

BTO 2019.205a (s) | Mei 2019

BTO rapport

Oorzaak piekverbruik
technisch memo

BTO

Oorzaak piekverbruik

technisch memo

BTO 2019.205a (s) | Mei 2019

Opdrachtnummer

402045/057

Projectmanager

drs. P.G.G. (Nellie) Slaats

Opdrachtgever

BTO - Bedrijfsonderzoek

Kwaliteitsborger(s)

dr. P (Peter) van Thienen

Auteur(s)

B (Bram) Hillebrand MSc., E (Erwin) Vonk MSc.,
drs. B.W. (Bernard) Raterman

Verzonden naar

Dit rapport is verspreid onder BTO-participanten.
Een jaar na publicatie is het openbaar

Jaar van publicatie
2019

Meer informatie

Bram Hillebrand MSc
T +31 (0)30 60 69 590
E bram.hillebrand@kwrwater.nl

Keywords

Postbus 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl



BTO 2019.205a (s) | Mei 2019 © KWR

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden veelevoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Inhoud

Inhoud	2	
1	Introductie	3
2	Data	4
2.1	Gebieden	4
2.2	KNMI-data	4
2.3	CBS data	5
2.4	Vakantieafwezigheid	5
2.5	Tuinoppervlak	5
2.6	Overige gebiedsindicatoren	5
2.7	Overige data	6
3	Scenario-ontwikkeling	7
4	Methode	9
4.1	Support Vector Regressie (SVR) model	9
4.2	Stappenplan methode	9
5	Resultaten	13
5.1	Oorzaken piekverbruik	13
5.2	Plasmolen als toeristisch gebied	15
5.3	Gebiedsonafhankelijk regressiemodel	15
6	Conclusie	17
	Bibliografie	19

1 Introductie

In deze studie wordt het effect onderzocht van klimaatverandering aan de hand van de verschillende KNMI '14 klimaatscenario's (KNMI 2014) op de watervraag. Specifiek gaat het om de dagfactor (het gemeten verbruik van een dag t.o.v. het jaargemiddelde) en de dagpiekfactor (dagfactor overeenkomend met een herhalingstijd van eens in de 10 jaar). Deze studie bouwt verder op bestaand BTO onderzoek uit het BTO-rapport "De gevolgen van klimaat verandering en vakantiespreiding voor de drinkwatervraag" (Vonk, Cirkel en Leunk 2017). De methode is in grote lijnen hetzelfde en wordt in dit memo dan ook kort toegelicht. Uitgebreidere toelichting is in het genoemde rapport te vinden.

Voor deze studie is een aangepaste versie van het EDWARD-model gebruikt en in een offline omgeving gedraaid. Het EDWARD-model is toegepast op de voorzieningsgebieden Plasmolen (WML) en Peelkant (Brabant Water). In dit memo staan kort de benodigde data beschreven, de methode beknopt uitgelegd en vervolgens de resultaten kort toegelicht. Tijdens het tussentijds projectoverleg is besloten dat er voor Roosendaal geen resultaten zouden worden doorgerekend. Dit in verband met het waterverbruik van de bietencampagne die in een deel van de dagafzet-metingen zat waardoor deze data niet bruikbaar waren voor analyse.

Onderdeel van deze studie is ook om meer inzicht te krijgen in de oorzaken van piekverbruik. Hiervoor is het nodig om een één enkel model te maken zodat onderzocht kan worden welke gebiedsspecifieke kenmerken zoals bevolkingssamenstelling en tuinarealen invloed hebben op het piekverbruik.

Dit memo is als volgt opgebouwd. Eerst wordt de input data voor het model beschreven, daarna een korte toelichting over de scenario's en de methode. Dan zijn de resultaten beschreven en tenslotte in een conclusie kort samengevat.

2 Data

2.1 Gebieden

In deze studie is gebruik gemaakt van data van Brabant Water (voorzieningsgebied Peelkant) en van WML (voorzieningsgebied Plasmolen). Tevens is er gebruik gemaakt van de data uit het vorige BTO-rapport (Vonk, Cirkel en Leunk 2017), namelijk van Waterbedrijf Groningen (Groningen Stad en Groningen Provincie), Waternet (Amsterdam, Heemstede) en PWN (HAU, 't Gooi en Texel). Dit geeft een totaal van 9 gebieden van verschillende grootte verspreid over het hele land. De weerstations waarvan de bij deze gebieden behorende KNMI-data verkregen zijn, zijn gegeven in Tabel 1. Hier zijn ook de keuzes voor de andere gebiedseigenschappen weergegeven.

Tabel 1: Gebiedsspecifieke informatie.

Gebied	Weerstation (naam en code)	Vakantiegebied	Toeristisch	Landelijk of stedelijk
Amsterdam	Schiphol, 240	Noord	Nee	Stedelijk
Groningen stad	Eelde, 280	Noord	Nee	Stedelijk
Groningen provincie	Eelde, 280	Noord	Nee	Landelijk
HAU	Schiphol, 240	West	Nee	Landelijk
't Gooi	Schiphol, 240 (onvoldoende straling gegevens bij Soesterberg)	West	Nee	Landelijk
Heemstede	Schiphol, 240	West	Nee	Stedelijk
Texel	De Kooy, 235	Noord	Ja	Landelijk
Peelkant	Eindhoven, 370	Zuid	Nee	Landelijk
Plasmolen	Volkel, 375	Zuid	Nee	Landelijk

Het vakantieveld wordt gebruikt voor de vakantieafwezigheidsdata. Daarnaast wordt er gebruik gemaakt van CBS-data: bevolkingsdichtheid (inwoners/m²) en het percentage inwoners dat valt in de leeftijdscategorieën 0-15, 15-25, 25-45, 45-65 en 65 en ouder. Deze CBS-data worden op jaarbasis gebruikt.

Als laatste is er per gebied ook het tuinoppervlak (m²/perceel) uitgerekend. Hieronder volgt een specifiekere uitleg.

2.2 KNMI-data

Via de door KNMI beschikbaar gestelde API (KNMI 2019) worden de daggegevens van elk weerstation behorend bij een gebied automatisch opgehaald. Het gaat om:

- neerslag (mm/dag);
- verdamping (mm/dag);
- gemiddelde temperatuur (°C);
- maximale temperatuur (°C);
- globale straling (J/cm²).

