



Herkenning kruidenrijk grasland middels Remote Sensing & foto's

Eindrapportage Fase 1 – PPS Herkenning Kruidenrijk Grasland

Wouter Meijninger, Janne Kool, Tim Visser, Ron Wehrens, Gert-Jan Noij, Natalie Hotrum

Herkenning kruidenrijk grasland middels Remote Sensing & foto's

Eindrapportage Fase 1 – PPS Herkenning Kruidenrijk Grasland

Wouter Meijninger, Janne Kool. Tim Visser, Ron Wehrens, Gert-Jan Noij, Natalie Hotrum

Wageningen University & Research

Dit onderzoek is in opdracht van Topsector Agri & Food (Samenwerkingsovereenkomst September 2020 TKI LWV20.296) uitgevoerd door de Stichting Wageningen Research (WR), business unit Agrosysteemkunde.

WR is een onderdeel van Wageningen University & Research, samenwerkingsverband tussen Wageningen University en de Stichting Wageningen Research.

Wageningen, februari 2025

Rapport WPR-1309

Dit rapport is gratis te downloaden op <https://doi.org/10.18174/650439>

Samenvatting

Het doel van dit onderzoek is om een wetenschappelijk onderbouwd monitoringssysteem te ontwikkelen waarmee de kruidenrijkdom van grasland kan worden herkend op basis van satelliet Remote Sensing (RS) data, al dan niet in combinatie met foto's uit het veld. De ontwikkelde RS-algoritmes kunnen de kruidenrijkdom (Q-index) van grasland voorspellen. De voorspelfout (RMSE) van de RS-modellen ligt rond de 0,7Q. Dat komt overeen met een 95%-betrouwbaarheidsinterval van $\pm 1,4Q$. Het foto-algoritme voor het bepalen van de kruidenrijkdom (Q-index) leverde de beste resultaten op met behulp van een classificatie model, waarbij de kruidenrijkdom wordt geschat in stapjes van 0,5 van $Q = 1$ tot 5. De fout daarvan correspondeert met een RSME van $\sim 0,89$. Los hiervan is de richting van de voorspelling wel kloppend; voor graslanden, waarvan bekend is dat ze een hogere kruidenrijkdom hebben, komt ook het model met relatief hogere voorspellingen. Dit rapport bevat aanbevelingen voor nader onderzoek om te bepalen welke verbeteringen in de modellen mogelijk zijn. Het foto-algoritme voor het onderscheiden van productief kruidenrijk grasland (PKG) t.o.v. regulier Engels raaigras (spoor 2b) laat daarentegen juist goede resultaten zien. Meer dan 95% van de foto's worden goed geclassificeerd.

Trefwoorden: Kringlooplandbouw, biodiversiteit, kruidenrijkdom, graslanden, Q-index, vegetatieopname, satelliet remote sensing, foto-herkenning, machine learning-algoritmen

© 2024 Wageningen, Stichting Wageningen Research, Wageningen Plant Research, Business unit Agrosysteemkunde, Postbus 16, 6700 AA Wageningen; www.wur.nl/plant-research

KvK: 09098104 te Arnhem
VAT NL no. 8113.83.696.B07

Stichting Wageningen Research. Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier zonder voorafgaande schriftelijke toestemming van Stichting Wageningen Research.

Stichting Wageningen Research is niet aansprakelijk voor eventuele schadelijke gevolgen die kunnen ontstaan bij gebruik van gegevens uit deze uitgave.

Rapport WPR-1309

Foto's omslag: Adrie van Heerden en Rob Geerts

Inhoud

Woord vooraf	7
Samenvatting	9
1 Inleiding	11
2 Methodiek	13
2.1 Overzicht project aanpak	13
2.1.1 Reden tweesporen aanpak	14
2.2 Typologie kruidenrijkdom	15
2.3 Inwinnen vegetatieopnames	17
2.3.1 Vegetatieopnames op plot-niveau	17
2.3.2 Vegetatieopnames op BRP-perceel niveau	19
2.4 Inwinnen foto's	20
2.4.1 Inwinning foto's via de Ellipsis Drive app	21
2.4.2 Inwinning foto's via de ODK app	21
2.4.3 Inwinning metadata via de app	21
2.4.4 Inwinning foto's van Pure Graze en andere projecten	21
2.4.5 Selectie van de bruikbare foto's	22
2.5 Inwinnen RS-data	22
2.5.1 Satelliet data	22
2.5.2 Verwerken RS-data	24
2.5.3 Selectie RS-markers	27
2.5.4 Selectie co-variabelen	31
2.5.5 Overzicht verzamelde RS-markers & co-variabelen	32
2.6 Ontwikkeling RS-algoritme	34
2.6.1 Model selectie	34
2.6.2 Koppeling Q-index en RS-data/markers	36
2.6.3 Dataselectie en voorverwerking, model strategie	37
2.7 Ontwikkeling foto algoritme	39
2.7.1 Fotoselectie en voorverwerking	39
2.7.2 Voorspellen kruidenrijkdom (spoor 2a)	40
2.7.3 Herkennen PKG-percelen (spoor 2b)	42
2.8 Ontwikkeling database (en ontsluiting resultaten)	43
2.8.1 RS-data	43
2.8.2 Foto's	43
3 Resultaten	45
3.1 Resultaten Remote Sensing algoritmes	45
3.1.1 Resultaten RS-algoritmes	45
3.1.2 Bijdrage RS-markers	48
3.1.3 Resultaten voor heel NL en opeenvolgende jaren	49
3.1.4 Twee-modellen en twee-klassen aanpak	53
3.2 Resultaten foto-algoritme	54
3.2.1 Voorspellen kruidenrijkdom (spoor 2a)	54
3.2.2 Herkennen PKG-percelen (spoor 2b)	58
3.3 Remote Sensing - versus foto algoritme	59
3.3.1 Productief kruidenrijke graslanden (PKG)	59
3.3.2 Kruidenrijkdom: RS- versus foto-algoritme	60
3.4 Stakeholder analyse workshop	62
4 Discussie	67

5	Conclusies en aanbevelingen	69
6	Literatuur	71
Bijlage A	Methodiek Q-index, voederwaarde, kruiden- en grassenindex	
Bijlage B	Toelichting RS-data	
Bijlage C	Toelichting RS-markers	
Bijlage D	Estimating biodiversity of grasslands from marker data and RS data	
Bijlage E	Prediction of Herb Richness in grasslands in the Netherlands with Remote Sensing	
Bijlage F	Rapportage PPS herkenning kruidenrijk grasland	

Woord vooraf

Bij de totstandkoming van dit project hebben is veel gesproken met de projectgroep binnen deze PPS, bestaande uit vertegenwoordigers van BoerenNatuur (Linda Romijn, David Kingma en Carleen Webers), NEO en Ellipsis Drive. We zijn hun veel dank verschuldigd voor hun tijd, inzet, kennis en suggesties.

In dit onderzoek is gebruik gemaakt van de vele foto's en veldgegevens die zijn verzameld door de betrokken organisaties (PPS consortium) en vrijwilligers. We zijn hun veel dank verschuldigd voor hun inzet, tijd en geduld (wanneer er weer eens problemen waren met de foto-apps).

De auteurs

Samenvatting

Er is veel aandacht voor de biodiversiteitswaarde van kruidenrijk grasland. Binnen talloze initiatieven en beleidsinstrumenten wordt beoogd het areaal kruidenrijk grasland in het Nederlands agrarisch gebied te vergroten. Echter het monitoring en borging van kruidenrijk grasland middels veldwerk is een tijdrovende onderneming. En op dit moment ontbreekt het aan een geschikt monitoringsinstrument waarmee de kruidenrijkdom van graslandpercelen in Nederland in beeld kan worden gebracht en kan worden gemonitord. Het doel van dit onderzoek is om een wetenschappelijk onderbouwd monitoringssysteem te ontwikkelen waarmee de kruidenrijkdom van grasland kan worden herkend op basis van satelliet Remote Sensing (RS) data (spoor 1), al dan niet in combinatie met foto's uit het veld (spoor 2) (Figuur 1).

In dit project is gewerkt met een maat voor kruidenrijkdom op een continue schaal, namelijk de Q-index. Deze index kan worden berekend op basis van de soortensamenstelling van de vegetatie en de abundantie waarmee de soorten voorkomen. Naast het feit dat deze index continue is, is deze ook minder gevoelig voor interpretatiefouten dan de methodiek van Schippers *et al.* (2015).

Voor het ontwikkelen en trainen van de algoritmes (zowel spoor 1 en 2) is een uitgebreide dataset opgebouwd, bestaande uit 1) vegetatieopnames (vertaald naar de Q-index), 2) satelliet RS-data en 3) foto's (overzicht + detailfoto's) van grasland percelen. De RS-data zijn 'vertaald' naar een set van RS-markers, die potentieel interessant zijn voor het kunnen voorspellen van kruidenrijkdom. De selectie van deze markers is gebaseerd op expert kennis (ecologie, RS) i.c.m. literatuuronderzoek. De RS-markers bevatten informatie die gerelateerd zijn aan o.a. management (intensiteit landgebruik), visuele kenmerken (kleur, heterogeniteit) of productiviteit van gras, en zijn verzameld voor alle grasland percelen in Nederland voor de jaren 2018 t/m 2022.

De ontwikkelde RS-algoritmes kunnen de kruidenrijkdom (Q) van grasland voorspellen. Echter, de voorspelfout (RMSE) van alle modellen ligt rond de 0,7Q. Dat komt overeen met een 95%-betrouwbaarheidsinterval van $\pm 1,4Q$. Mogelijke toepassingen zullen daarom eerder gezocht moeten worden in het herkennen van grote verschillen in kruidenrijkdom tussen graslanden. Op landelijke schaal zien de ruimtelijke patronen alsmede de relatieve verschillen er logisch uit, d.w.z. de modellen voorspellen een hoge kruidenrijkdom in Natura2000, weidevogel rustgebieden en andere natuurterreinen versus de 'reguliere' graslanden. Zowel het RR- en RF-model hebben de neiging om de kruidenrijkdom van productief grasland te overschatten, aangezien het zwaartepunt van de voorspelde kruidenrijkdom ligt tussen Q-index 2 en 3. Dit is niet het geval bij het SOM- en CNN-model, waar het zwaartepunt ligt tussen Q-index 1 en 2. Dat lijkt beter overeen te komen met de realiteit, aangezien de meerderheid van de graslanden in Nederland voornamelijk productieve graslanden zijn. Het SOM model heeft echter wel de neiging om de groep met Q-index 4-5, de groep zeer kruidenrijke graslanden, te overschatten in tegenstelling tot het CNN model (en RR- en RF-). Kijken we tot slot naar de jaarlijkse variabiliteit van de voorspelde kruidenrijkdom in de periode 2018 tot en met 2022), dan is met name het CNN-model behoorlijk stabiel, in tegenstelling tot de andere modellen.

Er zijn twee verschillende foto-algoritmes ontwikkeld. Het eerste algoritme (spoor 2a) is ontwikkeld om de kruidenrijkdom te bepalen. Het tweede algoritme (spoor 2b) om productief kruidenrijk grasland (PKG) percelen te kunnen onderscheiden van regulier Engels raaigrasland. Het foto-algoritme voor het bepalen van de kruidenrijkdom (Q-index) leverde de beste resultaten op met behulp van een classificatie model. De nauwkeurigheid van dit model is 52%, wat betekent dat het model in 52% van gevallen de mate van kruidenrijkdom correct heeft bepaald (dit correspondeert met een RMSE van $\sim 0,89$). Dit is een vergelijkbaar resultaat als van het RS-model en ook hiervoor staan aanbevelingen in dit rapport. Het foto-algoritme voor het onderscheiden van productief kruidenrijk grasland (PKG) ten opzichte van regulier Engels raaigras laat goede resultaten zien. Meer dan 95% van de foto's is goed geclassificeerd. Tot slot is er gekeken naar de voorspellingen van het RS-algoritme voor PKG-percelen. Bij aanvang van dit onderzoek was de verwachting dat het RS-algoritme de kruidenrijkdom van PKG-percelen zou onderschatten, gezien de hoge intensiteit in landgebruik. De Q-voorspelling laten inderdaad een behoorlijke range in Q-

waarden zien, variërend tussen 1,5 tot 4,5Q. Het foto-algoritme (spoor 2b) is daarom een nuttige aanvulling op het RS-model om PKG percelen te kunnen herkennen.

Een stakeholder analyse van de RS- en foto-algoritmes zijn tijdens een workshop geëvalueerd samen met de consortium partners en sponsors. Over het algemeen is de groep positief gestemd over de nut van deze modellen waar het gaat om stimuleren van de ontwikkeling van kruidenrijkdom en het inzichtelijk maken van de mate van kruidenrijkdom van grasland percelen. Echter, het belang van veldonderzoek wordt benadrukt, waarbij de kaarten in de huidige vorm ongeschikt lijken om te gebruiken voor directe beloning/afrekening. De datapositie van de eigenaar van het perceel en het (kunnen) integreren van de in dit project ontwikkelde RS- en foto algoritme modellen met bestaande systemen zijn ook benoemd als belangrijke aspecten voor de stakeholders. Het is zaak om rekening te houden met deze ethisch, juridisch en sociale aspecten bij de verdere ontwikkeling van de applicatie voor gebruik van de modellen.

1 Inleiding

Er is veel aandacht voor de grote biodiversiteitswaarde van kruidenrijk grasland. Binnen talloze initiatieven en beleidsinstrumenten wordt beoogd het areaal kruidenrijk grasland in het Nederlands agrarisch gebied te vergroten. Denk hierbij aan de biodiversiteitsmonitor melkveehouderij (KPI kruidenrijk grasland), het agrarisch natuurbeheer en initiatieven m.b.t. duurzame zuivel. De behoefte aan monitoring van kruidenrijk grasland neemt logischerwijs toe als gevolg van deze initiatieven, mede omwille van de noodzaak tot borging.

Het ontbreekt momenteel aan een geschikt monitoringsinstrument waarmee de kruidenrijkdom van graslandpercelen in Nederland in beeld kan worden gebracht en kan worden gemonitord. Monitoring en borging van kruidenrijk grasland middels veldwerk is een tijdrovende onderneming. **Het doel van dit (PPS) project is om een wetenschappelijk onderbouwd monitoringssysteem te ontwikkelen waarmee kruidenrijk grasland automatisch kan worden herkend op basis van satellietwaarnemingen (spoor 1), al dan niet in combinatie met foto's uit het veld (spoor 2a+b).** De toepassingsmogelijkheden van een functionerend algoritme zijn divers. Denk aan borging van beloningsinstrumenten, monitoring van de ontwikkeling van kruidenrijk grasland op verschillende ruimtelijke schalen, optimalisatie van natuurbeheer en wetenschappelijk onderzoek.

Binnen het project is voortgebouwd op eerder onderzoek, bijvoorbeeld bij het selecteren van Remote Sensing (RS) beelden die een voorspellende waarde hebben m.b.t. de kruidenrijkdom van een graslandperceel. Eerder onderzoek heeft aangetoond dat RS een zeer krachtige tool kan zijn voor het voorspellen van de kruidenrijkdom (Ali, Cawkell, Dwyer, Barret & Green, 2016). Er dient echter te worden opgemerkt dat de meerderheid van de onderzoeken de opgave om 'kruidenrijkdom te herkennen op basis van RS-beelden' sterk vereenvoudigden, en wel op 2 manieren:

1. Door gebruik te maken van een dataset die betrekking heeft op een relatief kleine regio (Fauvel *et al.*, 2020; Muro *et al.*, 2022). Hierdoor wordt de opgave vereenvoudigd, omdat variatie in abiotische factoren zoals bodemtype en grondwaterstand sterk worden beperkt, terwijl dat deze factoren kunnen zorgen voor verschillende vormen van kruidenrijk grasland, alsmede van invloed kunnen zijn op de wijze waarop relaties tussen Remote Sensing variabelen en kruidenrijkdom tot uiting komen (ruis). In dit opzicht is de uitdaging binnen dit onderzoek complexer, omdat een poging wordt gedaan om voor alle graslanden in Nederland een uitspraak te doen.
2. Enkele studies vereenvoudigen de werkelijkheid als het aankomt op de trainingsdata. In werkelijkheid vormt 'kruidenrijkdom' een continuüm, welke traditioneel wordt uitgedrukt als het aantal plantensoorten per vast oppervlakte (bijvoorbeeld 25 m²). In de meeste gevallen wordt geen recht gedaan aan dit continuüm. Bijvoorbeeld Bekkema & Eleveld (2018) hebben onderzocht of 'monocultuur Engels raaigras' kan worden onderscheiden van extensief gebruikt kruidenrijk grasland. Dit zijn twee klassen die in termen van kruidenrijk grasland relatief ver uit elkaar liggen; in werkelijkheid bestaan er meerdere tussenvormen. De resultaten – die soms tot een hoog percentage nauwkeurigheid komen – geven daardoor een vertekend beeld van de werkelijkheid.

Om het doel van dit onderzoek te realiseren, is geprobeerd de volgende onderzoeksvragen te beantwoorden:

- Kan op basis van een RS-algoritme met RS-data en/of RS-markers (verder aangeduid als spoor 1) de kruidenrijkdom van grasland op jaarbasis worden geschat voor heel Nederland?
 - Hoe nauwkeurig zijn de voorspellingen van het RS-algoritme?
 - Hoe consistent zijn de voorspelling jaar-op-jaar?
- Kan op basis van een foto-algoritme met foto's van een perceel (verder aangeduid als spoor 2a) de kruidenrijkdom van het perceel nauwkeurig worden geschat?
 - Hoe nauwkeurig zijn de voorspellingen van het foto-algoritme?
- Zijn de voorspellingen van spoor 1 en 2a consistent in het voorspellen van de kruidenrijkdom?

