiejaar is 2005 (\*: significant verschil t.o.v. 2005 op het 5% significantieniveau). Bovenaan staat het aantal meetplaatsen vermeld waarop de statistische analyse betrekking heeft.

Tabel 5.6: Resultaten ( $p$ -waarden) van CZV voor heel Vlaanderen. Referentiejaar is 2005 .

1990	1991	1992	1993	1994	1995	1996	1997	1998
0.0377	0.0313	0.2537	0.0063	0.0414	0.2273	0.0695	0.637	0.4216
	1999	2000	2001	2002	2003	2004	2005	
	0.3181	0.6936	0.65	0.7837	0.6948	0.5974	1	

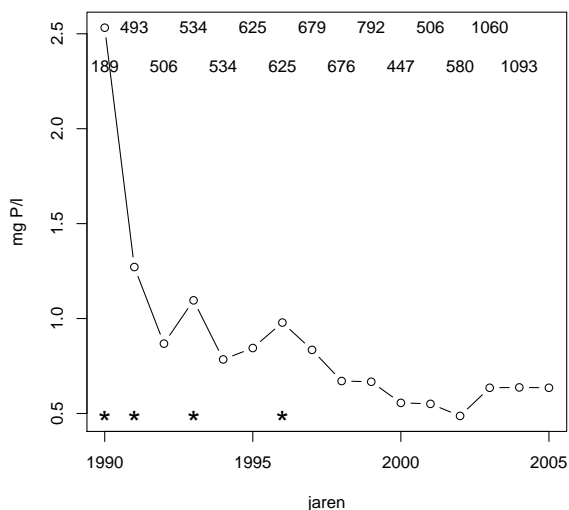


Figuur 5.13: Stijgende en dalende trends (% meetplaatsen) voor oPO4 in heel Vlaanderen. Referentiejaar is 2005. Donkere grijze balken geven het percentage meetplaatsen waarvoor een significante trend ( $\alpha=5\%$ ) gedetecteerd is. Onderaan wordt het totaal aantal meetplaatsen gegeven.

### Orthofosfaat

Figuur 5.13 toont een gelijkaardig beeld als Figuur 5.11 voor CZV. We zien nagenoeg geen meetplaatsen waarvoor in het verleden de gemiddelde orthofosfaatconcentratie significant lager was dan in 2005. In 1990 werden zelfs in meer dan 60% van de meetplaatsen een gemiddelde orthofosfaatconcentratie vastgesteld die significant hoger was dan in 2005. De jongste jaren is het aantal meetplaatsen met significante trends verminderd, en worden er zelfs tot in 40% van de meetplaatsen lagere gemiddelde orthofosfaatconcentratie geobserveerd dan in 2005 (niet-significant).

Figuur 5.14 toont dat de gemiddelde orthofosfaatconcentratie voornamelijk in de eerste helft van de jaren 1990 gedaald is, maar dan gestagneerd is. De significante verschillen met 2005 uit deze eerste helft van de jaren 1990 verdwijnen na correctie voor multipliciteit.



Figuur 5.14: Resultaten van oPO4 voor heel Vlaanderen. Referentiejaar is 2005 (\*: significant verschil t.o.v. 2005 op het 5% significantieniveau). Bovenaan staat het aantal meetplaatsen vermeld waarop de statistische analyse betrekking heeft.

Tabel 5.7: Resultaten ( $p$ -waarden) van oPO4 voor heel Vlaanderen. Referentiejaar is 2005 .

1990	1991	1992	1993	1994	1995	1996	1997	1998
0.0147	0.0091	0.4532	0.0396	0.1948	0.3861	0.0216	0.4326	0.631
	1999	2000	2001	2002	2003	2004	2005	
	0.9257	0.6045	0.822	0.8134	0.5331	0.8925	1	

# Hoofdstuk 6

## Conclusies

In dit project werden statistische testen ontwikkeld voor trenddetectie. In het bijzonder onderscheiden we twee niveaus voor trenddetectie:

1. op iedere individuele meetplaats, wordt ieder jaar vergeleken met het referentiejaar 2005. Hiervoor werden twee testen ontwikkeld en beschreven.
2. meetplaatsen worden vooraf geaggregeerd in bekkens of andere aggregaties. Voor ieder van deze aggregaties wordt eveneens een trendanalyse uitgevoerd waarbij ieder jaar met het referentiejaar 2005 wordt vergeleken. Dit betekent dat de conclusies hieruit betrekking hebben op de volledige aggregatie, i.p.v. op de individuele meetplaatsen. Hiervoor werden drie testen ontwikkeld en beschreven.

Alle statistische methoden hebben betrekking op gemiddelden.

Voor de analyse van de individuele meetplaatsen, werd één test (test 1) toegepast op alle data van het pakket basisparameters van het fysisch-chemisch meetnet (O<sub>2</sub>, NO<sub>3</sub>, NO<sub>2</sub>, NH<sub>4</sub>, BZV, CZV, oPO<sub>4</sub>). De resultaten zijn samengebracht in Appendices 1 t.e.m. 11. Aangezien dit een enorme hoeveelheid resultaten met zich meebrengt, werden de resultaten per bekken en per parameter samengevat in overzichtelijke figuren (Appendix 13).

Voor de analyses op de verschillende aggregatieniveaus (bekken, categorie, kwaliteit, saliniteit, stroming en waterloop) werd enkel test B2 toegepast. De resultaten zijn opgenomen in Appendix 14.



De statistische methoden kunnen overweg met onregelmatige steekproefname en met relatief beperkte aantallen observaties onder de detectielimiet. Appendix 12 geeft een overzicht van de hoeveelheden data boven en onder de detectielimiet. We denken dat voor de data van het pakket basisparameters zich weinig problemen voordoen. Van al deze parameters komt bij orthofosfaat het detectielimietprobleem het vaakst voor.

De ontwikkelde programmatuur in R is opgenomen in Appendix 15.

Ten slotte vatten we hier bondig de voornaamste conclusies voor heel Vlaanderen samen. Deze conclusies zijn gebaseerd op de resultaten uit Hoofdstuk 5. Voor de meeste variabelen werd ongeveer dezelfde conclusie bekomen: de gemiddelde concentraties van nitriet, ammonium, BZV, CZV en orthofosfaat daalden in vele meetplaatsen in de jaren 1990, maar vanaf eind jaren 1990 of 2000 werd een stagnatie vastgesteld. Hetzelfde fenomeen, maar iets minder uitgesproken, geldt ook voor zuurstof (maar hier gaat het over een stijging van de gemiddelde concentratie). De gemiddelde nitraatconcentratie vertoont geen duidelijke trend. Deze blijft meestal gesitueerd tussen 4 en 5 mg N/l (Vlaams gemiddelde), met enkel in 1990 een lagere concentratie, en in de jaren 1998, 1999 en 2000 een hogere gemiddelde concentratie, doch niet significant. Voornamelijk in 1990 stellen we vast dat er in een 10-tal % van de meetplaatsen een significant hogere gemiddelde nitraatconcentratie dan in 2005 gedetecteerd is.

# Referenties

Thas, O. en Ottoy, J.P. (1996) Gebruik van statistische analysemethoden voor doorgedreven interpretatie van de data uit de meetdatabank van de VMM. Eindverslag van TWO-project van de Vlaamse Milieumaatschappij.