Uit verdamping en neerslag wordt een neerslagtekort (mm/dag) berekend en in het groeiseizoen wordt dit cumulatief berekend tot Potentieel Doorlopend Neerslagtekort: PDN (mm). Iedere dag bevat ook waarden voor neerslag, verdamping, maximale temperatuur en globale straling van de dag daarvoor, twee dagen daarvoor en drie dagen daarvoor.. Alle KNMI-data worden toegepast op dagbasis.

2.3 CBS data

CBS-data zijn handmatig gedownload van de CBS-website open data statline (CBS 2019). Het gaat om inwoneraantallen op 1 januari per jaar in de categorieën: totaal, 0-15, 15-25, 25-45, 45-65 en 65 en ouder. Deze zijn vervolgens, met behulp van gebiedsgrootte, omgerekend tot bevolkingsdichtheid (inwoners/m²) en het percentage inwoners dat valt in de leeftijdscategorieën 0-15, 15-25, 25-45, 45-65 en 65 en ouder. Deze gegevens worden toegepast op jaarbasis.

2.4 Vakantieafwezigheid

De vakantieafwezigheidsgegevens per regio zijn door het CBS reeds aangeleverd voor Vonk, Cirkel en Leunk (2017). Deze zijn ongewijzigd gebruikt.

2.5 Tuinoppervlak

Voor het bepalen van het totale oppervlak aan tuinen is gebruik gemaakt van de volgende data:

1. begrenzing in de vorm van shapefiles van de distributiegebieden;
2. Digitale Kadastrale Kaarten (DKK);
3. Basisadministratie Adressen en Gebouwen (BAG);
4. Basisadministratie Grootschalige Topografie.

Vervolgens zijn in ArcGis Desktop 10.5 de volgende stappen toegepast:

1. Eerst zijn de kadastrale percelen (dataset 2) met een woonfunctie (dataset 3) in de distributiegebieden (dataset 1) geselecteerd.
2. De erven (dataset 4) binnen de geselecteerde percelen zijn bepaald.
3. Vervolgens is een gemiddelde per perceel bepaald en gebruikt als invoer voor het EDWARD-model.

Voor een erf is de definitie in het BGT-objectenhandboek gevolgd, namelijk een *“Terreindeel dat bij een pand of overig bouwwerk hoort, dat niet nader wordt ingewonnen en dat bestaat uit een mengvorm van begroeiing, verharding, en/of water.”* (Ministerie van Infrastructuur en Milieu 2019)

Verharde en niet verharde delen worden dus allemaal tot het tuinoppervlak gerekend. Verder is het belangrijk om aan te merken dat de gemeente en andere bronhouders nog werken aan de harmonisatie van de BGT. Soms is er geen erf vastgelegd terwijl dit duidelijk te zien is op een luchtfoto. Dit kan dus leiden tot een (kleine) onderschatting van het totale oppervlak. Het gebrek aan onderscheid tussen verhard en onverhard zal weer leiden tot een (kleine) overschatting van het totale oppervlak. De tuinoppervlak data is per gebied.

2.6 Overige gebiedsindicatoren

Zoals te zien is in Tabel 1, is de data voor ieder gebied aangevuld met enkele gebiedsindicatoren. Deze zijn bedoeld als invoer voor het machinelearningmodel (paragraaf 1.4) om beter onderscheid te kunnen maken tussen verschillende soorten gebieden. Het gaat hier om het onderscheid tussen al dan niet toeristische gebieden en landelijke of stedelijke gebieden. De waarden voor deze parameters zijn door de

onderzoekers gekozen. In de toekomst zouden hier meer objectieve criteria voor kunnen worden opgesteld. Deze data zijn per gebied toegepast.

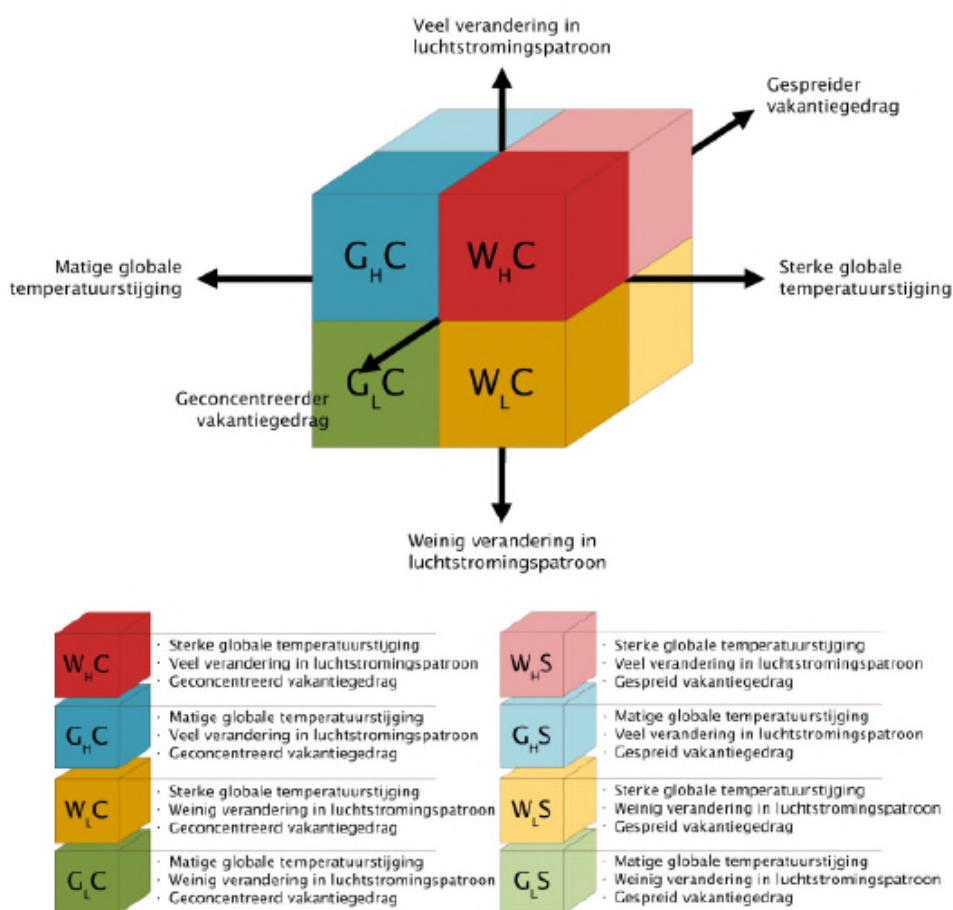
2.7 Overige data

Bovengenoemde data worden verzameld per dag. Iedere dag is dus een datapunt met:

- KNMI-data (elke dag anders) deze data wordt door het programma opgehaald;
- CBS data (varieert per jaar) deze data dient handmatig gedownload te worden;
- afwezigheidsdata (elke dag anders) deze data wordt door het programma gegenereerd;
- tuinooppervlak en overige gebiedsindicatoren (alleen anders per gebied, constant voor alle jaren) deze data dient handmatig te worden gegenereerd..

Deze data worden aangevuld door het script met informatie tot welke maand de betreffende dag behoort en of de dag in het weekend valt.

3 Scenario-ontwikkeling



Figuur 1: De onderlinge verhouding tussen de verschillende scenario's die gebruikt zijn in deze studie. De eerste twee letters verwijzen naar de KNMI '14 scenario's de derde naar de vakantieafwezigheidsscenario's. Hierbij is 'W' een sterke temperatuurstijging en 'G' een mildere. Subscript 'H' een grote verstoring van de luchtstroming boven Nederland en 'L' een relatief kleine. 'C' en 'S' geven een geconcentreerde of gespreide vakantieafwezigheid aan. Deze scenario's zijn identiek aan die van (Vonk, Cirkel en Leunk 2017)

In deze studie worden dezelfde toekomst scenario's toegepast als in Vonk, Cirkel en Leunk (2017). Het gaat hier om een combinatie tussen de 8 KNMI '14 klimaatscenario's (4 voor 2 horizonjaren (2050 en 2085) (KNMI 2014)) welke het mogelijke toekomstige weer beschrijven voor Nederland en 2 vakantieafwezigheidsscenario's (25% versterkt (C) en een 25% verzwakt (S) spreidingsmodel ten opzicht van de huidige situatie). Dit geeft een totaal van 16 scenario's.

In Figuur 1 zijn de verhoudingen van de verschillende scenario's weergegeven. Tijdens het tussentijdse overleg is besloten om vooral te focussen op 2050 omdat 2085 voor waterbedrijven te ver in de toekomst ligt.

Aan de hand van een KNMI-klimaattransformatieprogramma (Bakker 2015) kunnen historische tijdreeksen van meteogegevens worden omgerekend naar hoe deze er uit zouden zien onder de nieuwe klimaatomstandigheden van een van de klimaatscenario's.

4 Methode

4.1 Support Vector Regressie (SVR) model

Het voor deze studie gebruikte machinelearningalgoritme, *Support Vector Regression*, is gelijk aan dat van de vorige studie (Vonk, Cirkel en Leunk 2017) en voor een uitgebreidere beschrijving verwijzen we daarnaar. Het algoritme tracht een functie te fitten op de dataset. Hierbij is dagafzet de afhankelijke variabele en alle andere data de onafhankelijke variabelen. Support Vector Regression is toegepast met behulp van Scikit-Learn (Pedregosa, et al. 2011). Er is gekozen voor een 3^e graads polynoomkernel. De data zijn 80-20 gesplitst in een trainings- en een test dataset. Waar in de vorige studie het model geoptimaliseerd (parameterselectie) en getraind werd per gebied, zijn in deze studie alle data van de verschillende gebieden bij elkaar genomen en is getracht om één enkel model te optimaliseren en te trainen. Dit is gedaan met als doel een gebiedsonafhankelijk model te maken, hetgeen vervolgens toepasbaar zou moeten zijn op elk willekeurig gebied zonder dat daar 10 jaar aan trainingsdata van het betreffende gebied voor nodig is. In Tabel 2 zijn enkele in- en uitvoergegevens van het SVR model gegeven. Epsilon heeft betrekking tot de bandbreedte waarbinnen voorspellingen moeten vallen en C heeft betrekking op de “strafpunten” die worden toegepast wanneer een voorspelling buiten de bandbreedte ligt. Hiermee kan er worden bewogen tussen modellen die meer de overall trend volgen of (bij hoge C) die elke grilligheid mee zullen proberen te nemen.

De verschillende R^2 scores geven aan hoe goed de modellen de data voorspellen. Wanneer het model de observaties perfect voorspelt zal de score 1 zijn. Hoe lager de score hoe minder goed het model wordt voorspelt door de data. Er is geen simpele regel voor wat wel en wat niet acceptabele scores zijn omdat dit afhankelijk is van wat er gemodelleerd wordt en wat precies het doel is. In dit memo maken we gebruik van een kleurcode schaal. Rood voor minder dan 0.35, oranje voor 0.35 tot 0.5, lichtgroen voor 0.5 tot 0.7 en groen voor daarboven. Dit is een kwalitatieve schaal om een indicatie van de kwaliteit van de modellen weer te geven.

