Actualisatie van de drainageklasse van de bodemkaart door koppeling van grondwaterstatistieken aan de bodemkaart

Eindrapport

DEPARTEMENT
OMGEVING

omgeving.vlaanderen.be

ACTUALISATIE VAN DE DRAINAGEKLASSE VAN DE BODEMKAART DOOR KOPPELING VAN GRONDWATERSTATISTIEKEN AAN DE BODEMKAART

Het doel van deze opdracht is om aan de hand van recente grondwaterpeilmetingen en relevante omgevingsdata, de drainageklassen van de polygonen van de bodemkaart te actualiseren en te koppelen aan statistieken van grondwaterstanden (GxG) die een maat zijn voor de meerjarige klimatologische condities van het grondwaterpeil. De ontwikkelde methodiek laat toe dat, telkens wanneer er voldoende nieuwe data beschikbaar zijn, de drainageklasse van de bodemkaart kan geactualiseerd worden. Voorliggend rapport beschrijft eerst de eerdere studies waarop deze opdracht steunt, alvorens dieper in te gaan op de methodiek en de bevindingen.

Dit rapport bevat de mening van externe auteur(s) en niet noodzakelijk die van de Vlaamse overheid.

COLOFON

Verantwoordelijke uitgever
Ivo Palmers
Departement Omgeving
Koning Albert II-laan 15 bus 553, 1210 Brussel
www.omgevingvlaanderen.be

Een uitgave van het Departement Omgeving, Afdeling Vlaams Planbureau voor Omgeving
vpo.omgeving@vlaanderen.be

Auteurs
Martinus Van den Berg, Ferdinand Messens, Bos Debusscher, Ivo Van de Moortel en Symine Henkens – Antea Group

In samenwerking met
Prof. Peter Finke - UGent

Publicatiedatum
Mei 2024

Depotnummer
D/2024/3241/179

Wijze van citeren
Van den Berg, M., Messens, F., Debusscher, B., Henkens, S. en Van de Moortel, I. (2023). Actualisatie van de drainageklasse van de bodemkaart door koppeling van grondwaterstatistieken aan de bodemkaart. Studie in opdracht van het departement Omgeving, afdeling Vlaams Planbureau voor Omgeving. 111 pp.

PARTNERS

MANAGEMENTSAMENVATTING

De drainageklasse van de bodemkaart is nog steeds één van de meest gebruikte parameters in zeer diverse modelstudies. Deze drainageklassen zijn echter niet meer up-to-date vermits ze in het veld werden opgenomen tijdens de bodemkartering (1947-1974) en de hydrologie sindsdien plaatselijk sterk gewijzigd kan zijn. Bovendien baseert de beschrijving zich op het voorkomen van hydromorfe profielkenmerken die vaak na-ijlen op de hydrologische situatie. Op basis van de drainageklassen van de bodemkaart kan men dus mogelijk verkeerde modelresultaten bekomen en verkeerde conclusies trekken. Het Vlaams Planbureau voor Omgeving (Departement Omgeving - VPO) wenst hieraan tegemoet te komen door expertise met betrekking tot grondwater en bodem met elkaar te combineren en nieuwe drainageklassen voor de bodemkaart af te leiden op basis van actuele grondwaterstanden.

In voorliggende opdracht werd een nieuwe methodiek ontwikkeld om aan de hand van recente grondwaterpeilmetingen en relevante omgevingsdata, de drainageklassen van de polygonen van de bodemkaart te actualiseren en te koppelen aan statistieken van grondwaterstanden, de gemiddeld hoogste grondwaterstand (GHG) en gemiddeld laagste grondwaterstand (GLG), tezamen de "GxG" genoemd. De ontwikkelde methodiek laat bovendien toe om, telkens wanneer er voldoende nieuwe data beschikbaar zijn, de drainageklassen vlot te actualiseren; al dan niet voor heel Vlaanderen of per kaartblad.

Meerdere technieken werden onderzocht, maar uiteindelijk bleek een regressieanalyse op de grondwaterstatistieken de meest geschikte methode. Een data-assimilatie bleek onvoldoende robuust, terwijl het rechtstreeks simuleren a.d.h.v. remote sensed bodemvocht, alsook een regressieanalyse op modelparameters, onvoldoende accurate resultaten opleverde. Van de onderzochte regressietechnieken bleek de boosting-techniek *Extreme Gradient Boosting* het beste resultaat op te leveren. Hiervoor diende de pool aan trainingsdata weliswaar uitgebreid te worden met grondwaterstatistieken berekend op zowel incidentele metingen als tijdreeksmodellering met PIRFICT: in Vlaanderen bevinden er zich namelijk onvoldoende locaties waarop men de GxG's kan berekenen volgens de strikte definitie van Heesen (1970).

Het resultaat is drievoudig:

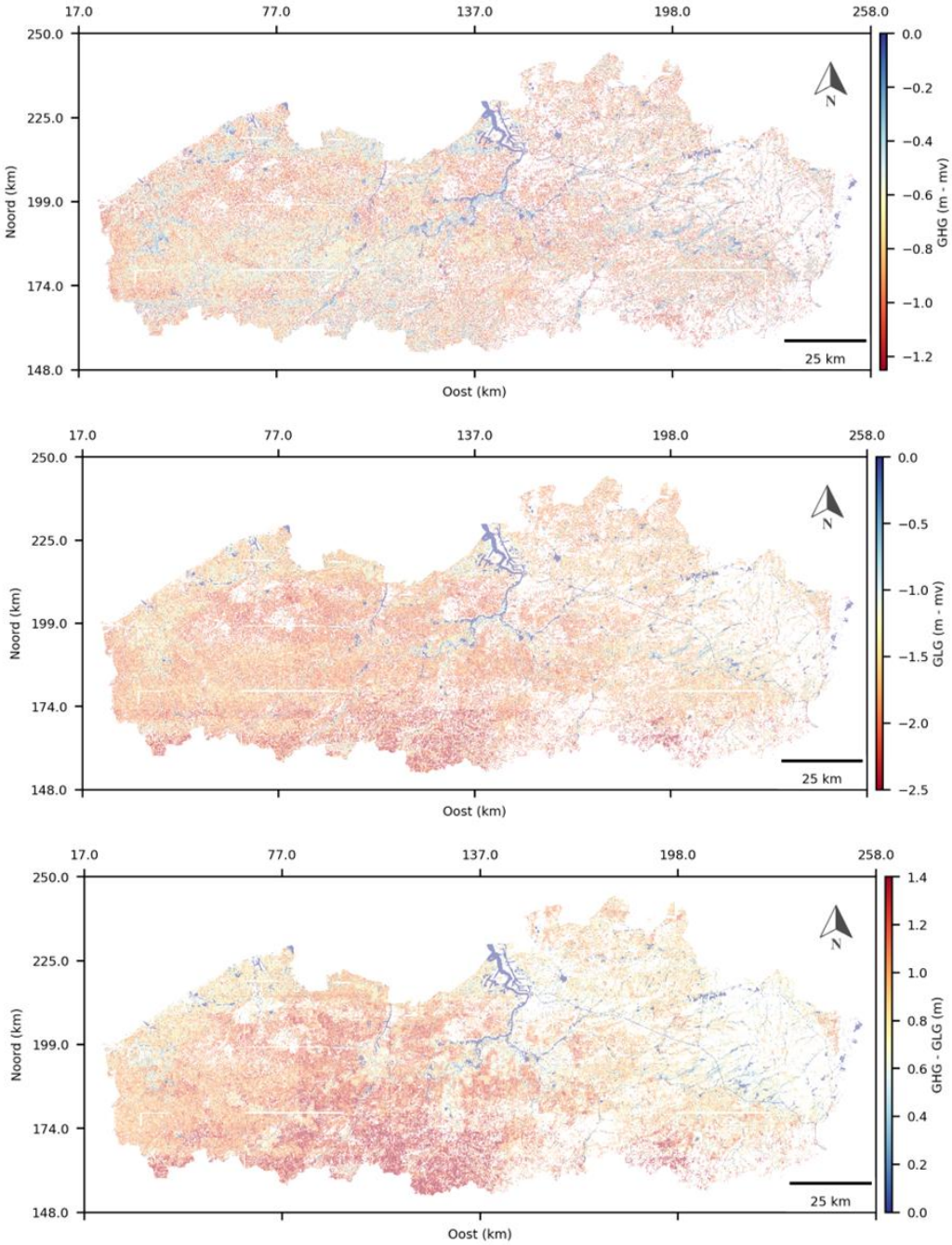
1. Een **Vlaanderendekkend raster van GHG en GLG met 10x10m resolutie**. De modelfout voor het simuleren van de GHG ligt, afhankelijk van de textuur- en drainageklasse, tussen de 20 en 35 cm. De fout op de GLG ligt hoger (tussen de 9 cm en 161 cm) voor de sterk drainerende zones waar het grondwaterpeil tot voorbij 3 m onder het maaiveld komt. Voor iedere rastercel wordt een maat van modelonzekerheid meegegeven: het 5de en 95ste percentiel van de kwantielregressie. Algemeen kan men opmerken dat het verschil tussen de twee uiterste percentielen groter is voor dieper gelegen voorspellingen (+/- 1m) vergeleken met minder diepe voorspellingen (+/- 0,5m) maar het blijft steeds nodig om dit op pixel niveau te evalueren, informatie die de gebruiker steeds kan raadplegen. Omdat waarden met een GHG dieper dan 125cm onder het maaiveld uiteindelijk allen worden geklassificeerd in een en dezelfde drainageklasse kozen we ervoor om die waarden, gezien hun grotere onzekerheid, te maskeren in de figuren van dit rapport.
2. **Geactualiseerde drainageklassen van de bodemkaart** op basis van de gemodelleerde GHG- en GLG-rasters Voor elke polygoon van de bodemkaart werd vervolgens de drainageklasse als volgt bepaald: we berekenden de mediaan van de GHG/GLG combinatie voor alle

rastercellen en vertaalden die naar een drainageklasse. De onzekerheid op die drainageklasse werd toegekend door ook het 25^{e} en 75^{e} percentiel voor die polygoon te rapporteren. Er worden ook drainageklassen voorzien voor de gebieden waar deze niet eerder voorhanden waren: de kustzone en verstedelijkte gebieden. In vergelijking met de oude drainageklassen zijn vooral de matig drainerende bodems (drainageklasse c) en de te sterk drainerende bodems (drainageklasse a) in aantal gestegen. Dit voornamelijk ten koste van de gunstig drainerende (drainageklasse b) en onvoldoende tot matig slecht drainerende bodems (resp. drainageklasse c en d). Dit wordt enerzijds veroorzaakt door een verschillende techniek om de drainageklasse af te leiden (gemodelleerde GxG's op basis van recent gemeten grondpeilen versus waargenomen historische bodemkenmerken) en anderzijds door verdroging van de ondergrond gedurende de laatste 50 jaar.

3. **Het volledige modelinstrumentarium** en de gebruikte data werden overgedragen aan de Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving. Het model draait op hun systeem. Wanneer er nieuwe informatie ter beschikking komt, kan men de kaart volledig, of kaartblad per kaartblad, geactualiseerd worden.

De nieuwe output komt tegemoet aan de vereisten die naar boven kwamen bij een eerdere gebruikersbevraging door Finke *et al.* (2010), die worden bevestigd door een nieuwe, beperkte, bevraging binnen deze studie: het kwantitatieve karakter, de resolutie, het gebiedsdekkende karakter en de meegegeven maat van onzekerheid worden als een aanzienlijke verbetering ten opzichte van de eerdere drainageklassenkaart aanzien.

Samengevat kan men stellen dat de bruikbaarheid van de nieuwe resultaten aanzienlijk hoger ligt dan de vorige drainageklassen van de bodemkaart. De nieuwe kaarten kunnen aangewend worden voor een eerste verkenning van het studiegebied, als hulpmiddel om de locatie van peilbuizen te bepalen of voor studies waar men geen nieuwe peilbuizen kan plaatsen. Bij gebruik en interpretatie van de nieuwe kaarten, dient de gebruiker echter nog steeds de lokale topografie en bodem in acht te nemen. Voor studies die een hoge nauwkeurigheid vergen (met de ecohydrologische studies als summum), zal men nog steeds een gedetailleerd grondwatermodel moeten afstemmen, gekalibreerd op de lokale situatie.



FIGUUR 1: Resulterende gesimuleerde GHG (BOVEN) en GLG (ONDER)

INHOUDSTAFEL

Managementsamenvatting 3
Inhoudstafel 6
Lijst van figuren 8
Lijst van tabellen 11
Lijst van afkortingen 12
1 Inleiding 14
1.1 Opdracht 14
1.2 Situering rapport & leeswijzer 14
2 Achtergrond 15
2.1 Nood aan geactualiseerde drainageklassen van de bodemkaart 15
2.2 Voorgaande studies 16
2.2.1 Finke *et al.* (2005) 16
2.2.2 Finke *et al.* (2010) & Van de Wauw (2012) 17
3 Dataverzameling en -correctie 23
3.1 Grondwaterpeilmetingen 23
3.1.1 Dataverzameling 23
3.1.2 Datakwaliteit 24
3.1.3 Datacorrectie 24
3.2 Neerslag & Potentiële evaporatie 26
3.2.1 Dataverzameling 26
3.2.2 Datacorrectie 27
3.3 Remote Sensed Bodemvocht 28
3.3.1 Dataverzameling 28
3.4 Databank Ondergrond Vlaanderen 30
3.4.1 Thema Bodem 30
3.4.2 Thema Geologie 33
3.4.3 Thema Grondwater 35
3.5 Digitaal Hoogtemodel Vlaanderen 38
3.6 Waterlopen 38
3.6.1 Vlaamse Hydrografische Atlas (VHA) 38
3.6.2 Grachtenlaag Universiteit Antwerpen 38
3.6.3 Project Agentschap Informatie Vlaanderen 38
3.6.4 Eerder intern onderzoek Antea 38
3.6.5 Open Street Map (OSM) 39
3.6.6 Datacorrectie: uiteindelijke grachtendetectie 39
3.7 Bodembedekkingskaart 40
4 Feature Engineering: Verklarende Variabelen 41
4.1 Afgeleiden van de Bodemkaart 42
4.2 Afgeleiden van de Bodemerodibiliteitskaart (k-factor) 43
4.3 Afgeleiden van het Digitaal Hoogtemodel 44
4.4 Afgeleiden van het Hydrogeologisch Model 46
4.4.1 Dagzomende hydrogeologische laag 46
4.4.2 Diepte tot de top van de bovenste 'Aquitard' 47
4.5 Afgeleiden van Waterlopen 48
4.5.1 De afstand tot de dichtstbijzijnde waterloop 48
4.5.2 Drainagediepte en drooglegging (GWT) 49
4.5.3 Drainagedichtheid 50
4.6 Afgeleiden van SWI 50
4.7 Grondwaterwinningen 51
4.8 Schaling en onderlinge correlatie 53
5 Methodologie 55

5.1 Definitie GxG 55
5.2 Tijdreeksmodellering 56
5.2.1 Inleiding 56
5.2.2 PIRFICT 56
5.3 Incidentele metingen: Relateren in ruimte 61
5.4 Zoektocht Optimale Herkarteringstechniek 61
5.4.1 Inleiding 61
5.4.2 Data-assimilatie op PIRFICT-parametrisatie 62
5.4.3 Link tussen bodemvocht en grondwaterpeil 63
5.4.4 Regressieanalyse PIRFICT-parameters 64
5.4.5 Interpolatie/regressie van GxG's 65
5.5 Opdeling Test- en trainingset 67
5.5.1 Inleiding 67
5.5.2 Verkenning Testset 67
5.5.3 Trainingset 70
5.6 Regressietechniek 71
5.6.1 Ordinary Least Squares (OLS) 72
5.6.2 Geregulariseerde OLS (Elastic Net) 72
5.6.3 Geografisch Gewogen Regressie (GWR) 73
5.6.4 Extreme Gradiënt Boosting (XGB) 73
5.6.5 Resultaat 74
5.7 Kwantielmodel 74
5.8 Van GxG's naar drainageklassen van de bodemkaart 74
6 Bevindingen 77
6.1 Bevindingen Methodologie 77
6.1.1 Data-gedreven methodiek 77
6.1.2 Herkarteringstechnieken 77
6.1.3 Regressietechnieken 77
6.2 Bevindingen Verklarende Variabelen 78
6.3 Bevindingen GxG-kaart 80
6.3.1 Vlaanderendekkende GHG- en GLG-kaarten 80
6.3.2 Histogrammen en residuen 83
6.3.3 RMSE 86
6.3.4 Onzekerheid 87
6.3.5 De pilootzones nader bekeken 89
6.3.6 Kwalitatieve vergelijking met studie VMM 92
6.4 Bevindingen Drainageklassen van de Bodemkaart 95
7 Aanbevelingen bij het gebruik van de gegenereerde kaarten 99
8 Aanbevelingen voor verdere ontwikkelingen 102
Referenties 104
Webpagina's 106
Appendix 1: Inventaris 107
Appendix 2: Waterinfo kwaliteitscodes 111

LIJST VAN FIGUREN

FIGUUR 1: RESULTERENDE GESIMULEERDE GHG (BOVEN) EN GLG (ONDER) 5
FIGUUR 2: WERKWIJZE FINKE *ET AL. (*2005) BIJ HUN KARTERING VAN DE GRONDWATERDYNAMIEK. 17
FIGUUR 3: SCHEMATISCHE WERKWIJZE BIJ DE HERKARTERING VAN DRAINAGEKLASSE VAN DE BODEMKAART VOLGENS FINKE *ET AL.* (2010) 18
FIGUUR 4: DE 1-DIMENSIONALE WATERBALANS UIT VAN DE WAUW (2012) 19
FIGUUR 5: PILOOTZONE BIJ DAMME (VAN DER WAUW, 2012) 22
FIGUUR 6: ECOREGIO'S IN VLAANDEREN (BRON: UIT VAN DE WAUW, 2012; NAAR SEVENANT *ET AL.*, 2002) 22
FIGUUR 7: VOORBEELDEN VAN ARTEFACTEN IN GRONDWATERPEILREEKSEN: UITSCHIETERS (LINKS), SPRONG (MIDDEN), HIAAT (RECHTS). 24
FIGUUR 8: VOORBEELD TIJDREEKS OP LOCATIE DBLP123 WAARBIJ METINGEN VERWIJDERD WORDEN (RODE VERTICALE LIJNEN) OMDAT ZE MEER DAN 3 MAAL DE STD AFWIJKING VAN DE VERWACHTE WAARDE (BLAUWE LIJN) BEDRAGEN. DE GROENE BAND IS 1.96 MAAL DE STD ROND DE VERWACHTE WAARDE (95% VAN DE DATA VALT HIERBINNEN). 25
FIGUUR 9: VOORBEELDEN VAN GWP-TIJDREEKSEN DIE GEMARKEERD WORDEN ALS EEN ANOMALIE EN NIET BESCHOUWD WORDEN BIJ GXG INSCHATTINGEN. 26
FIGUUR 10: MEETSTATIONS (RODE LOCATIES) WAARVAN NEERSLAG (BOVEN) EN POTENTIËLE EVAPORATIE (ONDER) WERDEN GEDOWNLOAD VAN WATERINFO.BE. PROVINCIE CONTOUREN ZIJN TOEGEVOEGD. 27
FIGUUR 11: PRECIPITATIEMETINGEN TE BOEKHOUTE ZONDER (BOVEN) EN MET (ONDER) IMPUTATIE VAN MISSENDE METINGEN. 28
FIGUUR 12: SCHEMATISCHE WEERGAVE VAN HOE DE SWI WORDT BEKOMEN UIT SAR SSM EN SCAT SSM 28
FIGUUR 13: DOV-BODEMDATABANK (BRON: HTTPS://WWW.MILIEUINFO.BE/CONFLUENCE/DISPLAY/DDOV/DOV-BODEMDATABANK, GERAADPLEEGD OP 2021-04-07) 31
FIGUUR 14: RUIMTELIJKE VERDELING TEXTUURKLASSEN IN VLAANDEREN. G=STENIGE GRONDEN (BRONDATA: DOV) 32
FIGUUR 15 – VERGELIJKING VAN DE METHODE VOOR GRACHTENDETECTIE VOOR EEN TESTGEBIED IN DE DIJLE MET: DE BRON, HET 1X1M DTM (LINKSBOVEN); DE GEÏDENTIFICEERDE GRACHTEN (RECHTSBOVEN); EN TER VERGELIJKING ZOWEL HET GRB (LINKSONDER); ALS DE WATERLOOPSEGMENTEN VAN HET TOP 10 VECTORBESTAND VAN HET NATIONAAL GEOGRAFISCH INSTITUUT (RECHTSONDER). 39
FIGUUR 16 :TOTAAL DRAINAGENETWERK VHA EN GRACHTEN. 40
FIGUUR 17 : CELLEN WAARVOOR DE HULPVARIABELEN AFZONDERLIJK WORDEN BEREKEND 42
FIGUUR 18: TEXTUURKLASSEN (BOVEN) EN DRAINAGEKLASSEN (ONDER) UIT DE BODEMKAART (BRON: DOV). 43
FIGUUR 19 : DTM KAART VLAANDEREN IN M TAW. 45
FIGUUR 20 : VOORBEELD VAN 10M RESOLUTIE DTM-RASTER IN DE DIJLEVALLEI. 45
FIGUUR 21 : HISTOGRAM VAN DTM-WAARDEN IN DE DIJLEVALLEI. 46
FIGUUR 22 : DAGZOMENDE HCOV LAGEN IN DE DIJLEVALLEI. 46
FIGUUR 23 : HISTOGRAM VAN DAGZOMENDE HCOV LAGEN IN DE DIJLEVALLEI. 47
FIGUUR 24 : DIEPTE VANAF HET MAAIVELD TOT AAN DE BOVENSTE AQUITARD. 47
FIGUUR 25 : DIEPTE TOT EERSTE AQUITARD IN DE DIJLEVALLEI. 48
FIGUUR 26 : HISTOGRAM VAN AQUITARDDIEPTES IN DE DIJLEVALLEI. 48
FIGUUR 27 :SCHEMATISCHE WEERGAVE VAN DE STIJGHOOGTEVERSCHILLEN DIE DE VERSCHILLENDE COMPONENTEN VAN DE DRAINAGEWEERSTAND BEPALEN (NAAR: VAN HOORN (1960)). 49
FIGUUR 28: AFGELEIDE VAN SWI MET HIEROP DUIDELIJK EEN AANTAL ONTBREKENDE LOCATIES (STERK VERSTEDELIJKTE GEBIEDEN). DE HOGERE WAARDEN (LICHTERE KLEUR) ZIJN DE HOGERE VOCHTGEHALTES OVER EEN LANGJARIG GEMIDDELDE (2015-2021) EN OVER ALLE DIEPTES T.E.M. 1M. DE LAGERE WAARDES (DONKERDERE KLEUR) KUNNEN WORDEN GEÏNTERPRETEERD ALS LAGERE ALGEMENE VOCHTGEHALTES. 50
FIGUUR 29: VOORBEELD VERLAGING TEN GEVOLGE VAN GRONDWATERWINNING. 52

FIGUUR 30: CORRELATIES TUSSEN VERSCHILLENDE VARIABELEN IN DE REGRESSIEANALYSE. 54
FIGUUR 31: GEBRUIKTE KLASSIEKE GXG DEFINITIE 55
FIGUUR 32: INTUÏTIE VAN PIRFICT-MODELLERING EN PARAMETERS A.D.H.V. 1 REGENVAL MOMENT OP EEN TOTALE SERIE VAN 100. A GEEFT DE TOTALE RESPONSWEGING, A EN N BEPALEN VERDER DE VORM VAN DE CURVE (ZIE RECHTERFIGUUR). DE PRECIPITATIE WORDT GECORRIGEERD VOOR DE EVAPORATIE OM TOT NEERSLAGOVERSCHOT TE KOMEN. DEZE AFWEGING WORDT BEPAALD DOOR DE 'ETP WEIGHT' PARAMETER. 57
FIGUUR 33: DE WEERHOUDEN LOCATIES VAN DE PIRFICT-MODELLERINGEN. DE KLEUR GEEFT DE PEARSON CORRELATIE VAN HOE GOED DE GESIMULEERDE DATA DE METINGEN KUNNEN BESCHRIJVEN. 59
FIGUUR 34: SIMULATIE VS GEMETEN GRONDWATERSTANDEN OP DE LOCATIE ZWAP253: EEN LOCATIE MET HOGE CORRELATIE VAN 0.97 (BOVEN), PIRFICT-LOCATIE 1-89079 MET LAGERE CORRELATIE VAN 0.47 (MIDDEN), PIRFICT-SIMULATIE OVER DE VOLLEDIGE TIJDSPANNE VOOR DE LOCATIE ZWAP253 (ONDER). 60
FIGUUR 35: TESTLOCATIES IN VLAANDEREN WAAR PIRFICT GESIMULEERDE GXG'S BESCHIKBAAR ZIJN (28 ZWARTE KRUISEN). 61
FIGUUR 36: BODEMVOCHTPROFIEL BOVEN DE GRONDWATERTAFEL BIJ 10, 5, 2 EN 1 M-MV EN VOOR 3 KARAKTERISTIEKE BODEMTYPES. (BRON: MIGUEZ-MACHO *ET AL.*, 2007, AANGEPAST UIT FAN *ET AL.,* 2007) 63
FIGUUR 37: VERGELIJKING TUSSEN TIJDSREEKSEN VAN ERS SWIT (RODE LIJNEN) EN GRONDWATERTAFEL HT (BLAUWE STIPPEN) OP MEERDERE LOCATIES: LENNISHEUVEL NL (A); HERKENBOSCH NL (B); FALMIGNOUL BE (C) EN GROOT-GENHOUT BEEK NL (D). DE ONDERSTE REEKS GEEFT EEN MEER GEDETAILLEERDE VERGELIJKING TUSSEN SWIT EN EEN AGGREGATIE VAN GESIMULEERDE/GEMETEN GRONDWATERTAFELS OP EEN SPECIFIEKE LOCATIE. (BRON: SUTANUDJAJA, 2012). 64
FIGUUR 38: ABSOLUTE CORRELATIE TUSSEN PIRFICT-PARAMETERS EN DE HULPVARIABELEN. 65
FIGUUR 39: SCHEMATISCHE WEERGAVE WERKMETHODIEK VOOR DATA-GEDREVEN MODELLERING. 66
FIGUUR 40 : LOCATIES DOV WAAR DE KLASSIEKE GXG DEFINITIE KAN WORDEN TOEGEPAST. 67
FIGUUR 41 : DE 154 LOCATIES VAN DE GRONDWATERINDICATOR. 68
FIGUUR 42 : BOXPLOTS VAN DE GXG'S PER OUDE DRAINAGE- EN TEXTUURKLASSE VAN DE BODEMKAART OP DE TESTSET. 69
FIGUUR 43 : BOXPLOTS VAN GXG'S TEGENOVER DE OUDE DRAINAGEKLASSEN EN TEXTUURKLASSEN VAN DE BODEMKAART. DE DATASET IS 2X/MAAND 8J +PIRFICT0.8+INCID. 71
FIGUUR 44 LEGENDE VERTALING GXG NAAR DRAINAGEKLASSEN (FINKE, 2010). DE AFKORTINGEN STAAN VOOR: Z=ZAND, S=LEMIG ZAND, P=LICHT ZANDLEEM, A=LEEM, L=ZANDLEEM, E=KLEI, U=ZWARE KLEI 76
FIGUUR 45: RESULTATEN VAN SHAP-WAARDEN IN DE GXG PREDICTIE MET XGB OP DE TRAININGSET (2X/M 8J + PIRFICT + INCIDENTELE METINGEN) (BOVEN) EN DE TESTSET (ONDER). DE VARIABELEN ZIJN GERANGSCHIKT VOLGENS AFNEMEND ALGEMEEN BELANG. 80
FIGUUR 46: RESULTERENDE GXG-KAARTEN: GESIMULEERDE, VLAANDERENDEKKENDE GHG (BOVEN), GLG (MIDDEN) EN VERSCHILKAART TUSSEN BEIDE (ONDER) 82
FIGUUR 47: DE HISTOGRAMMEN, WEERGEGEVEN ALS EEN KERNEL DENSITY PLOT, VAN DE GHG (BLAUW), EN DE GLG (ORANJE). DE LIJN STIJL GEEFT AAN OF HET DE OBSERVATIES (SOLIDE) OF DE VOORSPELLING (PUNTEN) BETREFT. MERK OP DAT DEZE HISTOGRAMMEN OP BASIS VAN DE GEHELE DATASET (TRAINING + TEST) BEPAALD ZIJN. 83
FIGUUR 48: BOXPLOTS VAN DE RESIDUEN PER DRAINAGE- EN TEXTUURKLASSE. 84
FIGUUR 49: VERGELIJKING VAN DE RESIDUEN MET DE MODELWAARDEN, VOOR DE GHG (BOVEN) EN GLG (ONDER). DE ORANJE PUNTEN ZIJN DE GWI-WAARDEN, DE BLAUWE ZIJN DE VOLLEDIGE DATASET. 85
FIGUUR 50 HET ABSOLUUT RESIDU OP DE GLG I.F.V. DE GAUSSIAN KERNEL DENSITY 86
FIGUUR 51: VOOR ZOWEL GHG (BOVEN) ALS GLG (ONDER): DE ONZEKERHEIDSBANDEN (5-95%), DE TOTALE SET WAARDEN (ORANJE PUNTEN), DE VOORSPELLING (BLAUWE LIJN). DE WAARDEN ZIJN GEORDEND OP BASIS VAN DE DOOR HET MODEL VOORSPELDE WAARDE. 88
FIGUUR 52 VOORBEELD VAN EEN GHG-VOORSPELLING WAARBIJ DE OVERGANG TUSSEN TWEE SWI-TEGELS DUIDELIJK ZICHTBAAR IS (LINKS). PROFIELKWANTIELEN (Q0.05 - Q0.95) VAN DE GHG-VOORSPELLING BIJ OVERGANG TUSSEN TWEE SWI-TEGELS (RECHTS) 89
FIGUUR 53 DHM (LINKS) EN TEXTUURKLASSE (RECHTS) - PILOOTZONE ANTWERPEN 90

FIGUUR 54 DHM (LINKS) EN TEXTUURKLASSE (RECHTS) - PILOOTZONE DAMME 90
FIGUUR 55 DHM (LINKS) EN TEXTUURKLASSE (RECHTS) - PILOOTZONE DIJLE 91
FIGUUR 56 DHM (LINKS) EN TEXTUURKLASSE (RECHTS) - PILOOTZONE LIEREMAN 91
FIGUUR 57 DHM (LINKS) EN TEXTUURKLASSE (RECHTS)-PILOOTZONE RODELAND 91
FIGUUR 58: GLG-VOORSPELLINGEN VAN SUMAQUA, B: GLG-VOORSPELLINGEN VAN ANTEA MET DE LOCATIE VAN DE TRAININGSDATA, C: VERSCHIL TUSSEN BEIDE PREDICTIES: POSITIEVE WAARDEN WIJZEN OP EEN LAGER GRONDWATERPEIL VOORSPELD DOOR SUMAQUA. 94
FIGUUR 59: VERGELIJKING SUMAQUA (A) EN ANTEA (B) MET DE RESULTATEN VAN ANTEA GEFILTERD VOOR DIE LOCATIES WAAR DE GHG OP EEN MAXIMALE DIEPTE VAN 125CM ONDER HET MAAIVELD LIGT. DE VERSCHILKAART VAN DE TWEE WORDT WEERGEGEVEN IN FIGUUR C. 95
FIGUUR 60 NIEUWE DRAINAGEKLASSE VAN DE BODEMKAART 97
FIGUUR 61 OUDE DRAINAGEKLASSE VAN DE BODEMKAART 97
FIGUUR 62: VERSCHILKAART TUSSEN DE OUDE EN DE NIEUWE DRAINAGEKLASSEN VAN DE BODEMKAART: NEGATIEF = NATTER, POSITIEF = DROGER. 97
FIGUUR 63: VERSCHILKAART TUSSEN DE TWEE UITERSTE KWANTIELEN VAN DE NIEUWE DRAINAGEKLASSEN VAN DE BODEMKAART 98

LIJST VAN TABELLEN

TABEL 1: MOGELIJKE GRONDWATERMODELLEN (C.L..) OM DE 1-D WATERBALANS OP TE LOSSEN. MET HIERBIJ: N HET AANTAL *THRESHOLDS* OF DRAINAGETERMEN DIE IN ACHT GENOMEN WORDEN, P^E_T HET NEERSLAGOVERSCHOT OP TIJDSTIP T, γh EN γv DE HORIZONTALE EN VERTICALE DRAINAGEWEERSTAND EN Q_B, Q_L DE HORIZONTALE EN VERTICALE GRONDWATERFLUXEN (BRON: VAN DE WAUW, 2012). 20
TABEL 2: EVALUATIE VAN DE VERSCHILLENDE STRATIFICATIEOPTIES (FINKE *ET AL.*, 2010) 21
TABEL 3: DIVERSE DATABRONNEN MET AANTAL LOCATIES EN METINGEN PER LOCATIE. 23
TABEL 4: VOORNAAMSTE EIGENSCHAPPEN VAN DE INPUT DATA (ASCAT & SENTINEL-1) EN DE GEFUSEERDE OUTPUT. 29
TABEL 5: OVERZICHT EIGENSCHAPPEN COPERNICUS SWI 29
TABEL 6 : BESCHRIJVING HULPVARIABELEN 41
TABEL 7: TABEL MET BODEMSERIES WAARVAN DE TEXTUURKLASSE ONTBREEKT. 44
TABEL 8: OMZETTING VAN K-FACTOR NAAR TEXTUURKLASSE. 44
TABEL 9 : OVERZICHT GEBRUIKTE TRAININGSETS EN AANTAL LOCATIES 70
TABEL 10: FOUTENSTATISTIEKEN BEREKEND OP DE TESTSET PER MODELTECHNIEK, INCL. AANDUIDING MEEST OPTIMALE TRAININGSET DIE HOORT BIJ DEZE MODELTECHNIEK. DE BESTE RESULTATEN ZIJN WEERGEGEVEN IN HET ROOD. 74
TABEL 11: NATUURLIJKE DRAINAGEKLASSE (VAN RANST & SYS, 2000) 75
TABEL 12: DE RMSE OP DE GHG-VOORSPELLING PER TEXTUUR- EN DRAINAGEKLASSE. DE EERSTE WAARDE IS DE RMSE VOOR DE VOLLEDIGE TRAININGSET, DE TWEEDE WAARDE ENKEL VOOR DE TESTSET. 86
TABEL 13: DE RMSE OP DE GLG-VOORSPELLING PER TEXTUUR- EN DRAINAGEKLASSE. DE EERSTE WAARDE IS DE RMSE VOOR DE VOLLEDIGE TRAININGSET, DE TWEEDE WAARDE ENKEL VOOR DE TESTSET. 87
TABEL 14: HET AANTAL OBSERVATIES PER TEXTUUR- EN DRAINAGEKLASSE. DE EERSTE WAARDE IS HET AANTAL OBSERVATIES IN DE VOLLEDIGE TRAININGSET, DE TWEEDE WAARDE HET AANTAL OBSERVATIES IN DE TESTSET. 87
TABEL 15: KRUISTABEL AANTAL POLYGONEN OUDE-VERSUS NIEUWE DRAINAGEKLASSE VAN DE BODEMKAART 96

LIJST VAN AFKORTINGEN

Afkorting	Betekenis
ALBON, dept. LNE	Afdeling Land en Bodembescherming, Ondergrond, Natuurlijke Rijkdommen, Departement Leefmilieu, Natuur en Energie (Vlaamse Overheid)
ANB	Agentschap voor Natuur en Bos
ASCAT	Advanced Scatterometer
BBK	Bodembedekkingskaart Vlaanderen
BBT	Best Beschikbare Technieken
BGD	Belgische Geologische Dienst
BVR	Besluit van de Vlaamse Regering
DFME	Difference From Mean Elevation
DHM	Digitaal Hoogtemodel
DHMV II	Digitaal Hoogtemodel Vlaanderen
DOV	Databank Ondergrond Vlaanderen
DSM	Digitaal Oppervlakte (Surface) Model
DTM	Digitaal Terreinmodel
reIDTM	Relatieve DTM (Digitaal Terreinmodel)
ELN	Elastic Net
ERS	Estimating Soil Moisture
G3Dv2	Geologisch 3D-lagenmodel van Vlaanderen - versie 2
G3Dv3	Geologisch 3D-lagenmodel van Vlaanderen - versie 3
GHG	Gemiddeld Hoogste Grondwaterstand
GLG	Gemiddeld Laagste Grondwaterstand
GRB	Grootschalig Referentiebestand / Basiskaart Vlaanderen
GVG	Gemiddeld Voorjaarsgrondwaterstand
GWI	Grondwaterstandindicator
GWP	Grondwaterpeil
GWR	Geografisch Gewogen Regressie
GWT	Grondwatertafel
GxG	GLH en GLG tezamen
H3Dv2	Hydrogeologisch 3D-model van Vlaanderen - versie 2
HCOV	Hydrogeologische Codering van de Ondergrond van Vlaanderen
HIC	Hydrologisch Informatiecentrum
IMJV	Integraal Milieujaarverslag
INBO	Instituut voor Natuur- en Bosonderzoek
K-factor	Bodemerodibiliteitsfactor / Bodemerosiegevoeligheidsfactor
K-Kaart	Bodemerodibiliteitskaart
Lasso	L1-regularisatie / "*Least Absolute Shrinkage and Selection Operator*"-Regularisatie
LiDAR	Light Detection And Ranging / Laser Imaging Detection And Ranging
MAE	Mean Absolute Error
MINA	Milieu- en Natuurraad van Vlaanderen II
m-MV	Meter t.o.v./onder Maaiveld

MSE	Mean Squared Error
m-TAW	Meter t.o.v. Tweede Algemene Waterpassing
NRT	Natuur, ruimte en techniek
OLS	Ordinary Least Squares
OSM	Open Street Map
QFLAG	Datakwaliteit - Kwaliteitsvlag
REA	Relative Elevation Attribute
RMSE	Root Mean Squared Error
RUSLE	Revised Universal Soil Loss Equation
SAR	Synhetic Aperture Radar
SCAT	Scatterometer
SHAP	Shapely Additive Explanations
SSM	Surface Soil Moisture
SWAP	Soil Water Atmosphere Plant (model)
SWI	Soil Water Index product van Copernicus (Copernicus Global Land Service, 2022)
SWSI	Sum of Weighted Squares Innovations
TAW	Tweede Algemene Waterpassing
VHA	Vlaamse Hydrografische Atlas
VITO	Vlaamse Instelling voor Technologisch Onderzoek
VLAKO	Vlaams Kenniscentrum Ondergrond
VLAREM	Vlaams Reglement betreffende de Milieuvergunning
VLM	Vlaamse Landmaatschappij
VMM	Vlaamse Milieumaatschappij
VPO	Vlaams Planbureau voor Omgeving
VVPW	Vereniging van Vlaamse Polders en Wateringen
WATINA	Water in Natuur (databank INBO)
WCS	Web Coverage Service
WGN	White Gaussian Noise
WRB	World Reference Base for Soil Resources / Internationale Bodemclassificatie
XGB	Extreme Gradiënt Boosting

1 INLEIDING

1.1 OPDRACHT

De drainageklasse van de bodemkaart is nog steeds één van de meest gebruikte parameters in zeer diverse modellen en toepassingen zoals modellen i.v.m. droogte en wateroverlast, berekening van infiltratievoorzieningen, transport van polluenten en nutriënten door de bodem, organische koolstofvoorraden in de bodem, ecosysteemdiensten, bodemgeschiktheid, enzovoort.

De drainageklassen van de bodemkaart, beschikbaar voor iedere polygoon van de bodemkaart, zijn echter niet meer up-to-date vermits ze in het veld werden opgenomen tijdens de bodemkartering (1947-1974) en de hydrologie sindsdien plaatselijk sterk kan gewijzigd zijn. Bovendien baseert de beschrijving zich op het voorkomen van hydromorfe profielkenmerken die vaak na-ijlen op de hydrologische situatie. Op basis van de drainageklasse van de bodemkaart kan men dus mogelijk een verkeerd beeld krijgen van de hydrologische conditie van de bodem, verkeerde modelresultaten bekomen en verkeerde conclusies trekken. Het departement Omgeving (Afdeling Vlaams Planbureau voor Omgeving, VPO) wenst hieraan tegemoet te komen door expertise met betrekking tot grondwater en bodem met elkaar te combineren en hieruit actuele grondwaterstanden (en dus ook de drainageklassen) af te leiden.

Het doel van deze opdracht is de drainageklassen van de polygonen van de bodemkaart te actualiseren. Hiervoor worden grondwaterstatistieken berekend op recente grondwaterpeilmetingen gekoppeld aan relevante omgevingsdata zoals bijvoorbeeld bodemtype. De ontwikkelde methodiek moet toelaten dat, telkens wanneer er voldoende nieuwe data beschikbaar zijn, de drainageklassen vlot kunnen worden geactualiseerd.

De opdracht focust op de grondwatertafels die ondiep genoeg voorkomen om een invloed uit te oefenen op de bodemkenmerken. De grens werd gelegd op locaties met gemiddeld hoogste grondwaterstanden ondieper dan 3 m onder het maaiveld (m-MV).

1.2 SITUERING RAPPORT & LEESWIJZER

Voorliggend rapport, *Actualisatie van de drainageklasse van de bodemkaart door koppeling van grondwaterstatistieken aan de bodemkaart*, in opdracht van *Departement Omgeving – Vlaams Planbureau voor Omgeving*, beschrijft de resultaten en bespreekt de bevindingen aan het einde van de opdracht.

Het rapport start (**Hoofdstuk §2**) met een beknopte uiteenzetting van de achtergrond waarop deze studie steunt: (a) de reden waarom de drainageklassen van de bodemkaart geactualiseerd dienden te worden en (b) een samenvatting van de voorgaande studies waarop deze opdracht is opgebouwd. Dit wordt gevolgd door de dataverzameling en toegepaste datacorrectie in **Hoofdstuk §3**, waarna **Hoofdstuk §4** dieper ingaat op de toegepaste *Feature Engineering*. Hierna focust **Hoofdstuk §5** op de methodologie: de zoektocht naar en onderbouwing van de meest optimale/toepasbare methode. **Hoofdstuk §6** presenteert de bevindingen, terwijl **Hoofdstuk §7** advies verleent over het gebruik van alle nieuwe info. Aanbevelingen voor verdere ontwikkelingen zijn terug te vinden in **Hoofdstuk §8**.

2 ACHTERGROND

2.1 NOOD AAN GEACTUALISEERDE DRAINAGEKLASSEN VAN DE BODEMKAART

De traditionele beschrijving van de drainageklasse van de bodemkaart volgens Van Ranst & Sys (2000) bepaalt de drainageklasse op basis van het ondiepste voorkomen van roestvlekken (begindiepte gley) en van reductiekleuren, samen ook wel de hydromorfe profielkenmerken genoemd:

- Het gedeelte van het profiel dat afwisselend verzadigd is met water en uitdroogt, vertoont roestvlekken (zgn. gleyverschijnselen).
- De zone die voortdurend met water verzadigd is, heeft een blauw- of grijsachtige kleur (zgn. reductiehorizont).

Omdat hydromorfe profielkenmerken vaak na-ijlen op de geohydrologische situatie, zijn deze kenmerken niet optimaal om een veranderde drainagesituatie vast te stellen. Daarom werd overgegaan op het hanteren van een begrippenkader gebaseerd op meetbare grondwatertafels om de drainageklassen opnieuw vast te stellen. Omwille van de vergelijkbaarheid wordt de ondiepste aanwezigheid van roestvlekken geacht te corresponderen met de gemiddeld hoogste grondwaterstand (GHG). De diepte waar beneden geen roestvlekken meer werden aangetroffen, of waar blauwgrijze reductiekleuren werden aangetroffen, wordt geacht te corresponderen met de gemiddeld laagste freatische grondwaterstand (GLG). Een dergelijke vergelijking van hydromorfe profielkenmerken met meetbare freatische stijghoogtes werd ook toegepast in Nederland (De Bakker, 1973; Locher & De Bakker, 1991) en België (Van Damme & De Leenheer, 1968). Hier werd geconcludeerd dat de GHG redelijk correspondeerde met de bovengrens van de roestvlekken en de GLG goed correspondeerde met de bovengrens van de gereduceerde zone. Een verschil tussen de Nederlandse en de Belgische situatie is, dat in Nederland hydromorfe profielkenmerken werden gebruikt om de GHG en GLG te karteren, terwijl in België hydromorfe profielkenmerken zelf werden gekarteerd en deze worden geïnterpreteerd in een hydrologische context. De vertaalslag van hydromorfe profielkenmerken naar GHG+GLG vond in Nederland dus plaats tijdens de kartering, in het veld. In België zijn de hydromorfe profielkenmerken daarentegen zelf genoteerd en gekarteerd.

Natuurlijke en door de mens veroorzaakte factoren hebben de hydrologische toestand in Vlaanderen sinds de kartering van Vlaanderen ook aanzienlijk beïnvloed. Klimaatverandering heeft geleid tot veranderingen in temperatuur- en neerslagpatronen, wat de beschikbaarheid van water en de frequentie van overstromingen en droogte beïnvloedt. De toename van verstedelijking heeft geleid tot meer verharde oppervlakken, wat op zijn beurt de oppervlakteafvoer en afvoer van regenwater heeft gewijzigd. Voorts hebben veranderingen in landbouwpraktijken, waterbeheermaatregelen en grondwaterwinning invloed gehad op de waterbalans. De veranderde variabiliteit van de hydrologische omstandigheden is een bijkomend argument om de drainageklasse van de bodemkaart opnieuw onder de loep te nemen.

Men kan dus besluiten dat er een noodzaak is om de drainageklassen rechtstreeks af te leiden van de toenemende set aan grondwaterpeilmetingen. Dit werd onderlijnd door de gebruikers-inventarisatie uitgevoerd door Finke *et al.* (2010) die volgende drie voornaamste beperkingen naar voor bracht: (1) het niet actueel zijn van de drainageklassen; (2) het ontbreken van een

kwantitatieve variabele en (3) te weinig detail. Verder gaf deze gebruikersinventarisatie de nood aan actualisatie voor de stuwwatertafels en de nood aan drainageklassen voor de kustvlakte duidelijk weer. De meerderheid van de respondenten verkoos een herkartering i.p.v. een her-etikettering.

