

## Nota

**Datum:** 19/12/17  
**Aan:** VLM  
**Auteur:** TFR  
**Nazicht:** GLE  
**Documentref:** I/NO/11504/17.373/TFR

---

**Betreft : Uitgebreide samenvatting statistische analyse waterkwaliteit**

---

### **Inhoudstafel**

---

<b>0. INLEIDING .....</b>	<b>2</b>
<b>1. VERKENNING VAN DE BESCHIKBARE DATA.....</b>	<b>2</b>
<b>2. GLOBALE ANALYSE.....</b>	<b>5</b>
<b>3. GERICHTE ANALYSE.....</b>	<b>7</b>

## 0. INLEIDING

Met het vijfde mestactieprogramma 2015-2018 (MAP5) in uitvoering van de Nitraatrichtlijn, wordt ingezet op een versterkte gebiedsgerichte aanpak. In de voorbije 15 jaar is er weliswaar een duidelijke verbetering gerealiseerd van de oppervlakte- en grondwaterkwaliteit in Vlaanderen, maar er blijven grote regionale verschillen. Daarom wenst de Vlaamse Overheid de achterliggende oorzaken verder te onderzoeken.

Deze studie heeft als doelstelling het verwerven van inzicht in de oorzaken van de goede of slechte waterkwaliteit in bepaalde gebieden, door het uitvoeren van een statistisch onderbouwde analyse van beschikbare data.

De opdracht bestaat uit verschillende onderdelen:

- Onderdeel 1: Verkenning van de beschikbare data
- Onderdeel 2: Globale statistische analyse van beschikbare data in Vlaanderen
- Onderdeel 3: Gerichtte statistische analyse van beschikbare data
- Onderdeel 4: Aanbevelingen voor implementatie van de onderzoeksresultaten

De eerste drie onderdelen hadden tot doel om de beschikbare, ruwe data te vertalen naar bruikbare informatie, die inzichten opleveren in de oorzaken van goede of slechte waterkwaliteit. Deze inzichten vormden het uitgangspunt voor onderdeel 4 waar ze werden voorgelegd aan een groep experts tijdens een uitgebreide workshop. Op basis daarvan werden enkele concrete aanbevelingen geformuleerd.

## 1. VERKENNING VAN DE BESCHIKBARE DATA

De gegevens die in deze studie werden gebruikt zijn afkomstig van verschillende instanties waaronder Vlaamse Milieumaatschappij (VMM), Vlaamse Landmaatschappij (VLM) alsook uit openbaar beschikbare databanken waaronder waterinfo.be en Geopunt Vlaanderen.

De beschikbare data werd opgesplitst in responsvariabelen en predictoren. Responsvariabelen zijn variabelen die betrekking hebben op de waterkwaliteit. Er zijn 5 verschillende responsvariabelen opgenomen in deze studie:

- Concentratie nitraat in het oppervlaktewater [mg N/l]
- Concentratie orthofosfaat in het oppervlaktewater [mg P/l]
- Concentratie nitraat in het grondwater [mg NO<sub>3</sub><sup>-</sup>/l]
- Concentratie fosfaat in het grondwater [mg PO<sub>4</sub><sup>3-</sup>/l]
- Nitraatresidu

De metingen in het oppervlaktewater zijn afkomstig van het MAP-meetnet. De metingen in het grondwatermeetnet zijn afkomstig van het freatisch grondwatermeetnet. Zowel het MAP-meetnet, alsook het freatische grondwatermeetnet, is voornamelijk gelokaliseerd in landbouwgebied.

Het nitraatresidu is de hoeveelheid nitraatstikstof per ha in de bovenste 90 cm van een landbouwperceel, gemeten in de periode van 1 oktober tot en met 15 november. Gewassen

nemen stikstof op in de vorm van nitraat om te groeien. De nitraten die niet opgenomen worden door de gewassen, blijven op het einde van het groeiseizoen achter in de bodem als residu, vandaar de term 'nitraatresidu'. Omwille van het verband tussen het nitraatresidu en het risico op uitspoeling van nitraten naar het oppervlakte- en grondwater tijdens de winter, wordt het nitraatresidu opgevolgd.