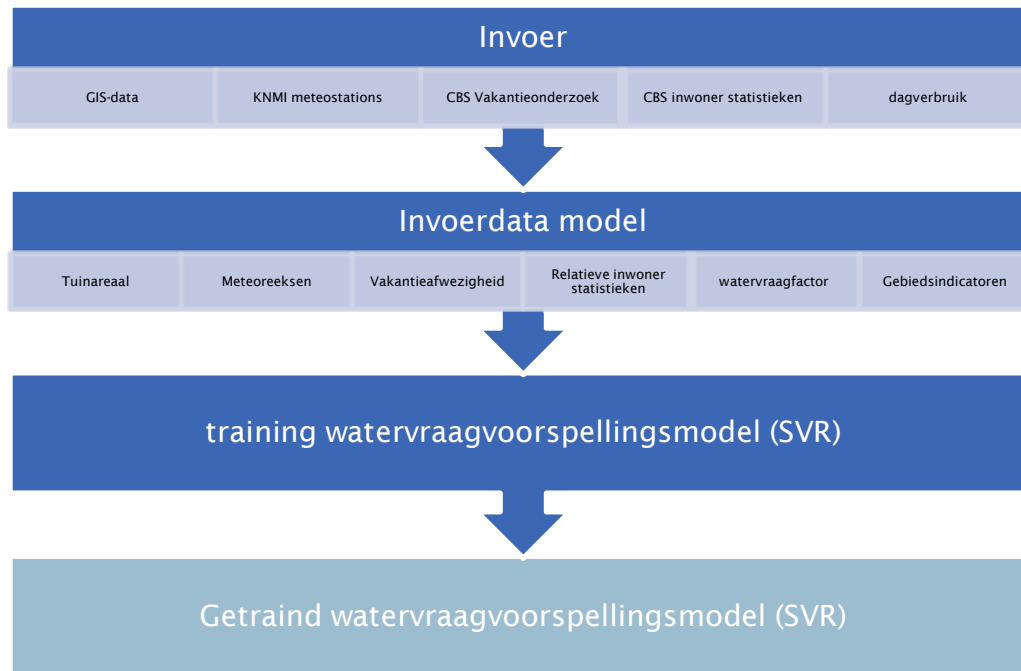
Tabel 2: Enkele in- en uitvoergegevens van het supportvectorregressiemodel.

C	epsilon	R^2 training	R^2 test
0,0075	0,015	0,71	0,69

4.2 Stappenplan methode

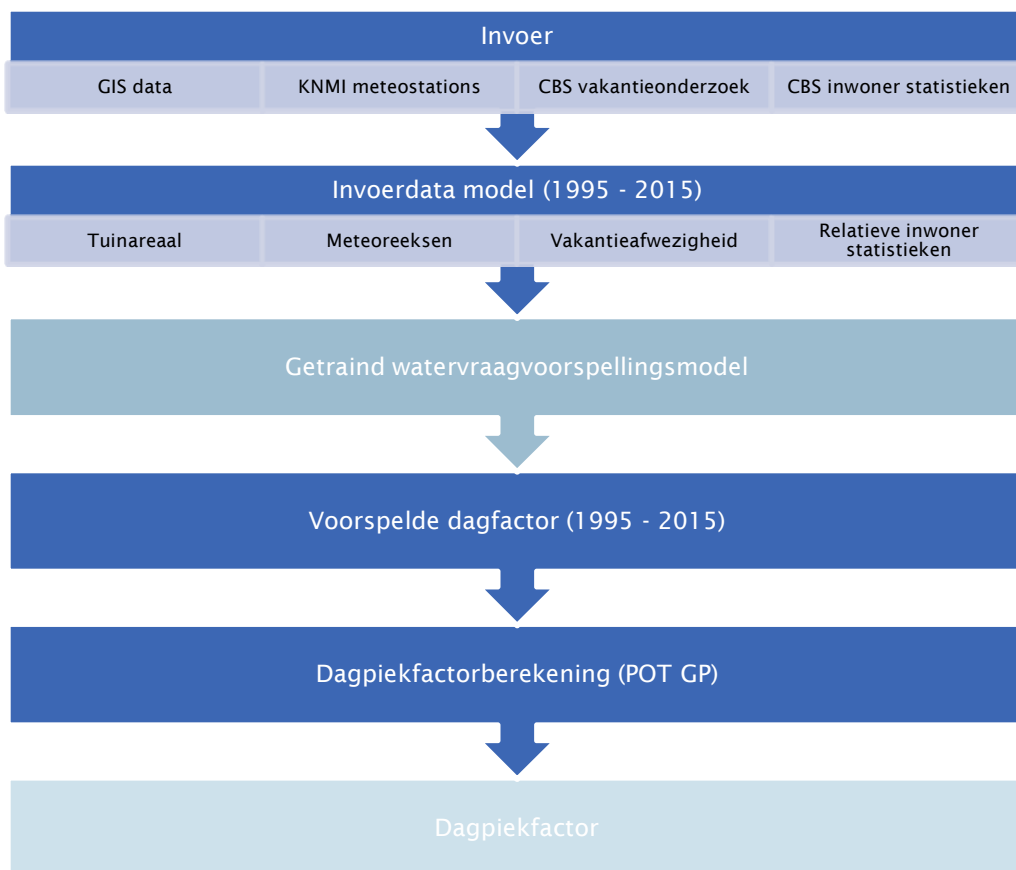
In Figuur 2 zijn schematisch de stappen weergegeven die leiden tot een getraind watervraagvoorspellingsmodel. Eerst worden alle input data (GIS, CBS, KNMI en dagverbruik) verzameld. Deze worden vervolgens omgerekend tot de daadwerkelijke invoerdata. Met deze data, de aangeleverde dagverbruikdata en de gebiedsindicatoren (toeristisch, landelijk/stedelijk) wordt vervolgens een supportvectorregressiemodel (SVR) getraind. Hieruit volgt een getraind watervraagvoorspellingsmodel, dat wordt getoetst aan de hand van een apart gehouden trainingsdataset. De testscore (R^2) van het getrainde model is 0,69 (zie ook Tabel 2).

Voordat het getrainde model gebruikt wordt, is er ook per gebied een zogenoemde R^2 -areascor berekend. Het gaat hier om de R^2 -score berekend voor alle invoerdata van een bepaald gebied (dus zowel de datapunten die in de trainings-, als die in de testset zaten). Dit geeft een indicatie van hoe goed het model werkt voor bepaalde gebieden. Voor Peelkant is deze score 0,71 voor Plasmolen 0,61. Het model geeft dus een betere voorspelling voor Peelkant dan voor Plasmolen.



