

# VLAAMSE LAND MAATSCHAPPIJ

# RAPPORT

VLAAMSE LANDMAATSCHAPPIJ

VLM Centrale Directie

Statistische Analyse Waterkwaliteit

Globale statistische analyse

14 december 17 - versie 4.0



**AGRO  
LOUVAIN**  
BIOINGÉNIEURS



FACULTE DES  
SCIENCES


---


## Colofon

---

International Marine & Dredging Consultants

Adres: Van Immerseelstraat 66, 2018 Antwerp, Belgium

: + 32 3 270 92 95

: + 32 3 235 67 11

Email: [info@imdc.be](mailto:info@imdc.be)

Website: [www.imdc.be](http://www.imdc.be)

---

## Document Identificatie

---

Titel	Globale statistische analyse
Project	Statistische Analyse Waterkwaliteit
Opdrachtgever	VLAAMSE LANDMAATSCHAPPIJ - VLM Centrale Directie
Besteknummer	APLM/2016/1
Documentref	I/RA/11504/17.101/TFR
Documentnaam	K:\PROJECTS\11\11504_P009704 - Statistische analyse waterkwaliteit\10-Rap\DO2_GlobaleAnalyse\RA17101_GlobaleStatistischeAnalyse_v4.0.docx

---

## Revisies / Goedkeuring

---

Versie	Datum	Omschrijving	Auteur	Nazicht	Goedgekeurd
1.0	26/06/2017	Eerste revisie	TFR, JAC, GLE, SDO	JWA	TFR
2.0	07/07/2017	Verwerking commentaar (I/EM/11504/17.167/TFR/)	VLM TFR, JAC, GLE, SDO	JSW	TFR
3.0	09/11/2017	Verwerking commentaar stuurgroep (I/NO/11504/17.343/TFR/)	TFR, JAC, GLE, SDO	GLE	TFR
4.0	14/12/2017	Verwerking bijkomende commentaar stuurgroep (I/EM/11504/17.302/TFR/)	TFR, JAC, GLE, SDO	GLE	TFR

---

## Verdeellijst

---

	Analoog	
1	Digitaal	VLM

---

## Contactpersoon IMDC

---

### Contactpersoon Tim Franken

Telefoonnummer +32 3 287 23 82

E-mail [tfr@imdc.be](mailto:tfr@imdc.be)

## Inhoudstafel

<b>0. SAMENVATTING.....</b>	<b>12</b>
<b>1. INLEIDING.....</b>	<b>13</b>
1.1 DE OPDRACHT .....	13
1.2 DOEL VAN HET RAPPORT .....	13
1.3 LEESWIJZER .....	13
<b>2. STATISTISCHE MODELLEN.....</b>	<b>15</b>
2.1 MULTIVARIATE REGRESSIE (MVR) .....	15
2.2 PRINCIPAAL COMPONENT REGRESSIE (PCR).....	16
2.3 PARTIAL LEAST SQUARE REGRSSIE (PLS) .....	16
2.4 NON-PARAMETRIC CAUSAL RANDOM FOREST .....	17
2.5 MULTISPATIAL CONVERGENT CROSS MAPPING (CCM).....	18
<b>3. METHODIEK.....</b>	<b>20</b>
3.1 NORMALISATIE.....	21
3.2 SELECTIE OPTIMAAL MODEL .....	21
3.3 AANTAL PREDICTOREN .....	22
3.4 AANTAL COMPONENTEN.....	24
3.5 OPTIMAAL MODEL.....	25
<b>4. RESULTATEN.....</b>	<b>27</b>
4.1 NITRAAT.....	28
4.1.1 <i>Oppervlaktewater</i> .....	28
4.1.2 <i>Grondwater</i> .....	35
4.2 FOSFAAT.....	41
4.2.1 <i>Oppervlaktewater</i> .....	41
4.2.2 <i>Grondwater</i> .....	47
4.3 NITRAATRESIDU.....	53
4.3.1 <i>Perceel</i> .....	53
4.3.2 <i>Gemeente</i> .....	54
<b>5. ANALYSE EN SAMENVATTING VAN RESULTATEN .....</b>	<b>57</b>
5.1 NITRAAT, OPPERVLAKTEWATER .....	58
5.2 NITRAAT, GRONDWATER .....	63
5.3 FOSFAAT, OPPERVLAKTEWATER .....	65
5.4 FOSFAAT, GRONDWATER .....	67
5.5 NITRAATRESIDU.....	69
5.6 VERGELIJKING OP GEMEENTELIJK NIVEAU.....	71
5.7 VALIDATIE MET RANDOM FOREST MODEL .....	73

<b>6. ALGEMENE CONCLUSIE .....</b>	<b>76</b>
<b>7. REFERENTIES .....</b>	<b>78</b>

---

## Bijlagen

---

<b>BIJLAGE A</b>	<b>OVERZICHT PARAMETERS.....</b>	<b>79</b>
A.1	WATERKWALITEITSPARAMETERS MAP-MEETPUNTEN .....	80
A.2	METADATA FREATISCHE GRONDWATERMEETPUNTEN .....	81
A.3	WATERKWALITEITSPARAMETERS FREATISCHE GRONDWATERMEETPUNTEN .....	82
A.4	EXPLOITATIE .....	83
A.5	IDENTIFICATIE EXPLOITATIE .....	89
A.6	PERCEEL.....	90
A.7	GEWASGROEPEN EN VANGGEWASSEN.....	93
A.8	OVERIGE PARAMETERS .....	94
<b>BIJLAGE B</b>	<b>CCM.....</b>	<b>95</b>
B.1	INTRO .....	96
B.2	DESCRIPTION OF CONVERGENT CROSS MAPPING.....	97
B.3	APPLYING CONVERGENT CROSS MAPPING.....	98
B.4	STUDY CASE: CCM APPLIED TO HYDROLOGICAL DATA.....	99
<i>B.4.1</i>	<i>Data and Study site.....</i>	<i>99</i>
<i>B.4.2</i>	<i>Motivation .....</i>	<i>100</i>
<i>B.4.3</i>	<i>Methodology .....</i>	<i>100</i>
<i>B.4.4</i>	<i>Results.....</i>	<i>100</i>
<i>B.4.5</i>	<i>Discussion .....</i>	<i>102</i>
<i>B.4.6</i>	<i>Conclusion and perspectives.....</i>	<i>103</i>
<b>BIJLAGE C</b>	<b>RESULTATEN GLOBALE STATISTISCHE ANALYSE.....</b>	<b>105</b>

## Lijst van tabellen

TABEL 4-1: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED. ....	28
TABEL 4-2: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAAT IN OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	28
TABEL 4-3: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU GEMEENTE. ....	30
TABEL 4-4: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAAT IN OPPERVLAKTEWATER OP NIVEAU GEMEENTE: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	30
TABEL 4-5: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU BEKKEN.....	32
TABEL 4-6: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAAT IN OPPERVLAKTEWATER OP NIVEAU BEKKEN: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	32
TABEL 4-7: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU KLASSENVERDELING .....	34
TABEL 4-8: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAAT IN OPPERVLAKTEWATER OP NIVEAU KLASSENVERDELING: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	34
TABEL 4-9: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAAT IN GRONDWATER OP NIVEAU INTREKGEBIED.....	35
TABEL 4-10: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAAT IN GRONDWATER OP NIVEAU INTREKGEBIED: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	36
TABEL 4-11: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAAT IN GRONDWATER OP NIVEAU GEMEENTE.....	37
TABEL 4-12: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAAT IN GRONDWATER OP NIVEAU GEMEENTE: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	37
TABEL 4-13: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAAT IN GRONDWATER OP NIVEAU HHZ.....	39
TABEL 4-14: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAAT IN GRONDWATER OP NIVEAU HHZ: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	39
TABEL 4-15: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED .....	41
TABEL 4-16: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	41

TABEL 4-17: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU GEMEENTE .....	43
TABEL 4-18: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU GEMEENTE: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	43
TABEL 4-19: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU BEKKEN .....	45
TABEL 4-20: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU HHZ: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	45
TABEL 4-21: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR FOSFAAT IN GRONDWATER OP NIVEAU INTREKGEBIED.....	47
TABEL 4-22: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN FOSFAAT IN GRONDWATER OP NIVEAU INTREKGEBIED: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	47
TABEL 4-23: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR FOSFAAT IN GRONDWATER OP NIVEAU GEMEENTE .....	49
TABEL 4-24: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN FOSFAAT IN GRONDWATER OP NIVEAU GEMEENTE: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	49
TABEL 4-25: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR FOSFAAT IN GRONDWATER OP NIVEAU HHZ.....	51
TABEL 4-26: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN FOSFAAT IN GRONDWATER OP NIVEAU HHZ: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	51
TABEL 4-27: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAATRESIDU OP NIVEAU PERCEEL .....	53
TABEL 4-28: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAATRESIDU OP NIVEAU PERCEEL: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	53
TABEL 4-29: SELECTIE AANTAL PREDICTOREN EN COMPONENTEN MET BIJHORENDE VERKLARENDE VARIANTIE VOOR NITRAATRESIDU OP NIVEAU GEMEENTE. ....	54
TABEL 4-30: OVERZICHT VAN DE GESELECTEERDE PREDICTOREN MET BIJHORENDE COËFFICIËNT VOOR HET MODEL VAN NITRAATRESIDU OP NIVEAU GEMEENTE: $Y = \text{CONSTANTE} + \sum (\text{COËFFICIËNT} * \text{PREDICTOR})$ .....	54
TABEL 5-1: OVERZICHT RESULTATEN GLOBALE STATISTISCHE ANALYSE.....	57
TABEL 5-2: VERSCHILLENDE CATEGORIEËN VAN GESELECTEERDE PREDICTOREN VOOR NITRAAT, OPPERVLAKTEWATER .....	58
TABEL 5-3: OVERZICHT VAN SAMENGEVOEGDE PREDICTOREN VOOR VERGELIJKING VAN RESPONSVARIABLEN OP NIVEAU VAN GEMEENTE.....	71

TABEL BIJLAGE A-1: OVERZICHT VAN DE PARAMETERS DIE GEREgistREERD KUNNEN WORDEN IN EEN MAP-MEETPUNT .....	80
TABEL BIJLAGE A-2: META-GEGEVENS MEETPUNTEN FREATISCH GRONDWATERMEETNET .....	81
TABEL BIJLAGE A-3: OVERZICHT VAN DE PARAMETERS DIE GEREgistREERD KUNNEN WORDEN IN DE MEETPUNTEN VAN HET FREATISCH GRONDWATERMEETNET.....	82
TABEL BIJLAGE A-4: OVERZICHT VAN DE BESCHIKBARE EXPLOITATIEGEGEVENS, MET AANDUIDING WELKE PARAMETERS OPGENOMEN WERDEN IN DE STATISTISCHE ANALYSE.....	83
TABEL BIJLAGE A-5: OVERZICHT VAN DE BESCHIKBARE IDENTIFICATIEGEGEVENS VAN DE EXPLOITATIES. ....	89
TABEL BIJLAGE A-6: OVERZICHT VAN DE BESCHIKBARE PERCEELSGEGEVENS, MET AANDUIDING WELKE PARAMETERS OPGENOMEN WERDEN IN DE STATISTISCHE ANALYSE.....	90
TABEL BIJLAGE A-7: OVERZICHT VAN DE GEWASGROEPEN IN DE PERCEELSGEGEVENS (2007-2014).....	93
TABEL BIJLAGE A-8: OVERZICHT VAN DE VANGGEWASSEN (2014).....	93
TABEL BIJLAGE A-9: OVERZICHT VAN DE OVERIGE BESCHIKBARE PARAMETERS.....	94



## Lijst van figuren

FIGUUR 2-1: ILLUSTRATIE ENSEMBLE SET REGRESSIE BOMEN .....	17
FIGUUR 3-1: WERKWIJZE VOOR OPMAAK OPTIMAAL MODEL.....	20
FIGUUR 3-2: WORKFLOW NORMALISATIE DATA. ((I)CDF = (INVERSE) CUMULATIEVE DENSITEIT FUNCTIE).....	21
FIGUUR 3-3: SELECTIE VAN EEN OPTIMAAL MODEL VOOR NITRAAT, OPPERVLAKTEWATER, AFSTROOMGEBIED .....	22
FIGUUR 3-4: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR NITRAAT, OPPERVLAKTEWATER, AFSTROOMGEBIED. OP DE X AS ZIJN ALLE PREDICTOREN WEERGEGEVEN. DE VOLGORDE VAN DE RELEVANTE PREDICTOREN WORDT VERDER TOEGELICHT.....	23
FIGUUR 3-5: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR NITRAAT, OPPERVLAKTEWATER, AFSTROOMGEBIED .....	24
FIGUUR 3-6: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR NITRAAT, OPPERVLAKTEWATER, AFSTROOMGEBIED .....	25
FIGUUR 3-7: OPTIMAAL PLS MODEL VOOR OPPERVLAKTE WATER OP AFSTROOM NIVEAU EN RESPONS NITRAAT MET 5 COMPONENTEN EN 20 PREDICTOREN.....	26
FIGUUR 4-1: OPTIMAAL PLS MODEL VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED.....	29
FIGUUR 4-2: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR NITRAAT IN HET OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED .....	29
FIGUUR 4-3: OPTIMAAL PLS MODEL VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU GEMEENTE .....	31
FIGUUR 4-4: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR NITRAAT IN HET OPPERVLAKTEWATER OP NIVEAU GEMEENTE .....	31
FIGUUR 4-5: OPTIMAAL PLS MODEL VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU BEKKEN .....	33
FIGUUR 4-6: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR NITRAAT IN HET OPPERVLAKTEWATER OP NIVEAU BEKKEN .....	33
FIGUUR 4-7: OPTIMAAL PLS MODEL VOOR NITRAAT IN OPPERVLAKTEWATER OP NIVEAU KLASSENVERDELING. ....	35
FIGUUR 4-8: OPTIMAAL PLS MODEL VOOR NITRAAT IN GRONDWATER OP NIVEAU INTREKGEBIED.....	36
FIGUUR 4-9: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR NITRAAT IN HET GRONDWATER OP NIVEAU INTREKGEBIED .....	37
FIGUUR 4-10: OPTIMAAL PLS MODEL VOOR NITRAAT IN GRONDWATER OP NIVEAU GEMEENTE.....	38
FIGUUR 4-11: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR NITRAAT IN HET GRONDWATER OP NIVEAU GEMEENTE .....	38
FIGUUR 4-12: OPTIMAAL PLS MODEL VOOR NITRAAT IN GRONDWATER OP NIVEAU HHZ.....	40
FIGUUR 4-13: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR NITRAAT IN HET GRONDWATER OP NIVEAU HHZ.....	40

FIGUUR 4-14: OPTIMAAL PLS MODEL VOOR FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED. ....	42
FIGUUR 4-15: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR FOSFAAT IN HET OPPERVLAKTEWATER OP NIVEAU AFSTROOMGEBIED .....	42
FIGUUR 4-16: OPTIMAAL PLS MODEL VOOR FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU GEMEENTE. ....	44
FIGUUR 4-17: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR FOSFAAT IN HET OPPERVLAKTEWATER OP NIVEAU GEMEENTE .....	44
FIGUUR 4-18: OPTIMAAL PLS MODEL VOOR FOSFAAT IN OPPERVLAKTEWATER OP NIVEAU BEKKEN. ....	46
FIGUUR 4-19: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR FOSFAAT IN HET OPPERVLAKTEWATER OP NIVEAU BEKKEN .....	46
FIGUUR 4-20: OPTIMAAL PLS MODEL VOOR FOSFAAT IN GRONDWATER OP NIVEAU INTREKGEBIED. ....	48
FIGUUR 4-21: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR FOSFAAT IN HET GRONDWATER OP NIVEAU INTREKGEBIED .....	48
FIGUUR 4-22: OPTIMAAL PLS MODEL VOOR FOSFAAT IN GRONDWATER OP NIVEAU GEMEENTE. ....	50
FIGUUR 4-23: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR FOSFAAT IN HET GRONDWATER OP NIVEAU GEMEENTE .....	50
FIGUUR 4-24: OPTIMAAL PLS MODEL VOOR FOSFAAT IN GRONDWATER OP NIVEAU HHZ. ....	52
FIGUUR 4-25: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR FOSFAAT IN HET GRONDWATER OP NIVEAU HHZ. ....	52
FIGUUR 4-26: OPTIMAAL PLS MODEL VOOR NITRAATRESIDU OP NIVEAU PERCEEL. ....	54
FIGUUR 4-27: OPTIMAAL PLS MODEL VOOR NITRAATRESIDU OP NIVEAU GEMEENTE. ....	55
FIGUUR 4-28: RUIMTELIJKE VERDELING VAN STATISTISCH RESIDU VOOR NITRAATRESIDU OP NIVEAU GEMEENTE. ....	56
FIGUUR 5-1: OVERZICHT VAN DE WAARDE VAN DE COËFFICIËNTEN VOOR ALLE PREDICTOREN BIJ NITRAAT, OPPERVLAKTEWATER. ....	60
FIGUUR 5-2: GEMIDDELD RESIDU VAN HET STATISTISCH MODEL VOOR NITRAAT, OPPERVLAKTEWATER, AFSTROOMGEBIED PER KLASSE BEPAALD OP BASIS VAN COMBINATIE TREND EN GROOTTEORDE .....	61
FIGUUR 5-3: CORRELATIE TUSSEN NETTO PRODUCTIE N VARKENS EN EMISSIEVERLIES N. ....	62
FIGUUR 5-4: CORRELATIE TUSSEN JAARLIJKSE MAXIMALE NITRAATCONCENTRATIE IN HET OPPERVLAKTEWATER EN EMISSIEVERLIES N. DE KLEURSCHAAL GEEFT DE GROOTTE VAN NETTO PRODUCTIE N VARKENS AAN. BLAUWE CIRKEL GEEFT PUNTEN WEER WAARVOOR HOGERE CONCENTRATIE VERWACHT KAN WORDEN OP BASIS VAN DE WAARDEN VAN EMISSIEVERLIES N. ....	62
FIGUUR 5-5: OVERZICHT VAN DE WAARDE VAN DE COËFFICIËNTEN VOOR ALLE PREDICTOREN BIJ NITRAAT, GRONDWATER. ....	64
FIGUUR 5-6: OVERZICHT VAN DE WAARDE VAN DE COËFFICIËNTEN VOOR ALLE PREDICTOREN BIJ FOSFAAT, OPPERVLAKTEWATER .....	66
FIGUUR 5-7: OVERZICHT VAN DE WAARDE VAN DE COËFFICIËNTEN VOOR ALLE PREDICTOREN BIJ FOSFAAT, GRONDWATER. ....	68

FIGUUR 5-8: OVERZICHT VAN DE WAARDE VAN DE COËFFICIËNTEN VOOR ALLE PREDICTOREN BIJ NITRAATRESIDU .....	70
FIGUUR 5-9: OVERZICHT VAN DE WAARDE VAN DE COËFFICIËNTEN VOOR ALLE PREDICTOREN OP GEMEENTELIJK NIVEAU. * BETREFT SAMENGEVOEGDE PREDICTOREN .....	72
FIGUUR 5-10: OPTIMALISATIE RANDOM FOREST PARAMETERS OP BASIS VAN OOB SCORE.....	73
FIGUUR 5-11: OPTIMALISATIE MINIMUM SAMPLES IN 1 OPDELING .....	74
FIGUUR 5-12: RESULTAAT RANDOM FOREST MODEL: VOORSPELDE WAARDEN TEN OPZICHTE VAN GEMETEN WAARDEN.....	75
FIGUUR 5-13: GESORTEERDE PREDICTOREN IN FUNCTIE VAN HET RELATIEF BELANG .....	75
FIGUUR BIJLAGE C-1: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR OW, N, AFSTROOM .....	106
FIGUUR BIJLAGE C-2: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR OW, N, AFSTROOM.....	106
FIGUUR BIJLAGE C-3: SELECTIE VAN EEN OPTIMAAL MODEL VOOR OW, N, AFSTROOM.....	107
FIGUUR BIJLAGE C-4: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR OW, N, AFSTROOM.....	107
FIGUUR BIJLAGE C-5: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR OW, P, AFSTROOM .....	108
FIGUUR BIJLAGE C-6: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR OW, P, AFSTROOM .....	108
FIGUUR BIJLAGE C-7: SELECTIE VAN EEN OPTIMAAL MODEL VOOR OW, P, AFSTROOM .....	109
FIGUUR BIJLAGE C-8: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR OW, P, AFSTROOM.....	109
FIGUUR BIJLAGE C-9: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR GW, N, INTREK .....	110
FIGUUR BIJLAGE C-10: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR GW, N, INTREK .....	110
FIGUUR BIJLAGE C-11: SELECTIE VAN EEN OPTIMAAL MODEL VOOR GW, N, INTREK .....	111
FIGUUR BIJLAGE C-12: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR GW, N, INTREK .....	111
FIGUUR BIJLAGE C-13: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR GW, P, INTREK .....	112
FIGUUR BIJLAGE C-14: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR GW, P, INTREK .....	112
FIGUUR BIJLAGE C-15: SELECTIE VAN EEN OPTIMAAL MODEL VOOR GW, P, INTREK .....	113
FIGUUR BIJLAGE C-16: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR GW, P, INTREK.....	113
FIGUUR BIJLAGE C-17: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR OW, N, GEMEENTE .....	114
FIGUUR BIJLAGE C-18: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR OW, N, GEMEENTE .....	114

FIGUUR BIJLAGE C-19: SELECTIE VAN EEN OPTIMAAL MODEL VOOR OW, N, GEMEENTE .....	115
FIGUUR BIJLAGE C-20: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR OW, N, GEMEENTE.....	115
FIGUUR BIJLAGE C-21: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR OW, P, GEMEENTE .....	116
FIGUUR BIJLAGE C-22: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR OW, P, GEMEENTE .....	116
FIGUUR BIJLAGE C-23: SELECTIE VAN EEN OPTIMAAL MODEL VOOR OW, P, GEMEENTE .....	117
FIGUUR BIJLAGE C-24: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR OW, P, GEMEENTE .....	117
FIGUUR BIJLAGE C-25: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR GW, N, GEMEENTE .....	118
FIGUUR BIJLAGE C-26: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR GW, N, GEMEENTE .....	118
FIGUUR BIJLAGE C-27: SELECTIE VAN EEN OPTIMAAL MODEL VOOR GW, N, GEMEENTE .....	119
FIGUUR BIJLAGE C-28: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR GW, N, GEMEENTE.....	119
FIGUUR BIJLAGE C-29: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR GW, P, GEMEENTE .....	120
FIGUUR BIJLAGE C-30: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR GW, P, GEMEENTE .....	120
FIGUUR BIJLAGE C-31: SELECTIE VAN EEN OPTIMAAL MODEL VOOR GW, P, GEMEENTE .....	121
FIGUUR BIJLAGE C-32: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR GW, P, GEMEENTE .....	121
FIGUUR BIJLAGE C-33: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR OW, N, BEKKEN.....	122
FIGUUR BIJLAGE C-34: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR OW, N, BEKKEN.....	122
FIGUUR BIJLAGE C-35: SELECTIE VAN EEN OPTIMAAL MODEL VOOR OW, N, BEKKEN.....	123
FIGUUR BIJLAGE C-36: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR OW, N, BEKKEN .....	123
FIGUUR BIJLAGE C-37: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR OW, P, BEKKEN.....	124
FIGUUR BIJLAGE C-38: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR OW, P, BEKKEN.....	124
FIGUUR BIJLAGE C-39: SELECTIE VAN EEN OPTIMAAL MODEL VOOR OW, P, BEKKEN.....	125
FIGUUR BIJLAGE C-40: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR OW, P, BEKKEN .....	125
FIGUUR BIJLAGE C-41: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR OW, N, KLASSE.....	126
FIGUUR BIJLAGE C-42: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR OW, N, KLASSE .....	126
FIGUUR BIJLAGE C-43: SELECTIE VAN EEN OPTIMAAL MODEL VOOR OW, N, KLASSE .....	127

FIGUUR BIJLAGE C-44: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR OW, N, KLASSE.....	127
FIGUUR BIJLAGE C-45: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR GW, N, HHZ .....	128
FIGUUR BIJLAGE C-46: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR GW, N, HHZ.....	128
FIGUUR BIJLAGE C-47: SELECTIE VAN EEN OPTIMAAL MODEL VOOR GW, N, HHZ .....	129
FIGUUR BIJLAGE C-48: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR GW, N, HHZ.....	129
FIGUUR BIJLAGE C-49: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR GW, P, HHZ.....	130
FIGUUR BIJLAGE C-50: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR GW, P, HHZ.....	130
FIGUUR BIJLAGE C-51: SELECTIE VAN EEN OPTIMAAL MODEL VOOR GW, P, HHZ.....	131
FIGUUR BIJLAGE C-52: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR GW, P, HHZ.....	131
FIGUUR BIJLAGE C-53: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR NRES, PERCEEL.....	132
FIGUUR BIJLAGE C-54: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR NRES, PERCEEL .....	132
FIGUUR BIJLAGE C-55: SELECTIE VAN EEN OPTIMAAL MODEL VOOR NRES, PERCEEL.....	133
FIGUUR BIJLAGE C-56: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR NRES, PERCEEL .....	133
FIGUUR BIJLAGE C-57: AANTAL COMPONENTEN IFV DE VERKLAARDE VARIANTIE VOOR NRES, GEMEENTE .....	134
FIGUUR BIJLAGE C-58: AANTAL PREDICTOREN IFV VERKLARENDE VARIANTIE VOOR NRES, GEMEENTE .....	134
FIGUUR BIJLAGE C-59: SELECTIE VAN EEN OPTIMAAL MODEL VOOR NRES, GEMEENTE .....	135
FIGUUR BIJLAGE C-60: VOLGORDE VAN DE PREDICTOREN ZOALS BEPAALD DOOR STEPWISE MVR VOOR NRES, GEMEENTE .....	135

## 0. SAMENVATTING

In dit rapport zijn globale statistische modellen opgesteld voor Vlaanderen met als doel om inzicht te verkrijgen in de achterliggende oorzaken van een goede of slechte waterkwaliteit. Er zijn modellen opgesteld voor verschillende responsvariabelen (nitraat en fosfaat), voor metingen in het oppervlaktewater, grondwater en op het perceel (nitraatresidu) en voor verschillende ruimtelijke aggregatieniveaus.

Er zijn initieel 5 verschillende statistische modellen in rekening genomen waarvan er 2 (CCM en Random Forest) niet volledig zijn uitgewerkt op ieder niveau. 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. Dit model werd vervolgens toegepast op iedere responsvariabele en op ieder ruimtelijk aggregatieniveau.

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, afhankelijk van de beschouwde combinatie van responsvariabele en schaalniveau waardoor er geen algemeen geldende conclusies voor het systeem als geheel kunnen worden afgeleid. Zoals blijkt uit de voorbeelden in Hoofdstuk 5 wordt een eenduidige interpretatie van de coëfficiënten bovendien bemoeilijkt door de onderlinge afhankelijkheid tussen de predictoren. De oppervlakte maïsteelt is de enige predictor die in alle modellen terug te vinden is. Al verandert het teken van de bijhorende coëfficiënt wel. Predictoren met betrekking tot de landbouwzone waarin het afstroomgebied gelegen is, zijn ook frequent terug te vinden in de vergelijkingen. Dit is een indicatie dat een differentiatie van de modelvergelijkingen per landbouwzone mogelijks bijkomende inzichten kan opleveren.

# 1. INLEIDING

## 1.1 DE OPDRACHT

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

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. Vanuit de verbeterde inzichten in de achterliggende redenen voor een goede of slechte waterkwaliteit in bepaalde gebieden, worden vervolgens aanbevelingen opgesteld voor implementatie van deze inzichten binnen de landbouwsector en de bijsturing van het beleid.

De studie bestaat uit 4 onderdelen:

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

## 1.2 DOEL VAN HET RAPPORT

Het voorliggende rapport beschrijft het tweede onderdeel, zijnde de globale statistische analyse van de beschikbare data. Het doel van dit onderdeel is inzicht te verkrijgen in de achterliggende oorzaken van een goede of slechte waterkwaliteit door het opstellen van statistische modellen die de waterkwaliteit voorspellen en vervolgens de modelstructuur en performantie van de modellen te analyseren.

## 1.3 LEESWIJZER

Dit rapport bouwt verder op de inzichten verworven in onderdeel 1 van deze studie: verkenning van de beschikbare data. Een uitgebreide beschrijving hiervan, alsook van de gebruikte data, variabelen en ruimtelijke aggregatieniveaus is terug te vinden in het desbetreffende deelrapport (I/RA/11504/17.100/TFR/). Om de leesbaarheid van dit rapport als afzonderlijk geheel te verhogen, wordt een samenvatting van de beschouwde variabelen gegeven in Bijlage A.

Hoofdstuk 2 geeft een theoretische beschrijving van de onderzochte modellen. Elk van de modellen heeft zijn sterktes en zwaktes en alvorens ze toe te passen op de volledige dataset is de toepasbaarheid van ieder model onderzocht, rekening houdend met het doel van deze globale analyse.

In Hoofdstuk 3 wordt vervolgens de methode beschreven om tot een optimaal en robuust statistisch model te komen.

Hoofdstuk 4 beschrijft de resultaten van de toepassing van deze werkwijze voor iedere combinatie van responsvariabelen en ruimtelijk aggregatieniveau. De interpretatie van het effect en het belang van iedere predictor op de responsvariabelen dient te gebeuren aan de hand van de tabellen met coëfficiënten die hier zijn weergegeven, in combinatie met de inzichten uit de bivariate analyse zoals beschreven in Onderdeel 1.

In Hoofdstuk 5 wordt een samenvatting en analyse gegeven van de bekomen resultaten. Deze analyse poogt geenszins om alle modellen tot in detail te ontleden, maar dient eerder aanzien te worden als een voorbeeld en leidraad op basis waarvan interpretatie van de tabellen in Hoofdstuk 4 kan gebeuren. Daarnaast worden enkele tendensen uit de resultaten weergegeven door middel van samenvattende figuren en tabellen.

Hoofdstuk 6 geeft tenslotte een overzicht van de algemene conclusies van de globale statistische analyse.



## 2. STATISTISCHE MODELLEN

Voor de globale statistische analyse zijn initieel 5 verschillende statistische modellen in rekening genomen:

- Multivariate regressie
- Principaal component regressie
- Partial Least Square regressie
- Non-parameteric causal random forest
- Multispatial convergent cross mapping

Elk van deze modellen heeft zijn sterktes en zwaktes. Alvorens deze modellen toe te passen op de volledige dataset is de toepasbaarheid van ieder model onderzocht, rekening houdend met het doel van deze globale analyse.

### 2.1 MULTIVARIATE REGRESSIE (MVR)

Bij het opstellen van een multivariate regressie model wordt de relatie tussen de responsvariabele en de predictorvariabelen uitgedrukt in functie van een functievoorschrift. De fit van de coëfficiënten in deze relatie gebeurt door de minimalisatie van de norm van de residue vector. Er wordt gestart met een univariaat regressiemodel, dat de best verklarende predictorvariabele bevat, en dit vervolgens stap voor stap uit te breiden met de andere predictorvariabelen, op volgorde van hun correlatie van de residuen uit de vorige stap. Deze stepwise regressie geeft ook een volgorde van de predictoren. Een multivariate regressie bevat de mogelijkheid om niet-lineaire verbanden op te nemen in de relatie. Deze mogelijkheid kan meegenomen worden als er fysische aanwijzingen zijn voor een niet-lineair verband.

Bij multivariate regressie wordt aangenomen dat een set van onafhankelijke variabelen een deel van de variantie van de afhankelijke variabele verklaart. Alle predictoren en de respons worden genormaliseerd om de multicollineariteit, de correlatie tussen de predictoren, te verminderen en om het vergelijken van de modelcoëfficiënten toe te laten. Met deze vergelijking kan de relatieve bijdrage van de verklarende variabelen kwantitatief worden vastgesteld.

De voordelen van multivariate regressie zijn het feit dat niet-lineariteit mee kan genomen worden en het model eenvoudig interpreteerbaar is. Daartegen is het model gevoelig aan outliers en multicollineariteit en dient aan de modelvoorwaarden zoals lineariteit van de regressievergelijking, de normaliteit van de errortermen en de constante variantie van de errortermen voldaan te zijn. Ook dient opgemerkt te worden dat een lineaire afhankelijkheid geen causaal verband impliceert tussen variabelen.

De Matlab statistical toolbox is gebruikt voor de multivariate regressie.

## 2.2 PRINCIPAAL COMPONENT REGRESSIE (PCR)

Principaal component regressie is een regressie analyse gebaseerd op een principaal component analyse (PCA<sup>1</sup>). 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 orthogonaal of onafhankelijk zijn. Dit resulteert in een set van onafhankelijke componenten. Het is immers zo dat bv. emissieverlies en veestapel wellicht gecorreleerd zijn, en dus de robuustheid van het model ondergraven wordt (principe van schaarste: extra parameters waar mogelijk vermijden). De respons wordt nu uitgedrukt als een functie van deze componenten, de lineaire combinaties van de predictoren. Deze lineaire combinaties zullen, in zoverre zij de variatie van de respons kunnen verklaren, meegenomen worden in het resulterende model. Om een robuust model op te stellen, wordt het aantal predictoren en componenten beperkt (principe van schaarsheid). Dankzij de opeenvolging van twee lineaire vergelijkingen kan het respons variabel model uiteindelijk worden uitgedrukt in functie van de oorspronkelijke predictoren. De regressiecoëfficiënten van deze finale vergelijking worden onderzocht waarmee de relatieve bijdrage van de verklarende variabelen kwantitatief kan worden vastgesteld.

De voordelen van PCR zijn een gereduceerde complexiteit en een reductie van de ruis. Aangezien het om een combinatie van lineaire relaties gaat, is de bijdrage van niet lineaire relaties klein.

De Matlab statistical toolbox is gebruikt voor het PCR model op te stellen.

## 2.3 PARTIAL LEAST SQUARE REGRESSIE (PLS)

PLS en PCR zijn verwante methodes. Dit houdt in dat bij PLS, net zoals bij PCR, een nieuwe predictor-ruimte wordt gezocht zodanig dat de predictoren onderling minimaal gecorreleerd zijn, maar 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.

---

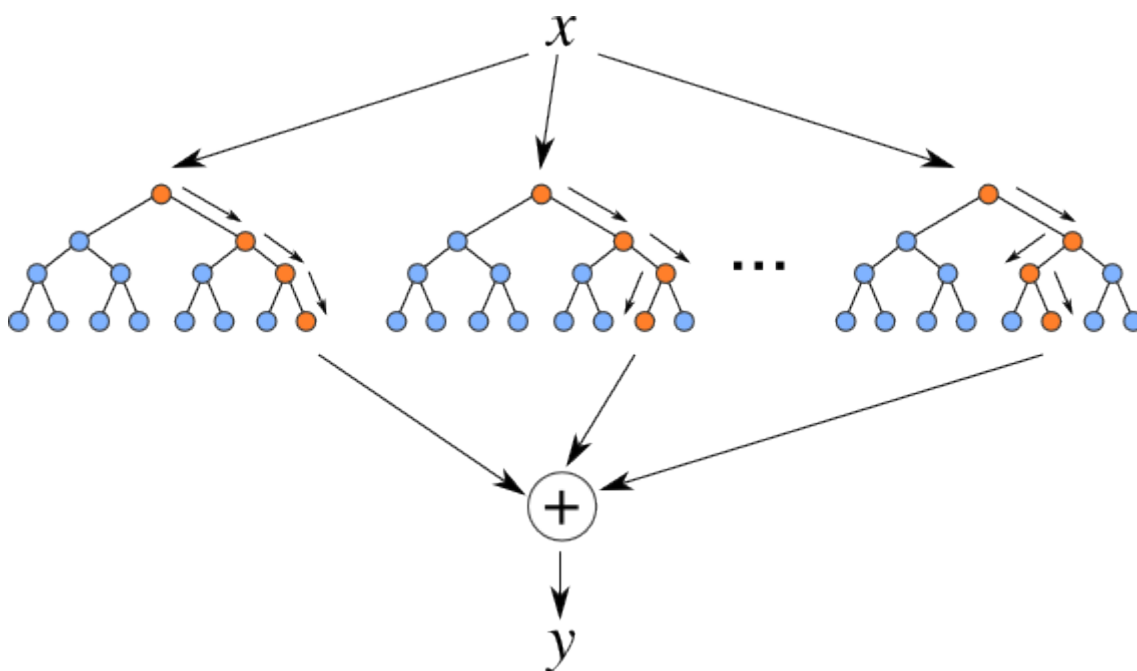
<sup>1</sup> Voor een intuïtieve illustratie van het concept PCA wordt verwezen naar <https://www.youtube.com/watch?v=BfTMmoDFXyE>

De voordelen van PLS zijn een gereduceerde complexiteit met oog op de verklaring van de variantie in de respons en een reductie van de ruis. Aangezien het om een combinatie van lineaire relaties gaat, is de bijdrage van niet-lineaire relaties klein.

De Matlab statistical toolbox is gebruikt voor het PLS model op te stellen.

## 2.4 NON-PARAMETRIC CAUSAL RANDOM FOREST

Het Random Forest algoritme creëert meerdere random beslissingsbomen en behoort tot de ensemble technieken. Alle beslissingsbomen worden gebruikt om de responsvariabele te classificeren. Het label of waarde van de respons dat door de meeste bomen gekozen wordt, is het uiteindelijke resultaat. Het algoritme berekent welke predictorvariabele de grootste variantie in de responsvariabele verklaart. Door de resultaten van meerdere beslissingsbomen te combineren, wordt de robuustheid van het model vergroot. De Random Forest techniek is bruikbaar voor zowel classificatie als regressie. De performantie van het model wordt geëvalueerd door een out of bag (OOB) score en een volgorde van predictoren volgens hun bijdrage aan de verklarende variantie van de respons.



Figuur 2-1: Illustratie ensemble set regressie bomen

Om een Random Forest model op te stellen, dienen volgende parameters geoptimaliseerd te worden

- Het aantal bomen
- Het aantal predictoren dat wordt meegenomen in 1 boom
- De maximale diepte van de boom
- Het minimaal aantal observaties om 1 node te splitsen

De voordelen van een Random Forest model zijn het niet-parametrische karakter van het model, waarbij niet wordt uitgegaan van een normale verdeling van de residu's, er geen aanname van lineariteit is, waardoor complexe interacties kunnen meegenomen worden, en de data niet gestandaardiseerd moet worden. Wel is het model gevoelig voor outliers en wordt de interpretatie bemoeilijkt bij gecorreleerde predictoren. Daarnaast is Random Forest een ensemble techniek. Dit impliceert dat er geen eenduidige modelstructuur kan worden afgeleid aangezien het finale resultaat wordt gekozen op basis van een groot aantal, mogelijks significant verschillende, beslissingsbomen. Wel kan het relatieve belang van de verschillende predictoren worden afgeleid op basis van de resultaten.

Een Random Forest model kan met andere woorden inzichten geven in welke predictoren van belang zijn, maar geeft weinig bijkomende informatie over hoe deze predictoren de waterkwaliteit beïnvloeden. Gezien dit net het doel is van deze globale analyse wordt het Random Forest model niet meegenomen in de algemene methodiek om tot een algemeen optimaal model te komen.

Random Forest wordt wel toegepast op één responsvariabele en aggregatieniveau in een vergelijkende analyse met de resultaten van het optimaal model. Deze vergelijking heeft tot doel om de sterktes en zwaktes van Random Forest verder te duiden en als verificatie van de resultaten van het optimaal model.

De python sklearn toolbox is gebruikt om het Random Forest model op te zetten.

## 2.5 MULTISPATIAL CONVERGENT CROSS MAPPING (CCM)

CCM is een statistische methode om causale effecten tussen zwak gerelateerde niet-lineaire systemen te bepalen. CCM is een algoritme dat het vermogen vergelijkt van lagged componenten om het dynamisch proces van elkaar in te schatten. De tijdreeksen van twee variabelen worden via CCM gekoppeld met als resultaat een causale relatie en een causale richting.

Om een CCM model op te stellen voor tijdsreeksen van twee variabelen X en Y worden deze tijdreeksen eerst opgedeeld in samples met toenemende lengte. Vervolgens worden modellen opgesteld met verschillende tijdsverschuivingen (State Space Reconstruction). Tenslotte worden predicties van de modellen opgesteld op verschillende sample lengtes en tijdsverschuivingen, en de voorspellende mogelijkheden geëvalueerd. Toename in deze voorspellende mogelijkheden wordt gebruikt om de causale relatie en richting te bepalen. De toepasbaarheid van deze methodologie in het kader van deze opdracht is door UCL onderzocht. In Bijlage B is het resultaat van dit onderzoek te vinden met een uitgebreide beschrijving van de methode alsook een toepassing van de methodologie op een hydrologische test dataset.

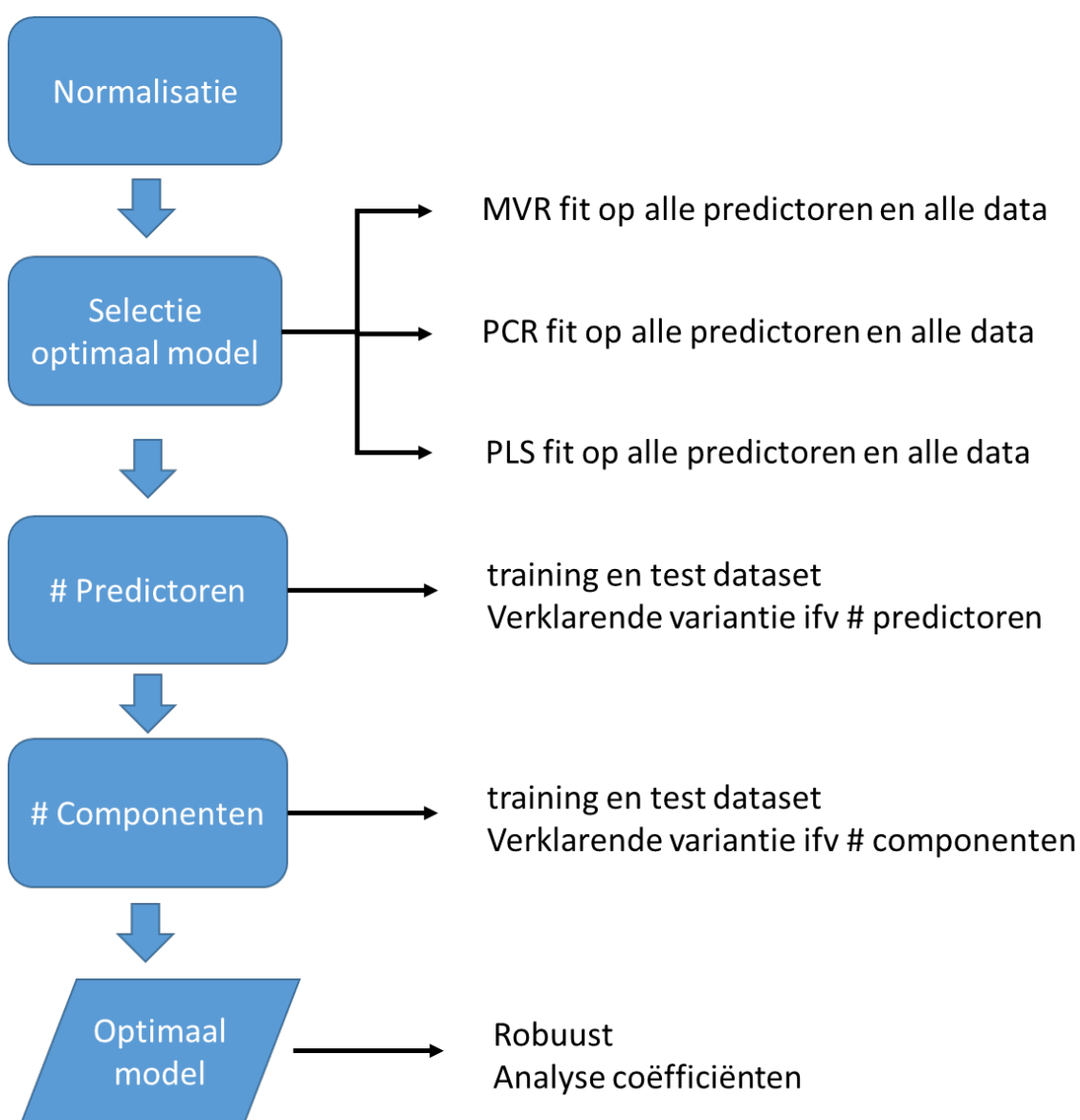
De CCM techniek is toepasbaar op relatief korte datasets. Door de tijdsverschuivingen en de verschillende lengtes van samples is echter wel een minimum lengte noodzakelijk. De voorliggende dataset heeft een lengte van 9 jaar (2007 tot 2015). Voor de respons variabelen zijn wel meerdere metingen per jaar beschikbaar maar de predictoren limiteren het maximaal aantal tijdstappen tot 9. Dit is te weinig om CCM succesvol toe te passen. Het "pairwise" karakter van de methode, waarbij steeds 2 variabelen vergeleken worden, bemoeilijkt ook het gebruik in een globale analyse waarbij zeer veel predictoren worden meegenomen.

Indien langere tijdreeksen beschikbaar komen kan CCM wel een toegevoegde waarde brengen in een detailanalyse. Omwille van bovenstaande limiteringen is CCM niet verder meegenomen in de globale analyse.

### 3. METHODIEK

Zoals hierboven beschreven zijn 3 statistische modellen (MVR, PCR en PLS) weerhouden voor de globale analyse. Voor iedere responsvariabele en voor ieder aggregatieniveau wordt het meest optimale model geselecteerd. Het niet-parametrische Random Forest model is eveneens toegepast op 1 dataset, namelijk van responsvariabele nitraat voor oppervlaktewater en het gemeentelijk aggregatieniveau voor het identificeren van de sterktes en pijnpunten van het model en ter verificatie van de resultaten uit het geselecteerde model. Het CCM model is niet weerhouden voor verdere toepassing in deze analyse.

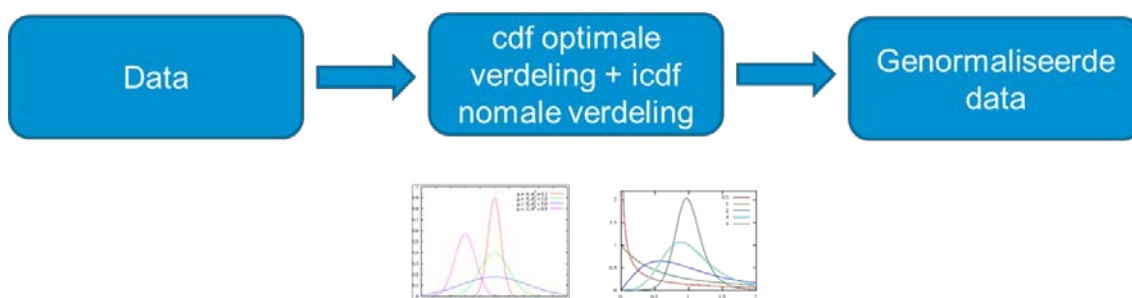
Om tot een optimaal en robuust statistisch model te komen dat de analyse en onderlinge vergelijking van coëfficiënten toelaat, is de werkwijze gevolgd zoals weergegeven in Figuur 3-1.



Figuur 3-1: Werkwijze voor opmaak optimaal model

### 3.1 NORMALISATIE

Vooraleer een multivariate relatie kan opgesteld worden, dienen de data genormaliseerd te worden, aangezien de verschillende variabelen een andere grootteorde en gemiddelde hebben (Figuur 3-2).



Figuur 3-2: Workflow normalisatie data. ((i)cdf = (inverse) cumulatieve densiteit functie)

Hiervoor wordt voor elke respons en predictor de best passende verdeling geselecteerd en worden de datapunten omgezet naar waarden tussen 0 en 1 via de cumulatieve densiteit functie (cdf). Vervolgens worden de waarden terug omgezet doormiddel van een inverse cumulatieve densiteit functie (icdf) van een normale verdeling met een gemiddelde van 0 en een standaard deviatie van 1. Hierdoor hebben alle predictoren en de respons een zelfde gemiddelde, nul, en bereik en kunnen de modelcoëfficiënten vergeleken worden. Voor meer informatie omtrent de mogelijke verdelingen en de selectie van de meest geschikte verdeling wordt verwezen naar Deelrapport 1.