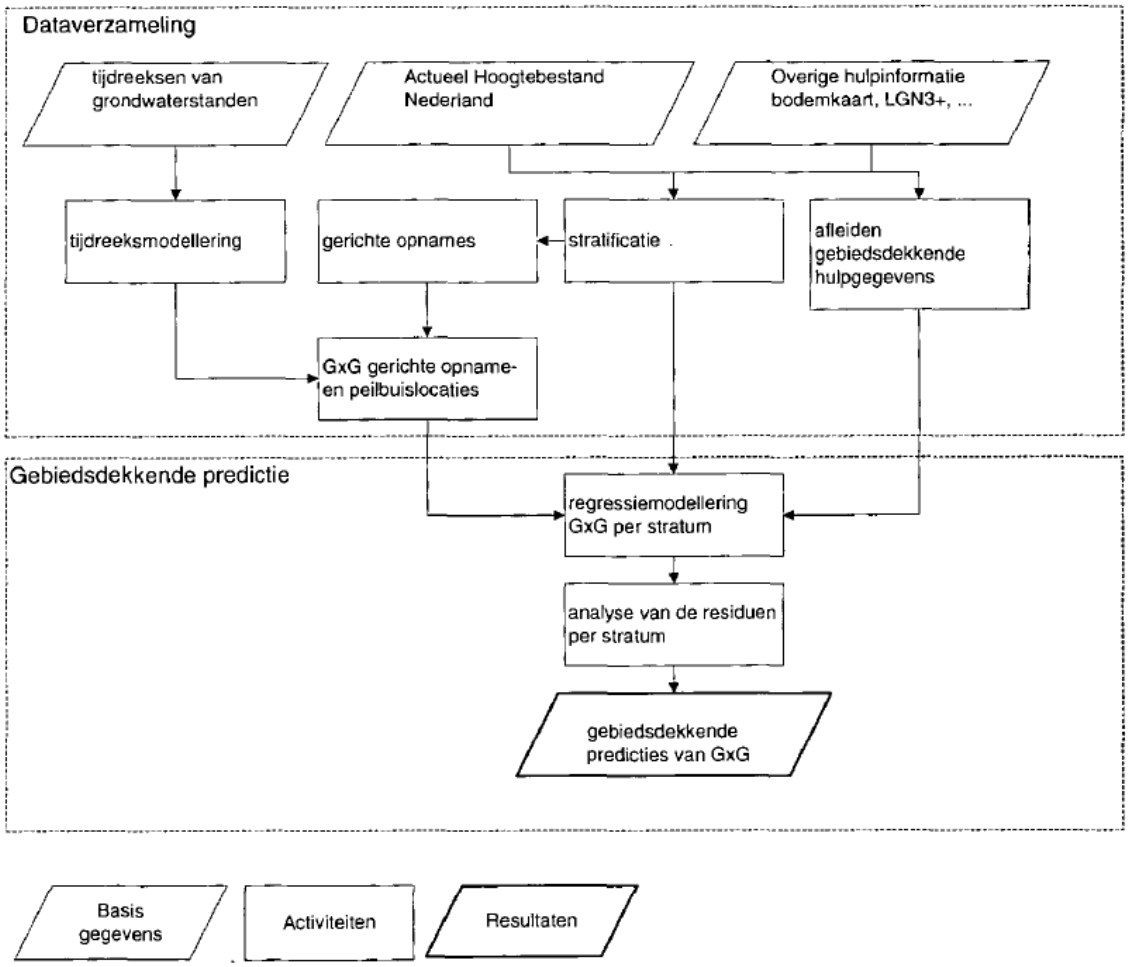
Tijdens de eerste stuurgroepvergadering kwam de terechte opmerking dat voorliggende opdracht eigenlijk geen actualisatie inhoudt, gezien de oude drainageklasse het resultaat is van historische bodemkartering en het eindproduct van voorliggende studie een kaart is van het grondwaterpeil, gerelateerd aan het huidige klimaat. Toch, gezien de titel van de opdracht al even vastlag – ook contractueel – en gezien eerdere communicatie naar buiten toe, is er besloten de titel van de rapporten niet te wijzigen.

2.2 VOORGAANDE STUDIES

Voorliggende studie bouwt voort op de voorgaande studies van Prof. Finke (o.m. Finke *et al.*, 2005; 2010; Van de Wauw, 2012). Deze studies worden hier beknopt samengevat.

2.2.1 Finke *et al.* (2005)

Finke *et al.* (2005) onderzochten in Nederland hoe de grondwaterdynamiek beter gekarteerd kon worden en dit met een gekwantificeerde nauwkeurigheid. Ze ontwikkelden een datagedreven werkwijze (Figuur 2) die minder kostte dan een traditionele Grondwatertrappen kartering, maar toch een complete set van parameters opleverde. De nauwkeurigheid van de GxG-voorspellingen werd ook in kaart gebracht, wat de mogelijkheid biedt tot verdere verbetering van de nauwkeurigheid van de kaart tegen gegeven kosten. Ze besloten dat de kaartkwaliteit tegen hoge kosten verbeterd kon worden door de waarnemingsdichtheid te vergroten. Een andere optie was nieuwe bronnen van hulpinformatie te gebruiken, zoals die bijvoorbeeld kunnen worden ontleend aan berekeningen met deterministisch-fysische hydrologische modellen.

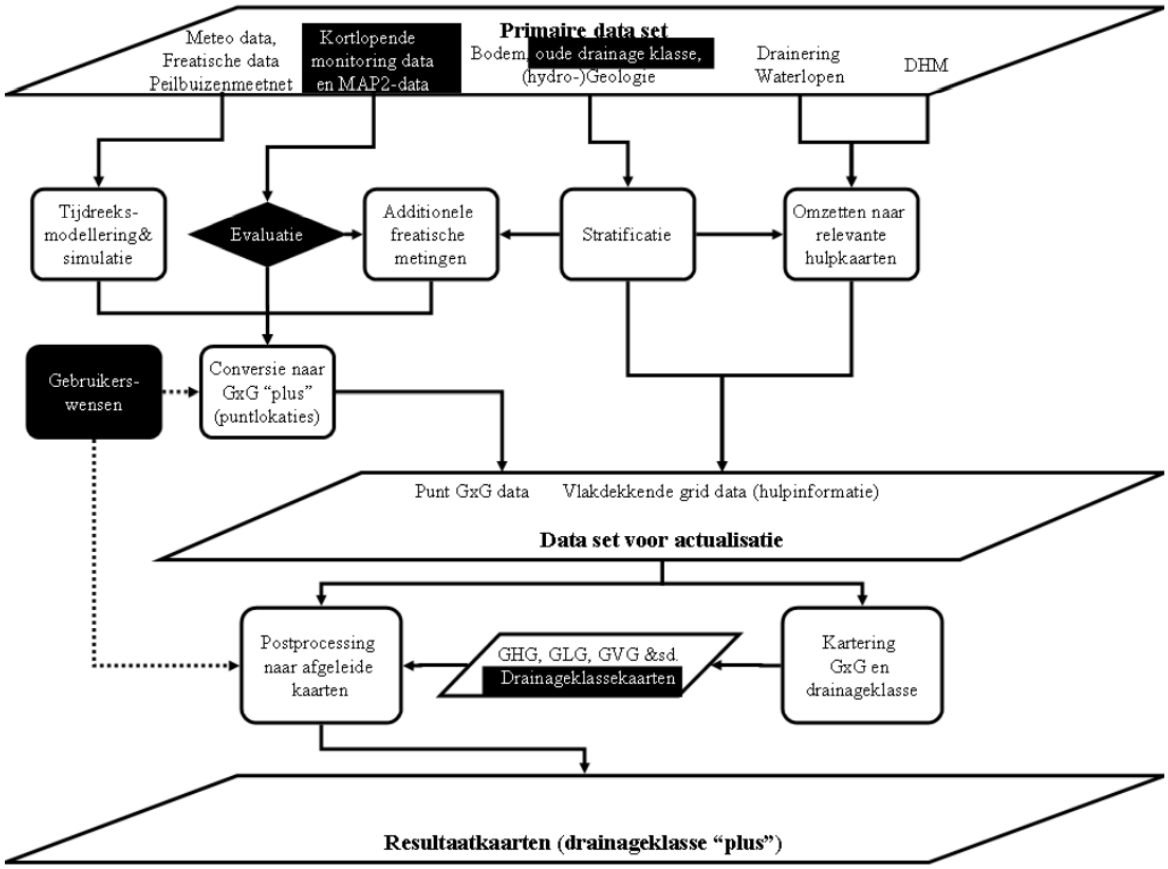


Figuur 2: Werkwijze Finke *et al.* (2005) bij hun kartering van de grondwaterdynamiek.

2.2.2 Finke *et al.* (2010) & Van de Wauw (2012)

Finke *et al.* (2010) bouwden voort op voorgaande bevindingen en ontwikkelden in opdracht van het toenmalig Departement Leefmilieu, Natuur en Energie een nieuwe methodiek voor het actualiseren van de drainageklassen van de bodemkaart van Vlaanderen. Hiervoor testten ze verscheidene (11) methoden uit van herclassificaties: (1) met behulp van louter peilbuisinformatie, (2) met een gerichte opname (meetcampagne) en (3) via geostatistische herkartering met gerichte opnames en ruimtelijke hulpinfo.

Uit hun studie bleek de **geostatistische herkartering met behulp van hulpdata de beste methode** (Figuur 3). Bovendien adviseerden ze in eerste instantie een waarnemingsdichtheid van 20 hectare per observatie. Deze waarnemingsdichtheid kan in tweede instantie lokaal worden vergroot na analyse van de resultaatkaarten per stratum, waarna de kaart kan worden bijgesteld. Met behulp van geïnterpoleerde residuen verkrijgt men dan een indicatie van waar de minder goed gekarteerde gebieden zich bevinden, die later bemonsterd kunnen worden tijdens een 2de campagne.



Figuur 3: Schematische werkwijze bij de herkartering van drainageklasse van de bodemkaart volgens Finke *et al.* (2010)

Tot slot werd ook de toenmalige kostprijs ingeschat om de drainageklassen te actualiseren, gerelateerd aan 3 scenario's:

- Scenario "regionaal streng" richt zich op het actualiseren van gehele ecodistricten die kwalitatief verbeterbaar) op de kaart staan, en besluit tot het actualiseren van 962 840 hectare tegen een geschatte kost van € 5,5M.
- Scenario "regionaal tolerant" richt zich eveneens op het actualiseren van gehele ecodistricten, maar alleen als de GxG-fouten een grenswaarde van 20 cm overschrijden. Het besluit is dan 764 681 hectare te actualiseren tegen een geschatte kost van € 4,427M.
- Scenario "lokaal tolerant" richt zich op het actualiseren van gebieden binnen elke ecoregio met een drainageklasse die kwalitatief onvoldoende op de kaart staat, waarbij de maximale getolereerde fout bij "natte" drainageklassen kleiner is (20 cm) dan bij droge drainageklassen (40 cm). Het besluit is dan 503 395 hectare te actualiseren tegen een geschatte kost van € 2,915M.

Bovenstaande studie vormde onderdeel van het doctoraatsonderzoek van Van de Wauw (2012). In zijn proefschrift gaat hij dieper in op een aantal relevante onderwerpen. Volgende hoofdstukken geven een thematische samenvatting van relevante items, behandeld in op zijn minst één van beide onderzoeken.

2.2.2.1 Te korte tijdsreeksen: extrapoleren en/of interpoleren in tijd

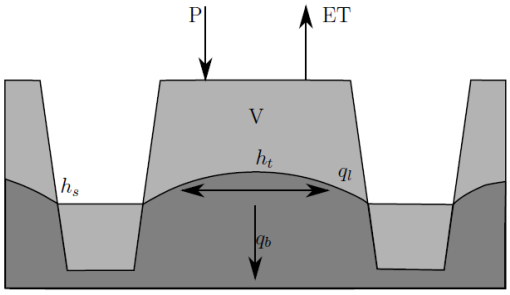
Voor een aantal peilmetingen met voldoende temporele resolutie en duur (verder referentiemetingen genoemd - vnl. deze van het primair meetnet) zijn de GxG's eenvoudig te

berekenen of zijn ze zelfs al beschikbaar. Echter voor een goede geostatistische herkartering dient men over een dichter meetnet (lees: hogere spatiale resolutie) te beschikken, terwijl veel meetreeksen niet voldoen qua temporele resolutie en/of duur.

Deze meetreeksen kan men uitbreiden met behulp van tijdreeksmodellering: modellen die de waterbalans trachten op te lossen via zgn. hulpdata:

$$\phi \frac{\partial h_t}{\partial t} = P_t^e + \frac{\partial V}{\partial t} - q_l - q_b$$

Met (zie ook Figuur 4): ϕ de bodemporositeit [m/m], h_t (grond)waterpeil op tijdstip t [m], P^e_t neerslagoverschot op tijdstip t [m/dag], V bodemvocht in de onverzadigde zone [m], q_l de laterale grondwaterflux naar drainerende oppervlaktewateren (m/dag), q_b regionale grondwaterflux zoals verticale drainage of opvulling [m/dag].



Figuur 4: De 1-Dimensionale waterbalans uit Van de Wauw (2012)

Men kan hierbij 2 grote types van modellen onderscheiden:

- De fysische (zgn. numerieke) modellen trachten een exacte oplossing te vinden van de waterbalans a.d.h.v. bodem-, vegetatie- en meteorologische parameters. Denk hierbij aan Hydrus-1D of SWAP.
- De empirische modellen: versimpelen de waterbalansvergelijking om zo minder parameters te bezitten (deterministisch deel) en fitten achteraf een foutenmodel (*noise-model* - stochastisch deel). Gezien de autocorrelatie in de fouten, wordt hier al snel een autoregressieve term in opgenomen.

Tabel 1 vat de eigenschappen van verschillende modellen samen. Uiteindelijk verkozen Finke *et al.* (2010) en Van de Wauw (2012) modellen met een stochastische component omdat deze beter de extreme waarden van het grondwaterpeil benaderden. Daarnaast verkozen ze ook een model met minder modelparameters. Tot slot verkozen ze empirische modellen boven de mechanistische gezien de eerste minder data vergen. Van de empirische konden ze enkel deze hanteren die geen *equidistante* tijdreeksen vergden – typisch geen eigenschap bij grondwatermodellen – en zo kwamen ze uiteindelijk bij de PIRFICT-modellen.

Tabel 1: Mogelijke grondwatermodellen (c.l..) om de 1-D waterbalans op te lossen. Met hierbij: n het aantal *thresholds* of drainagetermen die in acht genomen worden, P^e_t het neerslagoverschot op tijdstip t, γ_h en γ_v de horizontale en verticale drainageweerstand en q_b, q_l de horizontale en verticale grondwaterfluxen (Bron: Van de Wauw, 2012).

Model	Equidistant measurements required	Stoch.	Threshold	Data need	# Model parameters
Deterministic Models			✓	P_t^e	2
				Soil Hydraulics	6
				γ_h, γ_v	$2 \times n$
				q_l, q_b	2
Regressive modelling	✓			P_t^e	± 9
KalMAX		✓		P_t^e	3
KalTFN		✓		P_t^e	4
TARSO	✓	✓	✓	P_t^e	$n \times 4$
PIRFICT linear		✓		P_t^e	4
PIRFICT non linear		✓	✓	P_t^e	5

2.2.2.2 Incidentele metingen: relateren in ruimte

Overige metingen die te kort waren en/of men niet kon modelleren met PIRFICT, werden gerelateerd aan nabijgelegen langere meetreeksen om zo een GxG af te leiden. Hiervoor zijn 2 methodes toegepast:

- De eerste methode ("Gtkortereeks", Oude Voshaar & Stolp, 1997) koppelt informatie uit 1 referentiereeks met de korte reeks door middel van lineaire regressie. Dit gebeurt voor alle referentiereeksen in hetzelfde stratum (zie sectie §2.2.2.3). De relatie die de kleinste voorspelfout oplevert, wordt gebruikt om de GHG/GLG te voorspellen.
- Een tweede methode, de gerichte opname, wordt toegepast in het geval dat deze metingen zijn gedaan in een periode waarin de gemeten grondwatertafels de GxG benaderen. Deze metingen worden omgezet naar GxG door correlatie met gelijktijdige metingen en bekende GxG in freatische peilbuizen volgens de methode van Te Riele & Brus (1991). Hiertoe wordt, in de freatische peilbuizen, voor elke gerichte opname een regressierelatie bepaald tussen de watertafel op die datum en de GxG. Die relatie wordt vervolgens toegepast op de meting gedaan tijdens de gerichte opname, waardoor de gemeten watertafel wordt geconverteerd naar een GxG. De kwaliteit van de voorspelling van de GxG wordt ook hier gekwantificeerd met de standaardafwijking van de voorspelfout. De kwaliteit van een dergelijke voorspelling is uiteraard minder groot en ligt typisch tussen de 10 en 25 cm, afhankelijk van het aantal en de representativiteit van de beschikbare freatische peilbuizen.

2.2.2.3 Stratificatie

De regressieanalyse is opgedeeld in specifieke deelgebieden om zo de nauwkeurigheid van de output te optimaliseren. Beide onderzoeken hanteerden 2 stratificaties:

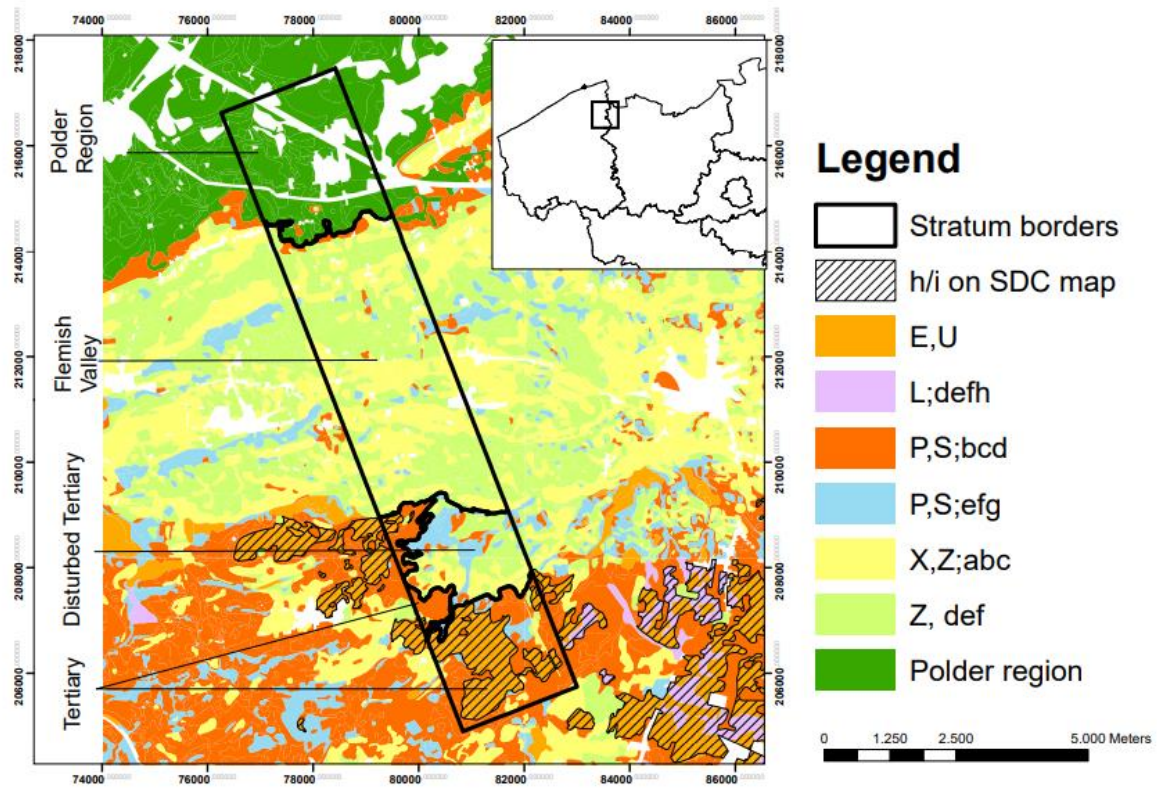
1. Op basis van de **ecoregio's** die gedefinieerd zijn door hun gelijkaardige geologie, geomorfologie, hydrologie en grondwaterregime (zie Figuur 6; Sevenant *et al.*, 2002). Deze kreeg de voorkeur in Vlaanderen ten opzichte van andere opties gezien deze voldoende groot zijn om voldoende meetpunten te bevatten, maar toch divers genoeg (Tabel 2). Bovendien produceert de

ecoregiokaart ruimtelijke eenheden waarvan de abiotische eigenschappen niet snel in de tijd veranderen. Deze ecoregio's hebben een oppervlak van 3 700 ha tot 372 000 ha.

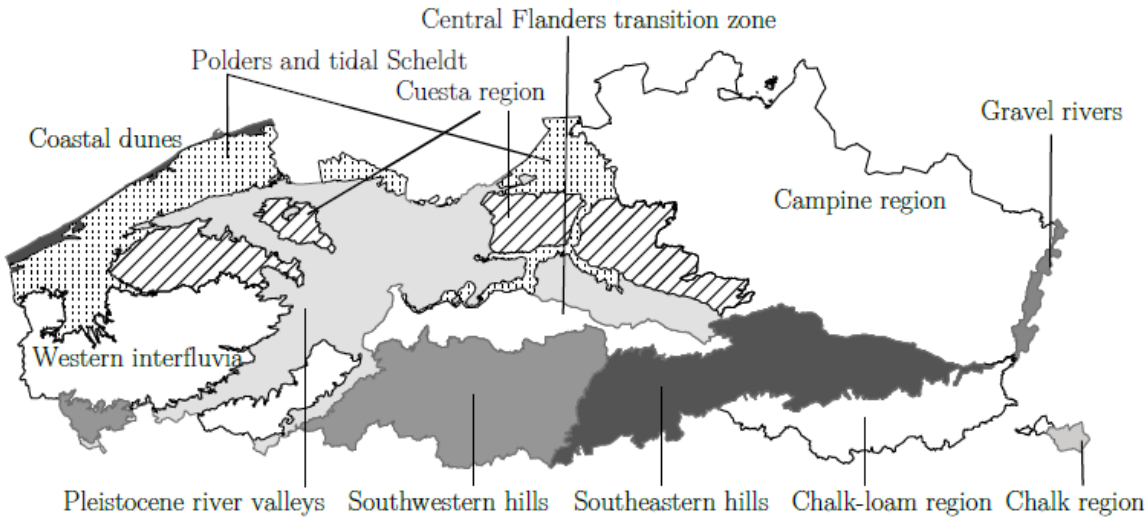
Tabel 2: Evaluatie van de verschillende stratificatieopties (Finke *et al.*, 2010)

Stratificatieoptie	Aantal eenheden	dekking
Landbouwstreken	7 (4 grote klassen)	11% niet gedekt
Bodemassociatie	47 (ongunstige spreiding)	31% niet gedekt
Freatische grondwaterlichamen	22 (ongunstige spreiding)	22% niet gedekt
Ecoregiokaart	12	15% niet gedekt
Ecoregiokaart gegeneraliseerd (Ecodistricten krijt-leemgebieden en zuid-oostelijke heuvelzone samengevoegd)	9	9% niet gedekt

2. Binnen de **pilootzones**, door de auteurs gekozen zones die de diversiteit aan landschappen in Vlaanderen capteren e.g. Figuur 5, identificeerde men eigen strata, gebaseerd op volgende stappen:
    a. Geohydrologische hoofdeenheden worden onderscheiden op basis van oppervlakkige geologische formaties en breuklijnen volgens de geologische kaart;
    b. Deze hoofdeenheden worden onderverdeeld in bodemkundig-hydrologische eenheden op basis van de bodemkaart en gerelateerde drainageklassen, gegevens over het afwateringspatroon en de maaiveldhoogte;
    c. De strata worden gevormd, waarbij plaatselijk bodemkundig-hydrologische eenheden worden samengevoegd om versnippering en kunstmatige discontinuïteiten op stratumgrenzen te vermijden. In Nederlandse toepassingsstudies waren de strata 1 000 tot 5 000 ha groot. In Vlaanderen zijn geomorfologische eenheden gebruikt die in grote lijnen overeenkomen met de ecoregio's (Figuur 5), weliswaar toegepast binnen de pilootzones.



Figuur 5: Pilootzone bij Damme (Van Der Wauw, 2012)



Figuur 6: Ecoregio's in Vlaanderen (Bron: uit Van de Wauw, 2012; naar Sevenant *et al.*, 2002)

2.2.2.4 Kuststreek

Voor de kuststreek bestaat er nog geen (oude) drainageklasse van de bodemkaart volgens de algemene classificatie in Vlaanderen. Om deze streek toch mee op te nemen, wordt het unibodemtype (3.4.1.3) gebruikt.

3 DATAVERZAMELING EN -CORRECTIE

Volgende hoofdstukken beschrijven de datasets die mogelijk een meerwaarde kunnen bieden voor dit project. Een aantal gegevens zijn cruciaal – e.g. grondwaterpeilmetingen, neerslag e.d. – terwijl andere potentiële verklarende variabelen – zgn. hulpdata zoals landgebruik e.d. – zijn.

Volgende hoofdstukken geven een beknopt, thematisch overzicht van deze informatie. De beschrijving baseert zich hoofdzakelijk op wat er op de respectievelijke websites aanwezig is (vaak letterlijk), aangevuld met de publicaties, dataverkenning en input vanuit overlegmomenten met de betrokken instanties.

3.1 GRONDWATERPEILMETINGEN

3.1.1 Dataverzameling

Tabel 3 geeft een overzicht van de verzamelde databronnen met betrekking tot grondwaterpeilen. Voor een gedetailleerd overzicht wordt verwezen naar de data-inventaris in *Appendix 1: Inventaris*. De focus van de dataverzameling ligt op meetgegevens sinds 2010. De grootste bron is nog steeds Databank Ondergrond Vlaanderen (DOV). De tijdreeksen van de grondwaterstandindicator (GWI) hebben gemiddeld de hoogste resolutie aangezien het hier gaat over de geassimileerde tijdreeksen van gemeten en gemodelleerde data.

Ook bij Aquafin en OVAM zijn meetgegevens opgevraagd, maar deze zijn uiteindelijk niet verkregen; gesprekken voor overdracht van desbetreffende gegevens naar DOV zijn namelijk nog lopende tussen de bevoegde instanties.

Tabel 3: Diverse databronnen met aantal locaties en metingen per locatie.

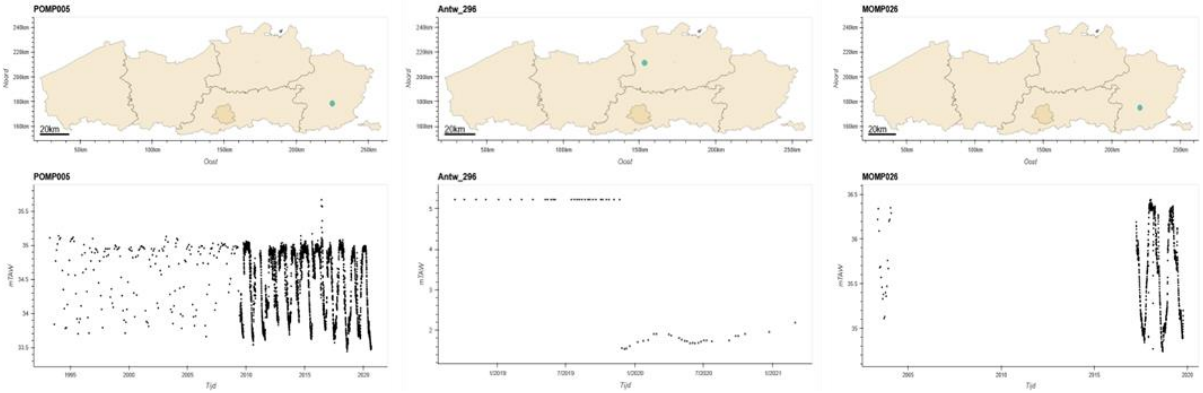
Bron	Aantal locaties	Gemiddeld aantal metingen per locatie
Stad Antwerpen	459	88
Provincie Antwerpen	20	188
DOV (diverse bronnen)	7346	339
GWI (gemeten + gesimuleerd)	154	10957
Stad Leuven	22	9
Maatschappij Linkerscheldeoever	10	74
Stad Mechelen	3	5
Provincie Oost-Vlaanderen	50	1445
Universiteit Antwerpen	25	151
Antea Group: peilbuizen met hoge meetfrequentie	125	193
Antea Group: peilbuizen met standaard meetfrequentie	8	77
Antea Group: ad hoc projectmetingen	1427	1

Om een databank aan te maken die eenvoudig raadpleegbaar is door verdere algoritmes, zijn verschillende bronnen naar een algemeen formaat omgezet. In eerste instantie werden de tijdseries

omgezet naar m-TAW en de x-y locatie geëxtraheerd (Lambert 72-coördinaten). Soms is de hoogte van het maaiveld beschikbaar (z) en wordt deze ook verder meegenomen. In een latere fase wordt met behulp van de z-waarde of het digitaal terreinmodel (DTM) op de locatie het grondwaterpeil omgezet naar meter t.o.v. maaiveld (m-MV, negatief getal).

3.1.2 Datakwaliteit

De datakwaliteit binnen de tijdreeksen is zeer variabel. Waargenomen artefacten zijn: tijdshiaten, sprongen, verkeerde metingen en uitschieters (zie Figuur 7).



Figuur 7: Voorbeelden van artefacten in grondwaterpeilreeksen: uitschieters (links), sprong (midden), hiaat (rechts).

3.1.3 Datacorrectie

De meetgegevens zijn niet vrij van artefacten. Meetfouten, menselijke fouten bij het invoeren (incl. schrijffouten), etc. kunnen de waarden significant verstoren. Hoewel de meeste algoritmes robuust zijn tegen kleine fouten, zien we dat dit minder het geval is bij grote fouten zoals bijvoorbeeld omzetting van meter maaiveld naar meter TAW of dubbele invoer (zogenaamde 'stotter') etc. De meetlocaties van de testset (incl. deze die onderdeel uitmaken van de grondwaterindicatoren), zijn allen manueel nagekeken. Echter, gezien de grote hoeveelheid aan peilbuizen in de testset, is het onmogelijk om deze allemaal manueel te controleren – ook met het oog op toekomstige herrekening is dit niet opportuun. Daarom werd er een automatische validatiestap opgesteld die eenvoudig de grootste problematische fouten wegwerkt:

1. De eerste stap was het omrekenen van peilmetingen ten opzichte van bovenkant peilbuis en/of TAW naar **de eenheid meter onder maaiveld** (m-MV). Bij het gros van de metingen, was dit al beschikbaar als afzonderlijke parameter. Echter, wanneer dit niet voorhanden was, werd de eenheid omgevormd door (a) de hoogte van de peilbuis boven het maaiveld in acht te nemen en/of (b) de hoogte uit het Digitaal Hoogtemodel Vlaanderen te halen wanneer dit niet voorhanden was in de metadata.
2. Er bestaan verschillende methodes om outliers te detecteren. Een veel voorkomende techniek is de zgn. ***Min-Max-corrector***, waarbij waarden uitgefilterd werden indien ze boven of onder een resp. boven- of ondergrens vielen. Metingen waarbij het grondwaterpeil onrealistisch hoog boven het maaiveld kwam, werden gefilterd: metingen die >20cm boven het maaiveld uitkomen, werden verwijderd en geïmputeerd naar 0m-MV. De beperkte tolerantie van 20cm diende toegevoegd te worden, gezien de onzekerheid rond de maaiveldhoogte nabij een aantal peilbuizen. Ook de punten waarbij de GHG dieper dan -3m-MV uitkwam, werden verwijderd, in overeenstemming met de opzet van deze opdracht. Tot slot, werden de grootste anomalieën verwijderd door de meetwaarden te vergelijken

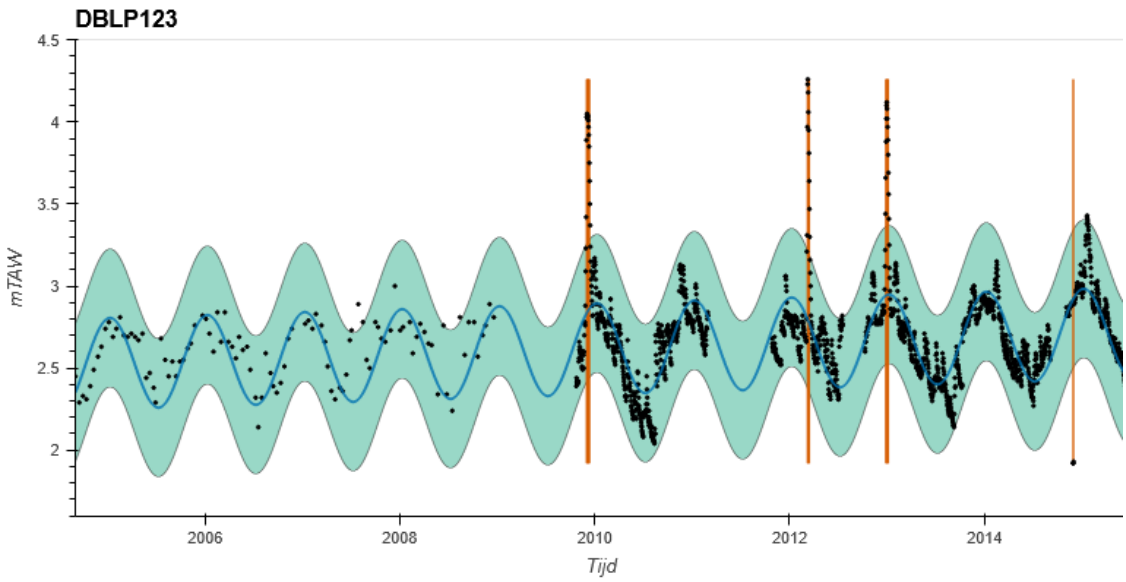
met een gefit, seizoenaal model gekoppeld aan een lineaire trend: een sinusoïde. De coëfficiënten worden bekomen met least-square optimalisatie.

$$GWP = A + B * \sin(2\pi * tj) + C * \cos(2\pi * tj) + D * t_j + \epsilon$$

Hierbij is/zijn:

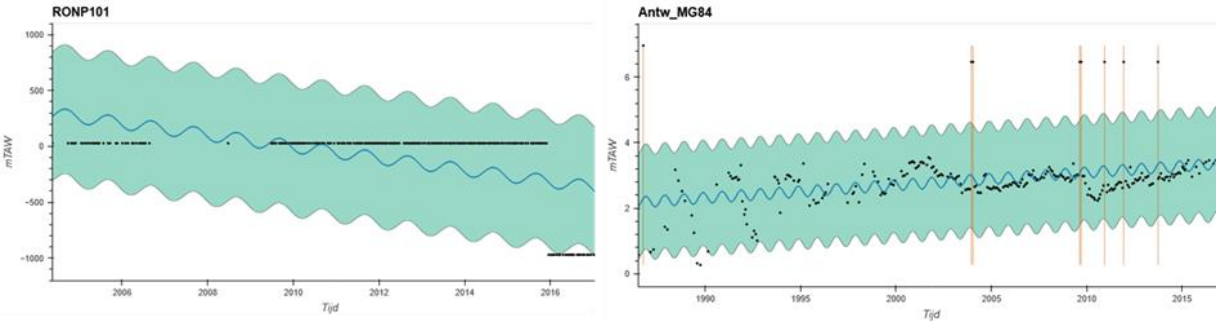
- GWP: het grondwaterpeil in m-TAW.
- A, B, C, en D: de geoptimaliseerde lineaire model coëfficiënten.
- t_j: de tijd uitgedrukt in jaren.
- ϵ: het residu.

Als de afwijking van een meting ten opzichte van de verwachte waarde (het residu) meer dan 3 maal de standaarddeviatie is, wordt (enkel) dit specifieke datapunt verwijderd (Figuur 8).



Figuur 8: Voorbeeld tijdreeks op locatie DBLP123 waarbij metingen verwijderd worden (rode verticale lijnen) omdat ze meer dan 3 maal de std afwijking van de verwachte waarde (blauwe lijn) bedragen. De groene band is 1.96 maal de std rond de verwachte waarde (95% van de data valt hierbinnen).

Verder, indien de standaarddeviatie van de residuen van een totale tijdreeks meer dan 0.5m is, dan wordt de gehele tijdreeks gemarkeerd als een anomalie. Deze wordt dan niet gebruikt bij de inschatting van GxG's.



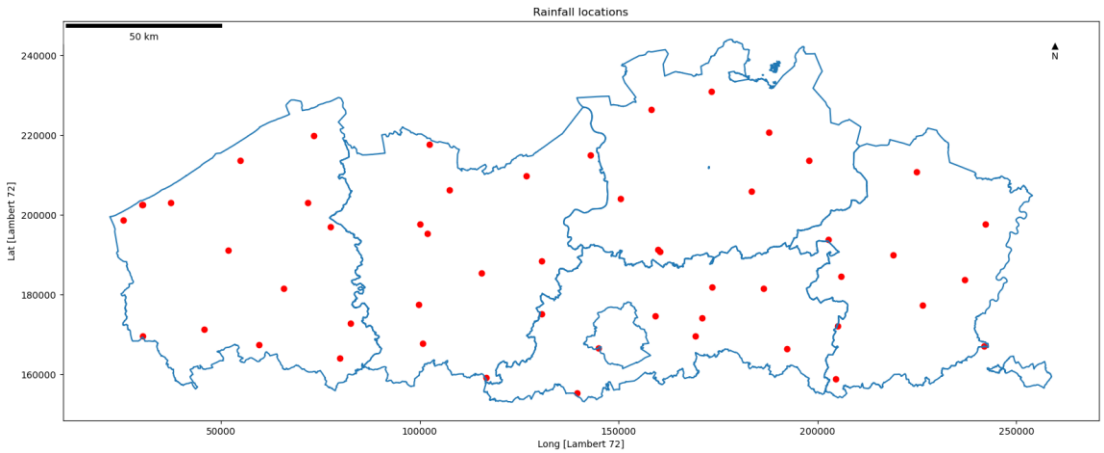
Figuur 9: Voorbeelden van GWP-tijdreeksen die gemarkeerd worden als een anomalie en niet beschouwd worden bij GxG inschattingen.

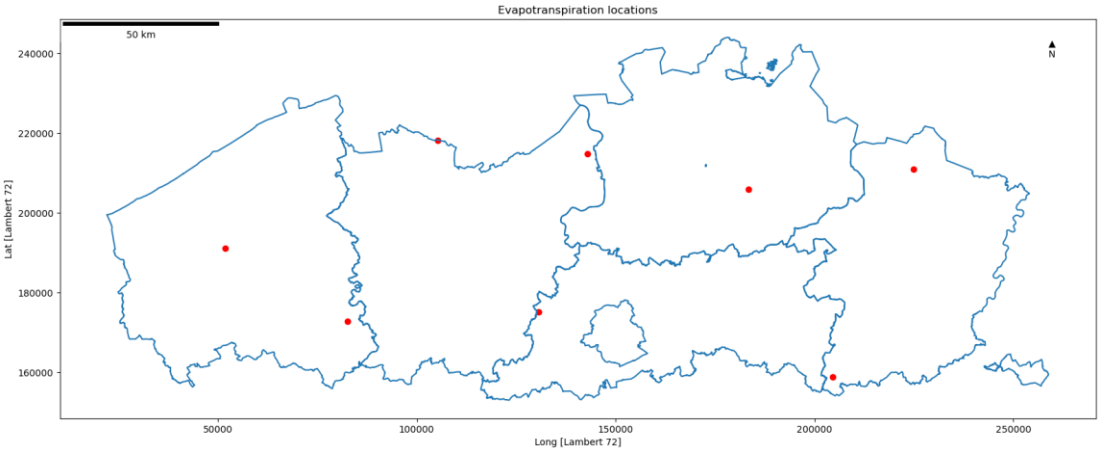
3.2 NEERSLAG & POTENTIËLE EVAPORATIE

3.2.1 Dataverzameling

Neerslag en evaporatie vormen de basis van de waterbalans in de ondergrond. De brondata voor deze studie werden gedownload vanaf waterinfo.be:

- Puntneerslag: in Vlaanderen staan er tientallen pluviometers opgesteld die registeren hoeveel neerslag er valt op een exacte locatie. Figuur 10 (Boven) situeert de 56 VMM-neerslagstations die aangewend zijn voor deze studie. Wij bevroegen de dagelijkse neerslagdata via waterinfo.be.
- Potentiële evaporatie: dit is de maximaal mogelijke waterafgifte door de bodem (verdamping) en de vegetatie (transpiratie) naar de atmosfeer, die optreedt indien er voldoende water beschikbaar zou zijn aan de oppervlakte of in de bodem. Als er minder water beschikbaar is in de bodem, zoals tijdens drogere periodes, is de werkelijke evaporatie lager dan de potentiële. Toch wordt meestal de potentiële opgevolgd in plaats van de werkelijke evaporatie, gezien de werkelijke sterk locatieafhankelijk en bijgevolg erg moeilijk te kwantificeren is. De potentiële evaporatie wordt berekend aan de hand van de Pennman-Monteith formule. Voor deze berekening worden de temperatuur op verschillende hoogtes boven de grond, de luchtdruk, de relatieve luchtvochtigheid, de windrichting- en snelheid, de instralingsflux door de zon, en de grondwarmteflux gemeten. Figuur 10 (Onder) situeert de acht locaties waarvan de potentiële evaporatie aangewend is voor deze studie.





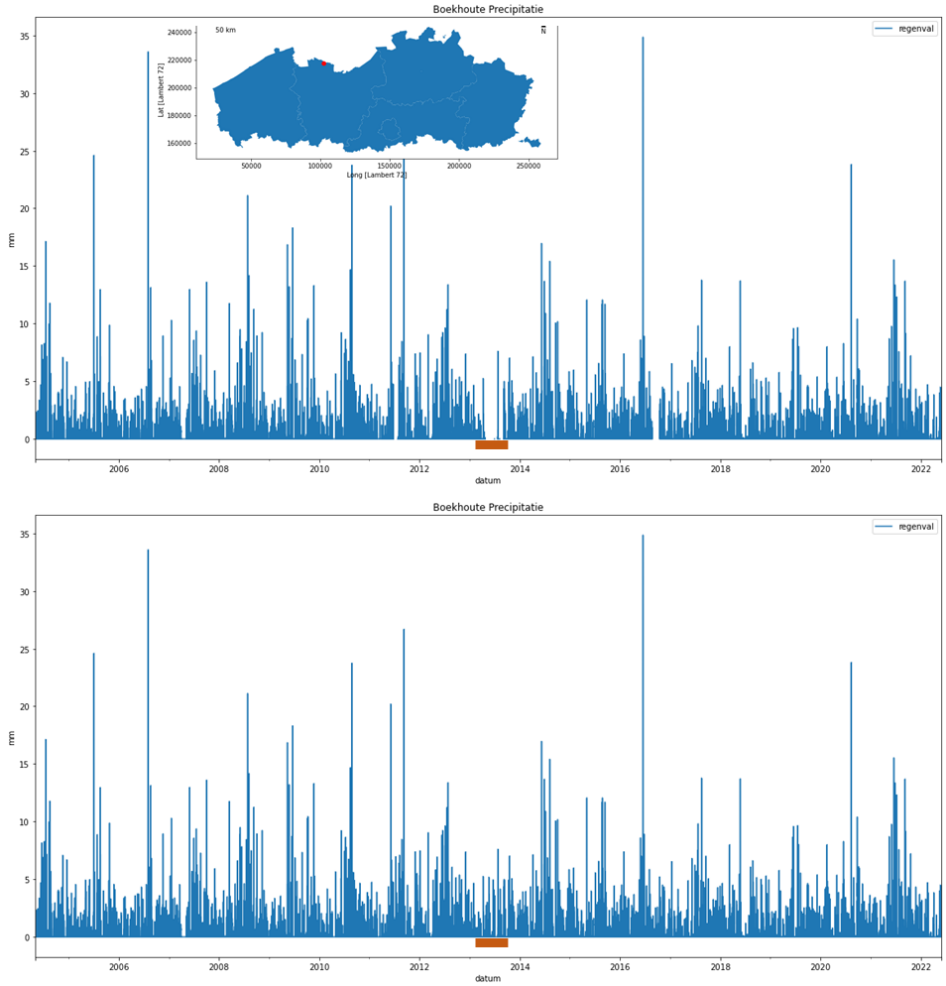
Figuur 10: Meetstations (rode locaties) waarvan neerslag (BOVEN) en potentiële evaporatie (ONDER) werden gedownload van waterinfo.be. Provincie contouren zijn toegevoegd.

3.2.2 Datacorrectie

De brondata van beide parameters bevatten ook kwaliteitscodes die zijn toegekend aan de metingen. Deze kwaliteitscodes worden toegekend na controle door de beheerinstanties van de meetstations (HIC en VMM). De meeste metingen hebben een kwaliteitscode 220 of 110: niet gecontroleerd. Een verder overzicht van de kwaliteitscodes wordt gegeven in *Appendix 2: Waterinfo kwaliteitscodes*. De meetwaarden met volgende kwaliteitscodes werden verwijderd: 70 & 170 (*Poor*) 90 & 190 (*Suspect*). Deze hiaten vormen echter een hindernis voor onze toepassing, die nood heeft aan continue invoerreeksen. Volgende data-imputatietechniek werd toegepast (Figuur 11):

1. De missende tijdstippen werden vervangen door de metingen van het dichtst bijgelegen station met beschikbare metingen. Een interpolatie met uitmiddeling van meerdere stations werd hier geweerd, gezien deze uitmiddeling surrealistische waarden kan veroorzaken m.b.t. neerslag of afgeleiden hiervan.
2. Ook na bovenstaande opvulling bleven er nog een aantal missende tijdstippen over. Deze laatste hiaten werden opgevuld met behulp van een lineaire interpolatie doorheen de tijd binnen hetzelfde meetstation.

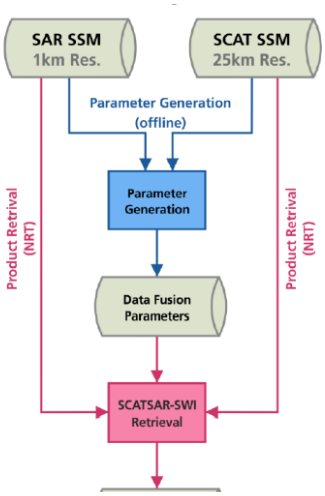
Na de data-imputatie zijn zowel regenval als potentiële evaporatie geaggregeerd tot op dag-niveau, gezien dit de resolutie is die we beogen met de toepassing. Tot slot zijn deze meteorologische stations rechtstreeks gelinkt aan de locaties met grondwaterpeilen. Hiervoor wordt louter de meting van het meest nabije station genomen, zonder enige interpolatie toe te passen.



Figuur 11: Precipitatiemetingen te Boekhoute zonder (BOVEN) en met (ONDER) imputatie van missende metingen.

3.3 REMOTE SENSED BODEMVOCHT

3.3.1 Dataverzameling



Figuur 12: Schematische weergave van hoe de SWI wordt bekomen uit SAR SSM en SCAT SSM

Het Soil Water Index product van Copernicus (Copernicus Global Land Service, 2022; hierna SWI) werd ontwikkeld om de behoeften vanuit de hydrologische sector te vervullen. De basis is een Scatterometer Surface Soil Moisture (SCAT SSM) product, gewonnen door de Advanced Scatterometer (ASCAT)-sensor, met een spatiale resolutie van 25 km en een temporele resolutie van 1 dag. Dit product laat toe de temporele dynamiek van het bodemvocht goed te beschrijven, maar ten koste van de spatiale resolutie. De grove resolutie van 40 km laat niet toe om hydrologische patronen op kleine schaal te detecteren. Voor deze behoefte werd een nieuw product ontwikkeld dat de voorgenoemde versie combineert met hoog-resolutie Synhetic Aperture Radar (SAR-data) van de in 2014 gelanceerde Sentinel-1 missie. De SAR-data hebben een aanmerkelijk hogere spatiale resolutie van 1 km, maar een lagere temporele resolutie van 1,5 tot 4 dagen. Door deze twee databronnen optimaal te

combineren, wordt een product bekomen met een spatiale resolutie van 1 km en een temporele resolutie van 1 dag. Een overzicht van deze eigenschappen wordt gegeven in Tabel 4.

Tabel 4: Voornaamste eigenschappen van de input data (ASCAT & Sentinel-1) en de gefuseerde output.