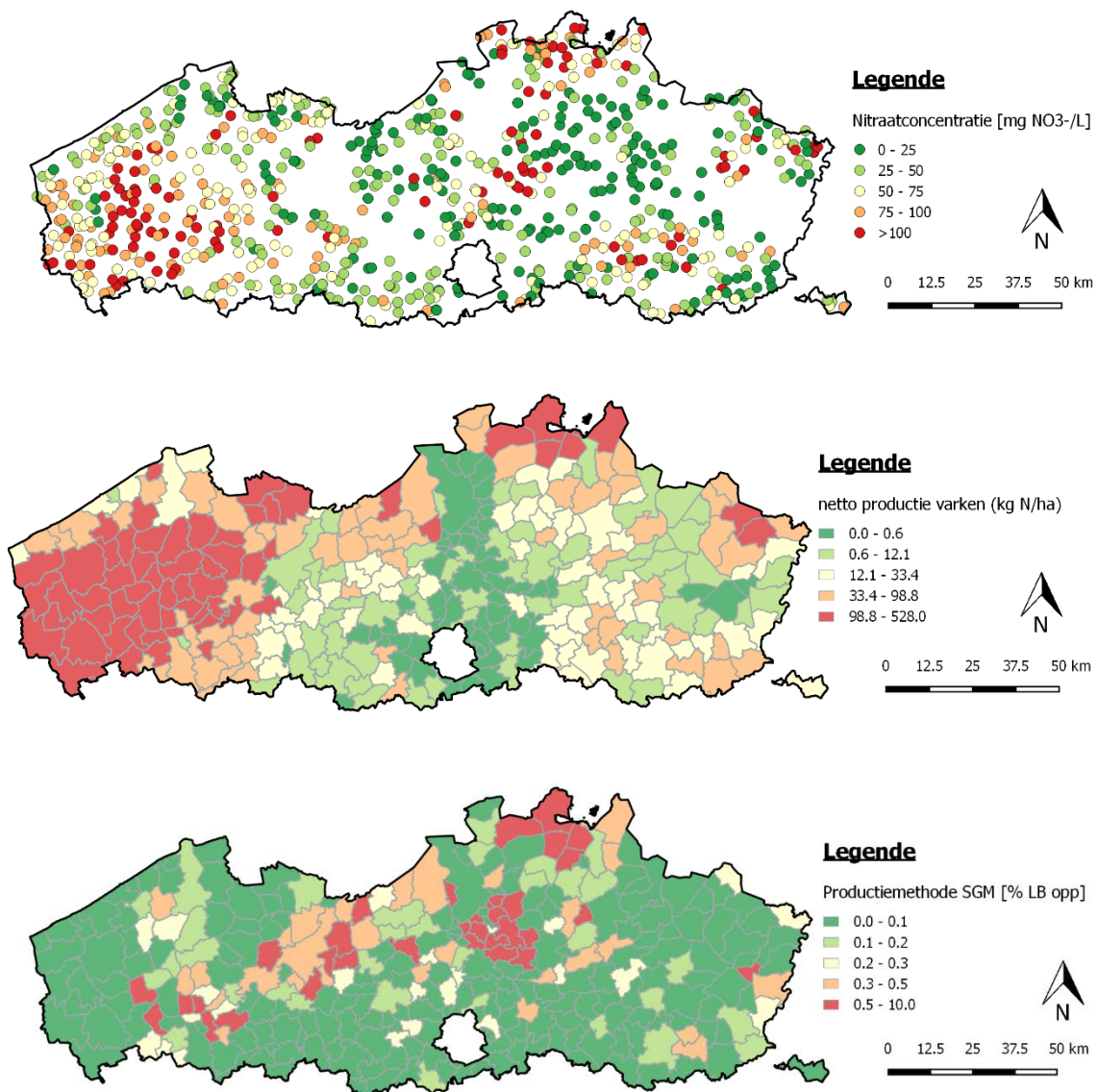
Predictoren zijn variabelen die mogelijks een invloed kunnen uitoefenen op de gemeten nitraat en fosfaatconcentraties en dus bijgevolg gebruikt kunnen worden in de statistische modellen als voorspellende variabelen. De volgende predictoren (of variabelen) werden o.a. opgenomen in deze studie:

- Landbouw en mestgegevens zoals geregistreerd in de mestbank
- Neerslag en afvoer
- Diepte grondwatertafel
- Stikstofdepositie
- Procesfactor oppervlaktewater zoals bepaald in de procesfactorstudie door middel van een voorspellend regressiemodel op basis van de dominante textuur van het afstroomgebied en de redoxpotentiaal van de onderliggende aquifer. De procesfactor wordt bepaald als de verhouding tussen de gemiddelde nitraatconcentratie over de winterperiode onderaan de wortelzone en de gemiddelde concentratie in het oppervlaktewater.

Alle relevante gegevens voor deze studie zijn verzameld in één overkoepelende meta-databank, waarin de gegevens gekoppeld worden. Deze koppeling is mogelijk op verschillende ruimtelijke aggregatieniveaus. Zo kan bv. voor ieder meetpunt of peilbuis het afstroom- of intrekgebied bepaald worden waarna alle predictoren gesommeerd kunnen worden over dat afstroomgebied. Op die manier kan dan bv. een meting rechtstreeks gekoppeld worden aan de hoeveelheid dierlijke mest of grasland in het afstroomgebied. Dat is het meest gedetailleerde ruimtelijke niveau maar hiervoor dienen wel enkele vereenvoudigingen en aannames gemaakt te worden. Een andere mogelijkheid is om deze koppeling uit te voeren op een hoger niveau zoals bv. de gemeente. Hierbij wordt de gemiddelde maximale concentratie van alle meetpunten in een gemeente gekoppeld aan bv. de hoeveelheid grasland per gemeente of de totale hoeveelheid dierlijke mest van alle exploitaties in de gemeente.

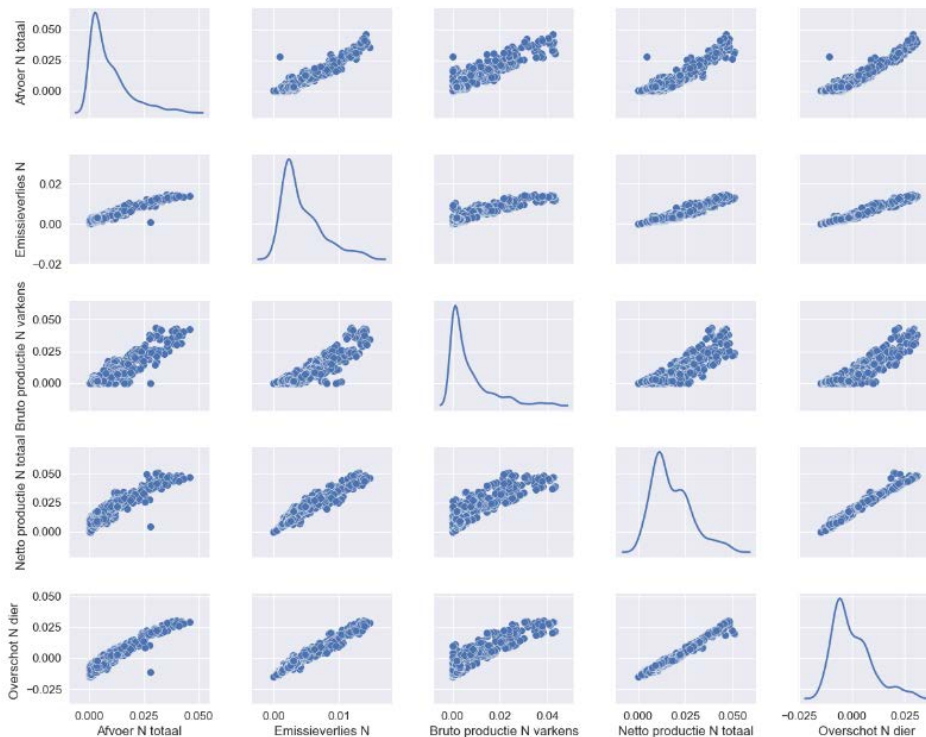
Op basis van een univariate data analyse zijn de belangrijkste eigenschappen zoals trends en ruimtelijke en temporele variatie van de beschouwde variabelen onderzocht. Zo werd voor de jaarlijkse maximale concentraties van nitraat bij 40 % (oppervlaktewater) en 44 % (grondwater) van de meetpunten een trendbreuk vastgesteld in de periode 2007-2014 en vertoont 74 % (oppervlaktewater) en 67 % (grondwater) een dalende trend. Voor fosfaat is in 32 % (oppervlaktewater) en 43 % (grondwater) van de meetpunten een trendbreuk vastgesteld voor de jaarlijkse maxima van de gemeten concentraties en vertoont 46 % (oppervlaktewater) en 40 % (grondwater) een dalende trend. Door middel van een combinatie van de grootteorde en trend zijn interessante meetlocaties (hoge concentraties en sterk stijgende trend) afgeleid. Het opnemen van een trendbreuk is van belang om een correcte inschatting te krijgen van de recente evoluties in zowel nitraat- als fosfaatconcentraties.