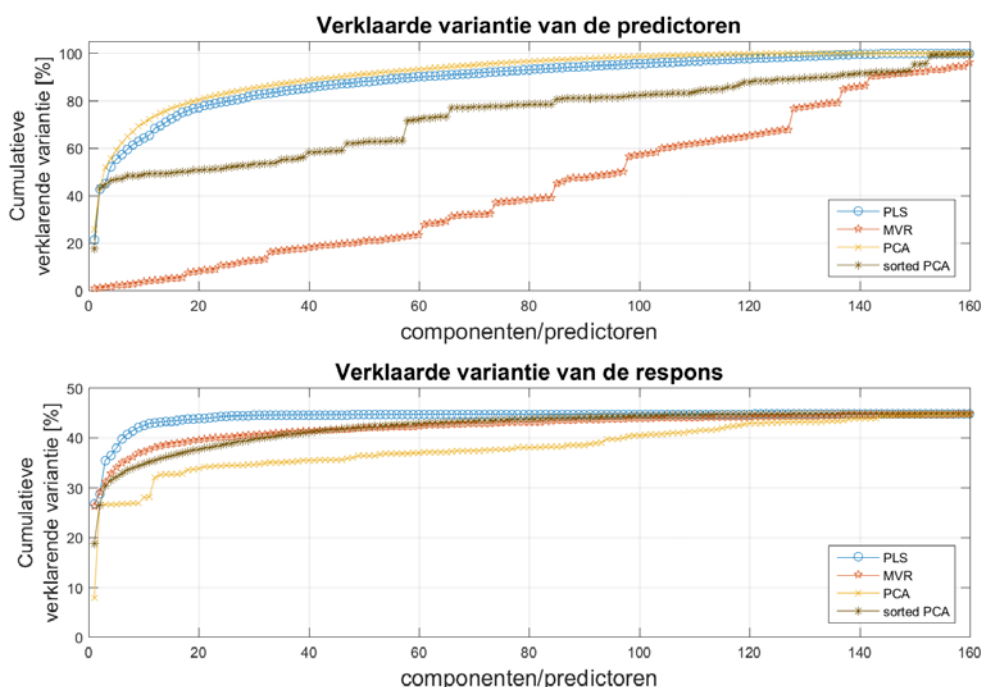
### 3.2 SELECTIE OPTIMAAL MODEL

Voor elke dataset is een MVR, een PCR en een PLS fit uitgevoerd op alle data. Voor de selectie van het optimaal model wordt geen opdeling gemaakt in een training en test dataset. Het optimaal aantal predictoren wordt geselecteerd door het aantal predictoren (MVR) of componenten (PCR en PLS) uit te zetten in functie van de cumulatieve verklarende variantie van het model. De verklarende variantie is een maat voor de variabiliteit die verklaard wordt door het statistisch model. Figuur 3-3 geeft een voorbeeld van de verklarende variantie van de predictoren weer (bovenste figuur) en de verklarende variantie in de respons voor een willekeurige responsvariabele en ruimtelijk aggregatieniveau. Voor de PCA/PCR zijn twee lijnen weergegeven, 1 waarbij de componenten gerangschikt zijn op volgorde van hun bijdrage tot de verklarende variantie van de predictoren (= PCA) en 1 waarbij de componenten zijn gerangschikt op volgorde van hun bijdrage tot de verklarende variantie van de respons (= sorted PCA). Hier is duidelijk dat via PCA met een beperkt aantal componenten het meeste van de verklarende variantie in de predictoren kan verklaren. Dit is logisch aangezien de componenten gekozen worden door een projectie in de multivariate ruimte met de bedoeling om de maximale variantie in 1 dimensie mee te nemen in 1 component. Ook PLS scoort goed voor de verklarende variantie van de predictoren met een relatief beperkt aantal componenten. De techniek is immers gelijkaardig aan PCA met het verschil dat hier ook de respons wordt meegenomen in de projectie.

De MVR heeft mindere verklarende variantie van de predictoren omdat hier geen projectie plaatsvindt en de stepwise regressie de predictoren toevoegt in volgorde van hun bijdrage in de verklaarde variantie van de respons.

Voor de verklarende variantie van de respons geeft PLS het beste resultaat met een beperkt aantal componenten. Het PLS model wordt in dit voorbeeld dan ook weerhouden als optimaal model. Uit het vervolg zal blijken dat dit zo is voor alle datasets. De PCA en MVR presteren ongeveer even goed.

Het aantal predictoren en componenten om tot een robuust PLS model te komen zal worden bepaald in volgende stappen.



Figuur 3-3: Selectie van een optimaal model voor nitraat, oppervlaktewater, afstroomgebied

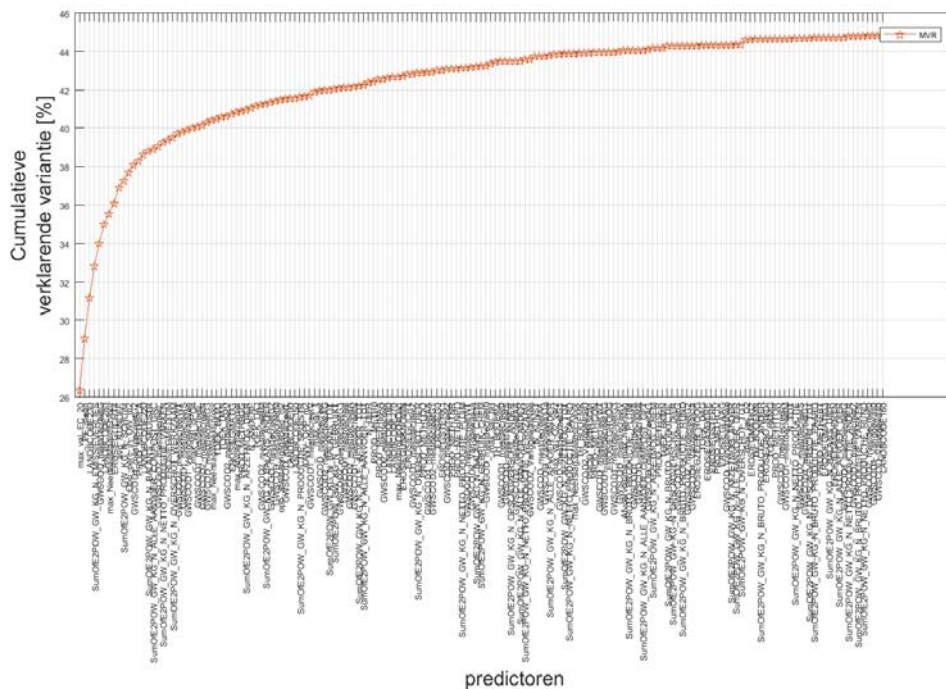
### 3.3 AANTAL PREDICTOREN

Om het aantal predictoren in het MVR en PLS model te bepalen wordt nagegaan wanneer de verklarende variantie van het model, opgesteld op een trainings dataset en toegepast op een test dataset, niet meer significant toeneemt.

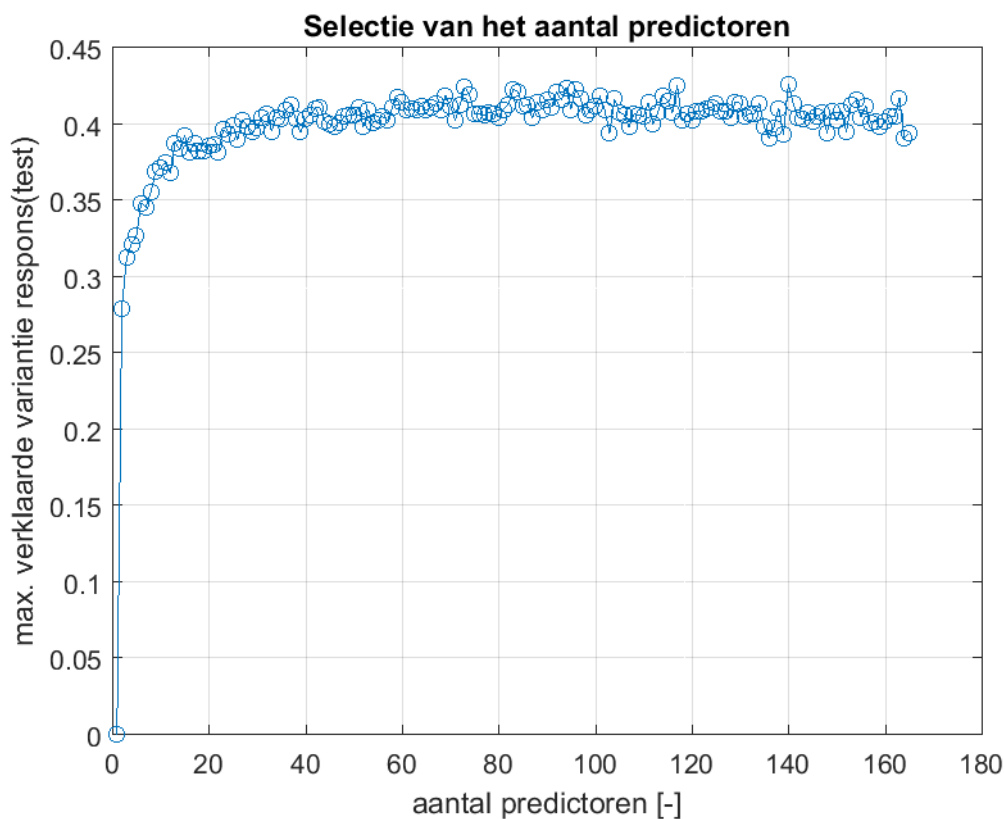
De beschikbare data worden eerst opgedeeld in 9 random verdeelde training (2/3 van de data) en test (1/3 van de data) datasets. Er wordt gekozen voor 9 datasets om mogelijke random effecten bij de opdeling in training- en testdataset te elimineren. De volgorde van de toe te voegen predictoren wordt allereerst bepaald volgens de stepwise multivariate regressie (§2.1 en Figuur 3-4). Vervolgens wordt stap voor stap een predictor toegevoegd aan de model fit. Deze fit wordt uitgevoerd op de 9 trainingsdatasets en geëvalueerd ten opzichte van de 9 test datasets. De maximale verklarende variantie over alle componenten wordt weerhouden als proxi. Vervolgens is de mediaan van deze 9 proxi's uitgezet in functie van het aantal predictoren.



Een voorbeeld van het resultaat van deze analyse wordt weergegeven in Figuur 3-5. In dit voorbeeld is duidelijk dat vanaf 20 predictoren de toename in verklarende variabiliteit marginaal is. Omwille van het principe van schaarsheid is het daarom niet wenselijk om nog verder predictoren toe te voegen. Er worden dan ook 20 predictoren meegenomen in de verdere stappen.



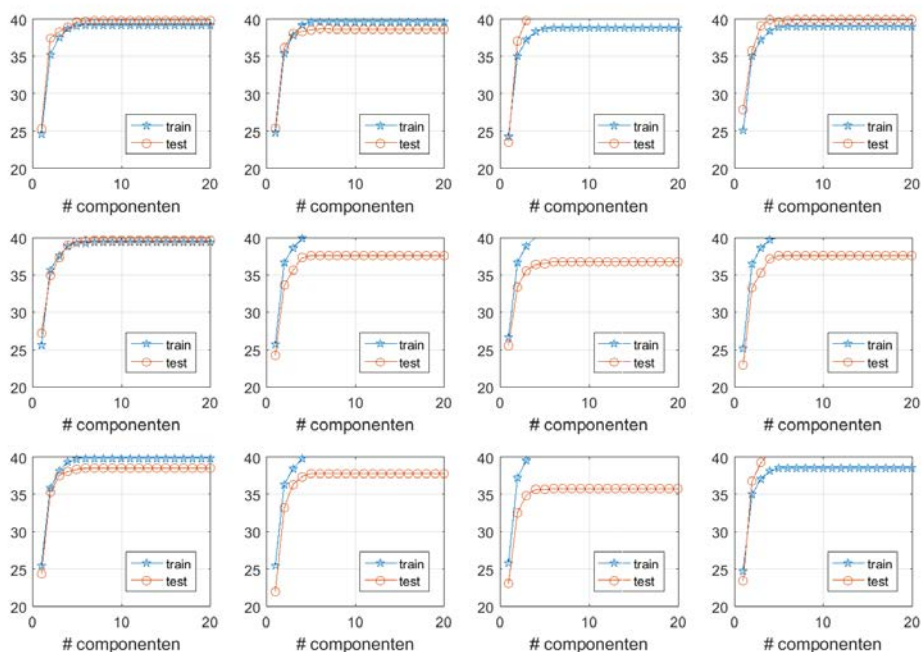
*Figuur 3-4: Volgorde van de predictoren zoals bepaald door stepwise MVR voor nitraat, oppervlaktewater, afstroomgebied. Op de x as zijn alle predictoren weergegeven. De volgorde van de relevante predictoren wordt verder toegelicht*



*Figuur 3-5: aantal predictoren ifv verklarende variantie voor nitraat, oppervlaktewater, afstroomgebied*

### 3.4 AANTAL COMPONENTEN

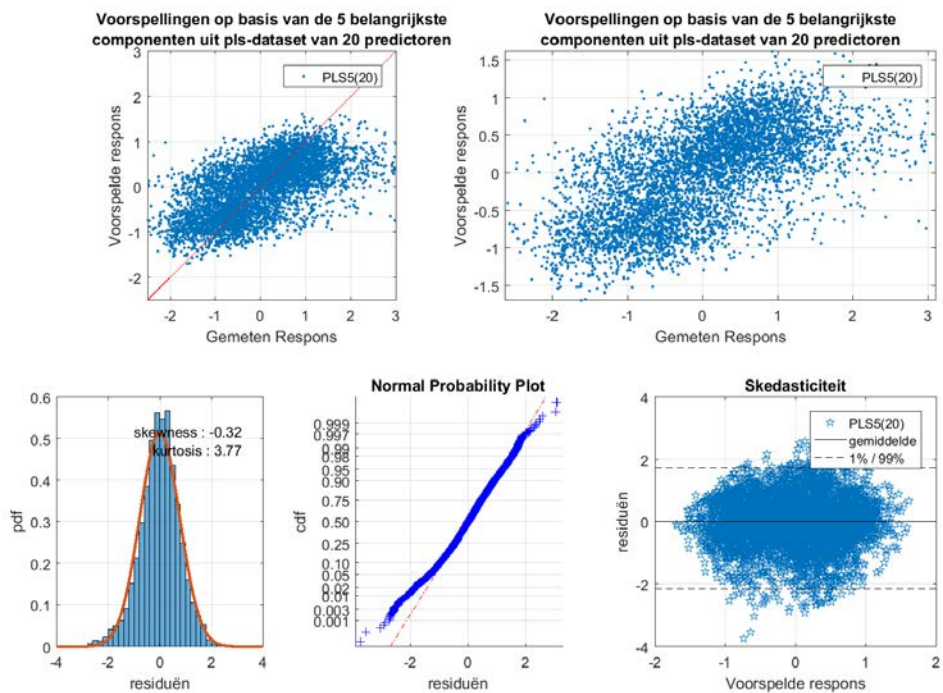
Nu het aantal predictoren is vastgelegd kan het optimaal aantal componenten voor een robuust model bepaald worden. De dataset is opgedeeld in 12 training en 12 test datasets zodoende mogelijke random effecten ten gevolge van de opdeling te minimaliseren. Vervolgens wordt stap voor stap een component toegevoegd aan de PLS fit. Het PLS model wordt 12 maal gefit op de training datasets en geëvalueerd op de test datasets (Figuur 3-6). Wederom wordt gezocht naar het punt waarop het meenemen van een extra component in de vergelijking geen meerwaarde meer biedt in functie van de verklarende variantie. In het onderstaande voorbeeld ligt dit knippunt op 5 componenten.



*Figuur 3-6: Aantal componenten ivf de verklaarde variantie voor nitraat, oppervlaktewater, afstroomgebied*

### 3.5 OPTIMAAL MODEL

Tenslotte wordt het finale model, met het geselecteerde aantal predictoren en componenten, opgesteld op basis van de volledige dataset om zo de maximale hoeveelheid beschikbare informatie mee te nemen en toch een robuust model te verkrijgen (Figuur 3-7). De verklaarde variantie voor de respons van dit model toegepast op alle data bedraagt 39.4 %. Uit de scatter plot blijkt dat het model de hogere gemeten waarden onderschat en de lagere gemeten waarden globaal overschat. De voorspelde wolk ligt bijgevolg geknikt ten opzichte van de bissectrice. Wel blijkt dat een redelijk groot deel van de variantie verklaard kan worden met de meegenomen predictoren en dat hieruit dus de nodige conclusies getrokken kunnen worden. De veronderstelling dat de residuen normaal verdeeld zijn is geldig.



Figuur 3-7: Optimaal PLS model voor oppervlakte water op afstroom niveau en respons nitraat met 5 componenten en 20 predictoren

## 4. RESULTATEN

Bovenstaande werkwijze is toegepast op de responsvariabelen nitraat, fosfaat alsook op de interne toestandsvariabele nitraatresidu. Voor iedere variabele zijn verschillende ruimtelijke aggregatieniveaus beschouwd zoals beschreven in rapport 1: verkennende data-analyse (I/RA/11504/17.100/TFR/).

In dit hoofdstuk worden de resultaten beschreven voor iedere combinatie van responsvariabelen en ruimtelijk aggregatieniveau. Telkens worden volgende gegevens getoond:

- Optimaal aantal predictoren;
- Optimaal aantal componenten;
- Verklarende variantie van het model;
- Overzicht optimaal PLS model. Hierbij worden telkens volgende figuren getoond:
  - Vergelijking van voorspelde en gemeten respons in een scatterplot met en zonder gestandaardiseerde assen;
  - Histogram van de residuen met weergave van coëfficiënt voor de scheefheid van de verdeling (skewness) en het staartgedrag (kurtosis);
  - Cumulatieve distributiefunctie van de residuen;
  - Scatterplot van residu in functie van de voorspelde respons ter validatie van de voorwaarde van homoskedasticiteit of homogeniteit van de variantie;
- Tabel met overzicht van geselecteerde predictoren met bijhorende coëfficiënten. De predictoren worden hierbij gerangschikt volgens prioriteit zoals bepaald door MVR;
- Ruimtelijke verdeling van het genormaliseerde statistisch residu. Het residu is hierbij gedefinieerd als de gemodelleerde concentratie – gemeten concentratie. Een positieve waarde impliceert m.a.w. een overschatting van het model, een negatieve waarde een onderschatting.

In Bijlage C zijn per combinatie respons variabele – ruimtelijk aggregatieniveau ook nog volgende figuren te vinden:

- Selectie optimaal model
- Volgorde predictoren
- Selectie aantal predictoren
- Selectie aantal componenten

Een samenvatting en bespreking van de resultaten volgt in hoofdstuk 5.

## 4.1 NITRAAT

### 4.1.1 Oppervlaktewater

#### 4.1.1.1 Afstroomgebied

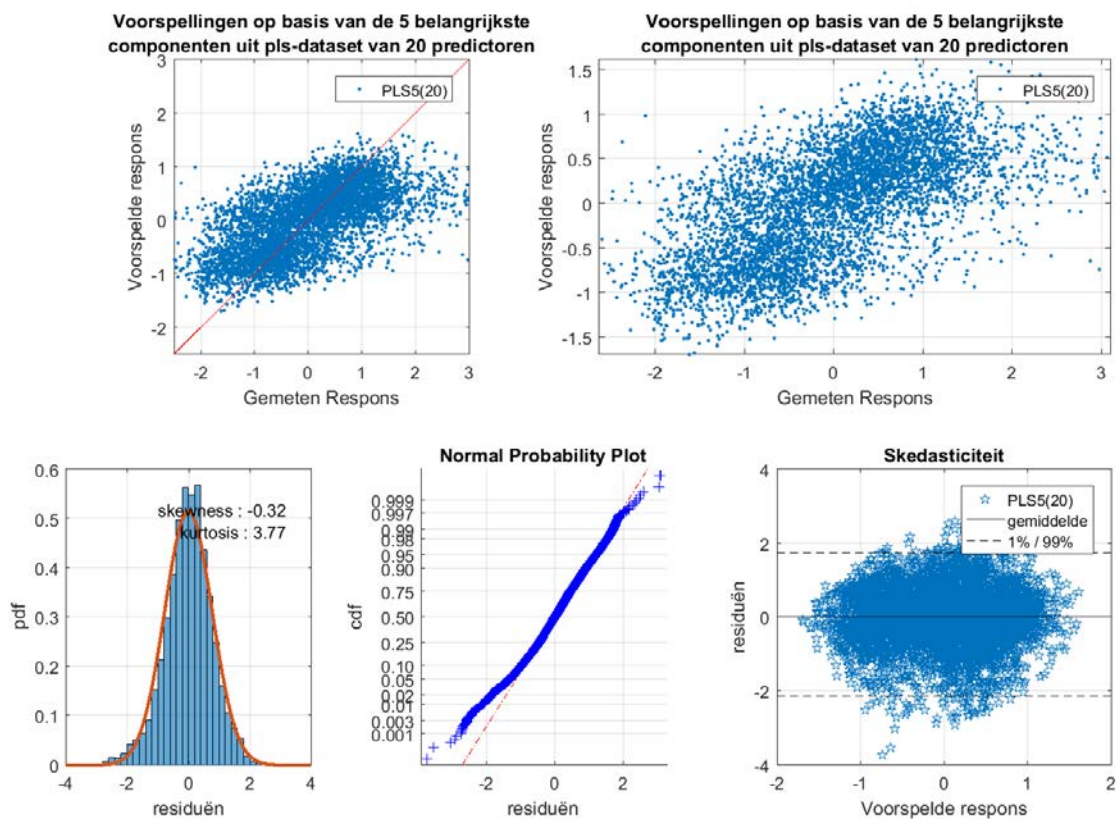
Tabel 4-1: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraat in oppervlaktewater op niveau afstroomgebied.

<b>Aantal predictoren</b>	20
<b>Aantal componenten</b>	5
<b>Verklarende variantie</b>	39.5 %

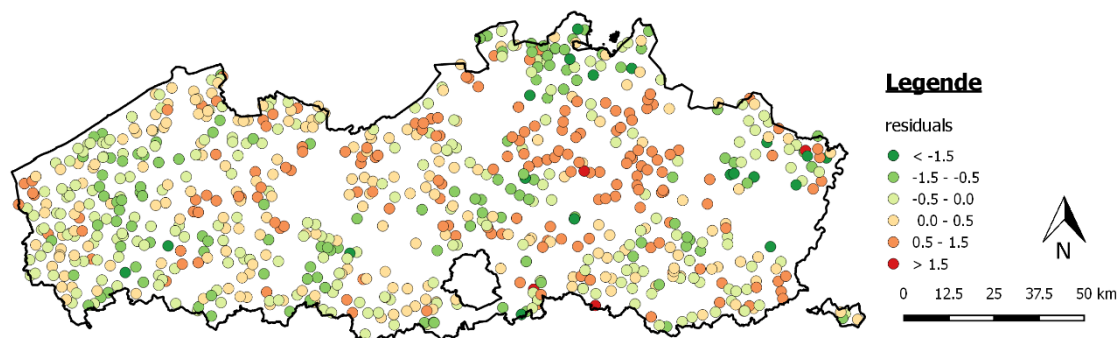
Tabel 4-2: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraat in oppervlaktewater op niveau afstroomgebied:

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.309
Elektrische geleidbaarheid	0.301
Procesfactor	-0.198
Zandleemstreek	0.133
Emissieverlies N	0.176
Hoofddeelt grassen	-0.143
Weidestreek	0.168
Neerslag maand	0.077
Erosiegevoeligheid licht	-0.081
Infiltratie coëfficiënt	0.146
Aangevoerde N schuimaarde	0.079
Oppervlakte percelen	-0.060
Voortelt Groenten groep II	0.056
Productiemethode PLA	0.045
Kempen	-0.084
Balans N spuistroom	0.050
Aanvoer N groencompost	-0.048
Nateelt Leguminosen	-0.050
Netto productie N varkens	-0.025
Erosiegevoeligheid medium	-0.048
Overschot N werkzaam	-0.040



Figuur 4-1: Optimaal PLS model voor nitraat in oppervlaktewater op niveau afstroomgebied.



Figuur 4-2: Ruimtelijke verdeling van statistisch residu voor nitraat in het oppervlaktewater op niveau afstroomgebied

#### 4.1.1.2 Gemeente

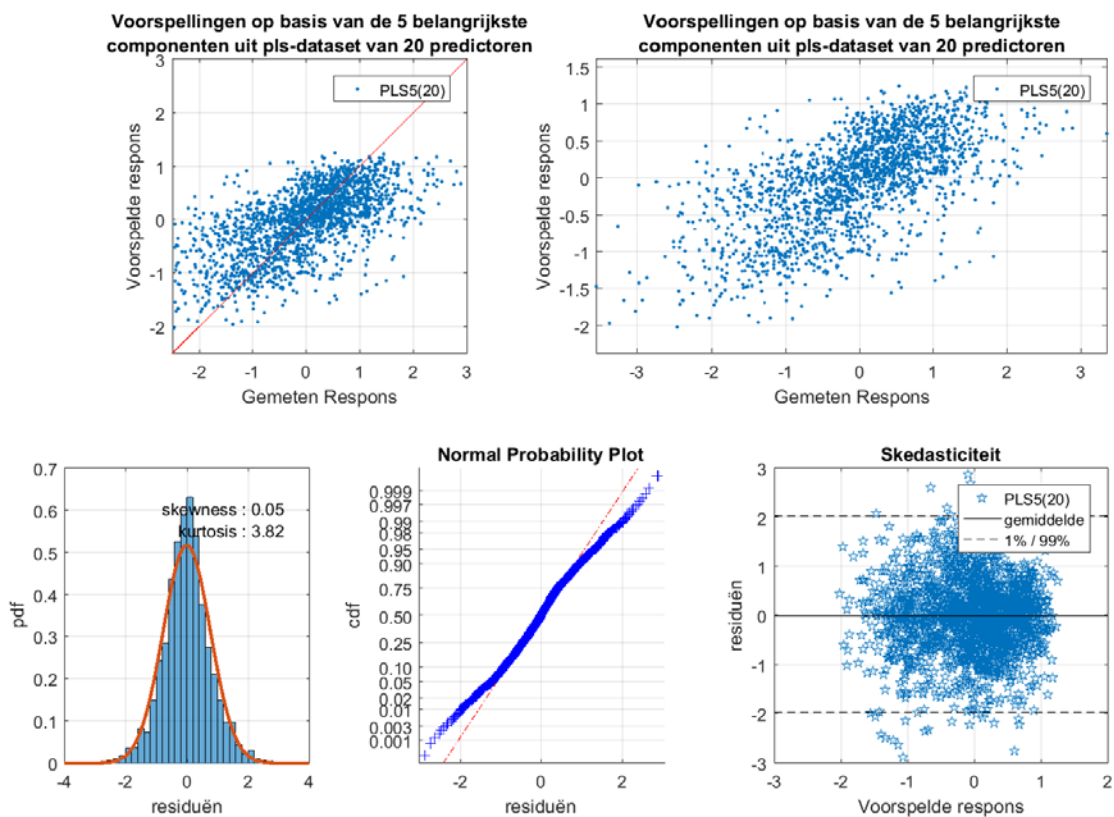
Tabel 4-3: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraat in oppervlaktewater op niveau gemeente.

<b>Aantal predictoren</b>	20
<b>Aantal componenten</b>	5
<b>Verklarende variantie</b>	41.7 %

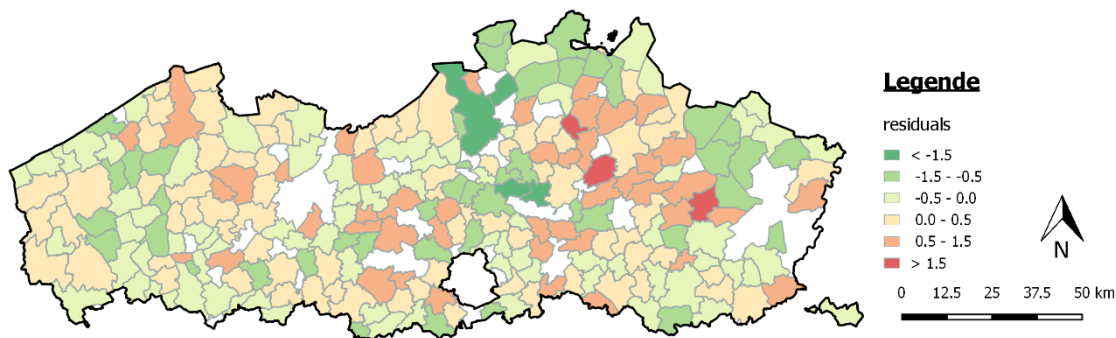
Tabel 4-4: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraat in oppervlaktewater op niveau gemeente:  $Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.108
Elektrische geleidbaarheid	0.119
Zandleemstreek	0.225
1 / Verbouwingsgraad	0.189
Overschot N totaal	0.093
Erosiegevoeligheid licht	-0.125
Bruto productie N andere dieren	-0.054
Neerslag maand	0.083
Procesfactor	-0.205
Kempen	-0.124
Balans N totaal	-0.031
Hoofddeelt Maïs	-0.129
Norm N dier	0.062
Leemstreek	-0.121
Landbouwstreek onbepaald	0.071
Productiemethode onbepaald	-0.032
Hoofddeelt Grassen	-0.021
Nateelt Grassen	-0.004
Balans N andere meststoffen	0.095
Emissieverlies N	0.132
Erosiegevoeligheid zeer hoog	-0.062





Figuur 4-3: Optimaal PLS model voor nitraat in oppervlaktewater op niveau gemeente



Figuur 4-4: Ruimtelijke verdeling van statistisch residu voor nitraat in het oppervlaktewater op niveau gemeente

#### 4.1.1.3 Bekken

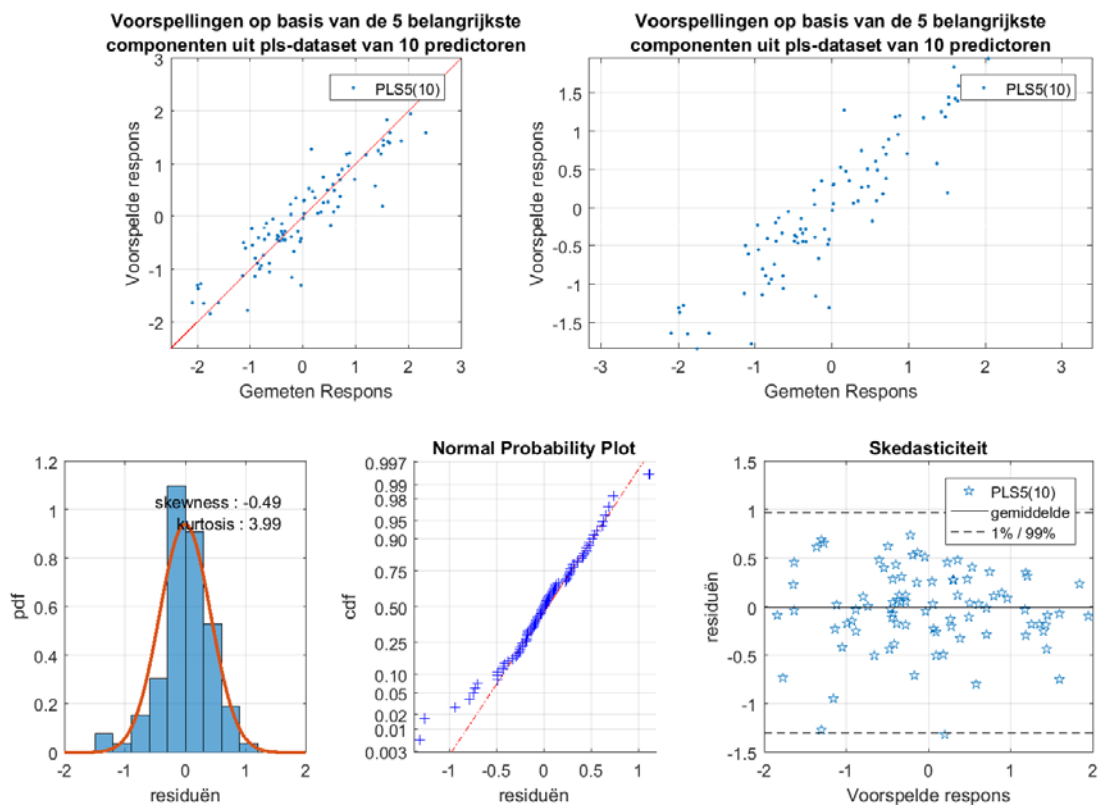
*Tabel 4-5: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraat in oppervlaktewater op niveau bekken*

<b>Aantal predictoren</b>	10
<b>Aantal componenten</b>	5
<b>Verklarende variantie</b>	82.3 %

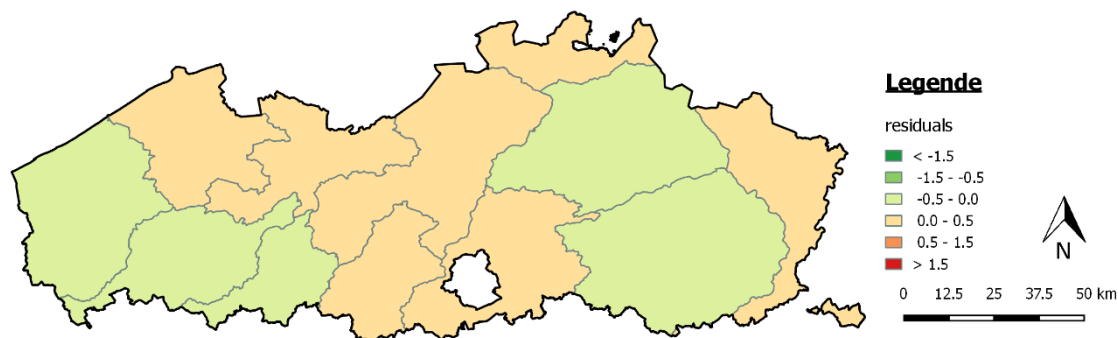
*Tabel 4-6: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraat in oppervlaktewater op niveau bekken:*

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.822
Zuurstofverzadiging	0.358
Netto Productie N varkens	0.414
Voorteelt Suikerbieten	0.056
Neerslag week	0.154
Hoofddeelt Gewassen met lage stikstofbehoefte	0.447
Productiemethode teelt niet op ruggen	0.371
Oppervlakte percelen	-0.368
Weidestreek	0.324
Orthofosfaat	0.298
Erosiegevoeligheid sterk	-0.164



Figuur 4-5: Optimaal PLS model voor nitraat in oppervlaktewater op niveau bekken



Figuur 4-6: Ruimtelijke verdeling van statistisch residu voor nitraat in het oppervlaktewater op niveau bekken

#### 4.1.1.4 Klassenverdeling

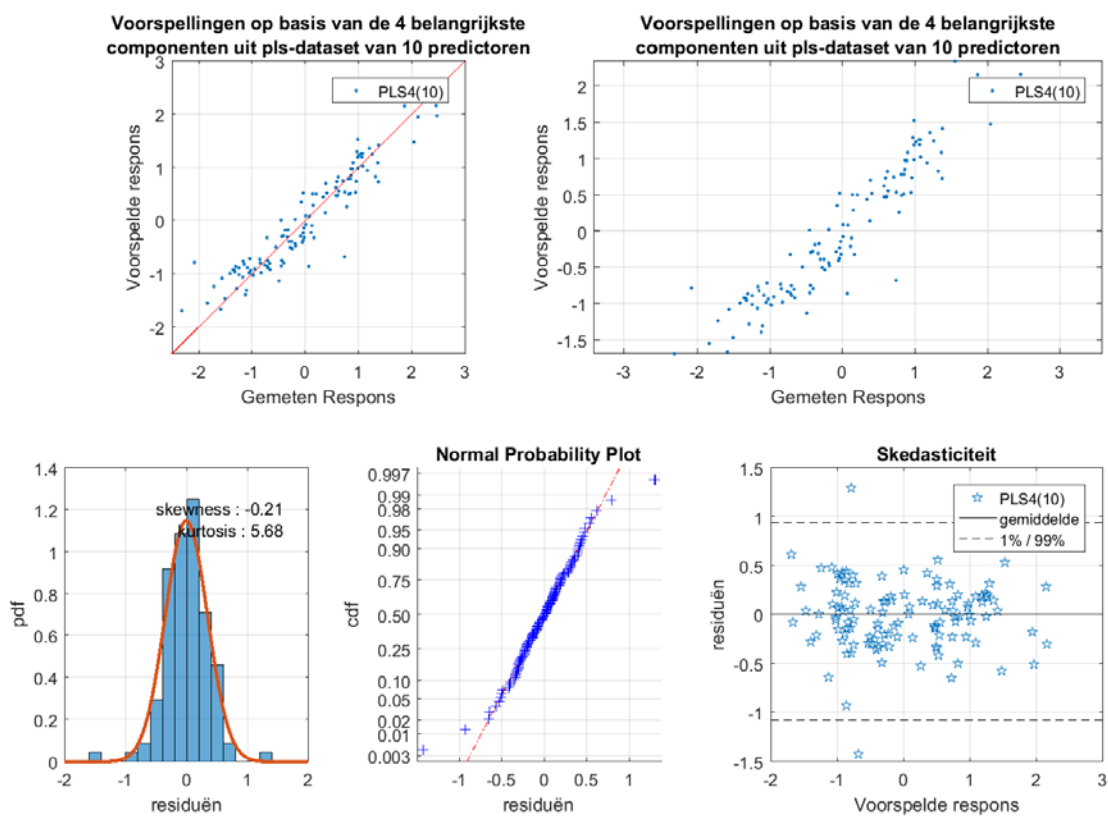
*Tabel 4-7: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraat in oppervlaktewater op niveau klassenverdeling*

<b>Aantal predictoren</b>	10
<b>Aantal componenten</b>	4
<b>Verklarende variantie</b>	88.0 %

*Tabel 4-8: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraat in oppervlaktewater op niveau klassenverdeling:*

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	-0.090
Emissieverlies N	0.472
Hoofdgewas grassen	-0.485
Leemstreek	-0.114
Orthofosfaat	0.274
Norm N Totaal	0.116
Weidestreek	0.120
Erosiegevoeligheid sterk	-0.191
Kempen	0.053
Erosiegevoeligheid zeer hoog	-0.158
Oppervlakte Percelen	-0.197



Figuur 4-7: Optimaal PLS model voor nitraat in oppervlaktewater op niveau klassenverdeling.

## 4.1.2 Grondwater

### 4.1.2.1 Intrekgebied

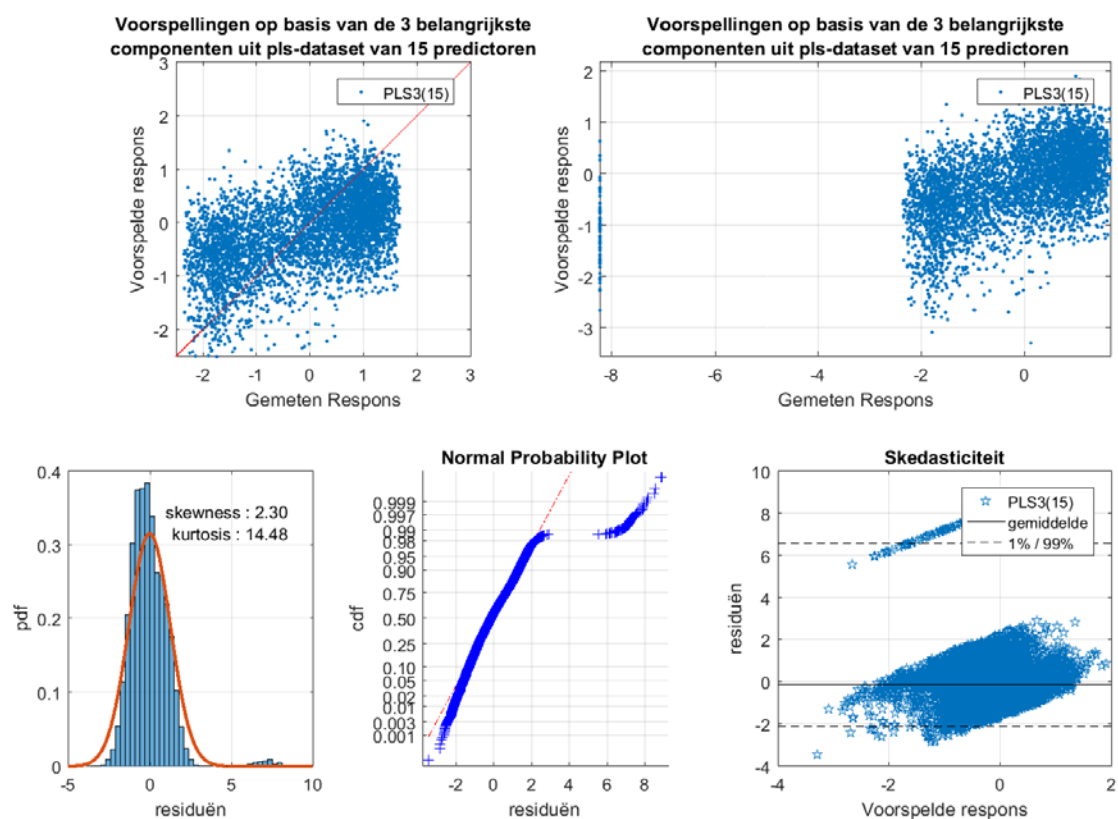
Tabel 4-9: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraat in grondwater op niveau intrekgebied.

Aantal predictoren	15
Aantal componenten	3
Verklarende variantie	24.1 %

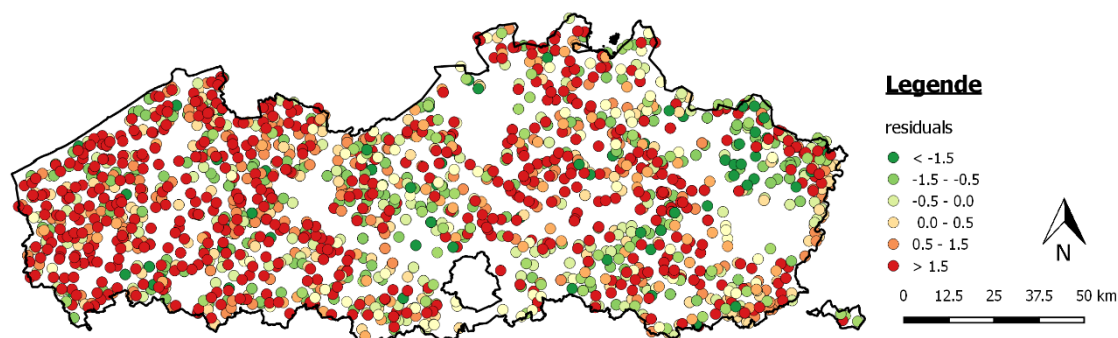
Tabel 4-10: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraat in grondwater op niveau intrekgebied:

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

predictor	coëfficiënt
Constante	-0.440
Gemiddelde grondwaterdiepte	0.424
Fosfaat	-0.334
Nitriet	0.169
Redoxpotentiaal	0.241
Kempen	-0.131
Geen BKM aangevraagd	0.021
Hoofddeelt Grassen	-0.177
Ijzer	0.015
Zuurtegraad	-0.079
Neerslag 3 Maand	-0.076
Elektrische geleidbaarheid	0.074
Nateelt Grassen	0.057
Thema Bodem zware klei	-0.177
Hoofddeelt Geen bemesting	-0.062
Bruto productie N runderen	-0.011



Figuur 4-8: Optimaal PLS model voor nitraat in grondwater op niveau intrekgebied.



Figuur 4-9: Ruimtelijke verdeling van statistisch residu voor nitraat in het grondwater op niveau intrekgebied

#### 4.1.2.2 Gemeente

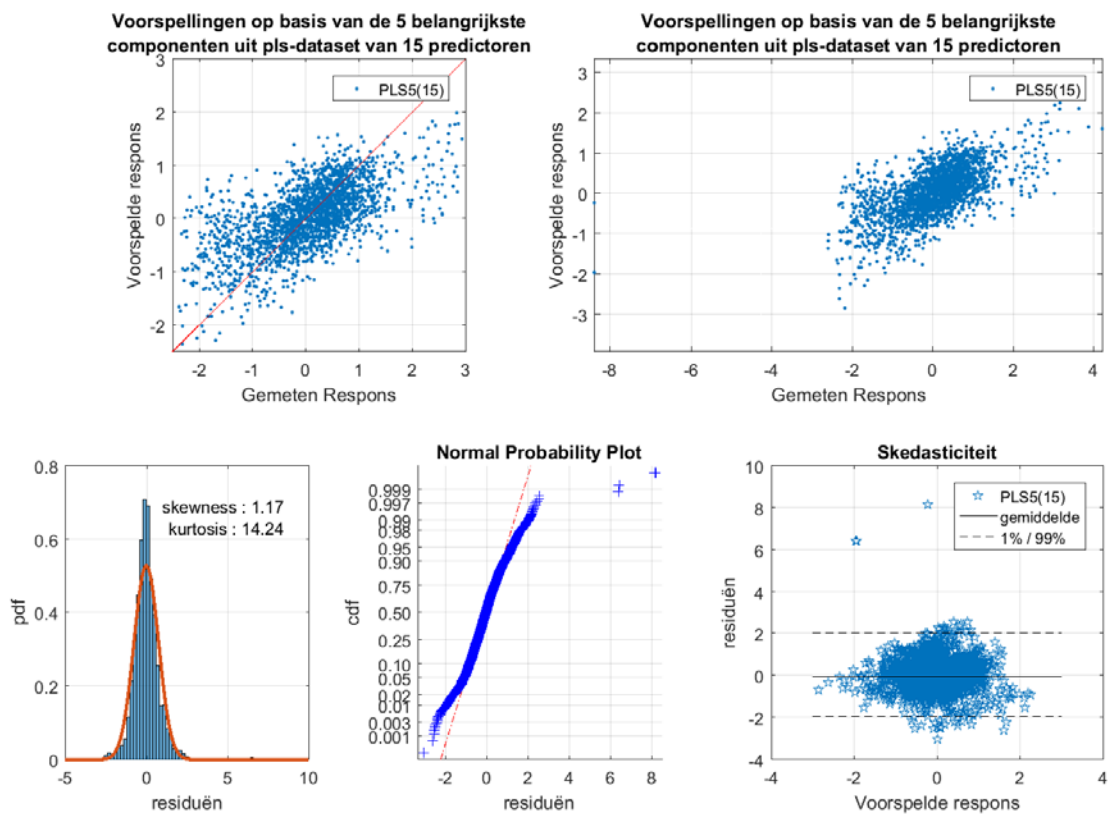
Tabel 4-11: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraat in grondwater op niveau gemeente

Aantal predictoren	15
Aantal componenten	5
Verklarende variantie	40.1 %

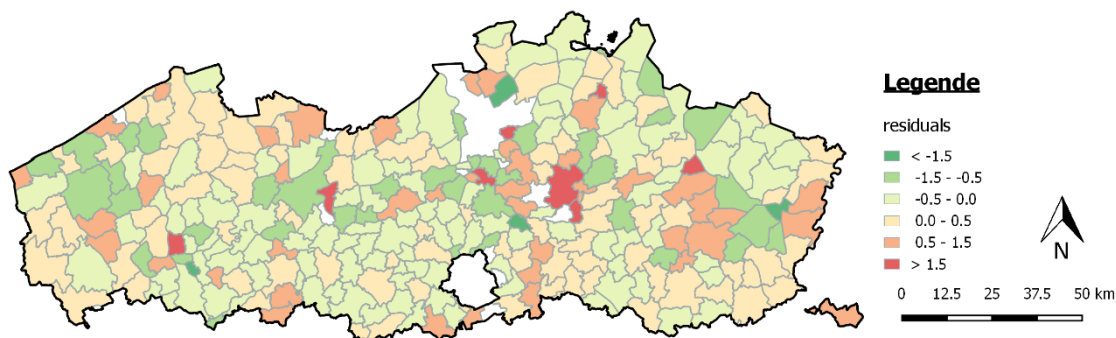
Tabel 4-12: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraat in grondwater op niveau gemeente:

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

predictor	coëfficiënt
Constante	0.076
Redoxpotentiaal	0.303
Gemiddelde grondwaterdiepte	0.230
Nitriet	0.302
Fosfaat	-0.119
Zandleemstreek	0.124
Bicarbonaat	-0.067
Hoofddeelt Maïs	0.158
Oppervlakte waarop teelt staat	-0.107
Netto Productie N andere dieren	0.054
Ijzer	-0.126
Sulfaat	0.172
Aangegeven N Kunstmest	0.067
Thema Bodem zware klei	-0.092
Hoofddeelt andere gewassen	0.042
Hoofddeelt onbekend	0.020



Figuur 4-10: Optimaal PLS model voor nitraat in grondwater op niveau gemeente.