Product	Databron	Datalevel	Temporele resolutie	Spatiale resolutie
SCAT SSM/SWI	• Metop ASCAT	L2	1 dag	25 km
Sentinel-1 SSM	• Sentinel-1	L2	1,5 – 4 dagen	1 km
SCATSAR-SWI	• Metop ASCAT • Sentinel-1	L3	1 dag	1 km

Het algoritme bestaat uit twee componenten, zoals geïllustreerd in Figuur 12. Eerst worden de dataparameters bepaald voor beide inputbronnen. Allereerst worden zowel de SAR SSM als de SCAT SSM op hetzelfde raster geprojecteerd. Daarna worden een groot aantal basis statistische eigenschappen bepaald voor deze data. Deze eigenschappen, de dataparameters en de datafusie parameters zijn nodig om de twee dataproducten samen te voegen.

Na de samenvoeging wordt de SWI bepaald. De methode die gebruikt wordt om de SWI te bepalen houdt geen rekening met lokale fysische parameters. In plaats daarvan kijkt ze naar historische data en schaalt zo de uitkomst van het gefuseerde dataproduct naar een bodemvocht-index [e.g. Bauer-Marschallinger *et al.*, 2018a m.b.t. Sentinel data]. Verder maakt ze gebruik van een aantal zogenaamde "Kritische T-waarden", waar T staat voor de lengte van het integratievenster. Dit zijn integratievensters doorheen de tijd voor filtering van het SAR-SSM backscatter signaal. Deze zijn afgestemd, a.d.h.v. in-situ metingen, om een correlatie te bekomen met bodemvocht op bepaalde dieptes [Albergel *et al.*, 2008 ; de Lange *et al.*, 2008; Paulik *et al.*, 2014]. Echter, de auteurs merken op dat dit dus modelresultaten betreft die niet expliciet als een diepte meegenomen worden, en dat we dus voorzichtig dienen te zijn met de interpretatie ervan.

Daarnaast wordt ook de datakwaliteit (QFLAG) bepaald voor de acht T-waarden. De QFLAG reflecteert de input SSM-data densiteit voor elke SWI-waarde en is dus een maat voor de data-beschikbaarheid op een gegeven moment en locatie. De fout of betrouwbaarheid is, in de zin van een statistische kansverdeling, nog niet (online) beschikbaar op het dataportaal van Copernicus. Echter, toekomstige versies van het SWI-product zullen wel over een dergelijke betrouwbaarheidsmaat beschikken [CGLOPS-1, 2019]. Tabel 5 geeft een overzicht van de voornaamste eigenschappen van de SWI.

Tabel 5: Overzicht eigenschappen Copernicus SWI

Eigenschap	
Soil water index (SWI)	m^3/m^3 (DN)
QFLAG	Dit is de kwaliteitsvlag - geen eenheid
Spatiele resolutie	1km x 1km
Temporele resolutie	1 dag
Geografische projectie	WGS84 in Lat-Lon
Positie coördinaat	Centrum van de pixel
Beschikbaarheid	Vanaf 2015

3.4 DATABANK ONDERGROND VLAANDEREN

Volgende hoofdstukken geven een beknopt, thematisch overzicht van de gebruikte data van Databank Ondergrond Vlaanderen (DOV). In eerste instantie werd er enkel gebruik gemaakt van de gegevens van de bodemkaart en het hydrogeologisch model. In een latere fase werd er nagegaan wat de meerwaarde is van het toevoegen van overige databronnen.

3.4.1 Thema Bodem

Zowel observaties als kaartlagen worden aangeboden voor gebruik.

3.4.1.1 Observaties in de Bodemdatabank

Voor een gedetailleerde beschrijving van de DOV-bodemdatabank wordt verwezen naar Oorts *et al.* (2020). Onderstaande alinea's omvatten een samenvatting van de informatie op DOV en verdere mondelinge toelichting (overleg d.d. 02-04-2021).

Historische en recente Vlaamse bodemdata zijn opgeslagen in de DOV-bodemdatabank. Deze bodemdatabank bevat gegevens van zowel profielputten als bodemboringen en is opgesteld om zeer diverse bodemdata uit Vlaanderen in één databank te kunnen integreren waarbij het eigenaarschap van de data bij de bron blijft. De DOV-bodemdatabank laat toe om samen in een open, multidisciplinair en sectoroverschrijdend netwerk van interne en externe partners, Vlaamse bodemdata te centraliseren, te valideren, te verwerken en geïntegreerd te ontsluiten op maat van de gebruiker. De data uit de DOV-bodemdatabank zijn ontsloten via de kaartlaag 'Bodemlocaties'. De meeste puntmetingen zijn niet zo recent (meetgegevens ten tijde van het opstellen van de bodemkaart).

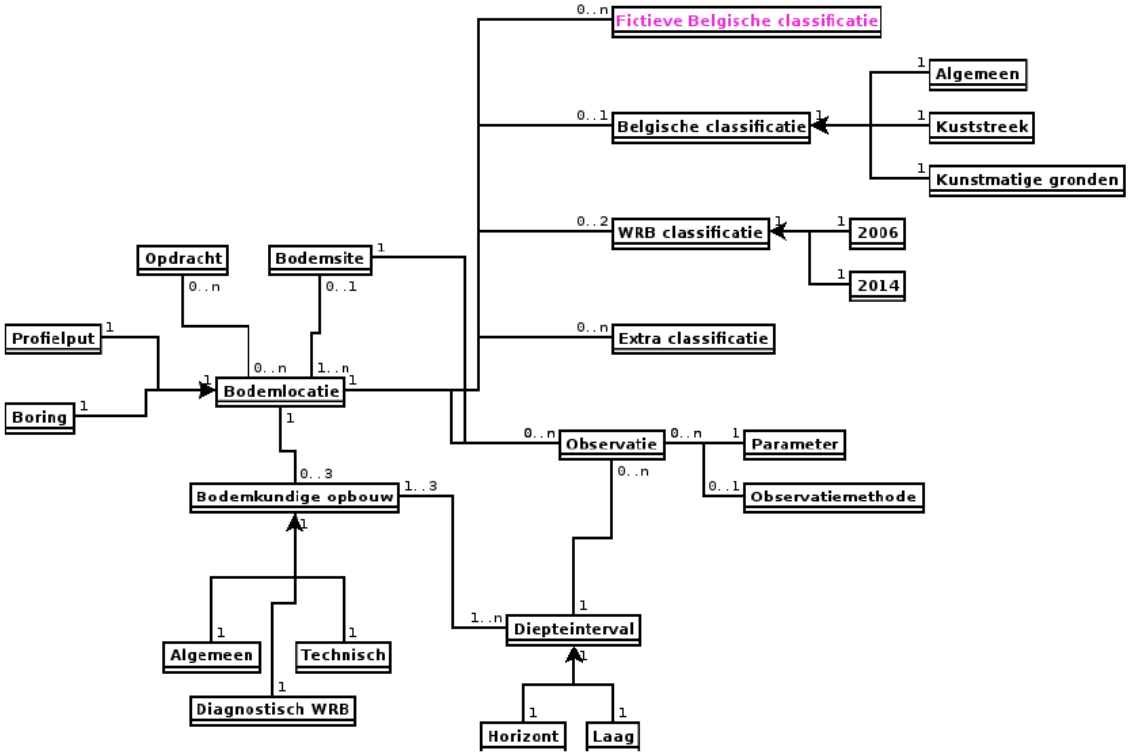
De algemene structuur is als volgt (zie ook Figuur 13):

- De kern van de DOV-bodemdatabank is de **bodemlocatie** met o.a. XY-coördinaten in Lambert 72, een z-waarde in mTAW, een datum en auteur(s). Er zijn 2 types bodemlocaties: profielputten en bodemboringen.
- Bodemlocaties kunnen gegroepeerd worden in **een bodemsite**. Dit is een vlak dat één of meer bodemlocaties groepeert en de nabije, homogene, omgeving ervan beschrijft.
- Een profielput of een boring wordt beschreven volgens **een bodemkundige opbouw**. Dit is een opeenvolging van **diepte-intervallen** die beschreven worden volgens een bepaald type: horizonten of lagen (staalnamelagen). In DOV zitten er nog niet veel lagen; de meeste data betreffen horizonten.
- De bodem op een bepaalde bodemlocatie kan geclassificeerd worden volgens verschillende classificatiesystemen. De belangrijkste is de **Belgische classificatie**. Daarnaast wordt ook de internationale **WRB-classificatie** aangegeven: 40k of 250k schaal, afgeleid van de Belgische Bodemkaart en bijgevolg geen toegevoegde informatie voor deze opdracht.
- Op een bodemlocatie kunnen in het kader van verder onderzoek of veldanalyses **bodemmonsters** genomen worden.
- Meer informatie over de bodem op een bepaalde locatie wordt vaak verkregen door **observaties**. Een observatie is de waarneming of meting van een bepaalde parameter op een

bepaalde locatie. Deze locatie kan ofwel een bodemsite, een bodemlocatie, een diepte-interval of een bodemmonster zijn. Niet alle parameters hebben numerieke meetwaarden; **sommige observaties zijn vrije tekstvelden of codetabellen**. Een observatie kan ook bestaan uit meerdere meetpunten en –waarden (een meetreeks).

- T**extuurmeting**: een meting van de bodemtextuur van een bepaald monster. Er kunnen op een bodemmonster meerdere textuurmetingen plaatsvinden met een bepaalde observatiemethode.
- Infiltratiemetingen: alle bodemobservaties met een Ksat-waarde. Dit zijn er wel maar een paar 100.

– Drainageklassen van de bodemkaart (en hydromorfe profielkenmerken) zitten ook in de bodemdatabank gerelateerd aan de bodemobservaties of in de horizontbeschrijving, maar dan als vrije tekst.



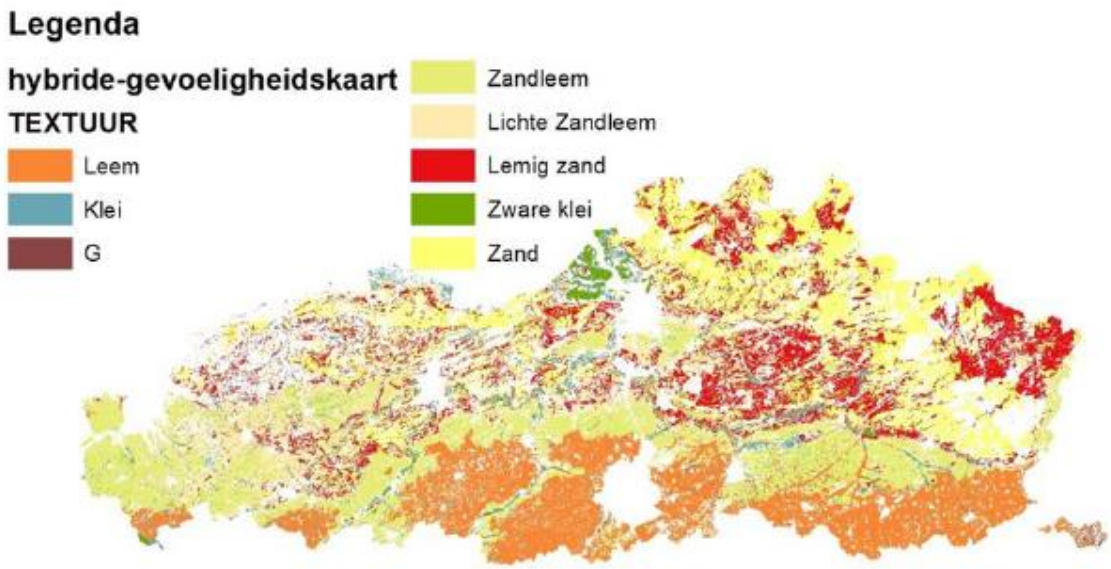
Figuur 13: DOV-bodemdatabank (Bron: https://www.milieuinfo.be/confluence/display/DDOV/DOV-bodemdatabank, geraadpleegd op 2021-04-07)

3.4.1.2 Digitale bodemkaart van het Vlaams Gewest

Op basis van de resultaten van een intensieve bodemkartering gedurende de jaren '50 tot '70 werd de Belgische bodemkaart opgesteld. Deze Belgische bodemkaart steunt op het Belgische bodemclassificatiesysteem. Het is een nationaal systeem dat uitsluitend voor de Belgische bodems werd opgesteld en bestaat uit een bodemtype dat overeenkomt met één van de drie classificatietypes: de algemene classificatie, en deze voor de kuststreek en de kunstmatige gronden. Het bodemtype van de algemene Belgische bodemclassificatie wordt opgebouwd uit verschillende onderdelen zoals een substraat, textuurklasse, drainageklasse, profielontwikkelingsgroep, fasen en varianten.

Voor de kuststreek is een apart classificatiesysteem toegepast waar er afgeweken wordt van de standaard bodemseries. Door middel van het Unibodemtype kunnen hieruit toch standaard textuur- en drainageklassen gehaald worden.

Ook voor bodemprofielen met een sterke menselijke verstoring, kunstmatige gronden, bestaat er binnen België een apart classificatiesysteem. Deze dataset is omwille van de visualisatie onderverdeeld in 5 kaartlagen: bodemtypes, substraten, fasen, varianten van het moedermateriaal en de profielontwikkeling. De kaartlaag 'bodemtypes' bevat alle attribuutinfo van de bodemkaart. Figuur 14 presenteert de ruimtelijke verdeling van textuurklassen in Vlaanderen.



Figuur 14: Ruimtelijke verdeling textuurklassen in Vlaanderen. G=stenige gronden (Brondata: DOV)

3.4.1.3 Bodemerodibiliteit (K-factor)

De bodemerosiegevoeligheidsfactor of bodemerodibiliteitsfactor K (ton.h.MJ^{-1}.mm^{-1}) geeft de mate van intrinsieke erosiegevoeligheid van de bodem weer. De K-factor wordt gebruikt bij de modellering van bodemerosie in de Revised Universal Soil Loss Equation (RUSLE) en wordt bepaald door verschillende bodemeigenschappen, waarvan de bodemtextuur en het gehalte aan organisch materiaal de belangrijkste zijn. Bij het opstellen van de bodemerosiegevoeligheidskaart of bodemerodibiliteitskaart (K-kaart) is enkel gebruik gemaakt van de bodemtextuur om de K-factor te bepalen. Er werd geen rekening gehouden met het gehalte organisch materiaal, omdat Vlaanderen niet over gebiedsdekkende data van het gehalte aan organische materiaal beschikt.

De bodemerosiegevoeligheidsfactor K (ton.h.MJ-1.mm-1) werd berekend aan de hand van de formule van Declercq & Poesen (1991), toegepast op de textuurklassen van de digitale bodemkaart van Vlaanderen. Hierbij werd het unibodemtype gebruikt, waarbij ook de zeepolders vertaald werden naar een bodemserie met textuurklasse volgens de Belgische bodemclassificatie. Waar er missing data waren (bv. de kunstmatige gronden) werden die ingevuld door interpolatie met een *nearest neighbour* algoritme. Aan de hand van een theoretische waarde voor de geometrisch gemiddelde korrelgrootte (mm) werd de K-factor voor elke textuurklasse berekend (Oorts *et al.*,

2019). Er dient opgemerkt dat deze berekening van de K-factor gebeurt aan de hand van een theoretische gemiddelde korrelgrootte per textuurklasse, zonder rekening te houden met de invloed van het gehalte aan organisch materiaal. Bovendien is de digitale bodemkaart zelf niet foutenvrij (foutieve kartering, wijziging van de textuur sinds de kartering door o.a. erosieprocessen of reliëfwijzigingen). Het is daarom mogelijk dat de werkelijke K-factor van een bodem verschilt van de berekende K-factor. Voor een uitgebreide discussie wordt verwezen naar Notebaert *et al.* (2006).

De ruimtelijke resolutie bedraagt 5x5 m. Deze brondata zijn afkomstig van DOV (beheerder: Vlaamse overheid, Departement Omgeving, Vlaams Planbureau voor Omgeving. Geraadpleegd op 20/09/2023, op https://www.dov.vlaanderen.be).

3.4.1.4 Gemodelleerde bodemeigenschappen (Bodemdatabank-STAT project)

Momenteel voert de KUL departement Aard- en Omgevingswetenschappen het project Bodemdatabank-STAT uit met als doel: het creëren van gebiedsdekkende rasterlagen en de koppeling van statistische bodemgegevens aan de polygonen van de bodemkaart gebaseerd op de bodemdata van de DOV-bodemdatabank. De Digital Soil Mapping zal gebiedsdekkende rasterlagen op 5 dieptelagen opleveren (0-10, 10-30, 30-60, 60-100 en 100-150 cm-MV) en met spatiale resolutie van 10x10m waarbij het raster overeenkomt met dat van de landgebruikskaart/landgebruiksbestand (Informatie Vlaanderen, 2013). Er komen rasterlagen ter beschikking voor de volgende gemodelleerde bodemeigenschappen:

- textuurklassen (klei-, zand- en leemfractie);
- gehalte aan organische koolstof;
- $CaCO_3$;
- pH;
- sorptie.

Deze opdracht is momenteel nog lopende waardoor deze niet aangewend kon worden als input voor deze voorspelling. Eens beschikbaar kan ze een aanzienlijke verbetering vormen t.o.v. de textuurklassen in de bodemkaart: een actuele analyse, hoge resolutie kartering, Vlaanderen dekkende kaart incl. verstedelijkte gebieden.

3.4.2 Thema Geologie

3.4.2.1 Geologisch 3D-model van Vlaanderen

Het Geologisch 3D-lagenmodel van Vlaanderen wordt opgemaakt door het Vlaams Kenniscentrum Ondergrond (VLAKO) bij VITO in opdracht van ALBON van het Departement LNE (tot 2016) / in opdracht van VPO van het Departement Omgeving (vanaf 2017). Zij voeren verscheidene geologische karterings- en modelleringsprojecten uit. Bij de aanmaak van de verschillende lagen werd steeds vertrokken van zoveel mogelijk bestaande (digitale) data zoals puntgegevens, kaartmateriaal, rasters, seismische data e.d. De puntgegevens uit DOV, gegevens van de Belgische Geologische Dienst (BGD), vroegere geologische karteringen, HCOV-rasters (Hydrogeologische Codering van de Ondergrond van Vlaanderen) e.d. vormen dus de basis van deze geologische 3D-modellen.

In 2013 werd de 2de versie opgesteld (G3Dv2), terwijl de derde versie (G3Dv3) gepubliceerd werd in 2019. Het Geologisch 3D-lagenmodel van Vlaanderen versie 3, G3Dv3, werd afgerond in 2019. In 2021 werd een verbeterde versie, G3Dv3.1, gepubliceerd in de DOV-verkenner. G3Dv3.1 is een

louter technische update van G3Dv3, waarbij enkele kleine technische fouten uit de oorspronkelijke publicatie verbeterd werden. Er werden geen inhoudelijke wijzigingen doorgevoerd bij de publicatie van G3Dv3.1. Het model werd voor alle lagen 'sluitend' gemaakt. Er werden controles ingebouwd zodat de lagen niet snijden of het model geen 'gaten' bevat. In de DOV-verkenner worden de voorkomenskaarten van al de eenheden, isohypsen en isopachen (vectorlagen) ter beschikking gesteld. Daarnaast worden er per gedefinieerd laagvlak ook een XYZ-tabel, breukenkaart, boringen, metadata en bijhorende rapporten ter beschikking gesteld. Ook de rasterlagen van 100*100 m die de basis, en in sommige gevallen ook de top, voorstellen van de verschillende geologische eenheden kunnen worden gedownload.

Een geologisch model is een versimpelde voorstelling van de geologische realiteit. Nuttige data werden geselecteerd en geïnterpreteerd om op die manier een consistente basis voor elke eenheid te kunnen creëren, gebruik makend van geschikte geologische concepten. De mate van nauwkeurigheid en betrouwbaarheid is verschillend voor elke laag van het model en zelfs voor elke rastercel van een gemodelleerde laag. Verscheidene zaken spelen een rol wanneer de nauwkeurigheid van een model moet ingeschat worden. De voornaamste zijn: de juistheid van een geologische interpretatie van een boring, de geografische nauwkeurigheid van de ligging van de informatiepunten, de afstand tot de dichtste boring met interpretatie van de hoogte van een bepaalde basis, de nauwkeurigheid van de gebruikte seismische gegevens en interpretaties, de datadensiteit, de modelleertechniek en de gebruikte softwarepakketten, inputdata en kaarten, de helling van de topografie en de eenheden zelf en de expertise van het modelleerteam.

De medegedeelde gegevens moeten steeds worden bevestigd door verder onderzoek. De gebruiker dient in te staan voor dit verder onderzoek waarbij hij rekening moet houden met het specifieke aspect van zijn project, studie of interesse.

3.4.2.2 Plastische gronden

DOV stelt thematisch kaartmateriaal beschikbaar in het themaloket 'Plastische Gronden' op basis van extracties uit de Digitale Bodemkaart van Vlaanderen, en op basis van het Geologisch 3D Model van Vlaanderen (G3Dv3 - Deckers *et al.*, 2019a) en het Hydrogeologisch 3D Model van Vlaanderen (H3Dv2 - Deckers *et al.*, 2019b). De beschikbare kaarten zijn dus afgeleid van kaarten en modellen, gebaseerd op waarnemingen, proeven, interpretaties en interpolaties. Deze kaarten geven het potentieel voorkomen van eenheden met klei, zand of leem als hoofdbestanddeel, samen met dikte-informatie, binnen de 5m onder de oppervlakte weer. Ze zijn niet gemaakt om een uitspraak te doen over de ligging van de freatische grondwatertafel.

3.4.2.3 Hydrogeologische kartering

De Hydrogeologische codering van de Ondergrond van Vlaanderen (**HCOV)** is een zuiver **hydrogeologische codering** die hiërarchisch is opgebouwd voor drie niveaus van detail, namelijk de hydrogeologische hoofd-, sub- en basiseenheden. Naargelang de noodzaak en de vereisten kan de HCOV meer in detail gebruikt worden. Het hiërarchische karakter komt de herkenbaarheid sterk ten goede, zodat men al aan het eerste cijfer herkent tot welke hoofdeenheid een bepaalde laag behoort.

Het eindproduct van de **hydrogeologische kartering** bestaat uit het voorkomen van de HCOV-eenheden en de verticale afbakening van de basis en de dikte van de voornaamste HCOV-eenheden in rasterformaat met een resolutie van 100m. Voor iedere HCOV-hoofdeenheid is de basis van de

desbetreffende laag gekarteerd. De dikte van een laag is te berekenen als het verschil tussen de basis van de betreffende laag en de basis(sen) van de direct bovenliggende la(a)g(en). De oorspronkelijke kartering is uitgevoerd in deelgebieden door Ecolas, Haskoning & Haecon[1] (2002-2003). De deelgebieden werden achteraf samengevoegd door de Vakgroep Hydrologie en Waterbouwkunde van de VUB (2005). De eerste herkartering is uitgevoerd door de Belgische Geologische Dienst (2006). In 2019 werd het Geologisch 3D-model (G3Dv3) en een nieuw **Hydrogeologisch 3D-model** (**H3Dv2**) gepubliceerd binnen de VLAKO-referentietaak.

Aan de verstrekte gegevens mag geen absolute nauwkeurigheid worden toegekend omwille van interpolaties en extrapolaties welke bij het interpreteren en opstellen ervan werden gemaakt. Ook was het niet steeds mogelijk voor de minder recente gegevens de coördinaten en hoogtepeilen nauwkeurig te bepalen. Er dient rekening gehouden te worden met het feit dat niet alle HCOV-eenheden gekarteerd zijn en verder zijn de gekarteerde HCOV-eenheden niet steeds gebiedsdekkend gekarteerd. Als een bepaald raster voor een bepaalde HCOV-eenheid op een bepaalde positie geen waarde vertoont, betekent dat dus niet noodzakelijk dat die eenheid daar niet voorkomt. Verder kan het lokaal voorkomen dat de hiërarchie van HCOV niet gerespecteerd werd (bv. subeenheid is uitgebreider gekarteerd dan de samenstellende basiseenheden en *vice versa*).

3.4.3 Thema Grondwater

3.4.3.1 Grondwatermeetnetten

In de Databank Ondergrond Vlaanderen zijn verschillende grondwatermeetnetten opgenomen. Deze meetnetten staan in functie van uitgebreide monitoringprogramma's met de bedoeling een goed beeld te krijgen van de beschikbare grondwaterkwantiteit en grondwaterkwaliteit van de watervoerende lagen in Vlaanderen. Hieronder een beknopt overzicht van de meetnetten:

- Meetnet 0: oorsprong/beheerder onbekend;
- Meetnet 1: meetputten VMM-afdeling Operationeel Waterbeheer - **primair meetnet**, vooral diepere meetputten, gebruikt voor kwantiteitsbeheer, kwalitatieve toestandsmonitoring en operationele monitoring, de meetfrequentie is maandelijks voor peilmetingen en jaarlijks voor kwaliteitsmetingen;
- Meetnet 2: meetputten VMM-afdeling Operationeel Waterbeheer - vooral diepere meetputten (initieel onderdeel van het primair meetnet) met onzekere kwaliteit van meetgegevens en beschikbare putinfo - zijn indicatief maar worden niet weerhouden in het kader van rapporteringsverplichtingen;
- Meetnet 3: meetputten VMM-afdeling Operationeel Waterbeheer - gebruikt voor tijdelijke projecten;
- Meetnet 4: meetputten van andere Vlaamse en Belgische overheden of instanties;
- Meetnet 5: meetputten van drinkwatermaatschappijen;
- Meetnet 6: meetputten van privébedrijven;
- Meetnet 7: grondwaterwinningsputten (pompputten);
- Meetnet 8: meetputten VMM-afdeling Operationeel Waterbeheer – **freatisch meetnet**, vooral ondiepe meetputten in de eerste watervoerende laag en hoofdzakelijk in landbouwgebied, maar ook in natuurgebieden (nummers beginnen met N/), gebruikt voor kwalitatieve

[1] Nu Antea Belgium nv.

toestandsmonitoring en operationele monitoring, de meetfrequentie is twee keer per jaar, voor peilmetingen en kwaliteitsmetingen;

- Meetnet 9: peilputten uit de **WATINA-databank** van het Instituut voor Natuur- en Bosonderzoek (INBO). WATINA (WATer IN NAtuur) is een databank waarin INBO zoveel mogelijk waterpeil- en hydrochemische gegevens van Vlaamse natuurgebieden tracht te verzamelen. Het kan gaan om peilputten in eigen beheer, maar ook peilputten afkomstig van andere overheidsinstanties (ANB, VLM, Provincies, …), natuurverenigingen (Natuurpunt, vzw Durme, …) en/of studiebureaus. Het gaat vrijwel altijd om ondiepe freatische peilputten met een beperkte diameter (meestal 5 cm), doorgaans met een filter van een halve meter die niet meer dan 5 meter diep zit. De peilputten in WATINA bestaan hoofdzakelijk uit piëzometers (onderste 30-50 cm geperforeerd onder het maaiveld) en een enkele keer uit grondwaterstandsbuizen (volledig geperforeerd onder het maaiveld). Dit type peilputten wordt in natuurgebieden vaak geplaatst om de ondiepe grondwaterstand op te volgen die direct van belang is voor de optimale ontwikkeling van grondwaterafhankelijke vegetatietypen.
- Meetnet 10: putten die worden aangelegd in de VLAREM-rubriek 55 (Verticale boringen ten behoeve van de aanleg van peilputten en voor andere doeleinden, andere dan deze bedoeld in de rubrieken 53, 54 en 55.2; diepteboringen)

Alle meetputten van meetnet 1 (primair meetnet), meetnet 8 (freatisch meetnet) en recent ook de WATINA-databank zijn op dit ogenblik publiek aanwezig in de databank. De niet-publieke metingen zijn opgedeeld in 2 secties:

- Niet standaard beschikbaar op DOV-verkenner, maar wel opvraagbaar (in kader van het integraal milieujaarverslag). Dit bevat ook tijdsreeksen met voldoende hoge resolutie voor ons doeleinde: e.g. metingen van drinkwatermaatschappijen.
- Niet opvraagbaar.

Bijkomende info over de meetpunten:

1. Een meetpunt kan meerdere filters hebben. Een grondwaterpunt is gerelateerd aan een boring met soms een bijhorende boorbeschrijving: niet altijd beschikbaar bij bedrijfsputten.
2. Alle info komt van de fiche zelf:
   a. Aquifer waarin de filter zich bevindt. Merk op: er bestaat ‘aquifer’ en ‘aquifercontrole’. Een bedrijf kan een vergunning gekregen hebben voor een bepaalde watervoerende laag, maar de filter zit soms verkeerd geplaatst. Als deze oud is, dient men deze niet per se opnieuw te installeren.
   b. Regime: e.g. Freatisch, maar dit kan op veel verschillende manieren ingevuld worden. DOV wenst dit mettertijd (lopend project) te vervangen door ‘afgesloten’ of niet op basis van gerelateerd grondwaterlichaam (HCOV-laag).
   c. Filteropbouw: e.g. stijgbuis, filter, zandvang etc.
   d. Referentiepunt: ten opzichte van wat wordt de peilmeting ingevuld? Maaiveld, bovenkant peilbuis etc. Zo wordt het waterpeil (intern) automatisch omgerekend naar GWP t.o.v. TAW.
   e. Betrouwbaarheid: goed, onbekend of twijfelachtig. Merk op: deze staat standaard op goed, maar indien er een bemaling in de buurt plaatsvond, dan staat het op twijfelachtig; of als er een spike is, dan is de meting ook twijfelachtig.
   f. Soms ook een waarneming i.p.v. daadwerkelijke peilmeting: e.g. de put stond droog of stond onder water, waardoor je geen peilmeting kon uitvoeren.

g. Uit de boringen kan men ook een aangenomen waterdiepte aangeven. Deze zijn minder betrouwbaar en minder van toepassing voor deze opdracht.

3.4.3.2 Waterwingebieden en beschermingszones

Deze laag bakent de waterwingebieden en beschermingszones af. Het Decreet van 24 januari 1984 houdende maatregelen over het grondwaterbeheer voorziet namelijk in de mogelijkheid tot afbakening van waterwingebieden en beschermingszones en tot het uitwerken van een specifieke regeling voor de grondwaterbescherming in deze gebieden. De te volgen procedure om de afbakening te realiseren, werd vastgelegd bij BVR van 27/03/1985 houdende reglementering en vergunning voor het gebruik van grondwater en de afbakening van waterwingebieden en beschermingszones. De individuele drinkwatermaatschappijen dienen het initiatief te nemen tot het opstarten van de procedure om tot afbakening van de zones te komen:

- Een waterwingebied heeft slechts betrekking op de installaties en een bufferzone van maximum 20 meter.
- De beschermingszones type I, II en III zijn veel ruimer:
  - Type I komt overeen met de 24-uurs grens;
  - Type II geldt als 'bacteriologische' zone en wordt begrensd door de 60-dagen-lijn met een maximum van 150 m voor artesische lagen en van 300 m voor de andere lagen;
  - Type III geldt als 'chemische' zone en komt overeen met het voedingsgebied van de waterwinning, evenwel beperkt tot een lijn gelegen op 2000 m van de grens van het waterwingebied in het geval van freatische waterwinningen. Voor artesische winningen wordt geen zone III afgebakend.

3.4.3.3 Grondwatervergunningen & Heffingendatabank

De gegevens betreffende de verleende **grondwatervergunningen** worden aan de overheid gerapporteerd. Zowel de vergunningen die nu actief zijn alsook de vervallen vergunningen kunnen gevisualiseerd en bevraagd worden via de puntenlaag 'Grondwatervergunningen (huidige) ' en ‘Grondwatervergunningen (historische en huidige)’.

De gegevens werden jaren terug reeds opgevraagd en verwerkt via een formulier op maat van de bedrijven. In 2005 werd deze geïntegreerd in het Integraal Milieujaarverslag (**IMJV**), samen met rapporteringen voor andere milieucompartimenten. De inhoud van het IMJV is afhankelijk van o.a. de milieuvergunningstoestand van de IMJV-plichtige exploitatie. Wegens privacyregels is deze info niet voor iedereen consulteerbaar op de Verkenner van DOV. De gegevens vormen echter wel een onmisbare gegevensbron voor de beoordeling van de impact van waterwinningen, zowel op lokale als op regionale schaal. Verwerving van deze gegevens dient door de bedrijven in samenwerking met de erkende laboratoria met de grootst mogelijke zorg, wetenschappelijk verantwoord en conform de wetgeving te gebeuren (VLAREM-rubriek 53).

Het **heffingenbeleid** is een toevoeging aan het vergunningenbeleid en moet zowel sturend als financierend zijn. Een verscherpt heffingenbeleid moet de sectoren aanzetten tot het investeren in een duurzaam watergebruik, het aanwenden van alternatieve bronnen en het investeren in Best Beschikbare Technieken (BBT) en waterbesparende technieken. De inkomsten uit het heffingenbeleid worden samengebracht in het MINA-fonds. Dit fonds financiert de uitgaven, ongeacht de aard, tot al wat dienen kan met betrekking tot het beleid van het Vlaamse Gewest over de preventie, de bescherming, de administratie, het beheer en de sanering van het leefmilieu.

3.5 DIGITAAL HOOGTEMODEL VLAANDEREN

Voor deze studie wordt gebruik gemaakt van het Digitaal Hoogtemodel Vlaanderen (DHMV) II, DTM, raster, 1 m. Het Digitaal Hoogtemodel Vlaanderen is de verzamelnaam voor alle gebiedsdekkende hoogtegegevens van Vlaanderen waarover Digitaal Vlaanderen beschikt. Het DHMV II werd ingewonnen in de periode 2013-2015. Deze datasetgroep omvat alle producten die als gratis open data aan een breed publiek aangeboden worden. Binnen deze producten wordt een onderscheid gemaakt tussen een digitaal terreinmodel (DTM) en een digitaal oppervlakte model (DSM) die in verschillende resoluties beschikbaar zijn. Deze studie maakt gebruik van het DTM dat werd afgeleid van LiDAR-hoogtegegevens die in het kader van het Digitaal Hoogtemodel Vlaanderen II (DHMV II) werden ingewonnen. Dit product werd aangevuld met een object 'vliegdagcontour DHMV' dat voor ieder deel van het DTM de precieze opnamedatum weergeeft.

3.6 WATERLOPEN

Er bleken meerdere datasets beschikbaar te zijn voor de waterlopen binnen Vlaanderen. Er is besloten om voort te bouwen (zie §3.6.5) op de voor deze opdracht beschikbare datalagen: de Vlaamse Hydrografische Atlas (VHA).

3.6.1 Vlaamse Hydrografische Atlas (VHA)

De VHA is een vectoren bestand met de assen van bevaarbare waterlopen (waterwegen), onbevaarbare-geklasseerde en een aantal niet-geklasseerde waterlopen. Dit bestand wordt bijgehouden in een samenwerkingsverband van de Vlaamse provincies en Vlaamse gewestelijke instanties, waarbij de Afdeling Operationeel Waterbeheer van de VMM als centrale beheerder optreedt.

3.6.2 Grachtenlaag Universiteit Antwerpen

Bij de opmaak van de ‘Watersysteemkaart’ heeft ook de Universiteit Antwerpen een grachtenlaag ontwikkeld. Gezien de benodigde effort om de grachtenlaag op te stellen, is deze niet vrij beschikbaar voor externe studies.

3.6.3 Project Agentschap Informatie Vlaanderen

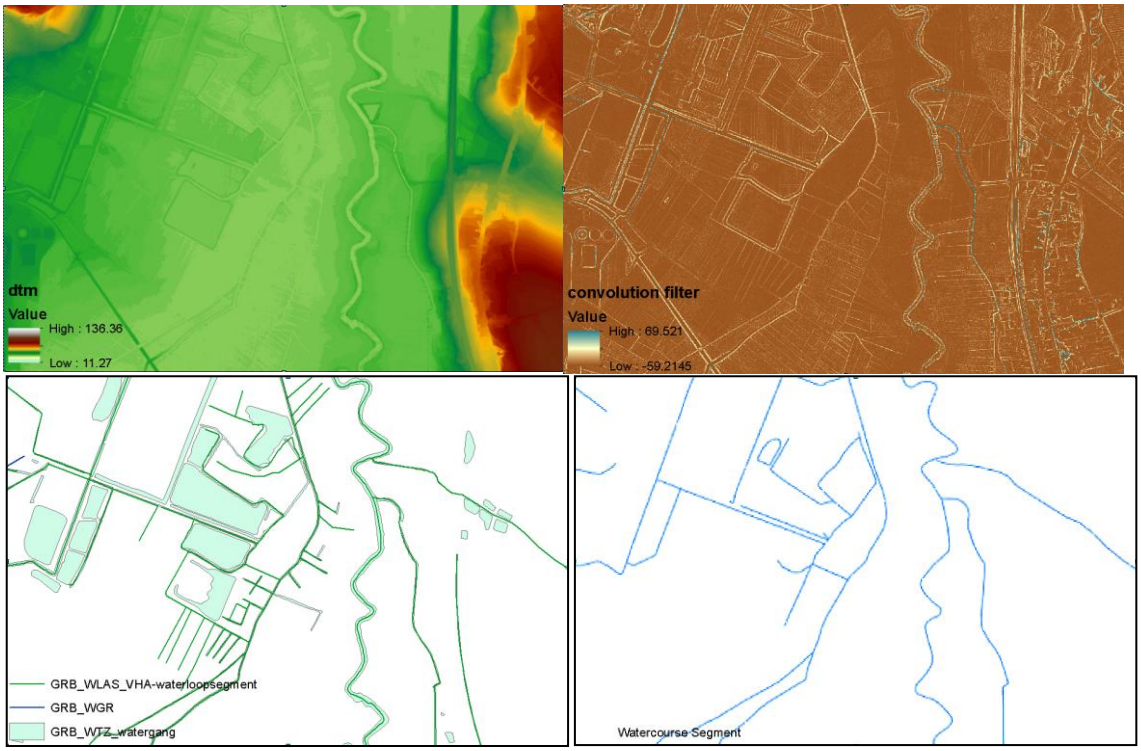
Er is in opdracht van de VMM een project lopende bij Digitaal Vlaanderen waarbij men een Vlaanderen dekkende grachtenkaart wenst op te stellen via A.I. op LIDAR-beelden. Momenteel is men deze nog aan het valideren.

3.6.4 Eerder intern onderzoek Antea

Er was in 2020-2021 een intern onderzoek lopende binnen Antea Group naar grachtendetectie in Vlaanderen. Meerdere technieken zijn uitgetest om te zoeken naar pixels die lager liggen dan de omgeving. Onder andere *Difference From Mean Elevation* (DFME) en *Relative Elevation Attribute* (REA). Dit omwille van de gunstige computationele kost van deze technieken. Verder zijn er verschillende convolutiefilters uitgetest. Hierbij wordt elke pixel omgezet naar een specifiek gewogen gemiddelde van zijn omgeving (kernel). Op die manier wordt een beeld gefilterd op basis van spatiale kenmerken.

Ter vergelijking toont Figuur 15 ook de waterlopen en grachten van respectievelijk het GRB en de top 10 vectorkaart. Uit de vergelijking blijkt duidelijk dat er op basis van de convolutiefilter een

verdere verfijning van de bestaande grachtendatasets mogelijk is. Toch zijn er nog een aantal artefacten die men moet wegwerken alvorens deze toepasselijk is in onze toepassingsgebieden: geconnecteerde pixels, geen stroomrichting etc. Dit was niet eenvoudig op te lossen.



Figuur 15 – Vergelijking van de methode voor grachtendetectie voor een testgebied in de Dijle met: de bron, het 1x1m DTM (LINKSBOVEN); de geïdentificeerde grachten (RECHTSBOVEN); en ter vergelijking zowel het GRB (LINKSONDER); als de waterloopsegmenten van het top 10 vectorbestand van het Nationaal Geografisch Instituut (RECHTSONDER).

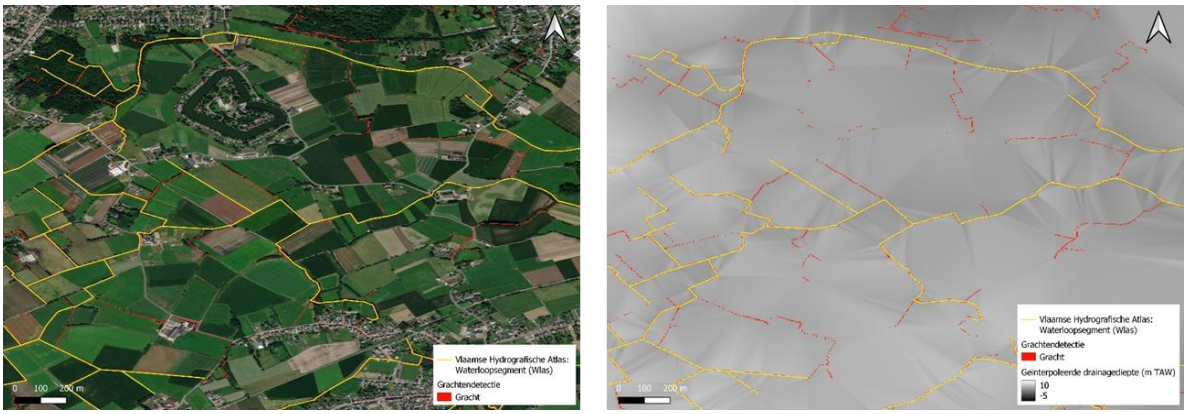
3.6.5 Open Street Map (OSM)

De watermassa's van Open Street Map werden gedownload (Geraadpleegd 23/03/2023).

3.6.6 Datacorrectie: uiteindelijke grachtendetectie

Het waterlopennetwerk van de Vlaamse Hydrografische Atlas (VHA) omvat de aslijnen van alle waterlopen die door openbare instanties beheerd worden, gelegen binnen het Vlaams Gewest. Omdat dit waterlopennetwerk niet alle drainerende grachten omvat, wordt een bijkomende grachtendetectie uitgevoerd op basis van het digitale hoogte model (DHM). Een operator (python-package PCRaster) creëert een lokale afvoerrichtingskaart met behulp van de stroomrichtingen van elke cel naar zijn steilste benedenbuur. Deze gekoppelde richtingen resulteren in een lokaal drainagenetwerk: het stromingspatroon op de kaart. Op basis van dit resultaat wordt aan elke cel een stroomorde toegekend. Deze stroomorde is gelijk aan 1 voor cellen waaraan geen stroomopwaartse cel verbonden is. Waar twee kanalen van orde 1 samenkomen, ontstaat stroomafwaarts een kanaal van orde 2. Op deze stroomorde wordt een drempelwaarde van 5 toegepast om te komen tot een identificatie van grachten. Als dit resultaat vergeleken wordt met hoge resolutie satellietbeelden, wordt er een overdetectie van de grachten vastgesteld: i.e. er worden grachten gedetecteerd waar er geen zijn. Door toepassen van een convolutiematrix worden gracht-cellen gefilterd die ondieper zijn dan 0,2m t.o.v. de omringende cellen.

De grachtendetectie werd uitgevoerd op basis van het DHM met resolutie 5m omdat deze methode op 1m resolutie computationeel te zwaar bleek. De uiteindelijk geselecteerde grachten worden gecombineerd met het waterlopennetwerk van de VHA (Figuur 16) en de watermassa's van Open Street Map (§3.6.5) om een totaalbeeld te verkrijgen van het oppervlaktewater.



Figuur 16 :Totaal drainagenetwerk VHA en grachten.

3.7 BODEMBEDEKKINGSKAART

Digitaal Vlaanderen heeft in navolging van de datasets van 2012 en 2015, in samenwerking met vertegenwoordigers van de Vlaamse Milieumaatschappij, Departement Omgeving en Waterbouwkundig Laboratorium de operationele aanmaak van een Bodembedekkingskaart Vlaanderen (BBK) en afgeleide datasets, namelijk Bodemafdekkingskaart en Waterondoorlaatbaarheidskaart, voor 2018 gerealiseerd. Aan de hand van een segmentatie op basis van de raster en vector datasets worden segmenten aangemaakt. Deze segmenten worden in een 2de fase geclassificeerd op basis van de informatie uit dezelfde raster en vector datasets.

De bodembedekkingskaart is een raster-dataset die de bodembedekking in Vlaanderen weergeeft, conform de toestand zoals opgenomen in 2018. Deze kaart op resolutie 1m heeft als doelpubliek de algemene gebruiker die een kaart met bodembedekking van Vlaanderen wil raadplegen als een basis voor diverse analyses. (Bron: agentschap Digitaal Vlaanderen).

4 FEATURE ENGINEERING: VERKLARENDE VARIABELEN

De hulpvariabelen zijn parameters die over het volledige gewenste gebied – hier Vlaanderen dekkend – aanwezig zijn en die door een model aangewend worden om een ontbrekende waarde – hier de GxG's – te berekenen. Ze worden geconstrueerd op hetzelfde raster als het beoogde resultaat: 10m x 10m resolutie raster. Een overzicht van de gebruikte hulpvariabelen wordt gegeven in Tabel 6.

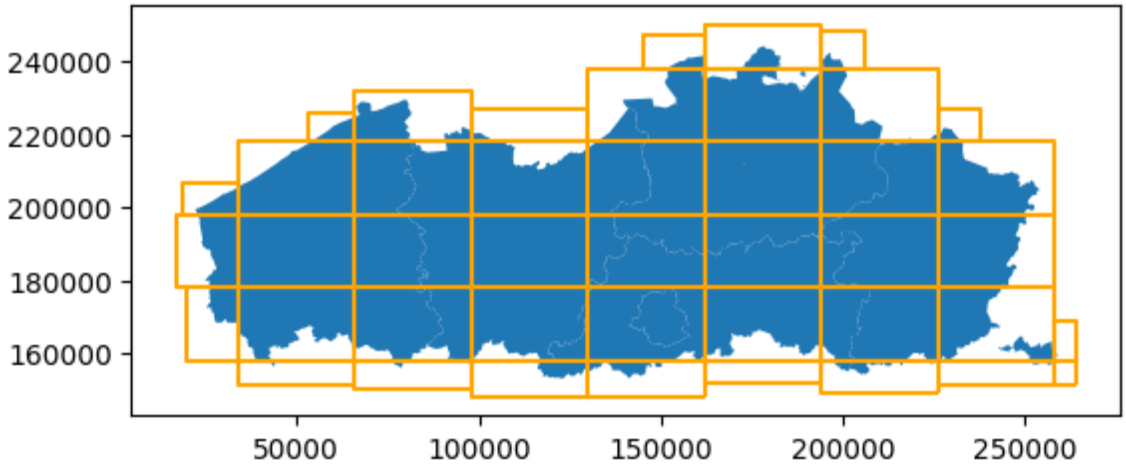
Tabel 6 : Beschrijving hulpvariabelen

Afkorting	Beschrijving
X	Lambert-x coördinaat
Y	Lambert-y coördinaat
Textuur_c_[klasse]	Textuurklasse van de bodemkaart. Dummyvariabele.
Drainage_c_[klasse]	De oude drainageklasse. Aangezien deze ook beschikbaar is voor grote delen van Vlaanderen kan dit een goede hulpvariabele zijn. Ook hier wordt een dummyvariabele geconstrueerd.
Substr	Enkele dummy variabele die aangeeft of substraat aanwezig is of niet (12% van de locaties in de databank).
DTM_abs	z-waarde mTAW van DTM (omgezet naar 10m resolutie zie §4.3)
relDTM	Relatieve maaiveldhoogte (1m resolutie) gebaseerd op het verschil tussen DTM van de locatie en het gemiddelde DTM in een cirkel met een straal van 20m. Eenheid: m. (beperkt) randeffect mogelijk.
Aqit_d	Aquitard diepte, maximum 100m onder het maaiveld. Eenheid m t.o.v. maaiveld. Details in *§4.4*
HCOV_surface_[klasse]	Dagzomende HCOV laag. In de modellering wordt hiervan een dummy variabele gemaakt waarbij elke klasse een aparte variabele is en de textuurklasse van toepassing op een locatie op 1 wordt gezet en de rest op 0.
BBK_[klasse]	Bodembedekking. Dummy variabele die aangeeft welke bodembedekking aanwezig is op de locatie.
Drain_dist	Afstand tot dichtstbijzijnde waterloop. Eenheid: m.
GWT	De drooglegging (§4.5.2).
Drain_dens	Drainagedichtheid (§4.5.3). Dit is de drainagedichtheid gebaseerd op een straal van 200m rond de locatie.
SWI_q[kwantiel]_d[diepte]	Soil Water Index. Verschillende SWI-variabelen zijn geconstrueerd uit de aggregatie van de Copernicus SWI data. Deze index is eenheidloos.
Verlaging (t.g.v. grondwaterwinning)	Een variabele die de ingeschatte grondwaterpeilverlaging weergeeft onder invloed van nabijgelegen grondwateronttrekking (meer detail in §4.7). Eenheid: m onder maaiveld.
[Databron]	De databron kan een effect hebben op de inschatting van de GxG. Verschillende databronnen hebben andere kwaliteit en meettechnieken. Deze wordt dus ook opgenomen als hulpvariabele. De databronnen worden weergegeven in Tabel 3

De ruwe versie van bepaalde informatiebronnen is echter vaak niet direct toepasbaar voor onze doeleinden. Zo zal een algoritme niet wijs geraken uit de drainageklassen a tot i, tenzij men informatie verschaft dat deze categorisch of ordinaal zijn, i.e. ze worden beschouwd als categorieën of ze volgen elkaar op. Anderzijds bevat een databron zoals het digitaal hoogtemodel een schat aan bijkomende contextuele informatie boven op de directe hoogte van een bepaald punt, die niet zomaar door een computer gebruikt kan worden. Deze vertaalslag is ook wel gekend als "preprocessing" (wanneer triviaal) of "feature engineering" (wanneer minder triviaal en meer arbeidsintensief).