Vanuit de gebruikersgroep was gevraagd om ook specifiek te kijken naar 'productief kruidenrijk grasland' (PKG). PKG is een alternatief voor de reguliere productiegraslanden. Deze graslanden worden per definitie

ingezaaid. De mengsels bestaan voornamelijk uit gras met witte klaver, rode klaver, smalle weegbree en cichorei. In dit project is een poging gedaan om deze graslanden te herkennen, met het ontwikkelen van een tweede foto-algoritme (spoor 2b).

- Kan het foto-algoritme onderscheid maken tussen regulier Engels raaigras en productief kruidenrijke graslanden (PKG)?
- Hoe gaat het RS-algoritme om met PKG-percelen?

Afbakening

Dit onderzoek is gericht op het voorspellen van de kruidenrijkdom voor graslandpercelen. Smalle grasstroken (akkerranden, bermen, e.d.) vallen buiten de scope van dit onderzoek. Dit omdat het detailniveau van de beschikbare RS-beelden niet toereikend is.

2 Methodiek

Dit hoofdstuk beschrijft in het kort de werkwijze van het project om het doel van deze PPS (fase 1) te realiseren, namelijk het "Ontwikkeling monitoringssysteem dat jaarlijks een uitspraak doet over kruidenrijkdom van graslandpercelen in Nederland". Vervolgens zal elke stap in de daaropvolgende secties nader worden beschreven.

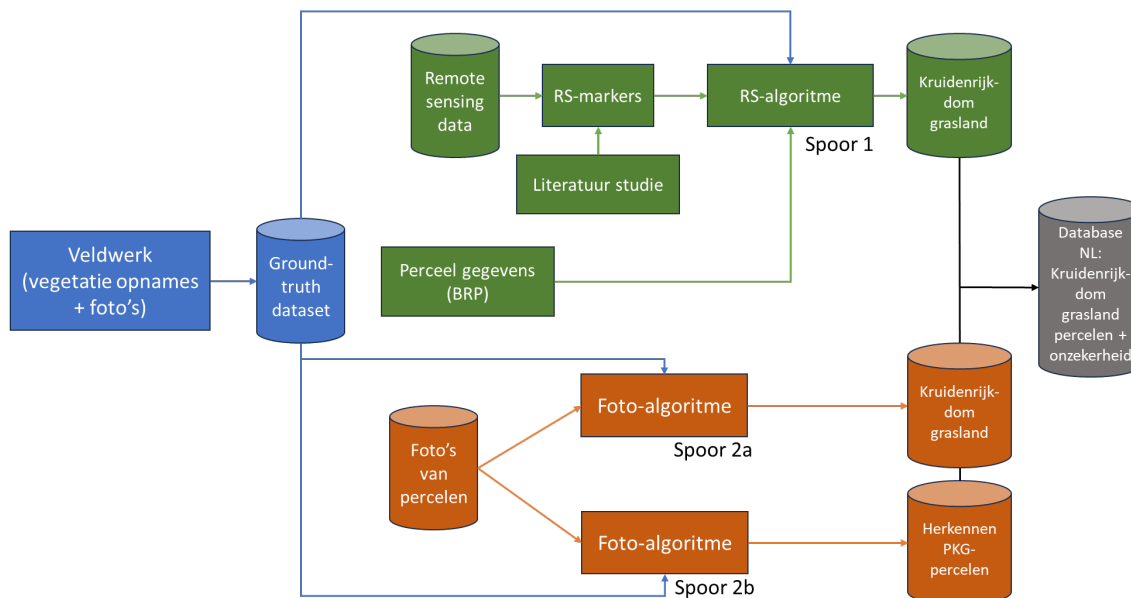
2.1 Overzicht project aanpak

Kort samengevat ziet de aanpak van het project (fase 1) er als volgt uit:

- Gezamenlijk vaststellen van een typologie van grasland op basis van kruidenrijkdom (zie sectie Typologie kruidenrijkdom).
- Creëren van een (centrale) database
 - Van grasland percelen waarvan zowel foto's als vegetatieopnames zijn verzameld in de afgelopen jaren (2016-2022). Verzameld uit bestaande nationale archieven zoals de Landelijke Vegetatie Databank (LVD), andere bronnen, en op basis van veldbezoeken en fotocampagnes georganiseerd vanuit dit project. Deze zogenaamde 'ground truth' data dienen als referentie voor de ontwikkeling van algoritmes voor de herkenning van kruidenrijk grasland (zie Inwinnen vegetatieopnames en Inwinnen foto's).
 - Satelliet Remote Sensing (RS) beelden en afgeleide RS-data van de afgelopen jaren (2017-2022) met voldoende ruimtelijke resolutie om op perceel niveau alle grasland in Nederland te kunnen monitoren (zie Inwinnen RS-data), zowel de percelen met vegetatie opnames, als de rest van de percelen in Nederland.
- Het ontwikkelen van twee soorten algoritmes waarmee de kruidenrijkdom kan worden geschat op basis van twee sporen:
 1. RS- en geodata (verder aangeduid als spoor 1)
 - Het ontwikkelen van RS-indicatoren of RS-markers. Er zal worden onderzocht welke satellietdata en daaruit afgeleide indicatoren (bijvoorbeeld maaifrequentie) gebruikt kunnen worden voor het schatten van de kruidenrijkdom. Op basis van wetenschappelijke literatuur en ervaring uit eerdere projecten zullen markers worden voorgesteld. Zo zal er worden onderzocht welke specifieke kenmerken van de (kruidenrijke)graslanden kunnen worden waargenomen met de satellietdata, zoals biomassa, kleur, heterogeniteit, gebruiksintensiteit (bv. maaien, begrazen), vegetatie-structuur, etc. (zie Selectie RS-markers).
 - Het ontwikkelen van een model/algoritme dat de kruidenrijkdom kan schatten op basis van de verzamelde RS-data en RS-markers voor alle grasland percelen in Nederland, en dat op jaarbasis een update kan geven van de status van de kruidenrijkdom op perceel niveau (om te beginnen op BRP-perceel niveau) (zie Ontwikkeling RS-algoritme).
 2. Foto's (detail en overzichtsfoto's) van percelen verzameld via een app.
 - Het ontwikkelen van een foto-algoritme dat op basis van een of meerdere foto's de kruidenrijkdom kan voorspellen (zie Ontwikkeling foto algoritme), verder aangeduid als spoor 2a.
 - Het ontwikkelen van een tweede foto-algoritme dat onderscheid kan maken tussen regulier Engels raaigras en productief kruidenrijke graslanden (PKG), verder aangeduid als spoor 2b.
- Op verzoek van gebruikersgroep zal ook worden onderzocht of 'productief kruidenrijk graslanden (PKG)' kunnen worden herkend, op basis van van het RS-algoritme of het foto-algoritme. PKG-percelen zijn een aparte groep kruidenrijke graslanden, aangezien deze zijn ingezaaid. De aanwezige kruiden zijn het resultaat van het zaadmengsel en niet van een meerjarig-specifiek-teeltplan.

- Beide algoritmes (RS en foto) zullen worden getoetst op basis van een deel van de 'ground truth' dataset. Ook zal worden onderzocht hoe beide algoritmes elkaar kunnen aanvullen of versterken (zie Remote Sensing - versus foto algoritme).

In Figuur 1 wordt een schets weergegeven van de werkwijze op hoofdlijnen.



Figuur 1 Schets van werkwijze op hoofdlijnen.

2.1.1 Reden tweesporen aanpak

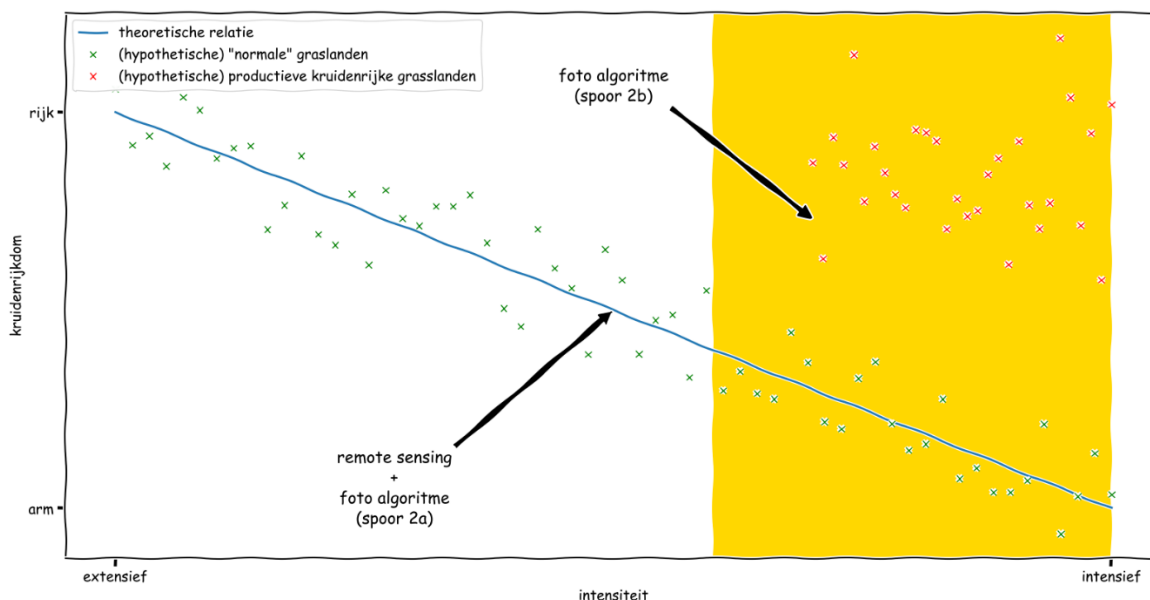
Zoals in de aanpak in 2.1 is beschreven zal in dit project twee sporen worden gevolgd, die op verschillende schalen werken. De eerste is op grote ruimtelijke en tijdsschaal, waarbij percelen gemonitord worden doormiddel van RS-data. Van een verscheidenheid aan percelen zijn de bijbehorende satellietbeelden verzameld, waaruit allerlei RS-markers zullen worden gedestilleerd, welke potentieel interessant zijn voor het kunnen voorspellen van kruidenrijkdom. Hierover meer verderop in het rapport.

De verwachting is dat de RS-data informatie bevat informatie die gerelateerd kan worden aan zowel het management als specifieke eigenschappen/karakteristieken van gras (intensiteit landgebruik, visuele kenmerken en productiviteit). De verwachting is dat gras dat intensief wordt gemanaged, grasgroen van kleur is en een hoge productie heeft, een lage kruidenrijkdom zullen hebben. En gras dat extensief wordt gemanaged, een afwijkende groentint heeft en een lage productie juist een hoge kruidenrijkdom heeft. Een schets hiervan is te zien in Figuur 2.

De laatste tijd is er veel aandacht voor zogenaamde productief kruidenrijke graslanden, ook wel bekend als PKG. Dit zijn ingezaaide percelen met zaadmengsels bestaande uit grassen en kruiden. Omdat deze percelen redelijk intensief worden gemanaged en een behoorlijke productie hebben, is de verwachting dat het RS-algoritme de kruidenrijkdom van deze percelen zal onderschatten (zie Figuur 2, gele vak).

Hypothese: Het RS-algoritme zal de kruidenrijkdom van ingezaaide productieve kruidenrijke grasland (PKG) onderschatten.

Daarom zal er een tweede spoor worden gevolgd (verder aangeduid als spoor 2b), waarbij onderzocht zal worden of het mogelijk is om met behulp van foto's onderscheid kan worden gemaakt tussen regulier Engels raaigras (groen kruisjes in gele vak Figuur 2) en productief kruidenrijk grasland (rode kruisjes in gele vak). Beide typen grasland worden behoorlijk intensief gebruikt.



Figuur 2 Schets van de rationale achter het gebruik van zowel Remote Sensing data als in het veld genomen foto's.

2.2 Typologie kruidenrijkdom

Het doel van dit project is om de kruidenrijkdom (het aantal verschillende soorten kruiden en grassen in een graslandperceel) te voorspellen. In de opstartfase van dit project is verkend of daarbij gebruik kan worden gemaakt van een klasseindeling. Classificaties zijn immers eenvoudig te hanteren en begrijpen. Verschillende opties voor classificaties zijn verkend. Een veel gebruikte klasseindeling betreft de typologie van Schippers *et al.* (2015). Binnen deze methodiek worden 6 verschillende fases onderscheiden, van intensief gebruikt, kruidenarm grasland (fase 0) tot extensief gebruikt kruidenrijk grasland (fase 4) en schraalland (fase 5). Deze methodiek is uitermate geschikt voor de beheerpraktijk, omdat binnen de systematiek beheeradviezen zijn gekoppeld aan iedere fase. Na vaststelling van de fase kan uit de gids worden afgeleid op welke wijze het grasland kan worden beheerd en/of doorontwikkeld.

De methodiek leent zich echter niet voor dit onderzoek, vanwege de volgende redenen:

1. In werkelijkheid betreft de kruidenrijkdom van agrarische graslanden een continuüm. De methodiek van Schippers slaagt erin om overkoepelende kenmerken te identificeren van bepaalde bandbreedtes langs de gradiënt van kruidenarm naar kruidenrijk. Dit vormen vervolgens de 'fases'. Bij gebruik van de systematiek in de praktijk blijkt echter dat een aanzienlijk deel van de graslanden niet goed kan worden ingedeeld, omdat het een ongedefinieerde tussenfase betreft. Dit is logisch, omdat in werkelijkheid talloze tussenvormen bestaan; kruidenrijkdom betreft immers een continuüm.
2. De methodiek van Schippers berust voor een belangrijk deel op beschrijvende kenmerken en richtgetallen. Denk hierbij aan het aanwezig zijn van kruiden in haarden (groepen) of de aanwezigheid van een homogene, ijle vegetatiestructuur. Uitsluitende, 'harde' criteria ontbreken in de systematiek. Het is dan ook niet ongebruikelijk dat gebruikers tot verschillende einduitslagen kunnen komen op hetzelfde perceel.

De combinatie van bovenstaande punten maakt de systematiek ongeschikt voor dit onderzoek. Voor het trainen van een statistisch model op basis van geclassificeerde data is absolute zekerheid over de juistheid van de toegekende klasse benodigd. Een bijkomend nadeel van het gebruik van de systematiek van Schippers betreft het feit dat – door het gebrek aan uitsluitende, harde criteria – de classificatie van Schippers niet achteraf kan worden 'berekend'. Daardoor zou – bij gebruik van deze methodiek – een grote hoeveelheid vegetatieopnames, die is opgeslagen in databases, onbruikbaar zijn.

Om alle bovenstaande redenen wordt in dit project gewerkt met maten voor de kruidenrijkdom op continue schaal en niet met classificaties. Hoofddoel is daarmee om continue waarden voor kruidenrijkdom zo goed mogelijk te voorspellen. Wij hanteren daarbij de volgende indicatoren voor kruidenrijkdom:

- **Rijkdom:** Het totale aantal verschillende soorten grassen en kruiden dat is aangetroffen in de vegetatieopname.

Toelichting: Dit betreft de meest eenvoudige maat voor kruidenrijkdom. Het probleem met deze indicator ligt voor de hand: het aantal aangetroffen soorten is mede afhankelijk van de grootte van het telvlak. Des te groter het telvlak, des te meer soorten worden aangetroffen. Uiteraard betreft dit geen lineaire relatie. Als men start met een proefvlak van 1 m² meter zal het toevoegen van nog eens 1 m² waarschijnlijk een flink aantal nieuwe soorten aan de uiteindelijke lijst toevoegen. Als men start met een proefvlak van 100 m² zal het toevoegen van 1 m² waarschijnlijk weinig nieuwe soorten opleveren. Er is dus sprake van een verzadigingscurve. In theorie kan deze curve worden berekend, waardoor vegetatieopnames met verschillende groottes aan proefvlakken met elkaar kunnen worden vergeleken. Dit gaat echter gepaard met een grote mate van onzekerheid. Wij kiezen ervoor om bij het trainen van het statistisch model op het voorspellen van deze indicator alleen gebruik te maken van vegetatieopnames met een proefvlakgrootte van 5 bij 5 meter; de meest voorkomende oppervlaktegrootte in onze dataset.

- **Shannon diversity H':** Een maat voor biodiversiteit waarbij niet alleen wordt gekeken naar het aantal soorten dat voorkomt maar ook naar de onderlinge verhoudingen waarmee de soorten voorkomen.

Toelichting: Hoe hoger de score, hoe te soortenrijker en gevarieerder het grasland is. Een lage score wijst op een soortenarme vegetatie die wordt gedomineerd door één of enkele soorten. Dit betreft een veel gebruikte maat voor het vergelijken van de biodiversiteit.

- **Q-index:** Indicatiewaarde op basis van de mate (fase) van ontwikkeling van soortenarm grasland naar soortenrijk grasland.

Toelichting: De index is ontwikkeld door van 't Veer (2020) en bouwt voort op de graslandfasen die worden onderscheiden door Bax & Schipper (1998), Schaminee & Jansen (1998) en Schaminee *et al.* (2000). Soortenarme graslanden (1-2), matig soortenrijke graslanden (3) en 4-6 (zeer kruidenrijke graslanden).

De methodiek van Van 't Veer bouwt voort op de veldgids van Schippers *et al.* (2015). Het voordeel van de Q-index ten opzichte van de methodiek van Schippers is tweeledig:

1. De Q-index kan worden berekend op basis van de soortensamenstelling van de vegetatie en de abundantie waarmee de soorten voorkomen. De methodiek is daardoor minder gevoelig voor interpretatiefouten dan de methodiek van Schippers *et al.* (2015).
2. De Q-index is een continue schaal, de classificatie van Schippers niet.