Figuur 4-11: Ruimtelijke verdeling van statistisch residu voor nitraat in het grondwater op niveau gemeente



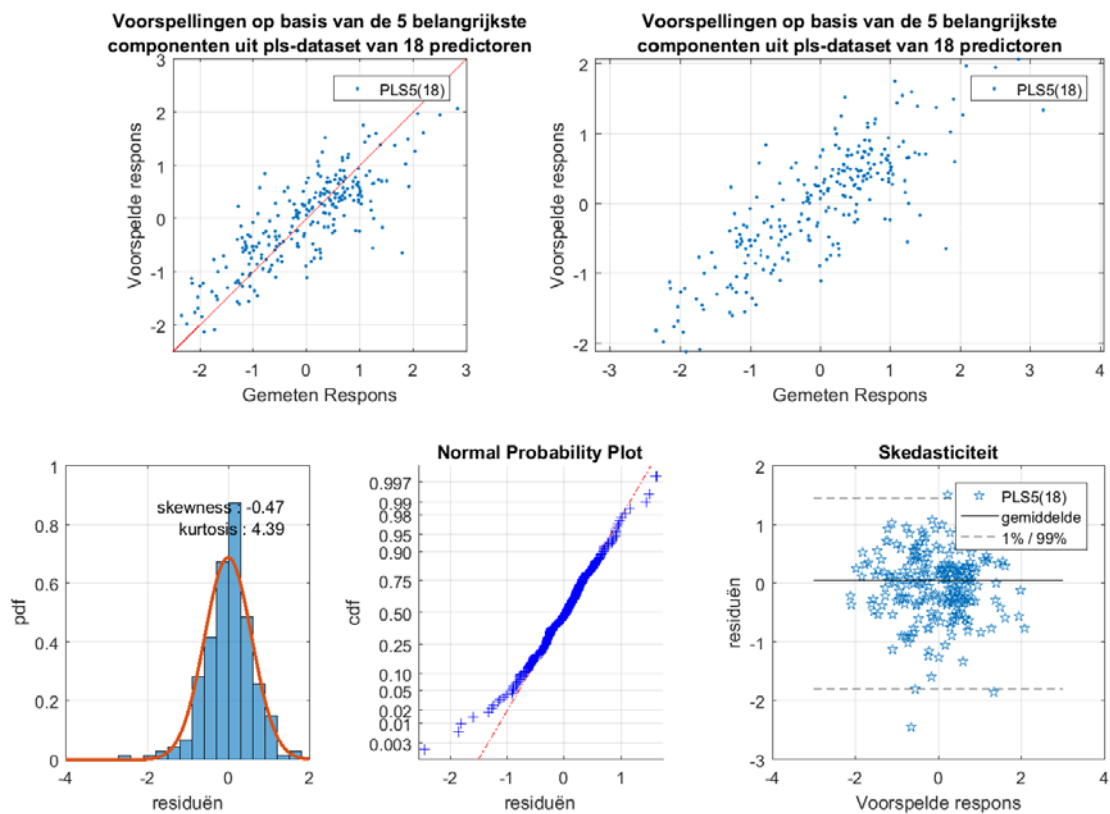
#### 4.1.2.3 HHZ

*Tabel 4-13: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraat in grondwater op niveau HHZ*

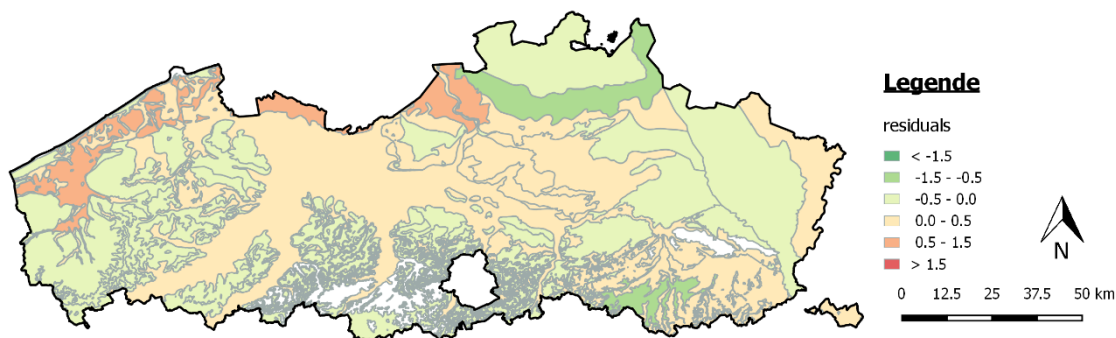
<b>Aantal predictoren</b>	18
<b>Aantal componenten</b>	5
<b>Verklarende variantie</b>	66.7 %

*Tabel 4-14: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraat in grondwater op niveau HHZ:  $Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$*

<b>predictor</b>	<b>coëfficiënt</b>
Constante	-0.351
Gemiddelde grondwaterdiepte	0.338
1 / Verbouwingsgraad	0.321
Fosfaat	-0.282
Nitriet	0.178
Hoofddeelt Maïs	-0.070
Productiemethode serres volle grond	0.126
Afvoer N andere meststoffen	0.106
Geen nateelt	0.051
Neerslag 3 Maand	-0.119
Balans N kunstmest	0.155
Aanvoer N totaal	-0.120
Aanvoer N andere meststoffen	0.153
Redoxpotentiaal	0.150
Zandleemstreek	0.130
Netto productie N pluimvee	-0.145
Erosiegevoeligheid niet gevoelig	0.024
Voorteeelt Leguminosen andere dan erwten	-0.057
Bruto Productie N andere dieren	0.079



Figuur 4-12: Optimaal PLS model voor nitraat in grondwater op niveau HHZ.



Figuur 4-13: Ruimtelijke verdeling van statistisch residu voor nitraat in het grondwater op niveau HHZ

## 4.2 FOSFAAT

### 4.2.1 Oppervlaktewater

#### 4.2.1.1 Afstroomgebied

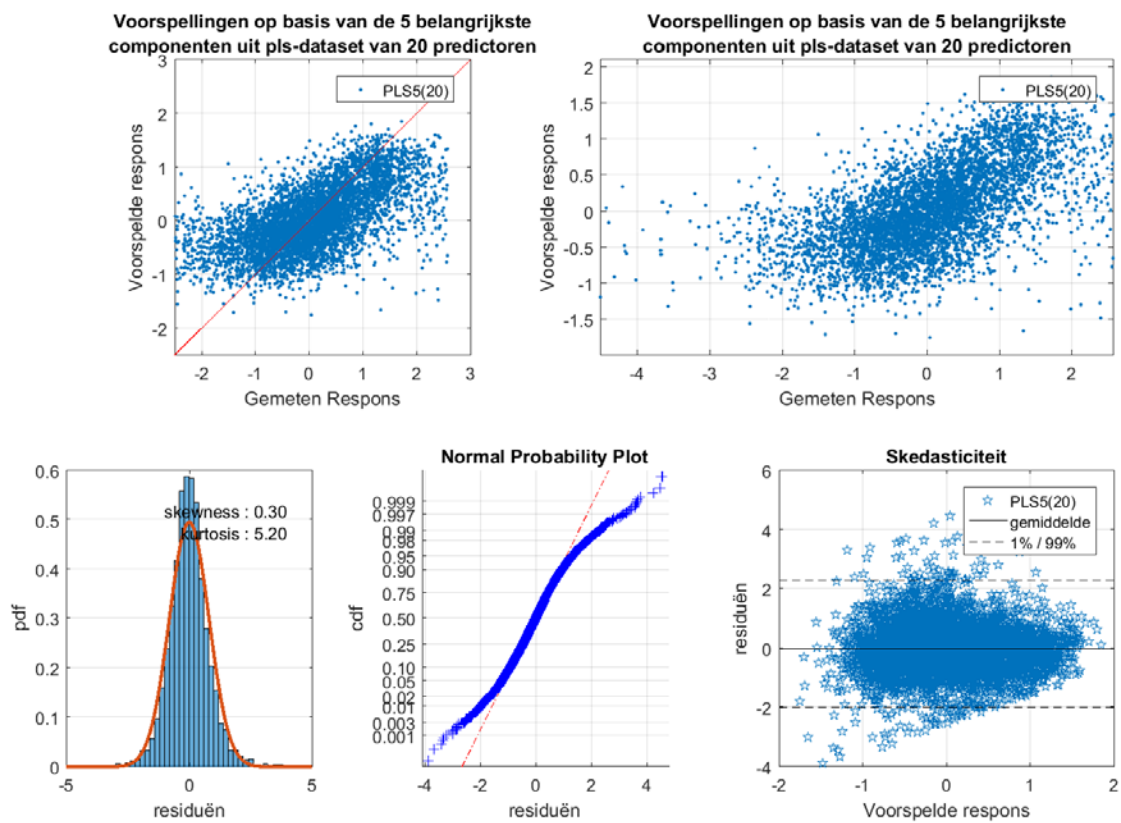
Tabel 4-15: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor fosfaat in oppervlaktewater op niveau afstroomgebied

<b>Aantal predictoren</b>	20
<b>Aantal componenten</b>	5
<b>Verklarende variantie</b>	35.1 %

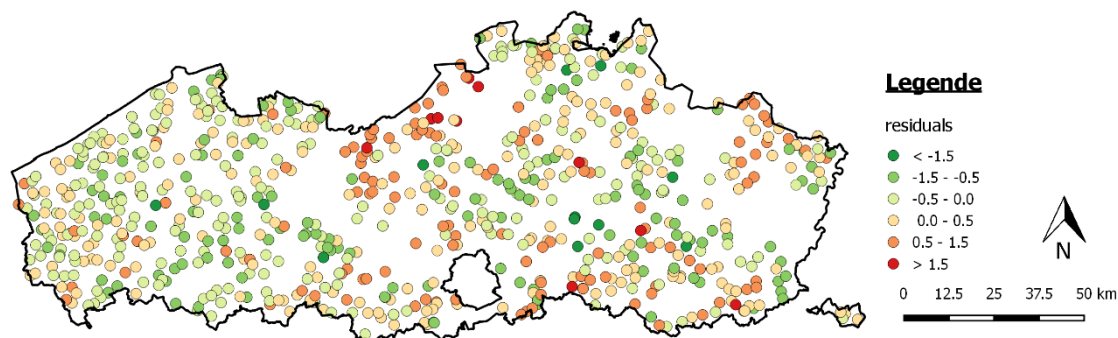
Tabel 4-16: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van fosfaat in oppervlaktewater op niveau afstroomgebied:

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.191
Nitraat	-0.292
Temperatuur	0.156
Kempen	-0.172
Aanvaarde oppervlakte	-0.124
Thema Bodem zware klei	0.150
Aangegeven P uit kunstmest	-0.048
Productiemethode serres groeimedium	0.066
Hoofdgewas groenten groep I	0.048
Afvoer percentiel	-0.018
Aangegeven P dierlijke mest	-0.061
1 / Verbouwingsgraad	0.056
Zandleemstreek	0.064
Nateelt andere gewassen	-0.044
Norm P Dier	-0.059
Hoofddeelt vanggewas	0.035
Nateelt wintertarwe en triticale	-0.061
Productie P totaal	0.047
Hoofddeelt Maïs	-0.092
Productie P paarden	-0.035
Productiemethode onbepaald	-0.028



Figuur 4-14: Optimaal PLS model voor fosfaat in oppervlaktewater op niveau afstroomgebied.



Figuur 4-15: Ruimtelijke verdeling van statistisch residu voor fosfaat in het oppervlaktewater op niveau afstroomgebied

#### 4.2.1.2 Gemeente

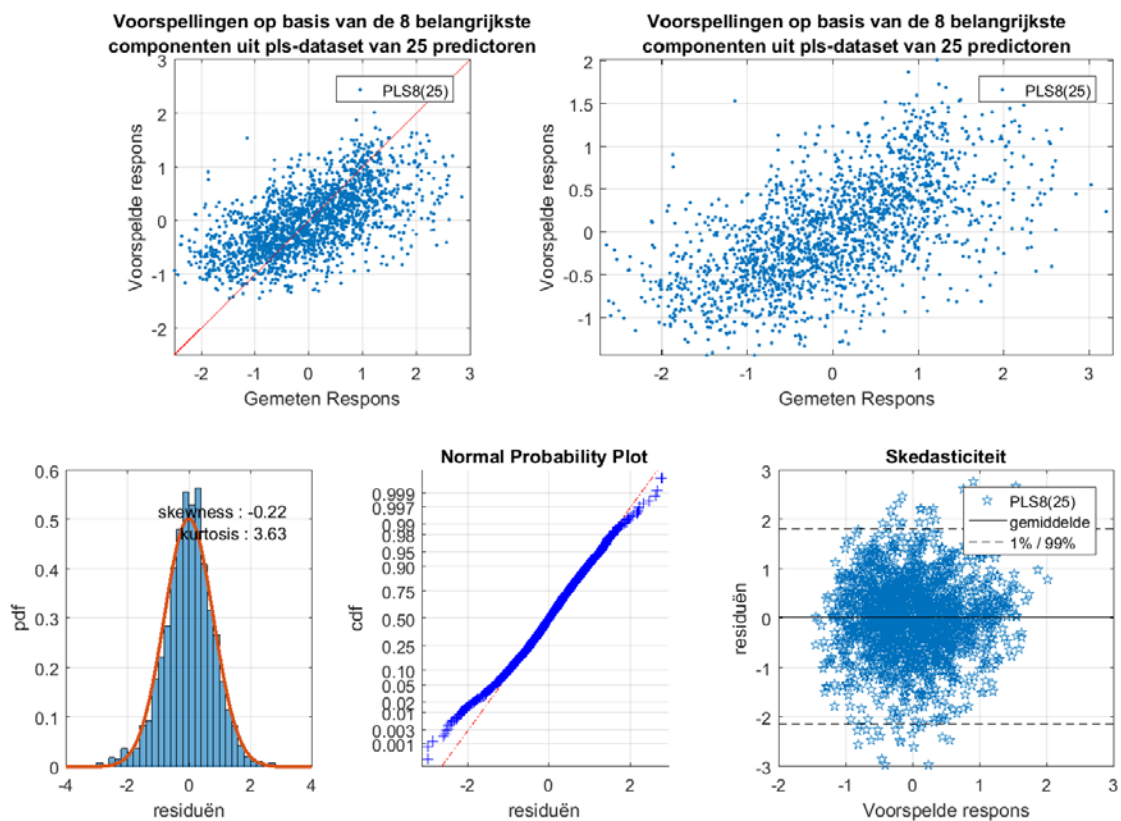
*Tabel 4-17: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor fosfaat in oppervlaktewater op niveau gemeente*

<b>Aantal predictoren</b>	25
<b>Aantal componenten</b>	8
<b>Verklarende variantie</b>	36.9 %

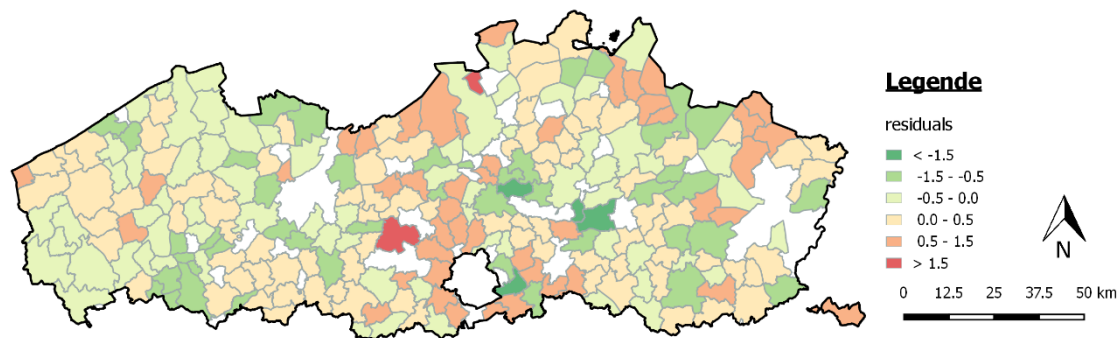
*Tabel 4-18: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van fosfaat in oppervlaktewater op niveau gemeente:*

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.210
Elektrische geleidbaarheid	0.060
Nitraat	-0.165
Productiemethode serres teelt volle grond	0.070
Aangegeven P uit kunstmest	-0.146
Thema bodem andere bodem	0.186
Vlaamse zandstreek	0.150
Balans P totaal	-0.135
Duinen	0.093
Productie P andere dieren	-0.057
Afvoer percentiel	-0.037
Landbouwstreek onbepaald	-0.045
Productiemethode plastic	0.071
Aanvaarde oppervlakte	0.061
Neerslag dag	0.042
Nateelt wintertarwe en triticale	-0.093
Hoofddeelt onbekend	0.028
Productiemethode teelt containers	0.050
Leemstreek	-0.066
Hoofddeelt maïs	-0.148
Weidestreek	-0.024
Productie P totaal	0.158
Productie P pluimvee	-0.089
Erosiegevoeligheid sterk	-0.053
Productie P spuistroom	0.052
Productie P paarden	0.061



Figuur 4-16: Optimaal PLS model voor fosfaat in oppervlaktewater op niveau gemeente.



Figuur 4-17: Ruimtelijke verdeling van statistisch residu voor fosfaat in het oppervlaktewater op niveau gemeente

#### 4.2.1.3 Bekken

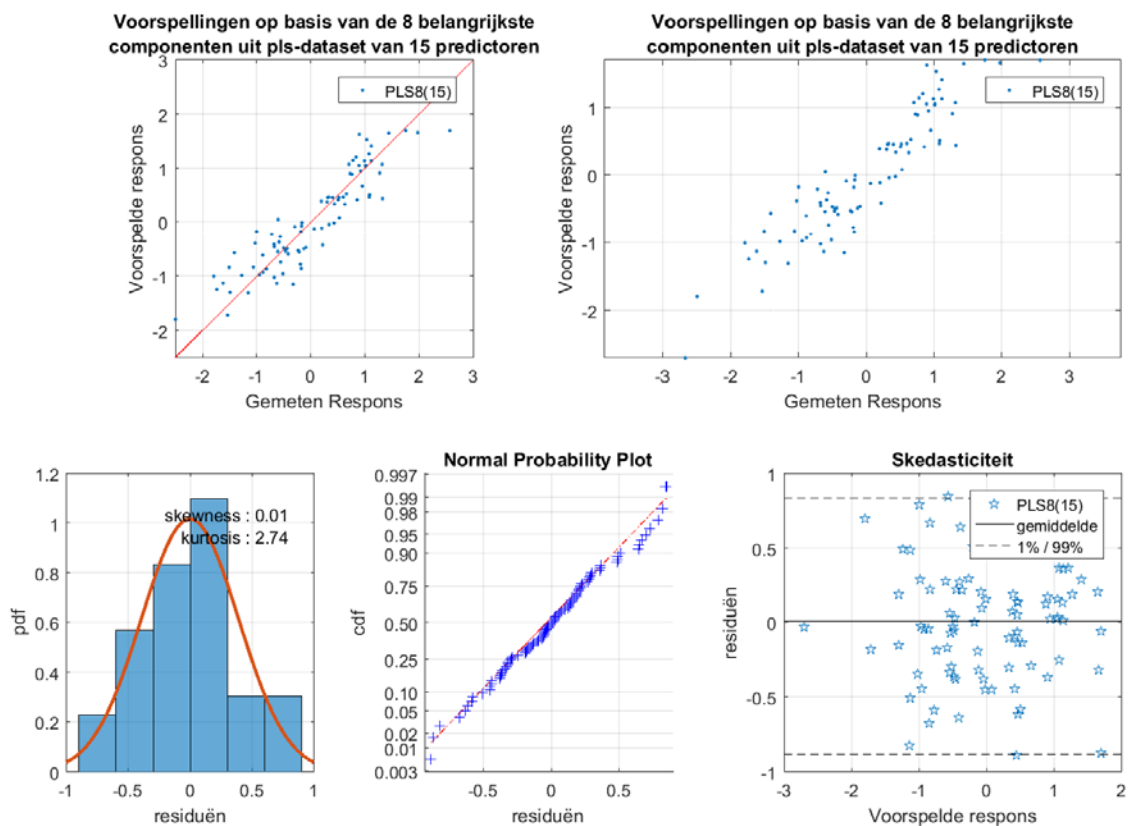
*Tabel 4-19: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor fosfaat in oppervlaktewater op niveau bekken*

<b>Aantal predictoren</b>	15
<b>Aantal componenten</b>	8
<b>Verklarende variantie</b>	84.6 %

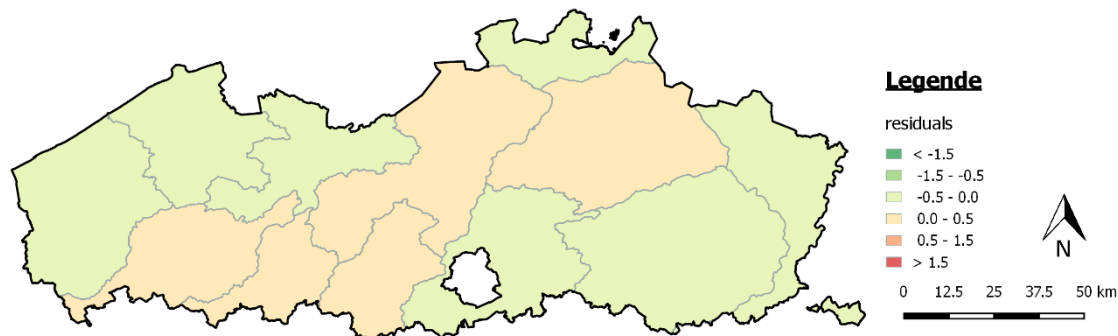
*Tabel 4-20: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van fosfaat in oppervlaktewater op niveau HHZ:*

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	-0.582
Elektrische geleidbaarheid	0.079
Kempen	-0.671
Productiemethode CON	0.288
Norm P kunstmest	-0.073
Productiemethode loods	0.123
Landbouwstreek onbepaald	0.105
Hoofddeelt Leguminosen andere dan erwten	0.005
Voortelt voederbieten	-0.231
Vlaamse zandstreek	0.640
Productiemethode NPO	0.071
Afvoer P andere meststoffen	0.344
Weidestreek	0.345
Thema bodem zware klei	-0.272
Productiemethode EBT	-0.136
Voortelt Maïs	-0.081



Figuur 4-18: Optimaal PLS model voor fosfaat in oppervlaktewater op niveau bekken.



Figuur 4-19: Ruimtelijke verdeling van statistisch residu voor fosfaat in het oppervlaktewater op niveau bekken



## 4.2.2 Grondwater

### 4.2.2.1 Intrekgebied

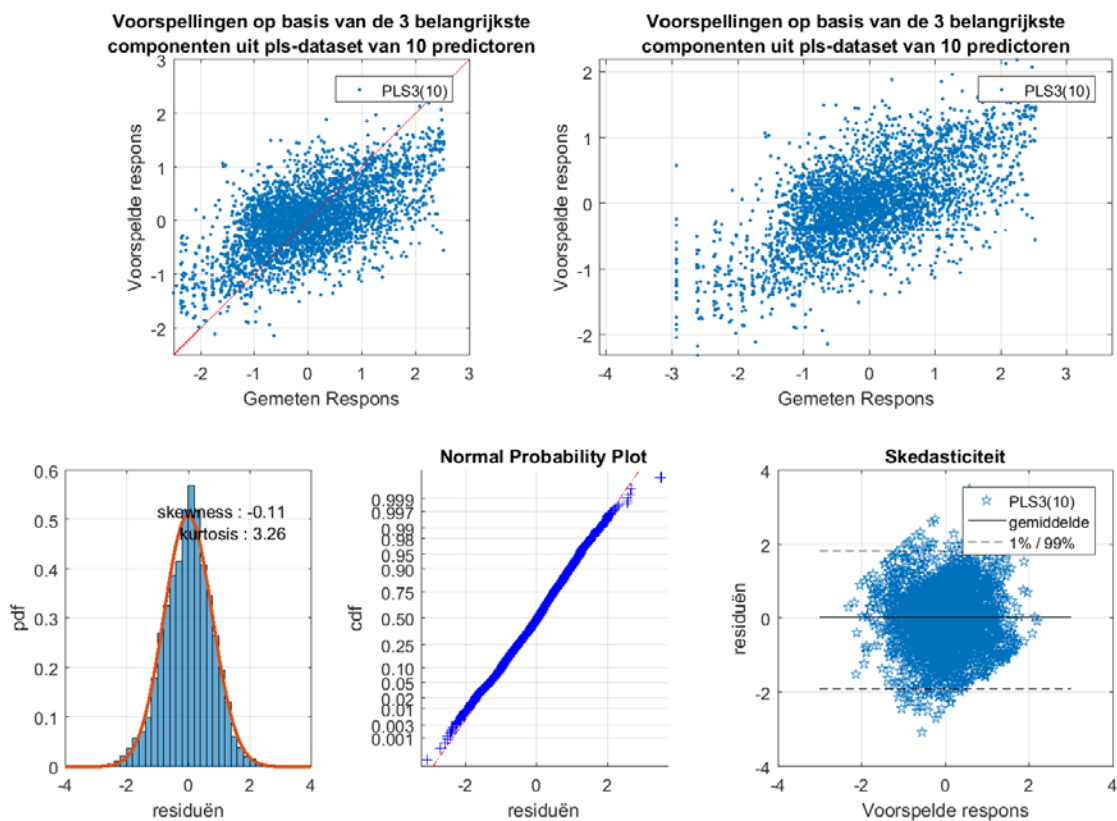
*Tabel 4-21: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor fosfaat in grondwater op niveau intrekgebied*

<b>Aantal predictoren</b>	10
<b>Aantal componenten</b>	3
<b>Verklarende variantie</b>	39.5 %

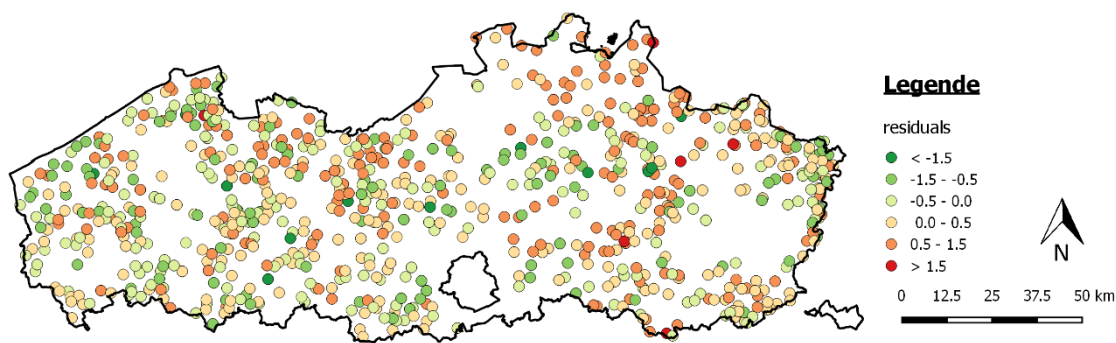
*Tabel 4-22: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van fosfaat in grondwater op niveau intrekgebied:*

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.006
Elektrische geleidbaarheid	0.141
Nitraat	-0.292
Nitriet	0.276
Thema bodem zware klei	0.159
Zuurtegraad	0.246
Kempen	-0.068
Geen derogatie aangevraagd	-0.107
Redoxpotentiaal	-0.086
Voortelt vanggewas	-0.040
Hoofddeelt groenten groep III	-0.052



Figuur 4-20: Optimaal PLS model voor fosfaat in grondwater op niveau intrekgebied.



Figuur 4-21: Ruimtelijke verdeling van statistisch residu voor fosfaat in het grondwater op niveau intrekgebied

#### 4.2.2.2 Gemeente

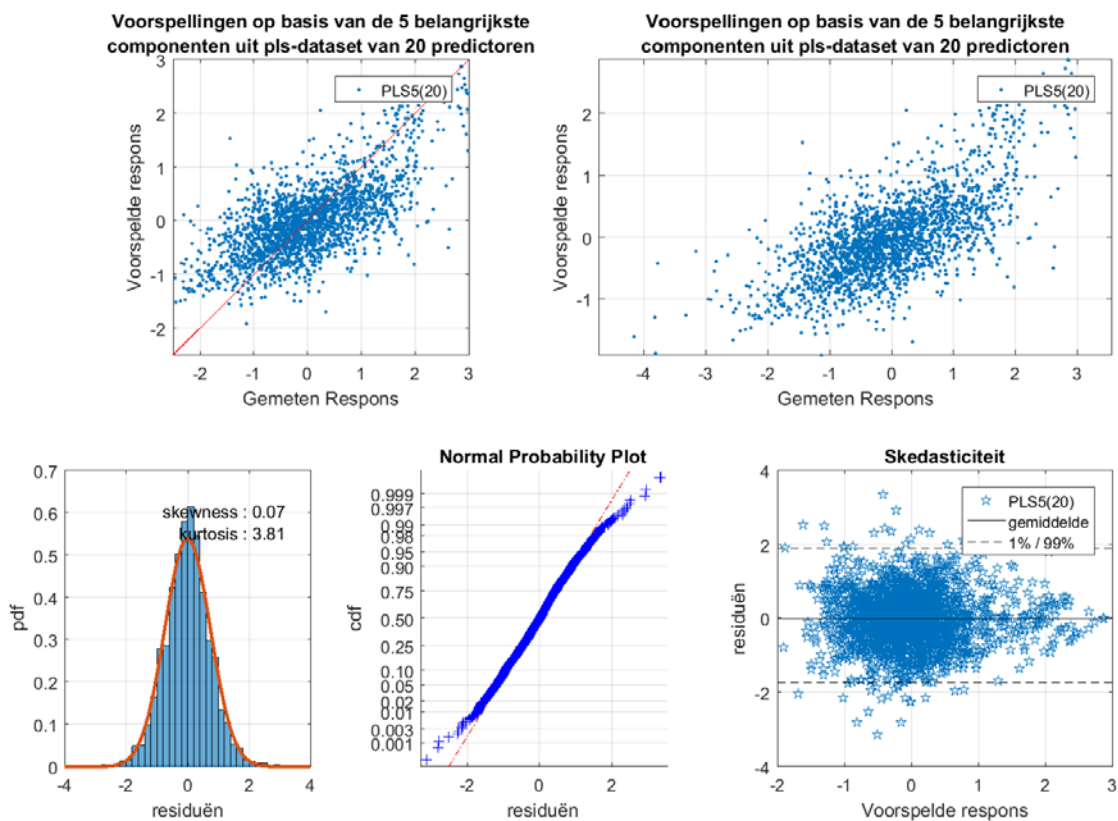
*Tabel 4-23: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor fosfaat in grondwater op niveau gemeente*

<b>Aantal predictoren</b>	20
<b>Aantal componenten</b>	5
<b>Verklarende variantie</b>	46.0 %

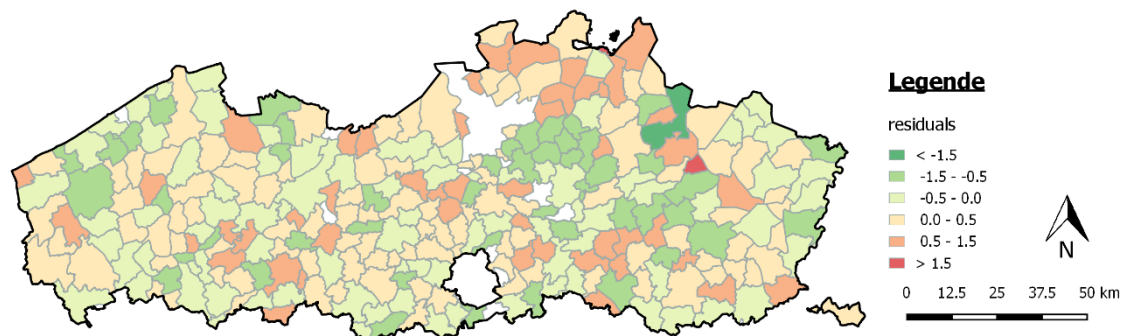
*Tabel 4-24: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van fosfaat in grondwater op niveau gemeente:*

$$Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$$

<b>predictor</b>	<b>coëfficiënt</b>
Constante	-0.183
Elektrische geleidbaarheid	0.222
Nitraat	-0.132
Nitriet	0.241
Duinen	0.145
Nateelt andere gewassen	-0.062
Leemstreek	-0.083
Hoofddeelt onbekend	-0.036
Productie P paarden	0.115
Thema bodem zware klei	0.117
Zuurtegraad	0.276
Redoxpotentiaal	-0.080
Ijzer	0.104
Voortelt andere granen	-0.018
Sulfaat	-0.133
Gemiddelde grondwaterdiepte	-0.058
Erosiegevoeligheid niet gevoelig	0.047
BKM goedgekeurd	-0.078
Landbouwstreek onbepaald	-0.058
Hoofddeelt Maïs	-0.077
Productie P varken	0.033



Figuur 4-22: Optimaal PLS model voor fosfaat in grondwater op niveau gemeente.



Figuur 4-23: Ruimtelijke verdeling van statistisch residu voor fosfaat in het grondwater op niveau gemeente

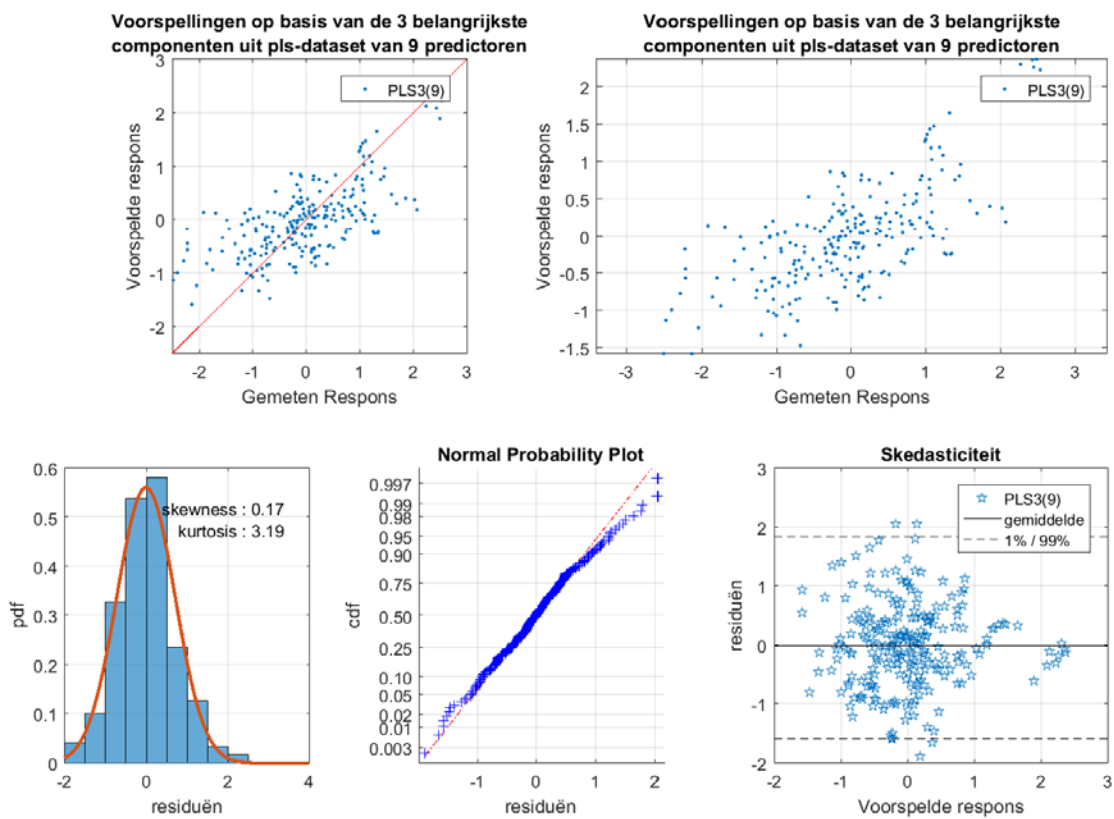
#### 4.2.2.3 HHZ

*Tabel 4-25: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor fosfaat in grondwater op niveau HHZ*

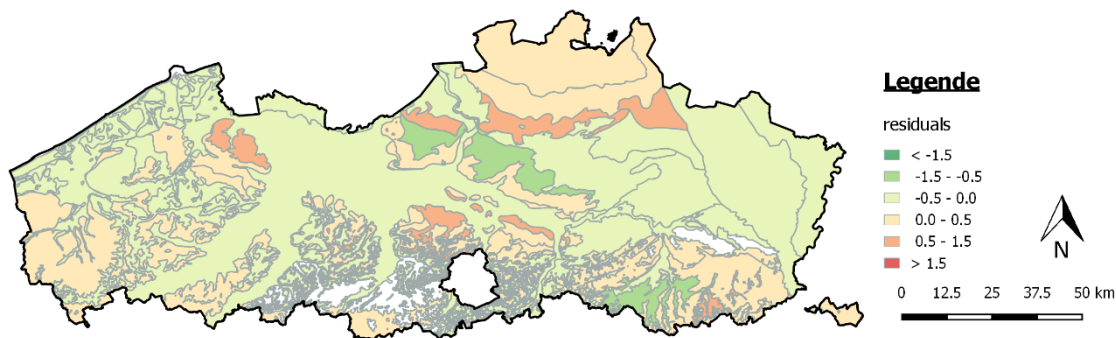
<b>Aantal predictoren</b>	9
<b>Aantal componenten</b>	3
<b>Verklarende variantie</b>	48.9 %

*Tabel 4-26: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van fosfaat in grondwater op niveau HHZ:  $Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$*

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.040
Thema bodem zware klei	0.275
Nitraat	-0.302
Nitriet	0.209
Productiemethode teelt containers	-0.172
Redoxpotentiaal	-0.167
Elektrische geleidbaarheid	0.217
Productiemethode NPO	-0.110
Productie P paarden	0.125
Leemstreek	-0.114



Figuur 4-24: Optimaal PLS model voor fosfaat in grondwater op niveau HHZ.



Figuur 4-25: Ruimtelijke verdeling van statistisch residu voor fosfaat in het grondwater op niveau HHZ

## 4.3 NITRAATRESIDU

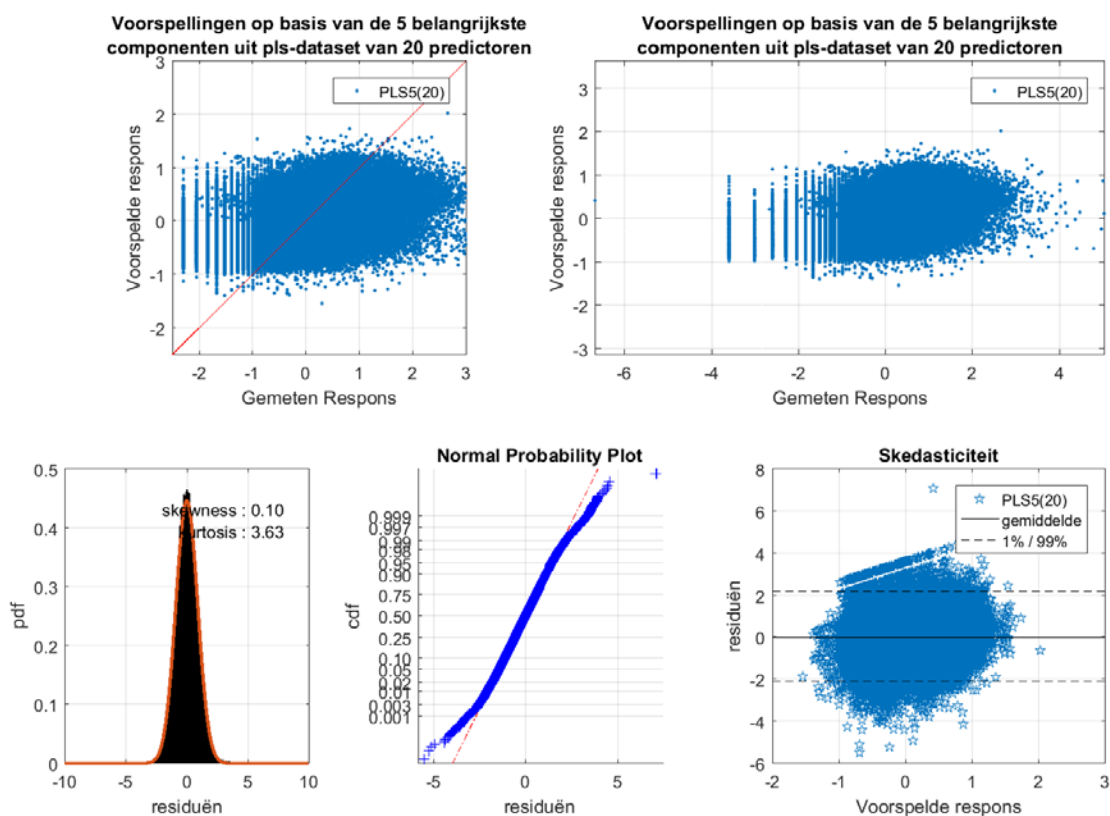
### 4.3.1 Perceel

*Tabel 4-27: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraatresidu op niveau perceel*

<b>Aantal predictoren</b>	20
<b>Aantal componenten</b>	5
<b>Verklarende variantie</b>	20.6 %

*Tabel 4-28: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraatresidu op niveau perceel:  $Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$*

<b>predictor</b>	<b>coëfficiënt</b>
Constante	0.080
Hoofddeelt maïs	0.340
Niet BKM goedgekeurd	0.099
Hoofddeelt groenten groep II	0.196
Totale N gebruikt	0.050
Hoofddeelt aardappelen geroid na 31 juli	0.179
Hoofddeelt andere gewassen	0.134
Thema bodem zware klei	0.078
Hoofddeelt groenten groep I	0.125
Norm N totaal	0.113
Hoofddeelt wintertarwe en triticale	0.190
Hoofddeelt groenten groep III	0.115
Overschot N dierlijke mest	0.064
Hoofddeelt Gewassen met lage stikstofbehoefte	0.086
Hoofddeelt andere granen	0.115
Nateelt andere gewassen	-0.090
Hoofddeelt geen vanggewas	0.010
Thema bodem zand	0.055
Hoofddeelt suikerbieten	0.043
Oppervlakte waarop teelt staat	-0.091
Afzetmogelijkheid N totaal	0.124



Figuur 4-26: Optimaal PLS model voor nitraatresidu op niveau perceel.

### 4.3.2 Gemeente

Tabel 4-29: Selectie aantal predictoren en componenten met bijhorende verklarende variantie voor nitraatresidu op niveau gemeente.

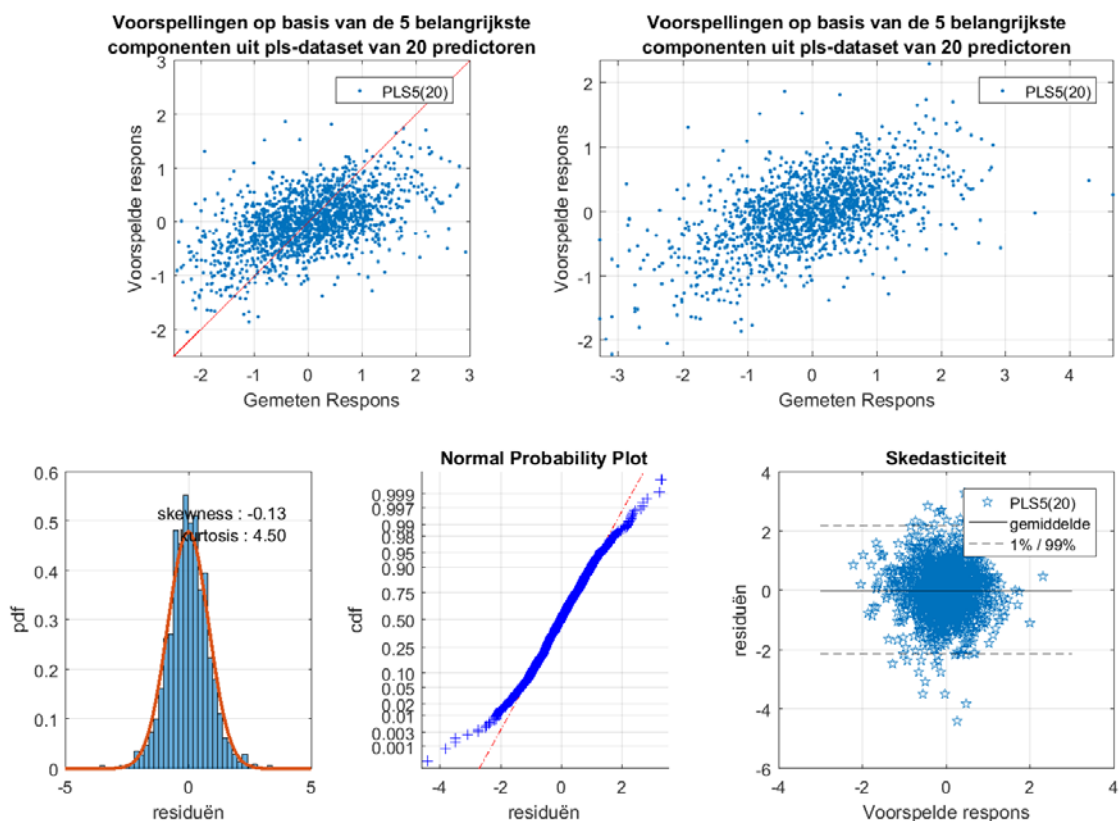
Aantal predictoren	20
Aantal componenten	5
Verklarende variantie	29.3

Tabel 4-30: Overzicht van de geselecteerde predictoren met bijhorende coëfficiënt voor het model van nitraatresidu op niveau gemeente:  $Y = \text{constante} + \sum (\text{coëfficiënt} * \text{predictor})$

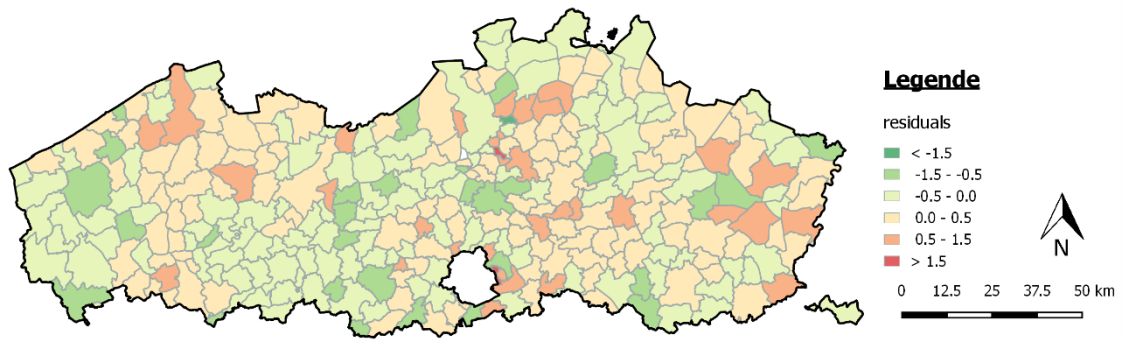
predictor	coëfficiënt
Constante	-0.076
Geen derogatie aangevraagd	-0.086
Hoofddeelt grassen	-0.064
Emissieverlies N	0.133
Thema bodem zand	-0.144
Erosiegevoeligheid andere	-0.080
Hoofddeelt Maïs	0.131
Overschot N werkzaam	0.115
Polders	0.059



predictor	coëfficiënt
Erosiegevoeligheid BO NK/DI	-0.043
Aanvoer N andere meststoffen	0.103
Hoofddeelt voederbieten	-0.076
Erosiegevoeligheid matig	0.081
Voortelt andere gewassen	-0.074
Duinen	0.073
Hoofddeelt groenten groep II	0.075
Hoofddeelt geen vanggewas	0.030
Overschot N dierlijke mest	-0.018
Opslagverschil N dierlijke mest	-0.042
Nateelt groenten groep III	-0.056
Voortelt gewassen met lage stikstofbehoefte	-0.045



Figuur 4-27: Optimaal PLS model voor nitraatresidu op niveau gemeente.