Uit een vergelijking van de ruimtelijke spreiding van de variabelen zijn reeds de eerste indicaties van onderlinge verbanden merkbaar.



*Figuur 1-1: Ruimtelijke spreiding van maximaal gemeten nitraatconcentratie (boven), nettoproductie stikstof door varkens (midden) en serres met teelt op groeimedium (onder)*

Deze correlaties zijn verder onderzocht in een bivariate data analyse, waarin telkens de overeenstemming of correlatie tussen twee variabelen werd onderzocht. Hieruit bleek dat er voor oppervlaktewater veel predictoren zijn die een significante correlatie vertonen met de gemeten nitraat- en fosfaatconcentraties. Eveneens kwam uit de bivariate analyse naar voor dat veel predictoren onderling ook sterk gecorreleerd zijn. Hiermee dient rekening gehouden te worden bij het opstellen van een globaal statistisch model aangezien deze collineariteit de robuustheid van het resulterende model kan verminderen. Voor de meetlocaties van het freatisch grondwatermeetnet zijn deze correlaties minder uitgesproken. Op basis van de volledige correlatiematrixes kan geconcludeerd worden dat globaal genomen de sterkste correlatie waarneembaar is door gebruik te maken van de jaarlijkse maxima van de gemeten concentraties. In het verdere verloop van de studie werd daarom gewerkt met deze jaarlijkse maximale waarden.



Figuur 1-2: Visuele weergave van correlatie tussen verschillende exploitatieparameters onderling.