Om de bewerkingen computationeel haalbaar te houden, wordt Vlaanderen opgedeeld in 48 kaartbladen (Figuur 17). Let wel: het is hierbij noodzakelijk om eventuele randeffecten te minimaliseren: bij enkele hulpvariabelen wordt er een buffer toegepast zodat informatie van net buiten het kaartblad ook in acht wordt genomen. Deze bufferafstand is op 400m gezet.

Volgende paragrafen beschrijven, thematisch, welke stappen werden uitgevoerd.



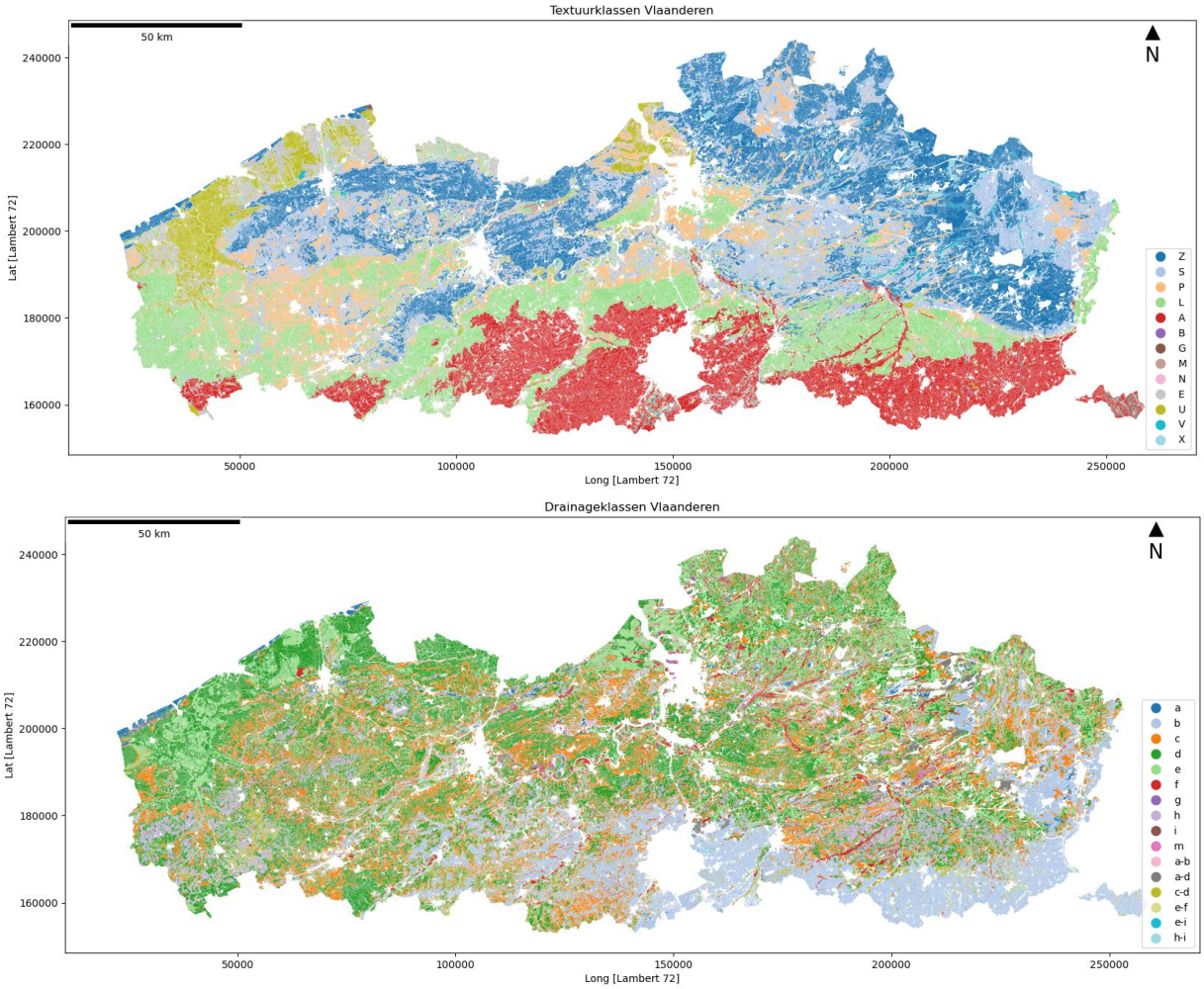
Figuur 17 : Cellen waarvoor de hulpvariabelen afzonderlijk worden berekend

4.1 AFGELEIDEN VAN DE BODEMKAART

Uit de bodemkaart halen we volgende afgeleiden (zie ook Figuur 18):

1. **De textuurklassen** worden uit de bodemkaart (DOV) gehaald. Indien de textuurklasse uit verschillende letters bestaat, wordt voorlopig enkel de eerste gebruikt (bv. A-U-S wordt A).
2. Op basis van de bodemkaart (DOV) wordt ook de **vorige drainageklasse** toegevoegd.

Aangezien bovenstaande variabelen categorisch zijn, worden deze omgezet naar zogenaamde 'dummy' variabelen. Hierbij wordt een variabele per klasse gedefinieerd. Als een bepaald punt behoort tot een bepaalde klasse wordt de waarde op 1 gezet, anders op 0.



Figuur 18: Textuurklassen (BOVEN) en drainageklassen (ONDER) uit de bodemkaart (Bron: DOV).

4.2 AFGELEIDEN VAN DE BODEMERODIBILITEITSKAART (K-FACTOR)

De bodemerodibiliteitskaart (§3.4.1.3) werd rechtstreeks afgeleid van de textuurklasse. Via *reverse engineering* wordt eenvoudig de geïnterpoleerde textuurklasse bekomen (Tabel). Waar de textuurklassen ontbreken in de bodemkaart (Tabel 7) wordt de getransformeerde K-factor aangewend om deze aan te vullen. Volgende stappen werden doorlopen:

- De VCS-service van DOV wordt bevraagd. Een WCS-service bevat rasters. Deze rasters worden opgehaald aan de hand van een omsluitingskader en resolutie. Hier wordt het omsluitingskader gebruikt van de polygonen met missende textuurklasse van de bodemkaart en een resolutie van 5m (resolutie van de brondata).
- Aansluitende pixels van deze rasters worden samengevoegd tot polygonen indien ze dezelfde waarde hebben.
- Deze polygonen worden vervolgens bijgesneden tot de vorm van de bodemkaart polygonen met missende textuurklassen.
- De K-factor van deze polygonen wordt vervolgens omgezet naar een textuurklasse met behulp van de weergegeven waarden in Tabel 8.

Tabel 7: Tabel met bodemseries waarvan de textuurklasse ontbreekt.

Bodemtype	Bodemserie
OB	Bebouwde zones
OE	Groeven
OT	Sterk vergraven gronden
ON	Opgehoogde gronden
OC	Verdwenen bewoningen
OH	Belten
OL	Doorbraak geulgronden
OA	Afgegraven gronden
OW	Gekarteerd als water

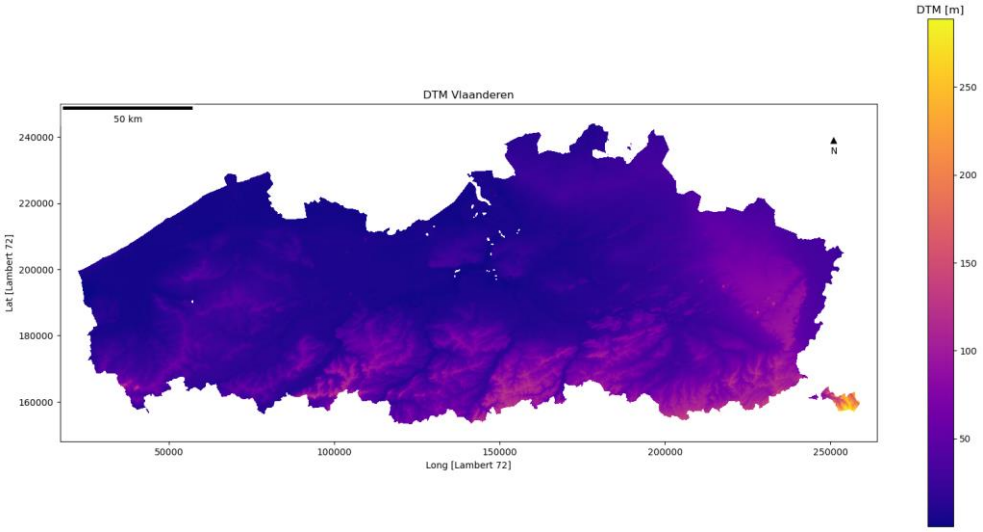
Tabel 8: Omzetting van K-factor naar textuurklasse.

Textuurklasse	K-factor
Leem (A)	0,0420
Zandleem (L)	0,0400
Licht zandleem (P)	0,0250
Lemig zand (S)	0,0200
Zand (Z)	0,0120
Klei (E)	0,0275
Zware klei (U)	0,0225

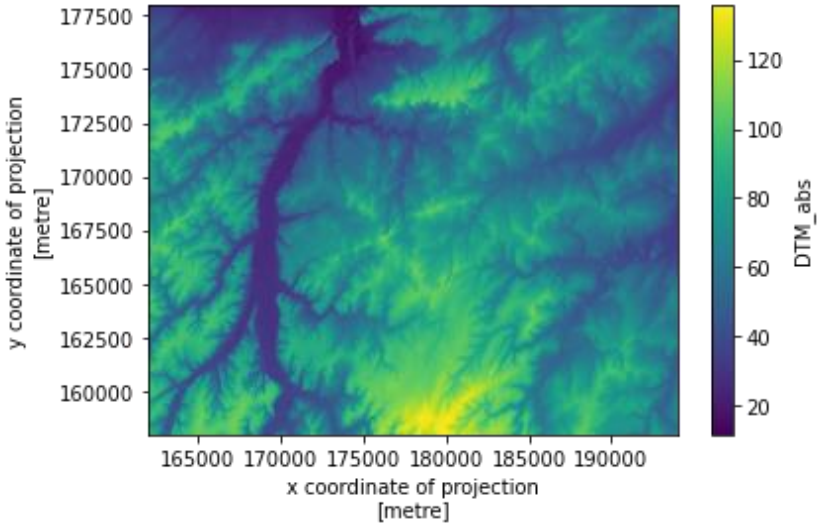
4.3 AFGELEIDEN VAN HET DIGITAAL HOOGTEMODEL

Het digitaal hoogtemodel is omgevormd tot een 10m resolutie raster waarbij de dichtstbijzijnde waarde van het 1m resolutie raster genomen is.

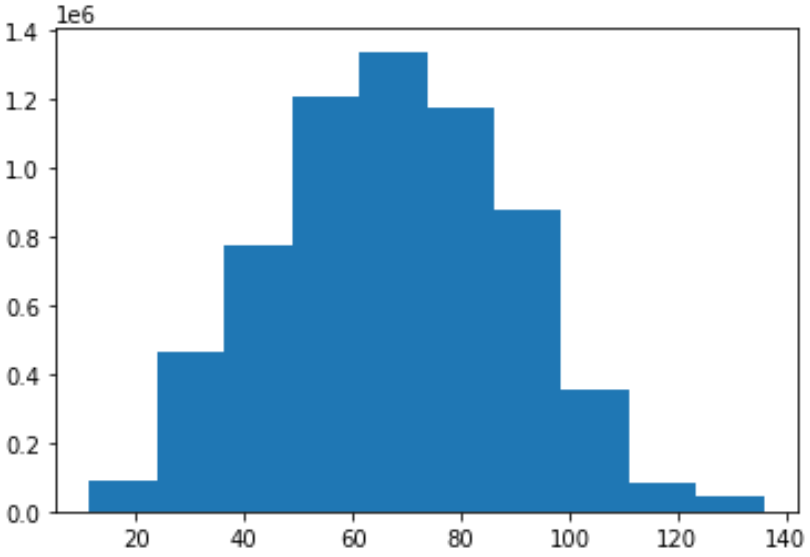
- **Maaiveld t.o.v. TAW**: de absolute maaiveldhoogte volgens het Digitaal Hoogtemodel Vlaanderen II dat Vlaanderendekkend beschikbaar is. Figuur 20 geeft het hoogtemodel weer in het Dijlegebied (pilootgebied) en een histogram van de verdeling van DTM-waarden in dit gebied is afgebeeld in Figuur 21.
- **Relatieve maaiveldhoogte**: de hoogte van een punt ten opzichte van de gemiddelde hoogte in een zoekgebied met een vaste straal rond dat punt.



Figuur 19 : DTM kaart Vlaanderen in m TAW.



Figuur 20 : Voorbeeld van 10m resolutie DTM-raster in de Dijlevallei.



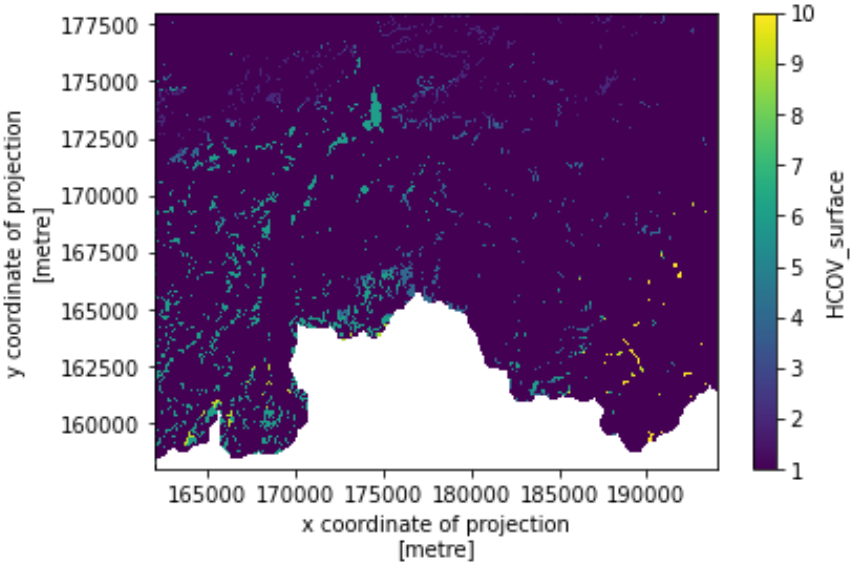
Figuur 21 : Histogram van DTM-waarden in de Dijlevallei.

4.4 AFGELEIDEN VAN HET HYDROGEOLOGISCH MODEL

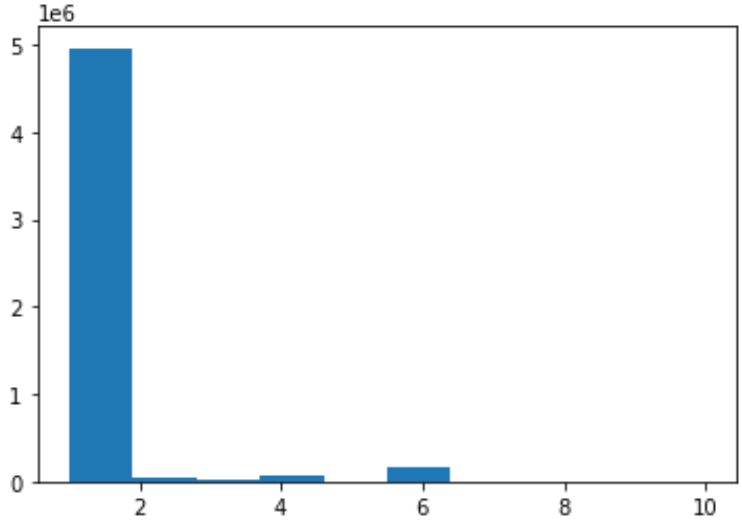
4.4.1 Dagzomende hydrogeologische laag

Naast de dagzomende bodem, omvat ook de onderliggende hydrogeologie nuttige informatie. Volgende verklarende variabelen worden berekend uit het hydrogeologisch model van Vlaanderen: **Dagzomende hydrogeologische laag**. Er is besloten om de dagzomende laag (Figuur 22 voor pilootgebied) als categorische variabele in te voegen in de regressieanalyse.

Een andere optie was één hydraulische permeabiliteit toe te kennen aan de hydrogeologische eenheid en deze zo om te vormen tot een pseudo-numerieke variabele. Na overleg is er besloten deze piste niet te volgen omdat deze stap een bijkomende fout kan veroorzaken in de regressieanalyse. Bovendien is het onmogelijk om één waarde toe te kennen aan het “Quartaire Aquifersysteem”.



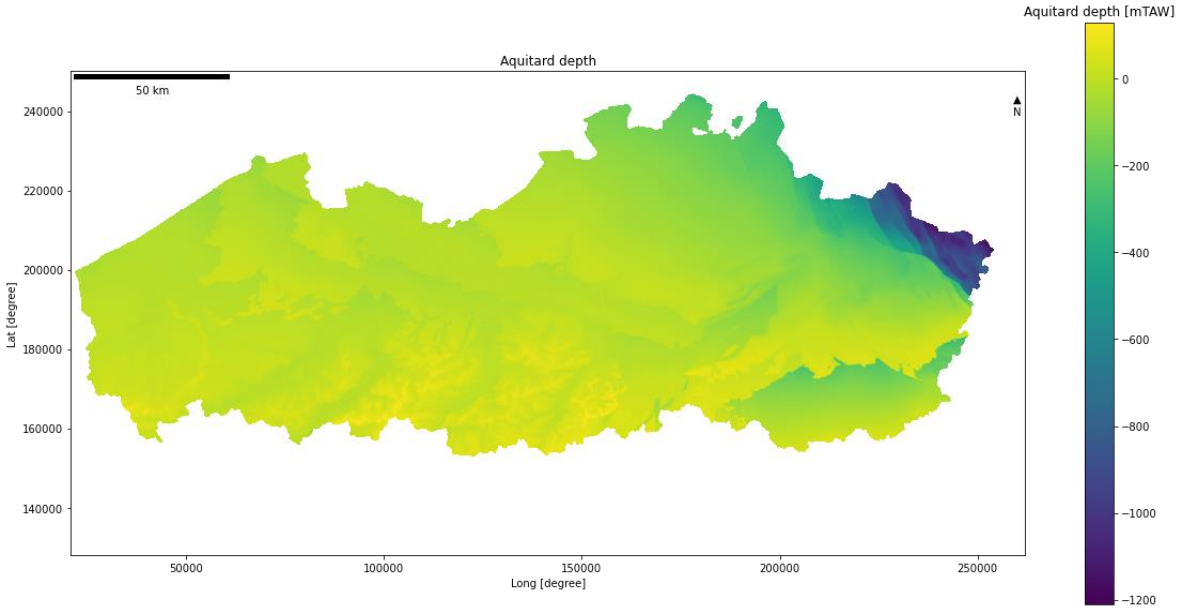
Figuur 22 : Dagzomende HCOV lagen in de Dijlevallei.



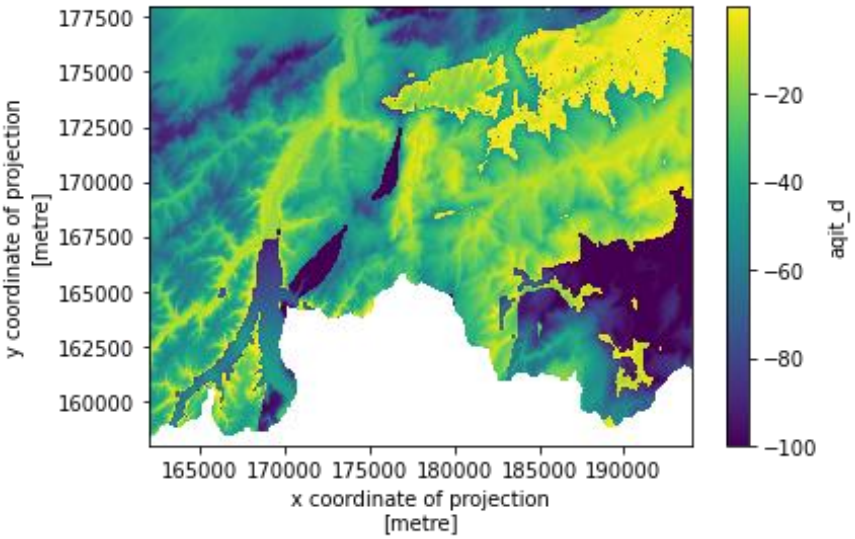
Figuur 23 : Histogram van dagzomende HCOV lagen in de Dijlevallei.

4.4.2 Diepte tot de top van de bovenste 'Aquitard'

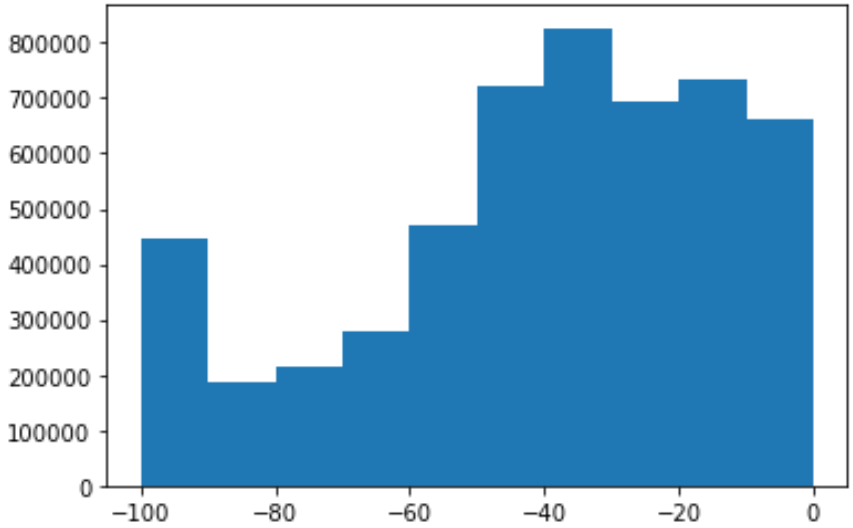
Een ondiepe laag met geringe hydraulische doorlatendheid (lees zgn. *aquitard*) kan een stuwwatertafel veroorzaken. We gaan na of de diepte tot de eerste aquitard een meerwaarde kan bieden voor het model. De definities van de *aquitards* zijn overgenomen van de hoofdeenheden uit het rapport en bijlagen van de HCOV-versie-2 (Deckers *et al.* 2019b). Het resultaat wordt voor Vlaanderen voorgesteld in Figuur 24 en voor het pilootgebied in Figuur 25. Figuur 26 geeft een histogram van de verdeling van aquitarddieptes in het pilootgebied weer.



Figuur 24 : Diepte vanaf het maaiveld tot aan de bovenste aquitard.



Figuur 25 : Diepte tot eerste aquitard in de Dijlevallei.



Figuur 26 : Histogram van aquitarddieptes in de Dijlevallei.

4.5 AFGELEIDEN VAN WATERLOPEN

Drainage naar oppervlaktewaterlichamen kan een sterke invloed uitoefenen op de grondwaterdynamiek. Meerdere afgeleiden zijn uitgetest. Voor het berekenen van deze features wordt gebruik gemaakt van de waterloopkaart zoals besproken in §3.6.5.

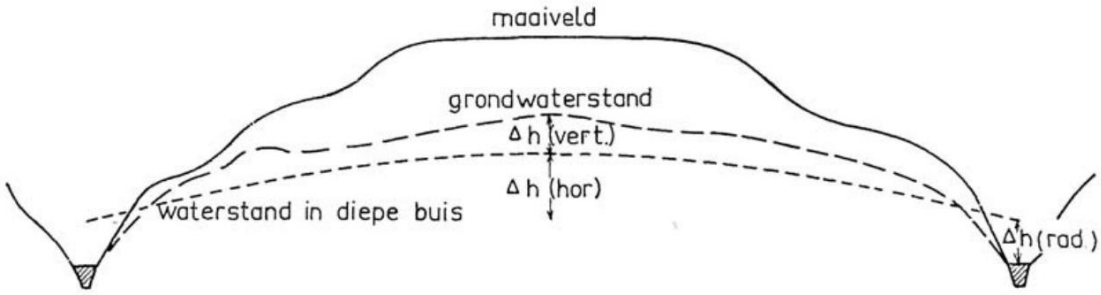
4.5.1 De afstand tot de dichtstbijzijnde waterloop

Voor iedere pixel werd de (euclidische) afstand tot de dichtstbijzijnde waterloop berekend. Voor grotere vlakdekkende datasets is hiervoor het 'kd-tree' algoritme (Maneewongvatana & Mount 1999) gebruikt voor snelle opzoeking van dichtstbijzijnde punten. Dit laat toe om efficiënt een

Vlaanderendekkende kaart op te stellen. Deze variabele wordt uitgerekend per kaartblad met een buffer van 400m.

4.5.2 Drainagediepte en drooglegging (GWT)

Gelaagdheid van de bodem heeft een invloed op de grondwaterstroming, waarbij de verticale weerstand van de slecht doorlatende lagen op geringe diepte van belang is (Ernst, 1958). De afvoer van overtollig regenwater gaat bij dergelijke lagen met een diepte van 3 à 4 meter verticaal naar beneden, aangezien het geleidend vermogen meestal pas in hieronder gelegen grove lagen voldoende is om een horizontaal transport over grote afstanden mogelijk te maken. Een verhoging van de opbolling van de grondwaterspiegel is hiervan het gevolg (Figuur 27).



Figuur 27 :Schematische weergave van de stijghoogteverschillen die de verschillende componenten van de drainageweerstand bepalen (naar: van Hoorn (1960)).

Om een drainagepeil te bepalen dat rekening houdt met een opbolling (formule van Bear, Hooghoudt, Wesseling), dienen de drainerende diepte en de dominant drainerende gracht geïdentificeerd te worden. Dit dient te gebeuren op perceelniveau, waarbij bovendien een rechthoekige perceelvorm vereist is. Deze methode werd uitgetest, maar bleek computationeel onhaalbaar op schaal van het huidige onderzoek (4819770 percelen in Vlaanderen).

De drainerende diepte wordt daarom geschat met een methode gebaseerd op Finke *et al.* (2010). Hierbij wordt de **drooglegging** berekend door eerst de waterpeilen[2] in de waterlopen en watervlakken te interpoleren via stuksgewijze lineaire interpolatie (convex hull interpolation: Barber *et al.*, 1997) en daarna de drooglegging te berekenen als het verschil tussen deze geïnterpoleerde peilen en de maaiveldhoogte. Deze methode bleek computationeel wel haalbaar. Merk op hoe deze – bij poldergebieden met artificieel peilbeheer – een beperkte fout kan veroorzaken.

Net zoals de andere variabelen wordt deze uitgerekend per kaartblad. Dit kan leiden tot randeffecten aangezien bij de kaartbladranden er geen complete informatie is van de naburige waterlopen. Om randeffecten zoveel mogelijk weg te werken is hier ook de buffer rond elk kaartblad in acht genomen van 400m.

Zelfs met deze buffer kan bij de hoeken van de cel in sommige gevallen geen convexe omsluiting worden gegenereerd voor de stuksgewijze lineaire interpolatie. Deze, kleine, oppervlaktes van de hoeken worden geïmputeerd met de dichtste beschikbare waardes. We beschouwen de kaartbladen en buffers echter als groot genoeg om deze effecten minimaal te houden.

[2] Ter vereenvoudiging, werd de 1x1m resolutie DTM aangewend op de waterlopen en vlakken om het waterpeil te bepalen.

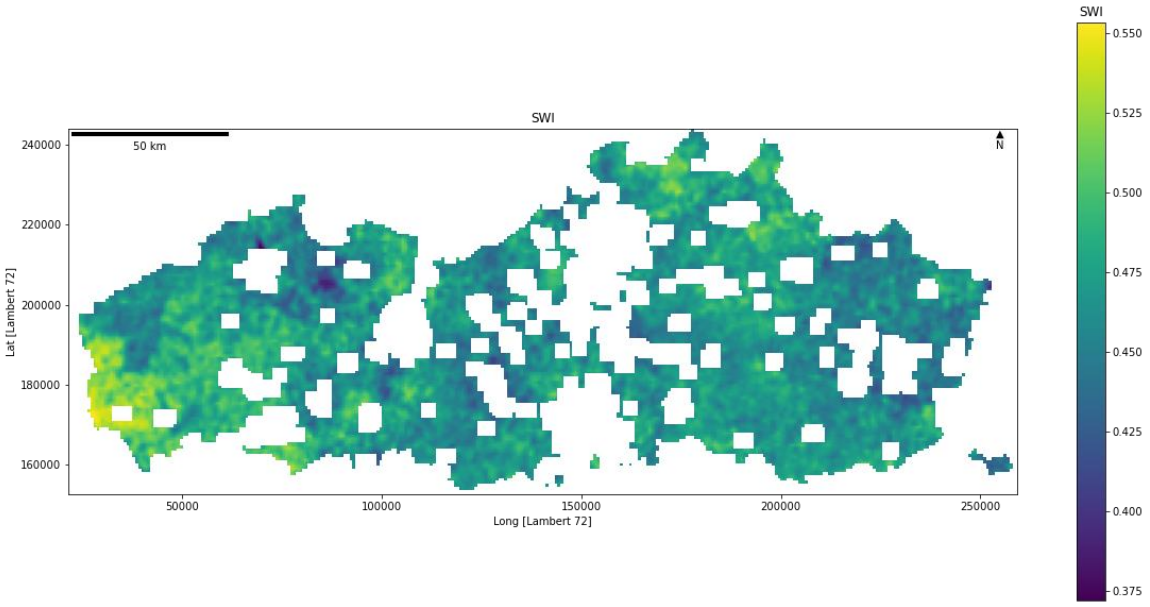
4.5.3 Drainagedichtheid

Voor elke cel werd berekend hoeveel ‘drainerende cellen’ er in een omgeving (van 200m bij 200m) aanwezig zijn. Deze parameter vormt de drainagedichtheid. Ook hier zal er een randeffect plaatsvinden. We beschouwen de cellen echter als groot genoeg om deze minimaal te houden. Ook hier wordt de variabele uitgerekend per kaartblad met de nodige buffer.

4.6 AFGELEIDEN VAN SWI

Remote sensed bodemvochtdata komende van de *Soil Water Index* van Copernicus (SWI) leveren een tijdreeks van gebiedsdekkende rasters op en dit bovendien voor verschillende dieptes tot 1m onder maaiveld (0.02, 0.05, 0.1, 0.15, 0.2, 0.4, 0.6 en 1m diepte). Hiervan kan men een wijd gamma aan verklarende variabelen afleiden, maar men dient steeds de focus van deze opdracht te behouden: het berekenen van een klimatologische variabele (GxG’s). In eerste instantie werd er gestart met louter het gemiddelde van alle dieptes (excl. ontbrekende metingen) over de gehele periode van 2015 tot 2021, die herschaald zijn naar een waarde tussen 0 en 1 (Figuur 28). Nadat bleek dat deze variabele een meerwaarde bood voor de regressieanalyse, werd overgegaan tot het capteren van een hele reeks afgeleiden: voor elke diepte (0,02m; 0,05m; 0,1m; 0,15m; 0,2m; 0,4m; 0,6m en 1,0 m) is het 0,1 en 0,9 kwantiel genomen voor elke SWI-rastercel. Al deze kwantielen worden opgenomen in de verdere regressieanalyses.

Alle meetlocaties worden steeds gelinkt aan de dichtstbijzijnde SWI-meting bepaald door het centrum van iedere rastercel. Op verschillende locaties echter in stedelijke gebieden zijn er echter geen SWI-waarden beschikbaar. Indien de regressietechniek ermee kan omgaan, wordt een missende SWI-waarde beschouwd als een aparte categorie. In andere geteste regressietechnieken worden ofwel deze variabelen niet gebruikt of worden er geen predicties uitgevoerd op deze locaties.



Figuur 28: Afgeleide van SWI met hierop duidelijk een aantal ontbrekende locaties (sterk verstedelijkte gebieden). De hogere waarden (lichtere kleur) zijn de hogere vochtgehaltes over een langjarig gemiddelde (2015-2021) en over alle dieptes t.e.m. 1m. De lagere waardes (donkerdere kleur) kunnen worden geïnterpreteerd als lagere algemene vochtgehaltes.

4.7 GRONDWATERWINNINGEN

Grondwaterwinningen in boorputten kunnen zorgen voor een verlaging van het grondwaterpeil. Aangezien de drainageklassen van de bodemkaart bepaald worden aan de hand van die grondwaterpeilen, is het van belang te kunnen inschatten hoe groot de verlaging in de omgeving zal zijn. Hieronder wordt een methode beschreven voor de berekening van de verlaging veroorzaakt door winning in freatische lagen. Winning in niet-freatische lagen valt buiten het bestek van dit rapport.

De grootte van de grondwaterpeilverlaging veroorzaakt door grondwaterwinning wordt beïnvloed door een aantal factoren, zoals afstand tot de pompput en grootte van het opgepompte debiet. Voor grondwaterwinning in freatische lagen kan een inschatting gemaakt worden aan de hand van de formule van Dupuit (Dupuit, 1863):

$$\Delta\phi = \frac{-Q}{2\pi kD} \cdot \ln\left(\frac{r}{R}\right)$$

Met:

- $\Delta\phi$: de verlaging op afstand r (m)
- Q: debiet van de winning (m^3/dag)
- k: doorlaatbaarheid aquifer (m/dag)
- D: dikte van de aquifer (m)
- r: afstand tot de pompput (m)
- R: invloedstraal (m)

De verlaging ($\Delta\phi$) voor een bepaalde peilbuis wordt berekend als de som van de verlagingen door alle pompputten waarvoor de peilbuis binnen hun invloedstraal ligt.

Het debiet van de winning wordt afgeleid uit de kolom 'vergund_jaardebiet' in vergunde_installaties_20210625.xlsx opgevraagd bij DOV. Hier wordt aangenomen dat het vergunde debiet gelijkgesteld kan worden aan het effectief opgepompt debiet, en dat het vergunde jaardebiet gelijkmatig wordt opgepompt doorheen heel het jaar. Het dagdebiet wordt berekend door het vergunde jaardebiet te delen door 365.

De code van de aquifer waaruit de winning gebeurt, wordt opgehaald uit de kolom 'aquifer_code' in vergunde_installaties_20210625.xlsx. De dikte van de aquifer wordt dan bepaald aan de hand van het H3Dv1.1 model uit de HCOV-kartering (Meyus *et al.*, 2000). Van dat model zijn alleen de diktes van de hoofdeenheden van de HCOV-hoofdeenheden en de subeenheden van HCOV 0200 beschikbaar.

De horizontale doorlatendheid van de aquifers wordt overgenomen uit de tabel van Lebbe *et al.* (2004), en wordt opnieuw gekozen aan de hand van de aquifer code uit vergunde_installaties_20210625.xlsx. De tabel stelt op vereenvoudigde wijze parameters voor per HCOVv1-eenheid. Hierbij moet wel rekening worden gehouden dat de hydraulische parameters in werkelijkheid sterk lateraal en met de diepte kunnen variëren, zeker in het Quartair (HCOV0100). De methode om deze parameters te bepalen, zou eventueel in de toekomst kunnen aangepast worden. De hydraulische parameters zijn beschikbaar voor de meeste hoofd- en subeenheden.

De horizontale afstand (r) van de locatie van de grondwaterpeilbuis tot de winning wordt bepaald door de euclidische afstand te nemen tussen de coördinaten in het systeem Lambert72 (EPSG: 31370).

De invloedstraal van een winning (R) wordt bepaald aan de hand van de volgende formule:

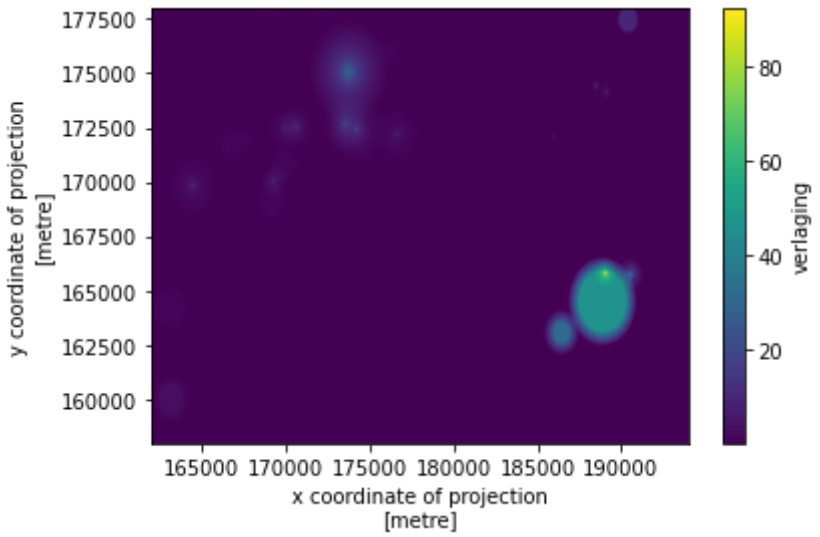
$$R = \sqrt{\frac{Q}{\pi \cdot N}}$$

Met:

- Q: het debiet van de winning (m^3/dag);
- N: de grondwatervoeding (m/dag)

Uit deze formule is af te leiden dat de grondwatervoeding de invloedstraal beïnvloedt. Echter, aangezien de grondwatervoeding niet homogeen is over Vlaanderen, bepaalt de invloedstraal deels ook weer de grootte van de grondwatervoeding. Hierdoor is er een iteratief proces noodzakelijk om beide te kunnen bepalen voor de grondwatervoeding. Er wordt gestart met de gemiddelde voeding voor Vlaanderen (0,00061 m/dag), waarna een invloedstraal berekend wordt. Hiermee wordt een 'buffer' berekend rond de pompput. De gemiddelde grondwatervoeding binnen die buffer wordt daarna geoptimaliseerd aan de hand van het Vlaams Grondwatervoedingsmodel, opgesteld door De Smedt *et al.* (2004). Deze gemiddelde N wordt opnieuw ingevuld in de formule hierboven en zo wordt de invloedstraal berekend. Om rekenkracht te sparen, houden we het voorlopig bij één iteratie.

Aangezien ook deze variabele uitgerekend wordt per cel en pompputten van naburige cellen niet worden opgenomen, kunnen hier randeffecten ontstaan en is er mogelijk een lichte onderschatting van de verlaging bij de celranden.



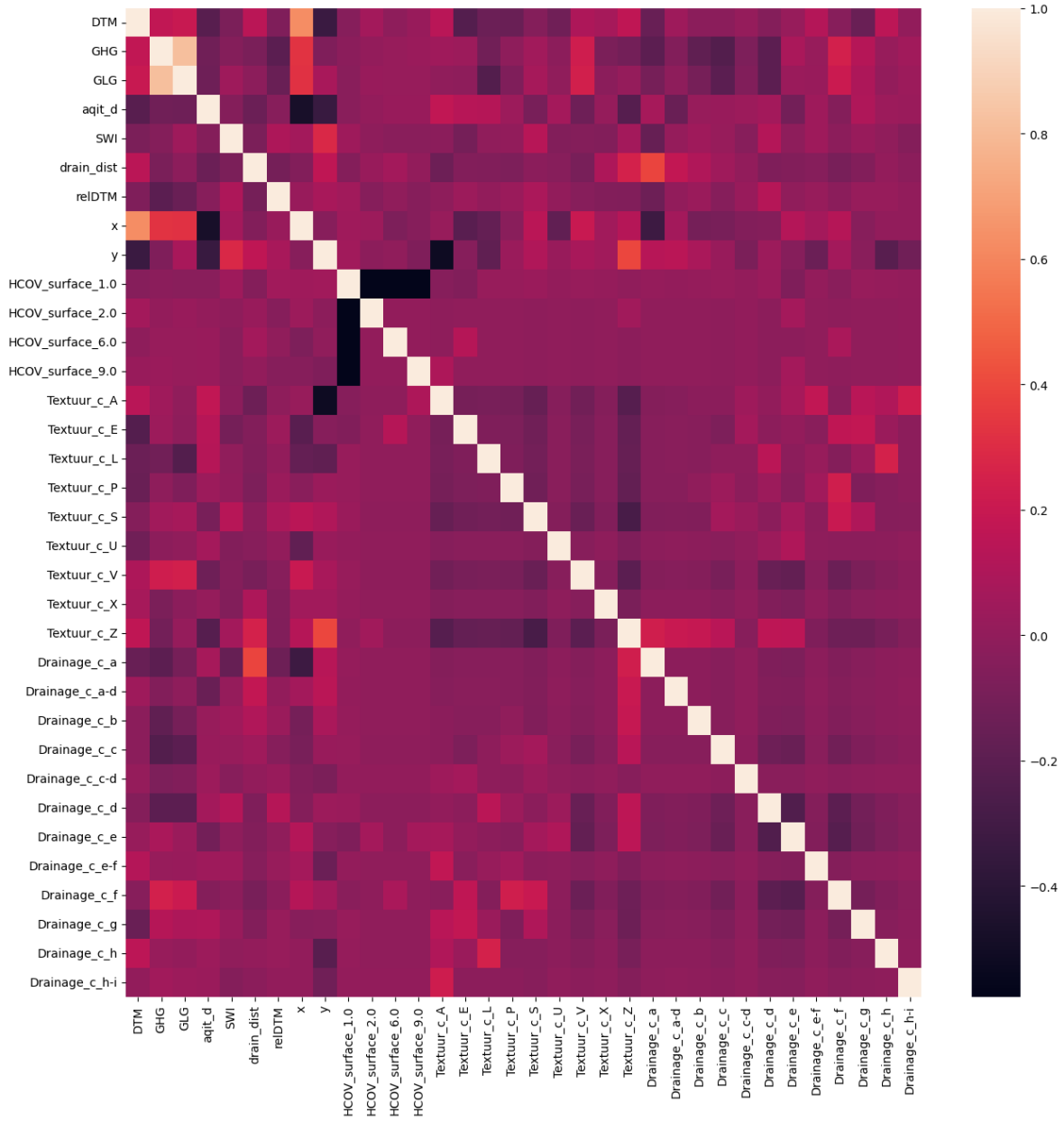
Figuur 29: Voorbeeld verlaging ten gevolge van grondwaterwinning.

4.8 SCHALING EN ONDERLINGE CORRELATIE

Voor de lineaire modeltechnieken was het noodzakelijk om de continue hulpvariabelen te herschalen. Hierbij wordt een standaard herschaling toegepast waarbij het verschil genomen wordt van de waarde en het gemiddelde van elke hulpvariabele over de trainingset afgetrokken en dit wordt gedeeld door zijn standaarddeviatie.

De *heatmap* in Figuur 30 toont de onderlinge correlatie tussen verschillende variabelen die aangewend zijn in de regressieanalyse. Het begrijpen van correlaties tussen verschillende features is essentieel, niet alleen in de context van klassieke statistiek maar ook voor machine learning-modellen. Dit is zeker relevant voor modellen zoals lineaire regressie, waar een dergelijke hoge correlatie tussen onafhankelijk variabelen kan leiden tot multicollineariteit, wat de interpretatie van de modelcoëfficiënten bemoeilijkt en eventueel onbetrouwbare extrapolaties teweegbrengt.

Het kennen van de correlaties helpt onder andere bij het identificeren van redundante hulpvariabelen. Het verwijderen van zulke hulpvariabelen kan niet alleen de modelinterpretatie vergemakkelijken maar kan ook overfitting voorkomen en de modelprestaties verbeteren.



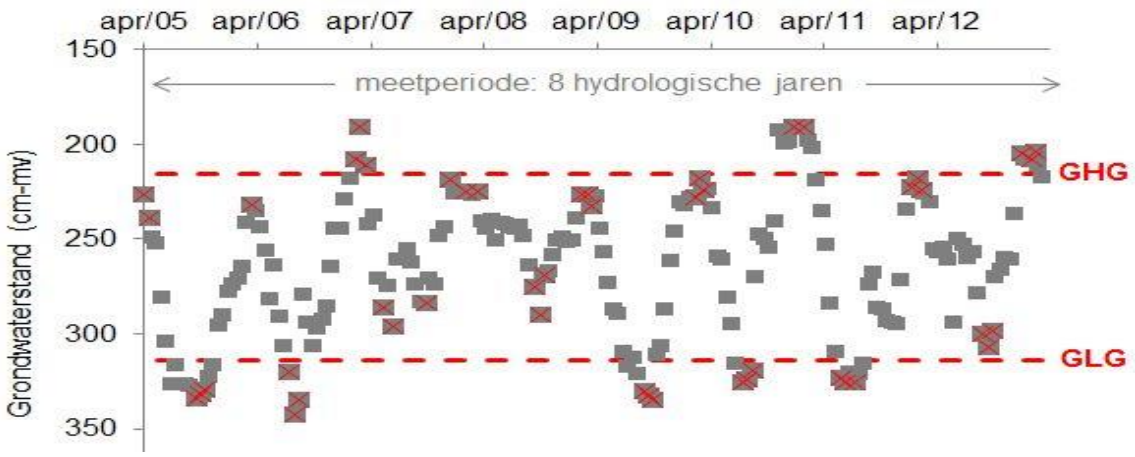
Figuur 30: Correlaties tussen verschillende variabelen in de regressieanalyse.

5 METHODOLOGIE

Dit hoofdstuk beschrijft beknopt het onderzoek dat gevoerd is om de meest optimale – gegeven de randvoorwaarden – methode te bekomen.