*Figuur 4-28: Ruimtelijke verdeling van statistisch residu voor nitraatresidu op niveau gemeente*

## 5. ANALYSE EN SAMENVATTING VAN RESULTATEN

Tabel 5-1 geeft een samenvatting van de resultaten van de globale analyse voor de verschillende responsvariabelen en ruimtelijke niveaus. Gemiddeld genomen stijgt het percentage verklarende variantie met opschaling van het ruimtelijk aggregatieniveau. Dit is te wijten aan het feit dat door opschaling veel variantie wordt uitgemiddeld.

Ondanks de grote onzekerheden op de afbakening van de afstroom- en intrekgebied en de verdeling van gegevens van exploitatieniveau naar perceel niveau, kunnen de modellen toch nog een significant gedeelte van de variantie op het laagste schaalniveau verklaren.

Op basis van de resultaten zoals weergegeven hoofdstuk 4 zien we dat de modellen voor de verschillende responsvariabelen grote verschillen vertonen. In onderstaande wordt daarom in eerste instantie de verschillende responsvariabelen afzonderlijk besproken. Vervolgens zal op gemeentelijk niveau de vergelijking gemaakt worden tussen de responsvariabelen onderling.

*Tabel 5-1: Overzicht resultaten globale statistische analyse*

Responsvariabele	Ruimtelijk niveau	# Predictoren	# Componenten	% VV	
Nitraat	OW	Afstroom	20	5	39
		Gemeente	20	5	41
		Bekken	10	5	82
		klasse	10	4	88
	GW	Intrek	15	3	24
		Gemeente	15	5	40
HHZ		18	5	67	
Fosfaat	OW	Afstroom	20	5	35
		Gemeente	25	8	37
		Bekken	15	8	85
	GW	Intrek	10	3	39
		Gemeente	20	5	46
		HHZ	9	3	49
Nitraatresidu	Perceel	20	5	21	
	Gemeente	20	5	29	

De interpretatie van het effect en het belang van iedere predictor op de responsvariabelen dient te gebeuren aan de hand van de tabellen met coëfficiënten die zijn weergegeven in hoofdstuk 4 in combinatie met de inzichten bekomen uit de bivariate analyse zoals beschreven in onderdeel 1. De weergegeven analyse hieronder poogt geenszins om alle bekomen modellen tot in detail te ontleden, maar dient eerder aanzien te worden als een voorbeeld en leidraad op basis waarvan interpretatie van deze tabellen kan gebeuren. Daarnaast worden enkele tendensen uit de resultaten weergegeven door middel van samenvattende figuren en tabellen.

## 5.1 NITRAAT, OPPERVLAKTEWATER

Figuur 5-1 toont de geselecteerde predictoren en resulterende coëfficiënten voor de verschillende ruimtelijke niveaus voor nitraat in het oppervlaktewater.

Hieruit blijkt dat er sterke verschillen zijn tussen de geselecteerde variabelen voor de verschillende aggregatieniveaus en dat er 11 variabelen zijn die in minimaal 2 van de 3 beschouwde aggregatieniveaus voorkomen. Het gaat om:

- Elektrische geleidbaarheid
- Erosiegevoeligheid licht
- Procesfactor
- Netto productie N varkens
- zandleemstreek
- Oppervlakte percelen
- Neerslag maand
- Emissieverlies N
- Weidestreek
- Hoofddeelt grassen
- Kempen

Hoewel er 41 variabelen voorkomen in de verschillende aggregatieniveaus kunnen deze wel worden teruggebracht tot een beperkte set van categorieën zoals weergegeven in Tabel 5-2.

*Tabel 5-2: Verschillende categorieën van geselecteerde predictoren voor Nitraat, oppervlaktewater*

Categorie	Predictoren
Gewasrotatie	Hoofddeelt maïs; Hoofddeelt grassen; Hoofddeelt Gewassen met lage stikstofbehoefte; ...
Exploitatiegegevens	Emissieverlies N; Netto productie N varkens; overschot N werkzaam; ...
Bodemgesteldheid	Procesfactor; Erosiegevoeligheid sterk; Zandleemstreek; Infiltratie coëfficiënt; ...
Waterkwaliteit	Elektrische geleidbaarheid; Orthofosfaat; Zuurstofverzadiging
Omgevingsfactoren	Neerslag maand; Neerslag week
Opgelegde normen	Norm N Dier

De predictoren met betrekking tot de landbouwzones zijn sterk vertegenwoordigd in de modelvergelijkingen. Dit is een indicatie dat er ruimtelijke verschillen zijn tussen de landbouwzones. Een differentiatie van de modelvergelijkingen volgens landbouwzones kan hier mogelijks een duidelijker beeld over scheppen.

Het valt op dat *NETTO PRODUCTIEN VARKENS* van teken verandert tussen het hoogste aggregatieniveau (bekken, positieve coëfficiënt) en het laagste aggregatieniveau (afstroomgebied, negatieve coëfficiënt). Hoewel een zekere variatie op de coëfficiënten binnen de verwachtingen valt, is het veranderen van teken niet logisch. De reden hiervoor dient echter gezocht te worden in de onderlinge afhankelijkheid tussen de predictoren. Bij het aggregatieniveau afstroomgebied is immers ook de predictor *EMISSIEVERLIEN N* opgenomen.

Deze vertonen onderling een sterke afhankelijkheid zoals geïllustreerd in Figuur 5-3. Een groot deel van de informatie die vervat zit in *NETTO PRODUCTIE N VARKENS* zit dus reeds vervat in de variabele *EMISSIEVERLIES N*. Een visuele vergelijking tussen de maximaal gemeten waarden voor nitraat in het oppervlaktewater en *EMISSIEVERLIES N* (Figuur 5-4) lijkt te suggereren dat er een verband is aangezien hogere emissieverliezen overwegend met hogere concentraties overeenstemmen. Dit verband is visueel uiteraard niet zeer duidelijk aangezien er meerdere predictoren een invloed uitoefenen op de gemeten concentraties. Er zijn echter een aantal punten waarneembaar waarvoor een hogere concentratie verwacht kan worden op basis van de waarden van *EMISSIEVERLIES N*. Deze zijn in de figuur aangeduid met een blauwe cirkel. Op basis van de kleurcode blijkt ook dat deze punten overeenstemmen met punten waarvoor de *NETTO PRODUCTIE N VARKENS* hoog is. De negatieve coëfficiënt dient hier dus bekeken te worden in combinatie met de overige parameters die meegenomen worden. Indien *EMISSIEVERLIES N* niet wordt meegenomen in de vergelijking zou de coëfficiënt voor *NETTO PRODUCTIE N VARKENS* immers positief zijn.

Dit voorbeeld, alsook het hoog aantal gecorreleerde variabelen dat op de verschillende niveaus wordt meegenomen, illustreert de moeilijkheid bij de interpretatie van de bekomen modellen. De selectie van de predictoren o.b.v. de MVR methodologie tracht zoveel mogelijk onafhankelijke variabelen te selecteren zoals toegelicht in hoofdstuk 3. Het is echter nog steeds mogelijk dat gecorreleerde variabelen geselecteerd worden indien deze nog bijkomende informatie bevatten. Door gebruik te maken van PLS elimineren we het probleem van de afhankelijkheid bij het opstellen van de modelvergelijkingen. Om de modellen te interpreteren, vertalen we de modelcomponenten van het PLS model echter terug naar de zuivere predictoren. Aangezien iedere component in deze vergelijking een lineaire combinatie is van alle meegenomen predictoren sluipt een mogelijke afhankelijkheid terug in onze vergelijking.

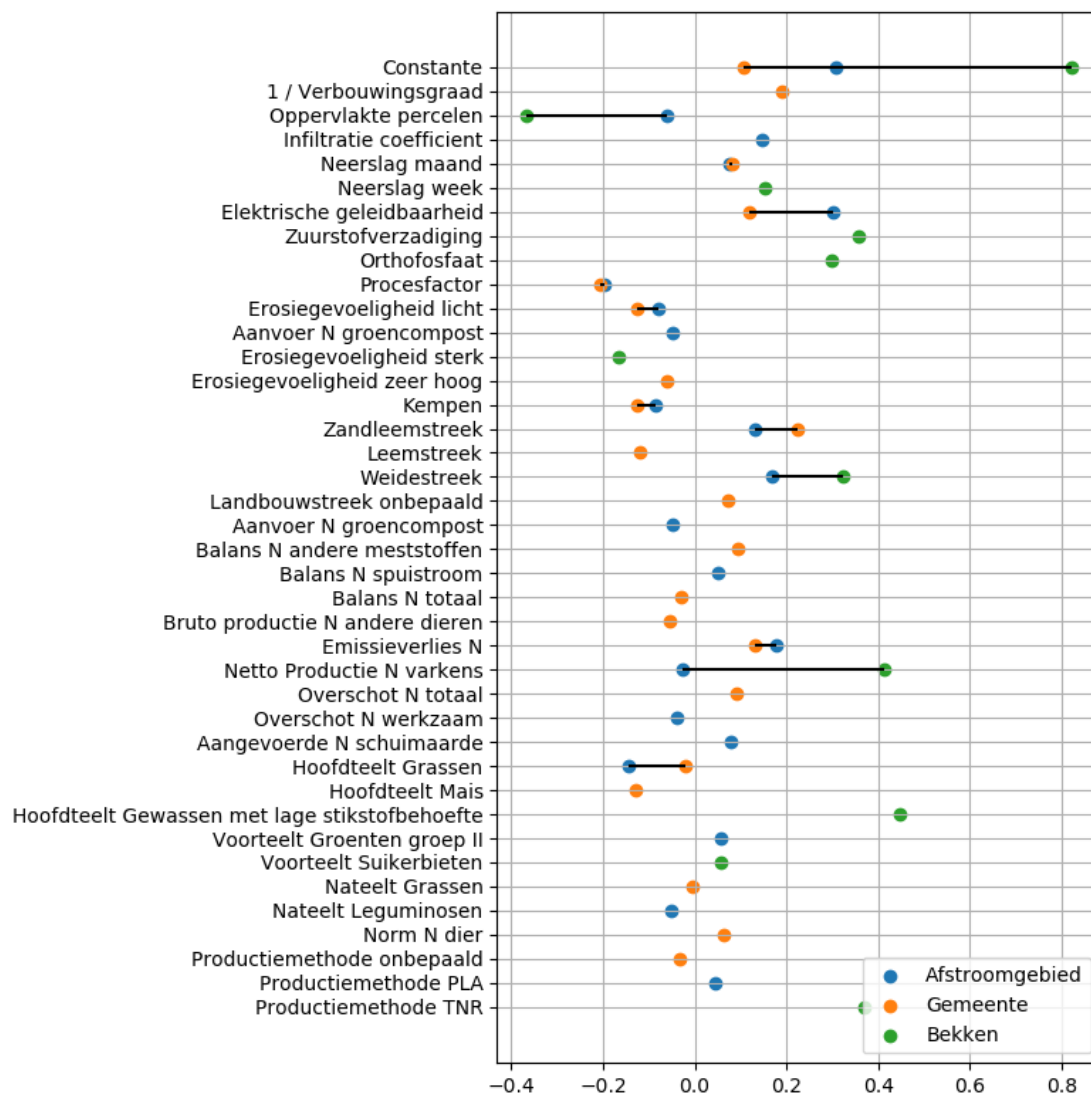
Dit impliceert dat de interpretatie van het belang van de predictoren en de bijhorende coëfficiënten niet eenduidig is en altijd in zijn geheel dient aanzien te worden.

De ruimtelijke verdeling van de statistische residuen vertoont een geclusterde overschatting van het model ter hoogte van het Netebekken voor het aggregatieniveau afstroomgebied en gemeente (Figuur 4-2 & Figuur 4-4).

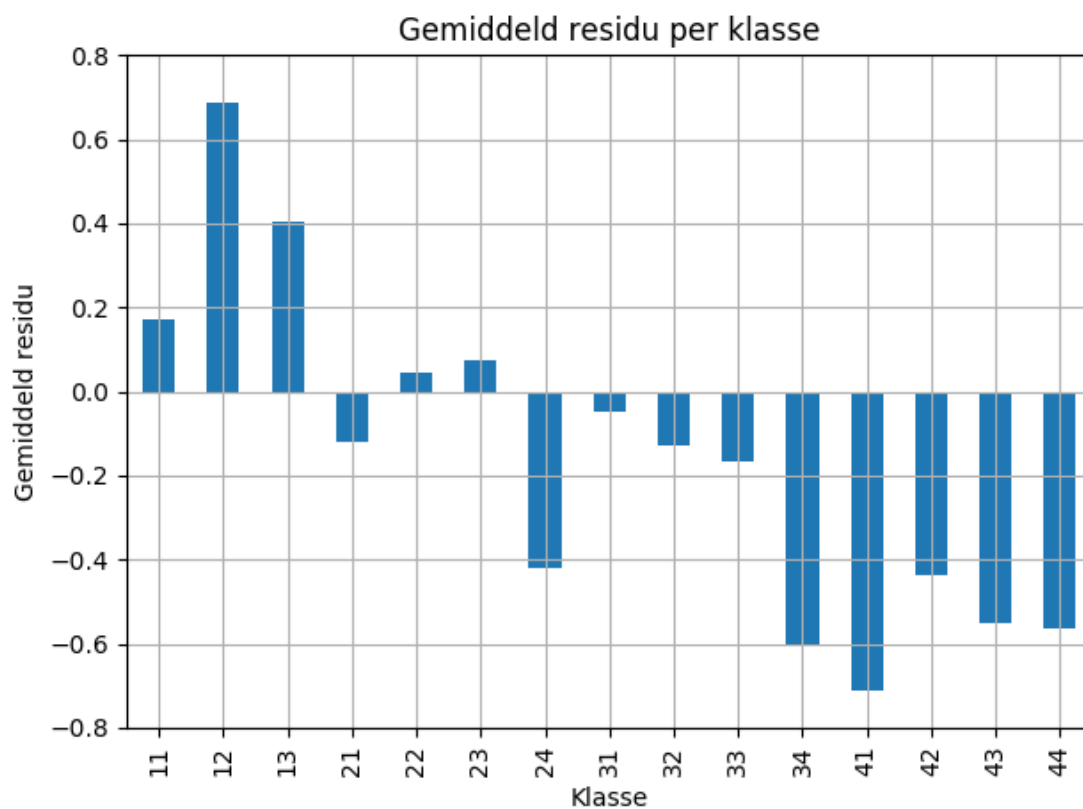
Het opstellen van een model op het niveau van de klassenverdeling heeft weinig toegevoegde waarde in de globale analyse en wordt bijgevolg niet herhaald voor de overige aggregatieniveaus. Het is echter wel interessant om de grootte van het residu te analyseren in functie van de klasse zoals is weergegeven in Figuur 5-2. Zoals beschreven in deelrapport 1 worden de klassen bepaald door onderstaande formule:

$$\text{Gecombineerde klasse} = 10 * \text{klasse grootteorde} + \text{klasse trend}$$

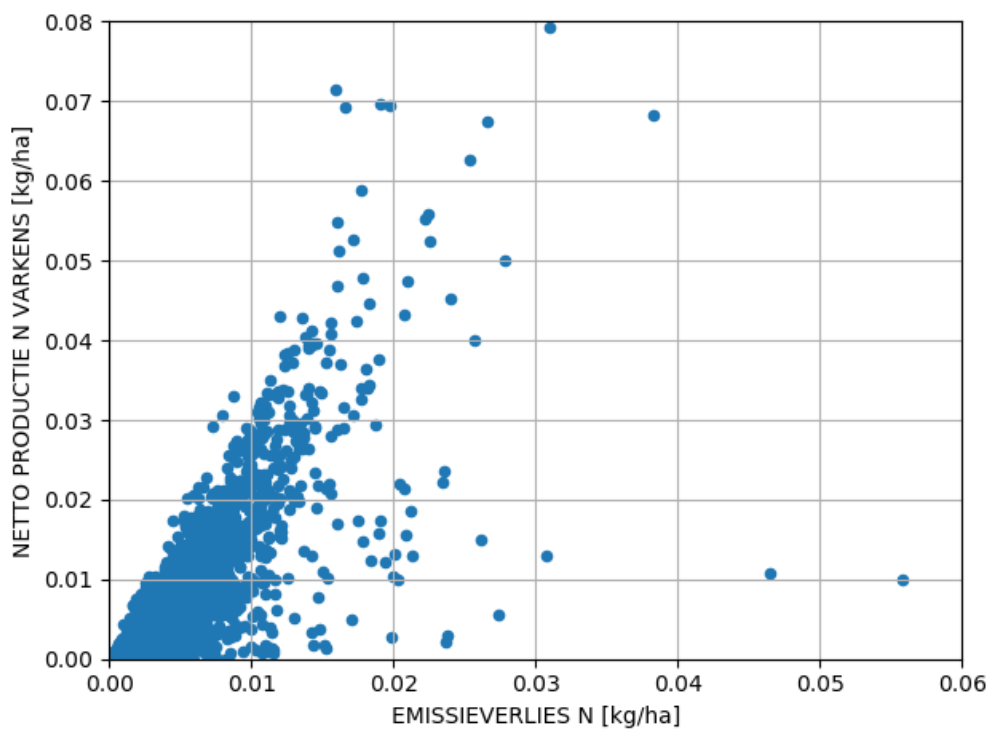
Uit de figuur kunnen we duidelijk afleiden dat het model bij lagere concentraties een overschatting geeft terwijl het bij hogere concentraties een onderschatting geeft. Dit is ook reeds zichtbaar op de verdeling van de residuen zoals weergegeven in Figuur 4-1. Voor grootteorde klasse 2 en 3 valt bovendien ook op dat bij een sterk positieve trend het model gemiddeld een sterkere onderschatting geeft. Al dienen we deze conclusie te nuanceren aangezien er slechts een beperkt aantal meetpunten aanwezig zijn in klasse 24 en 34. Verder valt er geen duidelijke verband te zien tussen de grootte van het gemiddeld residu en de trendklasse.



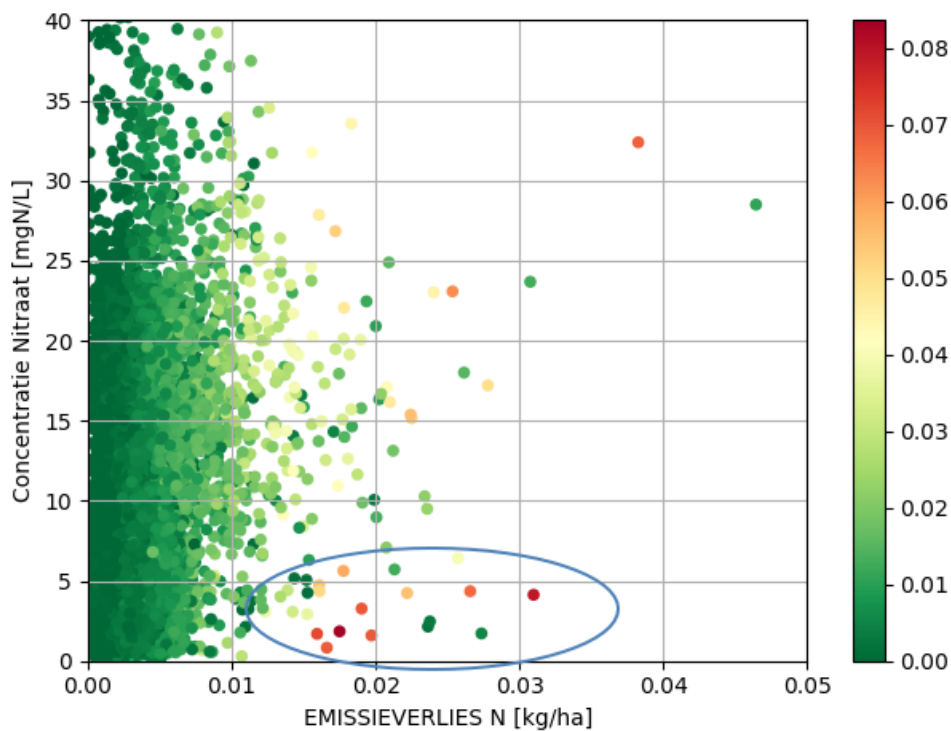
Figuur 5-1: Overzicht van de waarde van de coëfficiënten voor alle predictoren bij Nitraat, Oppervlaktewater.



*Figuur 5-2: Gemiddeld residu van het statistisch model voor Nitraat, oppervlaktewater, afstroomgebied per klasse bepaald op basis van combinatie trend en grootteorde*



*Figuur 5-3: Correlatie tussen NETTO PRODUCTIE N VARKENS en EMISSIEVERLIES N*



*Figuur 5-4: Correlatie tussen jaarlijkse maximale nitraatconcentratie in het oppervlaktewater en EMISSIEVERLIES N. De kleurschaal geeft de grootte van NETTO PRODUCTIE N VARKENS aan. Blauwe cirkel geeft punten weer waarvoor hogere concentratie verwacht kan worden op basis van de waarden van EMISSIEVERLIES N.*



## 5.2 NITRAAT, GRONDWATER

Figuur 5-5 toont de geselecteerde predictoren en resulterende coëfficiënten voor de verschillende ruimtelijke niveaus voor nitraat in het grondwater.

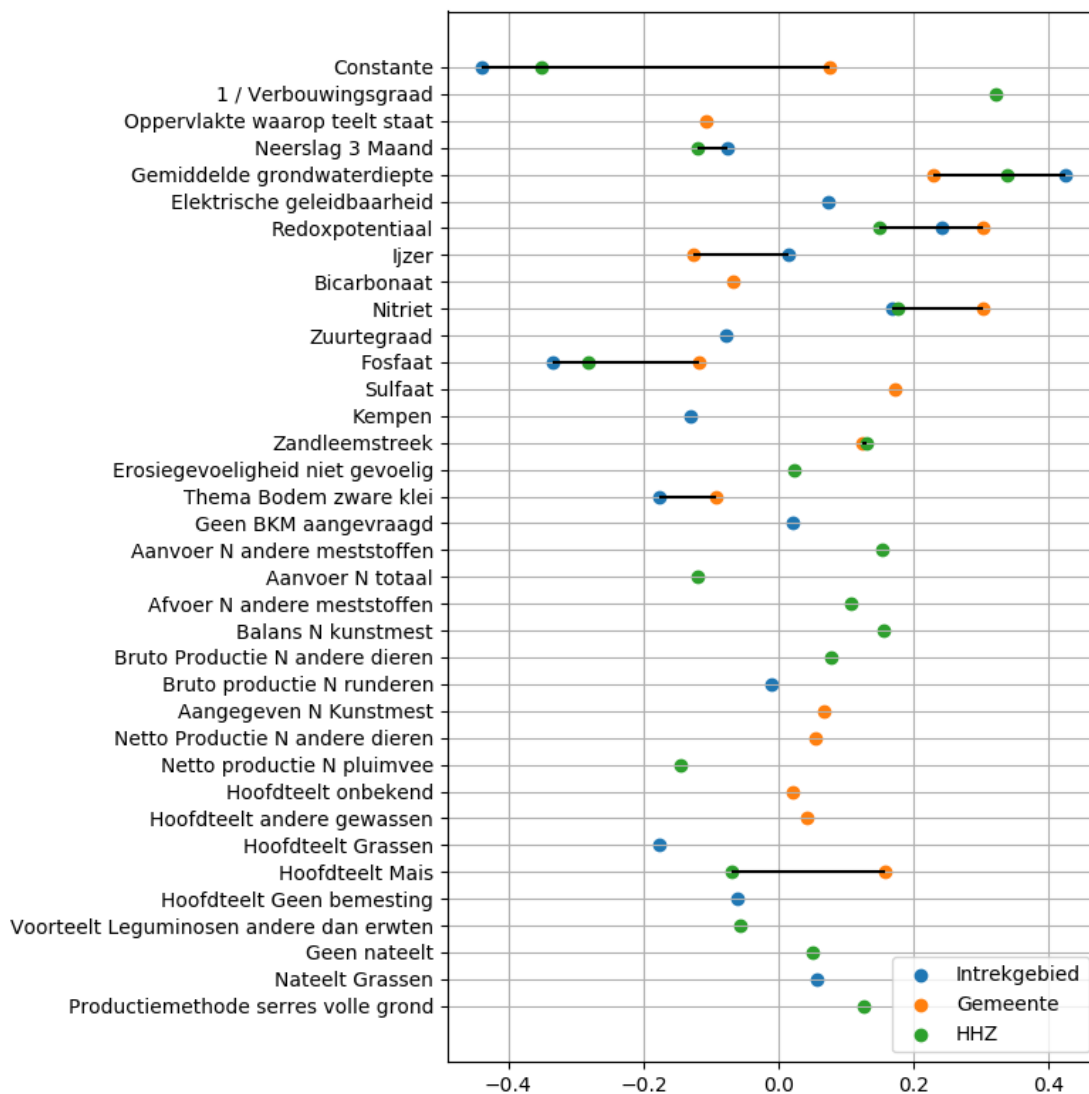
Hieruit blijkt dat er sterke verschillen zijn tussen de geselecteerde variabelen voor de verschillende aggregatieniveaus en dat er 9 variabelen zijn die in minimaal 2 van de 3 aggregatieniveaus voorkomen. Het gaat hierbij voornamelijk om gegevens m.b.t. waterkwaliteit, omgevingsfactoren & bodemgesteldheid:

- Nitriet
- Redoxpotentiaal
- Fosfaat
- Hoofddeelt Maïs
- Thema bodem zware klei
- Ijzer
- Gemiddelde grondwaterdiepte
- Zandleemstreek
- Neerslag 3 maand

Het valt op dat de waterkwaliteit- en omgevingsfactoren een groter deel van de variatie kunnen verklaren voor grondwater in vergelijking met N, oppervlaktewater. Zeker op het laagste schaalniveau (intrekgebied) wordt ruim 60 % van de maximaal te verklaren variatie verklaard door *GEMIDDELDE GRONDWATERDIEPTE* en *REDOXPOTENTIAAL*. Naarmate het aggregatieniveau verhoogt, worden meer punten uitgemiddeld en vermindert het belang van deze variabelen.

Variabelen met betrekking tot de gewasrotatie (vnl. grassen & maïs) en in mindere mate ook enkele exploitatieparameters verklaren het overige deel van de variatie.

De ruimtelijke verdeling van de statistische residuen vertoont geen duidelijke clustering (Figuur 4-9).



Figuur 5-5: Overzicht van de waarde van de coëfficiënten voor alle predictoren bij Nitraat, grondwater.

## 5.3 FOSFAAT, OPPERVLAKTEWATER

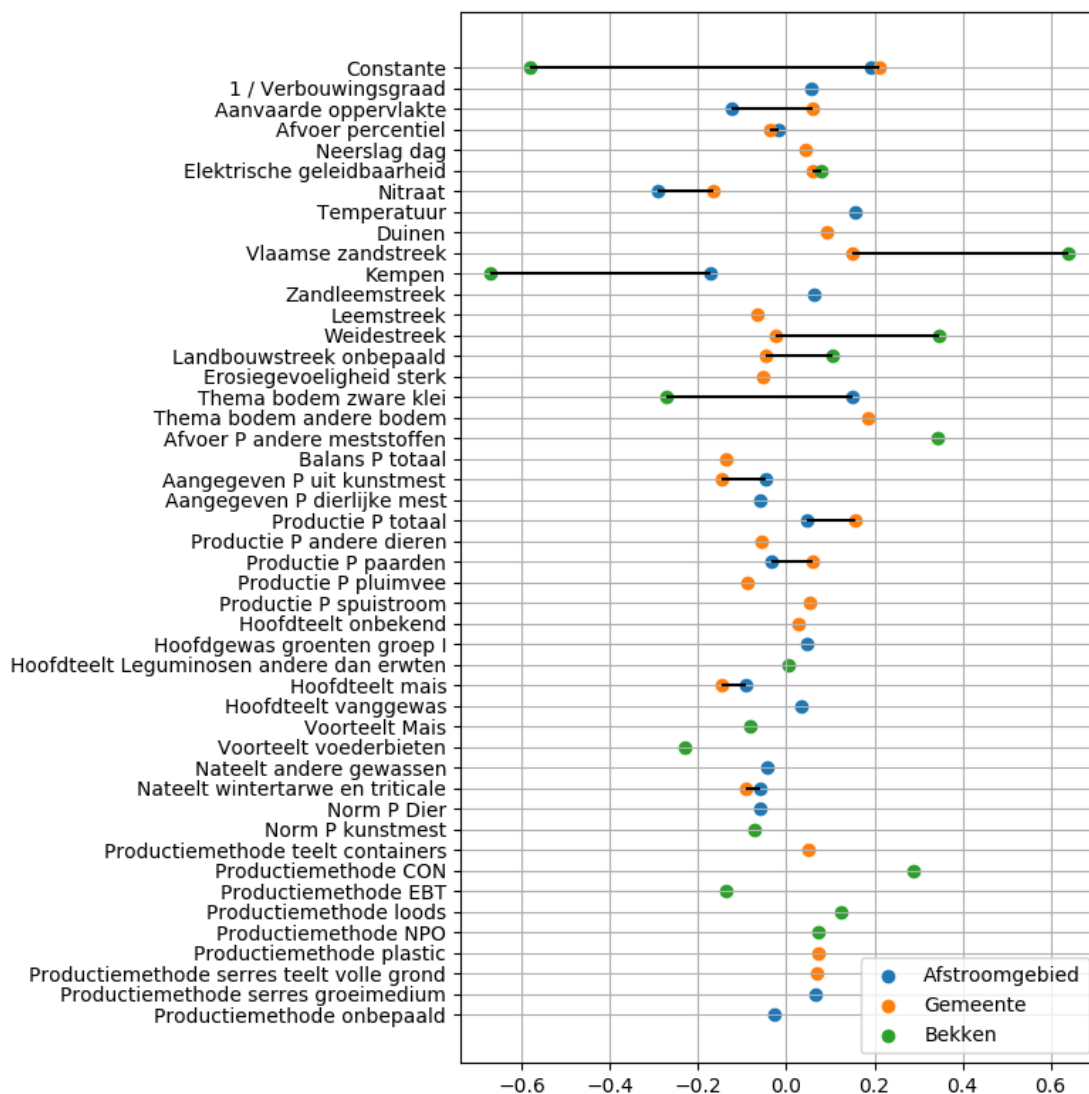
Figuur 5-6 toont de geselecteerde predictoren en resulterende coëfficiënten voor de verschillende ruimtelijke niveaus voor fosfaat in het oppervlaktewater.

Hieruit blijkt dat er sterke verschillen zijn tussen de geselecteerde variabelen voor de verschillende aggregatieniveaus en dat er 14 variabelen zijn die in minimaal 2 van de 3 aggregatieniveaus voorkomen. Het gaat hierbij voornamelijk om gegevens m.b.t. waterkwaliteit, omgevingsfactoren, gewasrotatie & exploitatiegegevens:

- Productie P paarden
- Afvoer percentiel
- Vlaamse zandstreek
- Aanvaarde oppervlakte
- Aangegeven P uit kunstmest
- Productie P totaal
- Nateelt wintertarwe en triticale
- Landbouwstreek onbepaald
- Elektrische geleidbaarheid
- Weidestreek
- Nitraat
- Kempen
- Hoofdgewas maïs
- Thema bodem zware klei

*NATEELT WINTERTARWE EN TRITICALE* en *PRODUCTIE P PAARDEN* zijn variabelen die niet voorkwamen in de modelvergelijkingen van nitraat. Het valt op dat veel landbouwcodes in deze lijst voorkomen wat een aanwijzing kan zijn van sterke regionale verschillen tussen de landbouwzones.

De ruimtelijke verdeling van de statistische residuen vertoont een geclusterde overschatting van het model ter hoogte van het Benden- en Bovenschelde bekken en het Denderbekken en tevens een (lichte) onderschatting voor de bekkens van Ijzer en Brugse polders voor het aggregatieniveau afstroomgebied en gemeente (Figuur 4-15 & Figuur 4-17).



Figuur 5-6: Overzicht van de waarde van de coëfficiënten voor alle predictoren bij Fosfaat, oppervlaktewater

## 5.4 FOSFAAT, GRONDWATER

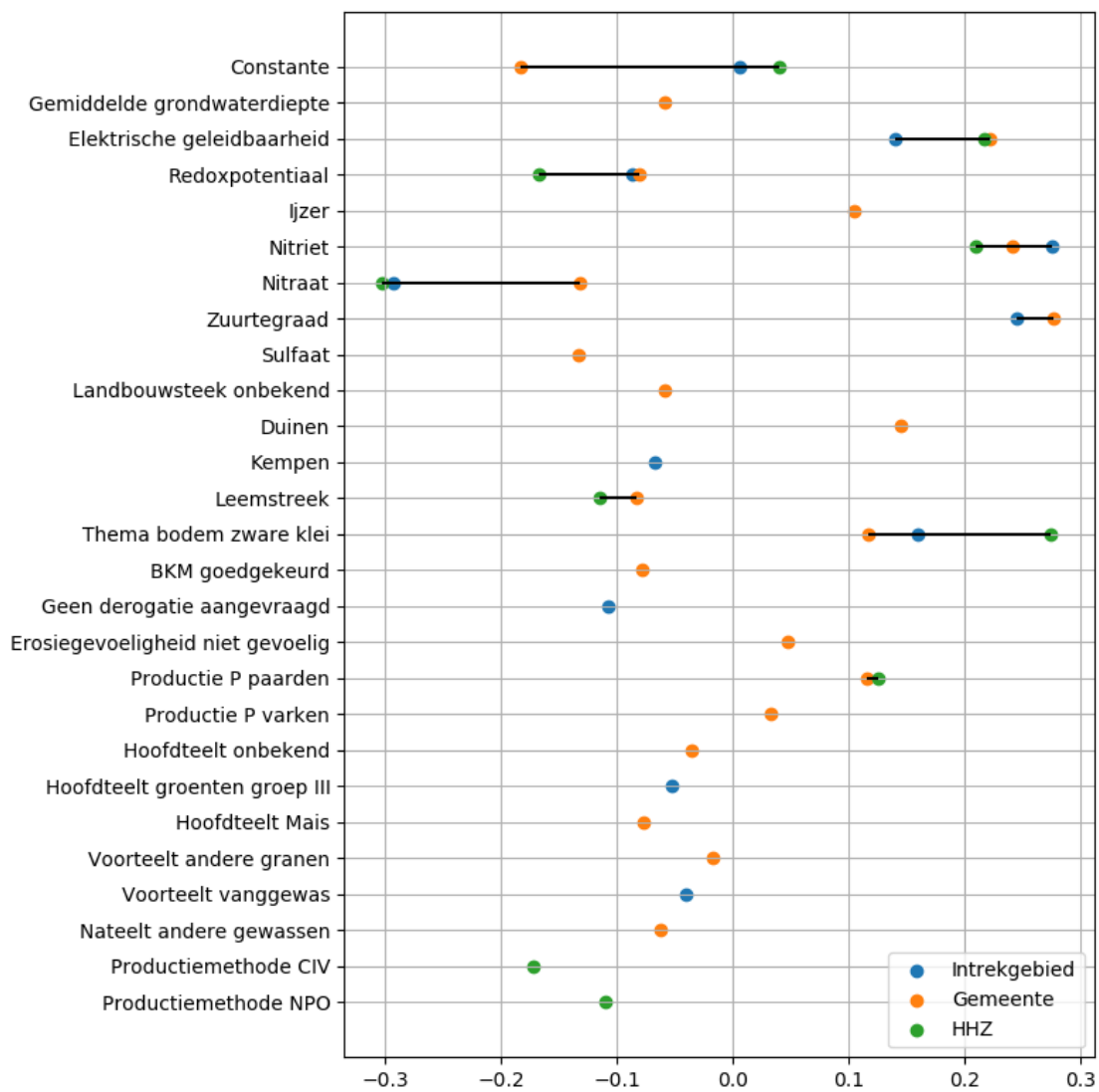
Figuur 5-7 toont de geselecteerde predictoren en resulterende coëfficiënten voor de verschillende ruimtelijke niveaus voor fosfaat in het grondwater.

Hieruit blijkt dat er sterke verschillen zijn tussen de geselecteerde variabelen voor de verschillende aggregatieniveaus en dat er slechts 8 variabelen zijn die in minimaal 2 van de 3 aggregatieniveaus voorkomen. Het gaat hierbij voornamelijk om gegevens m.b.t. waterkwaliteit, bodemgesteldheid en exploitatie:

- Thema bodem zware klei
- Nitriet
- Redoxpotentiaal
- Nitraat
- Elektrische geleidbaarheid
- Leemstreek
- Productie P paarden
- Zuurtegraad

Het valt ook hier op dat de waterkwaliteit- en omgevingsfactoren een groter deel van de variatie kunnen verklaren. Ook komt *PRODUCTIE P PAARDEN* hier weer duidelijk naar voor. *GEMIDDELDE GRONDWATERDIEPTE*, voor nitraat één van de belangrijkste factoren op het laagste schaalniveau (intrekgebied), is nu enkel nog aanwezig op het niveau van de gemeente.

De ruimtelijke verdeling van de statistische residuen vertoont geen duidelijke clustering (Figuur 4-21).



Figuur 5-7: Overzicht van de waarde van de coëfficiënten voor alle predictoren bij Fosfaat, grondwater.

## 5.5 NITRAATRESIDU

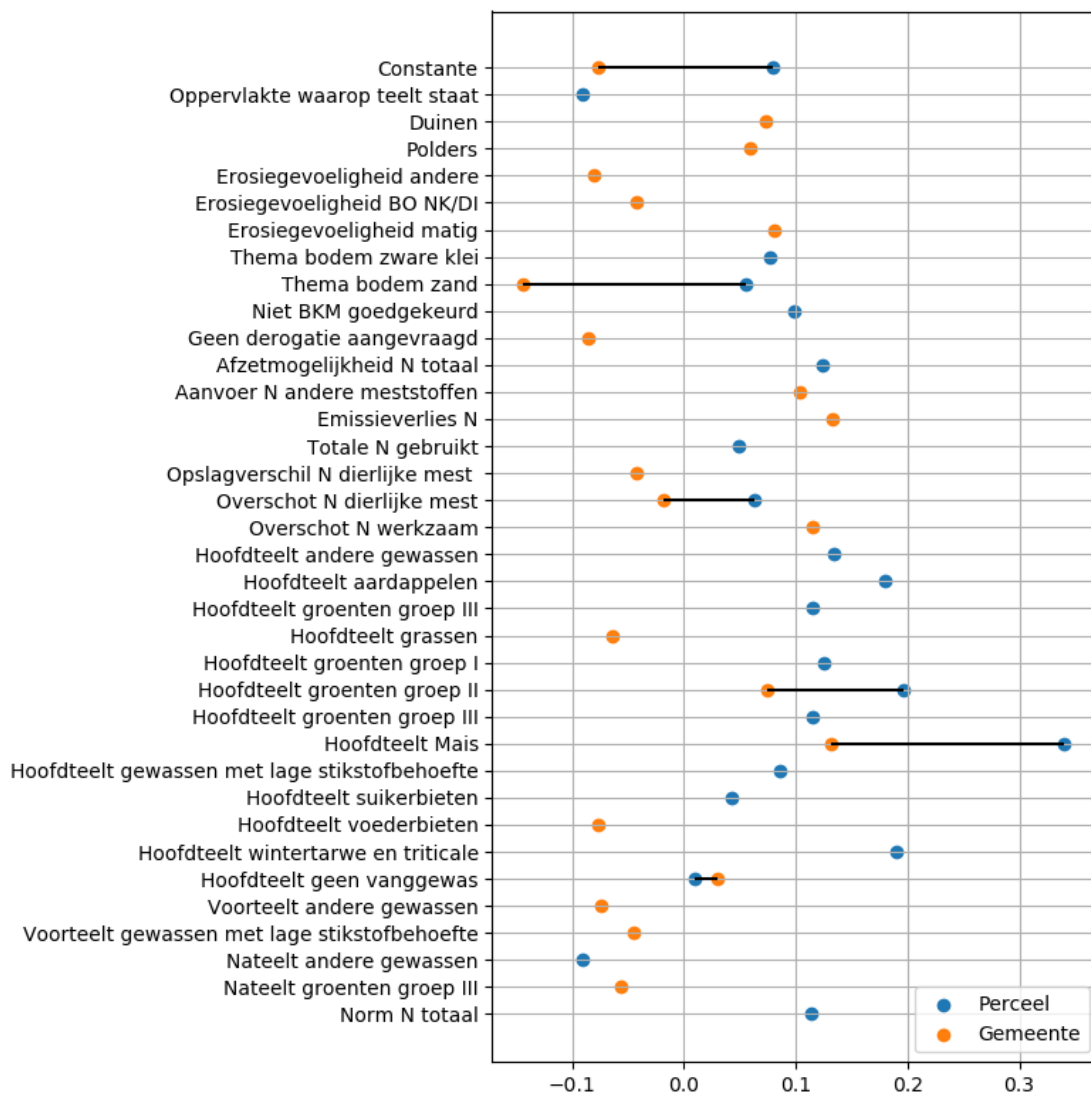
Figuur 5-8 toont de geselecteerde predictoren en resulterende coëfficiënten voor de verschillende ruimtelijke niveaus voor nitraat in het grondwater.

Hieruit blijkt dat er sterke verschillen zijn tussen de geselecteerde variabelen voor de 2 aggregatieniveaus en dat er slechts 5 variabelen zijn die in beide gevallen voorkomen:

- Overschot N dier
- Hoofddeelt geen vanggewas
- Hoofddeelt Maïs
- Hoofddeelt groenten groep II
- Thema bodem zand

Uit Tabel 5-1 blijkt dat de verklarende variantie van het model voor Nitraat residu lager is dan voor oppervlaktewater of grondwater en dit zowel op het kleinste schaalniveau als voor aggregatie op gemeentelijk niveau. Gezien nitraatresidu in essentie een meer directe meting is van het overgebleven nitraat in de bodem zouden we in eerste instantie een hogere verklarende variantie verwachten. Hierbij dienen we echter op te merken dat we bij nitraat residu, gegevens met betrekking tot neerslag en overige kwaliteitsparameters (bv. redoxpotentiaal) niet hebben meegenomen.

De ruimtelijke verdeling van de statistische residuen vertoont geen duidelijke clustering (Figuur 4-28).



Figuur 5-8: Overzicht van de waarde van de coëfficiënten voor alle predictoren bij Nitraatresidu



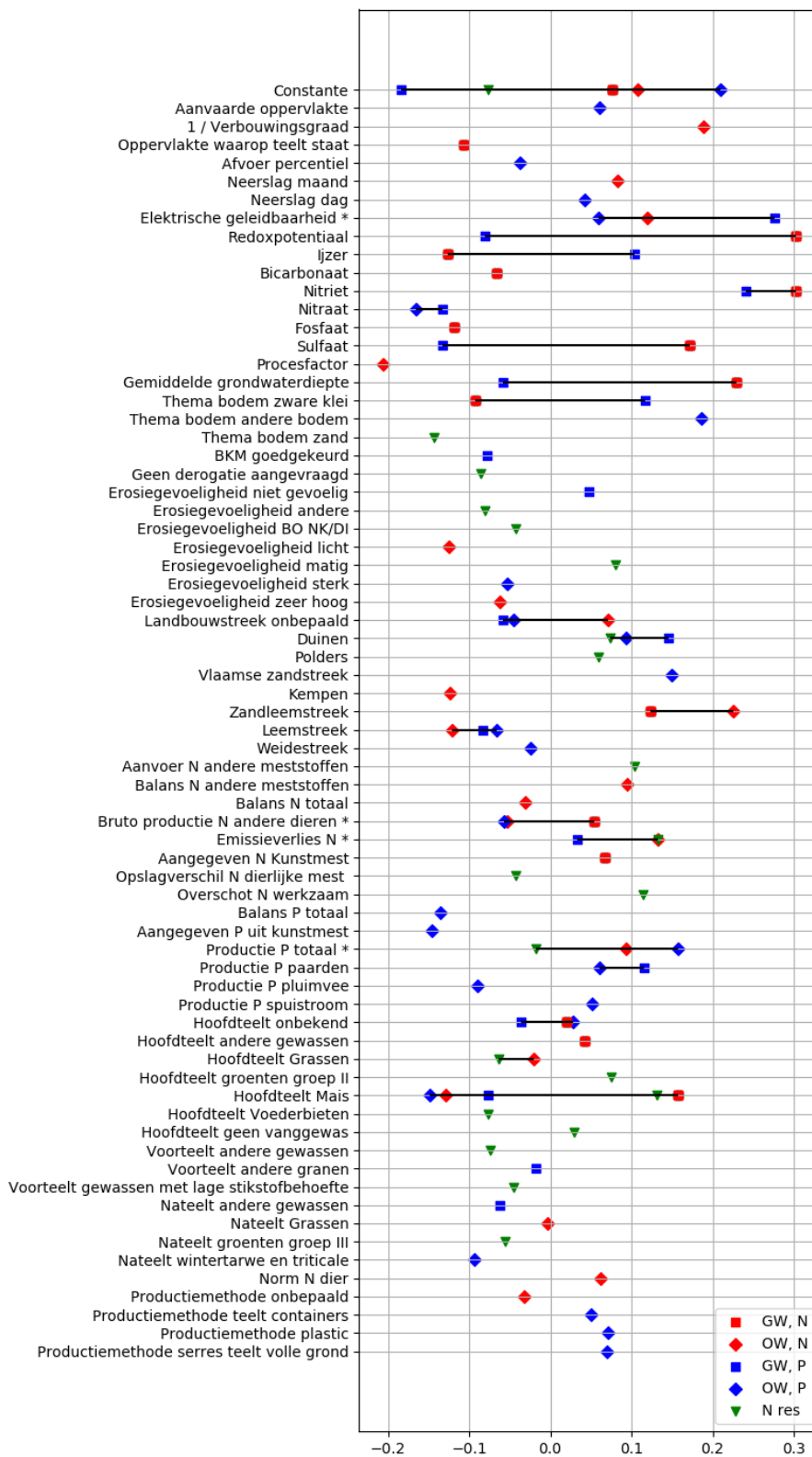
## 5.6 VERGELIJKING OP GEMEENTELIJK NIVEAU

Onderstaande Figuur 5-9 geeft een vergelijking tussen de coëfficiënten van de verschillende responsvariabelen op gemeentelijk niveau. Het gemeentelijk niveau is hier gekozen omdat dit niveau in iedere responsvariabele beschouwd is en daarom kan gebruikt worden als gemeenschappelijke basis voor vergelijking. Daarnaast wordt op gemeentelijk niveau steeds voldaan aan de veronderstellingen m.b.t. normale verdeling van de residuen en homoscedasticiteit. Om de vergelijking tussen de niveaus duidelijker te maken zijn predictoren met een onderlinge correlatiecoëfficiënt > 0.85 samengevoegd in Figuur 5-9. Concreet gaat het om de predictoren weergegeven in onderstaande Tabel 5-3.

*Tabel 5-3: Overzicht van samengevoegde predictoren voor vergelijking van responsvariabelen op niveau van gemeente*

Representatieve predictor	Predictoren
Emissieverlies N	Emissieverlies N; productie P varken
Productie P totaal	Productie P Totaal; Overschot N dier; overschot N totaal
Bruto productie N ander	Bruto productie N andere dieren; Productie P andere dieren; Netto productie N andere dieren
Elektrische geleidbaarheid	Elektrische geleidbaarheid; Zuurtegraad

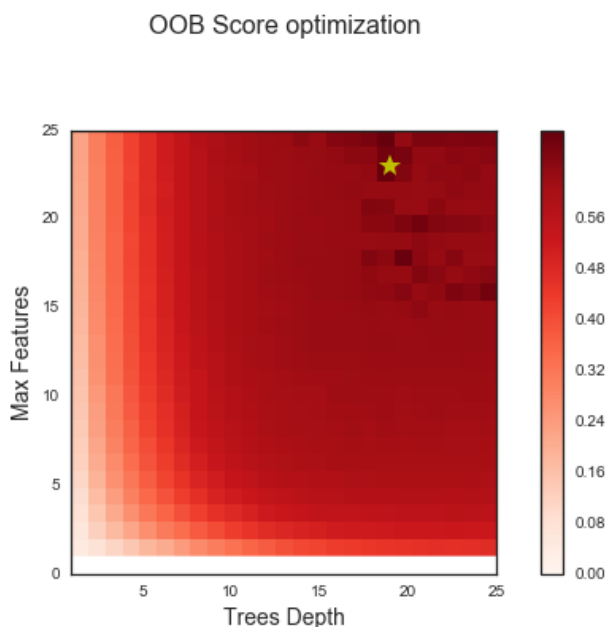
Uit Figuur 5-9 kunnen we bijvoorbeeld afleiden dat de oppervlakte van maïs een predictor is die voor alle responsvariabelen terugkomt. Echter de bijhorende coëfficiënt heeft niet steeds hetzelfde teken. Er zijn predictoren waarneembaar die terugkomen voor zowel oppervlaktewater als grondwater maar enkel voor fosfaat zoals bv. *PRODUCTIE P PAARD*. Er zijn ook predictoren waarneembaar die terugkomen bij zowel N als P maar enkel voor oppervlaktewater zoals de gecombineerde predictor *PRODUCTIE P TOTAAL* (waar ook overschot in vevat zit). We vinden eveneens predictoren terug waarvan de coëfficiënt van teken verandert tussen N & P zoals bv. *GEMIDDELDE GRONDWATERDIEPTE*, *THEMA BODEM ZWARE KLEI* of *IJZER*. Daarnaast is ook duidelijk dat de waterkwaliteitsgegevens voornamelijk van belang zijn bij grondwater en in mindere mate een rol spelen bij het oppervlaktewater, met uitzondering van de *ELEKTRISCHE GELEIDBAARHEID* en *NITRAAT* als predictor voor de concentratie fosfaat in het oppervlaktewater.