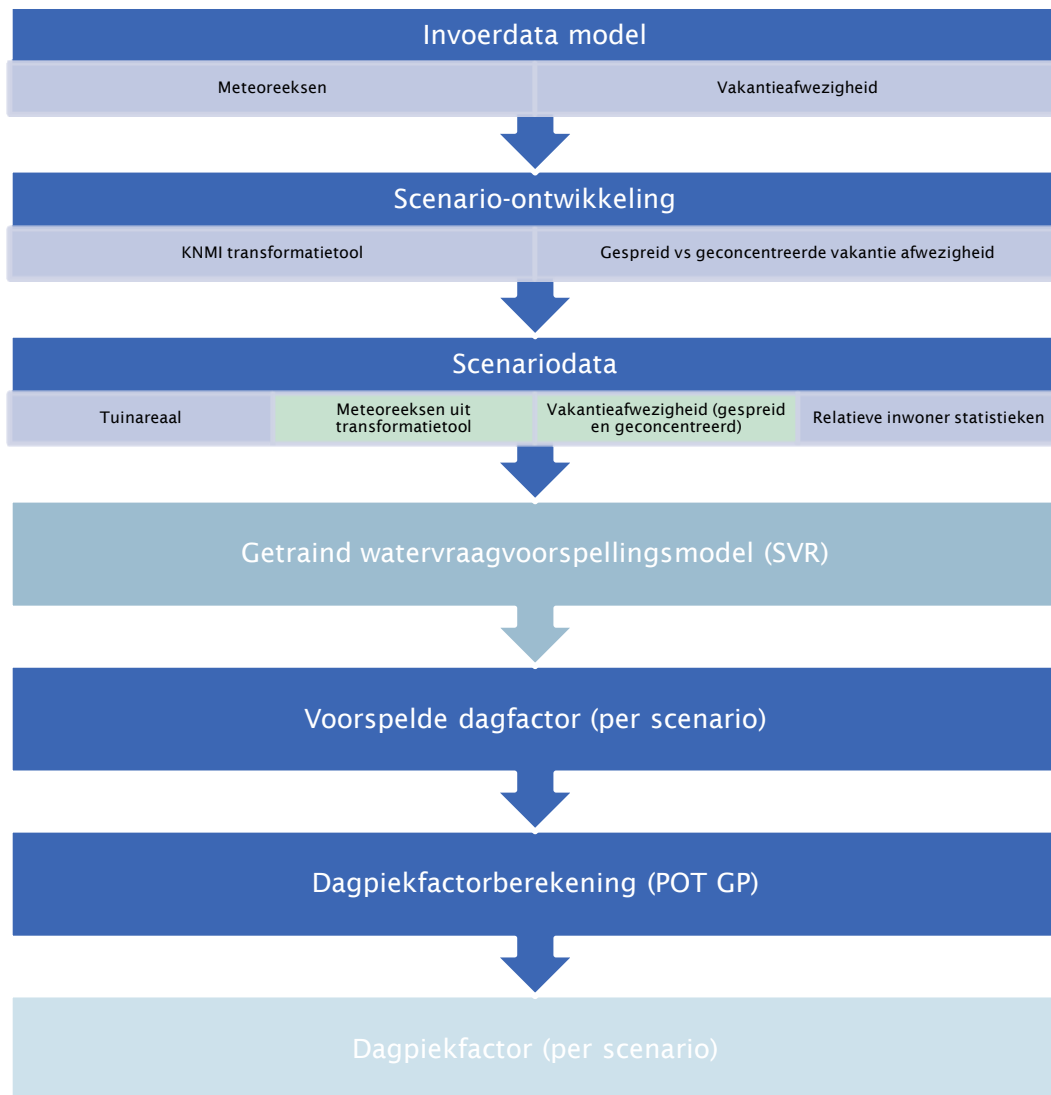
Figuur 2: Schematisch overzicht voor het verkrijgen van een getraind watervoorspellingsmodel

Met het getrainde watervraagvoorspellingsmodel wordt er vervolgens een referentieperiode doorgerekend (Figuur 3). Hiervoor worden als invoer de data gebruikt van 1995-2015 van het KNMI en het CBS. Tuinareaal en andere gebiedsindicatoren worden als constant beschouwd (tuinareaal van 1995 is hetzelfde als dat van 2015). Deze data worden omgerekend tot de daadwerkelijke invoerdata en hieruit volgt een dagfactor en vervolgens een dagpiekfactor voor 1995-2015. De dagpiekfactor is het resultaat van een statistische extremewaardenberekening, in dit geval de Peak-Over-Threshold (POT). Details over deze methode en hoe deze is toegepast zijn beschreven in Vonk, Cirkel en Leunk (2017).



Figuur 3: Schematisch overzicht voor het verkrijgen van een referentiepiekfactor (voor 1995-2015). Hiervoor wordt het getrainde model gebruikt dat verkregen is volgens Figuur 2.

Vervolgens worden de scenario's doorgerekend (Figuur 4). Als extra stap hierin zit de scenario-ontwikkeling waarin de KNMI-data door de KNMI-transformatietool (Bakker 2015) worden bewerkt om de meteorologische reeksen behorend bij een klimaatscenario te berekenen. Ook wordt er voor elk klimaatscenario een variant gemaakt met een gespreide vakantieafwezigheid en een geconcentreerde vakantieafwezigheid. Vervolgens wordt deze invoer aan het getrainde watervraagvoorspellingsmodel aangeboden waaruit, per scenario, een voorspelde dagfactor en vervolgens een dagpiekfactor volgt. Deze worden vervolgens vergeleken met de referentie-dagpiekfactor om de mate van toe- of afname te bepalen.



Figuur 4: Schematisch overzicht om tot een piekfactor per scenario te komen. Hiervoor wordt het getrainde model gebruikt dat verkregen was volgens Figuur 2. Dit schema verschilt met dat van Figuur 3 in de scenario-ontwikkelingsstappen.

5 Resultaten

In Figuur 7 zijn de resultaten voor Plasmolen en Peelkant voor de verschillende scenario's van 2050 weergegeven t.o.v. de referentieperiode. De resultaten laten een toename zien van de piekfactor voor ieder klimaatscenario en ieder vakantieafwezigheidsscenario. Afhankelijk van het scenario ligt die toename tussen de ~5% en de ~17%. Dit is vergelijkbaar met gebieden als HAU, Heemstede en 't Gooi uit de vorige studie (Vonk, Cirkel en Leunk 2017). De grootste toename wordt bereikt voor de WHC- en WHS- scenario's. Dit zijn respectievelijk een versie met geconcentreerde vakantieafwezigheid (C) en een met gespreide vakantieafwezigheid (S) van het W_n KNMI'14 klimaatscenario, welke gekarakteriseerd wordt door een sterke globale temperatuurstijging en veel verandering in het luchtstromingspatronen boven Nederland. Ook dit resultaat is gelijk aan dat van de vorige studie.