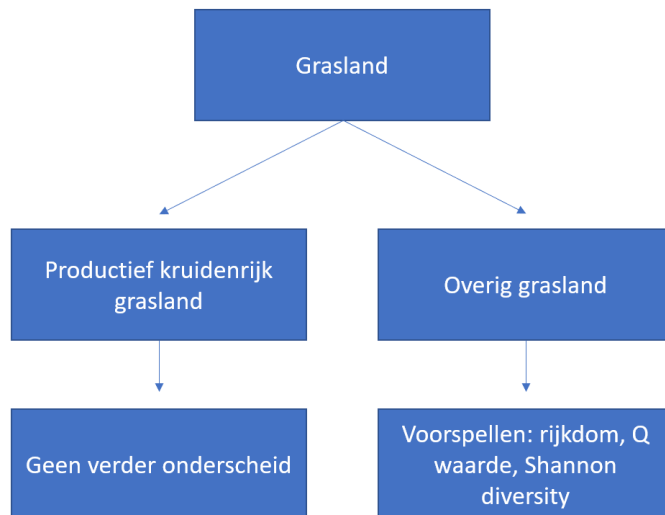
Uitzondering: productief kruidenrijk grasland (PKG)

Vanuit de gebruikersgroep is vroeg in het project gevraagd of een poging kan worden gedaan om ook 'productief kruidenrijk grasland' te herkennen. Dit zijn 'graslanden met een samengesteld soortenbestand, met de focus op functionele agrobiodiversiteit, met aanpalend een positief effect op de biodiversiteit en het landschap in algemene zin' (van Eekeren & Visser, 2019). Deze graslanden worden per definitie ingezaaid.

In dit project zal een poging worden gedaan om deze graslanden te herkennen. Wij gaan echter *geen* poging doen om bovengenoemde indicatoren voor kruidenrijkdom (rijkdom, Q-index en Shannon diversity) voor dit type grasland te voorspellen. Dit vanwege het volgende:

- Voor vrijwel alle graslanden in Nederland geldt dat een soortenrijke vegetatie het gevolg is van gunstige groeiomstandigheden voor de aanwezige kruiden. De kruiden zijn aanwezig omdat de groeiomstandigheden daarvoor geschikt zijn. Deze omstandigheden (zoals een lage productiviteit of hoge grondwaterstand) vormen vervolgens aanknopingspunten voor ons om te gebruiken in het trainen van een algoritme. De mogelijkheid om een grasland met 10 soorten te onderscheiden van een grasland met 30 soorten vloeit voort uit het feit dat beide graslanden naar alle waarschijnlijkheid zullen verschillen in groeiomstandigheden. Deze groeiomstandigheden kunnen wij vervolgens met Remote Sensing in beeld brengen.
- Voor productief kruidenrijke graslanden (PKG) geldt dat de soorten er staan omdat ze zijn ingezaaid. Er zijn geen groeiomstandigheden die een aanknopingspunt kunnen vormen voor het trainen van een model. Het voorspellen van bijvoorbeeld de rijkdom aan soorten is om die reden een onmogelijke opgave. Ter illustratie: In productief kruidenrijk grasland-1 staan 15 soorten. In productief kruidenrijk grasland-2 staan 25 soorten. Het verschil in het aantal soorten is het gevolg van het gebruik van een ander zaadmengsel en niet het gevolg van een verschil in groeiomstandigheden.

E.e.a. resulteert in de aanpak zoals in Figuur 3 weergegeven.



Figuur 3 Overzicht aanpak herkenning kruidenrijkdom.

2.3 Inwinnen vegetatieopnames

2.3.1 Vegetatieopnames op plot-niveau

Uitvoeren vegetatieopnames

In het kader van dit project zijn in 2022 in totaal 425 vegetatieopnames gemaakt. Toestemming voor het betreden van graslanden is geregeld via de agrarische collectieven en direct contact met boeren. De selectie van percelen heeft op dusdanige wijze plaatsgevonden dat een min of meer gelijke verdeling over de bodemtypen zand, klei en veen is bewerkstelligd.

Vegetatieopnames zijn gemaakt binnen één representatief telvlak van 5x5 meter per perceel. De abundantie van grassen en kruiden is genoteerd middels de Braun Blanquet methodiek (Westhoff & van der Maarel, 1978). Alle telvlakken waren tenminste op 5 meter afstand van de slootkant gepositioneerd.

Vegetatieopnames zijn uitgevoerd in de periode van april tot en met juni 2022.

Extraheren vegetatieopnames uit Landelijke Vegetatie Databank

De Landelijke Vegetatie Databank (LDV) wordt beheerd door Wageningen Environmental Research. Binnen de LDV wordt een groot deel van de in Nederland uitgevoerde vegetatieopnames bijeengebracht. Zodoende vormt de LDV een waardevolle bron voor dit project. Op grond van de volgende criteria zijn vegetatieopnames geselecteerd:

- Periode 2016 – 2022.
- Opname kan worden gelinkt aan BRP (meerdere redenen: praktisch voor Remote Sensing, nuttig om slootkantopnames uit te filteren).
- Alle opnames die binnen 5 meter van de slootkant zijn gemaakt zijn verwijderd. Dit omdat de plantenrijkdom in de slootkant veelal hoger is dan in de rest van het perceel.
- Vegetatieopnames waarin is genoteerd dat de bedekking van de struik- of boomlaag groter is dan 1% zijn verwijderd.
- Vegetatieopnames waarbij het telvlak kleiner was dan 4 meter zijn verwijderd.
- Vegetatieopnames waarbij de kruidlaag hoger was dan 70 cm zijn verwijderd. Dit betreft 'ruigtes'.

Extraheren vegetatieopnames uit overige bronnen

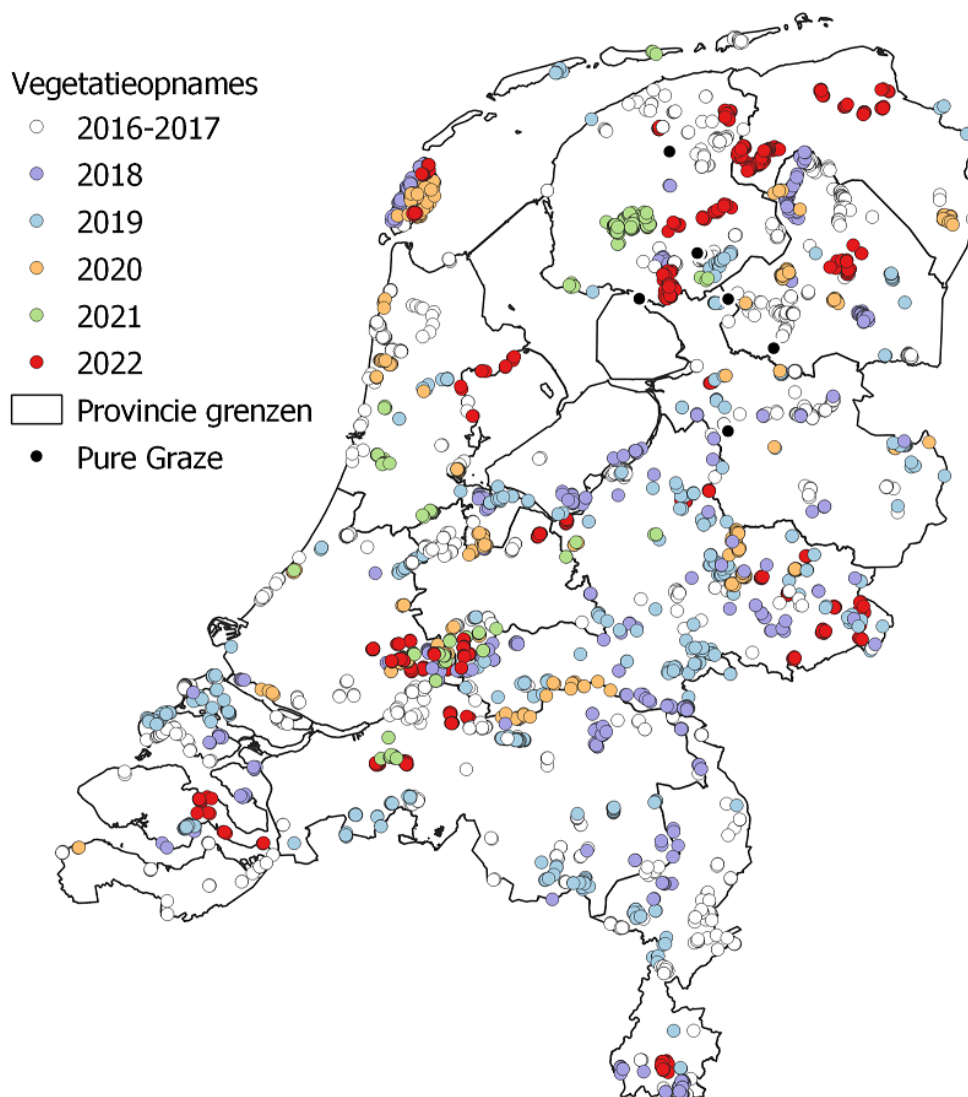
Overige bronnen vegetatieopnames:

- StaatsBosBeheer (SBB) - Uitbesteding SNL karteringen (500 opnames).
- NIL-project SBB.
- BO project habitatkwaliteit (2021).

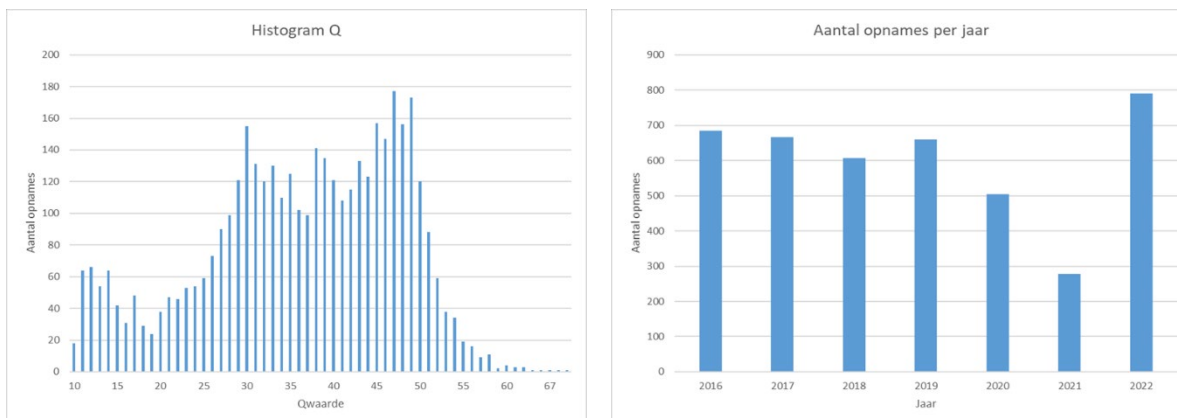
- Opnames Alblasserwaard.
- 300 à 400 via R. Ketelaar.

Van de totaal aantal verzamelde vegetatieopnames (11591) zijn er uiteindelijk 4189 overgebleven (volgens bovengenoemde selectiecriteria). Figuur 4 geeft een overzicht van de locaties van de verzamelde vegetatieopnames. Het histogram in Figuur 5 (links) laat de verdeling zien van de afgeleide Q-index voor elke vegetatie opname. Het aantal opnames per jaar is te zien in Figuur 5 (rechts). De range van Q-index ligt grofweg tussen 1 en 6, waarbij met name Q-waarden tussen 3 en 5 het meest voorkomen. Dit is logisch aangezien de vegetatieopnames uit de LVD vooral zijn gedaan in percelen met een hoge kruidenrijkdom. Daarom was besloten om bij de vegetatieopnames die binnen dit project in 2022 zijn verzameld te focussen op percelen met een lage Q-index (< 3).

NB: Figuur 4 laat ook wat locaties zien van Pure Graze percelen, dit betreft de locaties van foto's die zijn gemaakt van percelen met een ingezaaid productief kruidenmengsel (zgn. productief kruidenrijk grasland (PKG) percelen). Van deze percelen is de kruidenrijkdom niet bekend en zullen dus ook niet gebruikt worden voor de ontwikkeling en training van het RS-algoritme. Via de fotocampagne zijn in 2021 en 2022 foto's verzameld (met metadata) van percelen waarvan bekend is dat die zijn ingezaaid met kruidenmengsels. Deze foto's zullen worden gebruikt voor de ontwikkeling van het foto-algoritme.



Figuur 4 *Overzicht locaties vegetatieopnames.*



Figuur 5 Histogram Q-index van de geselecteerde vegetatieopnames (links) en het aantal vegetatieopnames per jaar (rechts).

2.3.2 Vegetatieopnames op BRP-perceel niveau

Bij de start van dit project is besloten dat de kruidenrijkdom van grasland percelen gemonitord zal worden op perceel niveau, waarbij gebruik gemaakt gaat worden maken van de jaarlijkse BRP-bestanden. De Basisregistratie Percelen (BRP) is een landelijk registratiesysteem van alle percelen in Nederland. Dat betekent dat inwinning van de Remote Sensing data zal worden vertaald naar perceel niveau, maar ook dat de vegetatieopnames moeten worden 'vertaald' naar perceel niveau. Dat laatste is als volgt gedaan:

- Koppelen van vegetatieopnames met BRP-percelen. Dit is gedaan op basis van de locatie gegevens (coördinaten) van de vegetatieopname met de BRP-perceel grenzen. Op deze manier is elke opname die binnen een BRP-perceel ligt meegenomen. Niet koppelbare opnames zijn niet meegenomen.
- In de meeste gevallen kon 1 vegetatieopname gekoppeld worden aan 1 BRP perceel. Er zijn ook BRP-percelen waarin meerdere vegetatieopnames liggen, vaak zijn dit wat grotere natuurterreinen. Voor deze gevallen zijn de indices van kruidenrijkdom gemiddeld (inclusief een standaarddeviatie).
- Belangrijke aanname hier is dat de gemeten kruidenrijkdom in het telvlak representatief is voor het hele BRP-perceel.

Tabel 1 laat zien hoeveel van de geselecteerde 4189 vegetatieopnames zijn gekoppeld aan een BRP-perceel. Het gemiddeld aantal percelen per jaar ligt rond de 2300. Het lagere aantal heeft te maken met het feit dat er percelen zijn waar meer dan 1 vegetatieopname in valt. Daarnaast zijn er ook wat gevallen dat er geen koppeling was tussen een opname en een perceel. De reden dat het aantal percelen iets varieert per jaar heeft te maken met de jaarlijkse veranderingen in het BRP-bestand. Soms komt het voor dat percelen niet zijn geregistreerd. Deze set van BRP-percelen zal uiteindelijk worden gebruikt om het RS-algoritme te ontwikkelen en te trainen.

Tabel 1 Aantal BRP-percelen per jaar met een index van kruidenrijkdom.

BRP	Aantal
2017	2415
2018	2364
2019	2321
2020	2273
2021	2261
2022	2250

Per perceel is uiteindelijk de volgende informatie verzameld:

- Perceel Identificatienummer (ID) en vegetatieopname ID. Het perceel-ID wordt jaarlijks opnieuw bepaald op basis van het nieuwe BRP-bestand.
 - We volgen hier de systematiek van de AgroDataCube, waarbij de RS-data worden verzameld op perceel-niveau (zie Verwerken RS-data).
- Indices kruidenrijkdom:

- Q-index, Shannon diversity, Richness, opname datum, grootte telvlak. In het geval deze zijn gebaseerd op meer dan 1 vegetatie opname, dan zijn de gemiddelde index + standaarddeviatie, aantal opnames, gemiddelde telvlak afgeleid.
- In het geval er 1 vegetatieopname ligt in een BRP-perceel: de afgeleide Q-index (en andere indices) wordt gekoppeld aan dat perceel. **NB:** dit wordt herhaald voor elk jaar op basis van de BRP van dat jaar.
- In het geval dat er meer dan 1 vegetatieopname ligt in een BRP-perceel: Het middelen van de vegetatieopnames wordt gedaan per jaar, d.w.z. vegetatieopnames van hetzelfde jaar worden gemiddeld. Het komt namelijk in een aantal gevallen voor dat percelen meer dan 1x zijn bezocht in een periode van meerdere jaren. Zo is het later mogelijk om vegetatieopname/kruidenrijkdom van jaar x te koppelen aan de Remote Sensing data van jaar x.
- Grondgebruik (bouwland/grasland), gewas naam en code (soort grasland: tijdelijk/permanent/natuurlijk grasland, etc.), omtrek en oppervlak volgens BRP-opgave.
 - Hierdoor is het mogelijk om in een later stadium, bijvoorbeeld bij de training van het RS-algoritme extra te kunnen filteren.

2.4 Inwinnen foto's

Van ieder perceel zijn minstens twee overzicht- en twee detailfoto's genomen. Een overzichtsfoto is een foto waarbij de bovenrand van de foto net boven de horizon uitkomt. Een detailfoto is een foto die schuin naar beneden is gemaakt. In Figuur 6 een voorbeeld te zien van beide type foto's.



Figuur 6 Voorbeeld van een overzichtsfoto (links) en een detailfoto (rechts) (Bron: Dataset natuurlijke graslanden).