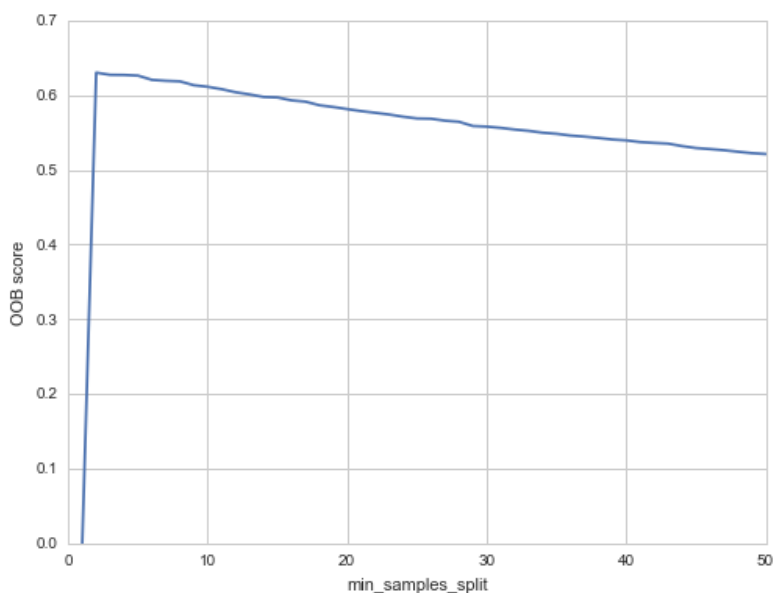
Figuur 5-9: Overzicht van de waarde van de coëfficiënten voor alle predictoren op gemeentelijk niveau. \* betreft samengevoegde predictoren

## 5.7 VALIDATIE MET RANDOM FOREST MODEL

Een Random Forest model, zoals beschreven in §2.4, is opgezet op de dataset voor nitraat, oppervlaktewater op gemeentelijk niveau. Het model bestaat uit een combinatie van 2500 individuele beslissingsbomen. Door middel van simultane optimalisatie van de Out Of the Box score (OOB) is het optimale aantal predictoren (max Features) in één beslissingsboom en de maximale diepte (Trees Depth) van één boom (Figuur 5-10) geselecteerd. Hier zijn 23 predictoren en een maximale diepte van 19 weerhouden. Het minimaal aantal observaties om 1 node te splitsen (min\_samples\_split) is vastgelegd op twee observaties door middel van een enkelvoudige OOB score optimalisatie (Figuur 5-11).

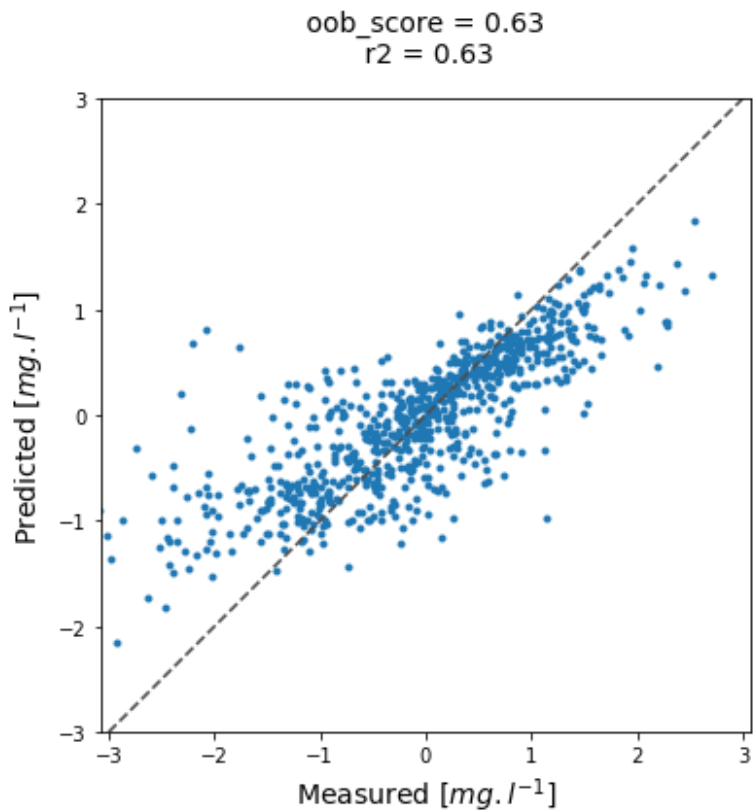


*Figuur 5-10: Optimalisatie Random Forest parameters op basis van OOB score*

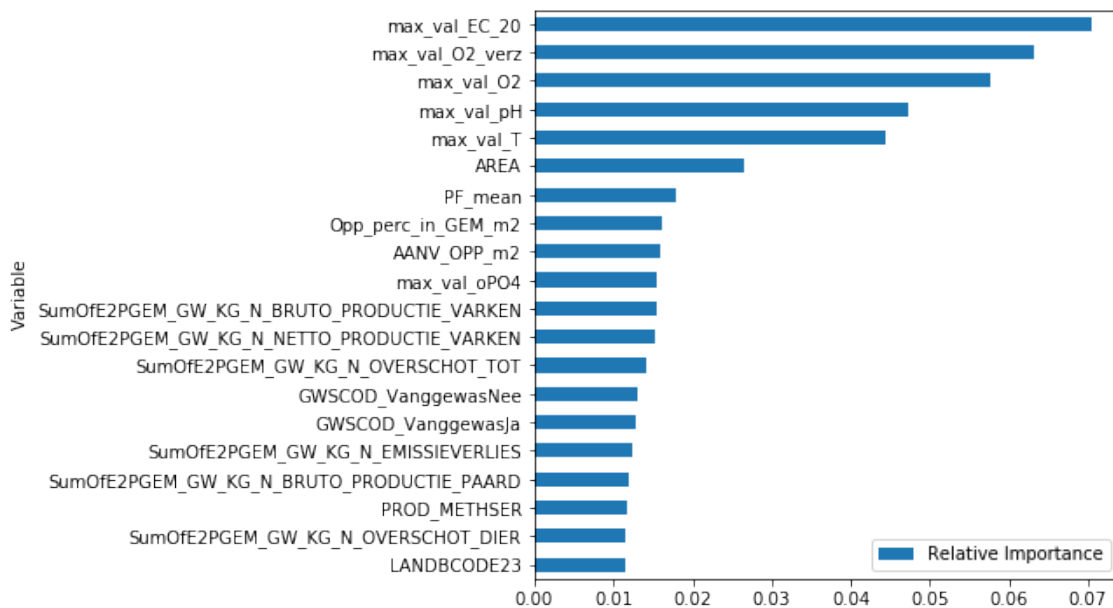


Figuur 5-11: Optimalisatie minimum samples in 1 opdeling

Het finale model is gebruikt om de genormaliseerde nitraat concentraties te voorspellen. Deze voorspelde waarden zijn uitgezet ten opzichte van de gemeten waarden (Figuur 5-12). Ook hier is een knik te zien van de puntenwolk ten opzichte van de bissectrice. Het model geeft dus een overschatting van de lage waarden en een onderschatting van de hogere waarden. Als we de predictoren sorteren op volgorde van hun relatief belang wordt duidelijk dat de hoog gecorreleerde waarden geclusterd voorkomen (Figuur 5-13). Dit kan verklaard worden door de inherente random selectie van predictoren in de aparte beslissingsbomen en de daaropvolgende samenvoeging voor het bepalen van het relatieve belang van de predictoren. Als variabele X1 in een aantal beslissingsbomen voorkomt en daar een sterke verklaring van de respons heeft, kan een sterk gecorreleerde variabele X2 (die dus grotendeels gelijkaardige informatie bevat) in een aantal andere beslissingsbomen voorkomen en daar ongeveer het zelfde belang hebben. Dit probleem komt minder sterk tot uiting in het PLS model aangezien de predictoren daar toegevoegd worden op basis van de multivariate ordening. Deze clustering bemoeilijkt een eenduidige vergelijking met het optimale PLS model. Toch zien we opvallende gelijkenissen tussen de geselecteerde predictoren voor beide modellen. Zo is *ELEKTRISCHE GELEIDBAARHEID* in beide modellen de predictor die het meeste variantie in de respons verklaart. Ook de drie daarop volgende predictoren uit het PLS model (*ZANDLEEMSTREEK, 1 / VERBOUWINGSGRAAD, OVERSCHOT N TOTAAL*) zijn terug te vinden in de top 20 van meest belangrijke variabelen uit het Random Forest model. Omgekeerd zijn ook veel van de geselecteerde variabelen uit het Random Forest, of een sterk gecorreleerde versie hiervan, ook terug te vinden tussen de predictoren van het PLS model. Deze vaststellingen vergroten het vertrouwen in de robuustheid en de resultaten van het optimale PLS model.



*Figuur 5-12: Resultaat Random Forest model:  
voorspelde waarden ten opzichte van gemeten waarden*



*Figuur 5-13: Gesorteerde predictoren in functie van het relatief belang*

## 6. ALGEMENE CONCLUSIE

In dit rapport zijn globale statistische modellen opgesteld voor Vlaanderen met als doel om inzicht te verkrijgen in de achterliggende oorzaken van een goede of slechte waterkwaliteit. Er zijn modellen opgesteld voor verschillende responsvariabelen (nitraat en fosfaat), voor metingen in het oppervlaktewater, grondwater en op het perceel (nitraatresidu) en voor verschillende ruimtelijke aggregatieniveaus.

Er zijn initieel 5 verschillende statistische modellen in rekening genomen waarvan er 2 (CCM en Random Forest) niet volledig zijn uitgewerkt op ieder niveau. Op basis van een analyse van de verklarende variantie in functie van het aantal predictoren of componenten werd het PLS model geselecteerd als meest optimale model. Dit model werd vervolgens toegepast op iedere responsvariabele en op ieder ruimtelijk aggregatieniveau.

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 het aggregatieniveau. Deze hogere percentages op niveau bekken dienen wel genuanceerd te worden. Door sterke uitmiddeling neemt de totale variantie van de responsvariabele sterk af en de procentuele verklaarde variantie bijgevolg toe. Daarnaast is de robuustheid en relevantie van deze modellen beperkt door een beperkt aantal datapunten. Er zijn geen consistente verschillen in modelperformantie waarneembaar tussen de responsvariabelen nitraat en fosfaat. Een ruimtelijke analyse van de statistische residuen vertoont duidelijke clustering van over- en onderschatting door het model voor de MAP meetpunten. Uit de verdeling van de residuen blijkt ook dat de modellen globaal genomen een overschatting geven van de lage concentraties en een onderschatting van de hoge concentraties.

De opgestelde vergelijking 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, afhankelijk van de beschouwde combinatie van responsvariabele en schaalniveau waardoor er geen algemeen geldende conclusies voor het systeem kunnen afgeleid worden. Zoals blijkt uit de voorbeelden in Hoofdstuk 5 wordt een duidelijke interpretatie van de coëfficiënten bovendien bemoeilijkt door de onderlinge afhankelijkheid tussen de predictoren. De oppervlakte maïsteelt is de enige predictor die in alle modellen terug te vinden is. Al verandert het teken van de bijhorende coëfficiënt wel. Predictoren met betrekking tot de landbouwzone waarin het afstroomgebied gelegen is, zijn ook frequent terug te vinden in de vergelijkingen. Dit is een indicatie dat een differentiatie van de modelvergelijkingen per landbouwzone mogelijks bijkomende inzichten kan opleveren.

De resultaten van dit onderdeel vormen de input voor onderdeel 3 van deze studie, nl. de gebiedsgerichte analyse. Door een meer gedetailleerde analyse van de residuen kunnen per responsvariabele interessante locaties bepaald worden. Onderzoek van deze locaties kan mogelijk resulteren in een verklaring waarom de huidige modellen de lage concentraties overschatten en de hoge concentraties onderschatten.

Daarnaast kunnen we het inzicht in de belangrijkste variabelen trachten te vergroten door een clustering uit te voeren op de predictorruimte alvorens over te gaan tot het opstellen van modelvergelijkingen. Door vervolgens per cluster maar één representatieve variabele op te nemen, vermijden we het voorkomen van sterk gecorreleerde waarden in de modelvergelijkingen. Dit kan resulteren in een duidelijkere weergave van het belang van iedere predictor (of groep van predictoren) op de resulterende waterkwaliteit.

## 7. REFERENTIES

Abdeslam A. (1998). Simulation numérique des écoulements souterrains et des relations nappes-rivière dans le bassin versant du Train (Belgique). Thèse, UCL - Institut Interfacultaire des Sciences.

Box G.E.P., Jenkins G.M., Reinsel G.C. & Ljung G.M. (2016). Time series analysis: forecasting and control Fifth edition. John Wiley & Sons, Inc, Hoboken, New Jersey.

Fabio Di Narzo A. (2017). mutual: Average Mutual Information in tseriesChaos: Analysis of nonlinear time series. Accessed 15 June 2017, <https://rdr.io/cran/tseriesChaos/man/mutual.html>.

Huang X., Hassani H., Ghodsi M., Mukherjee Z. & Gupta R. (2017). Do trend extraction approaches affect causality detection in climate change studies? *Physica A: Statistical Mechanics and its Applications*, 469, 604–624, doi: 10.1016/j.physa.2016.11.072.

Sugihara G., May R., Ye H., Hsieh C. -h., Deyle E., Fogarty M. & Munch S. (2012). Detecting Causality in Complex Ecosystems. *Science*, 338(6106), 496–500, doi: 10.1126/science.1227079.

Ye H. (2016). simplex: Perform univariate forecasting in rEDM: Applications of Empirical Dynamic Modeling from Time Series. Accessed 15 June 2017, <https://rdr.io/cran/rEDM/man/simplex.html>.

Ye H., Clark A., Deyle E., Keyes O. & Sugihara G. (2016). rEDM: Applications of Empirical Dynamic Modeling from Time Series.



## Bijlage A      **Overzicht parameters**

## A.1 Waterkwaliteitsparameters MAP-meetpunten

Tabel Bijlage A-1: Overzicht van de parameters die geregistreerd kunnen worden in een MAP-meetpunt

Opgenomen in analyse	Parameter <sup>2</sup>	Omschrijving	Eenheid
X	NO <sub>3</sub> <sup>-</sup>	Nitraat	mg N/L
X	oPO <sub>4</sub> <sup>3-</sup>	Orthofosfaat	mg P/L
X	Cl <sup>-</sup>	Chloride	Mg Cl/L
X	NH <sub>4</sub> <sup>+</sup>	Ammonium	mg N/L
X	NO <sub>2</sub>	Nitriet	mg N/L
X	EC_20	Elektrische conductiviteit bij 20°C	µS/cm
X	O <sub>2</sub>	Opgeloste zuurstof	Mg O <sub>2</sub> /L
X	O <sub>2</sub> _verz	Zuurstofverzadiging	%
X	T	Temperatuur	°C
X	pH	Zuurtegraad	-
	Secchi	Doorzicht	cm
	BZV <sup>5</sup>	Biochemisch zuurstofverbruik	mg O <sub>2</sub> /L
	CZV	Chemisch zuurstofverbruik	mg O <sub>2</sub> /L
	SO <sub>4</sub> <sup>2-</sup>	Sulfaat	Mg SO <sub>4</sub> <sup>2-</sup> /L
	KjN	Kjeldahl-stikstof	mg N/L
	N t	Totaal stikstof	mg N/L
	P t	Totaal fosfor	mg P/L
	ZS	Zwevende stof	mg ZS/L

<sup>2</sup> Na aggregatie over een jaartal kunnen deze parameters in het rapport volgende prefixen kennen:

- *max\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de maximale responswaarde geregistreerd werd
- *min\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de minimale responswaarde geregistreerd werd
- *median\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de mediaan responswaarde geregistreerd werd

## A.2 Metadata freatische grondwatermeetpunten

*Tabel Bijlage A-2: Meta-gegevens meetpunten freatisch grondwatermeetnet*

Opgenomen in analyse	Parameter	Bijkomende info
	put_id	Uniek identificatienummer van het meetput
	filternummer	Filternummer
X	onderkant_(peilfilter)	Diepte onderkant peilfilter [m-mv] <sup>3</sup>
	Gemeentenaam	Gemeente waarin meetpunt gelegen is
X	GEMGWD	Gemiddelde grondwaterdiepte (m-mv)
X (enkel op aggregatieniveau intrekgebieden)	opp_intrek	Oppervlakte van het intrekgebied van het meetpunt (m <sup>2</sup> )

---

<sup>3</sup> m-mv = meter onder maaiveld

## A.3 Waterkwaliteitsparameters freatische grondwatermeetpunten

Tabel Bijlage A-3: Overzicht van de parameters die geregistreerd kunnen worden in de meetpunten van het freatisch grondwatermeetnet.

Opgenomen in analyse	Parameter <sup>4</sup>	Omschrijving	Eenheid
X	Eh	Redoxpotentiaal	mV
X	Fe	Ijzer	Mg Fe/l
	Fe <sup>2+</sup>	Ijzer II	Mg Fe <sup>2+</sup> /l
	Fenoprop	Fenoprop	µg Fenoprop/l
X	HCO <sub>3</sub> <sup>-</sup>	Bicarbonaat	Mg HCO <sub>3</sub> <sup>-</sup> /l
X	Mn	Mangaan	Mg Mn/l
X	NO <sub>2</sub> <sup>-</sup>	Nitriet	Mg NO <sub>2</sub> /l
X	NO <sub>3</sub> <sup>-</sup>	Nitraat	Mg NO <sub>3</sub> <sup>-</sup> /l
X	PO <sub>4</sub> <sup>3-</sup>	Fosfaat	Mg PO <sub>4</sub> <sup>3-</sup> /l
X	SO <sub>4</sub> <sup>2-</sup>	Sulfaat	Mg SO <sub>4</sub> <sup>2-</sup> /l
X	pH	Zuurtegraad gemeten op het veld	Sörensen
	pH lab	Zuurtegraad gemeten in het labo	Sörensen
X	EC	Elektrische conductiviteit gemeten op het veld	µS/cm
	EC lab	Elektrische conductiviteit gemeten in het labo	µS/cm

<sup>4</sup> Na aggregatie over een jaartal kunnen deze parameters in het rapport volgende prefixen kennen:

- *max\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de maximale responswaarde geregistreerd werd
- *min\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de minimale responswaarde geregistreerd werd
- *median\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de mediaan responswaarde geregistreerd werd

## A.4 Exploitatie

*Tabel Bijlage A-4: Overzicht van de beschikbare exploitatiegegevens, met aanduiding welke parameters opgenomen werden in de statistische analyse*

Opgenomen in analyse	Parameter <sup>5</sup>	Shortname	Verklaring
	NR_LANDBOUWER_FICT	Landbouwnummer	Alle landbouwbedrijven hebben een uniek landbouwnummer, exploitantnummer(s) en exploitatienummer(s). Als een landbouwbedrijf uit meerdere exploitaties op verschillende locaties bestaat, dan beschikt dit landbouwbedrijf over één uniek landbouwnummer, met meerdere, unieke exploitatienummers. Eén exploitatie komt overeen met één bepaalde locatie (één adres), waar één of meerdere stallen gelegen zijn en/of waartoe één of meerdere landbouwpercelen toe behoren. Een landbouwperceel moet niet in de nabije omgeving van het exploitatieadres gelegen zijn.
	NR_EXPLOITANT_FICT	Exploitantnummer	
	NR_EXPLOITATIE_FICT	Exploitatienummer	
	JR_PRODUCTIE	Productiejaar	Vanaf 2007 t.e.m. 2014
X	GW_KG_N_AFZETMOG_ANDER	Afzetmogelijkheid N andere meststoffen	De hoeveelheid N uit andere meststoffen die maximaal kan afgezet worden op het geheel van percelen van de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van de afzetmogelijkheden voor N uit andere meststoffen op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg N (zie gegevens op perceelsniveau: <i>norm N andere mest</i> ).
X	GW_KG_N_AFZETMOG_CHEM	Afzetmogelijkheid kunstmest N	De hoeveelheid N uit kunstmest die maximaal kan afgezet worden op het geheel van percelen van de exploitatie, in kg N. Berekend als de som van de afzetmogelijkheden voor N uit kunstmest op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg N (zie gegevens op perceelsniveau: <i>norm N kunstmest</i> ).
X	GW_KG_N_AFZETMOG_DIER	Afzetmogelijkheid N dierlijke mest	De hoeveelheid N uit dierlijke mest die maximaal kan afgezet worden op het geheel van percelen van een de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van de afzetmogelijkheden voor N uit dierlijke mest op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg N (zie gegevens op perceelsniveau: <i>norm N dierlijke mest</i> ).
X	GW_KG_N_AFZETMOG_TOT	Afzetmogelijkheid N totaal	De totale hoeveelheid N uit meststoffen die maximaal kan afgezet worden op het geheel van percelen van de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van de afzetmogelijkheden voor totale N op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg N (zie gegevens op perceelsniveau: <i>norm N totaal</i> ).
X	GW_KG_N_ALLE_AANVOER_ANDER	Aanvoer N andere meststoffen	De hoeveelheid N uit andere meststoffen die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van alle geregistreerde transporten van andere meststoffen naar die bepaalde exploitatie in een bepaald productiejaar, in kg N.

<sup>5</sup> In de analyses kunnen parameters de volgende voorvoegsels hebben:

- E2P: De exploitatiegegevens (E) die omgezet werden naar perceelsniveau (P)
- SumOffE2POW: Som (SumOff) van exploitatiegegevens (E) omgezet naar perceelsniveau (P), en dit binnen een afstroomgebied van een MAP meetpunt (OW)
- SumOffE2PGW: Som (SumOff) van exploitatiegegevens (E) omgezet naar perceelsniveau (P), en dit binnen een intrekgebied van een grondwatermeetpunt (GW)
- SumOffE2PGEM: Som (SumOff) van exploitatiegegevens (E) omgezet naar perceelsniveau (P), en dit binnen een gemeente (GEM)

Opgenomen in analyse	Parameter <sup>5</sup>	Shortname	Verklaring
X	GW_KG_N_ALLE_AANVOER_COMPOST_GFT	Aanvoer N GFT-compost	De hoeveelheid N uit GFT-compost die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg N (van toepassing vanaf 2009). Berekend als de som van alle geregistreerde transporten van GFT-compost naar die bepaalde exploitatie in een bepaald productiejaar, in kg N.
X	GW_KG_N_ALLE_AANVOER_COMPOST_GRC	Aanvoer N groencompost	De hoeveelheid N uit groencompost die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg N (van toepassing vanaf 2009). Berekend als de som van alle geregistreerde transporten van groencompost naar die bepaalde exploitatie in een bepaald productiejaar, in kg N.
X	GW_KG_N_ALLE_AANVOER_DIER	Aanvoer N dierlijke mest	De hoeveelheid N uit dierlijke mest die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van alle geregistreerde transporten van dierlijke mest naar die bepaalde exploitatie in een bepaald productiejaar, in kg N.
X	GW_KG_N_ALLE_AANVOER_TOT	Aanvoer N totaal	De totale hoeveelheid N uit dierlijke of andere meststoffen die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van alle geregistreerde transporten van dierlijke of andere meststoffen naar die bepaalde exploitatie in een bepaald productiejaar, in kg N.
X	GW_KG_N_ALLE_AFVOER_ANDER	Afvoer N andere meststoffen	De hoeveelheid N uit andere meststoffen die via transportdocumenten wordt afgevoerd vanuit de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van alle geregistreerde transporten van andere meststoffen vanuit die bepaalde exploitatie in een bepaald productiejaar, in kg N.
X	GW_KG_N_ALLE_AFVOER_DIER	Afvoer N dierlijke mest	De hoeveelheid N uit dierlijke mest die via transportdocumenten wordt afgevoerd vanuit de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van alle geregistreerde transporten van dierlijke mest vanuit die bepaalde exploitatie in een bepaald productiejaar, in kg N.
X	GW_KG_N_ALLE_AFVOER_TOT	Afvoer N totaal	De totale hoeveelheid N uit dierlijke of andere meststoffen die via transportdocumenten wordt afgevoerd vanuit de exploitatie in een bepaald productiejaar, in kg N. Berekend als de som van alle geregistreerde transporten van dierlijke of andere meststoffen vanuit die bepaalde exploitatie in een bepaald productiejaar, in kg N.
X	GW_KG_N_BALANS_ANDER	Balans N andere meststoffen	Resultaat van de balansberekening voor andere meststoffen voor de exploitatie in een bepaald productiejaar, in kg N. Berekeningswijze: <i>Balans N andere meststoffen = + opslagverschil N andere meststoffen + aanvoer N andere meststoffen - afvoer N andere meststoffen - afzetmogelijkheid N andere meststoffen.</i>
X	GW_KG_N_BALANS_CHEM	Balans N kunstmest	Resultaat van de balansberekening voor kunstmest voor de exploitatie in een bepaald productiejaar, in kg N. Berekeningswijze: <i>Balans N kunstmest = + gebruik N kunstmest - afzetmogelijkheid N kunstmest.</i>
X	GW_KG_N_BALANS_DIER	Balans N dierlijke mest	Resultaat van de balansberekening voor dierlijke mest voor de exploitatie in een bepaald productiejaar, in kg N. Berekeningswijze (vereenvoudigd): <i>Balans N dierlijke mest = + netto productie N totaal + opslagverschil N dierlijke mest + aanvoer N dierlijke mest - afvoer N dierlijke mest - afzetmogelijkheid N dierlijke mest.</i>
X	GW_KG_N_BALANS_SPUISTR	Balans N spuistroom	Resultaat van de balansberekening voor spuistroom voor de exploitatie in een bepaald productiejaar, in kg N. Berekeningswijze: <i>Balans N spuistroom = + productie N spuistroom + opslagverschil N spuistroom + aanvoer N spuistroom - afvoer N spuistroom.</i>
X	GW_KG_N_BALANS_TOT	Balans N totaal	Resultaat van de totale balansberekening voor de exploitatie in een bepaald productiejaar, in kg N. Berekeningswijze (vereenvoudigd): <i>Balans N totaal = + netto productie N totaal + productie N spuistroom + gebruik N kunstmest + opslagverschil N dierlijke mest / andere meststoffen + aanvoer N dierlijke mest / andere meststoffen - afvoer N dierlijke mest / andere meststoffen - afzetmogelijkheid N totaal.</i>
X	GW_KG_N_BRUTO_PRODUCTIE	Bruto productie N totaal	Totale bruto dierlijke mestproductie op de exploitatie in een bepaald productiejaar, in kg N.

Opgenomen in analyse	Parameter <sup>5</sup>	Shortname	Verklaring
			Voor elke diercategorie (bv. melkkoeien) van elke diersoort (bv. runderen), wordt de bruto productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor N. De uitscheidingsnormen zijn afhankelijk van de diercategorie en de nutriëntenbalans (voor varkens en pluimvee). Door de bruto productie per diercategorie van alle diercategorieën van de exploitatie te sommeren, wordt de totale bruto dierlijke mestproductie op de exploitatie in een bepaald productiejaar berekend.
X	GW_KG_N_BRUTO_PRODUCTIE_ANDER	Bruto productie N andere dieren	Bruto dierlijke mestproductie van de diersoort andere dieren, op de exploitatie in een bepaald productiejaar, in kg N. Voor elke diercategorie van de diersoort andere dieren (dit zijn konijnen, schapen, geiten en nertsen), wordt de bruto productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor N. Vervolgens wordt de bruto productie van alle diercategorieën van de diersoort andere dieren gesommeerd.
X	GW_KG_N_BRUTO_PRODUCTIE_PAARD	Bruto productie N paarden	Bruto dierlijke mestproductie van de diersoort paarden, op de exploitatie in een bepaald productiejaar, in kg N. Voor elke diercategorie van de diersoort paarden, wordt de bruto productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor N. Vervolgens wordt de bruto productie van alle diercategorieën van de diersoort paarden gesommeerd.
X	GW_KG_N_BRUTO_PRODUCTIE_PLUIM	Bruto productie N pluimvee	Bruto dierlijke mestproductie van de diersoort pluimvee, op de exploitatie in een bepaald productiejaar, in kg N. Voor elke diercategorie van de diersoort pluimvee, wordt de bruto productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor N (afhankelijk van de nutriëntenbalans). Vervolgens wordt de bruto productie van alle diercategorieën van de diersoort pluimvee gesommeerd.
X	GW_KG_N_BRUTO_PRODUCTIE_RUND	Bruto productie N runderen	Bruto dierlijke mestproductie van de diersoort runderen, op de exploitatie in een bepaald productiejaar, in kg N. Voor elke diercategorie van de diersoort runderen, wordt de bruto productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor N. Vervolgens wordt de bruto productie van alle diercategorieën van de diersoort runderen gesommeerd.
X	GW_KG_N_BRUTO_PRODUCTIE_VARKEN	Bruto productie N varkens	Bruto dierlijke mestproductie van de diersoort varkens, op de exploitatie in een bepaald productiejaar, in kg N. Voor elke diercategorie van de diersoort varkens, wordt de bruto productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor N (afhankelijk van de nutriëntenbalans). Vervolgens wordt de bruto productie van alle diercategorieën van de diersoort varkens gesommeerd.
X	GW_KG_N_EMISSIEVERLIES	Emissieverlies N	Hoeveelheid stikstofemissieverliezen uit stal en opslag op de exploitatie in een bepaald productiejaar, in kg N. De emissieverliezen zijn afhankelijk van de diercategorie en van het staltype. Voor elke diercategorie wordt het emissieverlies berekend en vervolgens wordt gesommeerd over alle diercategorieën van de exploitatie.
X	GW_KG_N_GEBRUIK_CHEM	Aangegeven N kunstmest	Totale hoeveelheid kunstmest gebruikt op landbouwgrond van de exploitatie in een bepaald productiejaar, in kg N. Deze hoeveelheid wordt aangegeven door de landbouwer.
X	GW_KG_N_GEBRUIKT	Totale N gebruikt	Totale hoeveelheid meststoffen gebruikt op landbouwgrond van de exploitatie in een bepaald productiejaar, in kg N. Berekeningswijze (vereenvoudigd): <i>Gebruik N totaal = + netto productie N totaal + productie N spuistroom + gebruik N kunstmest + opslagverschil N dierlijke mest/andere meststoffen + aanvoer N dierlijke mest/andere meststoffen - afvoer N dierlijke mest/andere meststoffen.</i>
X	GW_KG_N_NETTO_PRODUCTIE	Netto productie N totaal	Totale netto dierlijke mestproductie op de exploitatie in een bepaald productiejaar, in kg N. Dit is de dierlijke mestproductie waarbij de emissieverliezen uit stal en opslag in mindering zijn gebracht. Berekend als het verschil tussen <i>Bruto productie N totaal</i> en <i>Emissieverlies N</i> .
X	GW_KG_N_NETTO_PRODUCTIE_ANDER	Netto productie N andere dieren	Netto dierlijke mestproductie van de diersoort andere dieren op de exploitatie in een bepaald productiejaar, in kg N. Dit is de dierlijke mestproductie waarbij de emissieverliezen uit stal en opslag in mindering zijn gebracht. Berekend als het verschil tussen <i>Bruto productie N andere dieren</i> en <i>Emissieverlies N andere dieren</i> .
X	GW_KG_N_NETTO_PRODUCTIE_PAARD	Netto productie N paarden	Netto dierlijke mestproductie van de diersoort paarden, op de exploitatie in een bepaald productiejaar, in kg N. Dit is de dierlijke mestproductie waarbij de emissieverliezen uit stal en opslag in mindering zijn gebracht. Berekend als het verschil tussen <i>Bruto productie N paarden</i> en <i>Emissieverlies N paarden</i> .
X	GW_KG_N_NETTO_PRODUCTIE_PLUIM	Netto productie N pluimvee	Netto dierlijke mestproductie van de diersoort pluimvee, op de exploitatie in een bepaald productiejaar, in kg N. Dit is de dierlijke mestproductie waarbij de emissieverliezen uit stal en opslag in mindering zijn gebracht. Berekend als het verschil tussen <i>Bruto productie N pluimvee</i> en <i>Emissieverlies N pluimvee</i> .
X	GW_KG_N_NETTO_PRODUCTIE_RUND	Netto productie N runderen	Netto dierlijke mestproductie van de diersoort runderen, op de exploitatie in een bepaald productiejaar, in kg N. Dit is de dierlijke mestproductie waarbij de emissieverliezen uit stal en opslag in mindering zijn gebracht.

Opgenomen in analyse	Parameter <sup>5</sup>	Shortname	Verklaring
			Berekend als het verschil tussen <i>Bruto productie N runderen</i> en <i>Emissieverlies N runderen</i> .
X	GW_KG_N_NETTO_PRODUCTIE_VARKEN	Netto productie N varkens	Netto dierlijke mestproductie van de diersoort varkens, op de exploitatie in een bepaald productiejaar, in kg N. Dit is de dierlijke mestproductie waarbij de emissieverliezen uit stal en opslag in mindering zijn gebracht. Berekend als het verschil tussen <i>Bruto productie N varkens</i> en <i>Emissieverlies N varkens</i> .
X	GW_KG_N_NETTO_PRODUCTIE_WERKZAAM	Netto productie N werkzaam	Totale netto dierlijke mestproductie op de exploitatie in een bepaald productiejaar, uitgedrukt in kg werkzame N. Werkzame stikstof is stikstof die in het jaar van toediening door de plant kan worden opgenomen. De hoeveelheid werkzame stikstof van een mestsoort (kg/ton) = totale stikstof (kg/ton) x werkingscoëfficiënt mestsoort (%). In het Mestdecreet is een werkingscoëfficiënt van 60% voor mengmest, 30% voor stalment en 20% voor beweiding vastgelegd. Voor elke diercategorie kan op basis van aangiftegegevens zoals onder meer de staltypes, berekend worden hoeveel mest geproduceerd wordt in de stal, onder de vorm van stal- of mengmest, of op de weide. Rekening houdend met de werkingscoëfficiënten, wordt vervolgens de dierlijke mestproductie berekend, uitgedrukt in kg werkzame N.
X	GW_KG_N_OPFLAGVERSCHIL_ANDER	Opslagverschil N andere meststoffen	Verskil tussen de aangegeven opslag van andere meststoffen op 1 januari van productiejaar X en de aangegeven opslag op 1 januari van productiejaar X+1, in kg N.
X	GW_KG_N_OPFLAGVERSCHIL_DIER	Opslagverschil N dierlijke mest	Verskil tussen de aangegeven opslag van dierlijke mest op 1 januari van productiejaar X en de aangegeven opslag op 1 januari van productiejaar X+1, in kg N.
X	GW_KG_N_OPFLAGVERSCHIL_TOT	Opslagverschil N totaal	Het totale opslagverschil tussen de aangegeven opslag op 1 januari van productiejaar X en de aangegeven opslag op 1 januari van productiejaar X+1, in kg N. Berekend als de som van het opslagverschil voor dierlijke mest en andere meststoffen.
X	GW_KG_N_OVERSCHOT_DIER	Overschot N dierlijke mest	Het mestoverschot van de exploitatie in een bepaald productiejaar, in kg N. Berekend als het verschil tussen <i>Netto productie N totaal</i> en <i>Afzetmogelijkheid N dierlijke mest</i> .
X	GW_KG_N_OVERSCHOT_TOT	Overschot N totaal	Het overschot van de exploitatie in een bepaald productiejaar, in kg N. Berekend als het verschil tussen <i>Netto productie N totaal</i> en <i>Afzetmogelijkheid N totaal</i> .
X	GW_KG_N_OVERSCHOT_WERKZAAM	Overschot N werkzaam	Het mestoverschot van de exploitatie in een bepaald productiejaar, in kg N. Berekend als het verschil tussen <i>Netto productie N werkzaam</i> en <i>Afzetmogelijkheid N werkzaam</i> .
X	GW_KG_N_PRODUCTIE_SPUISTR	Productie N spuiroom	Hoeveelheid spuiroomproductie op de exploitatie in een bepaald productiejaar, in kg N. Berekend o.b.v. de aangegeven hoeveelheid spuiroomproductie en de samenstelling.
X	GW_KG_N_SCHUIM	Aangevoerde N schuimaarde	Hoeveelheid aangevoerde schuimaarde in kg N
X	GW_KG_P_AFZETMOG_ANDER	Afzetmogelijkheid P andere meststoffen	De hoeveelheid P <sub>2</sub> O <sub>5</sub> uit andere meststoffen die maximaal kan afgezet worden op het geheel van percelen van een bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van de afzetmogelijkheden voor P <sub>2</sub> O <sub>5</sub> uit andere meststoffen op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> (zie gegevens op perceelsniveau: <i>norm P<sub>2</sub>O<sub>5</sub> andere mest</i> ).
X	GW_KG_P_AFZETMOG_CHEM	Afzetmogelijkheid P kunstmest	De hoeveelheid P <sub>2</sub> O <sub>5</sub> uit kunstmest die maximaal kan afgezet worden op het geheel van percelen van een bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van de afzetmogelijkheden voor P <sub>2</sub> O <sub>5</sub> uit kunstmest op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> (zie gegevens op perceelsniveau: <i>norm P<sub>2</sub>O<sub>5</sub> kunstmest</i> ).
X	GW_KG_P_AFZETMOG_DIER	Afzetmogelijkheid P dierlijke mest	De hoeveelheid P <sub>2</sub> O <sub>5</sub> uit dierlijke mest die maximaal kan afgezet worden op het geheel van percelen van een bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van de afzetmogelijkheden voor P <sub>2</sub> O <sub>5</sub> uit dierlijke mest op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> (zie gegevens op perceelsniveau: <i>norm P<sub>2</sub>O<sub>5</sub> dierlijke mest</i> ).
X	GW_KG_P_AFZETMOG_TOT	Afzetmogelijkheid P totaal	De totale hoeveelheid P <sub>2</sub> O <sub>5</sub> uit meststoffen die maximaal kan afgezet worden op het geheel van percelen van een bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van de afzetmogelijkheden voor totale P <sub>2</sub> O <sub>5</sub> op de verschillende percelen behorend tot die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> (zie gegevens op perceelsniveau: <i>norm P<sub>2</sub>O<sub>5</sub> totaal</i> ).



Opgenomen in analyse	Parameter <sup>5</sup>	Shortname	Verklaring
X	GW_KG_P_ALLE_AANVOER_ANDER	Aanvoer P andere meststoffen	De hoeveelheid P <sub>2</sub> O <sub>5</sub> uit andere meststoffen die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van alle geregistreerde transporten van andere meststoffen naar die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> .
X	GW_KG_P_ALLE_AANVOER_DIER	Aanvoer P dierlijke mest	De hoeveelheid P <sub>2</sub> O <sub>5</sub> uit dierlijke mest die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van alle geregistreerde transporten van dierlijke mest naar die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> .
X	GW_KG_P_ALLE_AANVOER_TOT	Aanvoer P totaal	De totale hoeveelheid P <sub>2</sub> O <sub>5</sub> uit dierlijke of andere meststoffen die via transportdocumenten wordt aangevoerd naar de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van alle geregistreerde transporten van dierlijke of andere meststoffen naar die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> .
X	GW_KG_P_ALLE_AFVOER_ANDER	Afvoer P andere meststoffen	De hoeveelheid P <sub>2</sub> O <sub>5</sub> uit andere meststoffen die via transportdocumenten wordt afgevoerd vanuit de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van alle geregistreerde transporten van andere meststoffen vanuit die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> .
X	GW_KG_P_ALLE_AFVOER_DIER	Afvoer P dierlijke mest	De hoeveelheid P <sub>2</sub> O <sub>5</sub> uit dierlijke mest die via transportdocumenten wordt afgevoerd vanuit de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van alle geregistreerde transporten van dierlijke mest vanuit die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> .
X	GW_KG_P_ALLE_AFVOER_TOT	Afvoer P totaal	De totale hoeveelheid P <sub>2</sub> O <sub>5</sub> uit dierlijke of andere meststoffen die via transportdocumenten wordt afgevoerd vanuit de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als de som van alle geregistreerde transporten van dierlijke of andere meststoffen vanuit die bepaalde exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> .
X	GW_KG_P_BALANS_TOT	Balans P totaal	Resultaat van de totale balansberekening voor de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekeningswijze (vereenvoudigd): <i>Balans P<sub>2</sub>O<sub>5</sub> totaal = + productie P<sub>2</sub>O<sub>5</sub> totaal + productie P<sub>2</sub>O<sub>5</sub> spuistroom + gebruik P<sub>2</sub>O<sub>5</sub> kunstmest + opslagverschil P<sub>2</sub>O<sub>5</sub> dierlijke mest / andere meststoffen + aanvoer P<sub>2</sub>O<sub>5</sub> dierlijke mest / andere meststoffen - afvoer P<sub>2</sub>O<sub>5</sub> dierlijke mest / andere meststoffen - afzetmogelijkheid P<sub>2</sub>O<sub>5</sub> totaal.</i>
X	GW_KG_P_GEBRUIK_CHEM	Aangegeven P uit kunstmest	Totale hoeveelheid kunstmest gebruikt op landbouwgrond van de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Deze hoeveelheid wordt aangegeven door de landbouwer.
X	GW_KG_P_GEBRUIK_DIER	Aangegeven P dierlijke mest	Hoeveelheid dierlijke mest die afkomstig is van dezelfde landbouwer van buiten Vlaanderen, op landbouwgrond van de exploitatie in Vlaanderen ("grensboerer") in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Deze hoeveelheid wordt aangegeven door de landbouwer.
X	GW_KG_P_GEBRUIKT	Totale P gebruikt	Totale hoeveelheid meststoffen gebruikt op landbouwgrond van de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekeningswijze (vereenvoudigd): <i>Gebruik P<sub>2</sub>O<sub>5</sub> totaal = + productie P<sub>2</sub>O<sub>5</sub> totaal + productie P<sub>2</sub>O<sub>5</sub> spuistroom + gebruik P<sub>2</sub>O<sub>5</sub> kunstmest + opslagverschil P<sub>2</sub>O<sub>5</sub> dierlijke mest / andere meststoffen + aanvoer P<sub>2</sub>O<sub>5</sub> dierlijke mest / andere meststoffen - afvoer P<sub>2</sub>O<sub>5</sub> dierlijke mest / andere meststoffen.</i>
X	GW_KG_P_OPSLAGVERSCHIL_TOT	Opslagverschil P totaal	Het totale opslagverschil tussen de aangegeven opslag op 1 januari van productiejaar X en de aangegeven opslag op 1 januari van productiejaar X+1, in kg N. Berekend als de som van het opslagverschil voor dierlijke mest en andere meststoffen.
X	GW_KG_P_OVERSCHOT_TOT	Mestoverschot P totaal	Het mestoverschot van de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Berekend als het verschil tussen <i>Productie P<sub>2</sub>O<sub>5</sub> totaal</i> en <i>Afzetmogelijkheid P<sub>2</sub>O<sub>5</sub> totaal</i>
X	GW_KG_P_PRODUCTIE	Productie P totaal	Totale dierlijke mestproductie op de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Voor elke diercategorie (bv. melkkoeien) van elke diersoort (bv. runderen), wordt de productie berekend op basis van het aantal dieren x de uitscheidingsnom per dier voor P <sub>2</sub> O <sub>5</sub> . De uitscheidingsnomen zijn afhankelijk van de diercategorie en de nutriëntenbalans (voor varkens en pluimvee).

Opgenomen in analyse	Parameter <sup>5</sup>	Shortname	Verklaring
			Door de productie per diercategorie van alle diercategorieën van de exploitatie te sommeren, wordt de totale dierlijke mestproductie op de exploitatie in een bepaald productiejaar berekend.
X	GW_KG_P_PRODUCTIE_ANDER	Productie P andere dieren	Dierlijke mestproductie van de diersoort andere dieren, op de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Voor elke diercategorie van de diersoort andere dieren (dit zijn konijnen, schapen, geiten en nertsen), wordt de productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor P <sub>2</sub> O <sub>5</sub> . Vervolgens wordt de productie van alle diercategorieën van de diersoort andere dieren gesommeerd.
X	GW_KG_P_PRODUCTIE_PAARD	Productie P paarden	Dierlijke mestproductie van de diersoort paarden, op de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Voor elke diercategorie van de diersoort paarden, wordt de productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor P <sub>2</sub> O <sub>5</sub> . Vervolgens wordt de productie van alle diercategorieën van de diersoort paarden gesommeerd.
X	GW_KG_P_PRODUCTIE_PLUIM	Productie P pluimvee	Dierlijke mestproductie van de diersoort pluimvee, op de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Voor elke diercategorie van de diersoort pluimvee, wordt de productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor P <sub>2</sub> O <sub>5</sub> . Vervolgens wordt de productie van alle diercategorieën van de diersoort pluimvee gesommeerd.
X	GW_KG_P_PRODUCTIE_RUND	Productie P runderen	Dierlijke mestproductie van de diersoort runderen, op de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Voor elke diercategorie van de diersoort runderen, wordt de productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor P <sub>2</sub> O <sub>5</sub> . Vervolgens wordt de productie van alle diercategorieën van de diersoort runderen gesommeerd.
X	GW_KG_P_PRODUCTIE_SPUISTR	Productie P spuistroom	Productie van spuistroom in kg P <sub>2</sub> O <sub>5</sub>
X	GW_KG_P_PRODUCTIE_VARKEN	Productie P varken	Dierlijke mestproductie van de diersoort varkens, op de exploitatie in een bepaald productiejaar, in kg P <sub>2</sub> O <sub>5</sub> . Voor elke diercategorie van de diersoort varkens, wordt de productie berekend op basis van het aantal dieren x de uitscheidingsnorm per dier voor P <sub>2</sub> O <sub>5</sub> . Vervolgens wordt de productie van alle diercategorieën van de diersoort varkens gesommeerd.
X	OPP_NETTO_m2	Oppervlakte teelt	Totale oppervlakte van de percelen behorend tot de exploitatie in een bepaald productiejaar, in m <sup>2</sup> .

## A.5 Identificatie Exploitatie

Tabel Bijlage A-5: Overzicht van de beschikbare identificatiegegevens van de Exploitaties.