De absolute waarden van de piekfactoren per scenario en voor de referentieperiode zijn gegeven in Tabel 3

Tabel 3: Absolute waarden van de piekfactoren voor de verschillende scenario's en de referentieperiode.

	1995-2015	GHC 2050	GHS 2050	GLC 2050	GLS 2050	WHC 2050	WHS 2050	WLC 2050	WLS 2050
Plasmolen	1,73	1,89	1,88	1,83	1,82	2,01	2,01	1,81	1,82
Peelkant	1,66	1,84	1,85	1,76	1,77	1,93	1,95	1,76	1,76

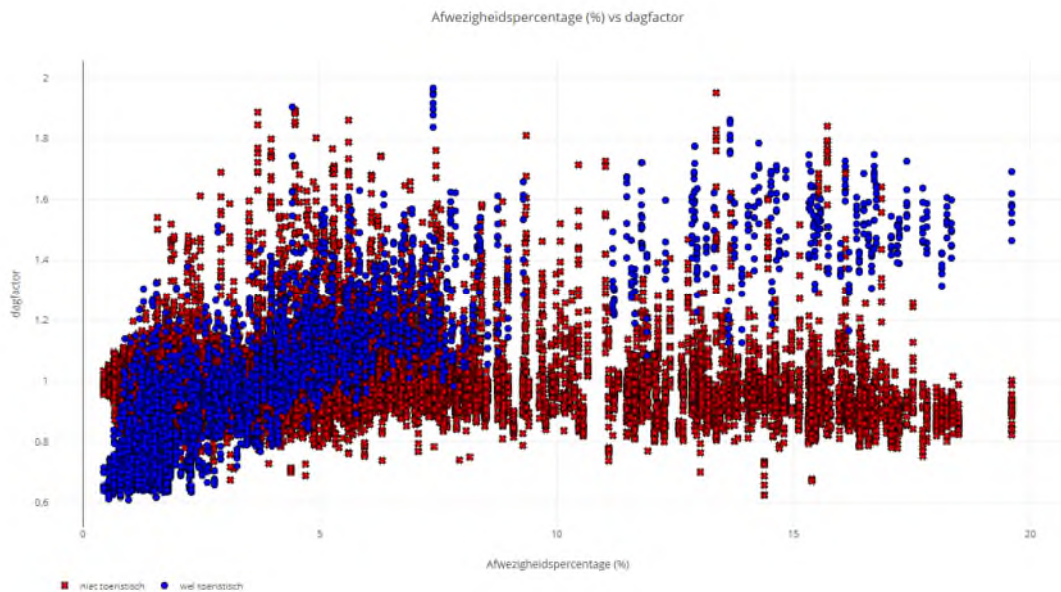
De te verwachten maximale piekfactor ligt rond de 2 voor beide voorzieningsgebieden.

5.1 Oorzaken piekverbruik

Doordat het regressiemodel gebiedsonafhankelijk is, is het niet alleen mogelijk te identificeren welke weerfactoren de grootste invloed hebben op het piekverbruik, maar is het ook mogelijk om verschillen tussen voorzieningsgebieden te verklaren. Uit analyse blijkt dat verdamping, toeristische functie, tuinoppervlak en afwezigheidspercentage belangrijke factoren zijn waarop het model zijn voorspelling baseert. Omdat het een niet-lineair probleem betreft dat gemodelleerd is met een niet-lineair regressiemodel, kunnen we uitsluitend kwalitatieve relaties ontleen.

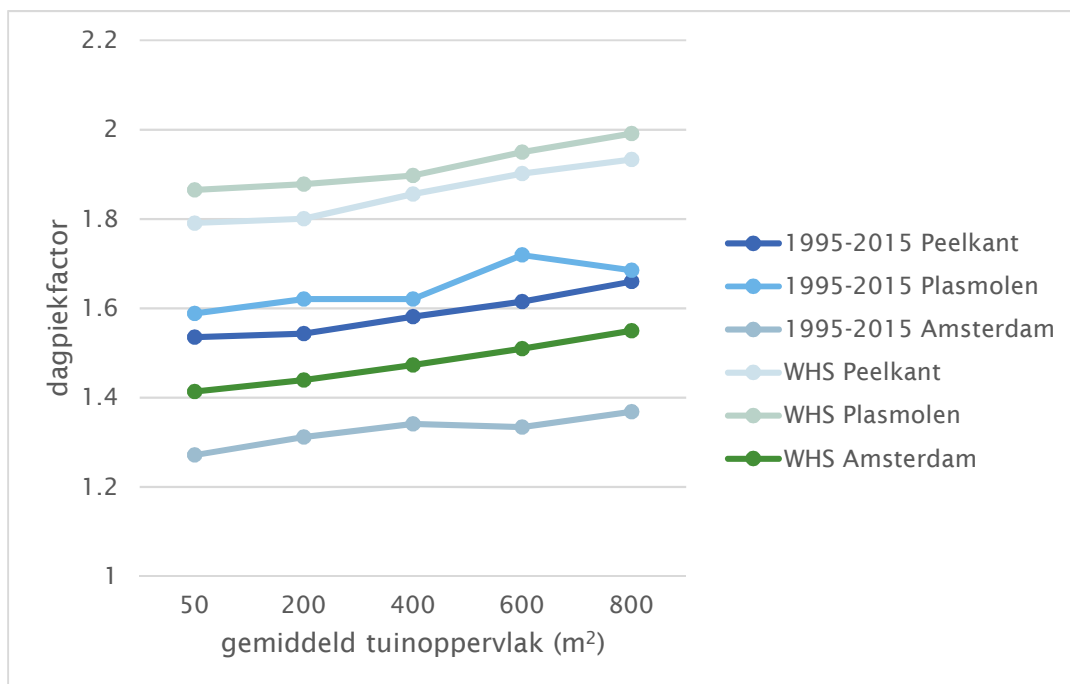
De indicator toeristisch geeft aan hoe een gebied reageert op vakantieafwezigheid. In een toeristisch gebied zal een hoge vakantieafwezigheid juist voor netto meer verbruik zorgen, dit in tegenstelling tot niet-toeristische gebieden. Deze relatie is ook weergegeven in Figuur 5. Voor niet-toeristische gebieden leidt een hoge vakantieafwezigheid tot een lagere watervraagfactor. Voor toeristische gebieden is dit precies andersom.