De foto's zijn ingewonnen door de verschillende partners van het project en ook vrijwilligers konden van 2021 t/m 2023 foto's inwinnen via de Ellipsis Drive-app (zie hieronder). Omdat de partners en vrijwilligers in 2021 voornamelijk foto's hadden gemaakt van de mooie kruidenrijke graslanden, is in 2022 een opdracht uitgezet om foto's en vegetatieopnames te maken van specifiek graslanden in een lagere Schippersfase. Deze opdracht is uitgevoerd door ATKB. In 2022 is naast de Ellipsis Drive-app ook een alternatieve app gemaakt en die in 2023 ook onder de vrijwilligers bekend gemaakt (zie hieronder).

2.4.1 Inwinning foto's via de Ellipsis Drive app

Voorafgaand aan het project is er een app ontwikkeld waarmee in het veld foto's genomen kunnen worden. Deze foto's, inclusief de metadata, worden dan direct geüpload naar de drive van Ellipsis. De metadata bestond uit de tijd, de locatie, de persoon en organisatie die de foto genomen had en een antwoord op een reeks vragen. Deze app is zowel in 2021 als in 2022 gebruikt. Nadeel van de app is dat de telefoon een goede internetverbinding moet hebben op de locatie. Dat bleek niet altijd het geval te zijn en heeft tot enige irritatie geleid, waardoor een gedeelte van de mensen die foto's gemaakt hebben, zijn afgehaakt.

2.4.2 Inwinning foto's via de ODK app

In 2022 is er een alternatieve app ontwikkeld in samenwerking met de IT-afdeling van WUR, verder aangeduid als ODK app. Dit is een opensource app, die de verzamelde foto's en metadata uploadt zodra er verbinding is met het internet. Dat hoeft dus niet op locatie te zijn. In 2022 en 2023 hebben mensen van WUR met de ODK-app van 108 percelen foto's verzameld, waarvan er in de afgelopen twee jaar een vegetatie opname gemaakt was. In 2023 is de app onder de andere gebruikers verspreid. In 2022 waren veel van de detailfoto's toch recht naar beneden genomen en op veel van de overzichtsfoto's was toch de horizon in het midden gekomen. Daarom zijn in de ODK-app per locatie een overzichtsfoto's, twee foto's recht naar beneden en twee detailfoto's zoals boven beschreven opgenomen. Daarmee kon zeker worden gesteld dat er goede detailfoto's genomen werden.

2.4.3 Inwinning metadata via de app

Met beide apps werden, naast de foto's en de locatie, ook overige data verzameld. Aan de hand van deze vragen kon vastgesteld worden of bekend was dat het grasland was ingezaaid met productieve kruiden, klaver, inheemse kruiden en of het begraasd werd op het moment. Zie Figuur 7 voor de precieze formulering van de vragen.

***Is dit grasland in de afgelopen 5 jaar ingezaaid of doorgezaaid?**

- ☐ Weet ik niet.
- ☐ Neen, niet ingezaaid of doorgezaaid in de laatste 5 jaar.
- ☐ Ja, met productief Engels raai gras en minder dan 5% klavers.
- ☐ Ja, met productieve voederkruiden, zoals cichorel, karwij, smalle weegbree en klavers.
- ☐ Ja, met natuurlijke inheemse kruiden (zaad van Biodivers of Cruydt-Hoeck of aangevoerd maaisel van extensief kruidenrijk grasland)
- ☐ Ja, met productief Engels raai gras en >20% klavers

***Wordt het perceel begraasd?**

- ☐ Weet ik niet
- ☐ Nee
- ☐ Er loopt nu graasvee
- ☐ Er zijn tekenen van begrazing (mestflatten, urineplekken, bossen)
- ☐ Het is mij bekend dat het perceel soms begraasd wordt

***Hoe productief schat u dit grasland in?**

- ☐ Weet ik niet
- ☐ Laag: < 4000 kg ds/ha/jaar
- ☐ Middel: 4000 - 7000 kg ds/ha /jaar
- ☐ Hoog: > 7000 kg ds/ha/jaar

Figuur 7 Een overzicht van de overige vragen die gesteld zijn tijdens de dataverzameling.

2.4.4 Inwinning foto's van Pure Graze en andere projecten

Om specifiek foto's te hebben van Productief Kruidenrijke Graslanden (PKG) zijn foto's verkregen van consortium lid Pure Graze, die zadenmengsels voor dergelijke graslanden verzorgt. Hiervoor zijn door Pure Graze zowel foto's gemaakt voor dit project als foto's verzameld vanuit hun archieven. Ook op WUR proefboerderij de Marke zijn in het verleden experimenten met productief kruidenrijke graslanden gedaan. Foto's hiervan zijn ook toegevoegd aan de gegevensverzameling.

2.4.5 Selectie van de bruikbare foto's

Net als de vegetatieopnames bleken ook niet alle verzamelde foto's bruikbaar. Bij veel overzichtsfoto's was er toch ver boven de horizon gefotografeerd en was het gedeelte grasland erg klein. Veel detailfoto's waren recht naar beneden genomen. Ook waren er veel foto's van bijvoorbeeld pas gemaaide graslanden en zelfs akkers. Alle foto's zijn daarom handmatig gecheckt op geschiktheid.

2.5 Inwinnen RS-data

2.5.1 Satelliet data

Met behulp van satelliet Remote Sensing (RS) is het mogelijk om vanuit de ruimte het aardoppervlak te verkennen en te monitoren. Tegenwoordig zijn er meerdere satellieten operationeel die op gedetailleerd niveau en met behulp van verschillende technieken het aardoppervlak scannen. Deze satellieten komen nagenoeg wekelijks over, waardoor gedurende het hele jaar (afhankelijk van de bewolking) opnames kunnen worden gemaakt.

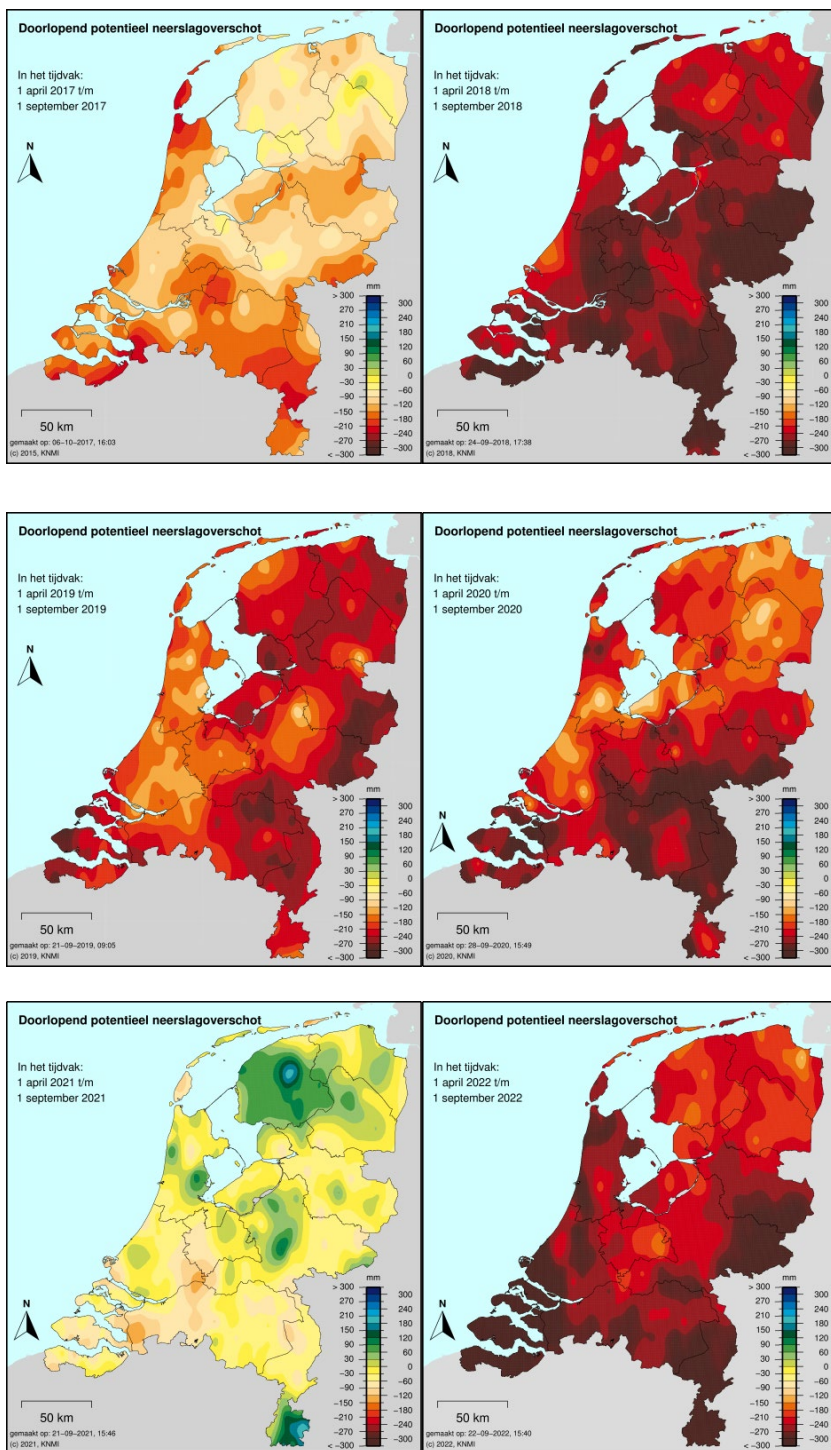
In dit project is gebruik gemaakt van Sentinel-1 en 2 satelliet data. Beide satellieten zijn onderdeel van het Europese Copernicus-programma. Daarnaast is ook gebruik gemaakt van Landsat-7 en 8 beelden (onderdeel van de NASA en USGS), als aanvulling op Sentinel-2. De data van deze satellieten is gratis en heeft een ruimtelijke resolutie van 10m tot 30m, waardoor het mogelijk is om de meeste grasland percelen in Nederland te kunnen monitoren. Tot slot wordt er binnen de ESA en NASA ook gekeken naar de continuïteit van de Sentinel- (1 & 2) en Landsat-missies, en staan er o.a. nieuwe Sentinels gepland.

Sinds 2018 zijn er voldoende Sentinel-2 beelden per jaar beschikbaar (het jaar waarin twee identieke Sentinel-2 satellieten operationeel zijn, namelijk 2A en 2B). Het Groenmonitor archief bevat wel beelden van daarvoor, echter de kwaliteit van deze beelden (spectrale resolutie, ruimtelijke resolutie) is onvoldoende. Daarom ligt de focus op de periode 2018-2022 en zullen alle bruikbare beelden (Sentinel en Landsat) worden verzameld.

Tabel 2 geeft een overzicht over de neerslag en de klimatologie voor de periode 2017 tot en met 2022 (bron: KNMI). In Figuur 8 wordt het doorlopend potentieel neerslagoverschot op 1 september voor de jaren 2017 tot en met 2022 weergegeven (bron: KNMI). Duidelijk is te zien dat deze periode bestaat uit jaren met vaak en zelfs extreme droogte. Eigenlijk is alleen 2021 een redelijk 'normaal' jaar.

Tabel 2 Neerslag totalen (landelijk gemiddeld) (Bron: KNMI).

Jaar	Neerslag [mm]	Klimatologie
2018	607	Extreem warm, extreem zonnig en zeer droog • Zachte lente • Zomer zeer warm en zeer droog
2019	783	Zeer warm, zeer zonnig en landelijk gemiddeld vrij droog • Zachte lente • Zomer afwisselend zeer warme en koele periodes, droog maar minder dan in 2018
2020	785	Extreem warm, zeer zonnig en aan de droge kant • Zachte lente met zeer veel zon • Zomer afwisselend zeer warme en koele periodes
2021	806	Gemiddeld normaal met recordaantal codes oranje en rood • Zeer koude lente • Zomer kende 2 gezichten: juni was warm gevolgd door een koele periode
2022	729	Extreem warm, recordzonnig en droog • De lente was aan de zachte kant, droog en zeer zonnig • Zeer warme, zonnige en droge zomer met het hoogtepunt in augustus
Gemiddeld	847	



Figuur 8 Doorlopend potentieel neerslagoverschot op 1 september voor de jaren 2017 t/m 2022 (bron: KNMI).

De verzamelde satelliet data kan onderverdeeld worden in 2 groepen: optische data en radar data. Hieronder volgt een kort overzicht van beide groepen. Gedetailleerde en aanvullende informatie is te vinden in Bijlage B.

2.5.1.1 Optische satelliet data

De Sentinel-2 en Landsat satellieten hebben een multi-spectrale sensor met meerdere spectrale banden. Deze banden liggen in verschillende delen van het elektromagnetisch (EM) spectrum, waaronder in het zichtbaar deel net als het menselijk oog (blauwe, groene en rode band). Daarnaast beschikken de satellieten nog over andere spectrale banden, zoals nabij-infrarode (NIR), Red-Edge en korte golf infrarode (SWIR) banden. Met behulp van deze spectrale banden is het mogelijk onderscheid te maken tussen verschillende landoppervlakken en vegetaties (zoals bv water, kale bodem, gras, bos, etc.) en kunnen verschillende

eigenschappen van de bodem en het gewas worden afgeleid. Zo zijn de rode (R) band en nabij-infrarode (NIR) band zijn zeer geschikt voor het monitoren van vegetatie. Planten absorberen grotendeels het rode zonlicht middels fotosynthese en reflecteren het merendeel van het nabij-infrarode licht. Dit principe wordt gebruikt om een groenindex te berekenen. De bekendste groenindex is de Normalised Difference Vegetation Index (NDVI), deze wordt aangeboden in de Groenmonitor. Daarnaast zijn er allerlei andere indices afgeleid, specifiek voor het monitoren van vegetatie (water, vocht in vegetatie etc.).

Optische sensoren kunnen niet door bewolking heen kijken. Om zoveel mogelijk wolkenvrije data te verzamelen wordt daarom gebruik gemaakt van zowel Sentinel-2 (2A en 2B) als Landsat (7 en 8). De opname dichtheid varieert per jaar, afhankelijk van de bewolking, en per locatie. Over het algemeen ligt de opname dichtheid tussen de 30 tot 50 opnames per jaar.

2.5.1.2 Radar satelliet data

Sentinel-1 is een zgn. actieve radar satelliet en zendt radar (RAadar Detection And Ranging) pulsen, met een bepaalde microgolf lengte (C-band, 5.405 GHz), naar het aardoppervlak en meet vervolgens de hoeveelheid radarstraling die door het aardoppervlak wordt teruggekaatst. De daaruit afgeleide microgolfbeelden (zgn. backscatter-beelden) geven informatie over de geometrische en diëlektrische eigenschappen van het oppervlak of volume, die vooral worden bepaald door oppervlakteruwheid én vochtigheidsgraad. Sentinel-1 emitteert verticaal (V) gepolariseerde pulsen en detecteert zowel verticaal (V) als horizontaal (H) gepolariseerde signalen die zijn teruggekaatst door het aardoppervlak, d.w.z. het meet backscatter-VV én VH. Zowel de bodem als vegetatie dragen bij aan het terugverstrooiingssignaal in zowel VV als VH. De interpretatie van radar beelden/signalen is moeilijker dan optische beelden. Kennis van het oppervlak is belangrijk om de radar backscatter-signalen te kunnen begrijpen, zoals hoogte (kort/lang) en dichtheid (open/gesloten) van de vegetatie, maar ook de vochtcondities van de grond (grondsoort) en de vegetatie (type), plas-dras omstandigheden (vegetatie deels/geheel onderwater) en tot slot eventuele managementactiviteiten (o.a. maaien, beweiden, bemesten, beregenen) (Zhang *et al.*, 2022). Naast de backscatter-beelden is het ook mogelijk om het zgn. coherentie-beelden af te leiden. Bij deze techniek (interferometrie) worden twee Sentinel-1 beelden (die vanuit hetzelfde punt in de ruimte zijn opgenomen, maar 12 dagen van elkaar verschillen) met elkaar vergeleken. Het afgeleide coherentie-beeld beschrijft per pixel de gelijkheid (coherentie of stabiliteit) van het gereflecteerde radarsignaal tussen de twee beelden, en zegt of het object (pixel of perceel) is veranderd of niet in die 12 dagen. Coherentieanalyse is bijzonder nuttig voor het detecteren van veranderingen. Zo is bijvoorbeeld het coherentie signaal zeer gevoelig voor het detecteren van ploegen van kale grond.

Voordeel van radar satellieten is dat deze door bewolking heen kunnen kijken, waardoor met een vast tijdsinterval (6 tot 12 dagen) beelden worden verzameld. De opname dichtheid van Sentinel-1A¹ ligt tussen de 30 en 60 beelden per jaar, en hangt alleen af van de locatie percelen in relatie tot de baan van de Sentinel-1 satelliet.