5.1 DEFINITIE GXG

Voor het berekenen van de GxG's wordt de definitie van DOV (cfr. Heesen, 1970) aangehouden (Figuur 31): de GHG en GLG worden berekend als het gemiddelde van de drie hoogste/laagste peilen per jaar van minimaal acht jaren, waarbij de grondwaterstand tweemaal per maand gemeten wordt (op of omstreeks de 14de en 28ste dag). Met 'jaren' worden hier geen kalenderjaren maar wel hydrologische jaren bedoeld, die beginnen op 1 april en eindigen op 31 maart.



Figuur 31: Gebruikte klassieke GxG definitie

Bovenstaande definitie is echter afgestemd op de traditionele meettechniek van de meeste grondwatermeetnetten: maandelijks zo dicht mogelijk bij de 14de en 28ste dag. Deze definitie is echter niet rechtstreeks toepasbaar op meetreeksen met hogere temporele resolutie. Het grondwaterpeil evolueert namelijk relatief traag: eerder op week- dan dagniveau. Voor meetreeksen met hogere resolutie (e.g. grootteorde van dagen of hoger) zijn er 2 gangbare methodes:

1. Selecteer de metingen op de 14de en 28ste van de maand.
2. Neem het 90ste en 10de percentiel als indicatie voor GHG en GLG.

Wanneer de meetreeksen een hogere resolutie hebben (e.g. na een tijdreeksmodellering), hanteren we de eerste methode: de dagwaarde op de 14de en 28ste van de maand. Dit opdat de resulterende GxG nog steeds vergelijkbaar zouden zijn met die van lagere resolutie meetreeksen. We voerden bovendien een korte analyse uit die de twee methodes met elkaar vergelijkt. De twee methoden leveren nagenoeg dezelfde resultaten, met een minimale correlatie van 99.9% en geen systematische verschuivingen tussen de bekomen GxG's. We concluderen hieruit dat het dus weinig uitmaakt welke methode we hanteren.

Voor deze studie werd gewerkt met zo recent mogelijke data, na 2010. Door steeds dezelfde strikte criteria te hanteren op de GxG berekening wordt een doelvariabele bekomen op basis van gelijkaardige tijdsreeksen en wordt ook een te grote heterogeniteit vermeden.

5.2 TIJDREEKSMODELLERING

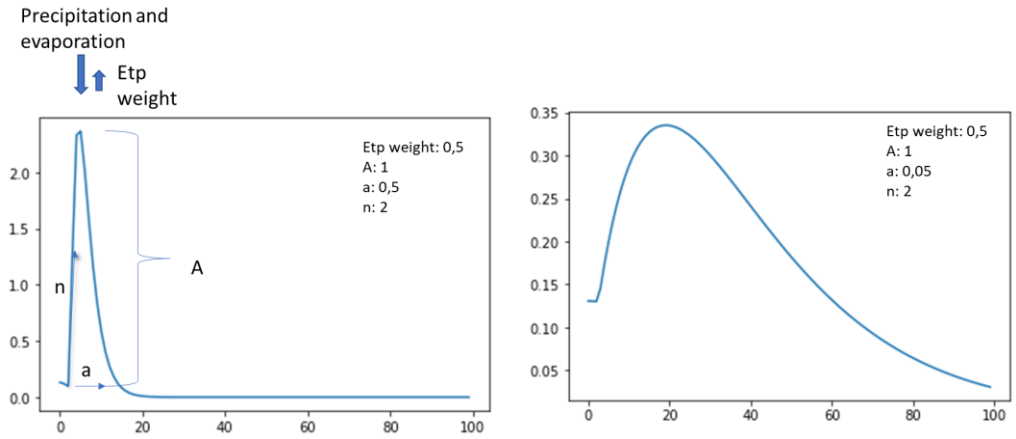
5.2.1 Inleiding

Zoals aangegeven zijn er te weinig peilmetingen met voldoende temporele resolutie en duur waarop de GxG's eenvoudig te berekenen zijn om een goede geostatistische herkartering toe te passen. Om deze dataset uit te breiden wordt er een tijdreeksanalyse toegepast. Tijdens de offerteopmaak werd ervan uitgegaan een KalMAX te fitten, maar dankzij terugkoppeling met Professor Finke en de resultaten van Van de Wauw (2012 – zie §2.2.2.1) – werd beslist om over te stappen op PIRFICT.

5.2.2 PIRFICT

5.2.2.1 Theorie

In een PIRFICT-modellering (Van de Wauw, 2012) worden de fysische processen die zich afspelen bij de drainage van de bodem afgestemd op de meetdata en kan bovendien makkelijk omgegaan worden met onregelmatige tijdintervallen. PIRFICT is in essentie een eenvoudig model. Het kan opgedeeld worden in een deterministisch en stochastisch onderdeel. De intuïtie van het deterministische deel wordt a.d.h.v. 1 regenvalmoment schematisch weergegeven in Figuur 32. Hierbij wordt de reactie op het doorsijpelen van water gemodelleerd aan de hand van een Pearson type III distributie. Verschillende parameters bepalen de vorm en hoogte van deze distributie als respons op neerslagoverschot. A, a, n bepalen de vorm van de curve. Deze parameters kunnen ook gelinkt worden aan hydraulische eigenschappen van de bodem. Hierbij wordt de A beschreven als de grootte van de respons op het neerslagoverschot (responsfactor). Dit is sterk afhankelijk van de lokale bodemsamenstelling en geomorfologie. Hierbij is het 'lokale' behoud van massa niet per se van toepassing. Deze parameter is typisch niet aanwezig in klassieke reservoirmodellering. De fysische onderlegging van n is het aantal gekoppelde reservoirs en a de inverse van de reservoir coëfficiënt (Nash, 1958; von Asmuth *et al.*, 2002). De 'etp weight' bepaalt het effect van de evaporatie op doorsijpelen van de regenval, en op die manier ook wat overblijft als neerslagoverschot.



Figuur 32: Intuïtie van PIRFICT-modellering en parameters a.d.h.v. 1 regenval moment op een totale serie van 100. A geeft de totale responsweging, a en n bepalen verder de vorm van de curve (zie rechterfiguur). De precipitatie wordt gecorrigeerd voor de evaporatie om tot neerslagoverschot te komen. Deze afweging wordt bepaald door de 'etp weight' parameter.

Er wordt een integratie (som) uitgevoerd over alle curves gebaseerd op de regenval en evaporatie in de volledige tijdspanne van regenval en evaporatie metingen (Van de Wauw, 2012).

$$h^{det}(t) = \overline{P_t^e} A\Gamma(\mathrm{a}(\mathrm{t}-t_0), n) + \sum_{i=t_0}^{t} (A\nabla_t(\gamma(a(t-i), n)))$$

Hierbij is t het tijdstip, h^{det} de deterministisch gesimuleerde grondwaterstand, P^e de precipitatie surplus, A, a en n zijn de parameters als eerder besproken. Γ en γ zijn respectievelijk de boven en onder incomplete gamma distributies:

$$\Gamma(\mathrm{z},\mathrm{x}) = \frac{1}{\Gamma(\mathrm{z})}\int_x^{+\infty} t^{z-1}\, e^{-t} dt$$

$$\gamma(\mathrm{z},\mathrm{x}) = \frac{1}{\Gamma(\mathrm{z})}\int_0^{x} t^{z-1}\, e^{-t} dt$$

Dit geeft dan de gesimuleerde grondwaterstand. De parameters worden geoptimaliseerd zodat de gesimuleerde grondwaterstanden zo goed mogelijk overeenkomen met de gemeten grondwaterstanden.

Aangezien er bij PIRFICT met tijdsreeksen gewerkt wordt, kan er uitgegaan worden van autocorrelatie van de residuen, met andere woorden, het residu, de afwijking van het model t.o.v. de gemeten data, zal geen random distributie volgen (WGN: *white gaussian noise*) maar zal gecorreleerd zijn over de tijd. Hiermee moet rekening gehouden worden tijdens de optimalisatie om een correcte inschatting van de parameters te kunnen maken. De modellering van deze residuterm vormt dan ook het stochastisch gedeelte van het model. Een exponentieel afnemende functie wordt mee geoptimaliseerd om het gewicht van vorige residuen te bepalen op het huidige residu (Schweppe criterium: Schweppe, 1973). Deze regularisatie wordt dan toegevoegd aan de standaardsom van de residu kwadraten minimalisatiefunctie. De minimalisatie (loss) functie wordt dan een functie van het type 'Sum of Weighted Squares Innovations (SWSI)' (Asmuth & Bierkens, 2005):

$$SWSI(\beta) = \sum_{j=1}^{N} \frac{\sqrt[N]{\prod_{i=1}^{N}(1-e^{-2a\Delta t_i})}}{1-e^{-2a\Delta t_j}} v^2(h_j, \beta)$$

Hierbij is β de set aan PIRFICT-parameters en v^2 een functie die het gekwadrateerde residu weergeeft tussen de deterministische modellering (a.d.h.v. de vergelijking hierboven) en de gemeten grondwaterstand h_j. N is het aantal gemeten neerslagmomenten voor het beschouwde tijdstip, Δt is het verschil in tijd tussen het beschouwde tijdstip en het beschouwde neerslagmoment.

5.2.2.2 Toegepast: Tijdreekssimulatie

Om een PIRFICT-modellering en simulatie te kunnen uitvoeren, zijn er evaporatie en regenvaldata nodig. Hiervoor wordt elke peilputlocatie gelinkt aan het dichtste evaporatie- en regenvalstation. Verder wordt een selectie gemaakt van data waarop PIRFICT wordt geoptimaliseerd op basis van volgende criteria:

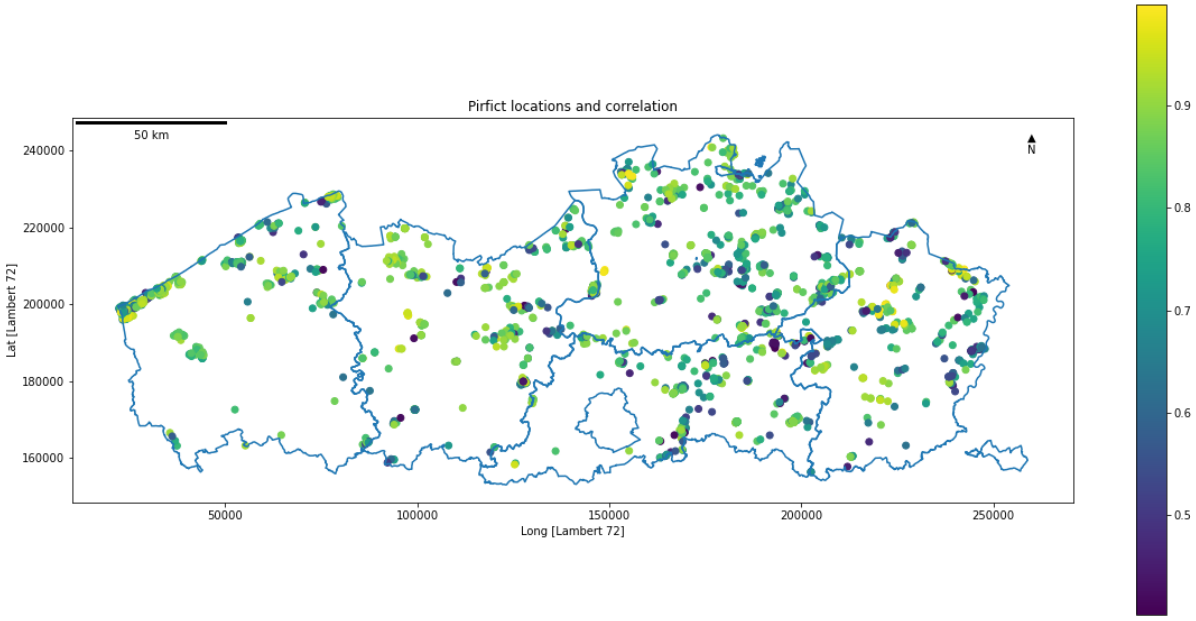
- Enkel data die binnen het tijdsvenster vallen waarvoor er beschikbare data zijn van de dichtste regenval- en evaporatiemeetposten worden geselecteerd (typisch vanaf 2004).
- Alle reeksen met minder dan 50 grondwaterstandmetingen worden verwijderd.
- De gesimuleerde dataset moet minstens een bepaalde correlatie hebben met de gemeten data om weerhouden te worden. Er worden verschillende datasets gemaakt waarbij dit criterium verschoven wordt van 0.5 naar 0.6, 0.7 en 0.8.

Op deze locaties wordt een PIRFICT-optimalisering uitgevoerd op basis van regenval en evaporatie om de PIRFICT-parameters te bekomen. Op basis van hoe goed deze optimalisatie en modellering slaagt, wordt er een bijkomende filtering uitgevoerd. Dit op basis van volgende criteria:

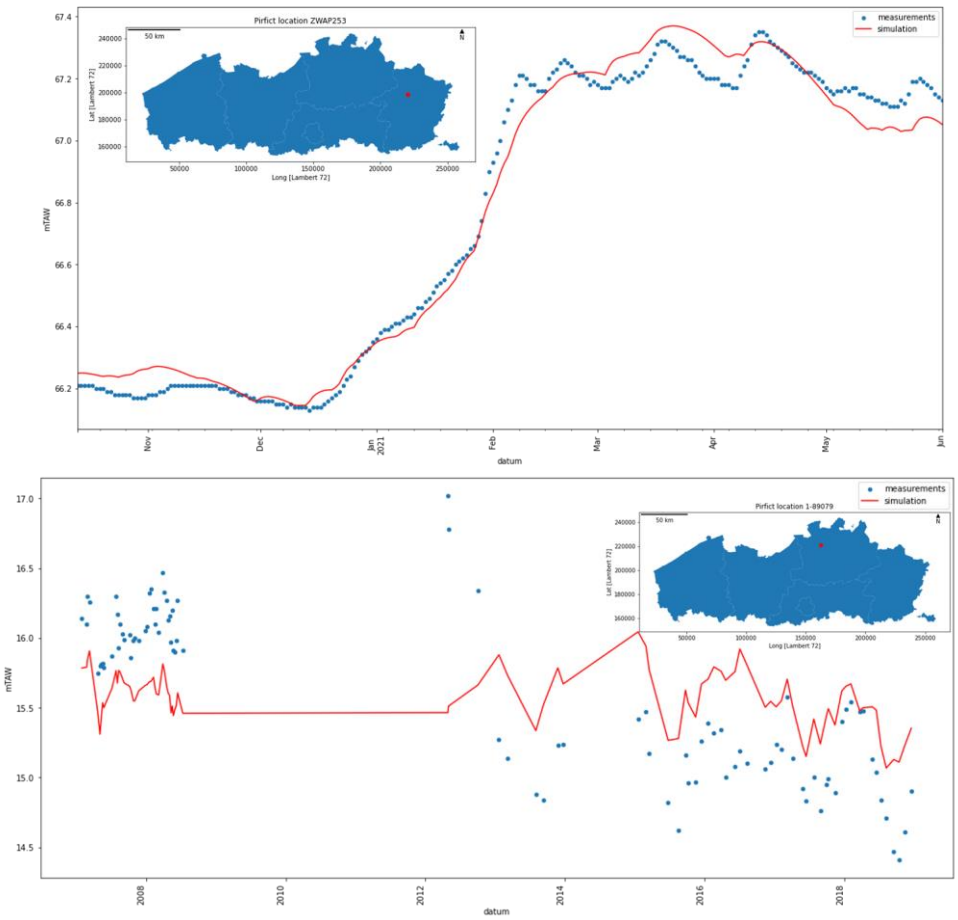
- De optimalisatie moet slagen. Vaak wordt er namelijk geen convergentie bekomen. De assumptie is dat (a) er te grote gaten in de datasets zitten die niet geïmputeerd kunnen worden en/of (b) de grondwaterstanden hier niet door regenval en evaporatie verklaard kunnen worden. Dit kan bijvoorbeeld zijn doordat er andere mechanismen spelen (zoals oppompen, artefacten in de metingen …).
- De bekomen PIRFICT-parameters moeten geloofwaardig zijn. Concreet betekent dit dat het gewicht gegeven aan de evaporatie en A groter moet zijn dan 0. Dit wil zeggen dat regenval een positief effect moet hebben op de grondwaterstand en de evaporatie een negatief effect op het doorsijpelen heeft. Verder moet A kleiner zijn dan 4, het etp gewicht kleiner dan 5, de 'a' parameter moet tussen 0 en 1 liggen, de n tussen 0 en 4, alpha tussen 0 en 2 en de basis mTAW tussen 0 en 200m (zie §5.2.2) . Als de optimalisatie hierbuiten treedt, wordt de interpretatie van dit model ongeloofwaardig. De uitschieters worden hierbij direct verwijderd.

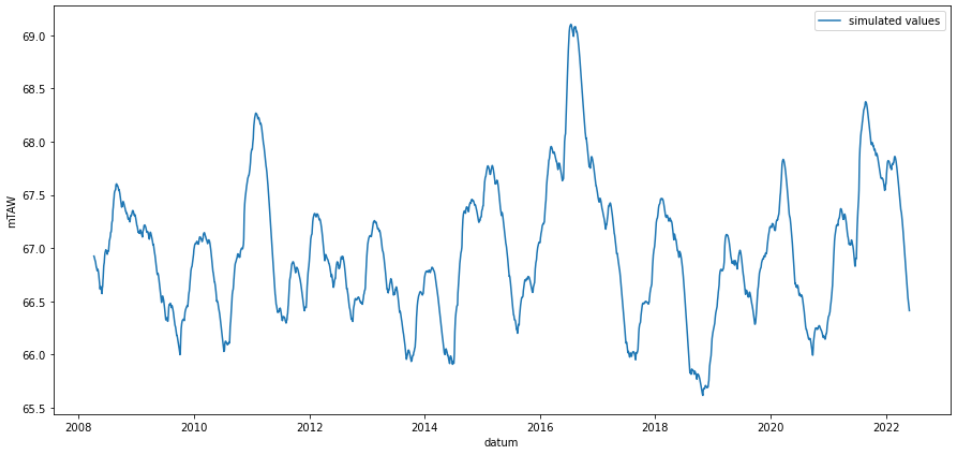
Deze overblijvende set aan locaties wordt gesimuleerd op basis van de PIRFICT-parameters op de tijdstippen van de grondwaterstandmetingen. Enkel de invloed van regenval en evaporatie wordt opgenomen in de simulatie.

Er blijven na de eerder vermelde filtering 2072 locaties over (Figuur 33). Sommige locaties kunnen zeer goed gesimuleerd worden zoals 'ZWAP253' (Figuur 34 - BOVEN). Door de PIRFICT-simulatie kan zelfs een korte periode gemakkelijk uitgebreid worden tot de hele tijdspanne waarin regenval en evaporatiedata aanwezig zijn (Figuur 34 - MIDDEN). Andere locaties hebben een lagere correlatie. Dit kan te wijten zijn aan bijvoorbeeld een sprong in de data of andere artefacten (Figuur 34 - ONDER). Dergelijke minder goede simulaties zullen gefilterd worden wanneer het criterium verhoogd wordt van 0.5 naar 0.8.



Figuur 33: De weerhouden locaties van de PIRFICT-modelleringen. De kleur geeft de Pearson correlatie van hoe goed de gesimuleerde data de metingen kunnen beschrijven.





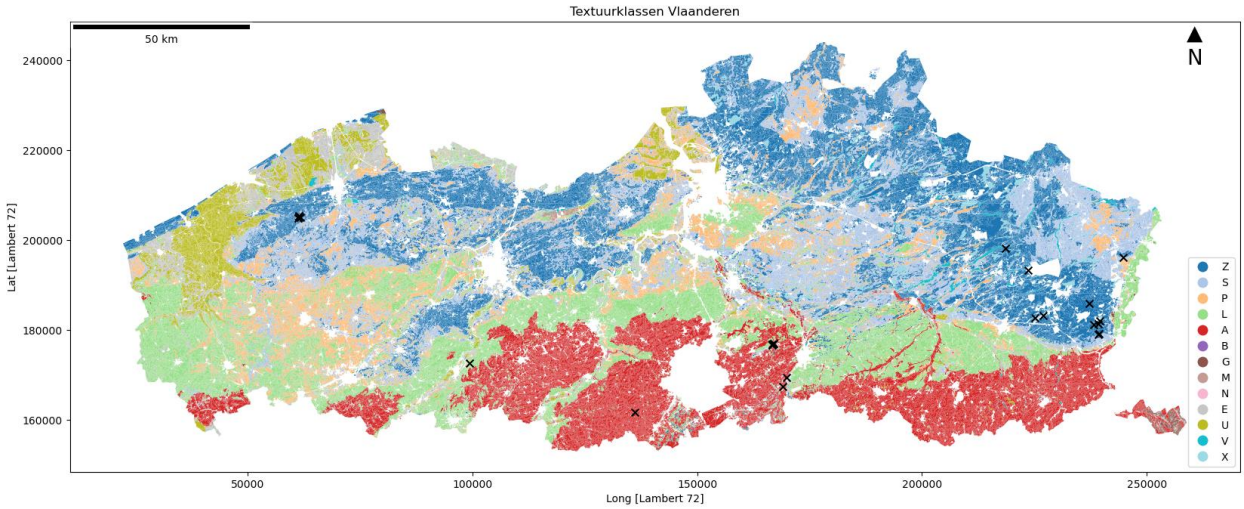
Figuur 34: Simulatie vs gemeten grondwaterstanden op de locatie ZWAP253: een locatie met hoge correlatie van 0.97 (BOVEN), PIRFICT-locatie 1-89079 met lagere correlatie van 0.47 (MIDDEN), PIRFICT-simulatie over de volledige tijdspanne voor de locatie ZWAP253 (ONDER).

5.2.2.3 Toegepast: GxG's berekend op PIRFICT

Vorig hoofdstuk gaf een vergelijking tussen de gesimuleerde en gemeten meetreeksen. Echter, het is ook belangrijk na te gaan hoe goed de GxG's berekend op gesimuleerde reeksen presteren ten opzichte van de daadwerkelijke GxG's (berekend op basis van meetreeksen cfr. de exacte definitie in Van Heesen, 1970).

Hiervoor dienden we een testset op te stellen, maar op slechts 28 locaties (Figuur 35) kon zowel (a) PIRFICT gedraaid worden met voldoende hoge resolutie, als (b) de GxG berekend worden volgens de exacte definitie in Van Heesen (1970). De foutenstatistieken leren ons dat er hier een beperkte fout in de grootteorde van ca. 10 cm geïntroduceerd wordt:

- GLG
    - MAE: 0.11 m
    - MSE: 0.02
    - RMSE: 0.16 m
- GHG
    - MAE: 0.09 m
    - MSE: 0.02
    - RMSE: 0.13 m



Figuur 35: Testlocaties in Vlaanderen waar PIRFICT gesimuleerde GxG's beschikbaar zijn (28 zwarte kruisen).

5.3 INCIDENTELE METINGEN: RELATEREN IN RUIMTE

De 2 methoden (§2.2.2.2) die vermeld zijn in Finke *et al.* (2010) zijn uitgetest om een bijkomende set aan GxG's berekend op incidentele metingen te bekomen. Uiteindelijk is ervoor gekozen om een regressieanalyse uit te voeren tussen de meetlocatie met incidentele metingen en een nabijgelegen meetlocatie met hoge temporele resolutie. Voor deze regressieanalyse zijn de met PIRFICT-gemodelleerde tijdsreeksen aangewend waarbij de PIRFICT-modellering een resultaat met correlatie van >0.5 opleverde. Voor de regressie werd een eenvoudige lineaire regressie toegepast met louter een offset in het geval van een éénmalige incidentele meting of, wanneer het meerdere metingen betrof, een offset en helling.

De andere methodiek, waarbij enkel metingen meegenomen worden die plaatsvonden tijdens de periode waarin de grondwatertafels in theorie de GxG zouden benaderen, stelde een te grote beperking waardoor de uitbreiding van de dataset onvoldoende was. Bovendien ligt de kwaliteit van deze methodiek lager.

5.4 ZOEKTOCHT OPTIMALE HERKARTERINGSTECHNIEK

5.4.1 Inleiding

Zoals vermeld bouwt de vooropgestelde methodiek voort op voorgaande studies van Prof. Finke (o.m. Finke *et al.*, 2005; 2010) waaruit bleek dat herkartering met behulp van geografische hulpdata de beste methode was om de drainageklassen van de bodemkaart te actualiseren.

De toenmalige kostprijs van ettelijke miljoenen euro's bleek niet haalbaar om uit te voeren. Recente vooruitgang qua beschikbare meetgegevens, *remote sensing*, (geo)statistiek, dataminingstechnieken en data-assimilatie en – niet in het minst – computationele rekenkracht, moeten evenwel toelaten om deze oorspronkelijke methodiek te actualiseren en zo niet enkel een accurater resultaat te bekomen, maar dit ook te behalen binnen de vooropgestelde termijn en budget.

De deeltaak herkartering werd dan ook aangevat als onderzoeksproject (onderzoek naar meest optimale methode): hoe kunnen we de recente schat aan remote sensed informatie (gebiedsdekkend beeld, maar met lage resolutie van 1x1km) het best koppelen aan de puntmetingen van het grondwaterpeil? Wordt deze informatie meegenomen als hulpvariabele in de geostatistische regressie van GxG's of is het mogelijk om deze eerst aan te wenden om een tijdreeks aan gebiedsdekkende rasters van het grondwaterpeil te bekomen en deze later statistisch te *downscalen* tot de gewenste resolutie? Na terugkoppeling/brainstorm met de leden van de stuurgroep bleek bovendien dat data-assimilatie waarbij de parametersetting van de 1-D PIRFICT-modellen gekalibreerd wordt via zowel grondwaterpeilmetingen als remote sensed bodemvochtdata, een veelbelovende methodiek is om gebiedsdekkende rasters van het grondwaterpeil te bekomen.

Volgende hoofdstukken behandelen kort iedere methode die is uitgetest, alvorens te landen met de finaal toegepaste methodiek (§5.5).

5.4.2 Data-assimilatie op PIRFICT-parametrisatie

Een eerste stap in het bepalen van de parameters is het fitten van individuele punten zodat we een inschatting kunnen maken van de eigenschappen van de parameters tegenover verschillende bodem- en landeigenschappen. De parameterruimte van de PIRFICT-modellen is echter bijzonder multimodaal (lees: dit is geen lineaire optimalisatie en bevat lokale minima) en de meeste praktische optimalisatiemethoden falen als gevolg hiervan. We kiezen voor een heuristische methodiek om de "correcte" schatting te vinden. We gebruiken een Kleinste Kwadraten methodiek (*scipy.optimize.Least_Squares*) om iedere peilbuis te fitten. Deze methodiek start typisch van een ruwe schatting van de correcte parameters en itereert dan om een minimum te vinden. Dit minimum is echter veelal lokaal.

Door ons model sterk te optimaliseren, kunnen we, praktisch gezien, vele honderdduizenden "runs" per minuut halen, en dus de LSQ-methodiek herhaaldelijk, met verschillende beginpunten, toepassen. Van al deze uitkomsten selecteren we dan diegene met de laagste score.

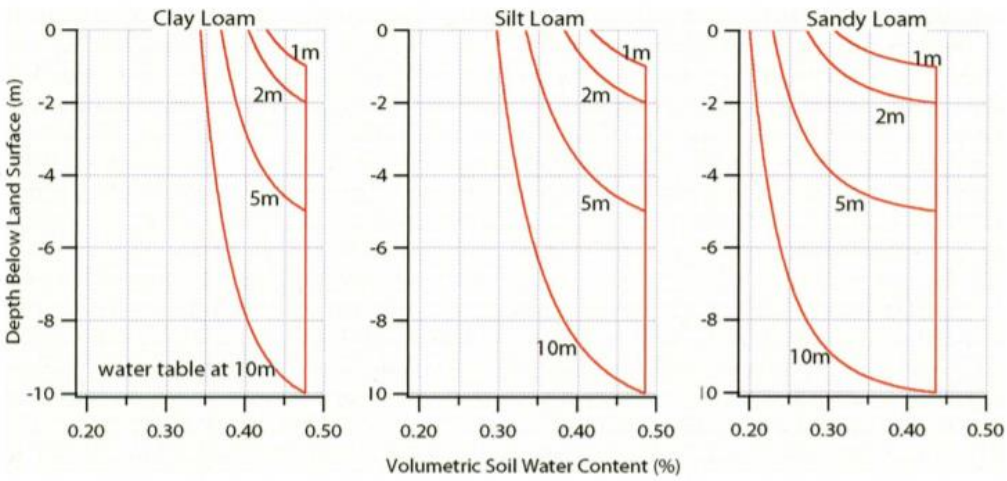
Via de bovenstaande methodiek bekomen we meerdere oplossingen, elk een representatie van een lokaal minimum. Hoewel deze theoretisch suboptimaal zijn, moeten we er rekening mee houden dat we in de praktijk vaak een lokaal minimum zullen hebben. Het valt op dat er enkele fits tussen zitten die bijzonder slecht presteren; de meeste fits vertonen echter een zeer vergelijkbaar resultaat.

Gegeven de bovenstaande resultaten is het opmerkelijk dat sommige geschatte parameters behoorlijk verschillen. We vermoeden dan ook dat het model ofwel niet gevoelig is aan deze parameters (zie hieronder) of dat deze parameters "vervangen" kunnen worden door één van de andere parameters in PIRFICT+.

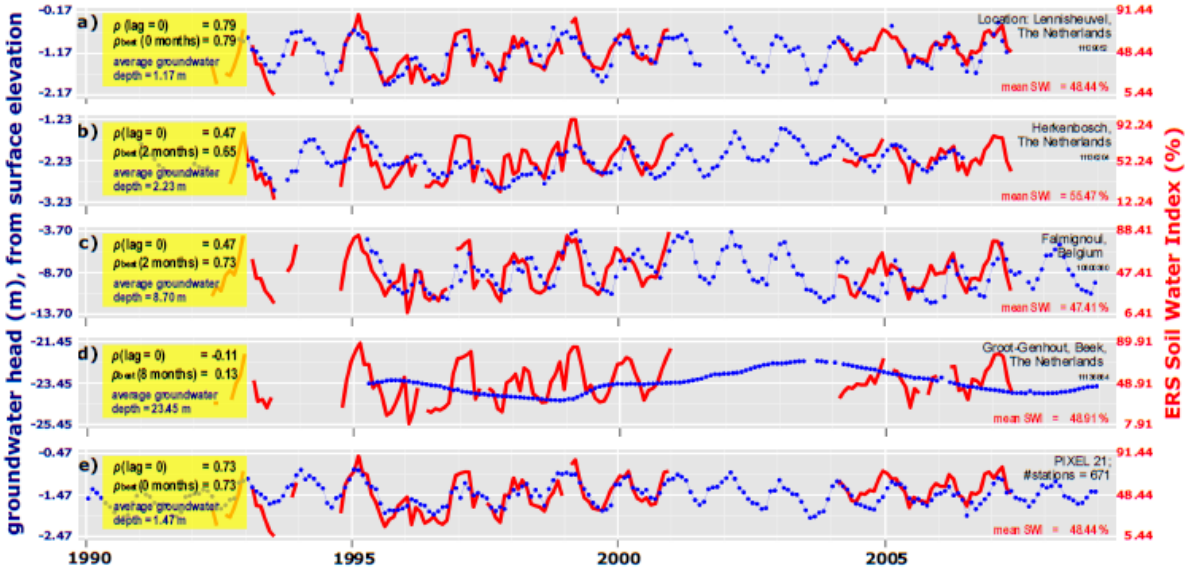
Uiteindelijk bleek deze methode niet robuust genoeg om efficiënt (lees: zonder veelvuldige manuele controle) de correcte parametrisatie te vinden voor een grote set aan grondwaterpeilmetingen, die bovendien te relateren zijn aan bodemkenmerken. Dit deed ons besluiten dat de eerste opzet, *het opstellen van een tijdreeks aan Vlaanderen dekkende rasters van het grondwaterpeil,* niet mogelijk is binnen de voorziene uitvoertermijn, budget en rekenkracht – een aantal van de kritische evaluatiecriteria op de te hanteren techniek.

5.4.3 Link tussen bodemvocht en grondwaterpeil

Men kan zich terecht afvragen of het bodemvochtgehalte wel een goede schatter is voor grondwaterpeilfluctuaties. Gebruikmakend van de Richards vergelijkingen toonden Fan *et al.* (2007) via een simulatiestudie al aan dat de verticale flux in een onverzadigde kolom dusdanig is dat in kleiige/lemige bodems de grondwaterstand het bodemvocht beïnvloedt, zelfs als de grondwaterstand 10 meter diep is (Figuur 36). Deze observatie heeft ervoor gezorgd dat veel *land-surface* modellen tegenwoordig ook een term voor het grondwater bevatten (e.g. Nju *et al.*, 2007). Als logisch gevolg kunnen we dus ook deze relatie omkeren en bodemvocht gebruiken als proxy voor grondwater (Sutanudjaja, 2012). Hij concludeerde dat bodemvocht vooral interessant is voor bodems met een ondiepe grondwaterstand, maar ook van waarde is voor de bodems met een diepere grondwaterstand (Figuur 37). Gezien voorliggende opdracht focust op de grondwatertafels die ondiep genoeg voorkomen om nog een effect uit te oefenen op de bodemkenmerken (afbakening op locaties met gemiddeld hoogste grondwaterstanden ondieper dan 3 m-MV), lijken in voorliggende opdracht deze bevindingen toepasbaar.



Figuur 36: Bodemvochtprofiel boven de grondwatertafel bij 10, 5, 2 en 1 m-MV en voor 3 karakteristieke bodemtypes. (Bron: Miguez-Macho *et al.*, 2007, aangepast uit Fan *et al.,* 2007)



Figuur 37: Vergelijking tussen tijdsreeksen van ERS SWIt (rode lijnen) en grondwatertafel ht (blauwe stippen) op meerdere locaties: Lennisheuvel NL (a); Herkenbosch NL (b); Falmignoul Be (c) en Groot-Genhout Beek NL (d). De onderste reeks geeft een meer gedetailleerde vergelijking tussen SWIt en een aggregatie van gesimuleerde/gemeten grondwatertafels op een specifieke locatie. (Bron: Sutanudjaja, 2012).

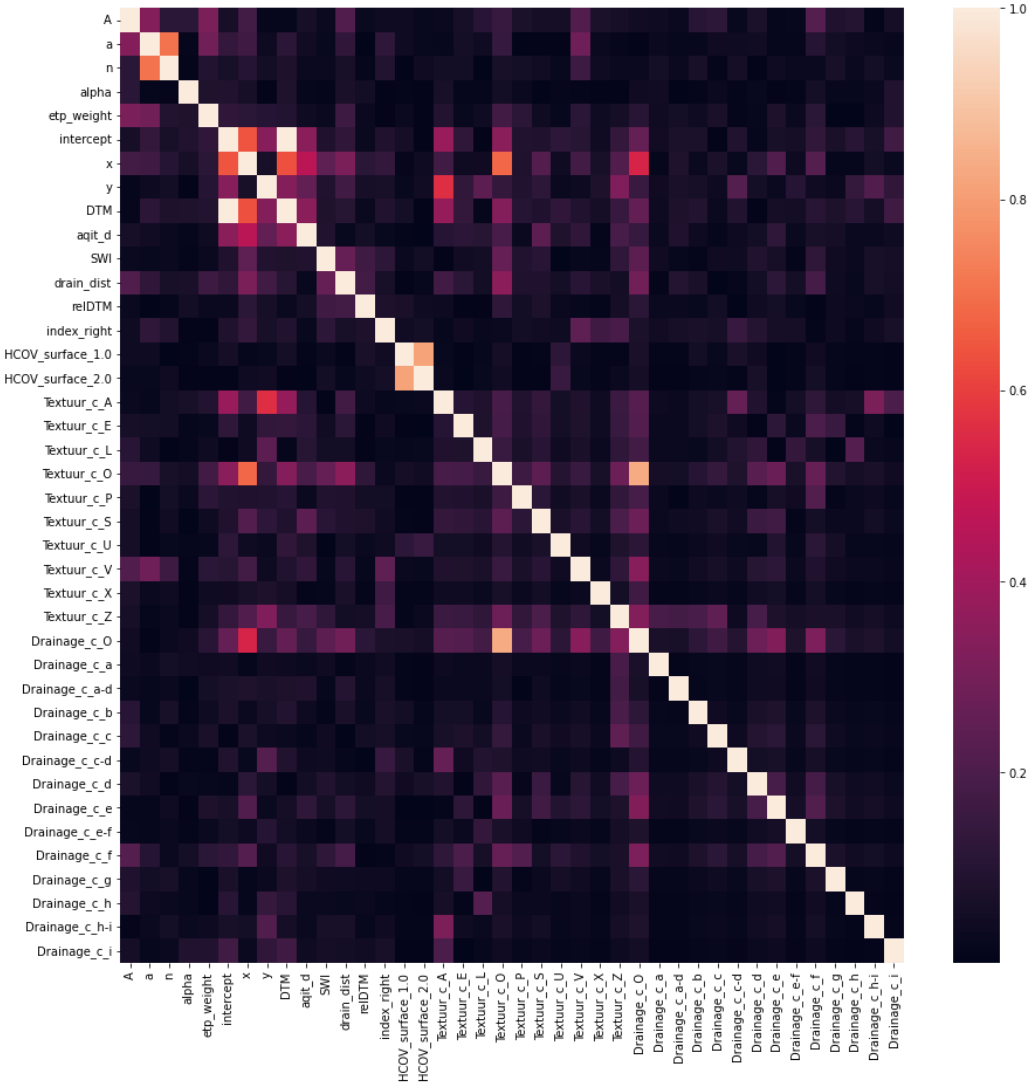
Doch, uiteindelijk bleek de resolutie van de remote sensed SWI te laag om te koppelen aan de meetlocaties: de diversiteit aan metingen binnen een remote sensed pixel bleek te groot om robuuste relaties tussen grondwaterpeil en SWI te bekomen. Er werd daarom besloten om SWI mee te nemen als verklarende variabele ('hulpvariabele') binnen de regressieanalyses.

5.4.4 Regressieanalyse PIRFICT-parameters

Een laatste, alternatieve piste omvatte het schatten van de PIRFICT-parameters a.d.h.v. regressieanalyses op hulpvariabelen. Merk op hoe hier dus geen data-assimilatie wordt opgezet zoals in §5.4.2. De stappen zijn:

1. Neem de parametrisatie van alle eerder gefitte PIRFICT-modellen.
2. Voer voor iedere parameter een regressieanalyse uit a.d.h.v. de geografische hulpdata: omgevingsvariabelen. Deze wordt sequentieel uitgevoerd waarbij de fit van één parameter ook meegenomen kan worden bij een andere.
3. Indien bovenstaande foutenstatistieken acceptabel zijn: maak een gestructureerd raster met vaste resolutie over heel Vlaanderen en draai deze PIRFICT-modellen.

De verkennende regressieanalyse ('Extreme Gradiënt Boosting', zie later) leverde geen goede correlaties tussen de gefitte parameters en de verklarende variabelen: de maximale correlatie was 0,4. De *heatmap* in Figuur 38 presenteert alvast de correlatie tussen de PIRFICT-parameters en de huidige verklarende variabelen.



Figuur 38: Absolute correlatie tussen PIRFICT-parameters en de hulpvariabelen.

5.4.5 Interpolatie/regressie van GxG's

Gezien bovenstaand tegenvallende resultaten, is er beslist om uiteindelijk over te schakelen op een regressiemodel waarmee men direct GxG's kan inschatten en waarbij het remote sensed bodemvocht opgenomen wordt als potentiële verklarende variabele. Het gebruik van statistische regressie om bodem- of grondwatereigenschappen te schatten, vereist (i) een set aan bodemparameters die als voorspellende variabelen (features) kunnen dienen (ii) een zorgvuldig geselecteerde regressiemethode.

Het GxG model werd iteratief ontwikkeld (zie ook Figuur 39):

A. Eerst worden alle databronnen verzameld en naar een standaardformaat gebracht. Dit omvat tijdreeksen van grondwaterpeilen (GWP) en de bijhorende hulpvariabelen. Een overzicht van de databronnen wordt gegeven in hoofdstuk §3.
B. De waarden van de hulpvariabelen op de meetlocaties worden samengebracht. Deze informatiebronnen moeten beschikbaar zijn over heel Vlaanderen om uiteindelijk geschikt

te zijn voor de voorspelling van Vlaanderendekkende GxG's. Op de hulpvariabelen wordt dieper ingegaan in hoofdstuk §4.

C. De GWP-data zijn echter niet altijd van goede kwaliteit of van toepassing voor deze studie. Daarom wordt een logica toegepast om deze data te filteren en, indien nodig, gehele tijdreeksen te verwijderen (hoofdstuk §3).

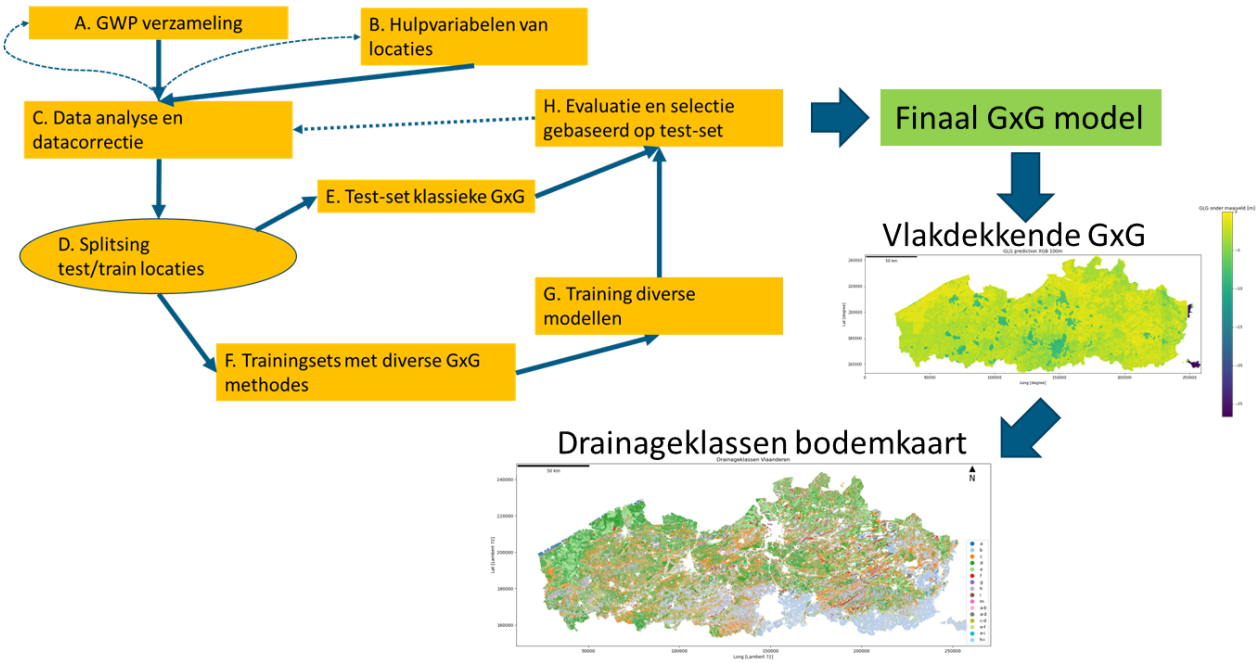
D. Data worden gesplitst in trainings- en testlocaties.

E. De testlocaties moeten zo gekozen worden dat we deze data kunnen aanvaarden als 'ground truth'. Idealiter is de kwaliteit van de GWP meetreeksen dus goed genoeg opdat we rechtstreeks GxG kunnen bekomen zonder af te wijken van de oorspronkelijke definitie. Verder moeten de locaties zo representatief mogelijk zijn voor heel Vlaanderen om de kwaliteit van het model te evalueren.

F. Alle locaties buiten die in de testset, kunnen gebruikt worden om de regressies te optimaliseren. Uit de meeste van deze trainingslocaties kunnen echter niet direct GxG's worden afgeleid omdat de tijdreeksen vaak te weinig data hebben (te lage resolutie) of niet (voldoende) bemeten zijn in de beoogde tijdspanne, na 2010. Hierdoor worden enerzijds technieken gebruikt om meetreeksen op te vullen, anderzijds worden de definities van GxG losser beschouwd, om meer locaties te weerhouden. Deze technieken werden reeds besproken in §5.2. Er worden verschillende technieken gebruikt en om ze te vergelijken worden er telkens nieuwe trainingsets aangemaakt. In de finale modelleringstap (punt 6) wordt dan de afweging gemaakt welke (combinatie van) trainingsdata best gebruikt wordt.

G. Naast verschillende trainingsets worden ook verschillende modeltechnieken vergeleken. Deze worden besproken in §5.6.

H. Alle combinaties van trainingsets en modeltechnieken worden vergeleken met behulp van de testset. Gebaseerd op deze analyse worden de beste modellen weerhouden. Indien, na verschillende iteraties, een model als finaal wordt beschouwd, kan het gebruikt worden om vlak dekkende GxG kaarten te maken voor Vlaanderen. Indien het model nog niet goed genoeg is, wordt gekeken waar het model verbeterd kan worden en wordt de iteratie terug gestart vanaf C.



Figuur 39: Schematische weergave werkmethodiek voor data-gedreven modellering.

5.5 OPDELING TEST- EN TRAININGSET

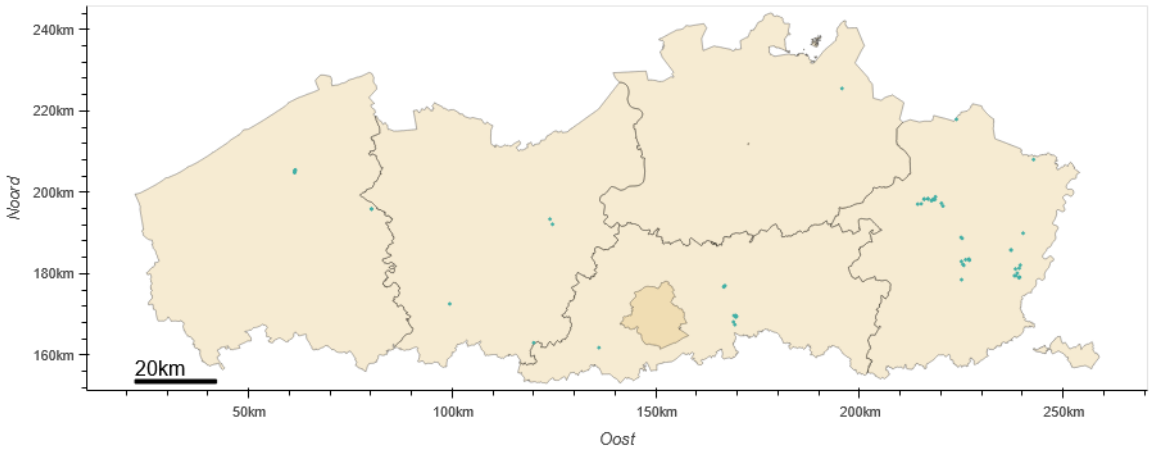
5.5.1 Inleiding

Het uiteindelijke doel van dit onderdeel is om een gebiedsdekkende GxG-kaart te bekomen. Gezien niet voor elk punt in het raster van de gewenste kaart een (bruikbare) meetreeks ter beschikking is, doen we beroep op een model dat geschikt is om een GxG te bepalen op eender welk rasterpunt van de kaart. Om een goede modelleermethode, en later een goed model, te bepalen is het noodzakelijk dat we een testset bepalen. Geteste modellen zullen geëvalueerd worden a.d.h.v. deze testset hetgeen ons toelaat ze met elkaar te vergelijken naar hun vermogen om grondwaterstatistieken te bepalen op onbemeten locaties. De training van deze modellen wordt gedaan op de trainingset, welke nooit overlapt met de testset. Zodanig kunnen betrouwbare en realistische resultaten behaald worden. Er is geopteerd om een beperkte, toch representatieve testset te selecteren om zeker te zijn dat deze meetreeksen bevat waarop de GxG nauwkeurig berekend kon worden (zonder te veel inter- of extrapolatie) zodat geen bijkomende ruis of onzekerheid wordt opgenomen.