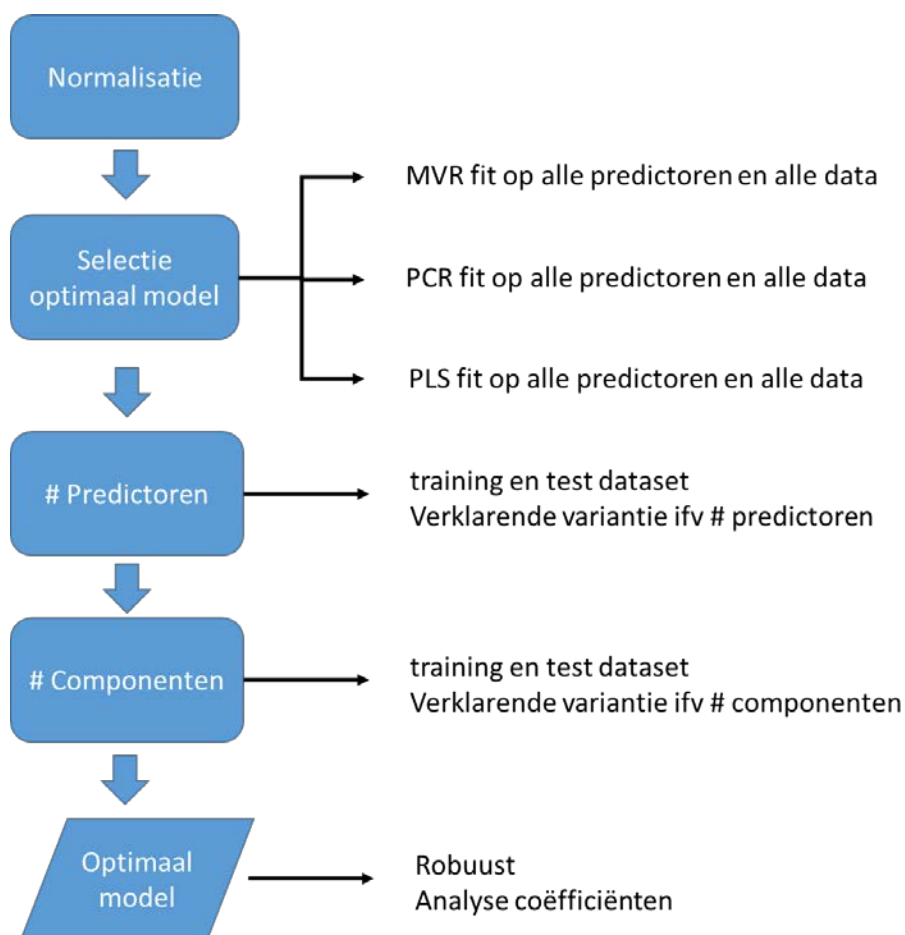
## 2. GLOBALE ANALYSE

Op basis van de gegevens en inzichten verzameld in de verkennende analyse werden in dit onderdeel globale statistische modellen opgesteld voor Vlaanderen. Door analyse van de resulterende modelstructuur werd getracht inzicht te verkrijgen in de oorzaken van goede of slechte waterkwaliteit. Er zijn modellen opgesteld voor verschillende responsvariabelen en voor verschillende ruimtelijke aggregatieniveaus.

Er zijn initieel 5 verschillende statistische modellen onderzocht waarvan er twee (Non-parametric causal random forest en Multispatial convergent cross mapping) niet volledig zijn uitgewerkt op ieder niveau omdat niet voldaan werd aan alle randvoorwaarden. Op basis van een analyse van de verklarende variantie in functie van het aantal predictoren of componenten werd het Partial Least Square regressie model (PLS) geselecteerd als meest optimale model. PLS is een regressie analyse gebaseerd op een principaal component analyse (PCA). Deze data mining techniek houdt in dat er correlaties worden gezocht tussen de respons en lineaire combinaties van de predictoren, de componenten. Hiertoe wordt eerst een nieuwe predictor-ruimte gezocht zodanig dat de predictoren onderling minimaal gecorreleerd zijn, met de extra voorwaarde dat de predictieve kracht van het model ten opzichte van de responsvariabele geoptimaliseerd wordt. PLS draagt hierbij zorg dat de covariantie van de componenten met de responsvariabele gemaximaliseerd wordt. De predictoren worden geprojecteerd naar een nieuwe predictor-ruimte, bestaande uit componenten die elk een lineaire combinatie vormen van de oorspronkelijke predictoren. Vervolgens wordt nagegaan in hoeverre deze

componenten de variatie van de respons kunnen verklaren. Om een robuuster model te krijgen, wordt het aantal predictoren en componenten beperkt (principe van schaarsheid). Vervolgens wordt, met het gekozen aantal componenten, een multidimensionale lineaire regressie uitgevoerd, waarbij de respons wordt voorspeld door de gekozen componenten. Uiteindelijk wordt het bekomen model terug uitgedrukt in functie van de oorspronkelijke predictoren. De regressiecoëfficiënten van deze finale vergelijking worden onderzocht waardoor de relatieve bijdrage van de verklarende variabelen kwantitatief kan worden vastgesteld.