Opgenomen in analyse	Parameter	Verklaring
	NR_LANDBOUWER_FICT	Alle landbouwbedrijven hebben een uniek landbouwnummer, exploitantnummer(s) en exploitatienummer(s). Als een landbouwbedrijf uit meerdere exploitaties op verschillende locaties bestaat, dan beschikt dit landbouwbedrijf over één uniek landbouwnummer, met meerdere, unieke exploitatienummers. Eén exploitatie komt overeen met één bepaalde locatie (één adres), waar één of meerdere stallen gelegen zijn en/of waartoe één of meerdere landbouwpercelen toe behoren. Een landbouwperceel moet niet in de nabije omgeving van het exploitatieadres gelegen zijn. In het kader van privacy, werden de landbouwer-, exploitant- en exploitatienummers gecodeerd en werden fictieve nummers uitgewisseld.
	NR_EXPLOITANT_FICT	
	NR_EXPLOITATIE_FICT	
	OMS_FUSIEGEMEENTE_EXPLOITATIE_UI	Gemeente van de exploitatie
	CO_NIS_FUSIEGEMEENTE_EXPLOITATIE	NIS code gemeente van de exploitatie
	STRAAT_NR_BUS_EXPLOITATIE_UITBAT	Adresgegevens van de exploitatie

## A.6 Perceel

*Tabel Bijlage A-6: Overzicht van de beschikbare perceelsgegevens, met aanduiding welke parameters opgenomen werden in de statistische analyse*

Opgenomen in analyse	Parameter	Shortname	Verklaring
	UNIEKEID_FICT	uniek perceelsnummer	uniek perceelsnummer opgebouwd uit jaartal-landbouwer-perceelssequentie. In het kader van privacy, werden de unieke perceelsnummers gecodeerd en werden fictieve nummers uitgewisseld.
	NR_LANDBOUWER_FICT	Landbouwnummer	Alle landbouwbedrijven hebben een uniek landbouwnummer, exploitantnummer(s) en exploitatienummer(s). Als een landbouwbedrijf uit meerdere exploitaties op verschillende locaties bestaat, dan beschikt dit landbouwbedrijf over één uniek landbouwnummer, met meerdere, unieke exploitatienummers. Eén exploitatie komt overeen met één bepaalde locatie (één adres), waar één of meerdere stallen gelegen zijn en/of waartoe één of meerdere landbouwpercelen toe behoren. Een landbouwperceel moet niet in de nabije omgeving van het exploitatieadres gelegen zijn. In het kader van privacy, werden de landbouwer-, exploitant- en exploitatienummers gecodeerd en werden fictieve nummers uitgewisseld.
	NR_EXPLOITANT_FICT	exploitant	
	NR_EXPLOITATIE_FICT	exploitatie	
	CMP_COD	campagnecode	Jaar waarin het perceel aangegeven is
X	AANV_OPP_M2	aanvaarde oppervlakte	Oppervlakte van het perceel, in m <sup>2</sup>
X	GWS_COD	hoofdteelt	Code van de hoofdteelt  In de analyse krijgt deze één van volgende achtervoegsel: <ul style="list-style-type: none"> <li>• <b>_Mestdecreet</b>: geeft aan van welke gewasgroep een gewas geteeld werd. <ul style="list-style-type: none"> <li>○ gewasgroepcodes zoals weergegeven in Tabel Bijlage A-7</li> <li>○ 0: niet gespecificeerd (het geval als er geen gewas gedefinieerd werd)</li> </ul> </li> <li>• <b>_Vanggewas</b>: geeft aan of het type gewas al dan niet een vanggewas is. Mogelijke achtervoegsel: <ul style="list-style-type: none"> <li>○ Ja: het betreft een vanggewas</li> <li>○ Nee: het betreft geen vanggewas</li> <li>○ 0: niet gespecificeerd (het geval als er geen gewas gedefinieerd werd)</li> </ul> </li> </ul> Voorbeeld: GWSCOD_MestdecreetG duidt op een hoofdteelt van Grassen
X	GWS_COD1	voorteelt	Code van de voorteelt  In de analyse krijgt deze één van volgende achtervoegsel: <ul style="list-style-type: none"> <li>• <b>_Mestdecreet</b>: geeft aan van welke gewasgroep een gewas geteeld werd. <ul style="list-style-type: none"> <li>○ gewasgroepcodes zoals weergegeven in Tabel Bijlage A-7</li> <li>○ 0: niet gespecificeerd (het geval als er geen gewas gedefinieerd werd)</li> </ul> </li> <li>• <b>_Vanggewas</b>: geeft aan of het type gewas al dan niet een vanggewas is. Mogelijke achtervoegsel: <ul style="list-style-type: none"> <li>○ Ja: het betreft een vanggewas</li> <li>○ Nee: het betreft geen vanggewas</li> <li>○ 0: niet gespecificeerd (het geval als er geen gewas gedefinieerd werd)</li> </ul> </li> </ul> Voorbeeld: GWSCOD1_MestdecreetG duidt op een voorteelt van Grassen
X	GWS_COD2	nateelt	Code van de nateelt  In de analyse krijgt deze één van volgende achtervoegsel:

Opgenomen in analyse	Parameter	Shortname	Verklaring
			<ul style="list-style-type: none"> <li>_Mestdecreet: geeft aan van welk gewasgroep een gewas geteeld werd. Mogelijke achtervoegsel: <ul style="list-style-type: none"> <li>gewasgroepcodes zoals weergegeven in Tabel Bijlage A-7</li> <li>0: niet gespecificeerd (het geval als er geen gewas gedefinieerd werd)</li> </ul> </li> <li>_Vanggewas: geeft aan of het type gewas al dan niet een vanggewas is. Mogelijke achtervoegsel: <ul style="list-style-type: none"> <li>Ja: het betreft een vanggewas</li> <li>Nee: het betreft geen vanggewas</li> <li>0: niet gespecificeerd (het geval als er geen gewas gedefinieerd werd)</li> </ul> </li> </ul> <p>Voorbeeld: GWSCOD2_MestdecreetG duidt op een nateelt van Grassen</p>
X	TH_BODEM	thema bodem	Code die aangeeft of een perceel volgens art. 3, §7 van het Mestdecreet onder 'zandgronden' of 'zware kleigronden' valt. K (of Zwarte klei): perceel is een zware kleigrond Z (of zand): perceel is een zandgrond N (of O): perceel is een andere bodem
	GEMEENTE	gemeente	Gemeente waarbinnen het grootste deel van het perceel gelegen is
X	LANDBCODE	code landbouwstreek	Code van de landbouwstreek 0: O, onbepaald 15(0): Duinen 16(0): Polders 18(0): Vlaamse Zandstreek 22(0): Kempen 23(0): Zandleemstreek 27(0): Leemstreek 28(0): Weidestreek
X	TOEK_DA	derogatie	Code die aangeeft of op het perceel derogatie werd toegekend of niet J: ja N: neen G: Geen derogatie aangevraagd
X	TOEK_BKM	BKM goedgekeurd	Code die aangeeft of de aanvraag tot beheerovereenkomst kunstmest werd goedgekeurd of niet J: goedgekeurd N: niet goedgekeurd G: Geen BKM aangevraagd
X	N_TOT	Norm N totaal	Totale afzetmogelijkheid voor meststoffen, in kg N, op het perceel. Berekend als de norm voor totale N, in kg N/ha, vermenigvuldigd met de oppervlakte van het perceel
X	N_DIER	Norm N dier	Afzetmogelijkheid voor dierlijke, in kg N, op het perceel. Berekend als de norm voor N uit dierlijke mest, in kg N/ha, vermenigvuldigd met de oppervlakte van het perceel
X	N_ANDERE	Norm N ander	Afzetmogelijkheid voor andere meststoffen, in kg N, op het perceel. Berekend als de norm voor N uit andere meststoffen, in kg N/ha, vermenigvuldigd met de oppervlakte van het perceel
X	N_KUNST	Norm N kunstmest	Afzetmogelijkheid voor kunstmest, in kg N, op het perceel Berekend als de norm N uit kunstmest, in kg N/ha, vermenigvuldigd met de oppervlakte van het perceel
X	P_TOT	Norm P totaal	Totale afzetmogelijkheid voor meststoffen, in kg P <sub>2</sub> O <sub>5</sub> , op het perceel. Berekend als de norm voor totale P <sub>2</sub> O <sub>5</sub> , in kg P <sub>2</sub> O <sub>5</sub> /ha, vermenigvuldigd met de oppervlakte van het perceel
X	P_DIER	Norm P dier	Afzetmogelijkheid voor dierlijke, in kg P <sub>2</sub> O <sub>5</sub> , op het perceel. Berekend als de norm voor P <sub>2</sub> O <sub>5</sub> uit dierlijke mest, in kg P <sub>2</sub> O <sub>5</sub> /ha, vermenigvuldigd met de oppervlakte van het perceel
X	P_KUNST	Norm P kunstmest	Afzetmogelijkheid voor kunstmest, in kg P <sub>2</sub> O <sub>5</sub> , op het perceel Berekend als de norm P <sub>2</sub> O <sub>5</sub> uit kunstmest, in kg P <sub>2</sub> O <sub>5</sub> /ha, vermenigvuldigd met de oppervlakte van het perceel
X	P_ANDERE	Norm P andere mest	Afzetmogelijkheid voor andere meststoffen, in kg P <sub>2</sub> O <sub>5</sub> , op het perceel. Berekend als de norm voor P <sub>2</sub> O <sub>5</sub> uit andere meststoffen, in kg P <sub>2</sub> O <sub>5</sub> /ha, vermenigvuldigd met de oppervlakte van het perceel

Opgenomen in analyse	Parameter	Shortname	Verklaring
X	EROSIEG	Erosiegevoeligheid	Indicatie van de potentiële erosiegevoeligheid van een perceel volgens erosiegevoeligheidskaart van Vlaanderen  LICHT MATIG HOOG ANDERE: percelen die volgens de erosiegevoeligheidskaart niet erosiegevoelig zijn, maar waarvan op terrein vastgesteld is dat ze wel bijdragen tot de erosieproblemen. 0: perceel is niet erosiegevoelig
X	PROD_METH	productiemethode	Productiemethode BGG: blijvend grasland met voorgedrukte status BG dat in 2017 gescheurd is of zal worden gescheurd CIV: teelt in containers op/in volle grond (bv. chrysanten) CON: teelt op groeimedium in open lucht (bv. containervelden) EBT: erosiebestrijdende teelttechniek (mulchzaai, strip-till of >80% waterdoorlatende bodembedekking) LOO: loods (voor plantaardige productie, bv. witloof) MAA: hoofdteelt gras wordt uitsluitend gemaaid en geoogst NPO: niet-permanent overkapte teelt op groeimedium PLA: niet-permanente overkapte teelt in volle grond (plastic) SER: serres met teelt in volle grond SGM: serres met teelt op groeimedium TOR: Teelt op ruggen TNR: Teelt niet op ruggen
X (enkel in het respectievelijke aggregatieniveau)	Opp_perc_in_OW_m <sup>3</sup> Opp_perc_in_GW_m <sup>3</sup> Opp_perc_in_GEM_m <sup>3</sup>		Oppervlakte aan percelen binnenin het respectievelijke afstroomgebied (OW), intrekgebied (GW) of Gemeente (GEM)

## A.7 Gewasgroepen en Vanggewassen

*Tabel Bijlage A-7: Overzicht van de gewasgroepen in de perceelsgegevens (2007-2014)*

Code Gewasgroep	Omschrijving
A	Andere gewassen
AA	Aardappelengeroid na 31 juli
AG	Andere granen
G	Grassen
G1	Groenten groep I
G2	Groenten groep II
G3	Groenten groep III
Grasland	Grassen
L	Leguminosen andere dan erwten en bonen
M	Mais
N	Gewassen met lage stikstofbehoefte
S	Suikerbieten
V	Voederbieten
WT	Wintertarwe en Triticale
Z	Geen bemesting

*Tabel Bijlage A-8: Overzicht van de vanggewassen (2014)*

Omschrijving teelt	Teelt code	Code Gewasgroep Mestdecreet VA	Omschrijving Gewasgroep Mestdecreet VA	Vanggewas
Bladrammenas	656	A	Andere gewassen	Ja
Eénjarige grasklaver	701	G	Grassen	Ja
Facelia	645	A	Andere gewassen	Ja
Festulolium	638	G	Grassen	Ja
Gele mosterd	643	A	Andere gewassen	Ja
Graszoden	955	G	Grassen	Ja
Haver	34	AG	Andere granen	Ja
Meerjarige grasklaver	702	G	Grassen	Ja
Nyger	649	A	Andere gewassen	Ja
Rammenas industriële verwerking	8550	G3	Groenten groep III	Ja
Rammenas vers gebruik	9550	G3	Groenten groep III	Ja
Snijrogge	639	AG	Andere granen	Ja
Soedangras	653	AG	Andere granen	Ja
Tagetes (Afrikaantje)	646	A	Andere gewassen	Ja
Tijdelijk grasland	62	G	Grassen	Ja
Voederkool (bladkool)	741	G1	Groenten groep I	Ja

## A.8 Overige parameters

Tabel Bijlage A-9: Overzicht van de overige beschikbare parameters

Opgenomen in analyse	Parameter <sup>6</sup>	Omschrijving
X	PF_mean	Procesfactor oppervlaktewater (enkel opgenomen in analyses betreffende oppervlaktewater)
	Ndepo	Stikstofdepositie
X	NeerslagOpDeDag	Neerslag op de dag van de meting (enkel OW)
X	NeerslagWeekErvoor	neerslag in de week voorafgaand aan de meting
X	NeerslagMaandErvoor	neerslag in de maand voorafgaand aan de meting
X	Neerslag3MaandErvoor	neerslag in de periode 3 maand voorafgaand aan de meting
X	AfvoerPercentiel	Afvoerpercentiel op het moment van de meting (enkel OW)
X	Avg_Runoff	De gemiddelde afvoercoëfficiënt van regenwater binnen een welbepaalde oppervlakte
X (enkel op aggregatieniveau gemeente)	AREA	Oppervlakte van de gemeentes (m <sup>2</sup> )
X (enkel op aggregatieniveau afstroomgebied)	opp_afwateringsgebied_m2	Oppervlakte van de afstroomgebieden (m <sup>2</sup> )

<sup>6</sup> Na aggregatie over een jaartal kunnen deze parameters in het rapport volgende prefixen kennen:

- *max\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de maximale responswaarde geregistreerd werd
- *min\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de minimale responswaarde geregistreerd werd
- *median\_val*: wijst op de meetwaarde die overeenstemt met de meting waar de mediaan responswaarde geregistreerd werd



## Bijlage B CCM

## B.1 Intro

When it comes to evaluate a causal relationship between two variables monitored by time series, the standard in science remains the Pearson correlation coefficient. However, such a measure is limited to the evaluation of the linear (proportional) relationships between variables and encounters the additional well-known drawback that “correlation does not imply causation”.

Recently, another pairwise causal detection method has been published to overcome these limits, namely Convergent Cross Mapping (CCM, (Sugihara et al., 2012). Its applicability extends to the case of weakly nonlinearly coupled time series. Besides, CCM provides a causal direction.

This study evaluates CCM applied to hydrological river level and groundwater level time series around the Train river basin in Wallonia. This data set has been used, since CCM needs detailed (i.e. high time resolution) and long term time series. Such long term time series from VLM were not available at the time of exploring the technique.

## B.2 Description of Convergent Cross Mapping

CCM provides empirical evidence that two time series belong to the same dynamical system and are therefore causally related. For two given time series  $X$  and  $Y$ , CCM test proceeds in three steps:

1. *Iterative sampling with increasing length:*

CCM samples many segments of  $X$  and  $Y$  having a given length  $L$  then increases  $L$  iteratively;

2. *State Space Reconstruction:*

On each sample of  $X$  and  $Y$ , CCM performs State Space Reconstruction. It consists in embedding the time series: lagging it several time ( $E$  times,  $E$  known as the embedding dimension) with a given time delay  $\tau$ . This process retrieves a space holding a geometric representation of the historical trajectories of the variable, i.e. a reconstructed state space in which the reconstructed geometric feature represents the trajectories;

3. *Cross Map predictions:*

CCM applies predictions between those reconstructed spaces and see if the prediction skills increases on average as the sampling length increases.

So the rise (convergence) in the cross-map prediction skills is used as the criterion for assessing a causal relationship. If convergence is observed when performing cross map predictions from the  $Y$  reconstructed space to the  $X$  one, it means that some information about the dynamic of  $X$  is included in the dynamic of  $Y$ , which is possible if  $X$  influenced  $Y$ . When convergence is also observed using cross map predictions from  $X$  space to  $Y$  space, the causal relationship is bidirectional. It is important to notice that the causal direction is the opposite of the cross-map prediction direction.

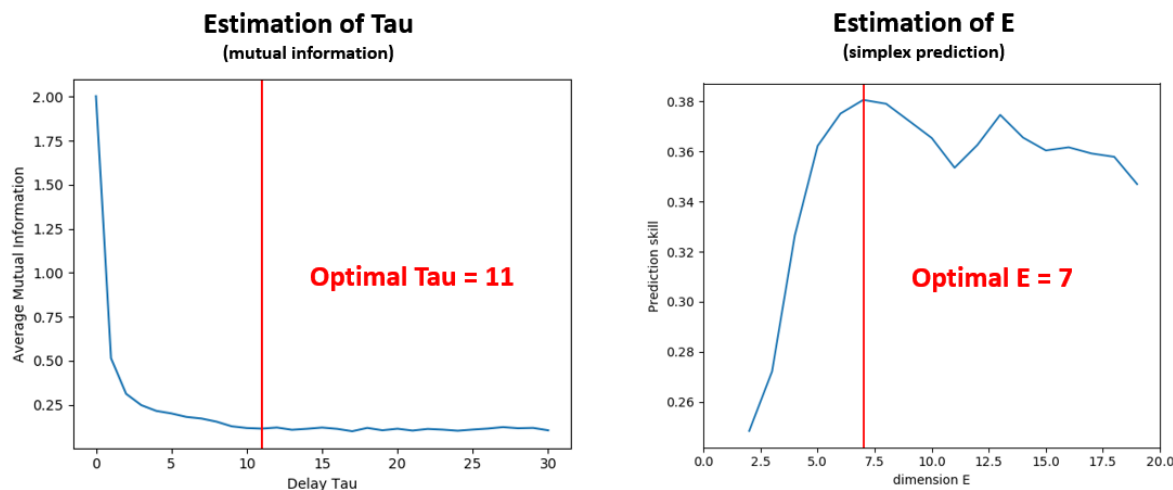
## B.3 Applying Convergent Cross Mapping

CCM is ideally applied to two long, noise free, stationary (without significant changes in the statistical properties of the time series across time), synchronized and evenly sampled time series. CCM therefore requires as first step a pre-processing of the time series with the objective of harmonizing the time steps, removing trends and reducing noise. Any method could be used to remove trend and noise but according to recent publications, *Singular Spectrum Analysis* exhibits better performance combined with CCM (Huang *et al.*, 2017) in climate change studies.

Then, as a second step, the two embedding parameters for state space reconstruction should be estimated: the time delay  $\tau$  and the dimension  $E$ . The strategy is to find  $\tau$  such that  $\tau$  is high enough to unrelate the lags but not too high to keep relevant trajectories in the reconstructed phase space. The dimension  $E$  should be also high enough to remove uncertainties in the trajectories directions. However, the higher  $E$ , the longer the time series should be as for each increasing dimension an amount of data is lost in the process of embedding.

There is a plethora of methods for estimating  $\tau$  and  $E$  separately or simultaneously as discussed in literature devoted to nonlinear dynamical systems. One of the mostly used method for estimating  $\tau$  is to use the first local minimum of the *average mutual information* applied to successive lags of the time series. Average mutual information is a statistical estimation of the nonlinear dependence between two dataset. In the *R package for Empirical Dynamic Modeling and Convergent Cross Mapping* (Ye *et al.*, 2016),  $E$  is found optimal regarding the performance of *simplex prediction*, a nonlinear prediction methods relying on state space reconstruction.

Both methods are illustrated hereafter using average mutual information from the R package *timeseriesChaos* (Fabio Di Narzo, 2017) and the simplex function of the rEDM package (Ye, 2016).



**Figure 1: Example of embedding parameters estimation using the first local minimum of the delayed mutual information as optimal  $\tau$  and the first local maximum of the simplex prediction skills as optimal  $E$**

## B.4 Study case: CCM applied to hydrological data

UCL provided a notebook where data preprocessing through Singular Spectrum Analysis and parameter estimation for Tau and E estimation methods are largely discussed. The present analysis intends to be more concise and straightforward regarding the conclusion.

### B.4.1 Data and Study site

The analysis is performed on data from the *Service Public de Wallonie* (SPW) from the river gauging data portal Aqualim and the Piez'eau data portal for groundwater level monitoring. The study site is the 1 / verbouwingsgraad of the Train river basin, tributary of the Dyle river, in the Province of Brabant Wallon, Northern Wallonia. The dataset includes two sets of limnimetric (river water level) data. The first one L1 is sited in Archennes on the Train river covering the basin over 73 km<sup>2</sup>. The second L2 is sited downstream on the Dyle river draining 430 km<sup>2</sup> including the Train river basin. The P1 and P2 are two piezometric stations sited in the Nethen river basin, tributary of the Dyle but outside of the Train drainage 1 / verbouwingsgraad.

Tertiary sands cover the whole 1 / verbouwingsgraad (Sables du Bruxeliens, Code SPW 051). But P1 and P2 are sited in the lower body made of clay (Craies du Brabant, Code SPW 080). P4 stands also in a lower geological unit (Socle du Brabant, Code SPW 160). This geological unit is not a regional aquifer but consists only of small local aquifers disconnected from each other. There is therefore no overall regional behavior for this groundwater body.

The dataset contains 1460 daily values from 2011 to 2015.

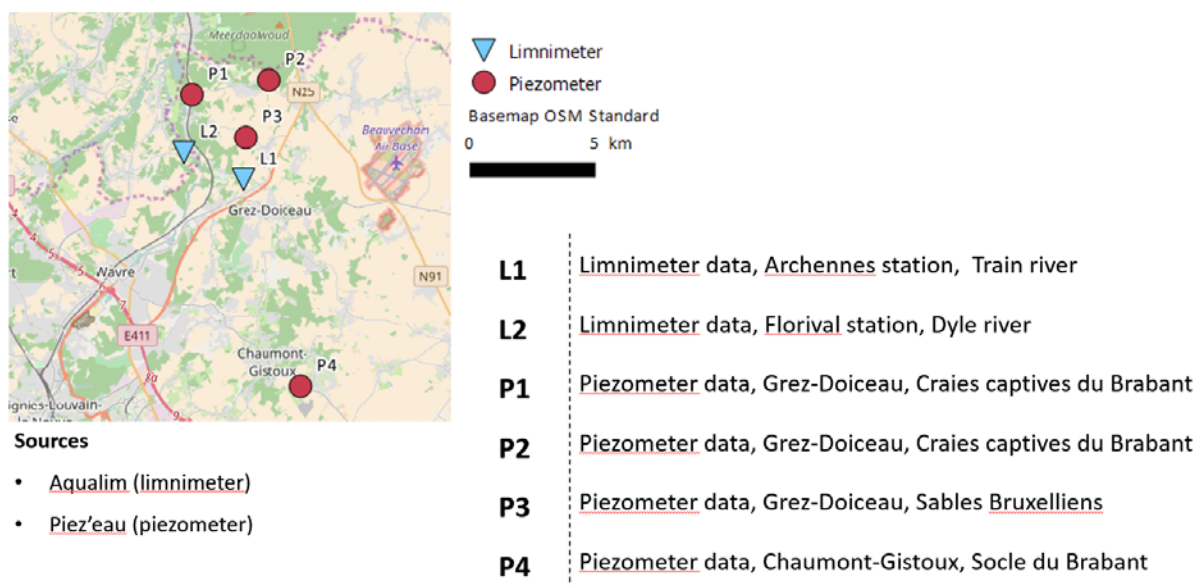


Figure 2: Study site and data

## B.4.2 Motivation

The Train river basin is actually an interesting river basin regarding causal analysis where UCL has already acquired some expertise across different studies (e.g. (Abdeslam, 1998)). Because the Train river basin lies on sandy materials, the Archennes gauging station L1 cannot have a proper rating curve for discharge estimates. The reason is that sand banks disrupt the profile of the riverbed. In addition, the topographical drainage basin do not reflect the real drainage as groundwater outside the Train surface catchment 1/ verbouwingsgraad contributes to the Train discharge. For instance, groundwater exchanges were reported between the Nethen catchment and the Train catchment. Finally, groundwater has a high contribution in the streamflow.

Considering these elements, traditional methods involving hydrological modelling of the mass balance over the topographically delineated drainage 1 / verbouwingsgraad and allocating high importance to rainfall runoff processes is a wrong approach for such a basin for investigating daily dynamics. Applying CCM on such a dataset is way to assess CCM potential to reveal these complex surface-groundwater interactions without the need of heavier methods as hydrogeological studies or hypothesis testing through hydrological modelling.

## B.4.3 Methodology

The six time series have been preprocessed using Singular Spectrum Analysis to remove the trend and the seasonal dynamics from the times series and we kept the residuals to perform the analysis. No noise has been removed, as the noise of the daily measurement cannot be easily distinguished from the daily dynamics.

The time delay parameter  $\tau$  was determined as the first local minimum of the mutual information and the optimal embedding dimension  $E$  as the first local maximum of the simplex prediction (see section 3).

Then, we applied CCM to perform cross-map predictions from the residuals of L1 to predict the other variables residuals (L2, P1 to P4) to evaluate which variables affect the daily discharge of the Train river. Finally, we performed cross-map predictions the other way by targeting residuals of L1 from the other residuals (L2, P1 to P4). The later aims to evaluate which variables are influenced by the Train river daily discharge. The time series length has been increased up to 550 days.

## B.4.4 Results

### *Signal preprocessing*

The following figure presents the results of the removal of trend and seasonality using Singular Spectrum Analysis. The time series presented are the residual and the correlation matrix for residuals.

### Correlations

	L1_res	L2_res	P1_res	P2_res	P3_res	P4_res
L1_res	1.00	0.84	0.07	0.08	-0.06	0.03
L2_res	0.84	1.00	0.08	0.15	0.05	-0.03
P1_res	0.07	0.08	1.00	0.34	0.04	-0.02
P2_res	0.08	0.15	0.34	1.00	0.48	0.30
P3_res	-0.06	0.05	0.04	0.48	1.00	0.01
P4_res	0.03	-0.03	-0.02	0.30	0.01	1.00

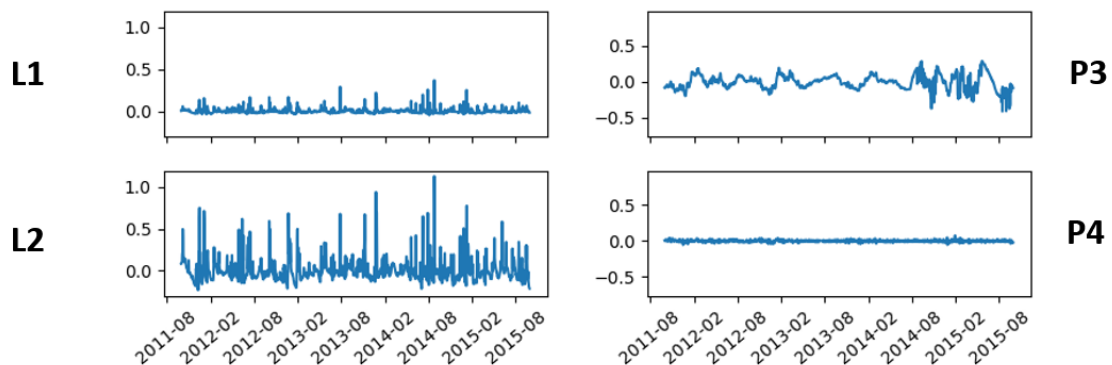


Figure 3: Correlation and residual time series after trend and seasonality extraction (unit: relative meter)

While the trend has disappeared, some residuals still contains a bit of seasonality mainly for the P3 time series. It is important to notice that the dynamic of P3 between August 2014 and 2015 is strange. This may have disrupted the signal pre-processing and it will probably affect the result of CCM.

Regarding the correlations, L1 and L2 are unsurprisingly correlated, as we would expect from two river gauging station on the same river network. Correlation between river level and groundwater level residual is inexistent. P2 and P3 are the most correlated groundwater levels (0.48), then P1 and P2 (0.34) but P1 and P3 are uncorrelated.

#### Parameters estimation

The following table shows the result of the parameters Tau and E estimation applied to the residuals. The results are variable. One may noticed that P3 (which has still a seasonal component in the residuals) has the longer time delay. In addition, the piezometer that shows the most variable daily dynamics after the preprocessing (P1) has the higher dimension.

Residuals	Tau	E
L1	11	7
L2	9	4
P1	4	11
P2	13	4
P3	20	4
P4	4	8

### Convergent Cross Mapping

The following figure shows the results of the CCM causal test. Convergence, i.e. a rise in the prediction skills with increasing time series length, is the criterion for causation.

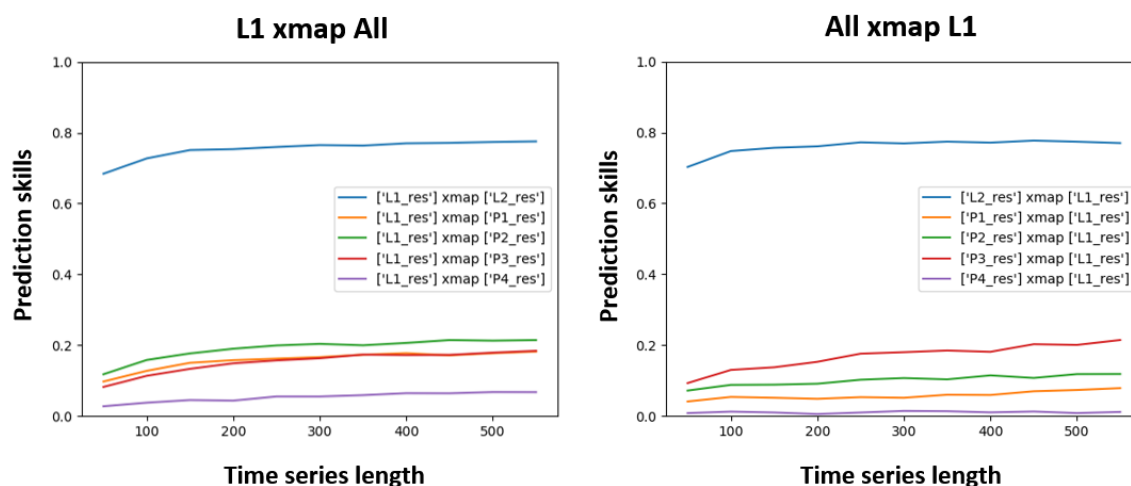


Figure 4: Results of the CCM analysis

A blind interpretation of CCM results suggests that:

- L1 and L2 have strong bidirectional relationship (L1 influences L2 and vice-versa)
- L1 and P3 share a bidirectional relationship of the same order of magnitude
- L1 is more influenced by P2 than the opposite
- L1 is more influenced by P1 than the opposite
- L1 is slightly influenced by P4 but not the opposite

These results are discussed further in the next section.

## B.4.5 Discussion

One of the most astonishing results is certainly the strong bidirectional relationship between L1 and L2. Indeed, as the Train river is a tributary of the Dyle, anyone would expect that L1 causes L2. Although a bidirectional relationship is not surprising from two close river monitoring level stations. The downstream flow certainly affects the dynamics of the upstream flow by changing the flow resistance at the river meeting point. Besides, both stations shared the same flood envelope. Nevertheless, this dynamics should only be noticeable during flood events and we still expect that L1 has more influence on L2 than the opposite.

The explanation lies in the strong correlation between the two time series (0.84). In such case, the information in the reconstructed state space are too similar because the same drivers force both dynamics: mainly the same rainfall pattern and the same recharge from groundwater. Moreover, the transit time from L1 to L2 is completely occluded by the daily time steps. Such a problem is named synchronization and it is mentioned in the supplementary materials of the original paper (Sugihara *et al.*, 2012).



This pathological case is hard to distinguish from bidirectional causality and does not provide much more information than the Pearson correlation coefficient. This raises the question: where stand the limits between strongly and weakly coupled dynamics?

Considering the other results, these are less surprising considering the prior knowledge about the basin. First, the three groundwater stations P1, P2, P3 exhibit a bidirectional relationship with respect to L1. It makes sense as we assume groundwater exchange from previous existing studies. The bidirectional relationship is more pronounced between L1 and P3. It makes sense considering the proximity of the stations and the fact that P3 is measuring the water table of the Brusselian sand. Finally, the behavior with respect to P4 also makes sense. As sited in the upstream part we expect that change in aquifer storage have less impact on the recharge of the river and vice-versa.

We would probably dive into causal speculation if we push the analysis further. First, it requires reviewing more precisely the hydrogeological processes involved in these groundwater bodies. Second, P3 ranking might be affected by the strange dynamics observed in figure 2. More importantly, we should consider the uncertainty and the lack of robustness associated to the preprocessing and the parameters estimations. Interpreting the relative ranking score of convergence between P1, P2, and P3 should be done with great caution. Indeed, changes in the preprocessing and the parameters selection would probably affect the relative ranking. This raises a second question: how sensitive is CCM to the preprocessing and the parameters selection?

#### B.4.6 Conclusion and perspectives

Results showed that CCM failed to interpret causal direction when two variables are too strongly coupled (high correlation) such as two river gauging station on the same river network. CCM revealed causal relationships between the Train river level and the surrounding water bodies where the Pearson correlation coefficient could not. The results are interpretable in such a way that they match the current knowledge about the hydrology of the Train river basin.

However, in order to apply CCM, it requires to preprocess the data and to select parameters. There exist different methods that are commonly accepted in the literature but leading to different preprocessed signal and parameters. These lead to uncertainty on the CCM results. Therefore, a fine interpretation of the results, especially about the relative ranking of causal variables, should be done with a low level of confidence.

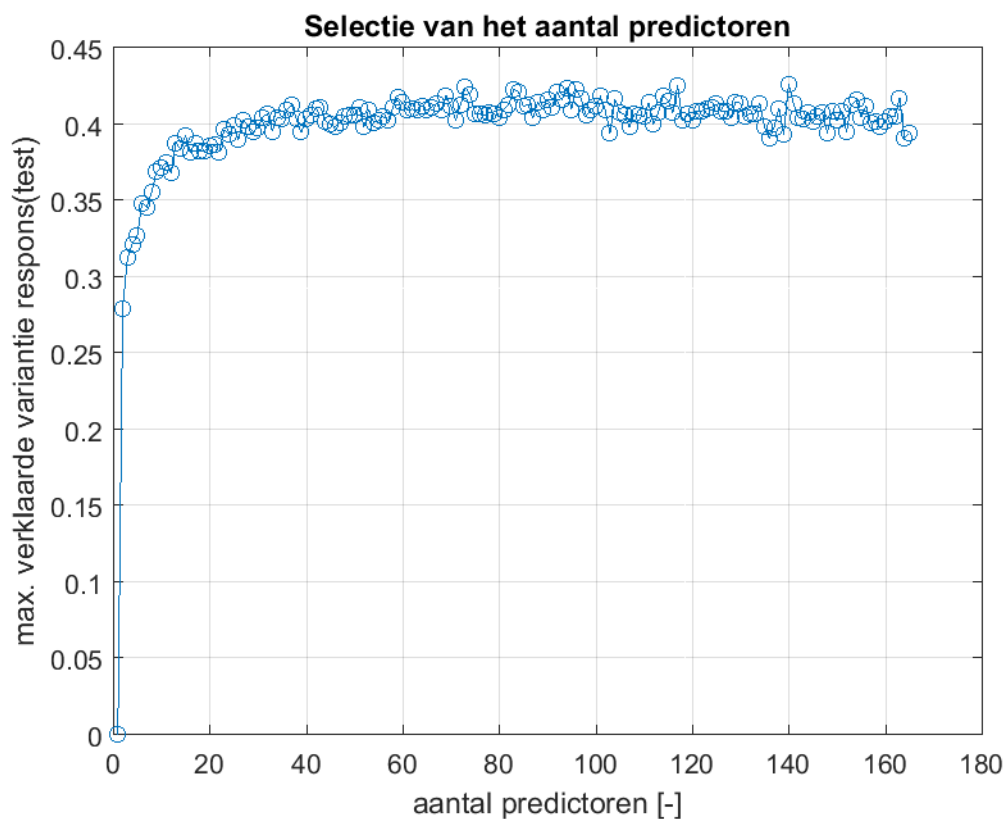
Given the results, CCM proved to be an interesting tool with potential applications in revealing unsuspected relationships or in validating hypothesis on an environmental systems dynamics. Nevertheless, regarding the absence of studies using CCM on hydrological data, the interpretation and the expertise over the method applied in the field of hydrology is still at its early stage.

In order to improve CCM interpretation when applied to hydrological data, UCL intends to perform virtual experiments within the scope of the PhD research of Damien Delforge (UCL). These would allow having control on time series lengths, noise level, optimal parameters selection and the structure of the virtual dynamical system. Applying CCM on such a controlled environment would allow CCM to perform between different couplings of the time series and provides recommendations on the proper preprocessing and parameters selection methods.

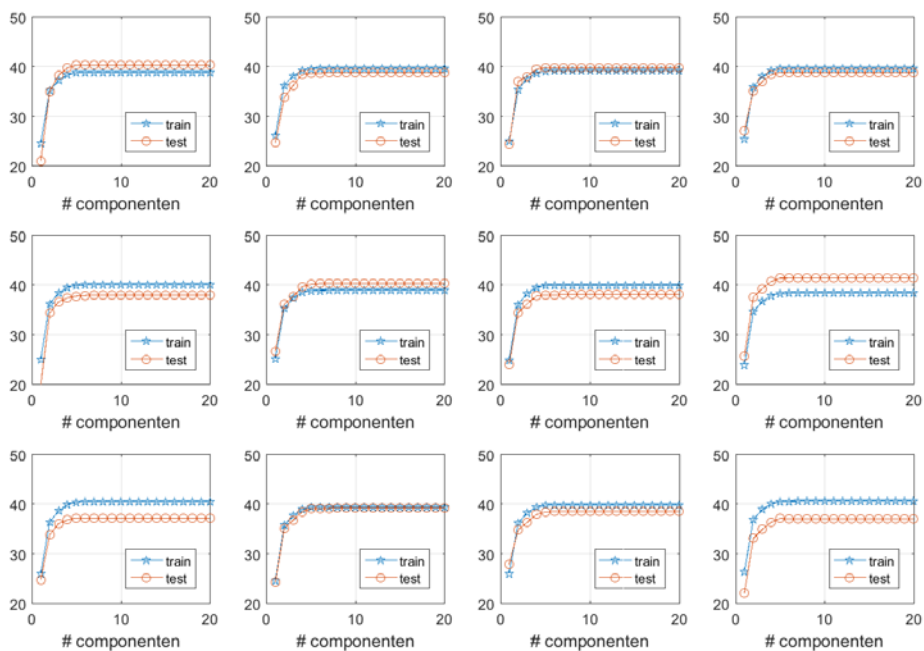
Regarding CCM applicability to a global analysis of water quality over Flanders, the main drawback stands in CCM need for long evenly sampled time series. For this purpose, the data is generally aggregated to bigger time steps (monthly or yearly) and time series lengths are reduced. There is no guarantee that CCM results are relevant over such a short time series length.

If the data become available at higher observed temporal resolution, than another conclusion could probably be drawn. Likewise, CCM should be applicable at a lower spatial scale where temporal data becomes available at higher resolution.

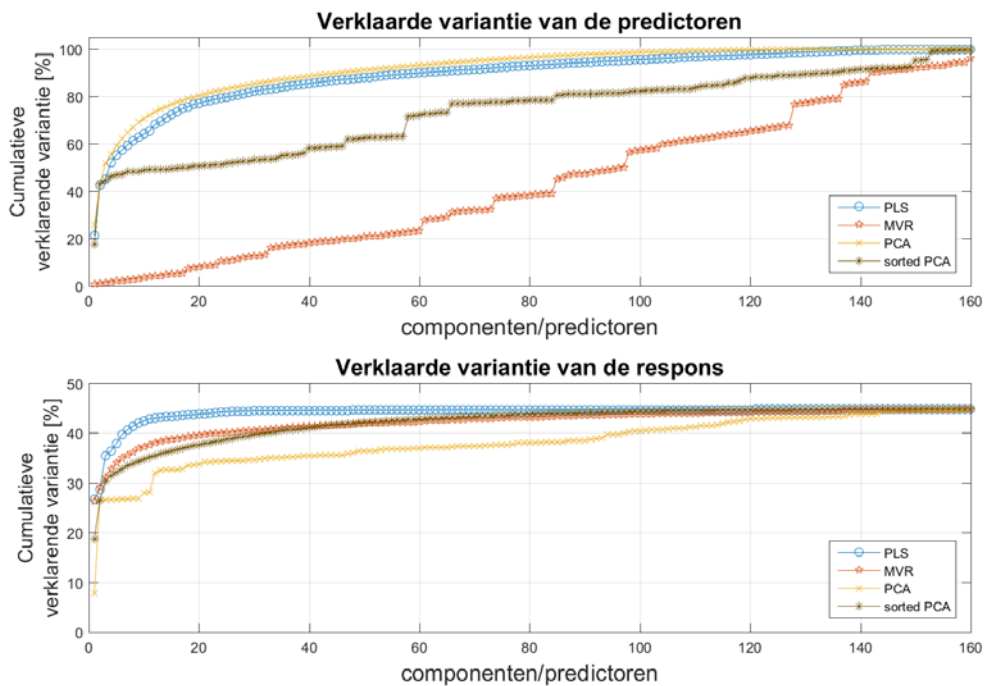
## Bijlage C      **Resultaten globale statistische analyse**



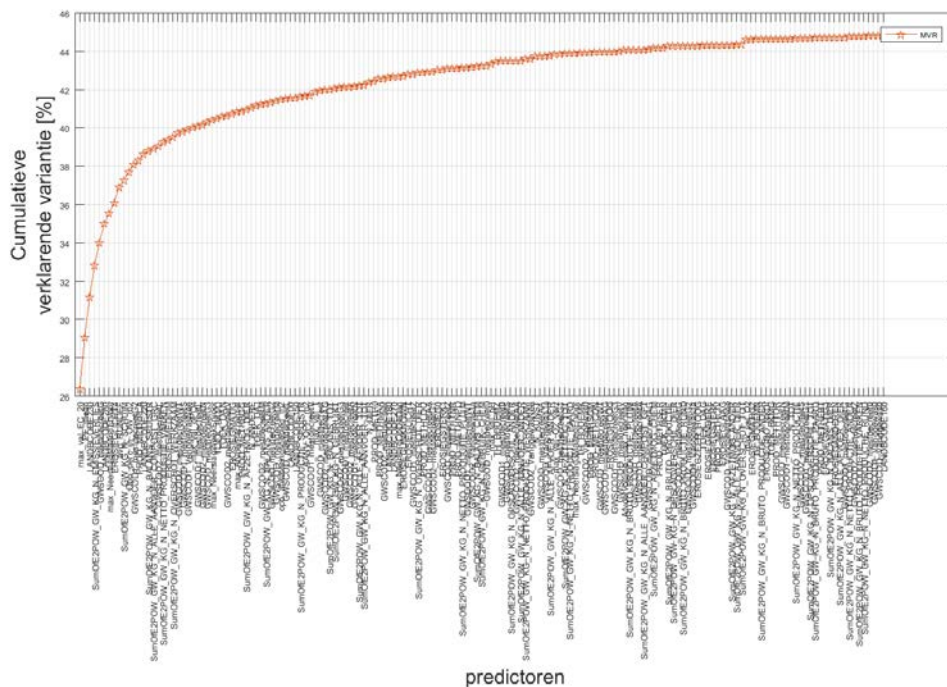
Figuur Bijlage C-1: Aantal predictoren ifv verklarende variantie voor OW, N, Afstroom



Figuur Bijlage C-2: Aantal componenten ifv de verklaarde variantie voor OW, N, Afstroom



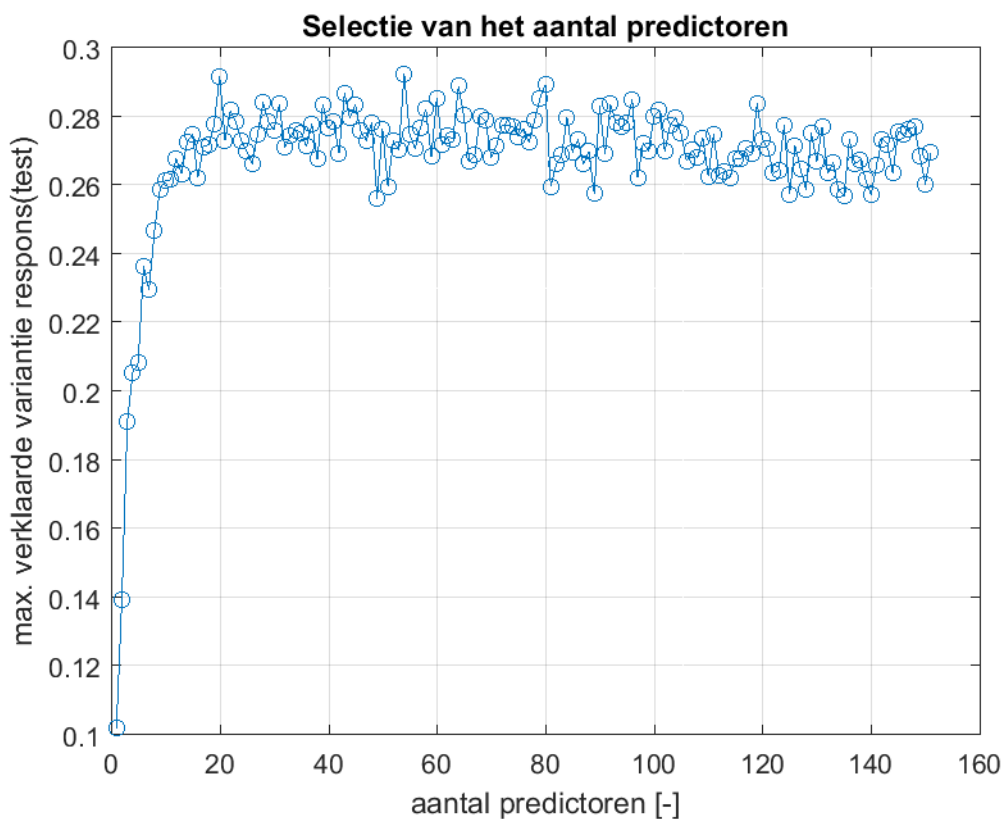
Figuur Bijlage C-3: Selectie van een optimaal model voor OW, N, Afstroom



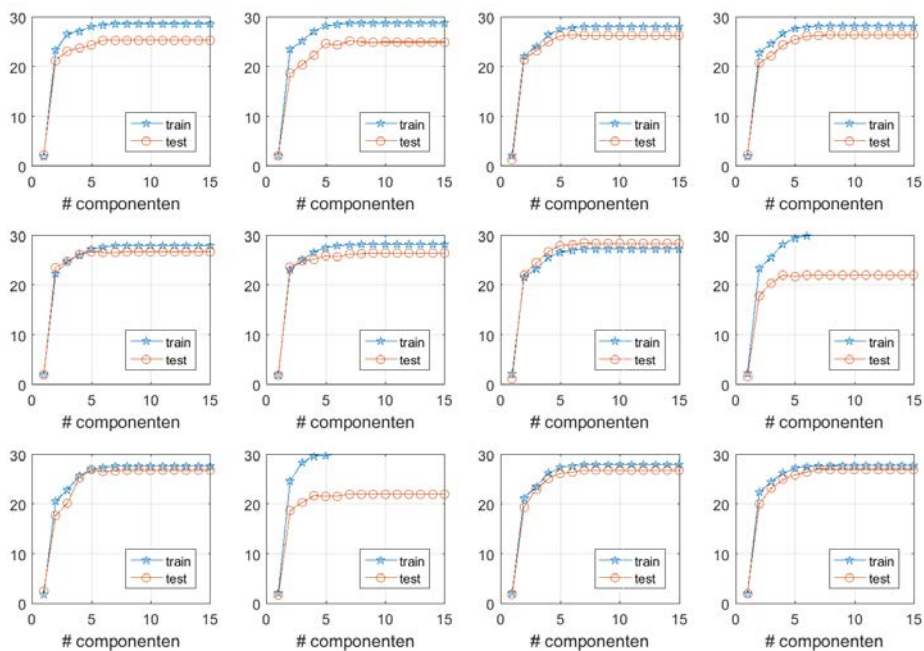
Figuur Bijlage C-4: Volgorde van de predictoren zoals bepaald door stapwise MVR voor OW, N, Afstroom