5.5.2 Verkenning Testset

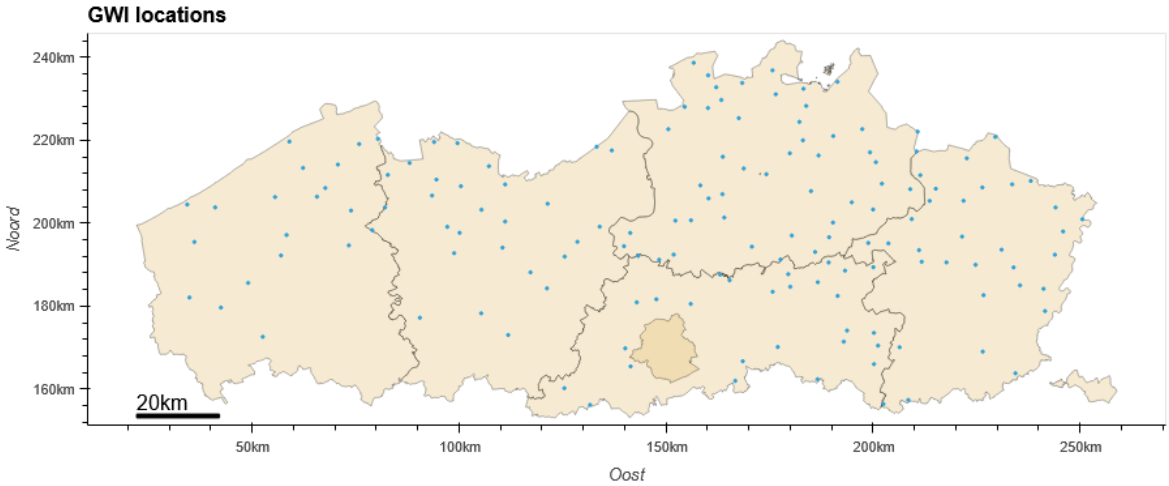
De testset moet een dataset zijn die kan beschouwd worden als een betrouwbare en representatieve 'ground truth'. Men zal namelijk onze modelresultaten steeds aftoetsen en rapporteren aan de hand van deze set.

In eerste instantie werd geopteerd voor locaties afkomstig van DOV waar de klassieke, strikte, GxG definitie kon worden toegepast zonder verdere tijdreeksopvulling. Echter, bij nadere manuele inspectie van deze set blijkt dat deze niet groot en kwaliteitsvol genoeg is om weerhouden te worden als een betrouwbare testset. Na filtering van de tijdreeksen, door middel van datacorrectie en een GHG minder dan 3m diep onder het maaiveld, worden er slechts 59 locaties weerhouden. Deze set is niet representatief voor heel Vlaanderen (Figuur 40).

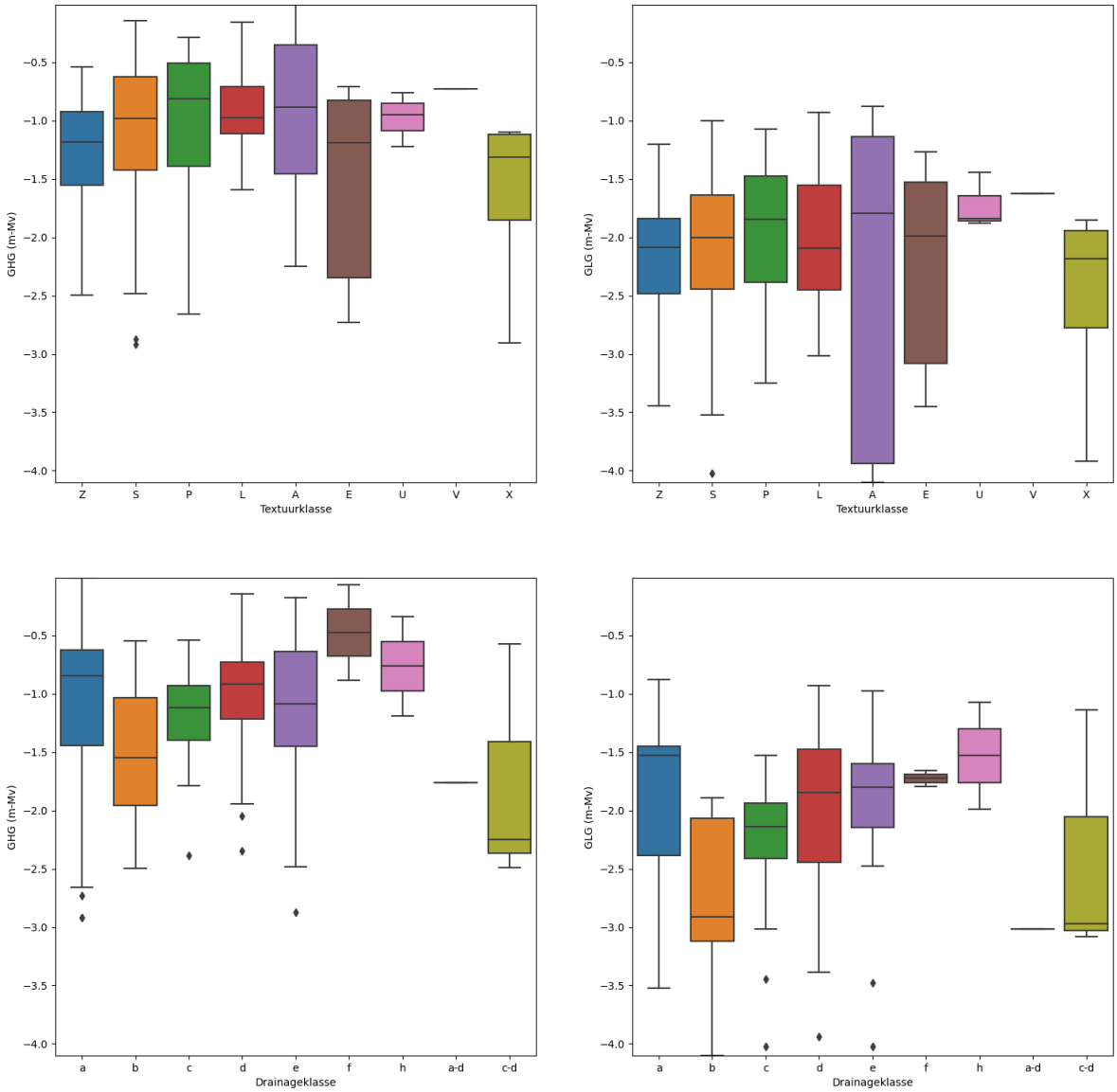


Figuur 40 : Locaties DOV waar de klassieke GxG definitie kan worden toegepast.

Een andere dataset die vaak gebruikt wordt als referentie voor grondwaterstanden in Vlaanderen is de Grondwaterstandsindicator (GWI). Dit is een set van regelmatig bemeten locaties waarop ook een numerieke SWAP modellering is uitgevoerd om, met behulp van data-assimilatie, tot een tijdreeks te komen met dagelijkse waarden. Deze hybride, gemodelleerde en bemeten, tijdreeksen worden beschouwd als een representatieve set om de 'natuurlijke' grondwaterstand, zonder artefacten, binnen Vlaanderen te monitoren. Deze set werd dan ook voorlopig gebruikt als testset voor onze modelevaluatie. In voorliggende studie werden 154 locaties weerhouden (Figuur 41). Figuur 42 geeft de verdeling weer van GxG's berekend op de GWI's in de testset over de oude drainageklassen van de bodemkaart en de textuurklassen.



Figuur 41 : De 154 locaties van de grondwaterindicator.



Figuur 42 : Boxplots van de GxG's per oude drainage- en textuurklasse van de bodemkaart op de testset.

De posities van de GWI's zijn zo geselecteerd dat deze zo min mogelijk beïnvloed zijn door menselijke ingrepen (cf. website DOV). Merk echter op dat er zich volgens de formule van Lebbe *et al.* (2004) toch 36 locaties in de theoretische invloedstraal van grondwaterwinningen bevinden en bovendien liggen er 16 locaties in het bodemtype OB (bebouwing of menselijke invloed). Dit betekent dat de testset dus toch posities omvat die representatief zijn voor andere locaties in Vlaanderen onderhevig aan menselijke beïnvloeding.

Verder werden er 60 punten toegevoegd om een betere spreiding te bekomen: 18 in drainageklasse a, 4 in e-f, 16 in g en 2 in h-i. De reeksen van al deze punten zijn manueel nagekeken om zeker te

zijn dat deze geen artefacten bevatten. Ten slotte werden er 17 GWI locaties naar de trainingsets overgeheveld, zodat er ook informatie van GWI meetreeksen in de trainingset zit.

5.5.3 Trainingset

Om een afweging te maken van hoever we kunnen afwijken van de klassieke GxG definitie ten voordele van meer weerhouden locaties in de trainingset – en bijgevolg meer metingen om het model op te trainen – werd er een gevoeligheidsanalyse uitgevoerd:

- Enerzijds worden er stapsgewijs minder strenge GxG definities gehanteerd, door bijvoorbeeld slechts 7 jaar metingen nodig te achten i.p.v. 8, of slechts 1 meting per maand.
- Anderzijds testten we verschillende opvultechnieken (zie §5.2).

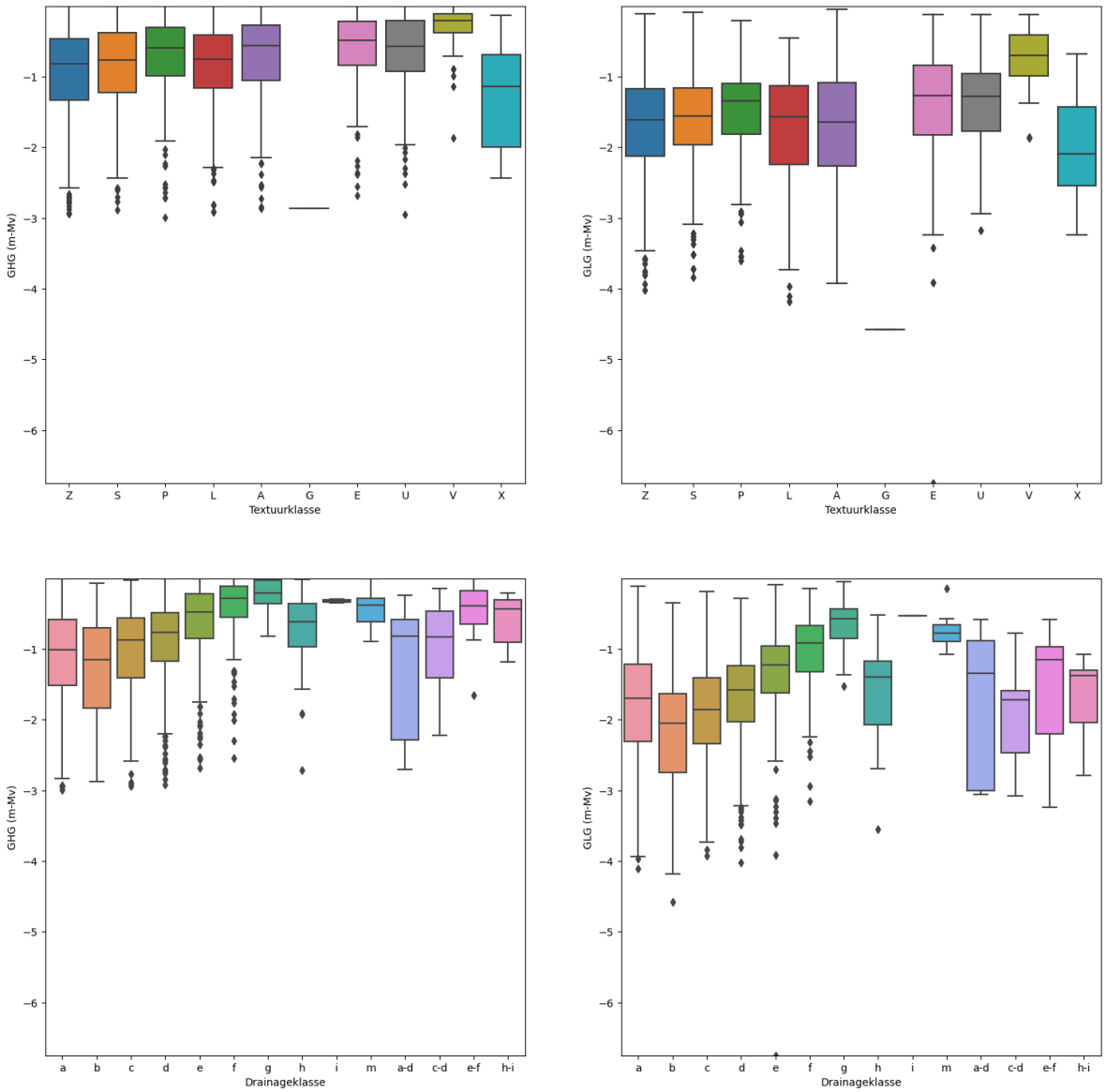
Een volledig overzicht van alle onderzochte trainingsets en het respectievelijk aantal locaties wordt gegeven in Tabel 9. Per dataset werd een beperkte exploratieve analyse uitgevoerd om te kijken naar de correlaties tussen de variabelen en de GxG verdelingen tussen de textuurklassen en oude drainageklassen van de bodemkaart. Hier viel de grote diversiteit direct op (Bijvoorbeeld Figuur 43).

Welke trainingset het beste resultaat oplevert, kan sterk afhangen van de toegepaste regressietechniek. Beide gevoeligheidsanalyses zijn daarom samen uitgevoerd. Voor de resultaten wordt dan ook verwezen naar hoofdstuk §5.6.

Tabel 9 : Overzicht gebruikte trainingsets en aantal locaties

Datasets	Aantal Meetlocaties
2X/maand 8j	59
2X/maand 7j	79
2X/maand 6j	118
2X/maand 5j	176
2X/maand 4j	293
2X/maand 3j	443
1X/maand 7j	146
1X/maand 6j	226
1X/maand 5j	364
PIRFICT filter 0.5	2121
PIRFICT filter 0.6	1889
PIRFICT filter 0.7	1571
PIRFICT filter 0.8	1145
Incidentele metingen	1043
Incidentele metingen filter 0.2	389
2X/maand 8j +PIRFICT0.5	2156
2X/maand 8j +PIRFICT0.5+Incid	2725
2X/maand 8j +PIRFICT0.6	1934
2X/maand 8j +PIRFICT0.6+Incid	2529
2X/maand 8j +PIRFICT0.7	1622
2X/maand 8j +PIRFICT0.7+Incid	2267

Datasets	Aantal Meetlocaties
2X/maand 8j +PIRFICT0.8	1200
2X/maand 8j +PIRFICT0.8+Incid	1932
2X/maand 8j +PIRFICT+Incid0.2	2294



Figuur 43 : Boxplots van GxG's tegenover de oude drainageklassen en textuurklassen van de bodemkaart. De dataset is 2X/maand 8j +PIRFICT0.8+Incid.

5.6 REGRESSIETECHNIEK

Deze stap omvat de kern de opdracht: de GxG's worden berekend op de gemeten grondwaterpeilen, waarna er een regressieanalyse wordt uitgevoerd om deze te koppelen aan Vlaanderendekkende

hulpinfo. Meerdere regressietechnieken zijn geëvalueerd om na te gaan welke techniek de beste resultaten oplevert: de foutenstatistieken werden berekend op de testset. De evaluatie focuste op de GLG, voor de GHG werd dezelfde techniek aangewend. Bij deze laatste wordt namelijk de GLG als bijkomende verklarende variabele toegevoegd, waardoor er vermeden wordt dat de GHG onder de GLG komt te liggen. Volgende hoofdstukken gaan kort in op de theorie achter deze technieken. Het resultaat wordt besproken in § 5.6.5.

De hieronder gehanteerde modellen bevatten allemaal zogenaamde 'hyperparameters'. Dit zijn parameters die de "vorm" van het model beïnvloeden, bijvoorbeeld het type regularisatie in §5.6.2 of het aantal schatters in §5.6.4 en zelfs de gekozen set hulp-variabelen. Deze vallen buiten de normale optimalisatie van een model, maar bepalen de prestaties van een model sterk. Hier zijn ze verkend doormiddel van herhaalde tests op gekozen sets van deze hyperparameters.

De gekozen sets zijn telkens een raster van alle mogelijke hyperparameters binnen een redelijke range, gekozen op basis van expert-judgement. Elk van deze parameters sets is getest door de trainingset op te splitsten in K blokken. De optimalisatie wordt dan K maal herhaald, waar elk van deze blokken eenmaal het blok is waartegen getest wordt, en de overige blokken de trainingset vormen. Het gemiddelde van de K test scores is dan de uiteindelijke score voor de gekozen set hyper-parameters, waar uiteindelijk de hyperparameters met de beste score gekozen zal worden. Wanneer twee sets hyperparameters nagenoeg dezelfde score hebben, zal de set die tot het minst complexe model leidt gekozen worden.

5.6.1 Ordinary Least Squares (OLS)

De eerste regressietechniek is een standaard lineaire regressie: Ordinary Least Squares (OLS). Hierbij wordt de gekwadrateerde fout van het model geminimaliseerd. De te minimaliseren functie is dus:

$$\sum_{i=1}^{n}(y_i - \sum_j x_{ij}\beta_j)^2$$

Waarbij y_i de eigenlijke waarde is (GxG), x_{ij} de waarden van de hulpvariabelen, β_j de te optimaliseren modelcoëfficiënten, i de index van de observatie en j de index van de variabele. Het voordeel van deze methode is haar simpliciteit en snelheid.

Het nadeel van deze techniek is dat er zich bij de optimalisatie problemen voordoen door de sterke multicollineariteit van hulpvariabelen. Dit was al zichtbaar in Figuur 30 waar verschillende verklarende variabelen onderling een hoge (absolute) correlatie hebben. Hierdoor zijn de coëfficiënten, standaardfouten en generaliseerbaarheid van deze modellen niet betrouwbaar. Daarom is er geopteerd om na de verkenning over te schakelen op de geregulariseerde lineaire techniek (§5.6.2) en OLS niet mee te nemen in de evaluatie.

5.6.2 Geregulariseerde OLS (Elastic Net)

Door de hoge multicollineariteit is er geopteerd voor een geregulariseerde lineaire regressie (*Elastic Net*). Dit is een standaardmethode om binnen de lineaire regressie een variabele selectie uit te voeren. Een aantal gewichten (coëfficiënten) zullen dan tot 0 herleid worden door een 'penalty' in te voegen bij optimalisatie voor de (absolute) grootte van de modelcoëfficiënten. Aangezien dit ook lineaire modellen zijn, kunnen de coëfficiënten nog steeds geraadpleegd worden en zijn ze meer betrouwbaar dan de gewone OLS indien multicollineariteit zich voordoet. Echter, hier valt geen betekenisvolle standaardfout op de coëfficiënten te onderscheiden.

Elastic Net is een lineaire regressiemethode die zowel L1- als L2-regularisatie combineert. Deze regularisatiemethoden worden ingezet om overfitting te verminderen, vooral wanneer er veel features zijn in verhouding tot het aantal observaties of wanneer veel features met elkaar gecorreleerd zijn.

- L1-regularisatie: Dit is ook bekend als Lasso (*Least Absolute Shrinkage and Selection Operator*) regularisatie. Het voegt een penalty toe aan de absolute waarden van de modelcoëfficiënten. De L1-regularisatie kan ertoe leiden dat sommige coëfficiënten van het model exact nul worden, wat betekent dat het model die specifieke features volledig negeert. Dit heeft als resultaat een model dat minder complex is en gemakkelijker te interpreteren.
- L2-regularisatie: Dit wordt ook *Ridge regularisatie* genoemd. Het voegt een penalty toe aan de kwadraatwaarden van de modelcoëfficiënten. In tegenstelling tot L1, zal L2 de coëfficiënten dicht bij nul duwen, maar ze zelden exact nul maken.
- Elastic Net combineert de regularisaties van L1 en L2 om van beide voordelen te profiteren. Indien pure L1 regularisatie wordt gebruikt, is de minimalisatiefunctie een uitbreiding op die besproken in §5.6.1 waarbij er een penalty toegevoegd wordt voor de absolute waarde van de coëfficiënten. Dit heet dan Lasso regressie:

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

De geregulariseerde regressie bevat standaard 2 hyperparameters die moeten afgestemd worden binnen de optimalisatie: de regularisatie penalty (λ) en de L1-L2 weging. Deze laatste geeft aan wat de verdeling is van de L1 t.o.v. de L2 regularisatie: 1 is 100% L1 (Lasso regressie), 0 is 100% L2 (Ridge regressie).

5.6.3 Geografisch Gewogen Regressie (GWR)

De Geografisch Gewogen Regressie (*Geographic Weighted Regression*, GWR) is een (semi-) lineair model waarbij de modelcoëfficiënten kunnen variëren in de ruimte. De ruimtelijke variatie van de coëfficiënten wordt apart gemodelleerd. Dit is een vaak gebruikte techniek in lineaire geografische regressie die typisch betere resultaten geeft dan OLS (*Ordinary Least Squares*).
Het nadeel van deze techniek is dat deze niet goed kan omgaan met multicollineariteit, het best werkt met herschaalde variabelen en de optimalisatie niet altijd slaagt indien de datamatrix niet inverteerbaar is.
Om deze obstakels te voorkomen is gewerkt met een beperkte set aan hulpvariabelen. Na het bekijken van de correlaties is geopteerd voor DTM, mTAW van de locatie (DTM_abs), de grondwatertafel (GWT) en relatieve DTM (relDTM)

5.6.4 Extreme Gradiënt Boosting (XGB)

'*Extreme Gradiënt Boosting'* (XGB) is een boostingtechniek gebaseerd op beslissingsbomen waarbij iteratief nieuwe beslissingsbomen worden getraind en toegevoegd op de residuen van het vorige model. Het voordeel van deze techniek is dat deze minimale datavoorbereiding vergt (bv. geen imputatie of herschaling), complexere interacties tussen de variabelen incorporeert, kan omgaan met niet-lineaire effecten, en kan omgaan met multicollineariteit.
Verder bevat de python implementatie van deze techniek een geoptimaliseerde integratie voor de berekening van shapely waarden ('SHAP'). SHAP-waarden zijn *Shapely Additive exPlanations*, afgeleid uit de speltheorie, die toelaten om een machinaal model enigszins verklaarbaar te maken. Deze shapely waarden zijn de marginale (lineaire) bijdragen van elke variabele op een

modelpredictie. Om dit te bekomen, wordt de predictie van het model vergeleken met een predictie voor een input waarbij een variabele is vervangen door zijn verwachte (gemiddelde) waarde. Zo kan dus voor elke predictie de bijdrage van elke variabele worden ingeschat. Deze analyse mitigeert de 'black box' ervaring van machine-learning technieken die hogere interacties en non-lineaire relaties bevatten. Deze shapely waarden bieden ook een handige basis om eenvoudig belangrijke lineaire effecten te herkennen en verdere feature-selectie uit te voeren bij verdere modelleringen.
Een nadeel van XGB is dat, vooral bij kleine datasets, er sterke over-fitting kan plaatsvinden. Hierbij zal, door het hoge aantal trainbare parameters, het model sterk geënt zijn op de trainingsdata en moeilijk generaliseren naar observaties hierbuiten. Met meer variabelen zal dit sneller voorkomen. Aangezien we hier een relatief groot aantal variabelen hebben en een niet al te groot aantal observaties moeten we hiervoor dus waakzaam blijven.

5.6.5 Resultaat

XGB komt, op basis van RMSE, MAE en MSE, als beste regressietechniek uit de bus (Tabel 10). Om betere foutenstatistieken te bekomen, diende de trainingset voldoende groot te zijn. Bij GWR en XGB geeft de grootste dataset de beste resultaten: zowel de incidentele meetputten, de grootste set van PIRFICT-simulaties (met lage filter) en de locaties met de klassieke GxG berekening worden opgenomen. Bij de elastic net (ELN) bleek het behoud van enkel de incidentele metingen tot het beste resultaat te leiden.

Tabel 10: Foutenstatistieken berekend op de testset per modeltechniek, incl. aanduiding meest optimale trainingset die hoort bij deze modeltechniek. De beste resultaten zijn weergegeven in het rood.

Model	MAE GLG [m]	MSE GLG	RMSE GLG [m]	Pearson R GLG	MAE GHG [m]	MSE GHG	RMSE GHG [m]	Pearson R GHG	Trainingset
XGB	0.49	0.41	0.64	0.72	0.41	0.32	0.56	0.65	StdformuleGxG+PIRFICT0.5+Incid
GWR	0.59	0.59	0.77	0.57	0.50	0.41	0.64	0.51	StdformuleGxG+PIRFICT0.5+Incid
ELN	0.62	0.68	0.82	0.42	0.47	0.40	0.64	0.35	Incidentele metingen

5.7 KWANTIELMODEL

Om een onzekerheid van GxG predicties te bekomen voor iedere rastercel, wordt gebruik gemaakt van kwantielregressies. Hierbij wordt een Gradient Boosting-type model gebruikt aangezien deze de beste resultaten gaf in §5.6.5. Ook worden dezelfde hulpvariabelen gebruikt in deze kwantielregressies. De optimalisatie (loss) functie wordt hier aangepast zodat in plaats van een verwachte waarde, een kwantiel van verwachte waardes wordt ingeschat. Bijvoorbeeld een 90% kwantiel model geeft inschattingen waar, over het algemeen, 90% van de echte waarden onder zou moeten zitten en 10% boven, gegeven de trainingsdata. In de praktijk kunnen er vele modellen gefit worden voor veel verschillende kwantielen. Hier beperken we ons tot 5%, 50% en 95% kwantielen om onzekerheden in te schatten.

5.8 Van GxG's naar drainageklassen van de bodemkaart

Deze stap betreft de vertaling van het GxG-raster naar een nieuwe drainageklasse van de bodemkaart. Eén van de wensen was om in deze kaart de bestaande polygonen uit de oude bodemkaart maximaal over te nemen. De intekening van de polygonen dateert van de analoge

bodemkaart. Gezien het algemeen geweten is dat de ruimte in Vlaanderen drastisch is veranderd, is het waarschijnlijk dat in de meeste polygonen de heterogeniteit sterk is toegenomen. Dit dient men in het achterhoofd te houden wanneer men de drainageklasse van de bodemkaart onderzoekt. In hoofdstuk §7 wordt dieper ingegaan op de bruikbaarheid van het GxG-raster versus de kaart gebaseerd op polygonen.

De drainageklasse van de bodemkaart wordt vervolgens als volgt bekomen:

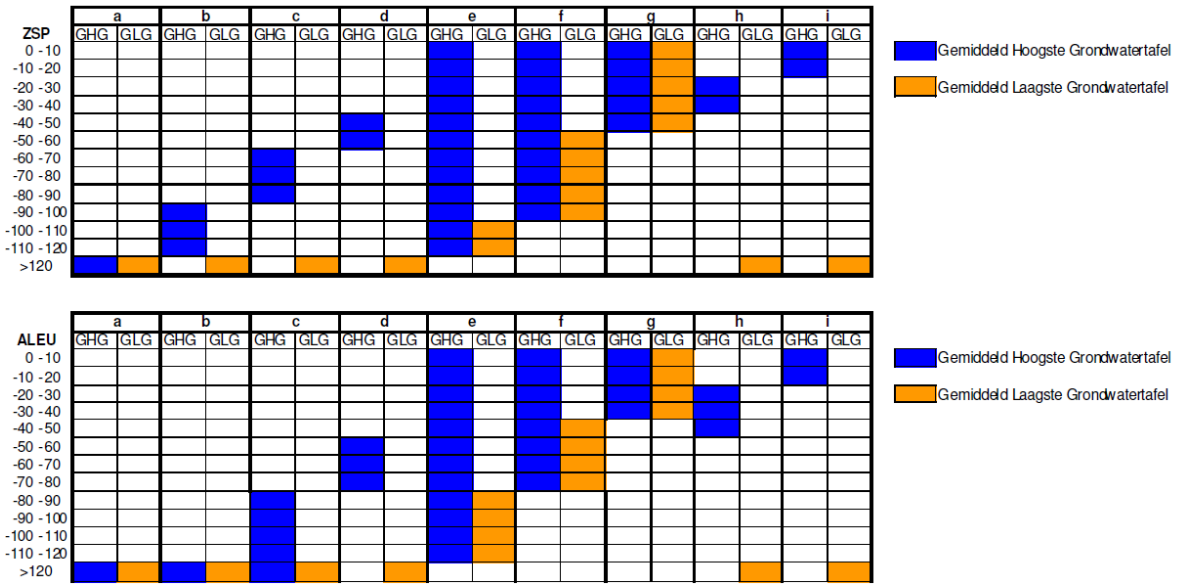
1. Voor iedere polygoon van de bodemkaart wordt een frequentietabel opgesteld van de GxG-cellen die binnen deze polygoon vallen.
2. Vervolgens wordt de mediaan van alle gesimuleerde GxG's overgenomen als GxG van de polygoon en de typische percentielen (25ste-, 75ste-, 5de-, en 95ste-) als bijkomende maat van onzekerheid. Deze laatste worden ook meegegeven als attribuut.
3. Om de drainageklassen te bepalen, werd in eerste instantie de legende van *Van Ranst & Sys* (2000; Tabel 11) gebruikt. Echter, omdat bepaalde combinaties van GHG en GLG met deze legende niet klassificeerbaar zijn (bijvoorbeeld: een GHG van 50 cm met een GLG van 75 cm diepte), is een aangepaste legende gebruikt naar het voorbeeld van Finke (2010; zie Figuur 44).

Tabel 11: Natuurlijke drainageklasse (Van Ranst & Sys, 2000)

Draineringsklasse	Diepte in cm op de welke beginnen			
draineringsgraad	de gleyverschijnselen		de reductiehorizont	
	in de volgende materialen			
symbool	lemig kleiig	zandig	lemig kleiig	zandig
. a . te sterke drainering	-	>120	-	-
. b .gunstige drainering	-	90-120	-	-
. c .matige drainering	> 80	60-90	-	-
. d .onvoldoende drainering	50-80	40-60	-	-
. e .matig slechte drainering (*)	20-50	20-40	> 80	> 100
. f . slechte drainering (*) 0-20	0-20	40-80	50-100	
. g .zeer slechte drainering (*)	0	0	< 40	< 50
. h .matig slechte drainering (**)	20-50	20-40	-	-
. i . slechte drainering (**) 0-20	0-20	-	-	

(*) Hydromorfe gronden met permanente grondwatertafel en reductiehorizont.

(**) Hydromorfe gronden met tijdelijke stuwwatertafel en zonder reductiehorizont.



Figuur 44 Legende vertaling GxG naar drainageklassen (Finke, 2010). De afkortingen staan voor: Z=Zand, S=Lemig Zand, P=Licht Zandleem, A=Leem, L=Zandleem, E=Klei, U=Zware Klei

De opdeling van (a) hydromorfe gronden met permanente grondwatertafel en met reductiehorizont en (b) deze met tijdelijke stuwwatertafel en zonder reductiehorizont werd tijdens de loop van het project regelmatig besproken. Volgende mogelijkheden bestaan:

1. Neem deze van de vorige kaart rechtstreeks over.
2. Relateer het voorkomen aan de nieuwe kaart met voorkomen van plastische gronden en meer informatie over verdichting.
3. Stel een kans op voorkomen op via een nieuwe regressieanalyse.

Het detecteren van hangwater is echter lastig. De diepte van filters is hierbij gevoelig omdat zelfs ondiepe filters eventueel aanwezig hangwater net zouden kunnen missen. Om bovendien hangwater op polygoonniveau te detecteren, is een fijnmazige sampling nodig, hetgeen niet beschikbaar is. Voorts werden de mogelijke oorzaken voor het ontstaan of verdwijnen van hangwater besproken. Informatie over bodemverdichting in Vlaanderen zou hiervoor nuttig kunnen zijn, maar deze info is niet met voldoende detail – en Vlaanderendekkend – voorhanden. Er is daarom besloten om deze uit de drainageklassen van de eerdere bodemkaart over te nemen.

6 BEVINDINGEN

6.1 BEVINDINGEN METHODOLOGIE

6.1.1 Data-gedreven methodiek

De oude drainageklasse van de bodemkaart kwam tot stand dankzij een dicht-bemeten dataset gebaseerd op reductiehorizonten. De techniek in voorliggende studie klasseren we onder een data-gedreven techniek die de drainage-eigenschappen afleidt uit grondwaterstanden: een herkartering, conform de voorkeur van de meerderheid aan respondenten tijdens de gebruikersinventarisatie uitgevoerd door Finke *et al.* (2010). Data-gedreven technieken hebben het voordeel van snelheid, schaalbaarheid en de mogelijkheid om patronen te identificeren uit grote datasets. Een data-gedreven techniek is derhalve goedkoper en niet beperkt in ruimtelijke dekking. Ze hangt anderzijds wel af van de kwaliteit en representativiteit van de gebruikte data. Bovendien kunnen categorische variabelen en/of variabelen met relatief lage resolutie, leiden tot harde grenzen en bijgevolg plotse sprongen tussen de GxG-waarden van nabijgelegen rastercellen. Een groot voordeel van de hier ontwikkelde techniek, is de maat van onzekerheid die berekend werd voor iedere individuele rastercel met gemodelleerde GxG's.

Reductiehorizonten verschaffen inzicht in historische anaerobe omstandigheden in de bodem, wat duidt op slechte drainage. Ze kunnen indicaties geven over de frequentie en duur van deze omstandigheden. Aan de andere kant hebben ook deze hun beperkingen; ze geven geen actuele informatie en sommige bodems kunnen anaerobe omstandigheden hebben zonder duidelijke reductiehorizonten. In tegenstelling hiermee bieden grondwaterstanden real-time informatie over de actuele diepte van het grondwater. Dit kan direct worden gerelateerd aan drainage-eigenschappen en kan over de tijd worden gemonitord. Echter, hun waarden kunnen variëren en worden beïnvloed door externe factoren zoals nabijgelegen waterstromen of waterwinning. In het licht van deze verschillen, zijn grondwaterstanden doorgaans een directere indicator voor actuele drainage-eigenschappen. Maar voor een alomvattende evaluatie is het combineren van beide indicatoren met aanvullende bodeminformatie raadzaam.

6.1.2 Herkarteringstechnieken

Tijdens deze opdracht zijn meerdere herkarteringstechnieken uitgetest. Uiteindelijk bleek de regressie op Grondwaterstatistieken (GxG's) de methode die het beste resultaat oplevert voor de vooropgestelde doeleinden: zo accuraat mogelijk, Vlaanderendekkend en op hoge resolutie de ruimtelijke verdeling van grondwaterstatistieken simuleren, mits de methodiek en gerelateerde computationele kost toelaten om, telkens wanneer er voldoende nieuwe data beschikbaar zijn, de drainageklassen van de bodemkaart vlot te actualiseren. De ontwikkelde techniek is zo opgesteld dat de (toekomstige her-)rekening voor ieder kaartblad afzonderlijk kan worden uitgevoerd (Figuur 17).

6.1.3 Regressietechnieken

Er zijn meerdere regressietechnieken, alsook test- en trainingsets geëvalueerd.

Voor iedere uitgeteste regressietechniek bleek het noodzakelijk om af te wijken van de GxG-definitie. Er blijven zo namelijk onvoldoende locaties over waar men GxG's voor kan berekenen. Een

dataset waar zoveel mogelijk trainingspunten werden weerhouden bleek onmisbaar om deze oefening te doen slagen. Dit kan door de onvoldoende (frequentie en/of duur) bemeten reeksen verder aan te vullen met oftewel PIRFICT-modelleringen oftewel door te relateren aan nabijgelegen gesimuleerde reeksen (§5.2).

Verder bleek het ook nodig om een geautomatiseerde techniek te ontwikkelen waarbij anomalieën in de dataset worden gedetecteerd. De uiteindelijk toegepaste techniek staat beschreven in §3.1.3.

Op basis van RMSE kwam de regressietechniek *Extreme Gradiënt Boosting* – getraind op de locaties met strikte GxG-definitie, aangevuld met GxG's berekend op PIRFICT-simulaties en incidentele metingen – als beste uit de bus (§5.6.5). Dit is een techniek die een enorm brede waaier aan patronen kan modelleren, waarbij weinig rekening moet gehouden worden met de vorm, distributies en correlaties van de input hulpvariabelen. Bovendien kan *Extreme Gradiënt Boosting* goed omgaan met non-lineaire effecten. Door gebruik te maken van SHAP-waarden, kan er een gewicht gegeven worden aan het belang van iedere hulpvariabele.

Een ander nadeel van deze techniek is dat er mogelijks overfitting kan gebeuren. Hierbij wordt een model te veel geoptimaliseerd op een bepaalde dataset waardoor deze te weinig generaliseert naar nieuwe observaties. Dit wordt voorkomen door te werken met een test- of validatieset om respectievelijk modelprestaties te evalueren buiten de trainingset en om de juiste hyperparameters te selecteren gebaseerd op generalisatievermogen. Om de dataset maximaal te benutten, werd er gebruik gemaakt van een testset om keuzes te maken over modeltechnieken en trainingsets en k-fold validatie voor hyperparameter selectie.

6.2 BEVINDINGEN VERKLARENDE VARIABELEN

De SHAP-waarden (§5.6.4), zoals gepresenteerd in Figuur 45, geven meer inzicht in het relatieve gewicht van iedere verklarende variabele. Merk op hoe, ten gevolge van multicollineariteit, features in onderlinge concurrentie gaan binnen dit model. Het model kan hiermee om, maar de SHAP-waarden kunnen enkel voor de individuele features berekend worden. Gezien het model meerdere afgeleiden van SWI omvat, zal de totale bijdrage van SWI hoger liggen dan hier aangegeven. Zoals eerder aangegeven wordt de GLG opgenomen om de GHG te voorspellen.

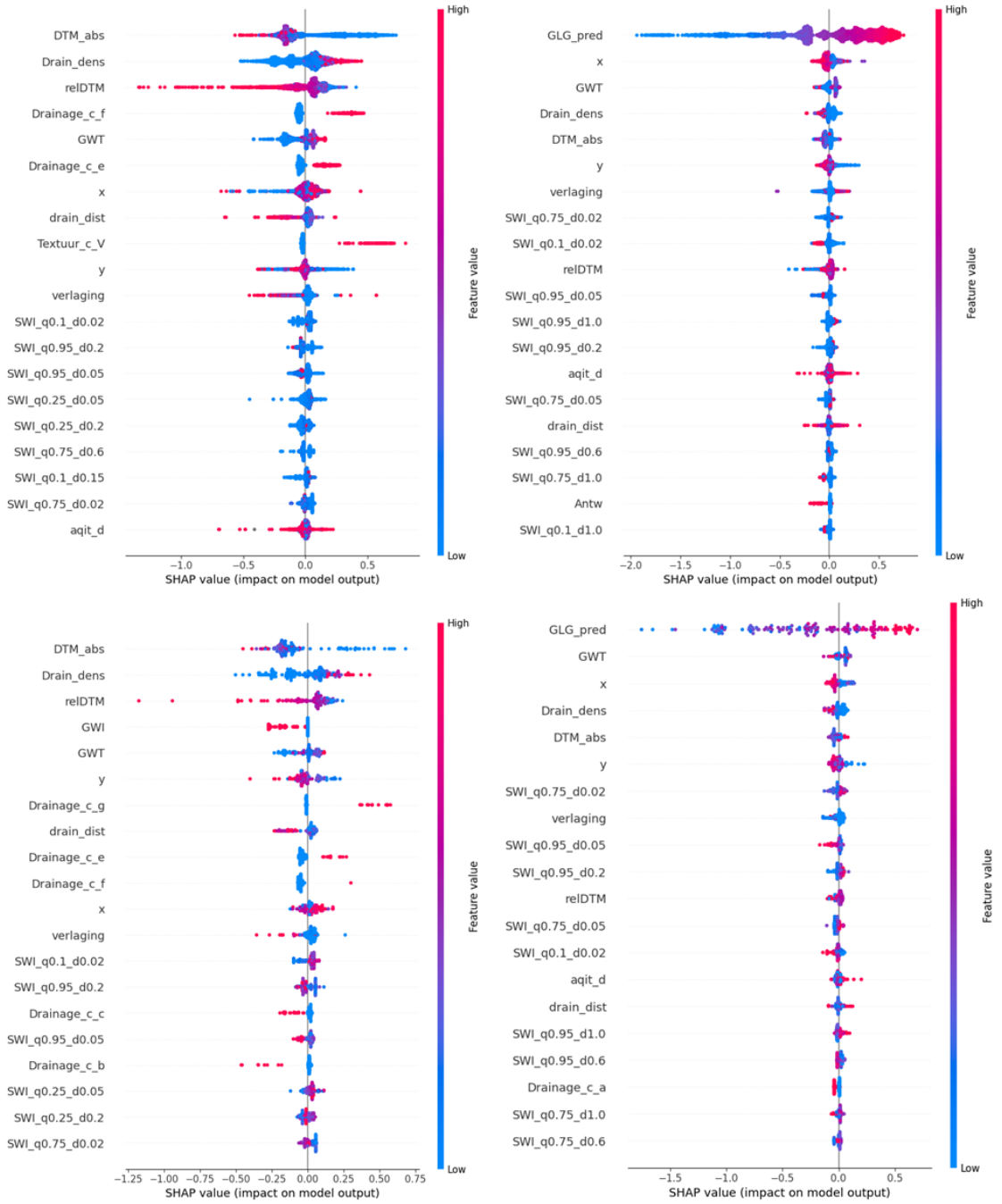
De belangrijkste features in het **XGB-model om GLG te simuleren** zijn (geordend van belangrijkst naar minder belangrijk):

1. Hoogte van het maaiveld (§4.3: *DTM_z*)
2. Drainagedensiteit (§4.5.3)
3. Relatieve maaiveldhoogte (§4.3: *relDTM*)
4. Specifieke drainageklasse 'f' (§4.1: *Drainage_c_f)*
5. Drooglegging (§4.5.2: *GWT*)

Merk op hoe in enkele observaties, dus lokaal, ook de grondwaterwinning (§4.7: *verlaging*) een duidelijke rol speelt.

Voor het **GHG-model** geven de belangrijkste features een ietwat vertekend beeld omdat de voorspelling van GLG al de impact van meerdere variabelen capteert. De GLG is dan ook de belangrijkste variabele. De features die bijkomend impact hebben op de afwijking tussen GHG en GLG zijn dan (zie Tabel 6 voor uitleg van de variabelen):

1. De cartografische variabelen (x- en y-coördinaat of positie langs de Noord-Zuid en Oost-West as). Dit betekent dat de features gecapteerd in het model, nog niet alle heterogeniteit omvatten die noodzakelijk is om de GxG's te simuleren. Mogelijke verklaringen zijn: (a) er zijn bijkomende features die hier niet vervat zitten, (b) de ruimtelijke variabiliteit binnen de gecapteerd features is onvoldoende ten gevolge van de spatiale resolutie of, vermoedelijk, (c) een combinatie van beide.
2. Diepte tot de bovenste hydrogeologische laag met geringe hydraulische doorlatendheid (§4.4.2: *aqit_d*)
3. Afstand tot de dichtstbijzijnde waterloop (§4.5.1: *drain_dist*)
4. De relatieve maaiveldhoogte (§4.3: *relDTM*)
5. Drooglegging (§4.5.2: *GWT*)
6. SWI met zijn componenten (§4.6: *SWI_...*).



Figuur 45: Resultaten van SHAP-waarden in de GxG predictie met XGB op de trainingset (2X/m 8j + PIRFICT + incidentele metingen) (Boven) en de testset (Onder). De variabelen zijn gerangschikt volgens afnemend algemeen belang.

6.3 BEVINDINGEN GxG-KAART

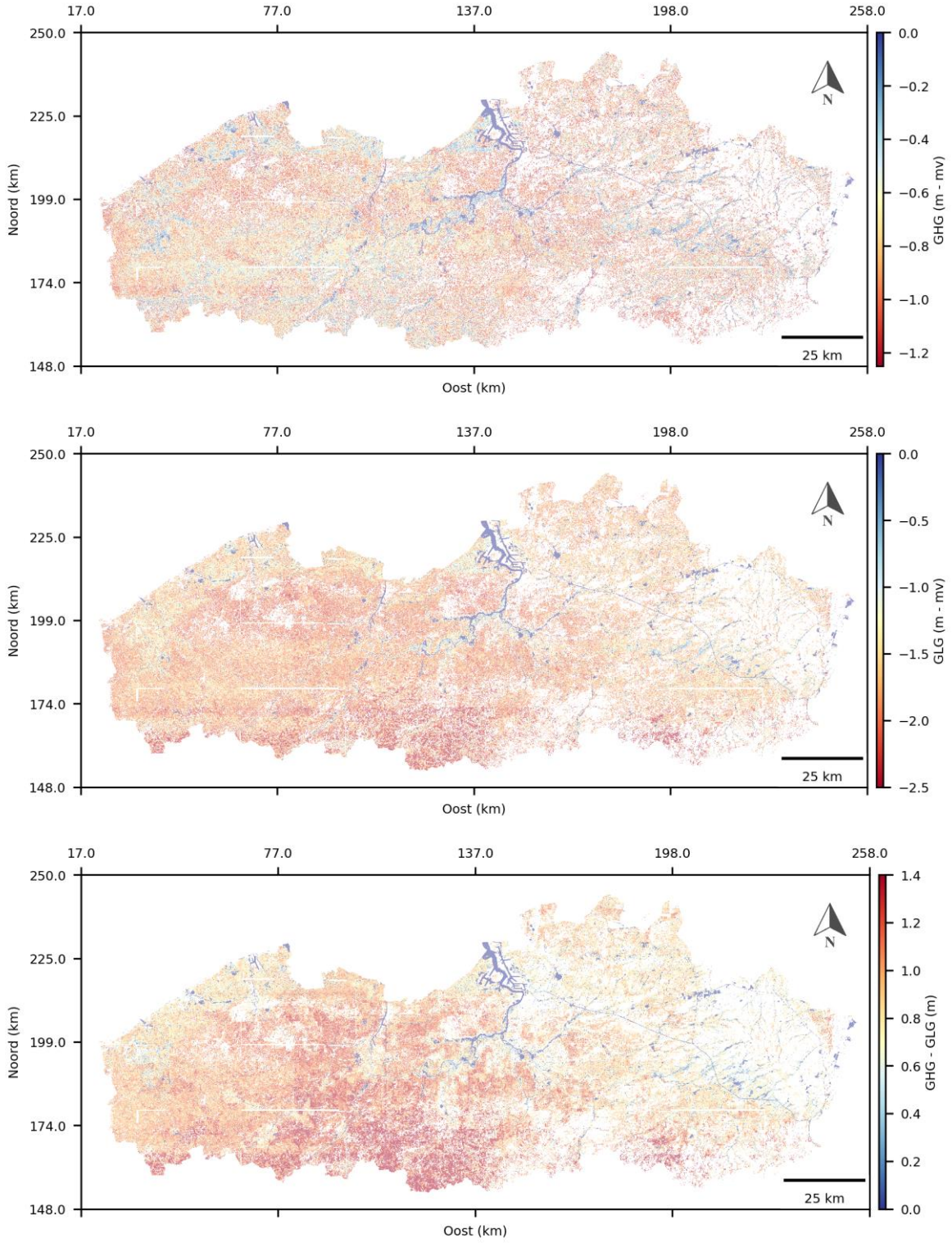
6.3.1 Vlaanderendekkende GHG- en GLG-kaarten

Het primaire doel van dit onderzoek was het vervaardigen van GxG-kaarten die volledig Vlaanderen dekken. Dit, in tegenstelling tot de drainageklassen van de bodemkaart, welke enkel informatie bood op polygoonniveau. In Figuur 46 worden de resulterende kaarten gepresenteerd die gefilterd werden voor die locaties waar de GHG-voorspelling ondieper is dan 125cm onder het maaiveld. Dit is namelijk de diepte tot waar het bodemprofiel wordt bestudeerd. Diepere grondwaterstanden

worden automatisch geclassificeerd onder een en dezelfde drainageklasse, zijnde drainageklasse “a”. De kaart heeft een resolutie van 10 bij 10 meter.