De volledige werkwijze is schematisch weergegeven in onderstaande figuur. Deze workflow werd toegepast op iedere responsvariabele en op ieder ruimtelijk aggregatieniveau.



Figuur 2-1: Werkwijze voor het opstellen van globale statistische modellen

De resulterende modellen slagen erin om een significant deel van de waargenomen variantie te verklaren. De verklaarde variantie varieert tussen 21 % en 88 %, afhankelijk van de beschouwde responsvariabele en aggregatieniveau. Er zijn geen consistente verschillen in modelperformantie waarneembaar tussen de responsvariabelen nitraat & fosfaat. Een ruimtelijke analyse van de statistische residuen vertoont duidelijk clustering van over- en onderschatting door het model voor de MAP meetpunten.

De opgestelde vergelijkingen bevatten over het algemeen een combinatie van variabelen met betrekking tot:

- Gewasrotatie
- Exploitatiegegevens
- Bodemgesteldheid
- Waterkwaliteit
- Omgevingsfactoren
- Opgelegde normen

Voor ieder opgesteld model is voor iedere predictor het belang van de predictor alsook de bijhorende coëfficiënt van de modelvergelijking bepaald. Er is een grote variatie aan geselecteerde predictoren waarneembaar, afhankelijk van de beschouwde combinatie van responsvariabele en schaalniveau waardoor er geen algemeen geldende conclusies voor het systeem als geheel kunnen worden afgeleid op basis van de globale analyse. Een eenduidige interpretatie van de coëfficiënten wordt bovendien bemoeilijkt door de onderlinge afhankelijkheid tussen de predictoren. Predictoren met betrekking tot de landbouwzone waarin het afstroomgebied gelegen is, zijn frequent terug te vinden in de vergelijkingen. Dit is een indicatie dat een differentiatie van de modelvergelijkingen per landbouwzone mogelijks bijkomende inzichten kan opleveren.

### 3. GERICHTE ANALYSE

Op basis van de resultaten van de globale statistische analyse werden er voor de gerichte analyse 3 specifieke doelstellingen geformuleerd.

Ten eerste werd het inzicht in de modelvergelijkingen vergroot door het aantal datapunten terug te brengen door middel van een weloverwogen selectie van het ruimtelijk en temporeel niveau. Daarnaast kan een beter inzicht in de modelstructuren ook verkregen worden door een clustering uit te voeren op de predictorruimte alvorens over te gaan tot het opstellen van modelvergelijkingen. Dergelijke clustering heeft als doel om variabelen te groeperen die onderling sterk gecorreleerd zijn. Deze dimensiereductie is uitgevoerd, gebruik makende van een factoranalyse.

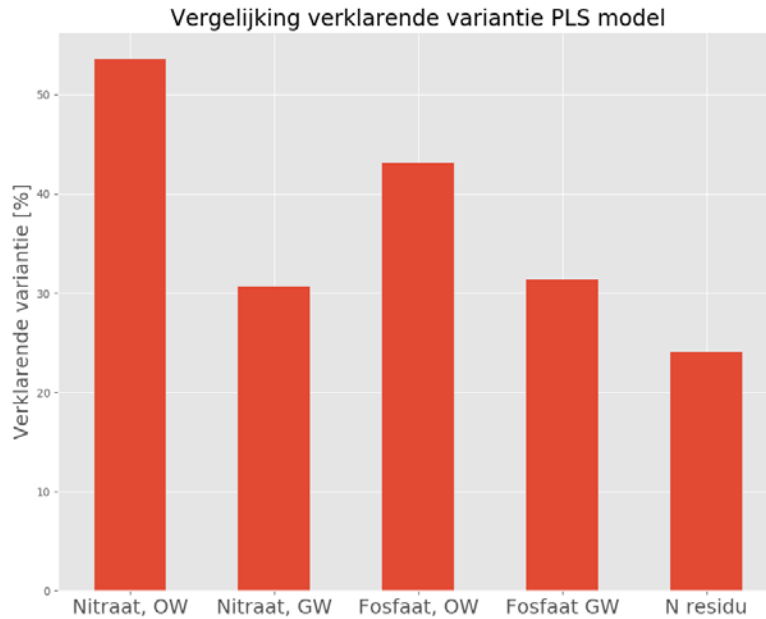
Daarnaast zijn pogingen ondernomen om de performantie van het model te verhogen. Het toevoegen van bijkomende variabelen en het in rekening brengen van niet lineaire verbanden bracht echter geen significante verbetering met zich mee.