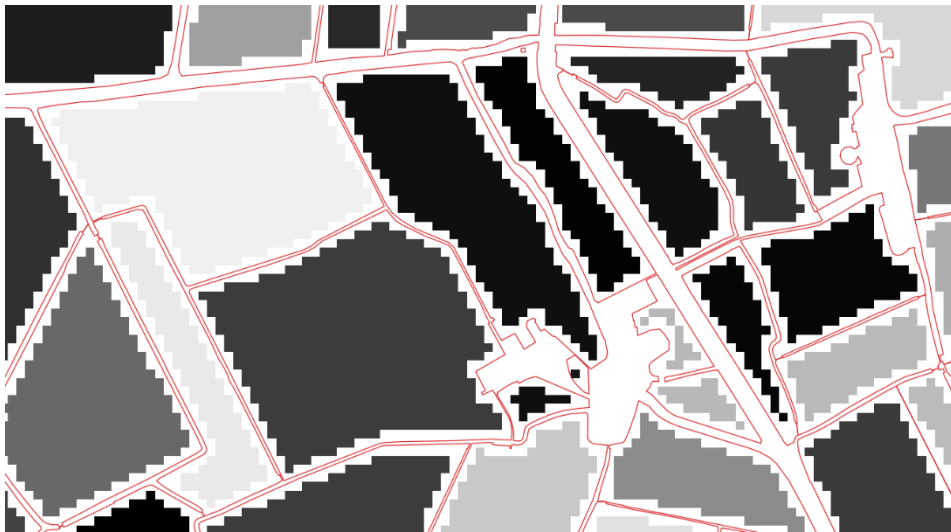
2.5.2 Verwerken RS-data

2.5.2.1 Van satelliet pixels naar BRP-perceel

Elk satellietbeeld (zowel optisch als radar) bestaat uit pixels (net als een digitale foto gemaakt met een smartphone). De grootte van de pixels varieert tussen de 10m en 25m, afhankelijk van de satellietsensor en de verwerking ervan tot bruikbare beelden. In deze studie is ervoor gekozen om niet op pixel-niveau te werken maar op perceel-niveau, d.w.z. op perceel niveau zal worden gekeken naar de kruidenrijkdom van grasland percelen.

Om van pixel- naar perceel-niveau te gaan wordt gebruik gemaakt van de Basisregistratie Percelen (BRP), een landelijk registratiesysteem van alle percelen in Nederland. Dit systeem bevat de locatie van landbouwpercelen, inclusief de grenzen, en daaraan gekoppeld het geteelde gewas. Door alle pixels die binnen de perceelgrenzen liggen (zgn. *pure pixels*, Figuur 9) statistisch te middelen kan voor elk perceel een gemiddelde pixel-waarde en bijbehorende standaarddeviatie worden afgeleid (RS-statistiek). De standaarddeviatie zegt iets over de variatie/heterogeniteit binnenin het perceel. De gemiddelde RS-waarde, de bijbehorende standaarddeviatie en het aantal pixels waarop de statistiek is gebaseerd, wordt voor elk RS-beeld bepaald. De grootte van het perceel (of aantal pure pixels) zal worden gebruikt bij de selectie van goede trainingsdata van het model.

¹ Sentinel-1B is eind 2021 door technische problemen uitgevallen.



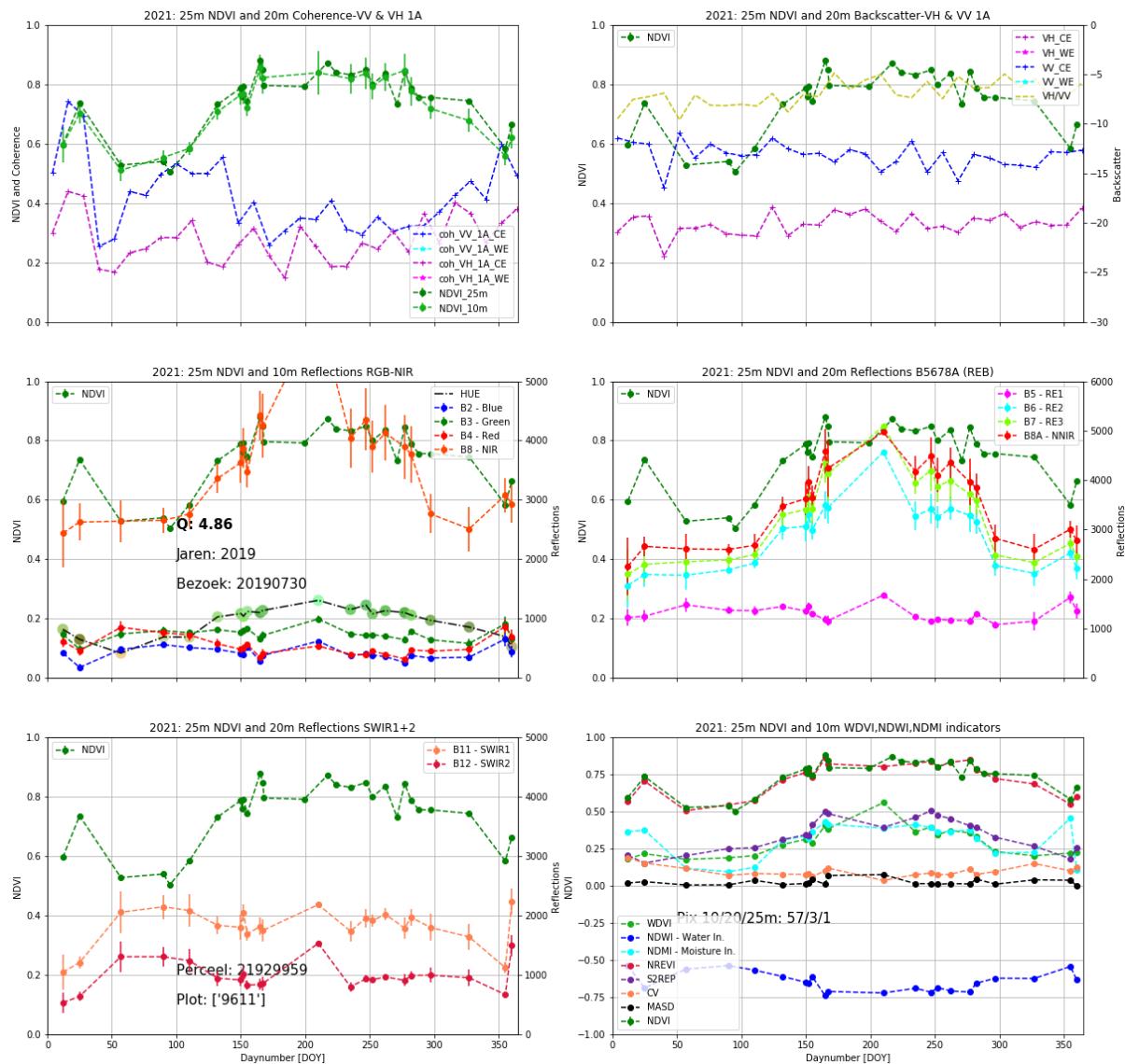
Figuur 9 Selectie van 'pure pixels', d.w.z. alleen de pixels die 100% binnen een BRP-perceel liggen.

2.5.2.2 RS-tijdseries

De RS-statistiek van elk BRP-perceel en voor alle opnames (beelden) die gedurende een heel jaar zijn verzameld en verwerkt, worden vervolgens per jaar samengevoegd (in een database zoals bv de [AgroDataCube](#)), resulterend in tijdseries van RS-data bestaande uit gemiddelde RS-waarden en bijbehorende standaarddeviaties voor dagen dat er een opname is.

Uiteindelijk zijn voor alle grasland percelen (geregistreerd in de BRP) in Nederland (~510.000 op jaarbasis) voor de jaren 2018 t/m 2022 deze RS-tijdseries verzameld. Deze verzameling aan tijdseries kunnen vervolgens worden gebruikt voor allerlei analyses op perceel-niveau. Figuur 49 geeft een voorbeeld van RS-tijdseries van zowel de optische als de radar data voor een enkel grasland perceel:

- De rode (R, band 4), groene (G, band 3), blauwe (B, band 2), 2x nabij-infrarood (NIR, band 8 en 8A), 3x Red-Edge (RE, band 5, 6 en 7) en 2x shortwave infrared (SWIR, band 11 en 12) reflecties van Sentinel-2.
 - De afgeleide de NDVI.
- De backscatter en coherentie van zowel VV als VH van Sentinel-1.
 - En de uit de backscatter afgeleide 'ratio' VH/VV (ook wel polarisatie ratio).



Figuur 10 Overzicht van tijdseries voor één perceel. Bovenstaande figuren tonen de gemiddelde waarden op perceel niveau voor Sentinel-2 (Red, Green, Blue, NIR, Red-Edge, SWIR) en afgeleide variabelen zoals NDVI. Daarnaast worden ook de gemiddelde waarden op perceel niveau voor Sentinel-1 getoond in de vorm van backscatter en coherentie (VH en VV). De punten betreffen de perceel-gemiddelde waarden en de fouten-balkjes de standaarddeviatie. De punten geven daarnaast ook de momenten aan waarop er bruikbare beelden beschikbaar waren.

2.5.2.3 Overzicht verzamelde RS-data

In Tabel 3 wordt een overzicht gegeven van de verzamelde RS-data (tijdseries). De RS-data (tijdseries) is door NEO omgezet in twee-wekelijkse tijdseries en zijn de eerste set van predictors die als input dienen voor de ontwikkeling van het RS-algoritme. Daarnaast is de RS-data gebruikt om zogenaamde RS-markers af te leiden die potentieel interessant zijn voor het voorspellen van kruidenrijkdom in grasland percelen.

Tabel 3 *Overzicht van verzamelde RS-data.*

	RS-data	Satelliet	Beschrijving
1	Reflecties v.d. 10 S2-banden (gemiddelde + standaarddeviatie)	Sentinel-2A/B	Tijdsreeks van reflecties: • Reflecties 4x 10m-banden: blauw (B2), groen (B3), rood (B4) en nabij infrarood NIR (B8) • Reflecties 6x 20m-banden: 1x nabij infrarood (NIR, B8A), 3x Red-Edge (RE, B5, B6, B7) en 2x shortwave infrarood (SWIR, B11, B12)
2	NDVI (gemiddelde + standaarddeviatie)	Sentinel-2A/B + Landsat-7/8	Tijdsreeks van NDVI (25m)
3	Backscatter en coherentie (gemiddelde + standaarddeviatie)	Sentinel-1A	Tijdsreeks van backscatter-sigitaal: • Backscatter-VH (20m) • Backscatter-VV (20m) Tijdsreeks van coherentie-sigitaal: • 12-daagse coherentie-VV (20m) • 12-daagse coherentie-VH (20m)

2.5.3 Selectie RS-markers

Voor alle percelen die onderdeel uitmaken van dit onderzoek zijn RS-gegevens ingewonnen; zowel van Sentinel 2 (optische satellietbeelden) als Sentinel 1 (radar satellietbeelden) (zie Verwerken RS-data). De ruwe Remote Sensing data - die bestaat uit talloze waarden voor verschillende spectrale banden - laat zich moeilijk interpreteren. Daartoe zijn zogenoemde 'RS-markers' ontwikkeld. Dit zijn indicatoren die worden afgeleid van de ruwe RS-data (tijdsreeks). Zo kan bijvoorbeeld worden berekend wat de verhouding is tussen bepaalde spectrale banden, waardoor men een beeld krijgt van de biomassa van het gewas, of kan worden gekeken naar de gemiddelde kleur. Er bestaan talloze mogelijkheden als het gaat om markers. Om in deze zoektocht structuur aan te brengen zijn de volgende categorieën gedefinieerd:

De verzamelde RS-data (tijdsreeks) vormt de basis dataset en is gebruikt om een set van RS-markers af te leiden, welke potentieel interessant zijn voor het kunnen voorspellen van kruidenrijkdom. De selectie van RS-markers is gebaseerd op expert kennis (ecologie, RS) i.c.m. literatuuronderzoek.

Kruidenrijke graslanden in Nederland onderscheiden zich van productieve graslanden op verschillende manieren. Ten eerste worden kruidenrijke graslanden anders gemanaged. Productief grasland wordt over het algemeen (zeer) intensief gemanaged, waarbij de focus ligt op maximale gras productie (biomassa). Deze percelen worden regelmatig gemaaid en/of beweid. De groeicondities voor het gras zijn daarvoor geoptimaliseerd, d.w.z. ideale vochtthuishouding (niet te nat of te droog) en optimale nutriënten condities (door toediening van kunst- en/of drijfmest). Over het algemeen zijn deze percelen homogeen grasgroen, waar het gras optimaal groeit en kruiden weinig tot geen kans krijgen, en die regelmatig worden gemaaid en/of beweid en bemest.

Kruidenrijke (met een natuurfunctie) graslanden daarentegen worden minder intensief gemanaged. De grascondities zijn minder optimaal. De percelen zijn natter of droger met minder nutriënten zodat ook andere grassen en kruiden een kans krijgen. Over het algemeen zijn deze percelen minder 'grasgroen' door de aanwezigheid van andere planten/kruiden en ligt de gras productie (biomassa) lager. Resultierend in minder frequent of zelfs laat maaien en/of minder intensief beweiding.

Tot slot zijn er ook een aantal co-variabelen geselecteerd die van invloed zijn op de uitingvorm van het type grasland, namelijk grondwaterstand en bodemsoort.

De RS-data (tijdsreeks) bevat informatie die gerelateerd kan worden aan zowel het management van grasland als de eigenschappen/karakteristieken van het gras:

- Intensiteit landgebruik
- Visuele kenmerken
- Productiviteit
- Co-variabelen

In de volgende secties zullen de geselecteerde RS-markers, waarmee een van de hierboven genoemde kenmerken beschreven kan worden, besproken. Gedetailleerde en aanvullende informatie van de RS-markers met voorbeelden zijn te vinden in Bijlage C.

2.5.3.1 Intensiteit landgebruik

Maaien, 1^{ste} maaidatum, scheuren

Management (of intensiteit landgebruik) van een grasland perceel, zoals maaien, scheuren of her-inzaaien, beweiden en andere maatregelen, zorgen over het algemeen voor fluctuaties in de RS-tijdseries. Zo is het bijvoorbeeld mogelijk om uit NDVI-tijdseries maaimomenten af te leiden, welke zichtbaar zijn als zgn.

'zaagtand' patroon (zoals is te zien in Bijlage C) in het NDVI-signaal (Kolecka *et al.*, 2018; Stumpf *et al.*, 2020). Deze 'zaagtand' is het gevolg van het plotseling verdwijnen van biomassa (als gevolg van maaien of zeer intensief beweiden) en vervolgens het weer (langzaam) aangroeien van gras.

Niet alleen in de NDVI-tijdserie maar ook in de radar-tijdseries (backscatter en/of coherentie) is het effect van maaien in meer of mindere mate zichtbaar. Meerdere papers zijn verschenen waarin algoritmes zijn ontwikkeld voor het detecteren van maaimomenten op basis van NDVI- of radar-tijdseries of een combinatie van beide (De Vroey *et al.*, 2021; Kolecka *et al.*, 2018; Reinermann *et al.*, 2022; Schwieder *et al.*, 2022; Stumpf *et al.*, 2020; Tamm *et al.*, 2016; Taravat *et al.*, 2019).

Hier is het NDVI-signaal gebruikt om maaimomenten te detecteren², zodat er per perceel een overzicht is van het aantal maaimomenten (per jaar) en de datum waarop het maaien plaats vond. Met name de 1^{ste} maaidatum is belangrijk. Percelen die al vroeg worden gemaaid in het voorjaar zijn over het algemeen productieve percelen. Daarnaast is gekeken naar het scheuren van percelen, waarbij de oude grasmat wordt vernietigd. Dit wordt (regelmatig) gedaan bij productieve graslanden wanneer de productie na verloop van tijd inzakt.

NB: Binnen deze studie is ook onderzocht of beweiding kan worden gedetecteerd. Daarvoor zijn een aantal grasland percelen geanalyseerd en vergeleken met maai- en beweidingsgegevens van het Koeien & Kansen project. Uit de analyse bleek dat beweiding in de meeste gevallen niet zichtbaar is in het NDVI-signaal omdat de afname van biomassa te klein is. De reden hiervoor is dat het aantal stuks vee in verhouding tot de grootte van het perceel te klein is, waardoor de afname van NDVI te klein is. In enkele gevallen, waarbij het aantal koeien per ha zeer groot was, was een afname in het NDVI-signaal zichtbaar, maar was qua patroon identiek aan een maaimoment, waardoor geen onderscheidt kan worden gemaakt tussen beweiden en maaien.

Mean Absolute Spectral Dynamics (MASD)

Een andere manier om te kijken naar de intensiteit van het landgebruik van een perceel is door te kijken naar de fluctuaties in de tijdseries van gemeten reflecties (van alle banden), zoals de MASD (Mean Absolute Spectral Dynamics). MASD geeft de spectrale veranderingen aan tussen twee opeenvolgende beelden, en kan worden gebruikt als indicator voor intensiteit landgebruik (Franke *et al.*, 2012; Bekkema en Eleveld, 2018). Een lage MASD, weinig fluctuaties, suggereert een laag intensiteit landgebruik (weinig maaien en/of beweiden). Een hoge MASD daarentegen is een indicatie voor een hoog intensiteit landgebruik.

De MASD is volledig gebaseerd op optische data, specifiek de reflecties (10 banden) gebaseerd op Sentinel-2. In deze studie is de MASD bepaald op jaarbasis alsmede voor het weidevogel broedseizoen (1 april – 15 juli). Om een goede MASD te kunnen bepalen zijn voldoende opnames nodig. Door zowel bewolking als de ligging van de baan van Sentinel-2 is de dekkingsgraad voor heel Nederland niet 100%. Daarom is als kwaliteitsindicator het aantal beelden toegevoegd waarop de MASD is gebaseerd.

Intensiteit landgebruik op basis van Radar

Howison *et al.* (2018) ontwikkelde een methodiek waarmee op basis van Sentinel-1 radar data vogelvriendelijke weiden kunnen worden gedetecteerd. Howison vond dat de gebruiksintensiteit (*land use intensity*) van percelen kon worden geschat met behulp van het Sentinel-1 radarsignaal. Daarbij wordt specifiek gebruik gemaakt van het backscatter VH-signaal. Door te kijken naar de veranderingen in het backscatter-signaal tussen twee opeenvolgende beelden (delta-VH) gedurende het weidevogel broedseizoen (1 april - 15 juli) en daarvan vervolgens de standaarddeviatie te nemen ($\sigma_{\text{delta-VH}}$), konden ze het intensiteit landgebruik van graslanden in Friesland nauwkeurig in kaart brengen.