Een bijkomend voordeel van de nieuwe GxG-kaarten is de introductie van kwantitatieve variabelen die fluctuaties in het grondwaterniveau weergeven. De drainageklasse van de bodemkaart biedt slechts categorische indicatoren die historische tendensen van droogte of nattigheid op een locatie aangeven. Er werd beslist om tijdens deze studie enkel de meetreeksen mee te nemen waarbij de GHG ondieper ligt dan 3m onder het maaiveld. Analyse van de trainingsdata heeft aangetoond dat deze beslissing effectief uitschieters in de voorspelde waarden elimineert. De implicaties, zowel positief als negatief, van deze beslissing zullen verder in dit hoofdstuk worden besproken.

In de bodemkaart werden verscheidene regio's (verstedelijkte gebieden, kustpolders) bovendien niet voorzien van een drainageklasse. Deze lacunes werd in de nieuwe versie overwonnen, waardoor de huidige kaarten een significant verbeterde informatievoorziening bieden in vergelijking met de bodemkaart.

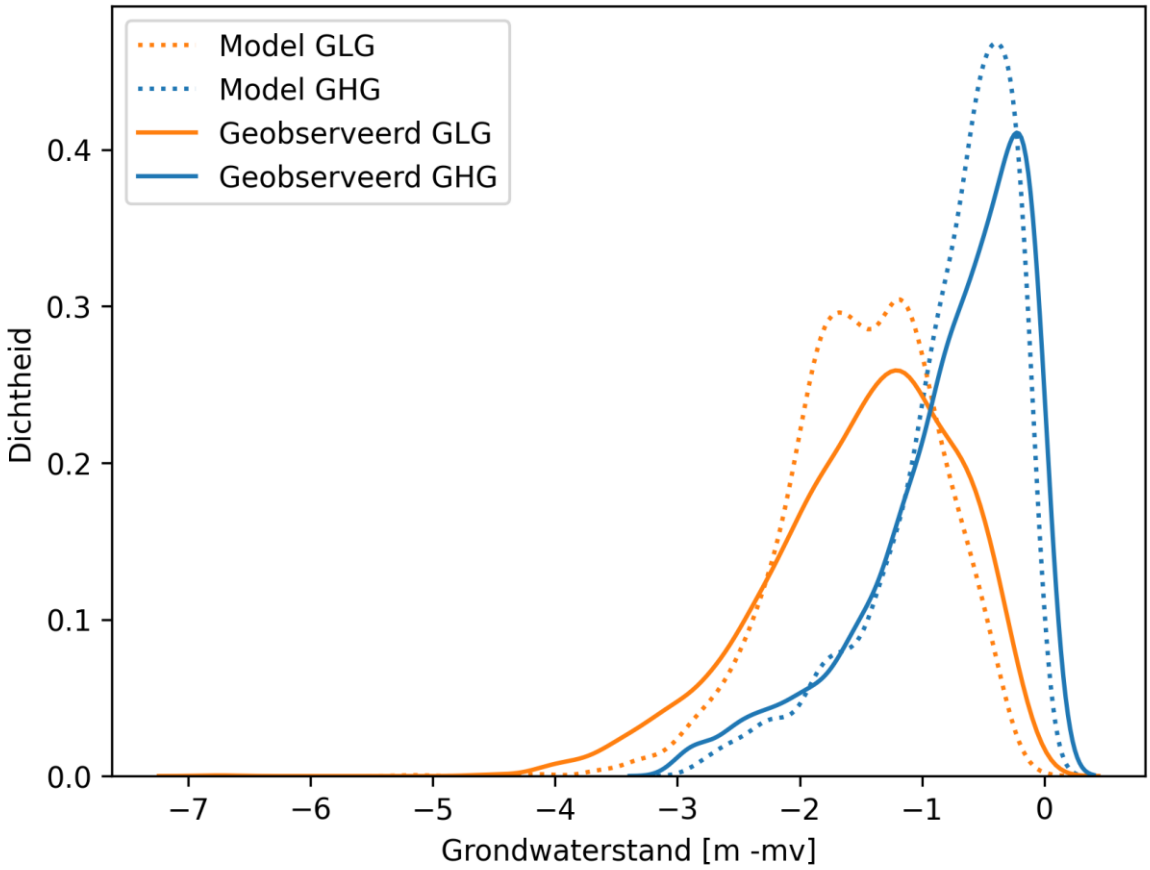


Figuur 46: Resulterende GxG-kaarten: gesimuleerde, Vlaanderendekkende GHG (BOVEN), GLG (MIDDEN) en verschilkaart tussen beide (ONDER)

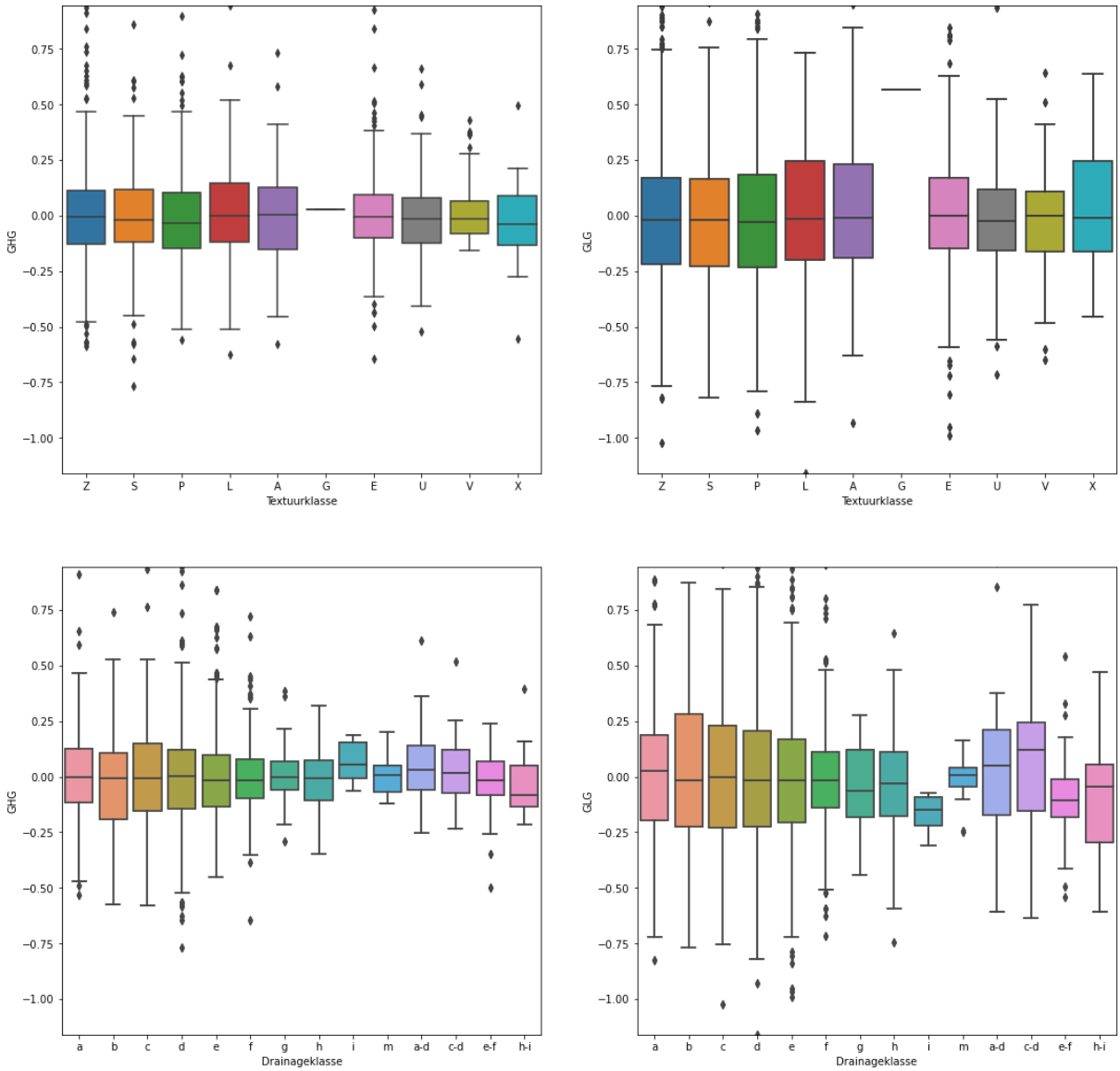
6.3.2 Histogrammen en residuen

We bekijken als eerste de histogrammen van de voorspellingen ten opzichte van de geobserveerde histogrammen. Figuur 47 presenteert de histogrammen, weergegeven als een Kernel-Density-plot, voor zowel de GHG als de GLG. Beide dichtheden zijn goed vergelijkbaar met de geobserveerde dichtheid van het GHG en GLG. Wel observeren we een lichte verschuiving naar ondiepere GHG's en een bredere kansverdeling voor de GLG's.

De residuen van de voorspellingen per drainage- en textuurklasse, van beide GxG's staan gepresenteerd in Figuur 48. Als eerste valt op dat zowel de GHG's als de GLG's nagenoeg geen systematische afwijking vertonen ten opzichte van de geobserveerde waarden, behalve voor de drainageklassen complexen. Dit is eenvoudig te verklaren door een bijzonder lage representatie in zowel de trainings- als testset. Interessant hier is om op te merken dat voor alle klassen de afstand tussen het 25ste en 75ste (de "box") percentiel veel kleiner is dan de afstand tussen het 5de en 95ste percentiel (de "whiskers"). Dit kan duiden op een sterk gepiekte verdeling.

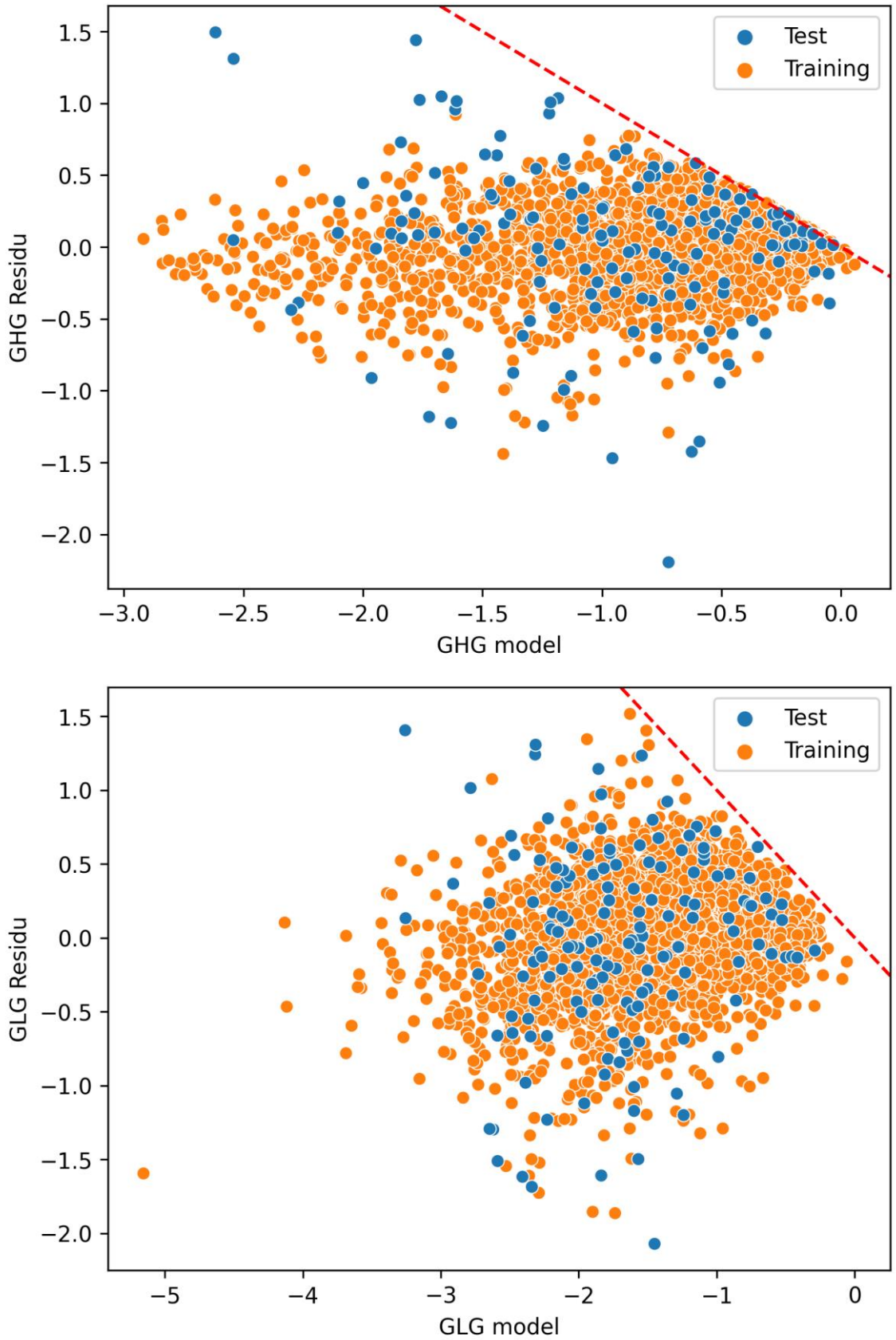


Figuur 47: De histogrammen, weergegeven als een kernel density plot, van de GHG (blauw), en de GLG (oranje). De lijn stijl geeft aan of het de observaties (solide) of de voorspelling (punten) betreft. Merk op dat deze histogrammen op basis van de gehele dataset (training + test) bepaald zijn.



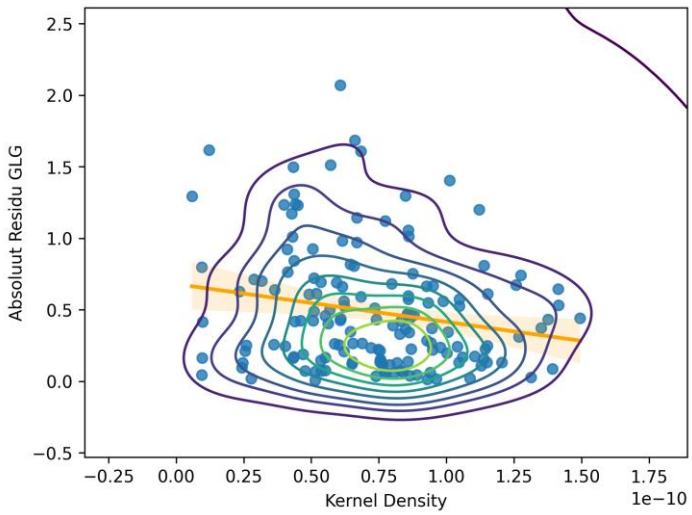
Figuur 48: Boxplots van de residuen per drainage- en textuurklasse.

Inderdaad, we zien in Figuur 49 dat er geen sterke trends zichtbaar zijn, met uitzondering van een heteroskedastische verdeling van de residuen.



Figuur 49: Vergelijking van de residuen met de modelwaarden, voor de GHG (BOVEN) en GLG (ONDER). De oranje punten zijn de GWI-waarden, de blauwe zijn de volledige dataset.

Verder zouden we graag de vraag beantwoorden of een dichter meetnet zou aanleiding geven tot betere modelinschattingen. Niettegenstaande grotere trainingsdatasets vrijwel altijd leiden tot betere resultaten willen we graag weten of we betere modelschattingen bekomen indien er een hoge densiteit aan geografisch naburige metingen zijn. Dit doen we door een *Gaussian kernel density* uit te rekenen op basis van de X- en Y-coördinaten van de trainingset voor X- en Y-coördinaten van de testset en te kijken of er een significante relatie is met het absoluut residu. In Figuur 50 is te zien dat er een beperkte relatie is, maar deze is niet-significant. De data wijzen dus op een beperkte verbetering indien het meetnet geografisch dichter zou worden.



Figuur 50 Het absoluut residu op de GLG i.f.v. de Gaussian Kernel Density

6.3.3 RMSE

Tabel 12 en Tabel 13 presenteren de RMSE per textuurklasse en drainageklasse voor respectievelijk de GHG en GLG. De fout op de GLG die het model maakt op de testset, valt veelal tussen de 20 en 35cm. Onder aanname van een normale fout-verdeling betekent dit dat 68% van de voorspelde waarden in de gekozen categorie een fout kleiner dan de RMSE heeft. Tevens valt het op dat de hoogste fouten meestal in categorieën vallen met weinig datapunten (Tabel 14). De fout op de GLG ligt hoger.

Gegeven het beperkt aantal meetpunten en gezien de diversiteit in drainage- en textuurklassen, wordt ook de fout op de trainingset meegegeven. Merk op: de trainingset omvat meer meetlocaties, doch omvat intrinsiek een fout op de GxG: de GxG van de trainingset is niet conform de strikte definitie berekend. De foutenstatistieken die berekend worden aan de hand van de trainingset, zijn bijgevolg de sommatie van modelfouten én fouten/onzekerheden in de berekening van GxG in de trainingset.

Tabel 12: De RMSE op de GHG-voorspelling per textuur- en drainageklasse. De eerste waarde is de RMSE voor de volledige trainingset, de tweede waarde enkel voor de testset.

	A	E	L	P	S	U	Z
a	0.48/0.25	0.16/0.37	0.64/0.37	0.62/0.32	0.95/0.30	0.27/0.29	0.57/0.29
b	0.13/0.21	-/-	-/-	0.08/0.23	-/-	-/-	0.27/0.28

	A	E	L	P	S	U	Z
c	0.04/0.25	-/-	0.24/0.3	0.15/0.29	0.71/0.30	-/-	0.31/0.27
d	0.59/0.33	0.28/0.28	0.34/0.31	0.5/0.22	0.82/0.31	-/-	0.56/0.24
e	0.15/0.28	-/-	-/-	0.41/0.31	0.71/0.2	-/-	0.40/0.26
f	0.51/0.28	-/-	-/-	-/-	-/-	-/-	-/-
g	0.09/0.11	0.11/0.09	0.20/0.07	-/-	0.10/0.12	-/-	0.23/0.08
h	-/-	-/-	-/-	1.44/0.19	-/-	-/-	-/-

Tabel 13: De RMSE op de GLG-voorspelling per textuur- en drainageklasse. De eerste waarde is de RMSE voor de volledige trainingset, de tweede waarde enkel voor de testset.

	A	E	L	P	S	U	Z
a	0.80/0.40	0.18/0.44	0.43/0.65	0.74/0.51	0.90/0.39	0.49/0.46	0.69/0.40
b	1.51/0.48	-/-	-/-	0.64/0.39	-/-	-/-	0.37/0.37
c	1.62/0.55	-/-	0.38/0.38	0.77/0.30	0.42/0.46	-/-	0.53/0.39
d	1.29/0.43	0.33/0.36	0.44/0.48	0.28/0.34	0.7/0.37	-/-	0.54/0.40
e	0.25/0.39	-/-	-/-	0.23/0.36	0.85/0.29	-/-	0.42/0.36
f	0.81/0.49	-/-	-/-	-/-	-/-	-/-	-/-
g	0.19/0.27	0.09/0.16	0.43/0.10	-/-	0.28/0.15	-/-	0.40/0.10
h	-/-	-/-	-/-	1.24/0.14	-/-	-/-	-/-

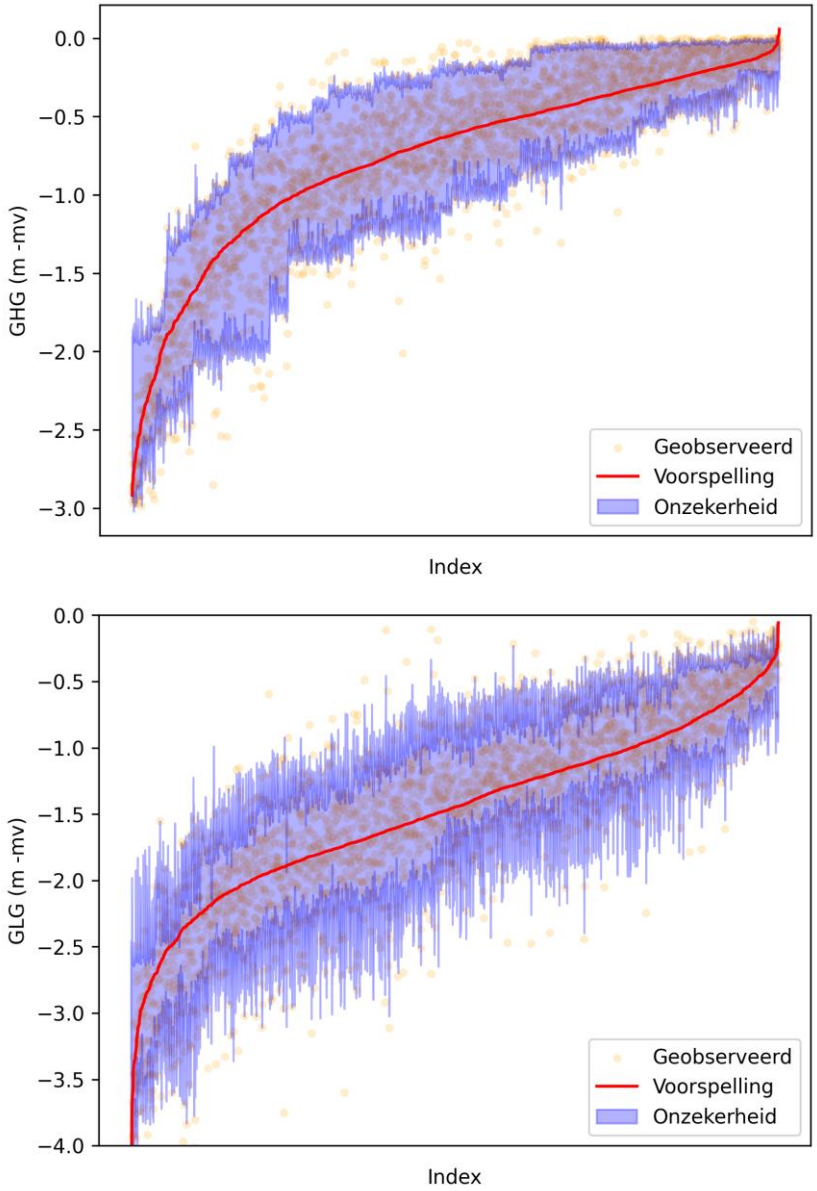
Tabel 14: Het aantal observaties per textuur- en drainageklasse. De eerste waarde is het aantal observaties in de volledige trainingset, de tweede waarde het aantal observaties in de testset.

	A	E	L	P	S	U	Z
a	10/1	31/3	28/1	52/9	28/6	31/3	234/2
b	7/1	-/-	-/-	2/1	-/-	-/-	59/6
c	15/1	-/-	31/2	18/1	49/5	-/-	150/1
d	46/1	68/1	79/3	66/3	133/1	-/-	173/11
e	43/1	-/-	-/-	54/4	107/5	-/-	91/4
f	52/1	-/-	-/-	-/-	-/-	-/-	-/-
g	9/6	22/1	1/1	-/-	3/7	-/-	2/1
h	-/-	-/-	-/-	4/1	-/-	-/-	-/-

6.3.4 Onzekerheid

Figuur 51 toont de resultaten van de kwantiel modellen (§.5.7) waarbij het model duidelijk een grotere fout vertoont voor de diepere GxG's. De figuur bewijst dat het model zijn eigen fout goed in kan schatten, en dat men dus bij de individuele voorspellingen van een GxG ook kan weten hoe betrouwbaar deze zijn.

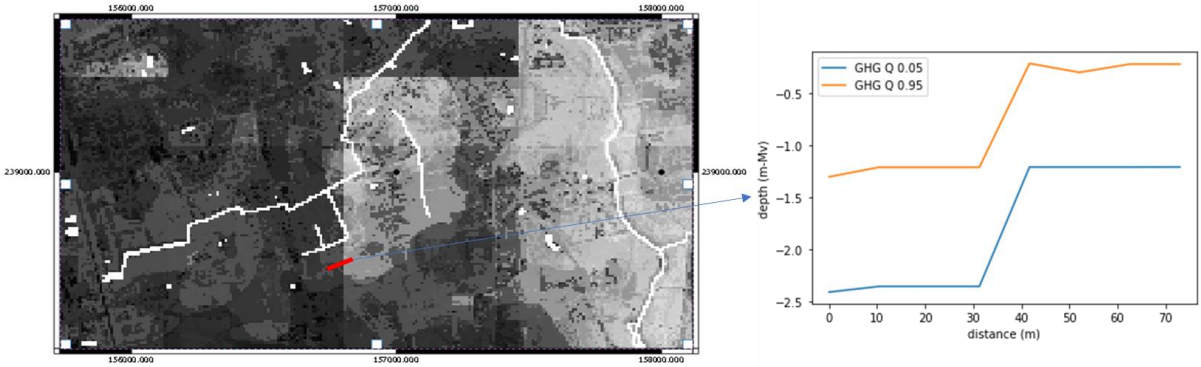
Een dergelijk model zorgt ervoor dat men niet enkel de overkoepelende foutenstatistieken moet vertrouwen (Tabel 12, Tabel 13), maar voor iedere pixel individueel de betrouwbaarheid kan beoordelen.



Figuur 51: Voor zowel GHG (Boven) als GLG (Onder): de onzekerheidsbanden (5-95%), de totale set waarden (oranje punten), de voorspelling (blauwe lijn). De waarden zijn geordend op basis van de door het model voorspelde waarde.

Ten slotte merken we op dat door het relatief grote belang van de SWI-feature in het maken van de voorspellingen tezamen met de relatief grove resolutie van deze feature (in vergelijking met de 10x10m resolutie van de GxG-kaart), er vrij harde overgangen voorkomen op de resolutie van de SWI. We zoomden in op <u>het meest extreme voorbeeld</u> hiervan en geven dit weer in Figuur 52. Het valt op dat er zich bij de overgang tussen twee SWI-tegels een sprong voordoet in GHG van ca. 1,75m. Dit geldt voor zowel het 5de als 95ste percentiel. Dit zorgt dus nog voor een artefact in de ordegrootte van de onzekerheid.

Na zorgvuldige overweging werd ervoor gekozen geen smoothing toe te passen op deze resultaten in de postprocessing stap. Smoothing kan leiden tot een verlies aan detail en variatie in de data. In sommige gevallen is het belangrijk om de exacte waarden op specifieke locaties te kennen, en kan smoothing deze details verbergen. Het zou in dit geval ook het onderliggende patroon maskeren. Bovendien is een smoothing techniek kiezen onderhevig aan subjectiviteit van de expert die de techniek toepast. Beter zou zijn om de SWI te downscalen naar hogere resolutie. De proefmatig uitgeteste methoden, leverden geen, computationeel-acceptabele, resultaten op.



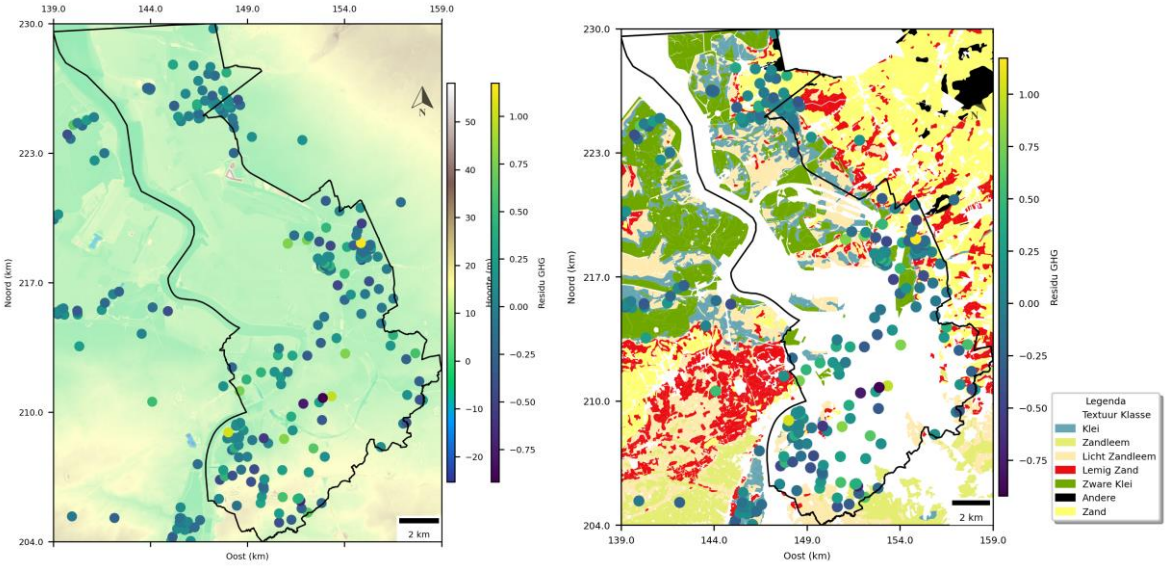
Figuur 52 Voorbeeld van een GHG-voorspelling waarbij de overgang tussen twee SWI-tegels duidelijk zichtbaar is (links). Profielkwantielen (q0.05 - q0.95) van de GHG-voorspelling bij overgang tussen twee SWI-tegels (rechts)

In de modelselectie hebben we geen rekening gehouden met mogelijke artefacten in ruimtelijke patronen. De predicties worden beschouwd als een onafhankelijke beste inschatting op elke locatie gegeven de onderliggende data. Indien die onderliggende data onrealistische ruimtelijke patronen bevatten, zullen deze te zien zijn in de modelpredicties.

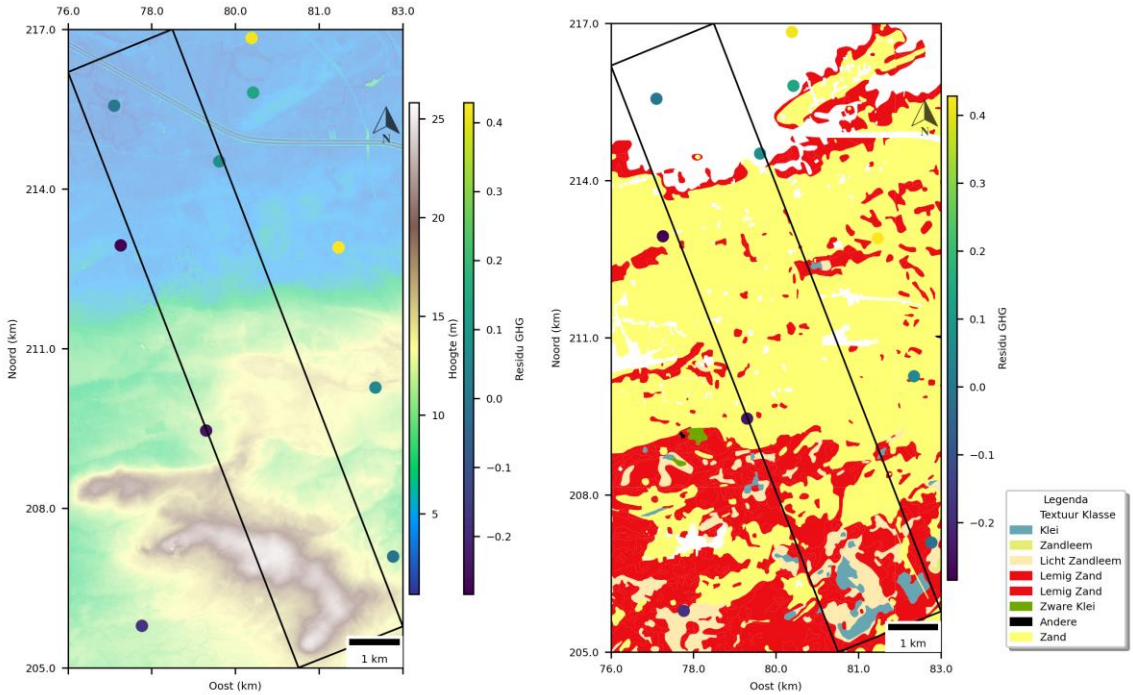
6.3.5 De pilootzones nader bekeken

In samenspraak met de Stuurgroep werden aan het begin van het project enkele pilootzones geselecteerd.

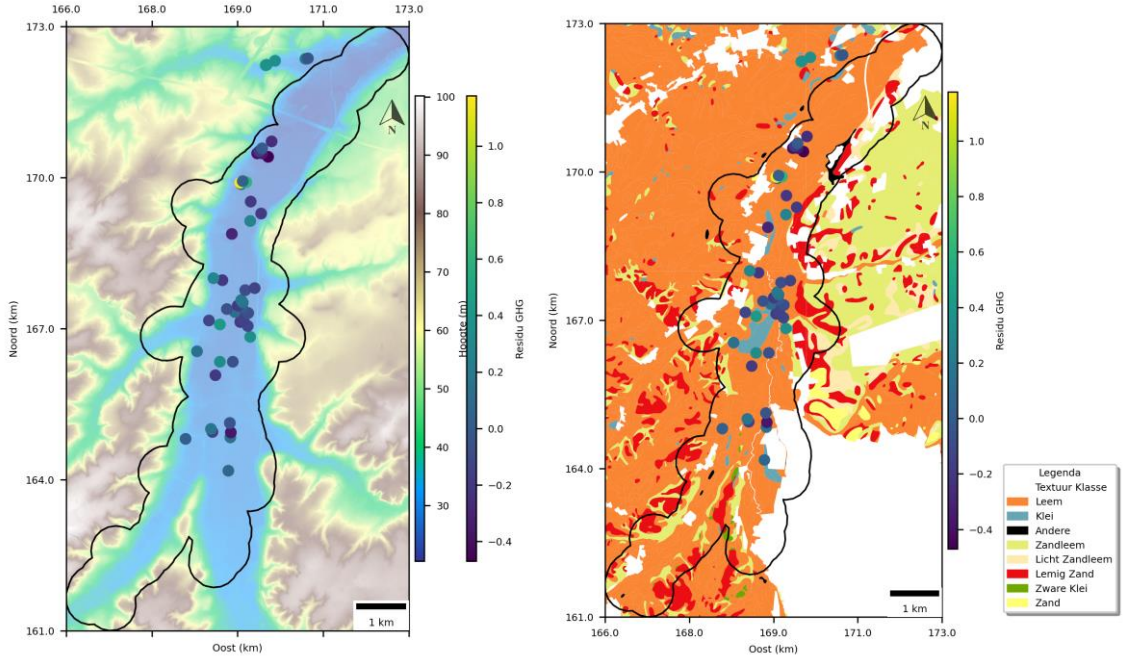
Er werd in Figuur 53 tot Figuur 57 ingezoomd op deze gebieden om nader te bekijken hoe de onzekerheid op de resultaten zich gedraagt in die gebieden en deze te relateren aan factoren die deze foutenmarge beïnvloeden: de textuurklasse en het digitaal hoogtemodel. Gezien het relatief lage aantal trainingspunten is het niet evident door middel van een visuele inspectie conclusies te trekken. Er zijn geen duidelijke conclusies af te leiden uit de grootte van het residu gerelateerd aan de hoogte/helling, noch gerelateerd aan de textuurklasse. M.a.w. uit deze inspectie kunnen we niet éénduidig afleiden dat met een bijkomende inspanning om in een specifieke textuurklasse of op een specifieke helling bijkomend te samplen, betere resultaten zullen worden verkregen.



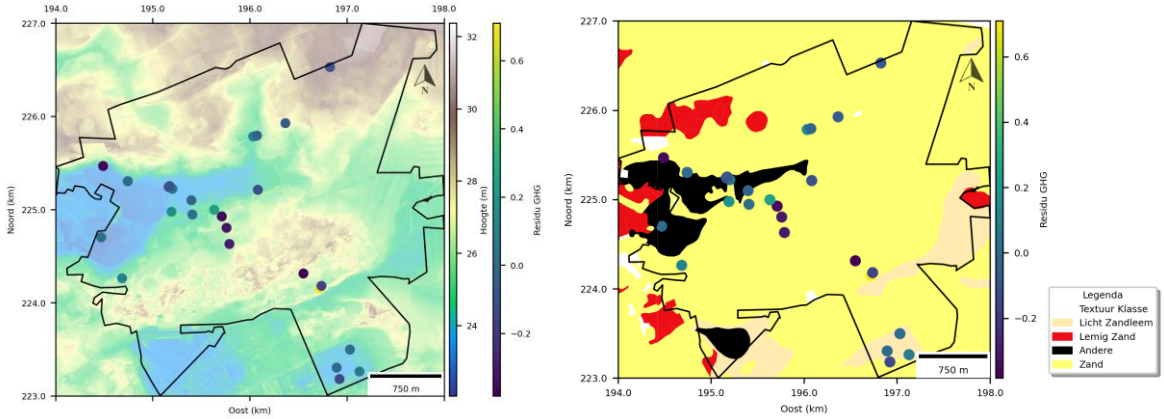
Figuur 53 DHM (links) en textuurklasse (rechts) - pilootzone Antwerpen



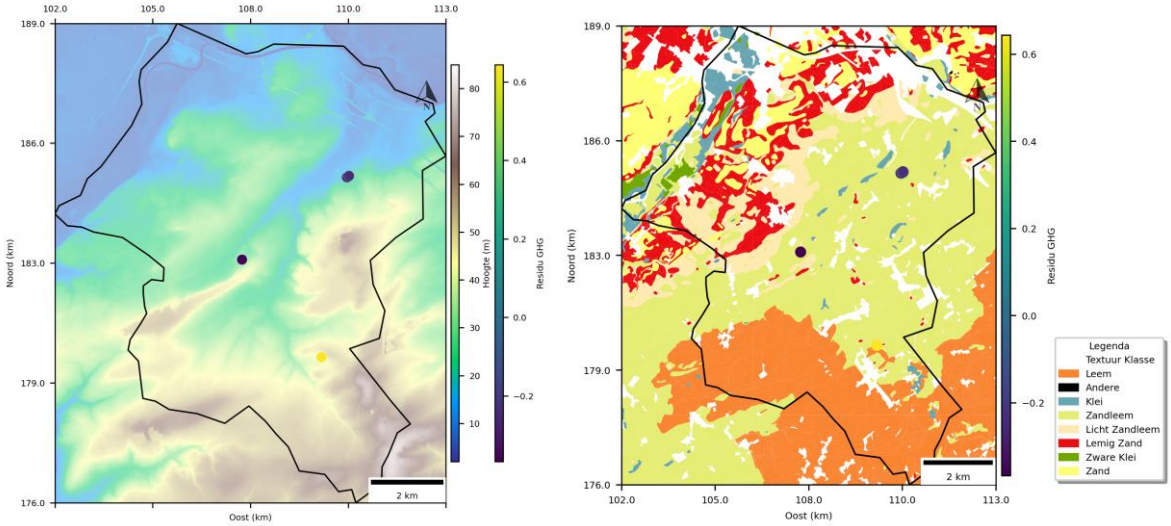
Figuur 54 DHM (links) en textuurklasse (rechts) - pilootzone Damme



Figuur 55 DHM (links) en textuurklasse (Rechts) - pilootzone Dijle



Figuur 56 DHM (links) en textuurklasse (rechts) - pilootzone Liereman



Figuur 57 DHM (links) en textuurklasse (rechts)-pilootzone Rodeland

6.3.6 Kwalitatieve vergelijking met studie VMM

In opdracht van de Vlaamse Milieumaatschappij (VMM) voerde Sumaqua een studie uit die het effect van klimaatverandering op de freatische grondwaterstanden begroot (Sumaqua, 2022). Om dit doel te bereiken was het noodzakelijk om ook eerst een GxG-kaart op te stellen. In wat volgt worden enkele belangrijke verschillen en gelijkenissen toegelicht.

6.3.6.1 Methode

Ondanks het feit dat beide studies onafhankelijk zijn uitgevoerd – en zelfs geen weet hadden van elkaars werk – zijn er opmerkelijke gelijkenissen qua aanpak.

De Machine Learning technieken om vanuit de trainingset de Vlaanderendekkende grondwaterpredicties te bepalen komen in beide methodologieën uit dezelfde familie. XGBoost (gehanteerd in huidige studie) en CatBoost (gebruikt voor de studie van VMM) zijn beide gerenommeerde gradiënt boosting bibliotheken die op brede schaal worden ingezet voor tal van machine learning taken.

Verder verschilt de keuze van het grondwatermodel dat gebruikt wordt om de trainingset te vergroten. Waar in deze studie een PIRFICT-model werd geselecteerd, verkoos men in de studie van de VMM een SWAP-model. Het SWAP-model staat bekend om zijn geïntegreerde aanpak, waarmee de interacties tussen bodem, water, atmosfeer en plant worden bestudeerd, wat het geschikt maakt voor een breed scala aan hydrologische en agronomische scenario's. Echter, deze veelzijdigheid komt met een prijs: SWAP kan complex en rekenintensief zijn, en het vereist gedetailleerde inputgegevens voor nauwkeurige simulaties. Er werden dan ook minder bijkomende punten gesimuleerd. Dit is een nadeel wanneer men resultaten berekent op fijnmazige schaal.

De resolutie van de bekomen GxG-kaart van huidige studie is fijnmaziger (10m) dan die van de VMM (100m). Dit is waarschijnlijk een gevolg van de kern van de opdrachten: waar de GxG-kaart van huidige studie van bij het begin van de opdracht bedoeld was om de grondwaterstatistieken rechtstreeks te gebruiken, was het doel van de VMM eerder deze informatie te gebruiken om de effecten van klimaatverandering te evalueren.

Er zit een belangrijk verschil in de uitfiltering van extreme waarden. Bij de output van de VMM komen waterstanden rond de 40 meter GHG onder het maaiveld voor in de trainingset terwijl we in de huidige studie besloten waterstanden lager dan 3 meter GHG onder het maaiveld niet mee te nemen, conform de opzet van de studie. Indien diepere grondwaterstanden toch als freatisch zouden beschouwd worden, zou de opname van deze extremen in de trainingset het model uiteraard robuuster maken tegen dergelijke diepe uitschieters. Extremen kunnen het model wel overmatig beïnvloeden, waardoor het te specifiek wordt afgestemd op deze waarden en zijn prestaties op ondiepere data verslechteren. Bovendien kan de aanwezigheid van extremen de modeltraining compliceren, wat resulteert in langere trainingstijden en potentieel verminderde betrouwbaarheid door onzekerheden omtrent zeldzame gebeurtenissen. Kortom, het integreren van extremen vraagt om een weloverwogen benadering om een evenwicht te vinden tussen robuustheid en algemene prestaties en houdt bijgevolg verband met de opzet en gevraagde output van iedere studie.

Ook de testset waarop beide modellen geëvalueerd worden, verschilt tussen beide studies. Deze gebruikt in voorliggende studie leunt minder hard aan tegen de trainingset dan wanneer er een

kruisvalidatie – cf. de andere studie – wordt uitgevoerd. Hier werd bovendien een specifieke testset gekozen om (a) zo dicht mogelijk aan te leunen bij de strikte definitie van GxG en zo geen bijkomende fouten te introduceren en (b) om de representativiteit te verhogen.

6.3.6.2 Fouten statistieken

Harde vergelijking tussen beide resultaten en beide foutenstatistieken wordt bemoeilijkt door de verschillende resolutie, verschillende test- en trainingsdatasets, alsook verschillende evaluatietechnieken en opzet, waarbij de VMM-output ook diepere GxG's opneemt.

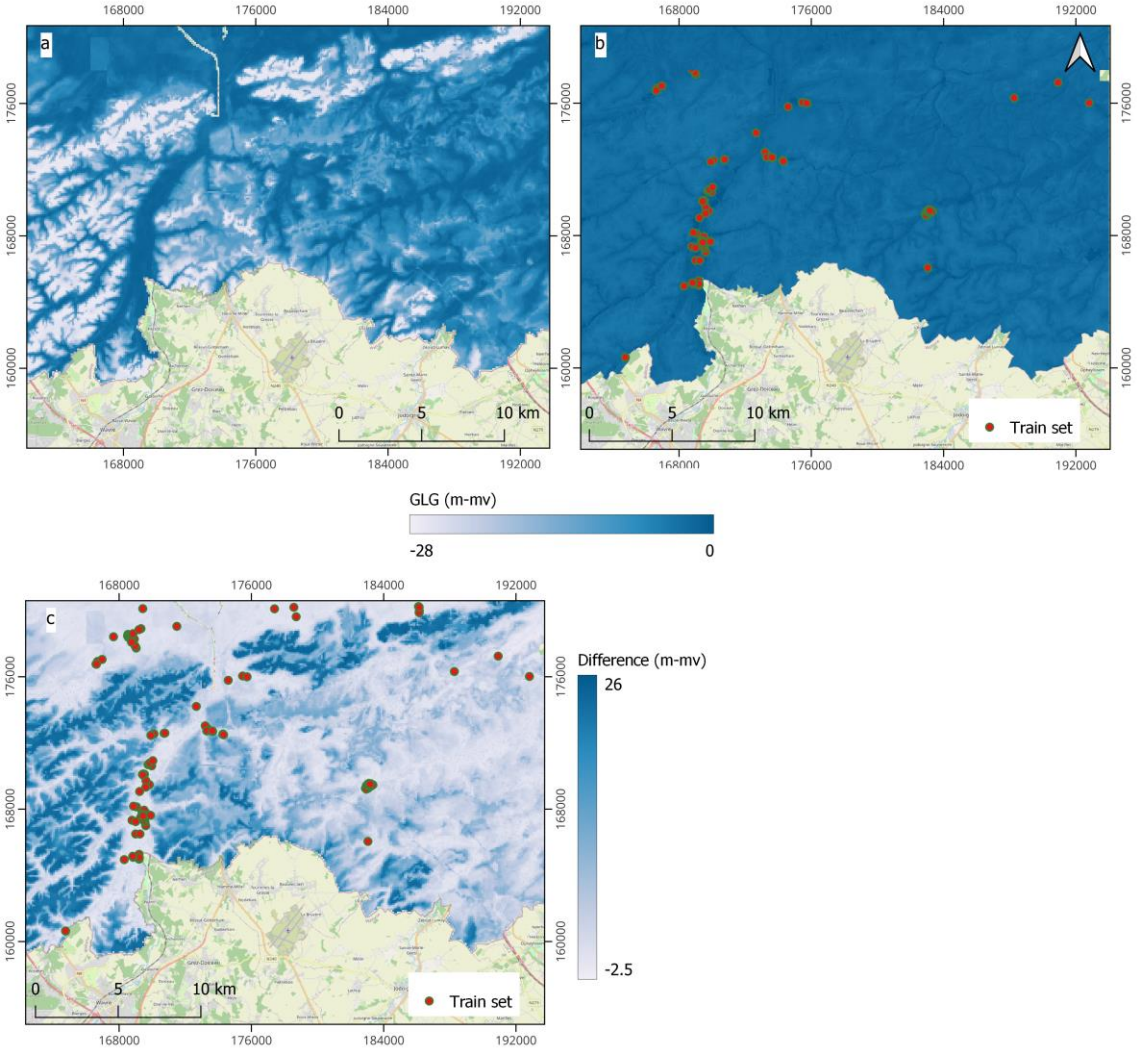
De studie uitgevoerd in opdracht van de VMM, bekomt via kruisvalidatie een MAE van ca. 60 cm voor zowel GLG als GHG. Bij onze resultaten wordt een MAE van respectievelijk 49 en 41 cm bekomen voor GLG en GHG. Deze resultaten liggen in dezelfde grootteorde, gegeven de verschillen in opzet en (evaluatie)methodiek. De performantie van het andere model op de testdataset die is aangewend in voorliggende studie onderzoeken, is niet aangewezen aangezien deze punten grotendeels zijn opgenomen in de trainingsdataset van het model.