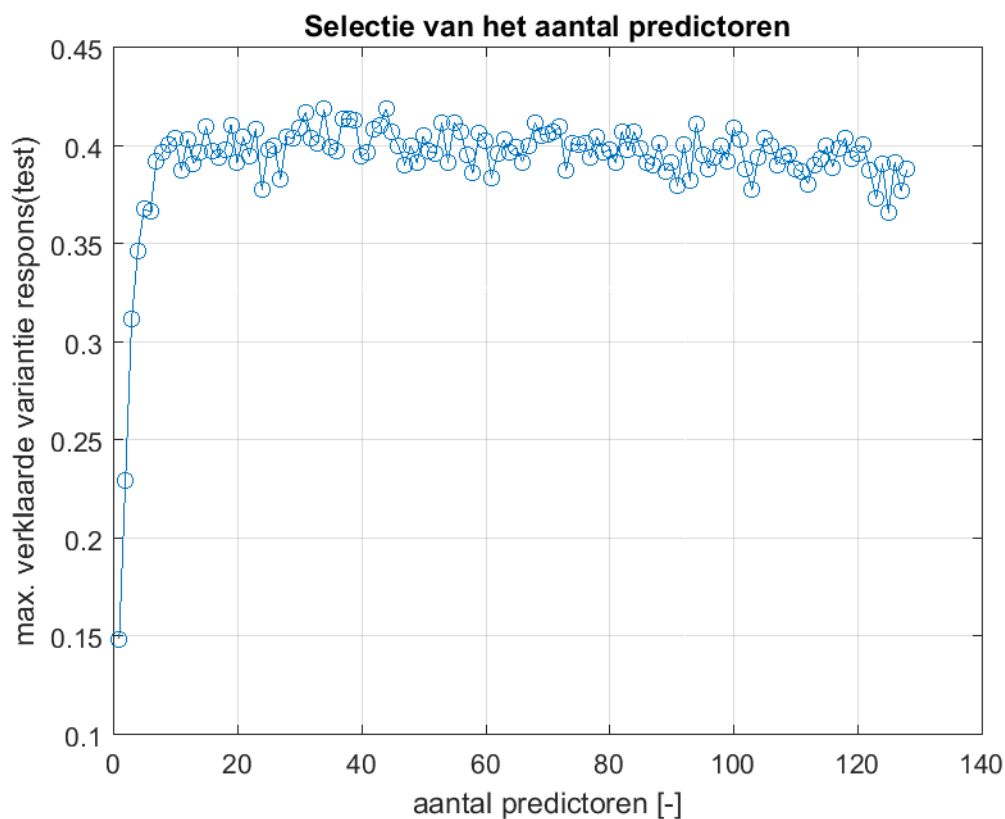
Figuur Bijlage C-9: Aantal predictoren *ifv* verklarende variantie voor GW, N, Intrek



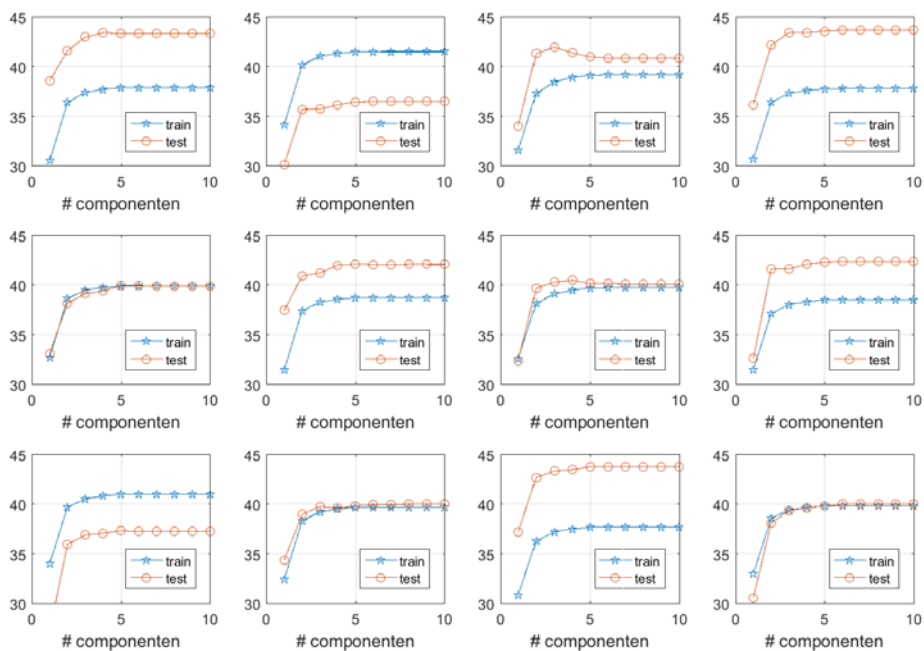
Figuur Bijlage C-10: Aantal componenten *ifv* de verklaarde variantie voor GW, N, Intrek



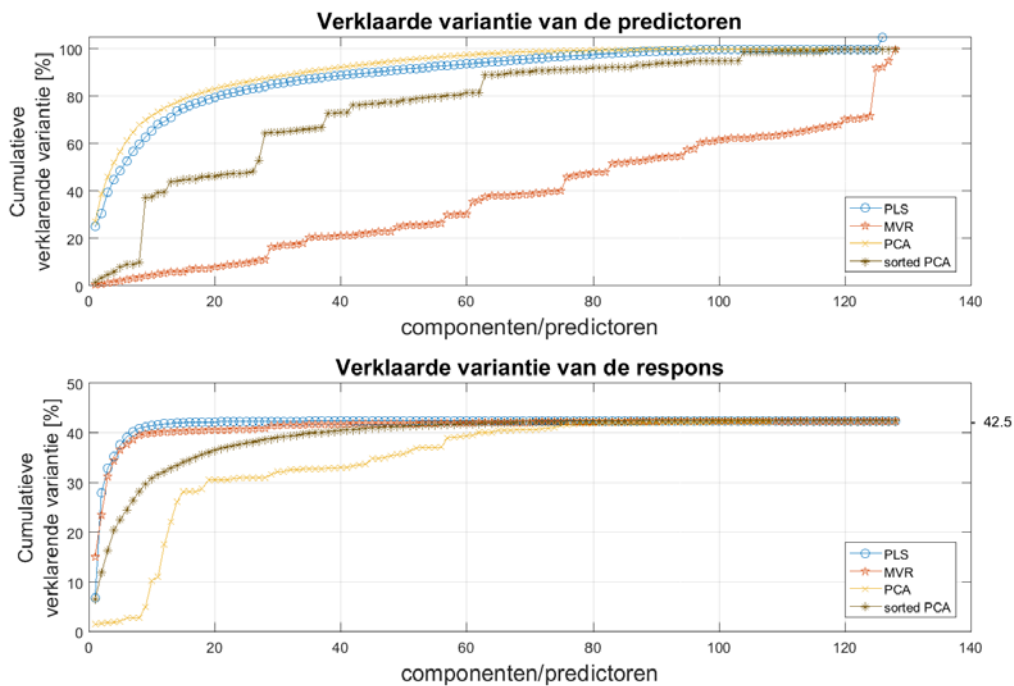




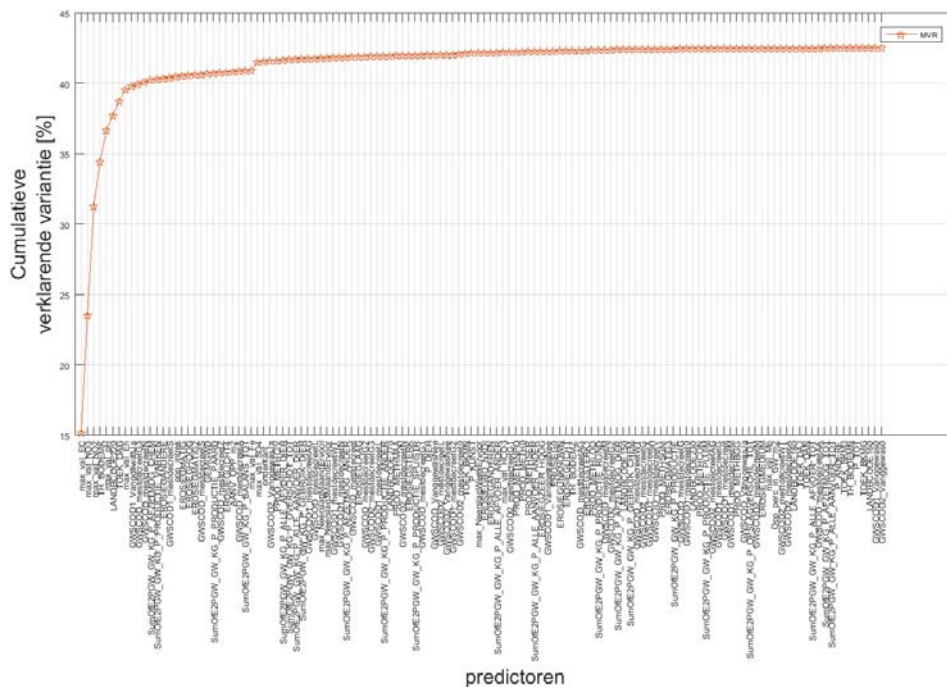
Figuur Bijlage C-13: Aantal predictoren ifv verklarende variantie voor GW, P, Intrek



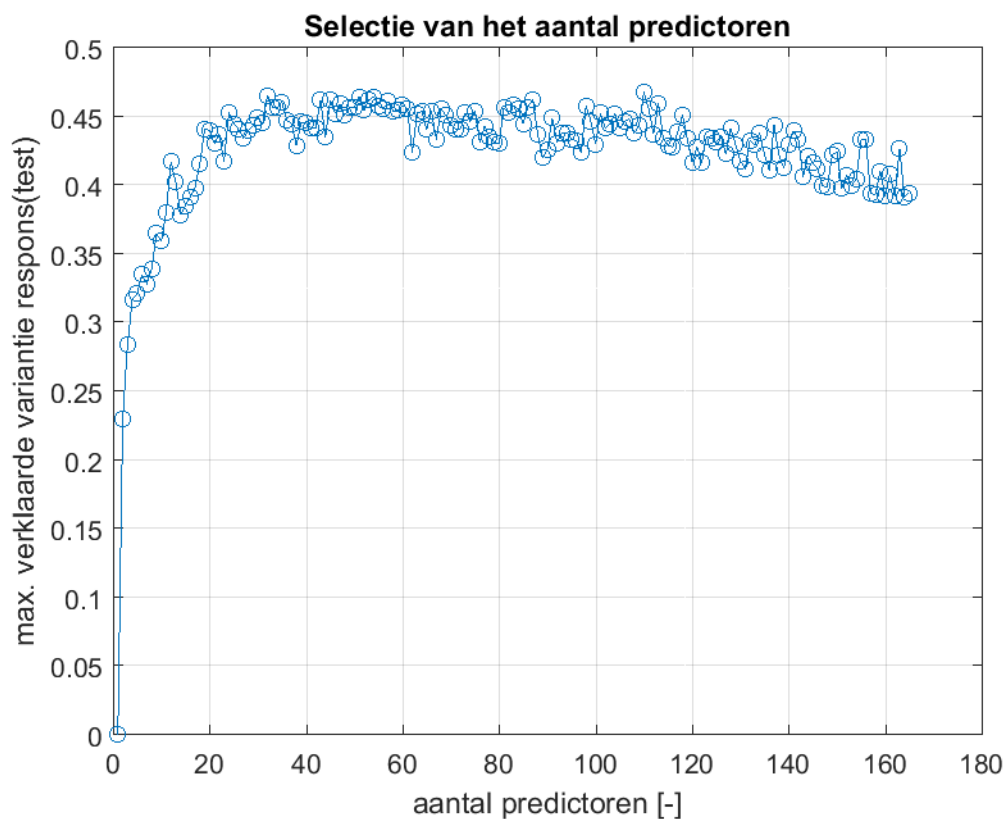
Figuur Bijlage C-14: Aantal componenten ifv de verklaarde variantie voor GW, P, Intrek



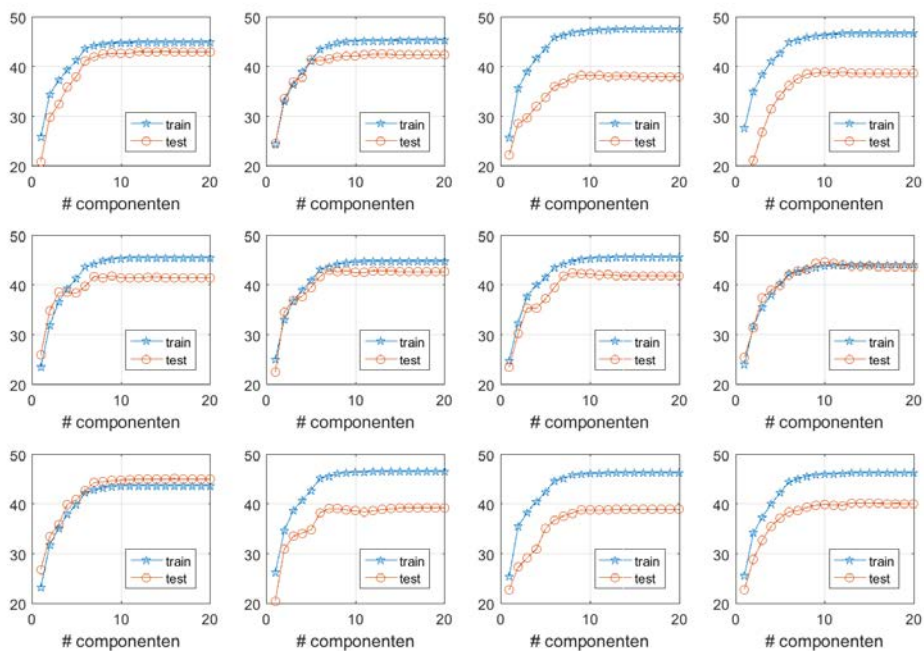
Figuur Bijlage C-15: Selectie van een optimaal model voor GW, P, Intrek



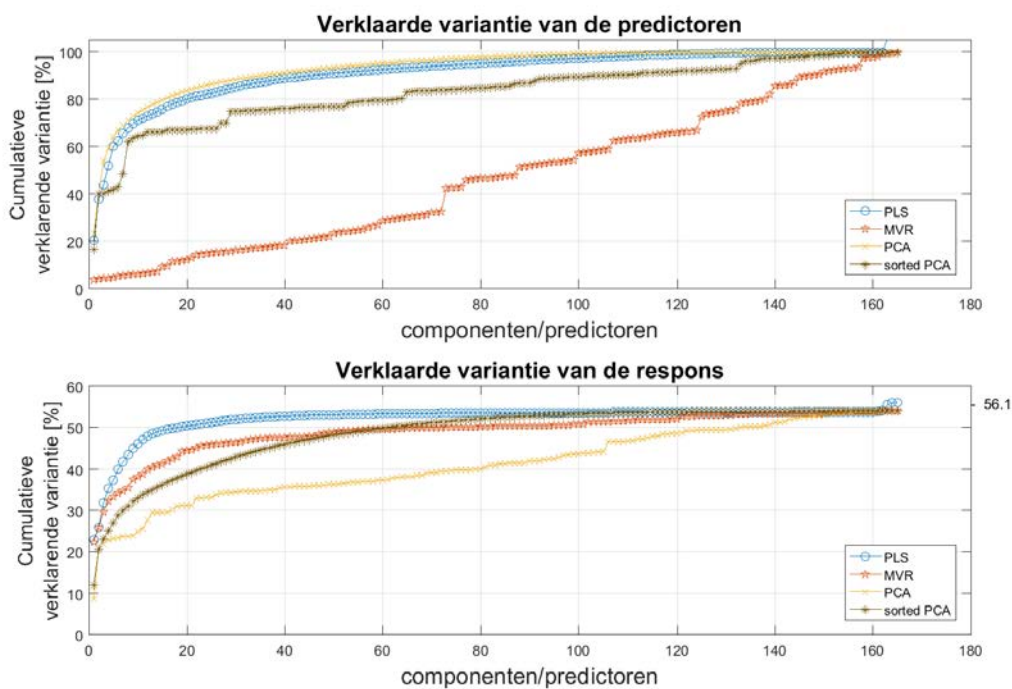
Figuur Bijlage C-16: Volgorde van de predictoren zoals bepaald door stepwise MVR voor GW, P, Intrek



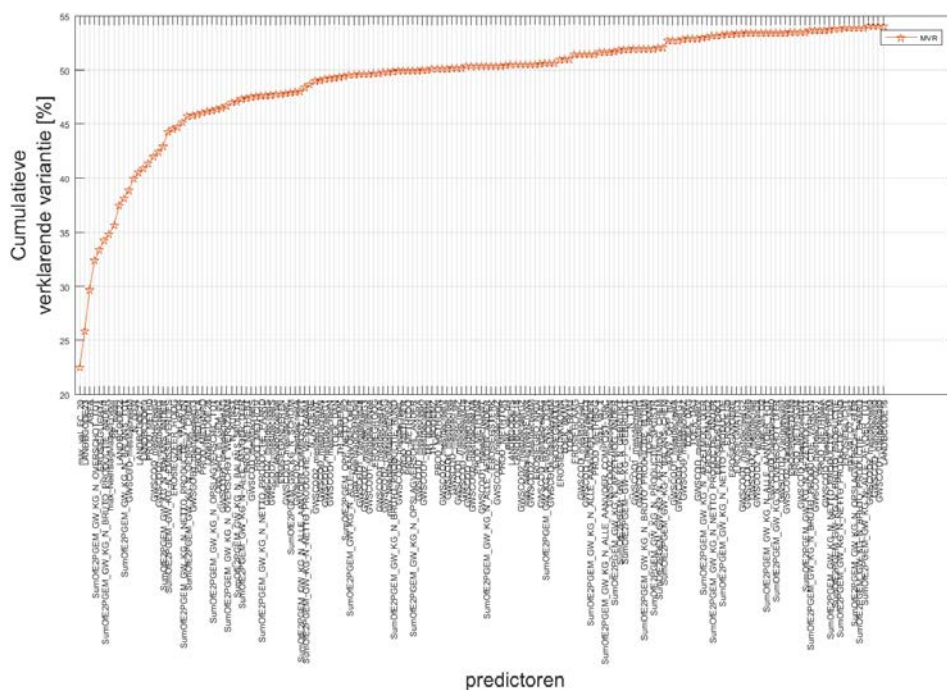
Figuur Bijlage C-17: Aantal predictoren ivf verklarende variantie voor OW, N, Gemeente



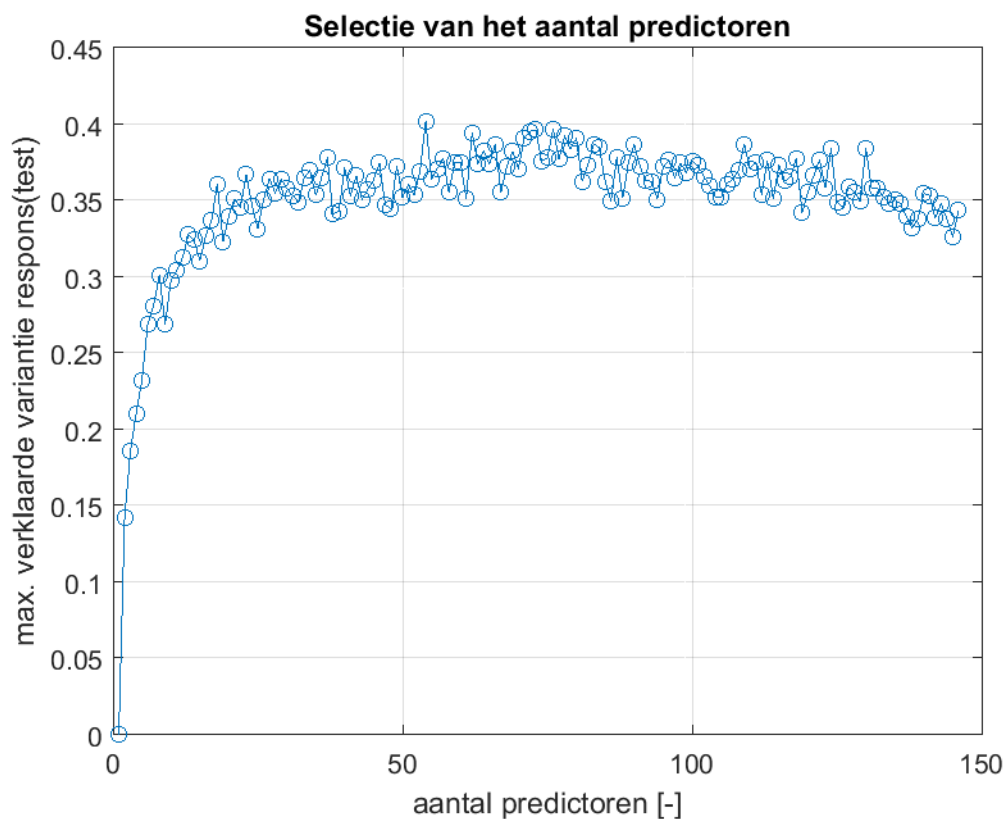
Figuur Bijlage C-18: Aantal componenten ivf de verklaarde variantie voor OW, N, Gemeente



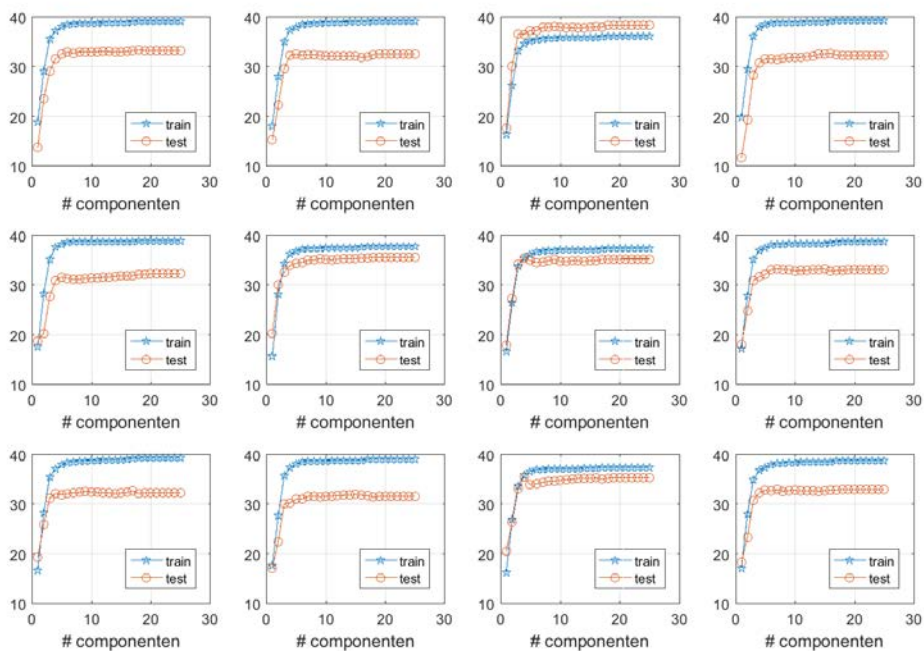
Figuur Bijlage C-19: Selectie van een optimaal model voor OW, N, Gemeente



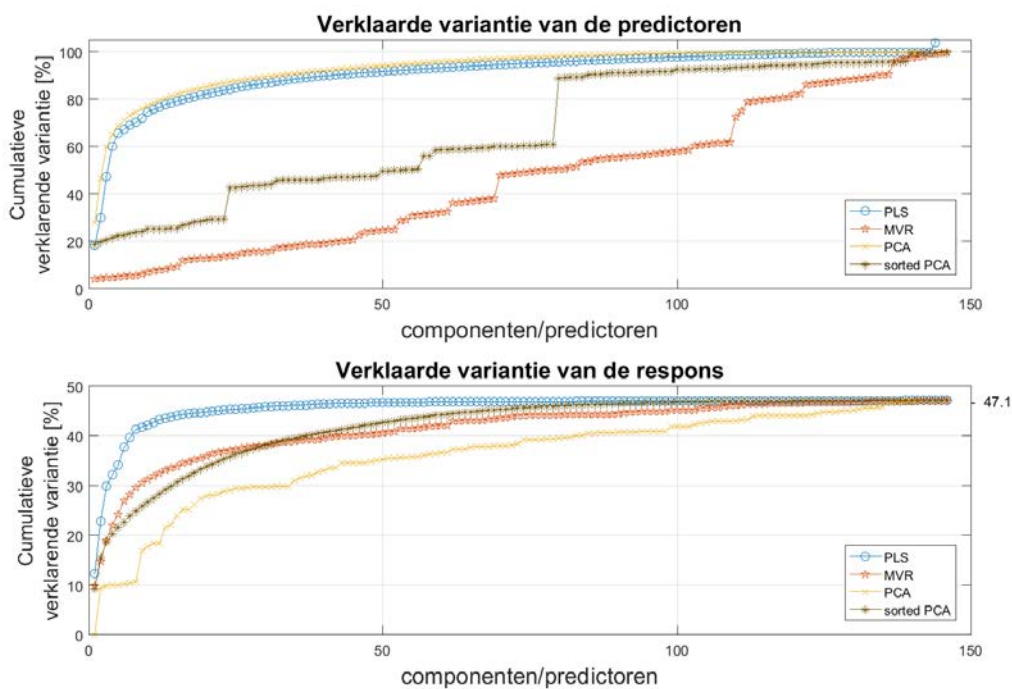
Figuur Bijlage C-20: Volgorde van de predictoren zoals bepaald door stepwise MVR voor OW, N, Gemeente



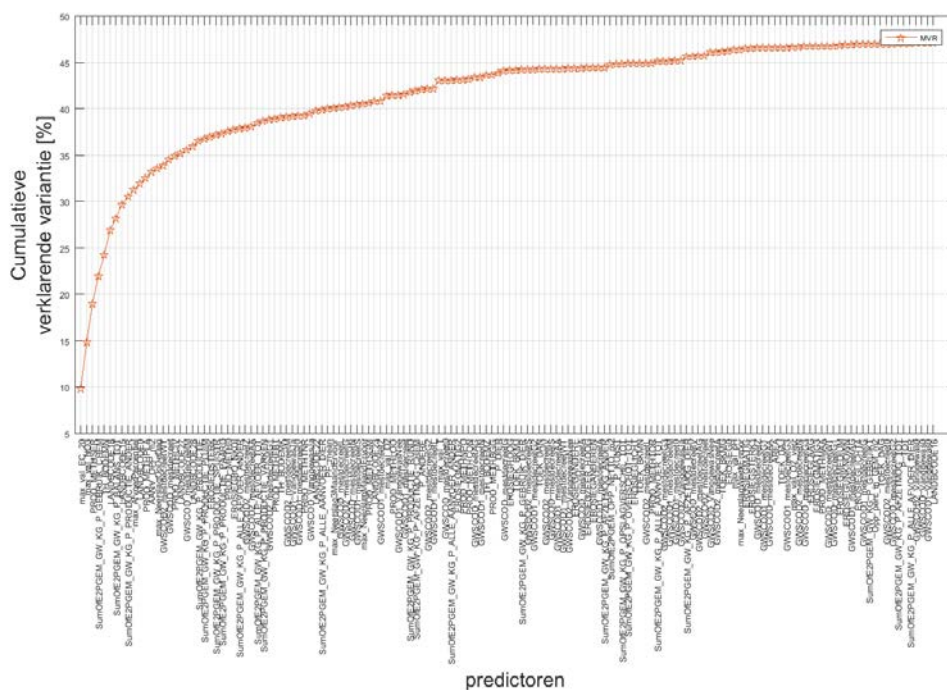
Figuur Bijlage C-21: Aantal predictoren ivf verklarende variantie voor OW, P, Gemeente



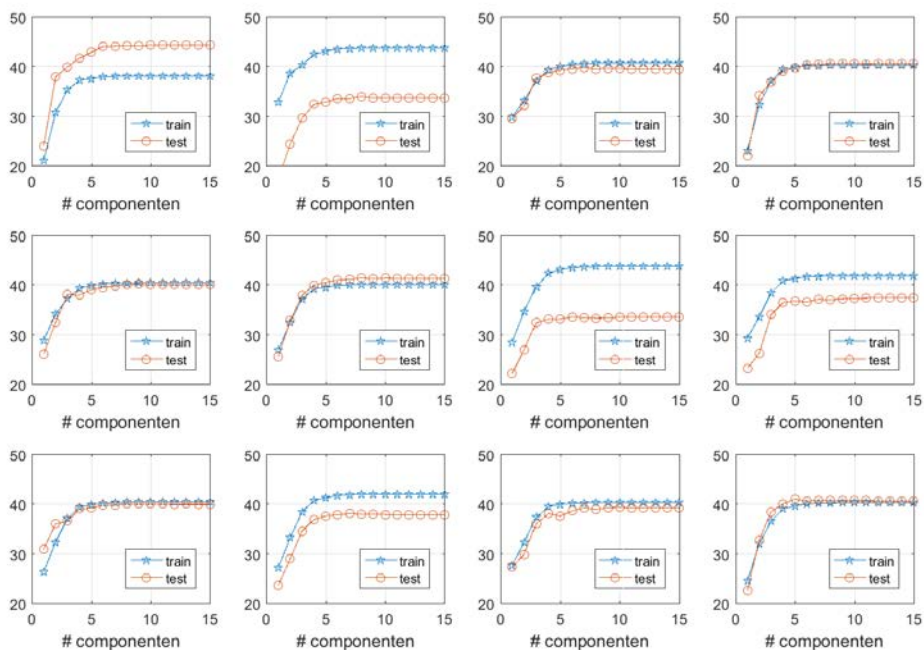
Figuur Bijlage C-22: Aantal componenten ivf de verklaarde variantie voor OW, P, Gemeente



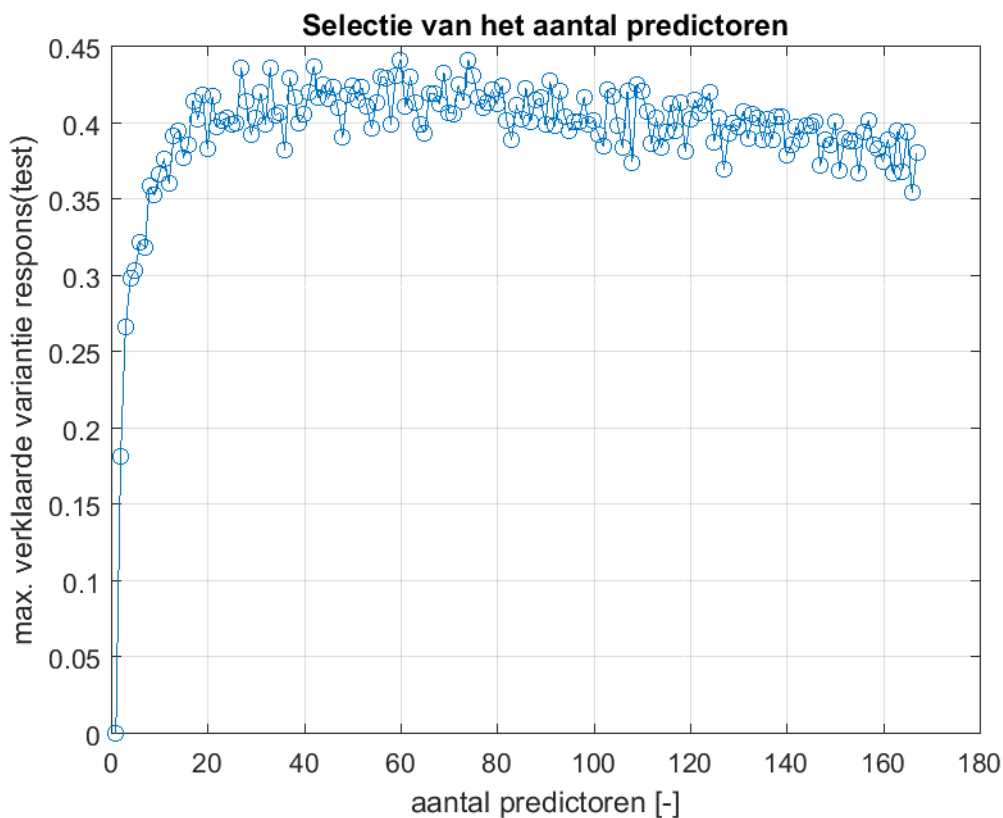
Figuur Bijlage C-23: Selectie van een optimaal model voor OW, P, Gemeente



Figuur Bijlage C-24: Volgorde van de predictoren zoals bepaald door stepwise MVR voor OW, P, Gemeente

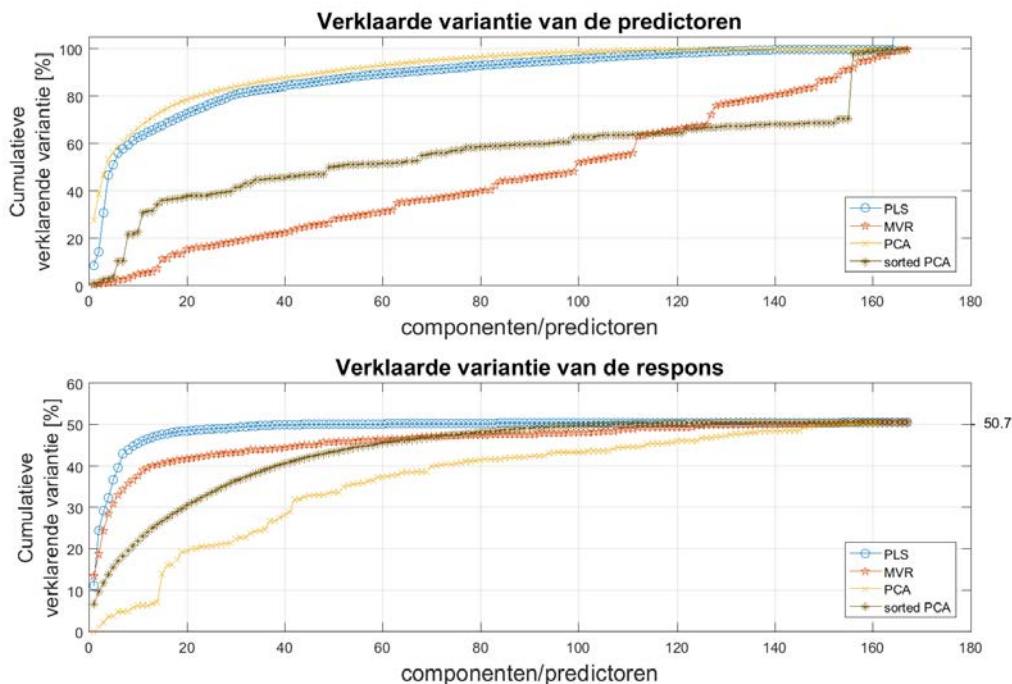


*Figuur Bijlage C-25: Aantal componenten ivf de verklaarde variantie voor GW, N, Gemeente*

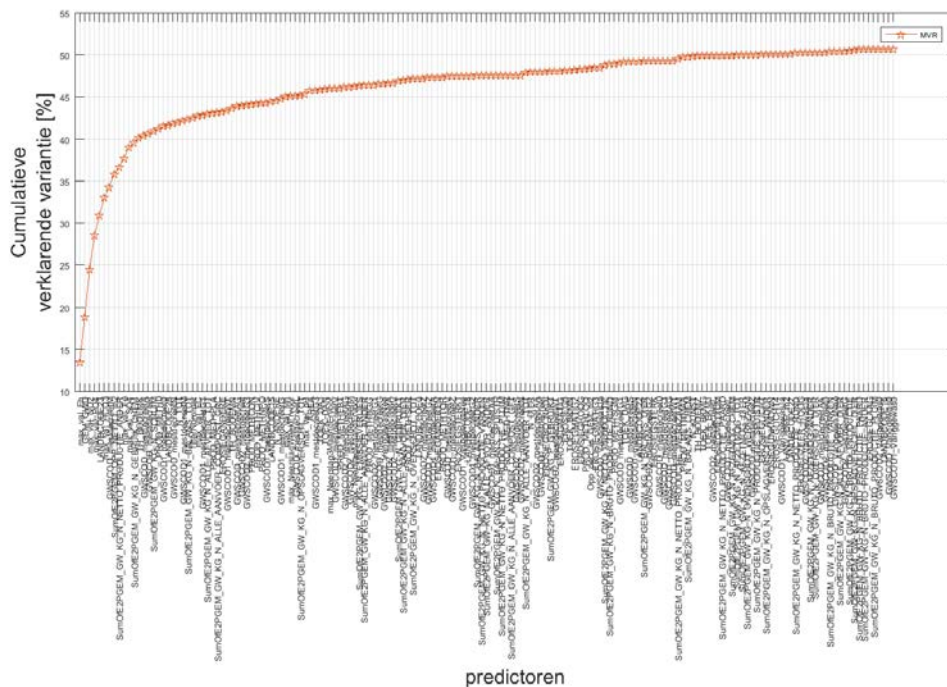


*Figuur Bijlage C-26: Aantal predictoren ivf verklarende variantie voor GW, N, Gemeente*

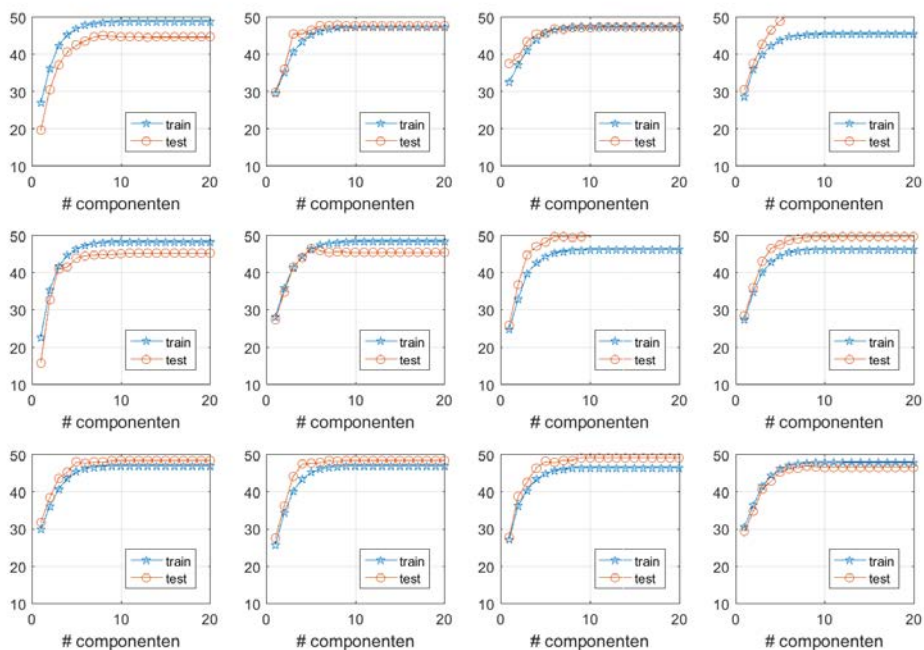




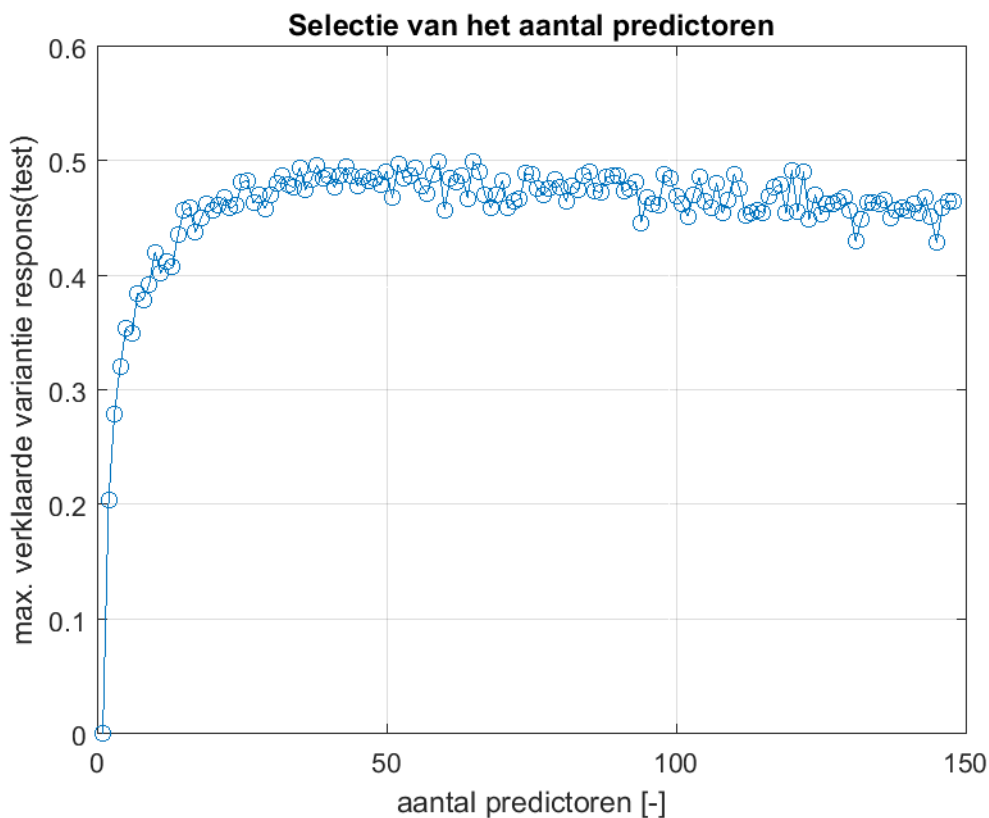
Figuur Bijlage C-27: Selectie van een optimaal model voor GW, N, Gemeente



Figuur Bijlage C-28: Volgorde van de predictoren zoals bepaald door stepwise MVR voor GW, N, Gemeente

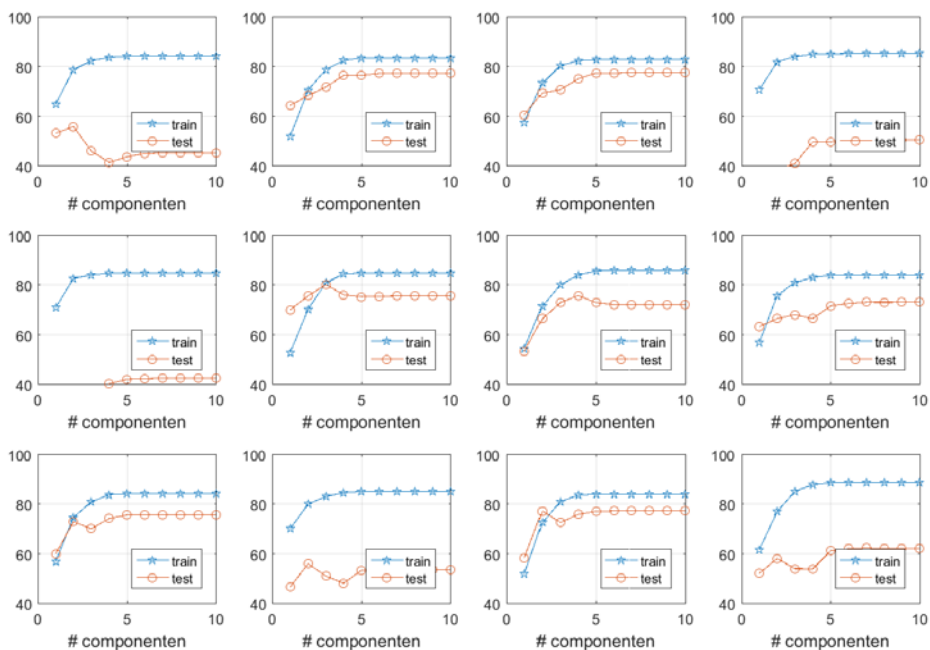


*Figuur Bijlage C-29: Aantal componenten ifv de verklaarde variantie voor GW, P, Gemeente*

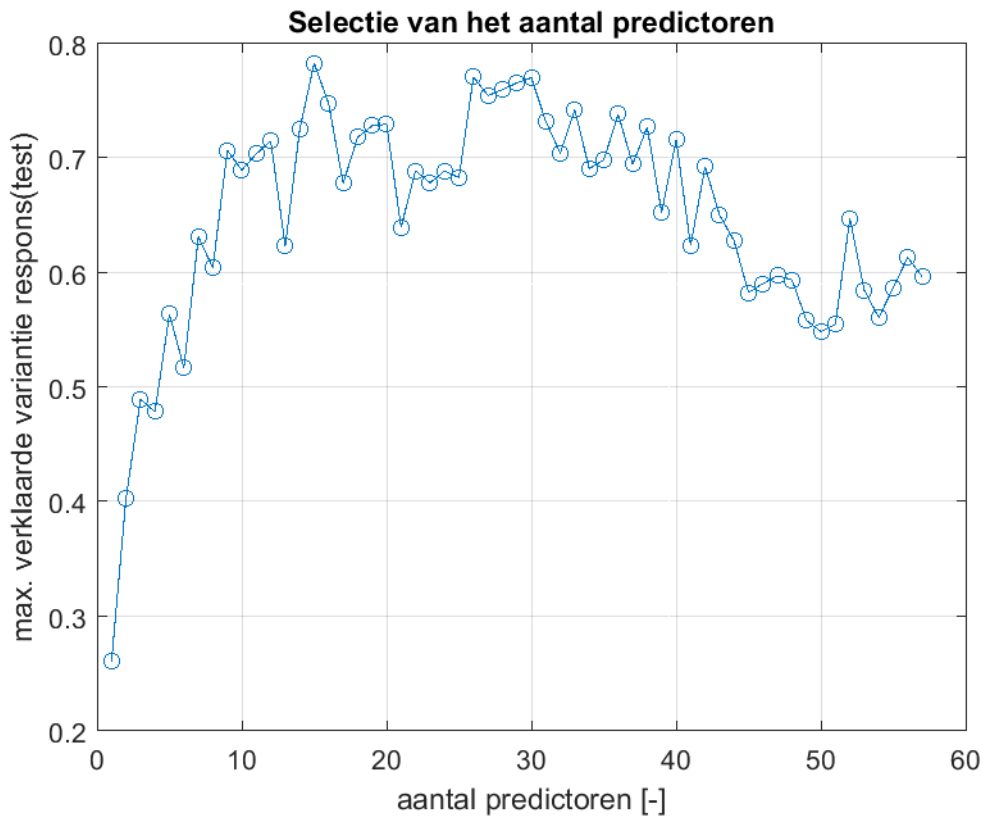


*Figuur Bijlage C-30: Aantal predictoren ifv verklarende variantie voor GW, P, Gemeente*



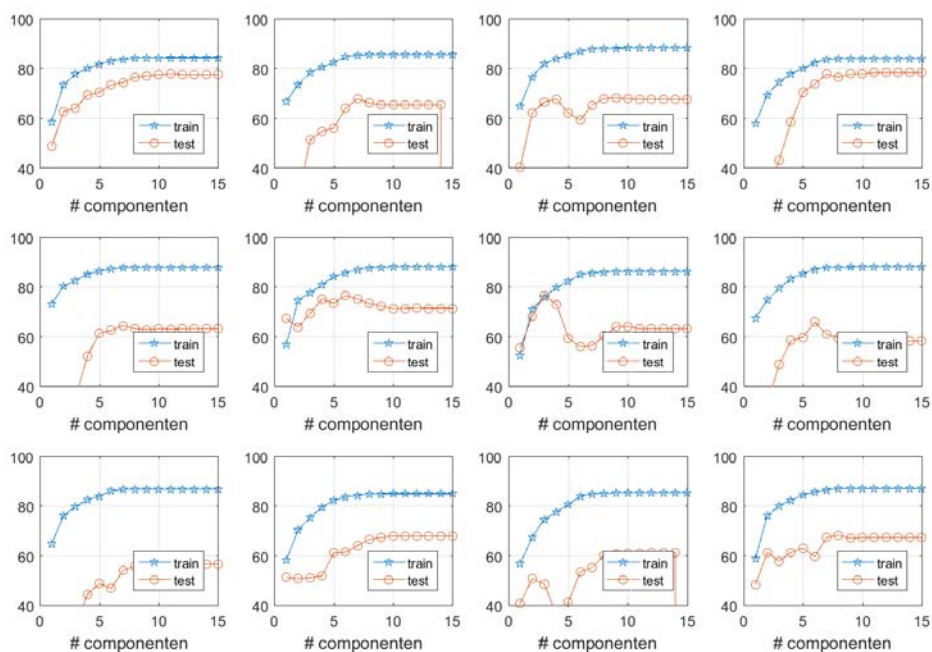


Figuur Bijlage C-33: Aantal componenten ifv de verklaarde variantie voor OW, N, Bekken