In deze studie is de methodiek van Howison *et al.* (2018) toegepast voor heel Nederland (net als Li *et al.*, 2023). Daarbij is $\sigma_{\text{delta-VH}}$ afgeleid voor het hele jaar (1 januari - 31 december), het weidevogel broedseizoen (1 april - 15 juli) en het groeiseizoen (1 april - 30 november) van gras.

Daarnaast is ook gekeken naar andere radarsignalen en zijn soortgelijke indicatoren voor intensiteit landgebruik afgeleid:

² NDVI is het meest gevoelig voor maaien, zolang er maar voldoende NDVI-beelden zijn.

- $\text{std}(\text{delta-VV})$ of $\sigma_{\text{delta-VV}}$, is identiek aan de Howison-indicator (op basis van VH-sigitaal) maar dan op basis van het VV-sigitaal
- Backscatter signalen:
 - Standaarddeviatie van het backscatter VH-sigitaal: $\text{std}(\text{VH})$ of S_{VH}
 - Standaarddeviatie van het backscatter VV-sigitaal: $\text{std}(\text{VV})$ of S_{VV}
 - Standaarddeviatie van depolarisatie ratio (VH/VV) of kruisverhouding (CR): $\text{std}(\text{CR})$ of S_{CR}
- Coherentie signalen:
 - Standaarddeviatie van het coherentie VH-sigitaal: $\text{std}(\text{coh_VH})$ of $S_{\text{coh-VH}}$
 - Standaarddeviatie van het coherentie VV-sigitaal: $\text{std}(\text{coh_VV})$ of $S_{\text{coh-VV}}$
 - Standaarddeviatie van veranderingen in het coherentie VH-sigitaal: $\text{std}(\text{dcoh_VH})$ of $S_{\text{delta_coh-VH}}$
 - Standaarddeviatie van veranderingen in het coherentie VV-sigitaal: $\text{std}(\text{dcoh_VV})$ of $S_{\text{delta_coh-VV}}$

NB: Het vermoeden is dat met name de indicatoren gebaseerd op VH meer relevant zullen zijn dan VV. Dat heeft te maken met het feit dat de bijdrage van vegetatie in het VH-sigitaal groter is dan in het VV-sigitaal.

Intensiteit landgebruik op basis van NDVI

De methodiek van Howison *et al.* (2018) is ook toegepast op het NDVI-sigitaal, in dit geval werd dus gekeken naar de verandering tussen twee opeenvolgende beelden (delta-NDVI). Vervolgens is uit tijdserie van delta-NDVI de standaarddeviatie bepaald ($\sigma_{\text{delta-NDVI}}$), en voor dezelfde periodes (jaar, groeiseizoen en weidevogel broedseizoen).

Samengevat overzicht RS-markers als proxy voor intensiteit landgebruik:

- Maaifrequentie, 1^{ste} maaidatum
 - Periode: 1 april – 30 november
- MASD op basis van alleen de 10 m reflecties ($\text{MASD}_{10\text{m}}$) of beide 10 m+20 m ($\text{MASD}_{10+20\text{m}}$)
 - Periode: jaar, 1 april – 15 juli
 - En aantal beelden waarop MASD is gebaseerd (kwaliteitsindicator)
- Intensiteit landgebruik op basis van Radar:
 - Howison *et al.* (2018) methodiek: $s_{\text{delta-VH}}$ (en $s_{\text{delta-VV}}$)
 - Periode: jaar, 1 april – 15 juli, 1 april – 30 november
 - Standaarddeviatie van:
 - Backscatter VH & VV (σ_{VH} , σ_{VV})
 - Delta_backscatter VH & VV ($\sigma_{\text{delta-VH}}$, $\sigma_{\text{delta-VV}}$)
 - Coherentie VH & VV ($\sigma_{\text{coh-VH}}$, $\sigma_{\text{coh-VV}}$)
 - Delta_coherentie VH & VV ($\sigma_{\text{coh-dVH}}$, $\sigma_{\text{dcoh-VV}}$)
 - Periode: jaar, 1 april – 15 juli
- Intensiteit landgebruik op basis van NDVI
 - $\sigma_{\text{delta-NDVI}}$ (Howison *et al.* (2018) methodiek, maar dan toegepast op de NDVI)
 - Periode: jaar, 1 april – 15 juli, 1 april – 30 november
 - Standaarddeviatie van:
 - NDVI, WDI, NDMI (σ_{NDVI} , σ_{WDI} , σ_{NDMI})
 - Periode: jaar, 1 april – 15 juli

2.5.3.2 Visuele kenmerken

Kleur

Kruidenrijke graslanden hebben vaak een afwijkende kleur t.o.v. de productieve graslanden. Deze kleurafwijking is het gevolg van veranderingen in de samenstelling van de vegetatie door de aanwezigheid van kruiden en andere grassoorten.

Op basis van de gemeten (Rood, Groen, Blauw) reflecties kan de kleur van het gras worden bepaald. Om de analyse te vereenvoudigen zijn de RGB-waarden vertaald naar een andere kleurenschaal, de zgn. Hue-angle, waarbij de kleur (kleurtint) wordt uitgedrukt als één waarde tussen 0 en 360 graden. Er is zowel gekeken naar de gemiddelde kleur van een perceel op jaarbasis als specifiek in het voorjaar. Om te voorkomen dat maai activiteiten de kleur van het perceel verstoren worden uitsluitend beelden tussen 1 maart en 31 mei meegenomen.

Naast de Hue-angle is ook gekeken naar de verhouding tussen de rode en groene reflectie (R/G-ratio). Voor de gangbare graslanden ligt deze R/G-ratio rond ~ 0.7 (< 1 , omdat de groen-reflectie dominant is). De R/G-ratio wijkt af zodra grasland percelen minder intens groen zijn t.g.v. de aanwezigheid van andere vegetatie

(bloemen, kruiden en andere grassoorten). Net als bij de Hue-angle is gekeken naar de gemiddelde ratio op jaarbasis en voor de periode 1 maart – 31 mei. Pluspunt van de rode en groene reflecties is dat deze minder gevoelig zijn voor atmosferische variaties (en de atmosfeercorrecties die worden gebruikt bij het verwerken van de beelden). Tot slot is ook de B/G-ratio is bepaald, echter omdat met name de blauwe band erg gevoelig is voor atmosferische variaties is het vermoeden dat deze minder bruikbaar zal zijn. Tot slot zijn ook de gemiddelde reflecties voor elke band afzonderlijk (R, G, B) verzameld voor dezelfde periodes.

Spectrale diversiteit/ spectrale heterogeniteit

Spectrale diversiteit (of optische diversiteit) wordt vaak gebruikt als indicator voor soortenrijkdom (Wang *et al.*, 2016; Wang *et al.*, 2018a, b; Gholizadeh *et al.*, 2018; Xu *et al.*, 2022) en is gerelateerd aan de ruimtelijke variaties van gemeten reflecties van een set RS-pixels, in dit geval binnen een grasland perceel. In feite is het gerelateerd aan de heterogeniteit van het perceel.

Een goede indicator voor de spectrale diversiteit van een perceel is de som van de variatiecoëfficiënt (coefficient of variation, CV) van elk S2-band, in dit geval 10 banden in totaal. Uiteindelijk is de gemiddelde CV bepaald per jaar en voor de periode 1 april – 15 juli.

Samengevat overzicht RS-markers als proxy voor visuele kenmerken:

- Hue-angle
 - Periode: jaar, 1 maart – 31 mei, 1 jan – 15 maart (winter)
- Rood-groen ratio, blauw-groen ratio
 - Periode: jaar, 1 maart – 31 mei, 1 jan – 15 maart (winter)
- Gemiddelde rode, groene en blauwe reflectie
 - Periode: jaar, 1 maart – 31 mei
- Gemiddelde spectrale diversiteit (CV)
 - CV_{10 banden} (op basis van reflecties v.d. 10 S2-banden)
 - CV_{NDVI} (op basis van NDVI)
 - Periode: jaar, 1 april – 15 juli

2.5.3.3 Productiviteit

Productieve graslanden hebben over het algemeen een hogere (gras)productie dan extensieve/kruidenrijke graslanden. Er zijn allerlei technieken en modellen ontwikkeld voor het schatten van de graslandproductie (droge stof): empirische relaties op basis van NDVI, gewasgroeimodellen als dan niet in combinatie met RS-data of op basis van gemeten (satelliet) reflecties (Rood, Groen, Blauw, Red-Edge, NIR, SWIR). Vaak zijn dat moment opnames. Hier is gekozen voor indicatoren/technieken die zoveel mogelijk zijn gebaseerd op de pure RS-data zonder beïnvloeding door (AI) modellen en/of andere algoritmes.

Cumulatie NDVI (i.c.m. maaien)

Een manier om te kijken naar de productie op jaarbasis is op basis van de NDVI-waarnemingen i.c.m. de waargenomen maaimomenten. De achterliggende gedachte is: hoe vaker een perceel wordt gemaaid hoe hoger de grasproductie. Als proxy is gekozen voor de cumulatieve-NDVI (Σ NDVI), waarbij alle NDVI-afnames bij elk maaimoment gedurende het groeiseizoen worden opgeteld (Stumpf *et al.*, 2020). Percelen die intensief gemaaid worden zullen dus een hogere productie hebben dan niet-gemaaide percelen. Daarnaast hebben productieve graslanden over het algemeen wat hogere NDVI-waarden dan extensieve en kruidenrijke graslanden, wat wordt meegenomen in de sommatie.

S2REP

S2REP (Sentinel-2 Red-Edge Position) maakt gebruik van de unieke eigenschappen van de Red-Edge banden van Sentinel-2. Deze banden zijn gevoelig voor chlorofyl en stikstof concentraties in het blad en de conditie van het gewas (Scheffler *et al.*, 2014; Bekkema en Eleveld, 2018). S2REP biedt daardoor de mogelijkheid om onderscheid te kunnen maken tussen bemeste en niet-bemeste graslanden, wat weer indirect is gerelateerd aan de productiviteit van het gras. Daarnaast is S2REP minder gevoelig voor verzanding dan de NDVI, echter omdat de ruimtelijk resolutie van de Red-Edge banden 20m is, zijn hele kleine/kleine percelen minder goed te monitoren. Bekkema en Eleveld (2018) vonden dat met name de beelden opgenomen in het voorjaar de beste resultaten opleverden, voordat het maaiseizoen begint. De S2REP is bepaald voor de maanden maart t/m juli.

Voorjaar-NDVI

Tijdens de winterperiode staat de groei van grasland over het algemeen stil. Soms is er zelfs een afname zichtbaar als het een strenge winter betreft. Zodra het voorjaar begint en de temperatuur gaat stijgen komt de groei langzaam weer op gang. De Groenmonitor beelden (Groenmonitor.nl) laten zien dat productieve graslanden aan het eind van de winter begin voorjaar (eind maart), over het algemeen een hogere start NDVI-waarden hebben dan natuurlijke grasland. Dit heeft o.a. te maken met de betere groeicondities (betere nutriënten- en vochtuithouding), waardoor deze graslanden beter door de winter komen en sneller van start kunnen, resulterend in een hogere biomassa (NDVI) waarden.

Samengevat RS-markers als proxy voor productiviteit:

- Cumulatieve NDVI of Σ NDVI (is gerelateerd aan maaifrequentie) over gras groeiseizoen
 - Periode 1 april – 30 november
- S2REP
 - Momenten: maart, april, mei, juni, juli
- Voorjaar-NDVI
 - ~eind maart

2.5.3.4 Ruwheid/biomassa/vocht

Zoals toegelicht in Satelliet data is de interpretatie van radar beelden/signalen moeilijk omdat zowel de ruwheid en vocht in de bodem als in de vegetatie invloed hebben op het scatteringsmechanisme en dus het gemeten backscatter-sigitaal. Om iets te kunnen zeggen over het vochtgehalte in de bodem dient het signaal gecorrigeerd te worden voor 'vegetatie-deel' en omgekeerd, met bijbehorende onzekerheden en aannames. Gezien de onzekerheden en aannames en de grote variëteit aan type graslanden (vochtige- en natte gras- en hooilanden versus droge graslanden op zandgronden) is ervoor gekozen om zoveel mogelijk te blijven bij het originele RS-sigitaal en wordt daarom gesproken van combineerde *ruwheid/biomassa/vocht*-beelden, aangezien de data zowel iets zegt over de ruwheid van het perceel van de bodem EN/OF vegetatie (biomassa) EN vochtcondities. In Bijlage B Toelichting RS-data **Error! Reference source not found.** staan een aantal figuren die laten zien dat de originele RS-data bruikbaar is voor het onderscheiden van kruidenrijke en productieve grasland, tgv variaties is de gecombineerde ruwheid/biomassa/vocht eigenschappen.

Naast de ruwe RS-data (tijdseries) van de backscatter en coherentie signalen zijn ook gemiddelden van de backscatter VH, VV en ratio van VH en VV afgeleid voor het hele jaar en een aantal specifieke periode, zoals voor de winter en het weidevogel broedseizoen.

Samengevat RS-markers als proxy voor ruwheid/biomassa/vocht:

- Gemiddelde VH, VV
 - Periode: jaar, winter (1 jan – 15 maart)
- Gemiddelde VH/VV-ratio
 - Periode: jaar, weidevogel periode

2.5.4 Selectie co-variabelen

Tot slot zijn ook een aantal (co-variabelen) variabelen geselecteerd die van invloed zijn op de uitingvorm van het type grasland dat tot stand komt, namelijk grondwaterstand, bodemsoort en hoogte.

2.5.4.1 Grondwaterstand

De grondwaterstand fluctueert in grote delen van Nederland op geringe diepte (0-200 cm beneden maaiveld), met ondiepe standen in de winter en diepe in de zomer, en wordt uitgedrukt met zogeheten grondwatertrappen (GT's). GT's zijn klassen van de niveaus waartussen de jaarlijkse fluctuatie van de grondwaterstand zich gemiddeld genomen afspeelt. De trappen zijn gedefinieerd op basis van de gemiddeld hoogste gemeten grondwaterstand (GHG, indicatie winter toestand) en gemiddeld laagste gemeten grondwaterstand (GLG, indicatie zomertoestand). GT's worden aangegeven met Romeinse cijfers, waarbij een hoger getal aangeeft dat het grondwater overwegend dieper onder maaiveld staat (bron).

2.5.4.2 Bodemsoort

Het BOFEK (BOdemFysische EenhedenKaart) 2012-bestand geeft de bodemfysische schematisatie weer van Nederland (Wosten *et al.*, 2013). Dit bestand is gecreëerd door aan de 315 bodemeenheden behorende bij de Bodemkaart van Nederland, schaal 1:50.000, waterretentie- en doorlatendheidskarakteristieken uit de Staringreeks toe te kennen. Met een model zijn voor deze eenheden vervolgens functionele kenmerken berekend. Op grond van verwantschap in functionele kenmerken zijn de 315 bodemeenheden geclusterd in 72 bodemfysische eenheden en afgebeeld in de nieuwe BOdemFysische EenhedenKaart (BOFEK2012).

2.5.4.3 Hoogte

Het Actueel Hoogtebestand Nederland (AHN) is een bestand met gedetailleerde hoogtegegevens voor heel Nederland. In dit geval is gebruik gemaakt van AHN2 (ingewonnen tussen 2007 en 2012).

2.5.5 Overzicht verzamelde RS-markers & co-variabelen

In Tabel 4 wordt een overzicht gegeven van de verzamelde RS-markers en co-variabelen gecategoriseerd per grasland eigenschap.

Tabel 4 Overzicht verzamelde RS-markers & co-variabelen gecategoriseerd per grasland eigenschap.

Grasland eigenschappen	RS-marker	Bron RS-data	Aanvullende opmerkingen
Intensiteit landgebruik	Maaien/scheuren: - Aantal maaimomenten - Datum maaimomenten (o.a. 1^{ste} maaidatum) - Scheuren	NDVI-tijdserie (25m)	Geen waarnemingen als het bewolkt is. Om zoveel mogelijk NDVI-opnames te hebben, wordt aanvullend gebruik gemaakt van Landsat.
	Periode: Groeiseizoen (1 april – 30 november)		Vanwege de 25m-pixel grootte zijn kleine/smalle percelen moeilijk te monitoren.
	Mean Absolute Spectral Dynamics (MASD): - MASD_{10m} - MASD_{10m+20m} - Aantal beelden	S2-tijdseries, de 10x spectrale banden (10m, 20m)	Geen waarnemingen als het bewolkt is.
	Periode: - Jaar - Weidevogel broedperiode (1 april – 15 juli)		MASD is niet landelijk dekkend, omdat bepaalde delen een te lage dekkinggraad hebben. Als kwaliteitscontrole voor MASD zijn het aantal beelden geteld.
	Intensiteit landgebruik op basis van Radar data: $\sigma_{\text{delta-VH}}$ (Howison methodiek)	S1-tijdseries van backscatter (20m) en coherentie (20m)	Vanwege de 20m-pixel grootte zijn kleine/smalle percelen moeilijk te monitoren.
	Alternatieven: $\sigma_{\text{delta-VV}}$ σ_{VV}, σ_{VH}, $\sigma_{\text{coh-VV}}$, $\sigma_{\text{coh-VH}}$, $\sigma_{\text{delta_coh-VV}}$, $\sigma_{\text{delta_coh-VH}}$		
	Periode: - Jaar - Weidevogel broedseizoen (1 april – 15 juli) - Groeiseizoen (1 april – 30 november)		
	Intensiteit landgebruik op basis van NDVI: $\sigma_{\text{delta-NDVI}}$ (volgens methodiek Howison)	NDVI-tijdserie (25m)	Geen waarnemingen als het bewolkt is. Om zoveel mogelijk NDVI-opnames te hebben, wordt aanvullend gebruik gemaakt van Landsat.
	Periode: - Jaar - Weidevogel broedperiode (1 april – 15 juli)		Vanwege de 25m-pixel grootte zijn kleine/smalle percelen moeilijk te monitoren.