6.3.6.3 Resultaten

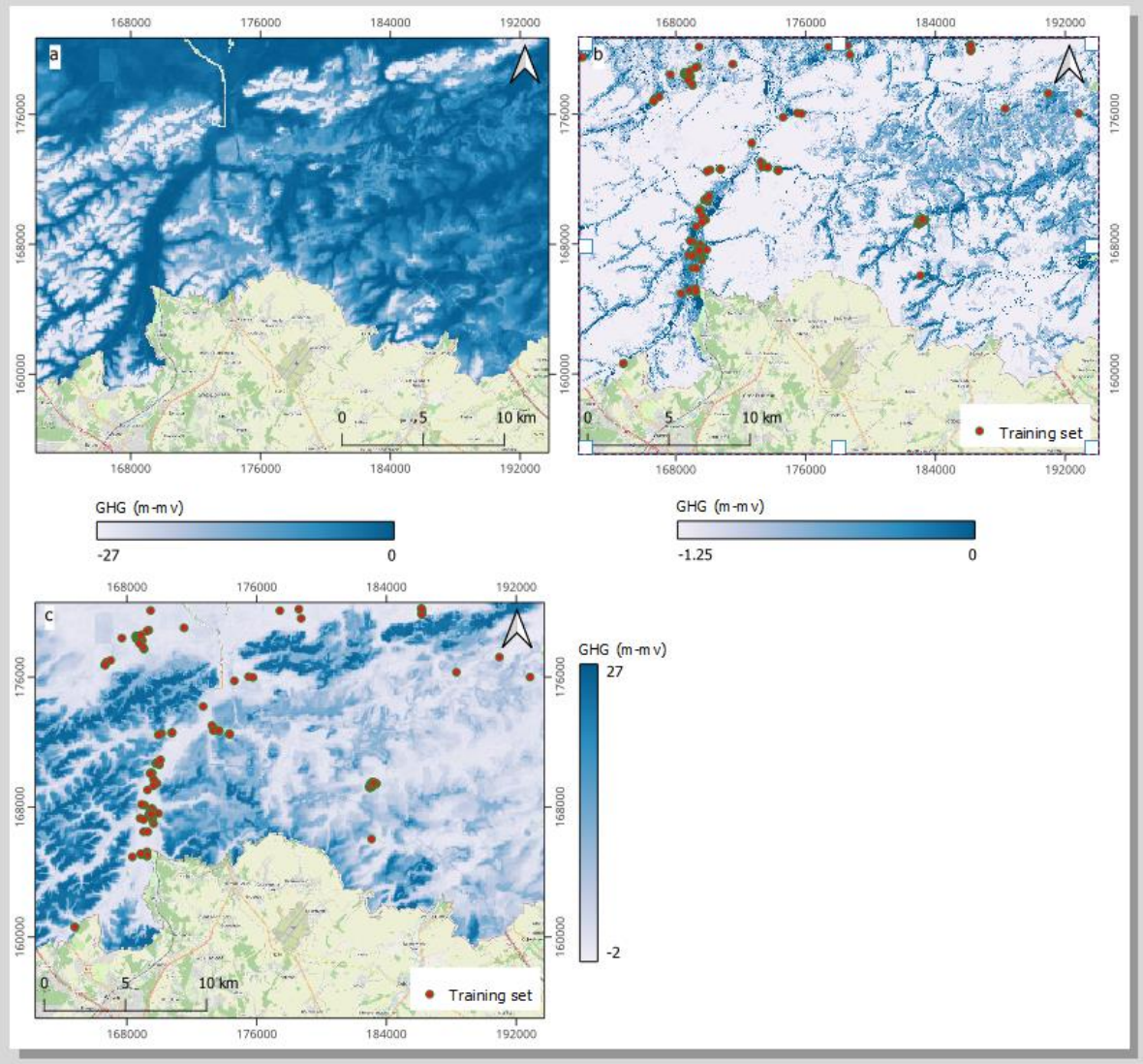
Een zorgvuldige handmatige inspectie van de GxG-kaarten is uitgevoerd, waarbij speciale aandacht is besteed aan enkele opmerkelijke verschillen tussen de voorspellingen van de twee methodes.

De analyse weergegeven in Figuur 58 wijst uit dat er significante discrepanties zijn tussen de voorspellingen van de twee data-gedreven methodes op specifieke geografische locaties. Deze verschillen zijn groter in gebieden waar de beschikbaarheid van trainingsdata beperkt is. Gezien we ervoor kozen om de resultaten van Antea te filteren op die locaties waar de GHG op een maximale diepte van 125cm onder het maaiveld ligt voegen we voor de volledigheid ook Figuur 59 toe.

Bovendien liggen de voorspellingen van Sumaqua consistent dieper voor hoger gelegen gebieden. Dit heeft ongetwijfeld te maken met de opname van diepere grondwaterstanden in hun trainingsdataset. Op de plekken waar wel data beschikbaar zijn, geven beide modellen zeer gelijkaardige voorspellingen. De keuze van de te gebruiken kaart moet dus afgeleid worden aan de hand van welke trainingsdataset en daardoor welke soort voorspelling representatief is voor de 'ground truth' of de uiteindelijke toepassing. Voor het doorvertalen naar drainageklassen is het minder relevant of de GLG uitkomt op 3 of 30m onder het maaiveld. Voor andere toepassingen kan dit verschil wel uitmaken.



Figuur 58: GLG-voorspellingen van Sumaqua, b: GLG-voorspellingen van Antea met de locatie van de trainingsdata, c: verschil tussen beide predicties: positieve waarden wijzen op een lager grondwaterpeil voorspeld door Sumaqua.



Figuur 59: Vergelijking Sumaqua (a) en Antea (b) met de resultaten van Antea gefilterd voor die locaties waar de GHG op een maximale diepte van 125cm onder het maaiveld ligt. De verschilkaart van de twee wordt weergegeven in figuur c.

6.4 BEVINDINGEN DRAINAGEKLASSEN VAN DE BODEMKAART

Tabel 15 geeft de kruisgewijze vergelijking tussen het aantal polygonen geclassificeerd in de oude kaart versus de nieuw bekomen drainageklassen weer. Hieruit blijkt dat vooral de matig drainerende bodems (drainageklasse c) en de te sterk drainerende bodems (drainageklasse a) in aantal zijn gestegen. Dit voornamelijk ten koste van de gunstig drainerende (drainageklasse b) en onvoldoende tot matig slecht drainerende bodems (resp. c en d).

Tabel 15: Kruistabel aantal polygonen oude-versus nieuwe drainageklasse van de bodemkaart

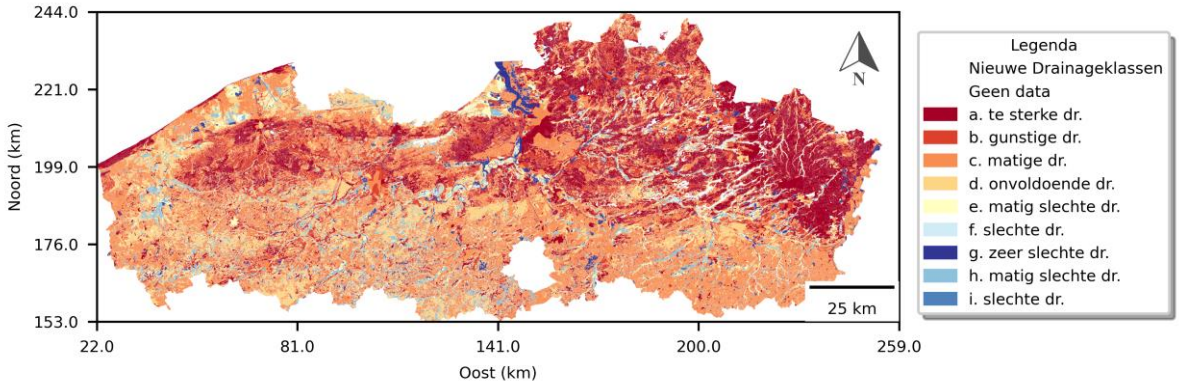
Oude drainage-klasse	Nieuwe drainageklasse										
	a	b	c	d	e	f	g	h	i	Totaal	%
a	1062	279	105	13	3	0	16	0	0	1478	0,7%
a-b	153	55	38	13	1	0	0	0	0	260	0,1%
a-d	1620	436	605	54	23	3	3	6	0	2750	1,3%
b	11150	2448	25466	1310	24	9	107	182	9	40705	18,6%
c	14409	7871	16434	6248	229	31	210	1646	112	47190	21,6%
c-d	194	174	5071	739	47	7	33	174	9	6448	2,9%
d	8863	10286	25100	15038	896	163	786	3482	232	64846	29,6%
e	1119	2625	8181	9681	4032	609	1086	5435	226	32994	15,1%
e-f	0	1	75	129	92	122	32	124	6	581	0,3%
e-i	0	0	17	14	0	0	0	1	0	32	0,0%
f	124	277	1134	1842	3111	1006	797	1678	194	10163	4,6%
g	24	23	112	126	357	305	240	126	15	1328	0,6%
h	199	314	3391	3044	698	163	92	1347	86	9334	4,3%
h-i	12	42	134	74	23	1	4	69	7	366	0,2%
i	0	0	14	24	115	33	7	60	5	258	0,1%
Totaal	38929	24831	85877	38349	9651	2452	3413	14330	901	218733	
%	17,8%	11,4%	39,3%	17,5%	4,4%	1,1%	1,6%	6,6%	0,4%		

Om de verschillen met de oude drainageklasse van de bodemkaart te bepalen werden de categorische drainageklassen omgezet naar een numerieke waarde als volgt: 'a': 0, 'b': 1, 'c': 2, 'd': 3, 'e': 4, 'f': 5, 'g': 6. Gezien de afwijkende definitie van de klassen h en i kunnen deze niet ordinair geordend worden zoals de andere klassen en laten we die buiten beschouwing bij deze interpretatie.

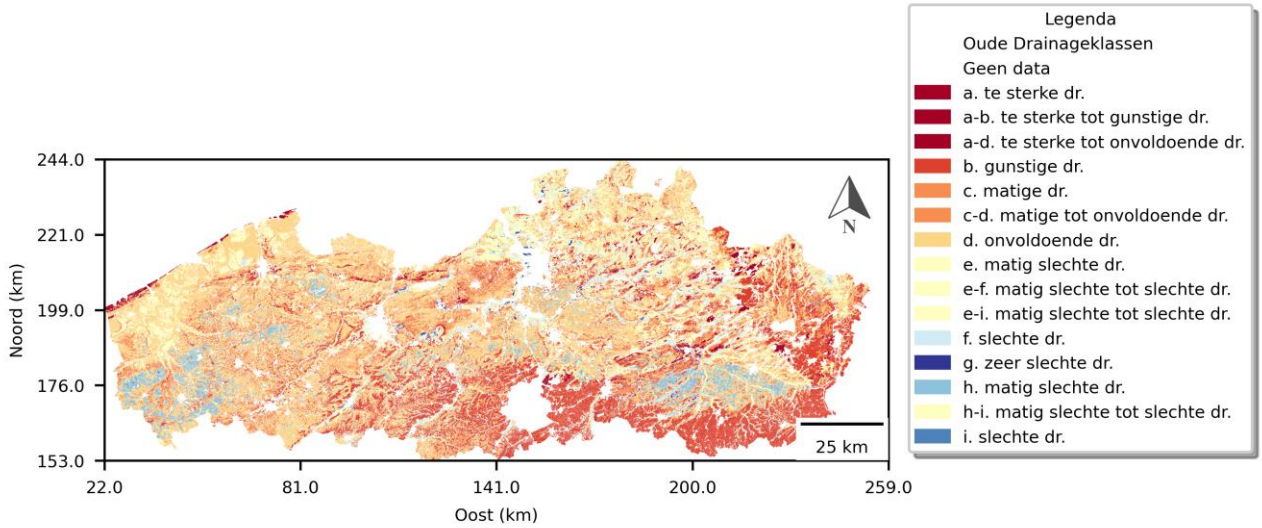
Figuur 60 en Figuur 61 bevatten de oude en de nieuwe drainageklasse van de bodemkaart respectievelijk. Figuur 62 presenteert hoeveel drainageklassen een polygoon is opgeschoven. Een negatieve waarde representeert de “nattere” richting, een positieve waarde een “drogere”. Het valt meteen op dat het grootste deel van de polygonen minimum één drainageklasse zijn opgeschoven en dat vooral in de drogere richting die verschuiving drastisch is. Een verschuiving van vier klassen is niet zeldzaam, wat drastisch is, maar niet volledig kan toegeschreven worden aan de modelfout. Dit kan enerzijds veroorzaakt worden door de verschillende techniek: afleiden van gemeten GxG’s versus van waargenomen bodemkenmerken en, anderzijds, door verdroging van de ondergrond gedurende de laatste 50 jaar.

Figuur 63 presenteert de interkwantielafstanden: het verschil in drainageklasse tussen het 5de en 95ste percentiel. Gecombineerd met Figuur 62, valt op hoe de grootste verdroging regelmatig samenvalt met de grootste onzekerheid. Dit zou de extreme verschuiving in een deel van de polygonen voor een stuk kunnen verklaren. Merk verder ook op dat niet alle drainageklassen dezelfde breedte hebben. In de nauwere drainageklassen is de kans groter dat bij men bij een verdroging of vernatting meteen één of meerdere drainageklassen opschuift. Omwille van dezelfde reden lijkt het er ook op dat onder verschillende steden, de onzekerheid bijzonder klein is (blauwe

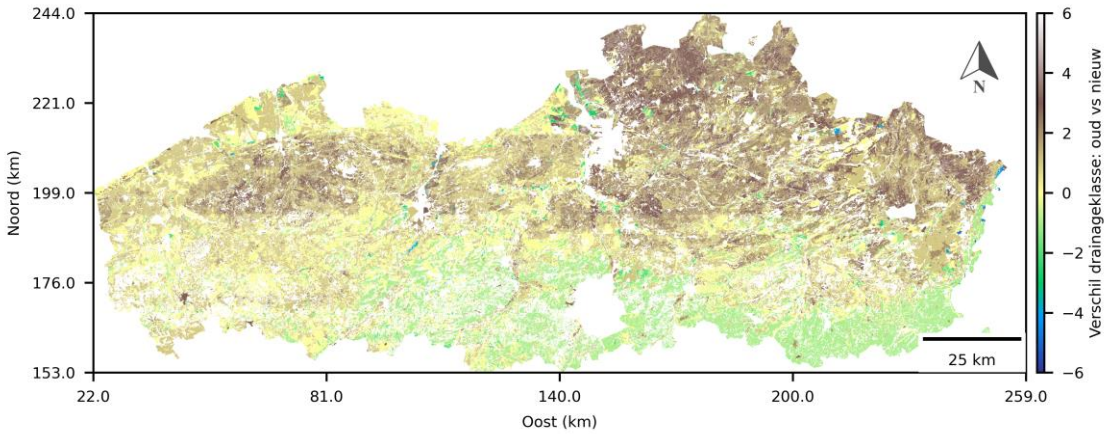
kleur in Figuur 63). Hier geldt de drainageklasse c, wat een relatief brede klasse is, hetgeen de kans verkleint dat men bij de uiterste kwantielen in een andere drainageklasse zou belanden.



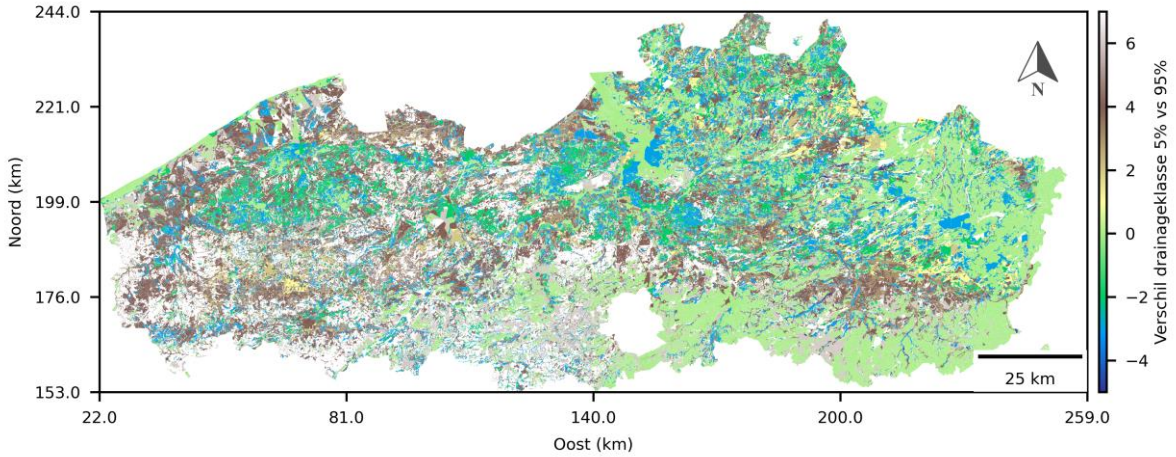
Figuur 60 Nieuwe drainageklasse van de bodemkaart



Figuur 61 Oude drainageklasse van de bodemkaart



Figuur 62: Verschilkaart tussen de oude en de nieuwe drainageklassen van de bodemkaart: negatief = natter, positief = droger.



Figuur 63: Verschilkaart tussen de twee uiterste kwantielen van de nieuwe drainageklassen van de bodemkaart

7 AANBEVELINGEN BIJ HET GEBRUIK VAN DE GEGENEREERDE KAARTEN

In 2010 voerden Finke, Van de Wauw & Baert een gebruikersinventarisatie uit onder gekende gebruikers van de oude drainageklassen van de bodemkaart met als doel de perceptie, kwaliteit en gebruikswaarde van zulke kaart af te leiden (§2.1). Het resultaat van voorliggende studie komt tegemoet aan de drie belangrijkste verzuchtingen:

1. Men gaf aan dat de drainageklassen niet meer actueel zijn, gegeven deze gebaseerd zijn op metingen in de jaren '70. Voorliggende studie actualiseerde de drainageklassen van de bodemkaart met behulp van recente metingen.
2. De gebruikers wensten een meer kwantitatieve variabele. Voorliggende studie levert gebiedsdekkende informatie op van zowel gesimuleerde GLG als GHG. Met andere woorden, de GxG-rasters bevatten voor de meeste gebruikers meer informatie dan de polygonen met één drainageklasse. We merken hier ook op dat een verdere abstractie van de GxG-waarden naar drainageklassen en polygonen, in de praktijk altijd een verlies aan informatie betekent, alsook een vergroting van de fout/ onzekerheid.
3. De gebruikers wensten meer detail: een hogere spatiale resolutie. De gesimuleerde GHG en GLG-rasters zijn opgesteld op een 10x10m resolutie.
4. Er was nood aan informatie voor de kustvlakte. Voorliggende studie levert een gebiedsdekkend resultaat op, waarbij er ook GLG, GHG en drainageklassen beschikbaar zijn voor zowel de kustvlakte als de bebouwde zones.

Bijkomend zijn er kwantielmodellen opgesteld, wat een maat van onzekerheid van de gesimuleerde GxG's voor iedere individuele rastercel oplevert. Hierdoor kan de eindgebruiker inschatten of de waarde, gegeven de geschatte fout op de GxG, voor zijn doeleinden geschikt is. Deze onzekerheden zijn doorvertaald naar de drainageklassen van de bodemkaart. Echter, dergelijke presentatie zal bemoeilijkt worden door het aggregeren naar polygonen, en nagenoeg onmogelijk zijn na vertaling naar drainageklassen. We merken hier ook op dat de vertaling naar drainageklassen per polygoon van de bodemkaart toentertijd gedaan is vanuit praktische overwegingen op basis van destijdse randvoorwaarden: o.a. de toen beschikbare informatie. De drainageklasse per polygoon in de bodemkaart speelt nog steeds een rol voor de eindgebruiker die hiermee snel en eenvoudig een ruimtelijk inzicht bekomt, maar voor meer specialistische eindgebruikers is het beter om de GxG's te hanteren.

Om verder de bruikbaarheid van de kaarten te toetsen, werden meerdere experts bevraagd, met jarenlange ervaring in grondwatermodelering en/of het opstellen van hemelwaterplannen, infiltratie- en bemalingsstudies of ecohydrologische studies. Dit leverde volgende inzichten op:

- De vorige drainageklasse van de bodemkaart wordt voornamelijk gebruikt om de grondwaterstand in te schatten, hetgeen een onnauwkeurig resultaat oplevert. De fijnmazigheid en de aanwezigheid van een kwantitatieve variabele wordt daarbij zonder uitzondering als groot voordeel naar voor geschoven zelfs met de foutenmarges en onzekerheden die verbonden zijn aan het resultaat uit voorliggende studie. Verder heeft men bij het gebruik van een drainageklasse, het raden naar hoe het peil spatiaal fluctueert. Dankzij de beschikbaarheid van een GHG en GLG op elke locatie is een GxG-kaart op dat vlak een enorme vooruitgang.

- De GxG-kaart door alle experts nuttig bevonden voor gebruik in de exploratieve fase van een project/studie. Men is in verschillende omstandigheden bij wet verplicht peilbuizen aan te leggen: een GxG-kaart kan helpen bij het selecteren van de locatie van nieuwe peilbuizen en geeft bijkomend inzicht in hoe diep men de peilbuis en/of filter moet plaatsen, wat een grote invloed kan hebben op het budget. Voor kleinere projecten en voor projecten die net een heel groot gebied beslaan, worden dikwijls nauwelijks peilbuizen geslagen omdat (a) het budget dit niet toestaat en/of (b) de wet dit niet strikt voorschrijft. Voor zulke studies is een gebiedsdekkende GxG-kaart een belangrijke informatiebron. Gegeven het karakter van deze studies, wordt de variërende foutenmarge van de nieuwe kaarten in die context dan ook niet beschouwd als een struikelblok.
- Men merkte ook op dat voor de textuurklasse "zand" de infiltratie een factor tien kan verschillen voor verschillende gradaties van zand. Het kan een goed idee zijn om bij een eventuele toekomstige kaart ook de porositeit van zand mee te nemen als verklarende variabele, van zodra deze gebiedsdekkend ter beschikking komt.
- Voor bemalingsstudies zijn indicaties van GHG onmisbaar. In praktijk wordt daarvoor de drainageklasse van de bodemkaart niet gebruikt. Voor infiltratie- en bemalingsstudies werkt men namelijk hoofdzakelijk in of nabij bebouwde gebieden waar de vorige bodemkaart geen informatie bevat met betrekking tot drainageklasse. Waar geen metingen voor handen zijn, kijkt men naar de dichtstbijzijnde peilbuis die niet zelden kilometers verderop ligt: een methode die in lijn ligt met de methodiek die hier gehanteerd wordt om GxG's van incidentele metingen vast te leggen (§5.3), maar met een grotere maximale zoekstraal. Het gebiedsdekkende karakter van de nieuwe kaarten levert op dat vlak een vooruitgang, zelfs rekening houdend met de grotere foutenmarge in deze zones. Men vermoedt namelijk dat de onzekerheid die het gevolg is van een extra-/ of interpolatie naar een verder gelegen peilbuis, groter is dan de gemiddelde fout van de nieuwe kaart. Bovendien geeft de nieuwe kaart een maat van onzekerheid voor iedere voorspelde rastercel. Bij gebruik zal men in eerste instantie kijken of er een peilbuis in de directe nabijheid van het studiegebied ligt en, zo niet, kan men de nieuwe GxG-kaarten aanwenden.
- Voor bemalingen beperkt de probleemstelling zich regelmatig ook tot de vraag of bemaling überhaupt nodig is. Dit zou men dankzij de fijnmazige GxG-kaart nu sneller kunnen afleiden. Hierbij moet men zich wel realiseren dat de extreem diepe grondwatertafels ook belangrijk zijn, hetgeen in voorliggende studie minder goed wordt voorspeld met voorliggende methode. In dat opzicht zou men kunnen opteren ook de gegenereerde waarden van de VMM-studie als informatie mee te geven.
- Ecohydrologische studies vergen de hoogste accuraatheid: de maximaal toegelaten onzekerheid op de GLG en GHG is veel kleiner. Voor sommige vegetatietypes is het zelfs nodig de GxG's op enkele centimeters nauwkeurig te kennen, wat zelfs uiterst lastig te halen is met speciaal hiervoor opgezette, lokale grondwatermodellen. De veelgebruikte modellen en software die worden gebruikt op het grondwaterniveau te berekenen, worden in de praktijk met een aantal metingen gekalibreerd. Echter valt het niet zelden voor dat men werkt met een tweestaps-kalibratie: in het studiegebied zijn nieuwe, nauwkeurige metingen om het grondwatermodel te kalibreren onmisbaar. Op grotere afstand van het studiegebied komt het voor dat men de modellen kalibreert met metingen die op grotere afstand liggen. Voor deze tweede kalibratie zou het zeer nuttig kunnen zijn om enkele fictieve peilbuizen te selecteren uit de GxG-kaart. Dit is echter enkel mogelijk dankzij de onzekerheden die als attributen zullen worden meegegeven aan de GxG-kaart. In de bredere context dan ecohydrologie, wordt van state-of-the-art grondwatermodellen een nauwkeurigheid van meer dan 20 cm verwacht.

- Ten slotte merken we nog op dat in de veelvuldig gebruikte grondwatermodellen, zoals MODFLOW, het de standaardpraktijk is om te werken met rastercellen die variëren van 10x10m tot 100x100m naargelang hoe ver men verwijderd is van het studiegebied. In het geval van steile topografie of verhang valt het voor dat de overgang tussen verschillende rastercellen relatief hard is (lees: niet glad). Harde overgangen zijn zoals eerder aangehaald, ook aanwezig in de nieuwe kaart met de grove resolutie van de SWI-feature als grootste oorzaak. Bij grondwatermodellen wordt meestal gesmooth met behulp van de isohypsen die een zeer glad resultaat opleveren. In voorliggende studie werd er echter beslist om geen smoothing toe te passen (§6.3.4).

Samengevat kan men stellen dat de bruikbaarheid van de nieuwe resultaten aanzienlijk hoger ligt dan de vorige drainageklassen van de bodemkaart. Het kwantitatieve karakter, de resolutie, het gebiedsdekkende karakter en de meegegeven maat van onzekerheid worden als een voordeel aanzien. De nieuwe kaarten kunnen aangewend worden voor een eerste verkenning van het studiegebied, als hulpmiddel om de locatie van peilbuizen te bepalen of voor studies waar men geen nieuwe peilbuizen plaatst. Bij gebruik en interpretatie van de nieuwe kaarten, dient de gebruiker echter nog steeds de lokale topografie en bodem in acht te nemen. Voor studies die een hoge nauwkeurigheid vergen (met de ecohydrologische studies als summum), zal men nog steeds een gedetailleerd grondwatermodel moeten afstemmen, gekalibreerd op de lokale situatie.

8 AANBEVELINGEN VOOR VERDERE ONTWIKKELINGEN

Gezien de data-gedreven methodiek, zal men bij verbetervoorstellen direct aan nieuwe grondwaterpeilmetingen denken. De analyse wijst echter uit dat dit nauwelijks, tot maar een beperkte verbetering van de modelonzekerheid met zich meebrengt. Er is namelijk nood aan **meer kwalitatief hoogstaande metingen met voldoende hoge resolutie en duur** om de GxG volgens de correcte definitie te berekenen (cf. Heesen, 1970): de GHG en GLG worden berekend als het gemiddelde van de drie hoogste/laagste peilen per jaar van minimaal acht jaren, waarbij de grondwaterstand tweemaal per maand gemeten wordt (op of omstreeks de 14de en 28ste dag). Voornamelijk de duur en resolutie zijn niet eenvoudig, noch snel op te lossen. Er wordt daarom aangeraden om bij het opstellen van meetcampagnes en/of meetnetten rekening te houden met deze vereiste. De nieuwe IoT-meetnetten met uiterst hoge resolutie, zullen een aanzienlijke meerwaarde bieden. Doch het is noodzakelijk om een voldoende lange periode (meerdere jaren) te meten zodat de grondwaterstatistieken (GxG) niet beïnvloed worden door toevallige meteorologische condities. Tijdens deze studie werd een uiterst diepgaande dataverzameling en -bevraging uitgevoerd. Er kon echter nog geen gebruik gemaakt worden van alle meetgegevens die beschikbaar zijn binnen de Vlaamse Overheid, lokale overheden en/of publiek-private instanties. Het ter beschikking krijgen van deze bijkomende punten lijkt een eenvoudige en goedkope verbeterstap. Momenteel lopen er dan ook gesprekken om meer data ter beschikking te krijgen op DOV.

De analyse wijst uit dat de modelvoorspellingen nog meer gebaat zullen zijn met **een verbetering van de geografische hulpvariabelen**. Er lopen momenteel 2 andere projecten in opdracht van de Vlaamse Overheid die een aanzienlijke verbetering met zich mee kunnen brengen:

1. Momenteel voert de KUL departement Aard- en Omgevingswetenschappen het project Bodemdatabank-STAT uit met als doel: de koppeling van statistische bodemgegevens aan de bodemkaart gebaseerd op de bodemdata van de DOV-bodemdatabank. Deze opdracht is momenteel nog lopende waardoor deze niet aangewend kon worden als input voor deze voorspelling. Eens beschikbaar kan ze een aanzienlijke verbetering vormen t.o.v. de textuurklassen in de bodemkaart: granulometrische textuuranalyses, hoge resolutie kartering, Vlaanderendekkende kaart incl. verstedelijkte gebieden.
2. Er is in opdracht van de VMM een project lopende bij Digitaal Vlaanderen waarbij men een Vlaanderendekkende grachtenkaart wenst op te stellen via A.I. op LIDAR-beelden. Momenteel is men deze nog aan het valideren. Gezien het belang van de verklarende variabelen die afgeleid zijn van deze grachtenkaart, wordt er aangenomen dat deze geoptimaliseerde grachtenkaart een aanzienlijke verbetering met zich mee zal brengen. Idealiter zal dit ook de watermassa's (vijvers, meren etc.) omvatten.

Het digitaal hoogtemodel en zijn afgeleiden zijn uiterst belangrijk voor de modelvoorspellingen. Gezien de computationele kracht, in relatie tot de dichtheid aan beschikbare meetpunten, kon er bij deze studie maar gewerkt worden met een resolutie van 10x10m. Toekomstige verbetering van de computationele rekenkracht zou toelaten om ook de 1x1m variaties in hoogteligging in rekening te brengen. Een optie zou zijn om de Vlaamse Supercomputer aan te wenden. Het opstellen van een script dat een supercomputer vergt, was echter niet in overeenstemming met de krijtlijnen van deze opdracht: de ontwikkelde methodiek moet toelaten om, telkens wanneer er voldoende nieuwe data beschikbaar zijn, de drainageklassen vlot te actualiseren. Een belangrijke lacune aan beschikbare

informatie is de bathymetrie onder de waterlopen en -massa's. Deze informatie zou een aanzienlijke optimalisatie van meerdere hulpvariabelen met zich meebrengen.
De remote sensed informatie aan bodemvocht bleek ook een nuttige predictor. De beschikbare resolutie speelt echter parten en zorgt soms lokaal voor scherpe overgangen tussen naburige rastercelen. Het downscalen van deze schat aan informatie naar een hogere resolutie zou een aanzienlijke meerwaarde bieden voor de modelvoorspellingen. Tijdens deze studie werd dit uitgetest via een data-assimilatietechniek met PIRFICT, maar deze opzet bleek niet haalbaar binnen de krijtlijnen van deze opdracht. Het zou nuttig zijn hier een specifiek onderzoeksproject voor op te zetten.

Voorliggende studie bouwt voort op de bevindingen van Finke *et al.* (2010) waarbij de data-gedreven herkarteringstechnieken als beste uit de bus kwamen. Er bestaan echter nog methodes, elk met hun voor- en nadelen. Via data-assimilatie kan men meerdere predicties optimaal – gegeven hun modelonzekerheid – combineren om een beter resultaat te bekomen: een techniek die succesvol wordt toegepast bij o.m. weersvoorspellingen. In het ideale geval zou er een specifiek onderzoeksproject opgezet worden waarbij men via data-assimilatie voorliggende grondwaterpredicties combineert met een numeriek grondwatermodel, mogelijk aangevuld met bijkomende modelvoorspellingen zoals het omrekenen van downscaled remote sensed bodemvocht of een raster aan 1D PIRFICT-modellen. Voor het succesvol beëindigen van dergelijke grote studie, zijn er echter nog meerdere (onderzoeks)stappen te zetten.

REFERENTIES

Albergel, C. ; Rüdiger, C. ; Pellarin, T. ; Calvet, J. - C. ; Fritz, N. ; Froissard, F. ; Suquia, D. ; Petitpa, A. ; Piguet, B. ; Martin, E. From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. Hydrol. Earth Syst. Sci. 2008, vol.12, 1323-1337.

Deckers J., De Koninck R., Bos S., Broothaers M., Dirix K., Hambsch L., Lagrou, D., Lanckacker T., Matthijs, J., Rombaut B., Van Baelen K. & Van Haren T. (2019a). Geologisch (G3Dv3) en hydrogeologisch (H3D) 3D-lagenmodel van Vlaanderen. Studie uitgevoerd in opdracht van: Vlaams Planbureau voor Omgeving (Departement Omgeving) en Vlaamse Milieumaatschappij 2018/RMA/R/1569, 286p. + bijlagen, https://archief.onderzoek.omgeving.vlaanderen.be/Onderzoek-1999911.

Deckers J., De Koninck R. & Rombaut B. (2019b). Update Hydrogeologische Codering van Vlaanderen: HCOV-versie 2. Studie uitgevoerd in opdracht van: Vlaamse Milieumaatschappij 2019/RMA/R/1899, 21p. + bijlagen

Declercq, F., & Poesen, J. (1991). Erosiekarakteristieken van de bodem in Laag- en Midden-Belgie. Tijdschrift van de Belgische vereniging voor Aardrijkskundige Studies (BEVAS). 1:29-46.

De Bakker, H. (1973). Hydromorphic soils in the system of soil classification for the Netherlands. In: Pseudogley and Gley, Transactions of Commissions V and VI, Int. Soc. Soil Sci. Stuttgart 1971: 404-412.

Fan, Y., Miguez-Macho, G., Weaver, C. P., Walko, R., & Robock, A. (2007). Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations. Journal of Geophysical Research: Atmospheres, 112(D10).

Finke, P.A. , D.J. Brus, M.F.P. Bierkens, T. Hoogland, M. Knotters, F. de Vries. (2004). Mapping ground water dynamics using multiple sources of exhaustive high resolution data. Geoderma 123: 23-39.

Finke, P.A., Brus, D.J., Bierkens, M.F.P., Hoogland, T., Knotters, M. & de Vries, F. (2005). Kartering van de grondwaterdynamiek met behulp van geo-informatie van hoge resolutie

Finke, P.A., Van de Wauw, J. & Baert, G. (2010). Ontwikkelen en uittesten van een methodiek voor het actualiseren van de drainageklasse van de bodemkaart van Vlaanderen.

Locher, W.P. en H. De Bakker. (1991). Bodemkunde van Nederland Deel 1, Algemene bodemkunde. Tweede druk. Malmberg Den Bosch.

Miguez-Macho, G., Fan, Y., Weaver, C. P., Walko, R., & Robock, A. (2007). Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation. Journal of Geophysical Research: Atmospheres, 112(D13).

JE Nash. Determining runoff from rainfall. Proc. Inst. Civ. Eng., 10:163–184, 1958

Niu, G. Y., Yang, Z. L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. Journal of Geophysical Research: Atmospheres, 112(D7).

Nock, R & Nielsen, F. (2004) Statistical Region Merging, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 26, No 11, pp 1452–1458, 2004.

Notebaert, B., Govers, G., Verstraeten, G., Van Oost, K., Ruysschaert, G., Poesen, J., & Van Rompaey, A. (2006). Verfijnde erosiekaart Vlaanderen: eindrapport. Leuven: K.U. Leuven, Onderzoeksgroep Fysische en Regionale Geografie. https://publicaties.vlaanderen.be/view-file/20203

Oorts, K., Huybrechts, R., Buyle, S. & Van Damme, M. (2020). De DOV-bodemdatabank versie 1.0: objectenmodel. Departement Omgeving, Brussel, 49 pp

Oorts, K., Deproost, P., Buyle, S. & Swerts, M. (2019). Eindrapport potentiële bodemerosiekaart per perceel (2018). Departement Omgeving, Brussel. https://archief-algemeen.omgeving.vlaanderen.be/xmlui/bitstream/handle/acd/269122/Eindrapport%20potentiele%20bodemerosiekaart%20per%20perceel%202018.pdf

Oude Voshaar, J. and Stolp, J., 1997. Schatting van GHG en GLG van tijdelijke peilbuizen met korte meetreeksen. Technical Document 30, Winand Staring Centre for Integrated Land, Soil and Water Research, Wageningen.

Poelmans L., Janssen L. & Hambsch L. (2021), Landgebruik en ruimtebeslag in Vlaanderen, toestand 2019, uitgevoerd in opdracht van het Vlaams Planbureau voor Omgeving.

Sutanudjaja (2012). The use of soil moisture - remote sensing products for large-scale groundwater modeling and assessment, Proefschrift, Universiteit Utrecht

Shih, F. Y., & Cheng, S. (2005). Automatic seeded region growing for color image segmentation. Image and vision computing, 23(10), 877-886.

Song, Y. & Hao, Y. (2017). "Image segmentation algorithms overview." arXiv preprint arXiv:1707.02051.

Sumaqua (2022). Effecten van klimaatverandering op de freatische grondwaterstanden. Syntheserapport. Opdracht uitgevoerd in opdracht van VMM.

Te Riele, W.J.M. en D.J. Brus. 1991. Methoden van gerichte grondwaterstandmetingen voor het schatten van de GHG. Rapport 158, DLO-Staring Centrum, Wageningen.

Van de Wauw, J. (2012). Spatial Prediction of Water Table Dynamics in Flanders. Proefschrift voorgelegd tot het verkrijgen van de graad van Doctor in de Wetenschappen: Geologie.

Van Damme, J. and De Leenheer, L. (1969). Variations du niveau de la nappe phréatique au cours de cinq années (1963-1968) dans les sols de la Campine anversoise (province d'Anvers-Belgique). Pedologie 19(3): 275-320.

Van Heesen, H.C. (1970). Presentation of the seasonal fluctuation of the water table on soil maps. Geoderma 4, 257-278.

Van Ranst, E. & Sys, C. (2000). Eenduidige legende voor de digitale bodemkaart Van Vlaanderen (schaal 1:20 000).

Von Asmuth, J.R., M.F.P. Bierkens and C. Maas. 2002. Transfer function noise modeling in continuous time using predefined impulse response functions. Water Resour. Res. 38(12): 23.1-23.12.

WEBPAGINA'S

VITO (2022). Copernicus Global Land Service: Soil Water Index. https://land.copernicus.eu/global/products/swi. Bezocht op: 31 augustus 2022.

APPENDIX 1: INVENTARIS

Instantie	Reactie ontvangen?	Aangeschreven instanties Opmerking	Mag ook na project gedeeld worden?	Bijkomende data aangeleverd
Stad Aalst	nee		nvt	nee
ANB	nee		nvt	nee
Antea Intern	ja	Data van ons IoT netwerk voor meerdere projecten beschikbaar. Kan van meeste projecten aangewend worden.	ja	ja
Stad Antwerpen	ja	Zeer behulpzaam. Veel info ontvangen: - neerslagdata 8 stadspluvio's, grondwaterpeilmetingen en vijverpeilen + shapes met overzicht van locaties; - studie van ABO voor de opmaak van grondwaterkaarten voor de stad; - raden aan om ook LANTIS te contacteren: meetnet aan het uitbouwen voor de Oosterweelverbinding.	ja	ja
Aquafin	ja	Beschikken over een IoT netwerk. Wordt aangeleverd.	ja	nee
ASGO	ja	Meetnet rond winning onder golfgebied Knokke, was al beschikbaar via IMJV	nvt	nee
Belgische Geologische Dienst (BGD)	ja	Historische data reeds gedeeld met DOV. Huidige gemonitorde meetnetten in functie van diepe geothermie (niet op DOV) --> niet bruikbaar voor de studie	nvt	nee
Stad Brugge	ja	Geen gegevens die gedeeld mogen worden	nvt	nee
De Watergroep	ja	Gros van data beschikbaar via VMM. Ze bezitten nog een aantal nog hogere resolutie reeksen. We starten met dump van VMM.	nvt	nee

Instantie	Reactie ontvangen?	Aangeschreven instanties Opmerking	Mag ook na project gedeeld worden?	Bijkomende data aangeleverd
		Wanneer er tijdens de kalibratiefase detailinfo nodig is omtrent waterwinningsgebieden, kunnen we een vervolgoverleg inplannen.		
Farys	nee	Nieuwe contactpersoon benaderd via collega	nvt	nee
Stad Genk	ja	Geen bijkomende info beschikbaar	nvt	nee
Stad Gent	ja	Geen bijkomende info beschikbaar, metingen van Provincie Oost-Vlaanderen mogen aangewend worden	nvt	nee
Stad Hasselt	ja	Geen info beschikbaar	nvt	nee
ILVO	ja	Enkel oppervlaktewaterpeilmetingen Rodeland gebied	???	nee
INBO	ja	WATINA volledig gesynchroniseerd met DOV	nvt	nee
IWVA	ja	Ja, rond winningen in Westhoek en Sint André, raadpleegbaar via IMJV	nvt	nee
Stad Kortrijk	ja	Geen metingen	nvt	nee
KUL	ja	Geen info beschikbaar Nieuwe contactpersoon benaderd	nvt	nee
Stad Leuven	ja	Data aangeleverd	ja	ja
Stad Mechelen	ja	Beperkte info maar reeds aangeleverd	ja	ja
Middenkustpolder	ja	Geen grondwater, oppervlaktewaterpeilen staan allemaal op waterinfo.	nvt	nvt
Maatschappij Linkerscheldeoever (MLSO)	ja	Data aangeleverd	nee, enkel raadplegen via DOV op aanvraag	ja
Natuurpunt (Gent)	ja	Geen data	nvt	nee
Natuurpunt (Antwerpen - zie ook Havenbedrijf Antwerpen)	ja	Vragen vergoeding. 50-tal peilbuizen op LSO en RSO. Gestart In 2011-2012, daarna meetnet langzaam uitgebreid en overgedragen aan Havenwachter. Nu opnieuw overgedragen aan Natuurpunt dat de gegevens	nvt	nee

Instantie	Reactie ontvangen?	Aangeschreven instanties Opmerking	Mag ook na project gedeeld worden?	Bijkomende data aangeleverd
		herstructureert, controleert en omzet naar mTAW). Zal niet aangewend worden gezien de gevraagde vergoeding		
Nieuwe Polder van Blankenberge	ja	Geen metingen die niet op DOV staan	nvt	nee
Stad Oostende	ja	Geen metingen die niet op DOV staan	nvt	nee
Oostkustpolder	ja	Geen grondwater, oppervlaktepeilen staan allemaal op Waterinfo. Wel zoutmetingen beschikbaar (overige Polders waarschijnlijk zelfde)	nvt	nee
OVAM	ja	MISTRAL database kan aangeleverd worden, mits een kleine vergoeding	nvt	nee
Pidpa	nee		nvt	nee
PoA (Havenbedrijf Antwerpen)	ja	Meetnet recent overgedragen aan Natuurpunt Antwerpen (zie Natuurpunt Antwerpen)	nvt	Nee
Provincie Antwerpen	ja	Beperkte info maar wordt aangeleverd.	ja	Ja
Provincie Limburg	ja	Data van projecten Antea mag aangewend worden.	ja	Ja
Provincie Oost-Vlaanderen	ja	Overleg ingepland. Data wordt aangeleverd via externe harde schijf. Data van Antea's project mag aangewend worden	ja, mits voorafgaande terugkoppeling	Ja
Provincie Vlaams-Brabant	ja	Nieuwe contactpersoon benaderd via collega: oppervlaktewaterpeilen staan allemaal op Waterinfo. Geen grondwatermetingen	nvt	Nee
Provincie West-Vlaanderen	ja	Peilmetingen in waterlopen zijn gekoppeld aan Waterinfo, houden ze niet apart bij	nvt	Nee
Stad Roeselare	ja		ja	ja
SCK-CEN	nee			nee
Stad Sint-Niklaas	ja		ja	ja

Instantie	Reactie ontvangen?	Aangeschreven instanties Opmerking	Mag ook na project gedeeld worden?	Bijkomende data aangeleverd
Stad Turnhout	ja	Geen bijkomende metingen	nvt	nee
UAntwerpen	ja	Data beschikbaar in Sigma gebied, mag niet publiek gemaakt worden gezien onderzoek nog niet gepubliceerd is.	nee	nee
UGent	ja	Meerdere contacten benaderd. Indien bijkomende data aanwezig is: meestal niet in bruikbaar formaat.	nvt	nee
Vereniging van Vlaamse Polders en Wateringen (VVPW)	nee	We contacteren de belangrijkste Polders afzonderlijk	nvt	nee
VITO	ja	Doorverwezen naar collega	nvt	nee
VLM	ja	Doorverwezen naar VMM	nvt	nee
VUB	ja	Geen data	nvt	nee
Westkustpolder	ja	Geen grondwatermetingen, wel oppervlaktepeilen die niet op Waterinfo staan	ja	ja
Zuidijzerpolder	ja	Geen bijkomende metingen	nvt	nee

APPENDIX 2: WATERINFO KWALITEITSCODES

VMM Quality Code Interpretation

10/110 - Excellent
30/100/130 - Good
50/150 - Moderate
70/170 - Poor
80/180 - Estimated
90/190 - Suspect
220 - Default - Unchecked
-1 - Missing

HIC Quality Code Interpretation

6-8 External validated data (6=good, 7=estimated, 8=suspect)
10-19 Good measurements
20-29 Good calculations
30-39 Estimate measurements
40-49 Estimate calculations
60-69 Suspect measurements
70-79 Suspect calculations
110-179 Unchecked
221-223 Unknown (import)
255 of -1 Missing