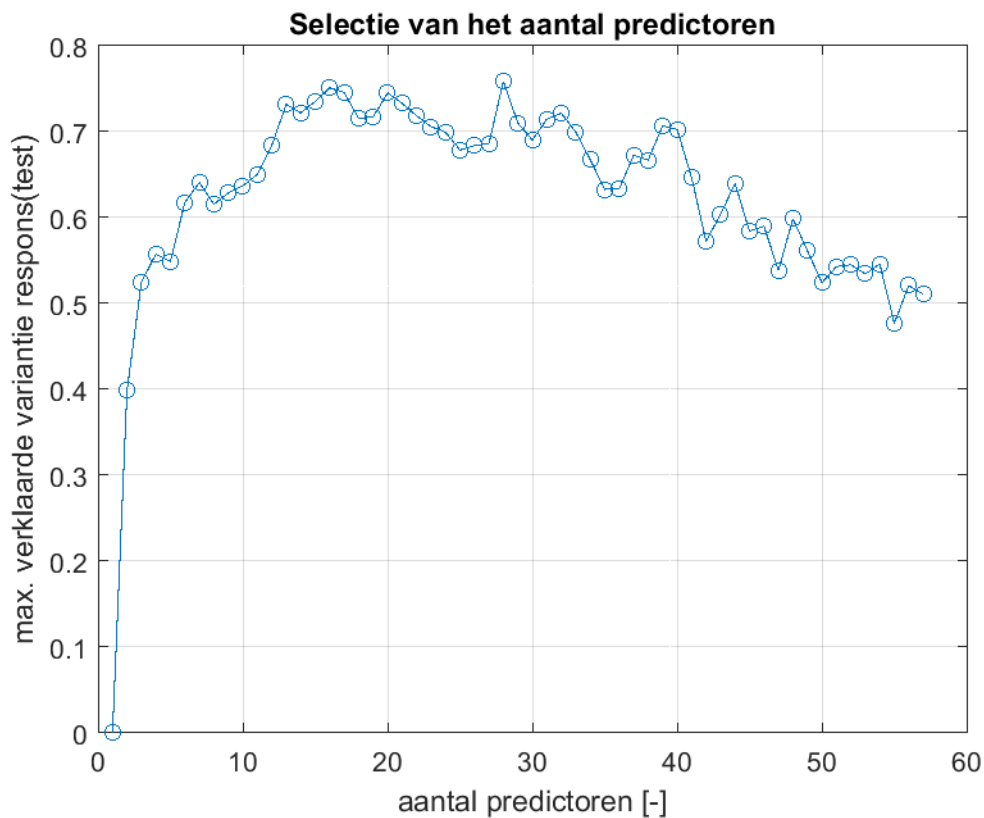


Figuur Bijlage C-34: Aantal predictoren ifv verklarende variantie voor OW, N, Bekken



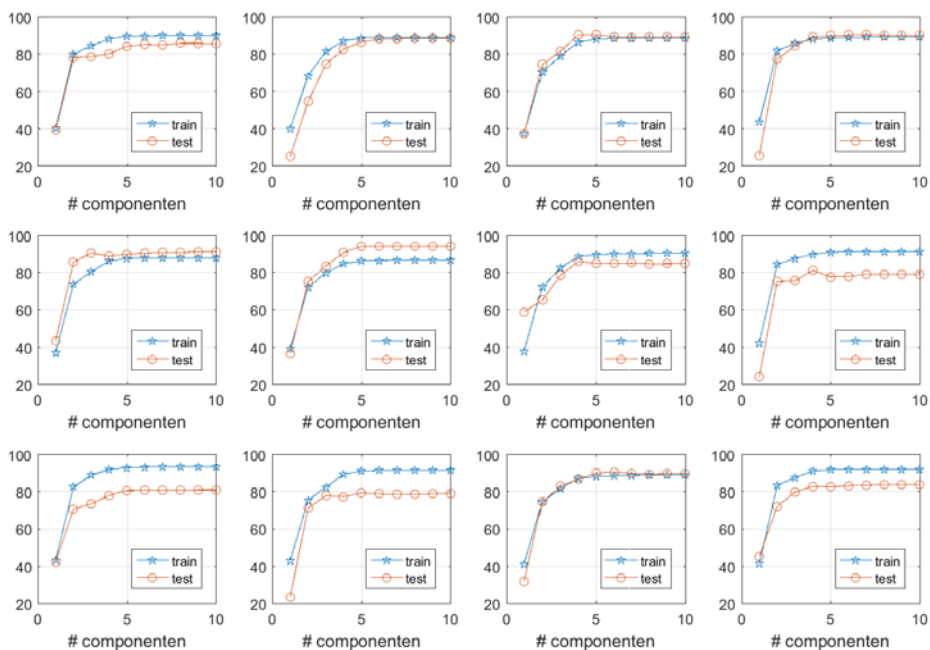


Figuur Bijlage C-37: Aantal componenten ivf de verklaarde variantie voor OW, P, Bekken

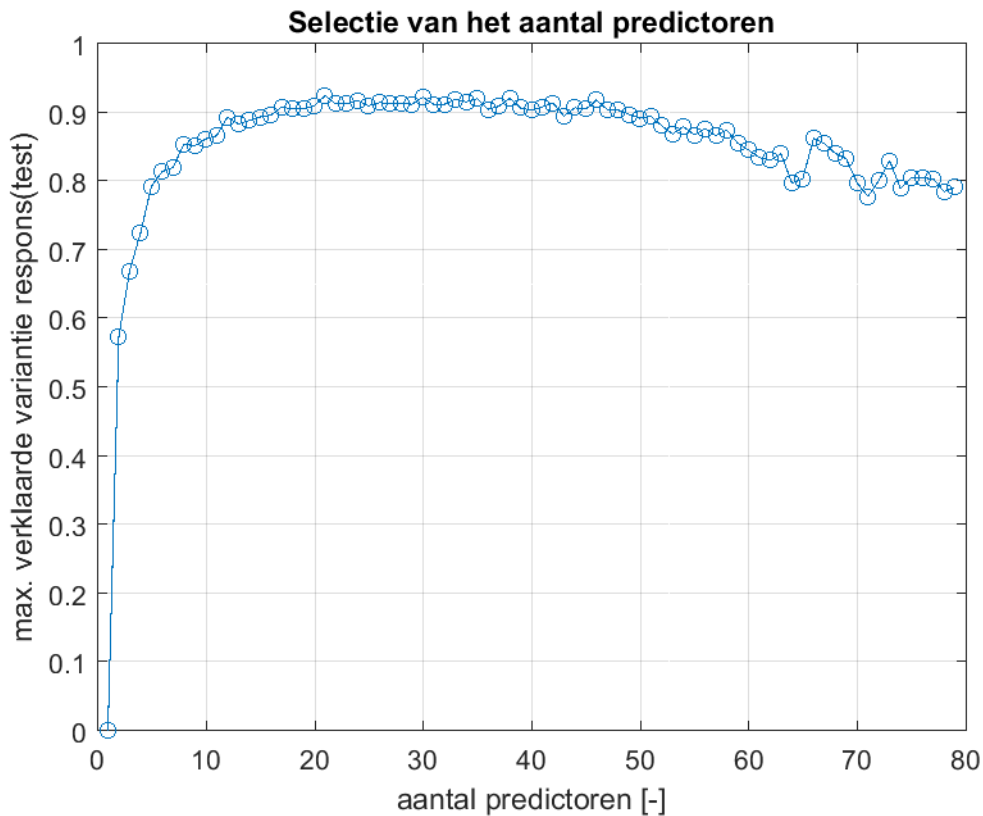


Figuur Bijlage C-38: Aantal predictoren ivf verklarende variantie voor OW, P, Bekken



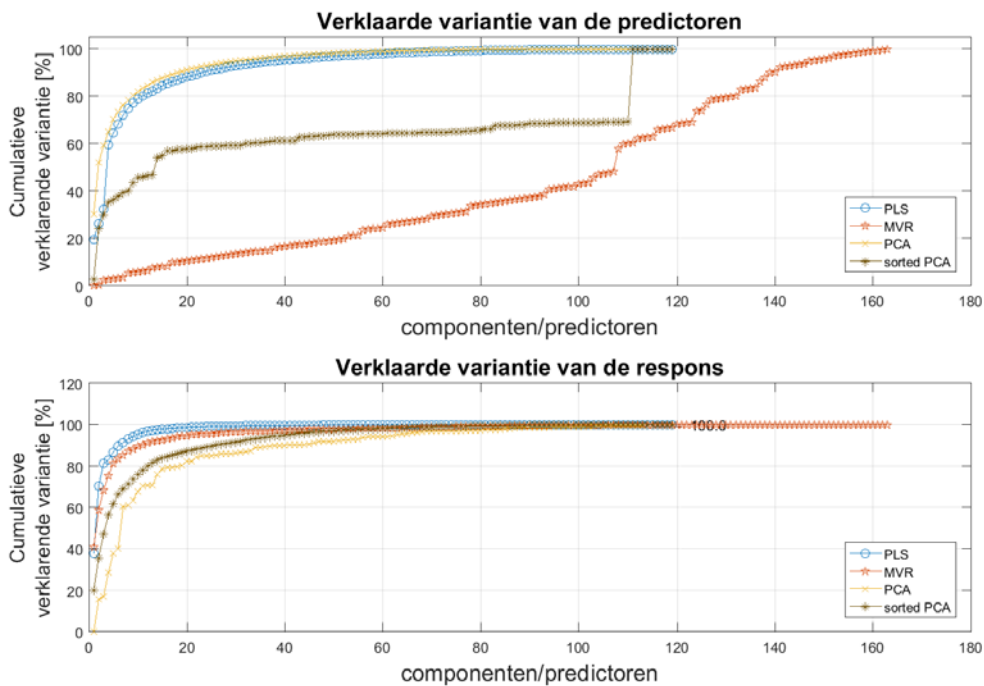


Figuur Bijlage C-41: Aantal componenten ifv de verklaarde variantie voor OW, N, Klasse

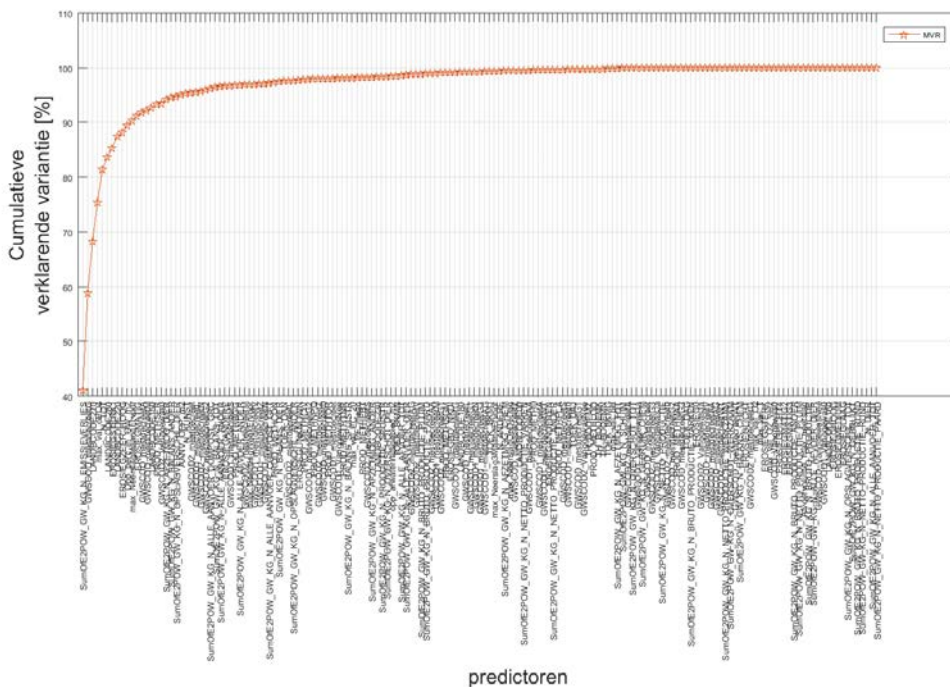


Figuur Bijlage C-42: Aantal predictoren ifv verklarende variantie voor OW, N, Klasse

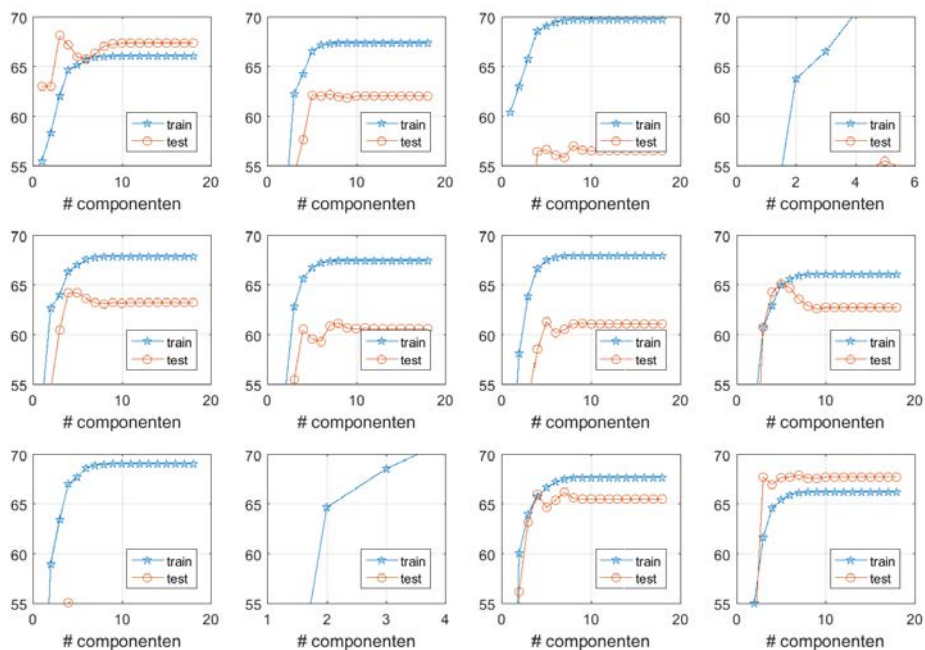




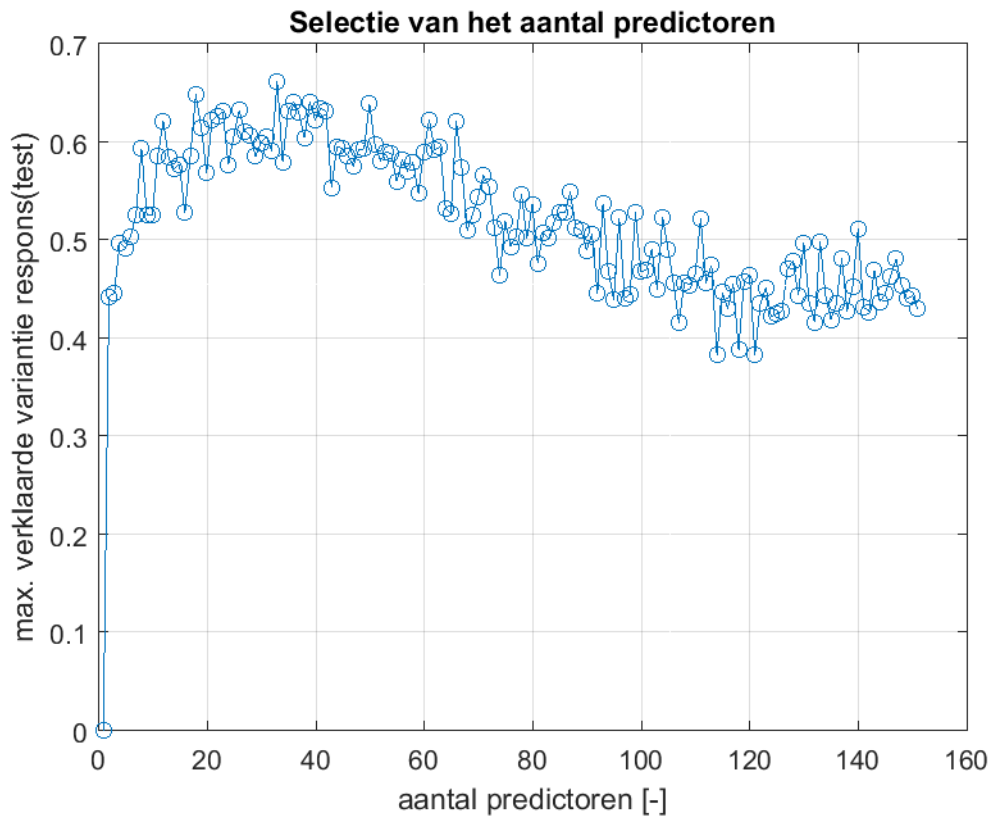
Figuur Bijlage C-43: Selectie van een optimaal model voor OW, N, Klasse



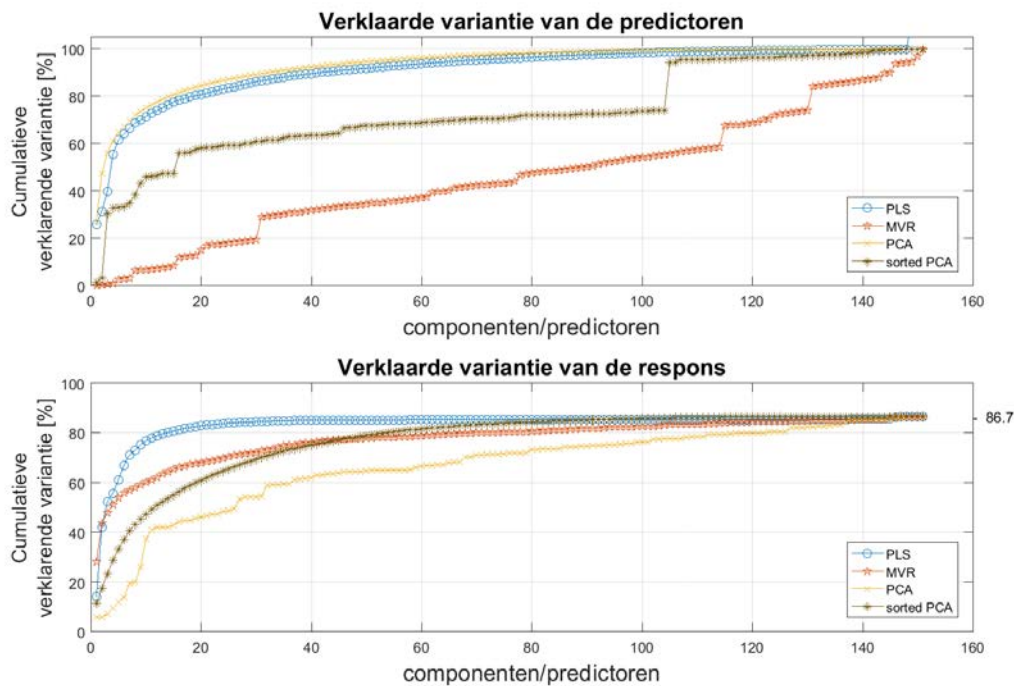
Figuur Bijlage C-44: Volgorde van de predictoren zoals bepaald door stepwise MVR voor OW, N, Klasse



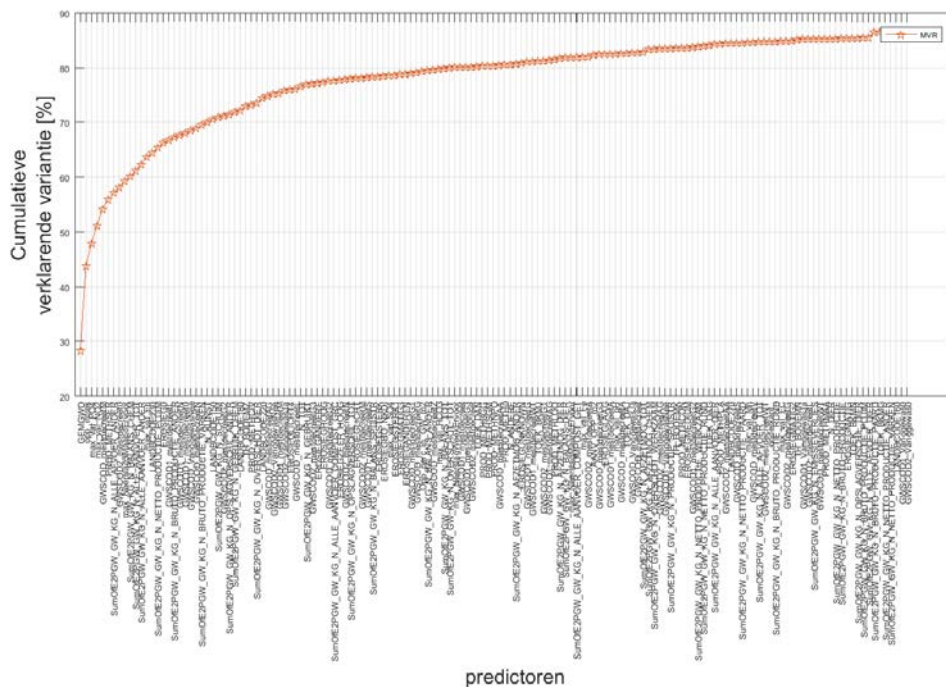
Figuur Bijlage C-45: Aantal componenten ifv de verklaarde variantie voor GW, N, HHZ



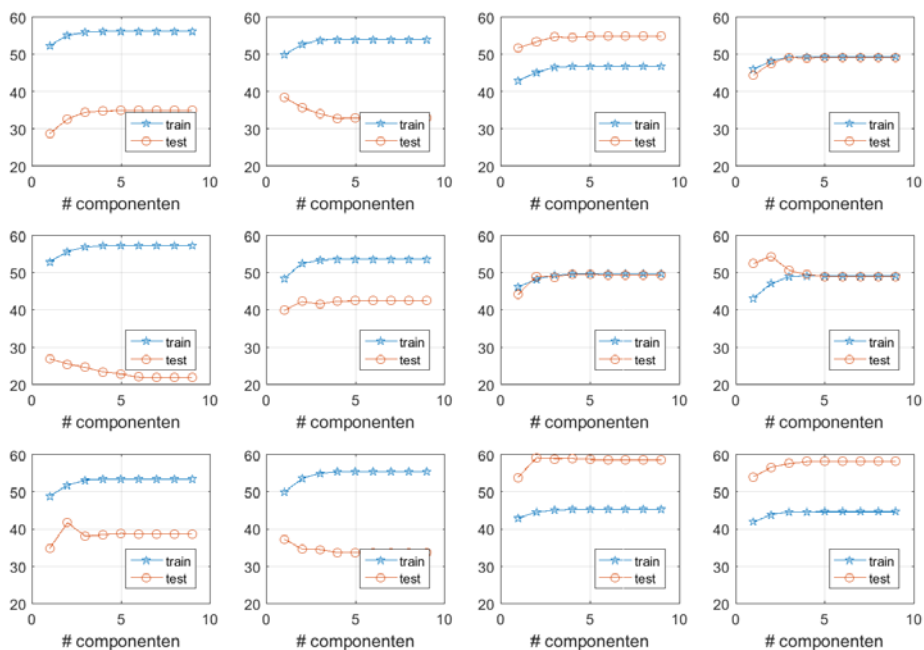
Figuur Bijlage C-46: Aantal predictoren ifv verklarende variantie voor GW, N, HHZ



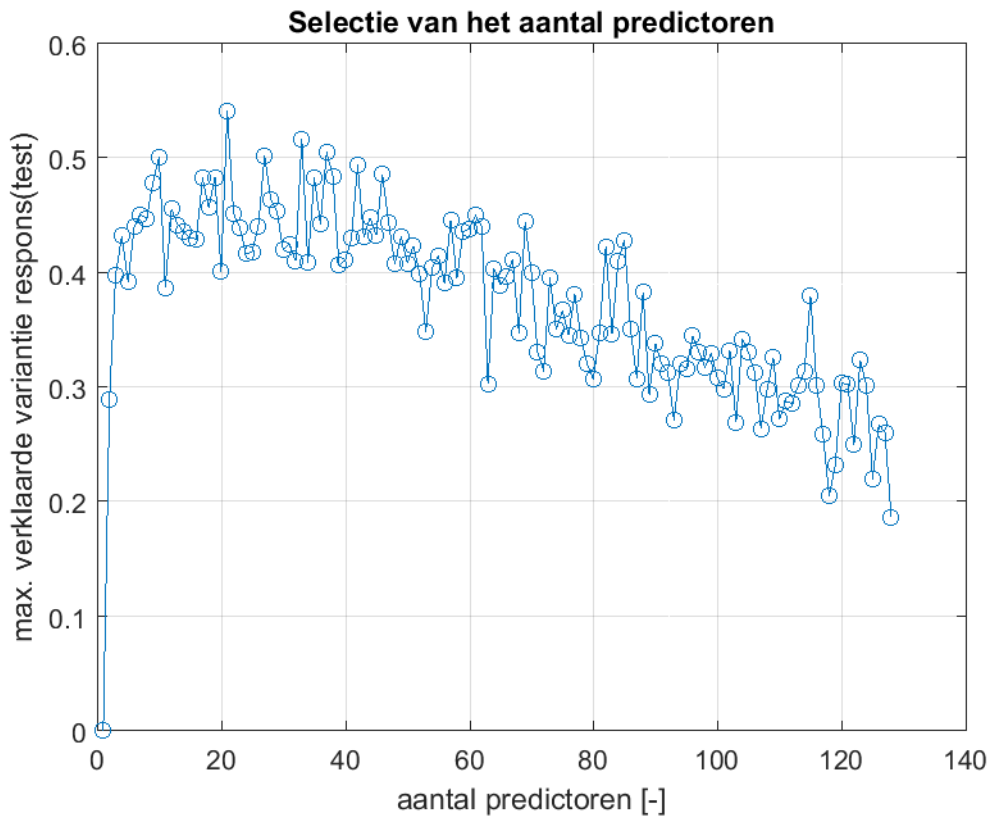
Figuur Bijlage C-47: Selectie van een optimaal model voor GW, N, HHZ



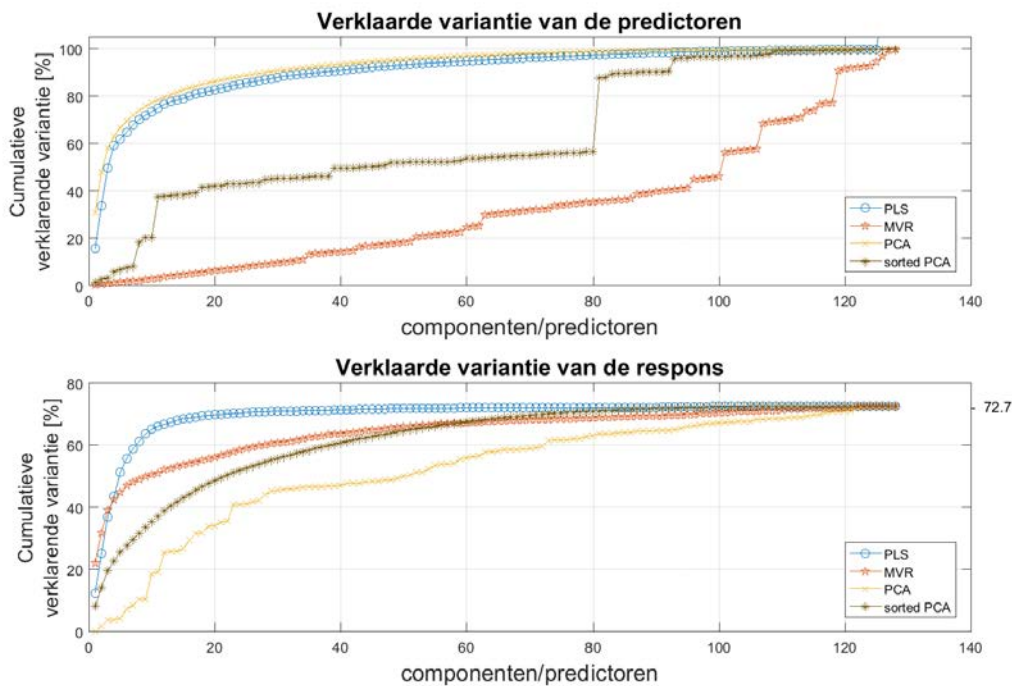
Figuur Bijlage C-48: Volgorde van de predictoren zoals bepaald door stepwise MVR voor GW, N, HHZ



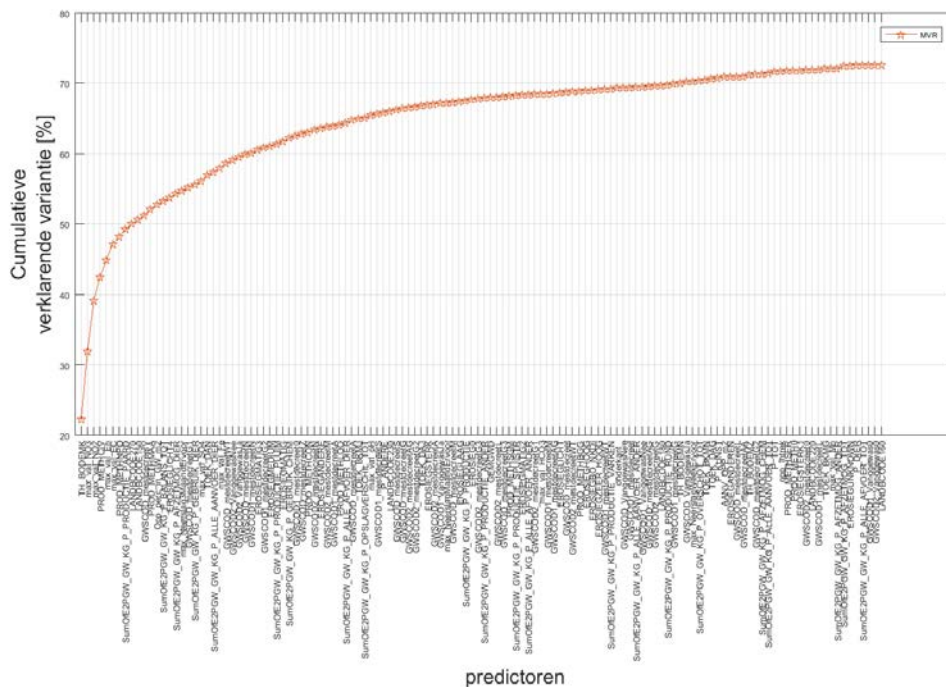
Figuur Bijlage C-49: Aantal componenten ifv de verklaarde variantie voor GW, P, HHZ



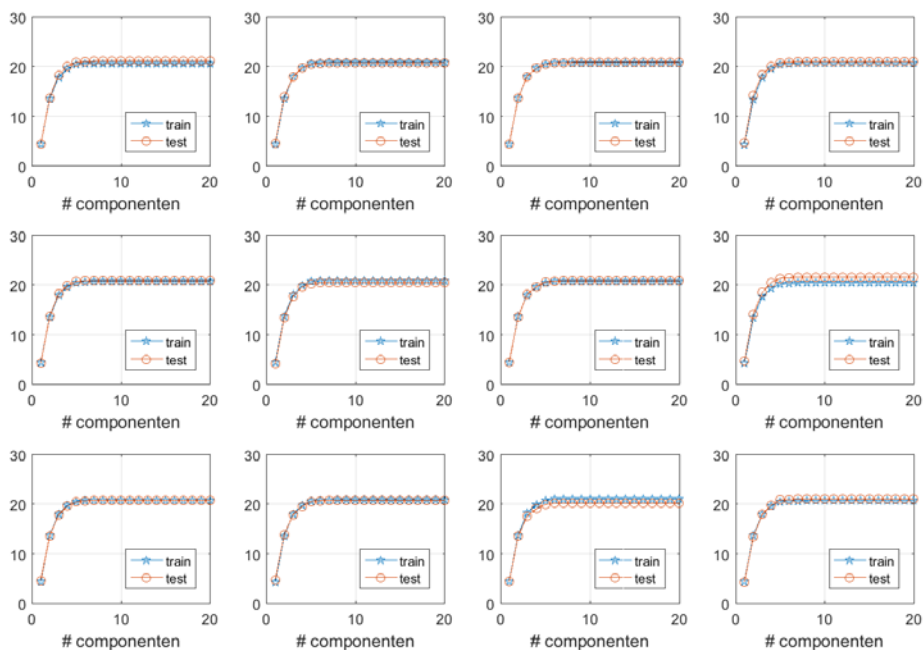
Figuur Bijlage C-50: Aantal predictoren ifv verklarende variantie voor GW, P, HHZ



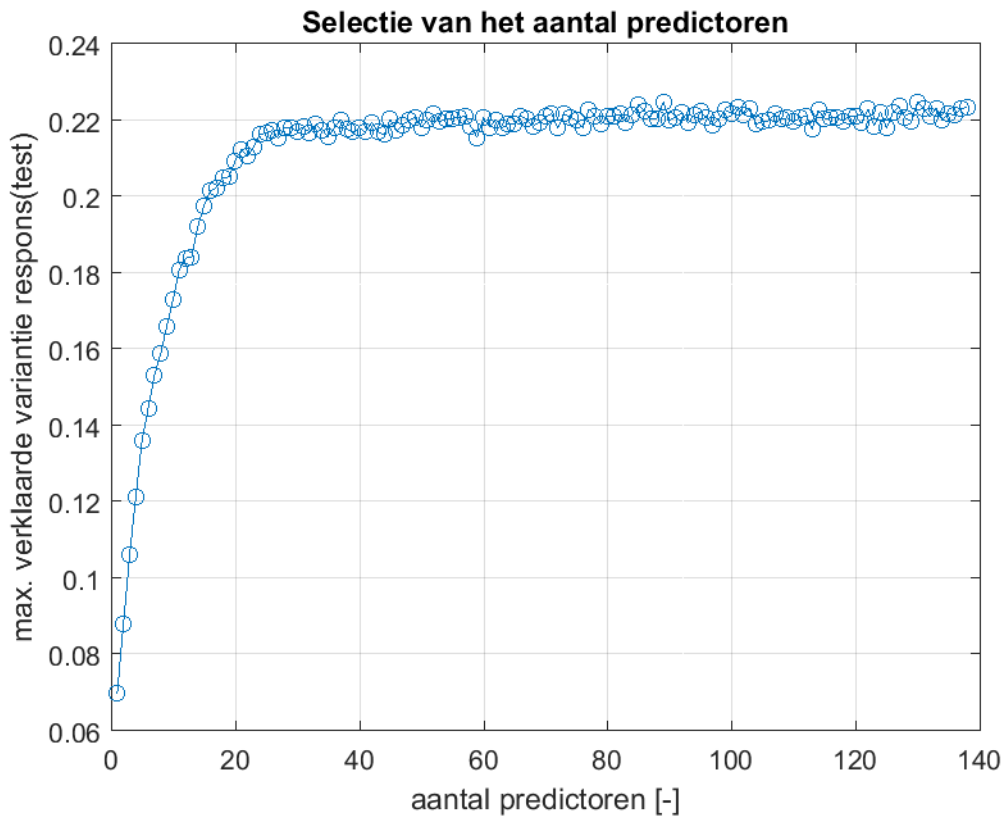
Figuur Bijlage C-51: Selectie van een optimaal model voor GW, P, HHZ



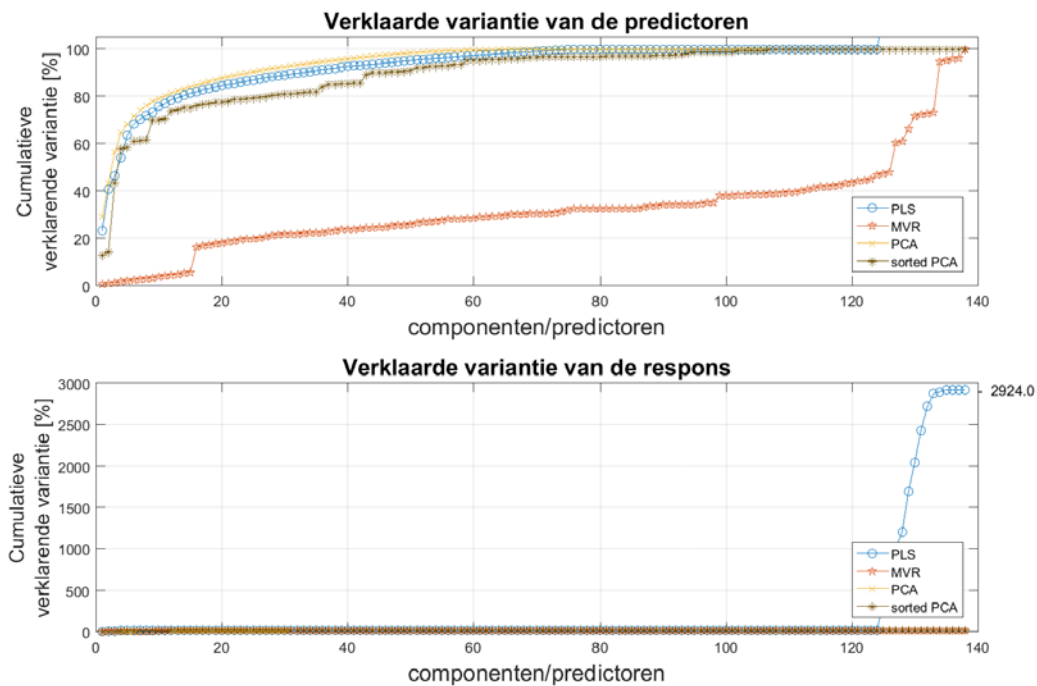
Figuur Bijlage C-52: Volgorde van de predictoren zoals bepaald door stepwise MVR voor GW, P, HHZ



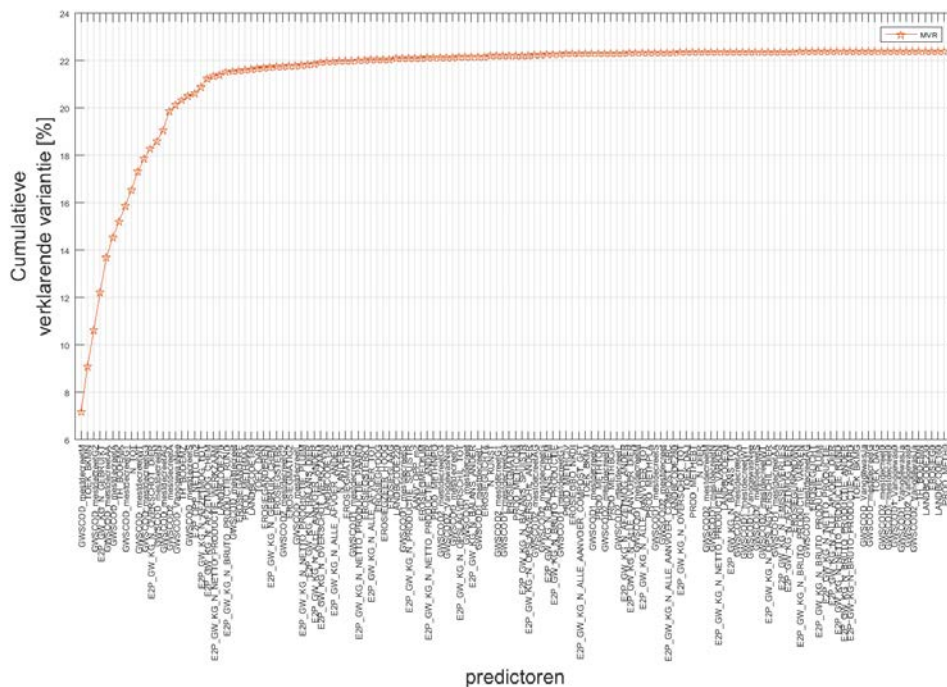
Figuur Bijlage C-53: Aantal componenten ivf de verklaarde variantie voor NRES, Perceel



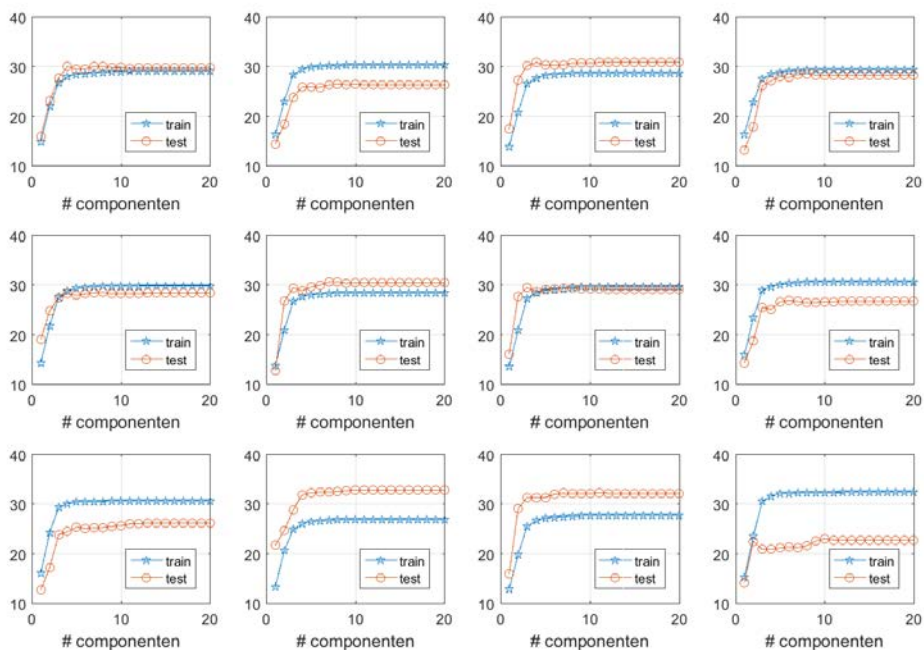
Figuur Bijlage C-54: Aantal predictoren ivf verklarende variantie voor NRES, Perceel



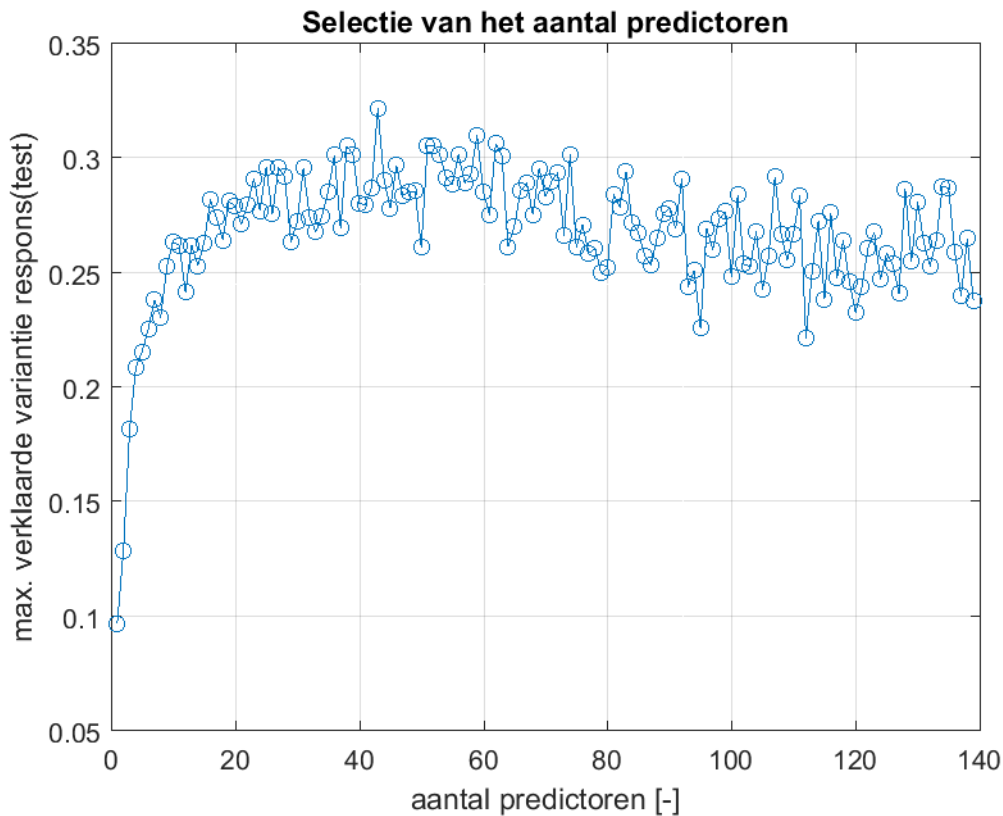
Figuur Bijlage C-55: Selectie van een optimaal model voor NRES, Perceel



Figuur Bijlage C-56: Volgorde van de predictoren zoals bepaald door stepwise MVR voor NRES, Perceel

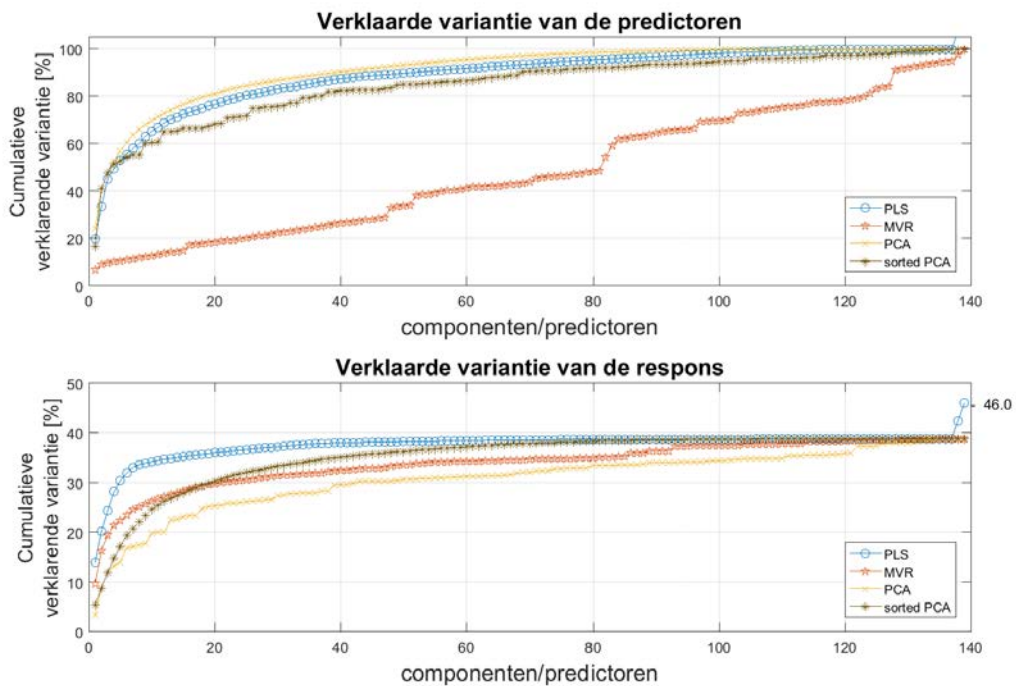


Figuur Bijlage C-57: Aantal componenten ifv de verklaarde variantie voor NRES, Gemeente

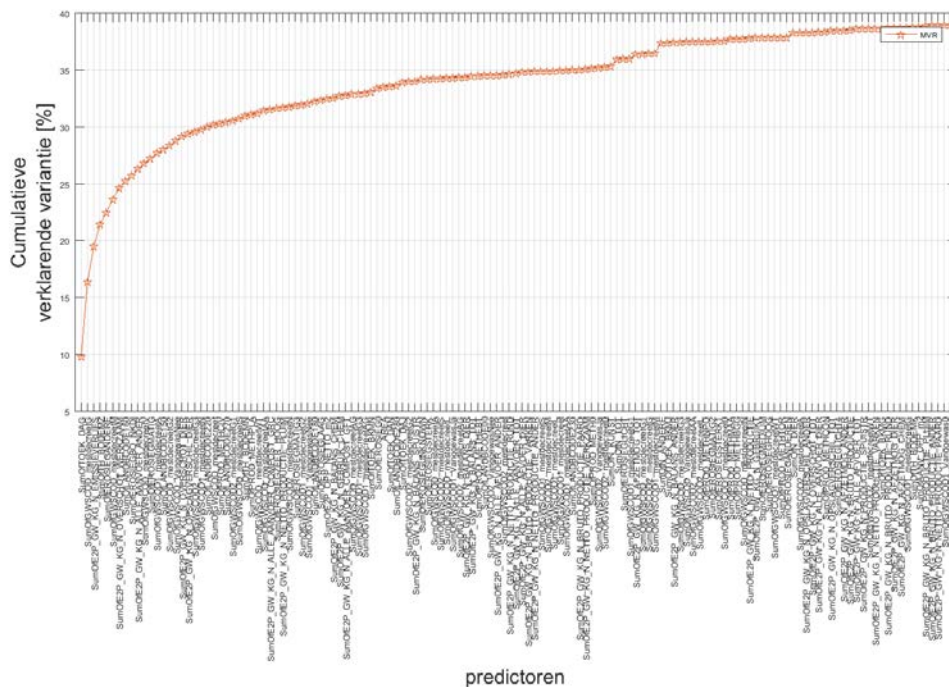


Figuur Bijlage C-58: Aantal predictoren ifv verklarende variantie voor NRES, Gemeente





Figuur Bijlage C-59: Selectie van een optimaal model voor NRES, Gemeente



Figuur Bijlage C-60: Volgorde van de predictoren zoals bepaald door stepwise MVR voor NRES, Gemeente