Grasland eigenschappen	RS-marker	Bron RS-data	Aanvullende opmerkingen
	Intensiteit landgebruik op basis van fluctuaties NDVI, WDI en NDMI: ◦ σ_{NDVI}, σ_{WDI}, σ_{NDMI} Periode: ◦ Jaar ◦ Weidevogel broedperiode (1 april – 15 juli)	S2-tijdseries, NDVI-tijdseries	Geen waarnemingen als het bewolkt is.
Visuele kenmerken	Kleur: ◦ HUE ◦ Rood-groen ratio, blauw-groen ratio ◦ Gemiddelde R, G, B Periode: ◦ Jaar ◦ 1 maart – 31 mei (voordat het maaiseizoen echt begint) ◦ Winter (1 jan – 15 maart)	S2-tijdseries, 3x 10m banden (R, G, B)	Geen waarnemingen als het bewolkt is. Variaties in de atmosfeer correctie kunnen een verstorend effect hebben op met name de blauwe reflectie.
	Spectrale diversiteit/heterogeniteit (CV): ◦ $CV_{10\text{banden}}$ ◦ CV_{NDVI} Periode: ◦ Jaar ◦ Weidevogel broedperiode (1 april – 15 juli)	S2-tijdseries, de 10x spectrale banden (10m, 20m) en NDVI-tijdserie (25m)	Geen waarnemingen als het bewolkt is.
Productiviteit	NDVI-voorjaar (eind maart) ◦ $NDVI_{\text{jaar}}$	NDVI-tijdserie (25m), specifiek eind maart	Geen waarnemingen als het bewolkt is.
	Jaar gemiddelde WDI, NDVI	NDVI-tijdserie (25m) en S2-tijdseries	Geen waarnemingen als het bewolkt is.
	Sentinel-2 Red-Edge Position (S2REP): ◦ Maart ◦ April ◦ Mei ◦ Juni ◦ Juli	S2-tijdseries, Red-Edge banden (20m), specifiek maanden maart t/m juli	Geen waarnemingen als het bewolkt is.
	Cumulatieve NDVI ($\Sigma NDVI$) Periode: ◦ 1 april – 30 november	NDVI-tijdserie (25m)	Geen waarnemingen als het bewolkt is. Om zoveel mogelijk NDVI-opnames te hebben, wordt aanvullend gebruik gemaakt van Landsat. Vanwege de 25m-pixel grootte zijn kleine/smalle percelen moeilijk te monitoren.
Ruwheid/biomassa/vocht	Gemiddelde VH, VV Periode: ◦ Jaar ◦ Winter (1 jan – 15 maart) Gemiddelde ratio VH/VV Periode: ◦ Jaar ◦ Weidevogel broedperiode (1 april – 15 juli)	S1-tijdseries van backscatter (20m)	Vanwege de 20m-pixel grootte zijn kleine/smalle percelen moeilijk te monitoren.
Co-variabelen	◦ Grondwatertrappen (GT – 8 klasse) ◦ Bodemsoort (BOFEK, 5 klasse) ◦ Hoogte (AHN2)	Grondwatertrappen kaart, BOFEK 2012 en AHN2	

2.6 Ontwikkeling RS-algoritme

Doel van het RS-algoritme (spoor 1) is om de kruidenrijkdom (als Q-index) te schatten van grasland percelen in Nederland op basis van de verzamelde RS-data en/of afgeleide RS-markers zoals beschreven in Inwinnen RS-data.

De verzamelde RS-data en -markers voor de jaren 2018 t/m 2022 met de bijbehorende vegetatieopnames (vertaald naar Q-index) zal deels gebruikt worden voor het ontwikkelen en trainen van het algoritme en deels voor de validatie van het algoritme. Wanneer de resultaten van het algoritme goed genoeg zijn kan het gebruikt worden om op jaarbasis voor toekomstige jaren (2023 en later) een uitspraak te doen over de kruidenrijkdom van alle BRP-grasland percelen in Nederland.

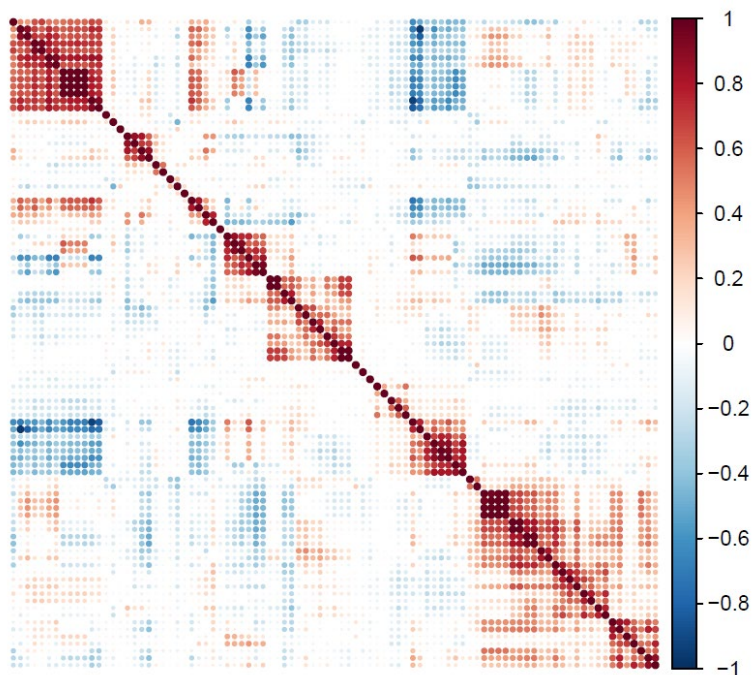
De volgende paragrafen zijn gebaseerd op de onderzoeksrapporten van Ron Wehrens (Biometris, RR-, RF-, RFC- en SOM modellen, zie Bijlage D) en NEO (CNN-model, zie Bijlage E). Deze paragrafen geven een korte samenvatting, meer informatie is terug te vinden in de Bijlagen.

2.6.1 Model selectie

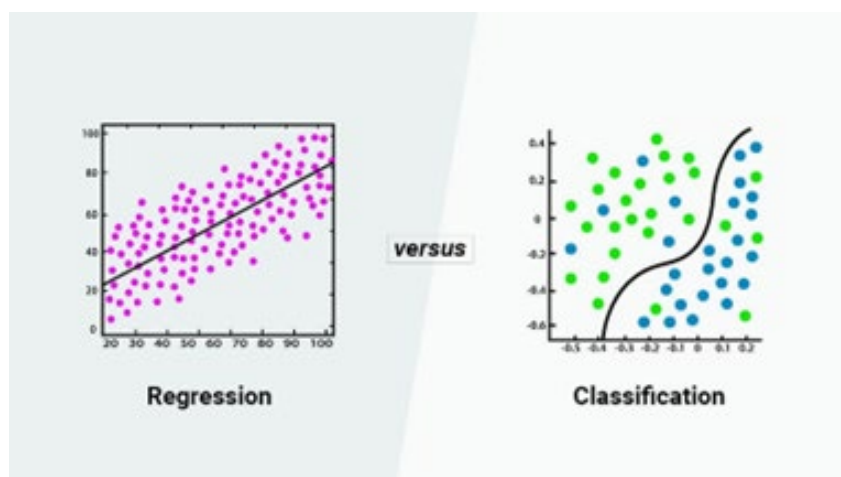
Op basis van de soort data (in dit geval de verzamelde RS-data en RS-markers) waarmee de kruidenrijkdom te voorspellen, is gekozen voor vier verschillende type modellen: een Ridge Regression (RR) model, een Random Forest (RF) model, een Self Organizing Maps (SOM) model en een Convolutional Neural Network (CNN) model. In soortgelijke studies, ook gericht op het voorspellen van soortenrijkdom (Simpson, Shannon) op basis van RS-data, zijn dezelfde soort modellen (RF, deep learning) gebruikt (Fauvel *et al.*, 2020; Muro *et al.*, 2022).

2.6.1.1 Ridge Regression

Ridge Regression (RR, Hoerl en Kennard, 1970) is een methode voor het schatten van de coëfficiënten van een meervoudig lineair regressie model waarbij niet alleen de beste fit wordt gezocht, maar waarbij de model coëfficiënten zo klein mogelijk worden gehouden. Dat maakt het mogelijk de techniek toe te passen in situaties waar normale lineaire regressie faalt, bijvoorbeeld in situaties waarin de onafhankelijke variabelen sterk gecorreleerd zijn. Dat sommige variabelen (RS-markers) zeer sterk gecorreleerd zijn, is te zien in Figuur 11. Dit figuur toont clusters van RS-markers die zeer sterk zijn gecorreleerd (donkerrood) of anti-gecorreleerd (donkerblauw). Om te voorkomen dat het model gaat over-fitten is het van belang dat de zgn. 'penalty parameter' van het model wordt geoptimaliseerd. Dat gebeurt typisch door kruisvalidatie. Vervolgens kan het RR-model vervolgens uitspraken doen over de kruidenrijkdom in nieuwe data (Figuur 12, links).



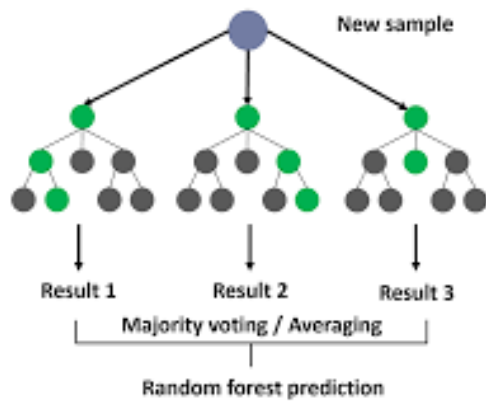
Figuur 11 Correlatie tussen de RS-markers. Met name de donkerrode en blauwe blokken zijn clusters van RS-markers die zeer sterk (anti-)gecorreleerd zijn.



Figuur 12 Regressie versus classificatie.

2.6.1.2 Random Forests

Het Random Forest (RF, Breiman, 2001) model bestaat uit een verzameling beslisbomen, die getraind zijn op verschillende stukjes van de dataset (Figuur 13). Groot voordeel van een RF-model is dat het goed kan omgaan met grote hoeveelheden data en niet erg gevoelig is voor overfitten. Daarnaast kan het model zowel worden gebruikt als regressie model (zoals het RR-model) of als classificatie model (Figuur 12, rechts). In het laatste geval zal het model voorspellen in welke (sub) Q-index klasse/categorie het resultaat valt, bijvoorbeeld in (de vooraf gedefinieerde) categorieën Q-index 1 tot 1,5, 1,5 tot 2, 2 tot 2,5, etc.



Figuur 13 Random forests model bestaande uit een set van beslisbomen. De uitspraak van het model wordt bepaald door de grootste groep beslisbomen met dezelfde uitkomst (bron: Yehoshua, 2023).

2.6.1.3 Self-organising maps

Self-organising maps (SOM's, Kohonen 2001) zijn clusteringsmethoden, waarbij op elkaar lijkende records uit een data set bij elkaar worden gevoegd, en gevisualiseerd in een twee-dimensionale map. In deze context worden ze gebruikt om database entries die voor zowel de RS markers als Q waardes erg op elkaar lijken te groeperen, volgens de methodologie van Wehrens en Kruisselbrink (2018). Nieuwe gegevens, waarvoor alleen RS markers beschikbaar zijn, kunnen op dezelfde map worden afgebeeld – de Q waarde die bij de plek hoort waar ze terechtkomen is dan de schatting voor de kruidenrijkdom voor de nieuwe records.

2.6.1.4 Convolutional Neural Network

De voorgaande modellen zijn alleen geschikt voor de RS-markers dataset. Aangezien de ruwe RS-data tijdseries dezelfde informatie bevat als de RS-markers en wellicht zelfs nog meer, is besloten om ook nog een model te kiezen dat wél kan omgaan met de RS-data tijdseries (zoals te zien in Figuur 10). In dit geval is gekozen voor een Convolutional Neural Network (CNN) model. Een neurale netwerk is een methode binnen de 'Machine Learning' waarmee standaard vraagstukken zoals regressie en classificatie opgelost kunnen worden en zeer geschikt is om in te zetten voor complexere taken zoals beeld- geluid- en taalherkenning. In dit geval zal het model worden getraind op de ruwe RS-data tijdseries van zowel de optische data en radar data. Uiteindelijk is nog wel besloten om de RS-markers dataset aan te vullen met jaarlijkse statistiek van de ruwe RS-data, d.w.z. de jaarlijkse gemiddelde waarden en de standaarddeviaties.

Tabel 5 Overzicht model keuze en bijbehorende databronnen (RS-data en -markers).

Model	RS-data	RS-markers	Training-modus
RR, SOM	Nee, alleen jaarlijks gemiddelde + standaarddeviatie	Ja	Regressie
RF	Nee, alleen jaarlijks gemiddelde + standaarddeviatie	Ja	Regressie (RF) + Classificatie (RFC)
CNN	Ja	Nee, alleen co-variabelen (bodem, GT en hoogte)	Regressie + Classificatie

2.6.2 Koppeling Q-index en RS-data/markers

Zowel de vegetatieopnames als de RS-data/markers zijn 'vertaald' naar BRP-perceel niveau (zie Vegetatieopnames op BRP-perceel niveau, en Van satelliet pixels naar BRP-perceel). De koppeling van de Q-index en de RS-data/markers is als volgt gedaan:

Perceel met Q-index van jaar 20xx is gekoppeld met de RS-data/markers van dat perceel van jaar 20xx.

De uiteindelijke training set bestaat dus uit 5 jaar, namelijk 2018 t/m 2022, waarbij de percelen die in jaar 2018 zijn bezocht (dus met een Q-index 2018) zijn gekoppeld met de RS-data/markers van 2018, etc.

2.6.3 Dataselectie en voorverwerking, model strategie

Voordat de modellen/algoritmes kunnen worden getraind zijn de zowel de RS-markers, RS-data en Q-indices nogmaals nagelopen, beoordeeld en waar nodig aangepast en verder opgeschoond (in het geval ze bij de selectie van de vegetatieopnames niet waren opgevallen).

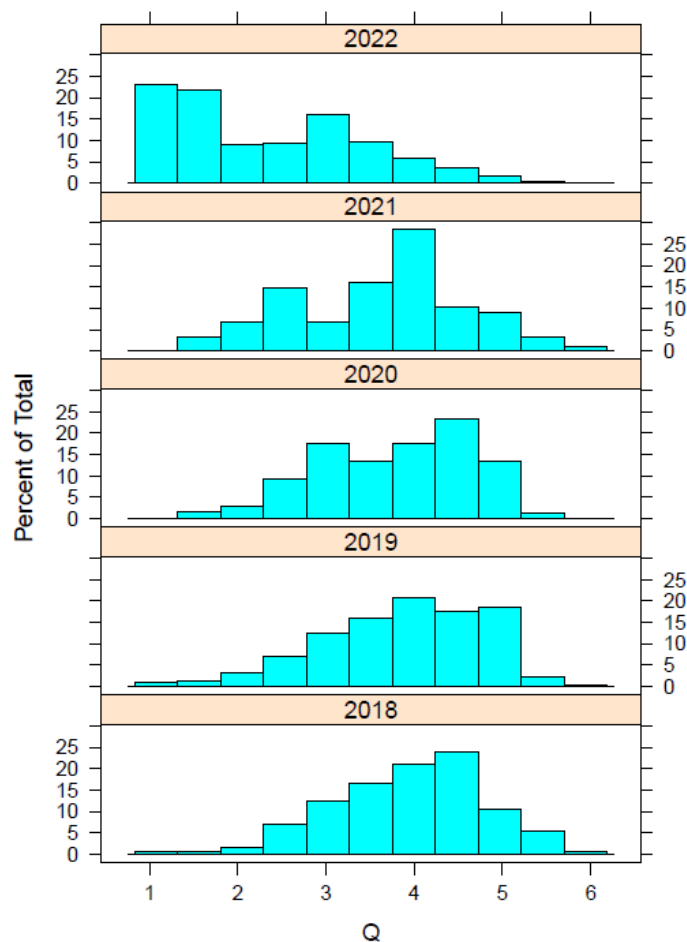
Dataselectie en voorverwerking:

- Representativiteit van de percelen met een vegetatieopname (Q-index). Percelen waar sprake is van sterke verruiging (aanwezigheid van struiken en lage vegetatie) en dus meer 'natuurterrein' dan 'kruidenrijk grasland' voorstellen, zijn verwijderd uit de dataset.
- Ontbrekende waarden in de RS-markers. RS-markers met te veel ontbrekende data (0-waarden of geen data) zijn verwijderd. RR en RF modellen mogen helemaal geen ontbrekende waarden bevatten, waardoor enkele belangrijke optische gebaseerde RS-markers, zoals S2REP en MASD, bij deze modellen niet gebruikt worden. Bij SOM's mogen er wel waarden ontbreken in de data (tot op zekere hoogte). Daarnaast zijn de 'productiviteit'-markers gebaseerd op maai-activiteiten ook vervallen omdat deze niet beschikbaar zijn voor kleine grasland percelen.
- Correlaties tussen variabelen. Met name bij SOM modellen kunnen hoge correlaties invloed uitoefenen: een groep van hoog-gecorrleerde variabelen oefent dus een buitensporig grote invloed op het eindresultaat uit. Daarom zijn hele hoge correlaties bij SOM modellen vermeden door van zo'n groep gecorrleerde variabelen slechts 1 variabele te selecteren – in principe die variabele die het minste ontbrekende waarden heeft.
- De RS-data tijdseries. De originele RS-tijdseries bestaan uit een verzameling datapunten (datums waarop beelden verzameld/beschikbaar zijn), echter de frequentie van deze punten is onregelmatig, zeker bij de optische data als gevolg van bewolking. Deze 'onregelmatige' tijdseries zijn omgezet naar 'regelmatige' wekelijkse tijdseries.
- Normalisatie: data zijn genormaliseerd zodanig dat elke variabele een gemiddelde van nul heeft, en een standaard deviatie van 1.