Tenslotte werden de bekomen modellen ook gedifferentieerd volgens landbouwzone. Enkel de landbouwzones Kempen, Zandstreek en Zandleemstreek bevatten voldoende datapunten om een robuust model op te bouwen.

Met behulp van de predictoren zoals geselecteerd in de factoranalyse zijn statistische modellen opgesteld die de waarde van de verschillende responsvariabelen op gemeentelijk niveau voorspellen voor de periode 2011 – 2014 en dit zowel voor Vlaanderen alsook voor de landbouwstroken Kempen, Zandstreek en Zandleemstreek.

Zowel voor nitraat als fosfaat is de verklarende variantie van de resulterende modellen voor het oppervlaktewater significant hoger dan van de modellen voor het grondwater. De verklarende variantie van nitraat in het oppervlaktewater is eveneens significant groter dan deze van fosfaat in het oppervlaktewater. Daarnaast kunnen we ook afleiden dat de verklarende variantie van nitraat in het oppervlaktewater significant groter is dan deze van

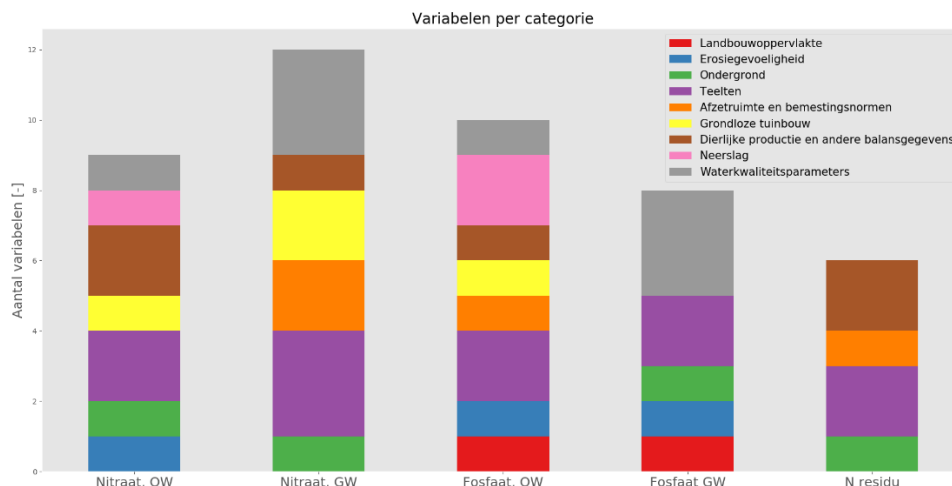
nitraatresidu. Er zijn bovendien duidelijke verschillen vast te stellen in de verklarende variantie van de modellen op niveau van Vlaanderen en deze van de verschillende landbouwstreken.



*Figuur 3-1: Samenvatting van verklarende variantie van de modellen voor Vlaanderen*

Op basis van de geselecteerde predictoren voor de verschillende vergelijkingen kunnen we concluderen dat variabelen met betrekking tot teelten in alle vergelijkingen zijn terug te vinden. Ook dierlijke mestproductie en balansgegevens zijn terug te vinden in alle vergelijkingen, met uitzondering van de vergelijking ter voorspelling van fosfaat in het grondwater. Variabelen met betrekking tot grondloze tuinbouw zijn terug te vinden in de vergelijkingen van nitraat (zowel oppervlaktewater als grondwater) en fosfaat (enkel oppervlaktewater). De vergelijking van fosfaat in het grondwater bevat slechts twee variabelen waarop ingegrepen kan worden (beide gerelateerd aan teelten). De overige variabelen zijn gelinkt aan de ondergrond of parameters m.b.t. waterkwaliteit waarop geen invloed kan uitgeoefend worden. Binnen een responsvariabele is er een sterke variatie per landbouwzone waarneembaar in zowel aantal als categorieën van de geselecteerde predictoren.