Verdamping blijkt de belangrijkste weerfactor te zijn waarop het model zijn voorspelling baseert. Een hogere verdamping leidt tot een hogere watervraagfactor.



Figuur 5: Watervraagfactor afgezet tegen afwezigheidspercentage voor toeristische en niet-toeristische gebieden.

Om in te schatten op welke manier tuinoppervlak belangrijk is, zijn er voorspellingen gemaakt voor fictieve gebieden waarbij alles gelijk is aan het daadwerkelijke voorzieningsgebied behalve het tuinoppervlak (m^2 /perceel). Dit is op 50, 200, 400, 600 en $800 m^2$ gezet voor Amsterdam, Peelkant en Plasmolen. De resultaten zijn weergegeven in Figuur 6. Omdat het hier om fictieve gebieden gaat moeten we voorzichtig zijn om harde conclusies te trekken, maar de trend is dat bij een groter tuinoppervlak (m^2 /perceel) een grotere dagpiekfactor wordt voorspeld. Het is belangrijk om te noemen dat het model is getraind met 9 verschillende gebieden en dus met 9 verschillende gemiddelde tuinarealen. Deze beperkte hoeveelheid data verklaard waarschijnlijk waarom de "1995-2015 Plasmolen" en "1995-2015 Amsterdam" data voor $600 m^2$ beide een afwijking laten zien van de algemene opgaande trend. Om een beter begrip te krijgen voor de exacte invloed van tuinareaal is het wenselijk om andere gebieden met verschillende gemiddelde tuinarealen toe te voegen in de trainingsfase van het model.



Figuur 6: Piekfactor voor drie bestaande gebieden, voorzien van vergrote fictieve tuinoppervlakken.

5.2 Plasmolen als toeristisch gebied

In de vorige studie liet elk gebied behalve Texel een grotere toename zien van de piekfactor voor de gespreide (S) vakantieafwezigheid versie van een klimaatscenario dan de geconcentreerde (C) vakantieafwezigheid versie. Voor voorzieningsgebied Texel was dit precies andersom. Dit werd verklaard door het feit dat vakantieafwezigheid voor Texel in praktijk een toename in het aantal gebruikers betekent door het toeristische karakter van Texel. Het is daarom interessant om te zien dat, ondanks dat Plasmolen als niet toeristisch is aangemerkt in deze studie, voor sommige klimaatscenario's variant met geconcentreerde vakantiespreiding tot een hogere toename van de piekfactor leidt dan die met een grotere mate van vakantiespreiding (analoog aan Texel in de vorige studie). Dit suggereert dat Plasmolen wellicht beter als toeristisch gebied aangemerkt zou kunnen worden vanwege de aanwezigheid van bijvoorbeeld camping Eldorado.

Hiervoor is een korte test uitgevoerd waarin er een nieuw watervoorspellingsmodel getraind is waar Plasmolen als toeristisch was aangemerkt om vervolgens de R^2 -areascor te bepalen. Voor het in deze studie gebruikte watervoorspellingsmodel was de R^2 -areascor van Plasmolen 0,61. Voor dit nieuwe watervoorspellingsmodel waarin Plasmolen als toeristisch wordt aangemerkt is deze ook 0,61. Dit duidt erop dat Plasmolen zowel toeristisch als niet toeristisch reageert. Het is dan ook aan te bevelen om een andere meer geleidelijke schaal voor toerisme te ontwikkelen in plaats van de binaire onderscheid 'wel' of 'niet' (toeristisch).

5.3 Gebiedsonafhankelijk regressiemodel

Een van de doelen van dit project is om tot een gebiedsonafhankelijk regressiemodel te komen. Dit is nodig om de invloed van de hierboven besproken gebiedsindicatoren te onderzoeken, maar het biedt ook de mogelijkheid om het model toe te passen op gebieden waarvoor geen 10 jaar aan trainingsdata beschikbaar zijn.

Om te testen of dit mogelijk is, zijn er verschillende nieuwe regressiemodellen getraind. In elk regressiemodel ontbrak een van de negen gebieden tijdens de trainingsfase. Vervolgens is er een voorspelling gedaan voor het waterverbruik in dat gebied en vergeleken met de daadwerkelijke waterverbruik. Hieruit is een R^2 areascore bepaald. Deze wordt vergeleken met de R^2 areascore van het desbetreffende gebied van het watervraagvoorspellingsmodel (waarbij alle negen gebieden in de trainingsset aanwezig waren). De resultaten hiervan staan in Tabel 4.

Tabel 4: De R^2 areascores voor de verschillende regressiemodellen

Gebied	R^2 areascore van het watervraagvoorspellingsmodel	R^2 areascore voor het nieuw getrainde model waarbij het desbetreffende gebied ontbrak tijdens de training
Heemstede	0.56	0.37
Amsterdam	0.66	< 0
Texel	0.88	0.15
Plasmolen	0.61	0.53
Peelkant	0.71	0.64
Groningen Stad	0.49	< 0
Groningen Provincie	0.58	0.23
HAU	0.63	0.18
't Gooi	0.65	0.55

De resultaten geven een gemengd beeld. Voor alle gebieden is er een betere R^2 areascore voor het model waarbij hun data ook in de training aanwezig was dan wanneer dit niet het geval was. Voor de voorzieningsgebieden Peelkant, Plasmolen en 't Gooi en in mindere mate Heemstede (omdat hier de originele fit ook niet heel hoog is) zijn de verschillen echter niet groot en deze resultaten lijken te suggereren dat een gebiedsonafhankelijk model zeker mogelijk is. Voor andere voorzieningsgebieden scoort het model dat getraind was zonder hun data echter zeer slecht. Voor bijvoorbeeld Texel is dit zeer begrijpelijk. Omdat Texel het enige toeristische gebied is van de negen. Dat betekent dat wanneer deze data weggelaten wordt het model geen enkel notie heeft hoe een toeristisch gebied anders reageert op vakantieafwezigheid (zie Figuur 5) en dit dus ook verkeerd zal voorspellen. Voor Amsterdam geldt dat door het weglaten van deze data het model niet langer getraind wordt op kleine tuinen en grote bevolkingsdichtheid cijfers. Voor het voorzieningsgebied Groningen provincie geldt weer dat dit het enige gebied is van deze grootte. Voor Groningen stad in HAU is het niet geheel duidelijk wat de reden is dat deze gebieden blijkbaar anders reageren dan andere gebieden. Wat deze resultaten wel laten zien is dat 1) een gebiedsonafhankelijk model zeker tot de mogelijkheden hoort maar 2) dat de huidige hoeveelheid aan verschillende gebieden nog niet voldoende is.

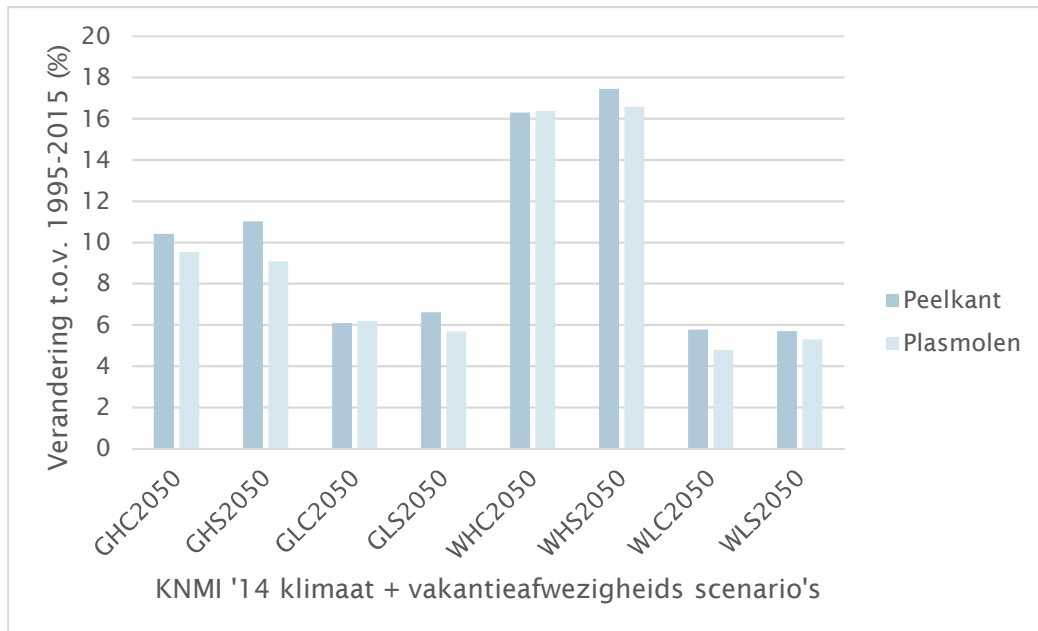
6 Conclusie

Voor Plasmolen wordt voor 2050 een piekfactor verwacht van tussen de 1,8 en 2,0, hetgeen neerkomt op een procentuele toename van ~5 - ~17%. Voor Peelkant wordt een piekfactor verwacht van tussen de 1,75 en 1,95, hetgeen ook neerkomt op een toename van ~5 - ~17%. Voor beide gebieden geldt dat het WHS-scenario, het W_H KNMI '14 klimaatscenario in combinatie met een gespreide (S) vakantieafwezigheid, tot de hoogste voorspelde piekfactor leidt. Dit is in lijn met de bevindingen van Vonk, Cirkel en Leunk (2017). Het is belangrijk om hierbij op te merken dat enkel alleen de hoogte van de piekfactor in 2050 is onderzocht. Hoe de piekfactor de komende jaren zal toenemen tot 2050 (linear of anderszins) is niet onderzocht.

Een groter tuinoppervlak leidt tot een grotere piekfactor. Verder is verdamping een belangrijke voorspellende factor voor de watervraagfactor. Doordat toeristische gebieden anders reageren op vakantieafwezigheid (toename in plaats van afname van watervraagfactor voor hogere afwezigheid), is ook de toeristische functie een voorspellende factor. Het is echter voor toekomstige studies aan te raden om een meer geleidelijke schaal voor de toeristisch functie van een bepaald gebied in te voeren in plaats van het huidige binaire onderscheid.

Initiële testen laten zien dat een gebiedsonafhankelijk watervraagvoorspellingsmodel tot de mogelijkheden behoort. Het is wel nodig om de trainingsdataset uit te breiden

met meer verschillende gebieden om tot een robuust gebiedsonafhankelijk model te komen.



Figuur 7: Procentuele verandering van de piekfactor t.o.v. de referentiepiekfactor (1995-2015) voor Plasmolen en Pealkant voor alle klimaatscenario's van 2050.

Bibliografie

- Bakker, A. 2015. *Time series transformation tool: description of the program to generate time series consistent with the KNMI'14 climate scenarios*. De Bilt: Technical report TR-349.
- CBS. 2019. *CBS Open data StatLine*. Februari 25. https://opendata.cbs.nl/statline/portal.html?_la=nl&_catalog=CBS.
- KNMI. 2019. *Data ophalen vanuit een script*. Februari 25. <https://www.knmi.nl/kennis-en-datacentrum/achtergrond/data-ophalen-vanuit-een-script>.
- Ministerie van Infrastructuur en Milieu. 2019. *Objectenhandboek BGT | IMGeo`*. Maart 11. <http://imgeo.geostandaarden.nl/def/imgeo-object/onbegroeidterreindeel/erf>.
- Pedregosa, F, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 2825-2830.
- Vonk, E, D.G Cirkel, and I Leunk. 2017. *De gevolgen van klimaatverandering en vakantiespreiding voor de drinkwatervraag*. Nieuwegein: KWR 2017.043.