Figuur 3-2: Overzicht van de variabelen responsvariabele, opgedeeld in categorieën

Op basis van de resultaten van de gerichte analyse kunnen er enkele variabelen uitgelicht worden waarvoor duidelijke verbanden zijn blootgelegd tussen de variabele in kwestie en de waterkwaliteit. Een ongunstig effect in onderstaande bespreking wil zeggen dat een toename van deze variabele resulteert in een toename van de concentratie van de responsvariabele. Een gunstig effect daarentegen wil zeggen dat een toename van de variabele in kwestie resulteert in een daling van de concentratie van de responsvariabele. Het is hierbij belangrijk om op te merken dat de termen gunstig en ongunstig in deze context niet noodzakelijk een causaal verband tussen de variabelen in kwestie impliceren.

Hoofddeelt grassen is een belangrijke variabele, die een gunstig effect heeft op de responsvariabele nitraat in het oppervlaktewater en grondwater en nitraatresidu.

Nateelt groenten groep 1 is een belangrijke variabele, die een ongunstig effect heeft op de responsvariabele nitraat in het oppervlaktewater. Groenten groep 1 is een verzamelterm van meerdere soorten groenten. De belangrijkste hiervan zijn bloemkool en prei.

Emissieverlies, als representatieve variabele voor dierlijke mestproductie, is de belangrijkste predictor in de vergelijking van nitraatresidu en komt eveneens voor in de vergelijking van nitraat in het oppervlaktewater. In beide gevallen heeft de predictor een ongunstig effect op de gemeten concentraties.

Productiemethode SGM (serres met teelt op groeimedium), als representatieve factor voor grondloze tuinbouw, heeft een ongunstig effect op nitraat en fosfaat in het oppervlaktewater en een gunstig effect op nitraat in het grondwater. Het ongunstige effect in het oppervlaktewater wijst erop dat het voornamelijk directe lozingen zijn die een invloed uitoefenen op de waterkwaliteit.

Afzetmogelijkheid dier heeft een ongunstig effect op nitraat in het grondwater en op nitraatresidu. De norm fosfaat heeft dan weer een gunstig effect op fosfaat in het oppervlaktewater. Indien we ook de vergelijkingen van de landbouwstroken bestuderen, zien we een gelijkaardig beeld naar voor komen. Afzetmogelijkheden en normen blijken een ongunstig effect te hebben op de nitraatconcentratie en een gunstig effect op de fosfaatconcentratie.

De procesfactor heeft een gunstige invloed op de nitraatconcentratie in het oppervlaktewater. Dat valt binnen de verwachtingen aangezien de procesfactor wordt bepaald als de verhouding tussen de gemiddelde nitraatconcentratie over de winterperiode onderaan de wortelzone en de gemiddelde concentratie in het oppervlaktewater. Hoe hoger de procesfactor hoe meer verdunning zal optreden en hoe lager de verwachte concentraties in het oppervlaktewater. Het blijkt echter ook dat de procesfactor een ongunstige invloed heeft op de concentratie van fosfaat in het grondwater en op nitraatresidu. De procesfactor die hier in rekening wordt gebracht is de procesfactor voor oppervlaktewater bepaald op basis van een voorspellend regressiemodel met als verklarende variabelen de dominante textuur van het afstroomgebied en de redoxpotentiaal van de onderliggende aquifer. De ongunstige invloed van de procesfactor op nitraatresidu en fosfaat in het grondwater dient dan ook eerder begrepen te worden in functie van deze achterliggende factoren zoals bv. de dominante bodemtextuur.