Model strategie:

- Verkenning en beoordeling van de onafhankelijke variabelen (RS-markers). Welke zijn gecorrleerd, wat is hun range? Omzetten van categorische variabelen naar indicator variabelen.
- Splitsen dataset voor training en validatie. De uiteindelijke dataset bestaande uit goedgekeurde percelen met bijbehorende overgebleven RS-markers en tweewekelijkse RS-data tijdseries is vervolgens gesplitst in een trainingsdataset (~80%) en een validatie dataset (20%).
- Optimaliseren penalty-parameter: In het geval van het RR-model is het van belang dat de penalty-parameter wordt geoptimaliseerd ter voorkoming van overfitten van het model.

Figuur 14 laat per jaar de verdeling van Q-index zien (histogram) van de uiteindelijke trainings- en validatiedataset. Duidelijk is te zien dat de jaren 2018 t/m 2021 voornamelijk hoge Q-waarden (> 3) bevatten. Daarom was besloten dat de vegetatieopnames die in 2022 zijn gedaan binnen dit project waren gericht op percelen met een lage Q-index (< 2).



Figuur 14 Verdeling van de verzamelde Q-index voor de jaren 2018 t/m 2022.

In Tabel 6 wordt een overzicht gegeven van het aantal overgebleven percelen (1022) en variabelen (113, bestaande uit RS-markers en jaar-statistiek RS-data).

Tabel 6 Overzicht van overgebleven RS-markers voor de training van het RR- en RF-algoritme.

Grasland eigenschappen	RS-marker
Intensiteit landgebruik	Maaien: - Datum maaimomenten [dag nummer] (1^{ste} t/m 5^{de}) Intensiteit landgebruik op basis van Radar data: $\sigma_{\text{delta-VH}}$ (Howison methodiek) $\sigma_{\text{delta-VV}}$ $\sigma_{\text{VV}}, \sigma_{\text{VH}}, \sigma_{\text{coh-VV}}, \sigma_{\text{coh-VH}}, \sigma_{\text{delta_coh-VV}}, \sigma_{\text{delta_coh-VH}}$ Intensiteit landgebruik op basis van fluctuaties NDVI, WdVI en NDMI: $\sigma_{\text{NDVI}}, \sigma_{\text{WdVI}}, \sigma_{\text{NDMI}}$
Visuele kenmerken	Kleur: - HUE - Rood-groen ratio, blauw-groen ratio - Gemiddelde R, G, B Periode: - Jaar 1 maart – 31 mei Spectrale diversiteit/ heterogeniteit (CV): $\text{CV}_{10\text{banden}}$ CV_{NDVI} Periode: - Jaar - Weidevogel broedperiode (1 april – 15 juli)
Productiviteit	NDVI-voorjaar (eind maart) $\text{NDVI}_{\text{vjaar}}$ Jaar gemiddelde WdVI, NDVI

Grasland eigenschappen	RS-marker
Ruwheid/biomassa/vocht	Gemiddelde VH, VV Periode: - o Jaar - o Winter (1 jan – 15 maart) Gemiddelde ratio VH/VV Periode: - o Jaar - o Weidevogel broedperiode (1 april – 15 juli)
Co-variabelen	- o Grondwatertrappen (GT, 8 klasse) - o Bodemsoort (BOFEK 5 hoofdklassen: veengrond, moerige grond, zandgrond, kleigrond, leemgrond) - o Hoogte [meters]
Jaarlijkse gemiddelde en standaarddeviaties van de RS-data tijdseries	S1, S2 banden

2.7 Ontwikkeling foto algoritme

Doel van het foto algoritme is om de kruidenrijkdom (als Q-index) te schatten van grasland percelen op basis van een foto (spoor 2a). Daarnaast zal er een tweede foto-algoritme worden ontwikkeld specifiek getraind op het kunnen onderscheiden van PKG-percelen van regulier Engels raaigras (spoor 2b).

2.7.1 Fotoselectie en voorverwerking

De verzameling van foto's was erg gevarieerd, zowel qua dimensies als qua belichting; soms is de foto op een zonnig moment genomen andere foto's zijn juist op een erg sombere dag genomen. Zeker bij de overzicht foto's waren erg veel foto's waarbij de horizon zich niet bijna boven in het beeld bevond maar juist midden het beeld, waarbij ook boerderijen, bosschages en dergelijke in beeld kwamen. Veel van deze foto's waren daarom ook niet bruikbaar. Daarom zijn de detailfoto's voor het algoritme gebruikt. Het aantal foto's om het algoritme op te trainen is vergroot door middel van zogenaamde gegevens verrijking. Hierbij zijn uit oude beelden nieuwe beelden gemaakt, door beelden te spiegelen, het contrast te wijzigen en de verzadiging van het beeld te veranderen. Ook zijn de beelden verkleind zodat de kortste zijde 200 pixels groot werd. Uit de overgebleven afbeelding is vervolgens een stuk van 200x200 pixels geknipt. Een illustratie hiervan is te zien in Figuur 15.



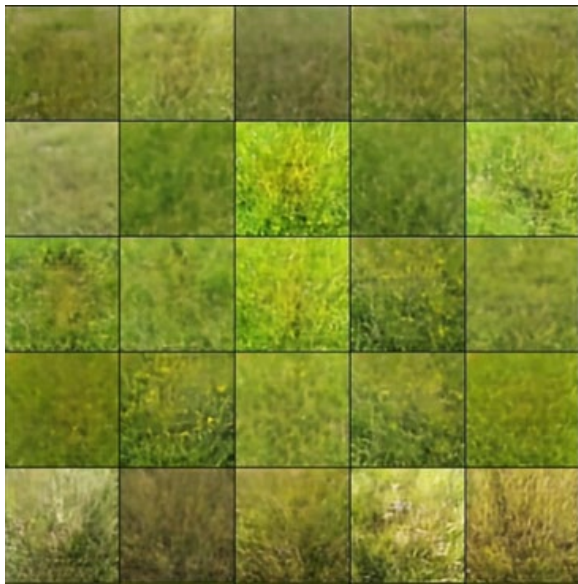
Figuur 15 Een voorbeeld van hoe uit een foto vier verschillende foto's zijn gemaakt om de trainingsgegevensverzameling te verrijken.

De gebruikelijke methode om een model te trainen is om de data punten op te delen in een train-, test- en validatie dataset. Zo'n algoritme wordt dan 'getraind' met behulp van de train dataset. Er wordt dan gezocht naar de optimale parameters. Bij de training worden vooraf heel veel keuzes gemaakt. Deze keuzes worden vastgelegd in de zogenaamde hyperparameters. De prestatie voor de verschillende hyperparameters wordt dan gemeten met de test dataset. Uiteindelijk wordt het beste model geëvalueerd op basis van de validatie dataset. De percelen zijn ingedeeld in een trainings-, test- en validatie dataset, die respectievelijk 60%, 20% en 20% van de totale verzameling percelen bevatten. Er is hier op gelet dat niet de verschillende foto's van eenzelfde perceel in een andere dataset terecht komen. De foto's in de validatie dataset zijn niet verrijkt, maar alleen verkleind tot een beeld van 200x200 pixels.

2.7.2 Voorspellen kruidenrijkdom (spoor 2a)

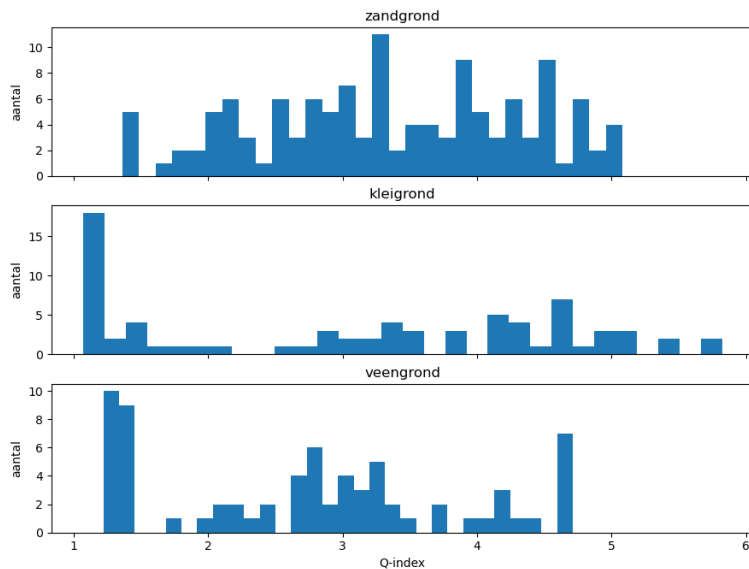
De foto's van de percelen waarvan een vegetatie opname beschikbaar is, en dus een Q-index hebben, zijn gebruikt om een algoritme te ontwikkelen waarmee de kruidenrijkdom kan worden geschat (Q-index). Twee soorten algoritmen zijn hiervoor verkend: een regressie en een classificatie (net als bij het RS-algoritme) algoritme, waarbij de Q-index op een continue schaal of in subklasse wordt voorspeld.

Voor het eerste algoritme is een regressie methode gebruikt. Na de voorverwerking van de foto's bestaat iedere foto nog altijd uit $200 \times 200 \times 3 = 120.000$ waarden. Dat zijn er veel meer dan het aantal beschikbare verrijkte foto's. Het is daarom belangrijk om, voordat er een regressie uitgevoerd kan worden, het aantal dimensies sterk te verminderen. Omdat het hier om beeldmateriaal gaat liggen klassieke methoden hiervoor, zoals bijvoorbeeld hoofdcomponentenanalyse niet voor de hand. Om het aantal dimensies te reduceren hebben is een zogenaamde variational autoencoder (VAE) gemaakt (Kingma *et al.* 2013). Hiermee zijn de afbeeldingen met een neuraal netwerk met ResNet18 (He et al 2016) architectuur naar 225 dimensies teruggebracht. Met een VAE kunnen de originele beelden weer worden gereconstrueerd. In Figuur 16 zijn een aantal van die reconstructies ter illustratie weergegeven.



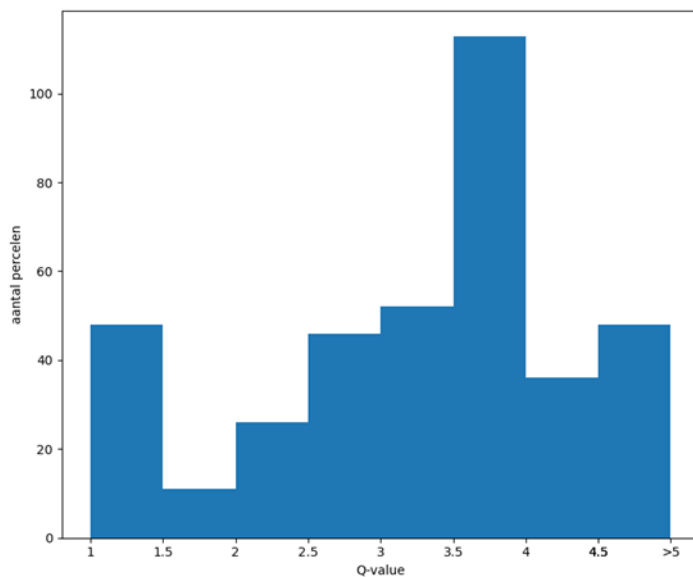
Figuur 16 Een aantal gereconstrueerde verrijkte beelden. De reconstructie is gedaan vanuit een 225 dimensionale ruimte.

Vanuit de 225 dimensionale ruimte is er met een neuraal netwerk, zoals beschreven in (Specht, 1996), een regressie algoritme gemaakt. Veel van de vegetatie hangt af van het bodemtype; veel soorten gedijen beter op het ene bodemtype dan op het andere. Daarom zijn de foto's opgesplitst per bodemtype, zand, klei en veen. Dit is gedaan met de foto's zoals die er waren eind 2022. Er waren toen foto's en Q-indices van 127, 71, 75 percelen op respectievelijk zand-, veen- en kleigrond. Figuur 17 toont de histogrammen van de Q-indices van de verschillende percelen per grondsoort.



Figuur 17 Histogrammen van de Q-indices van de percelen waarvan eind 2022 foto's waren.

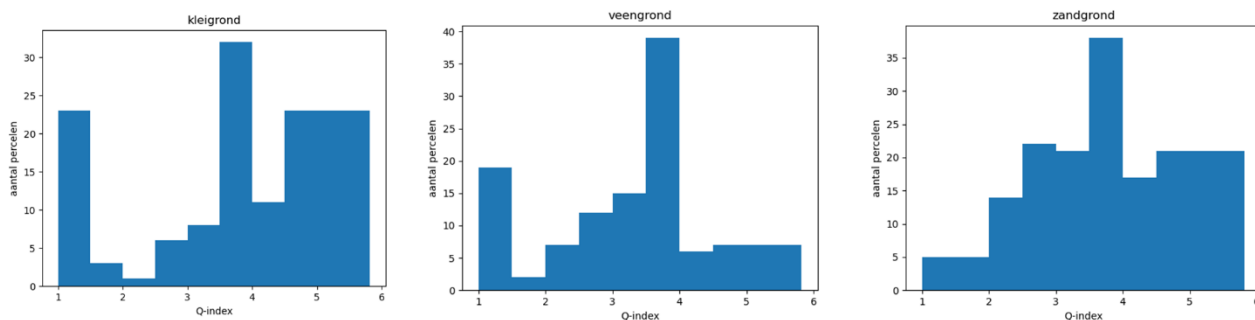
De tweede benadering bestond uit het opdelen van de beelden in classes. Een klasse bestaat uit die beelden die een Q-index hebben tussen de $(x, x+0,5)$, voor x in $\{1, 1,5, 2, 2,5, 3, 3,5, 4\}$ en een laatste klasse waarin alle percelen opgenomen zijn met een Q-index groter dan 4,5. Dit laatste is gedaan omdat er maar enkele percelen met een Q-index van groter dan 4,5. In Figuur 18 is de verdeling van de Q-index van de percelen waarvan er goede foto's zijn. Opmerkelijk is dat het aantal percelen met Q-index tussen de 1,5 en 2 aanmerkelijk lager is dan bij de overige waarden.



Figuur 18 De verdeling van de Q-index over de percelen. In totaal waren er eind 2023 383 percelen met goede foto's en een Q-index.

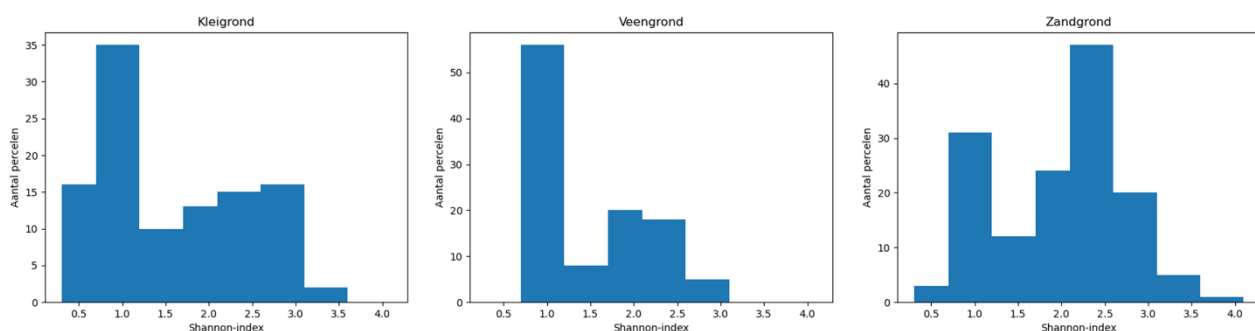
Op de beelden verdeeld in de verschillende classes is vervolgens een classificatie algoritme getraind. Hiervoor is het zogenaamde yolo8 algoritme, dat is op dit moment het state-of-the-art algoritme voor classificatie. Vegetaties op verschillende grondsoorten zijn nogal verschillend, en daarom zijn, in derde instantie, de percelen verdeeld in 3 grondsoorten op basis van de bodemkaart van Nederland (Figuur 71): klei-, veen- en zandgrond. Veel foto's zijn genomen op plekken waar geen grondsoort op de kaart staat, bijvoorbeeld in de stad, of soms net in het water. Aan deze foto's zijn toch grondsoorten toegekend door de meest naburige grondsoort te kiezen. Voor deze drie grondsoorten zijn vervolgens weer yolo8 classificatie algoritmes gemaakt. Omdat op ieder perceel meerdere detailfoto's zijn genomen, is er als gehele

voorspelling voor het perceel de voorspelling die het meest voor kwam genomen. In Figuur 19 wordt per grondsoort de verdeling van de Q-index voor 383 percelen weergegeven.



Figuur 19 De verdeling van de Q-index over de percelen per grond soort. In totaal waren er eind 2023 383 percelen met goede foto's en een Q-index.

Dit is nog eens gedaan door de Shannon index te gebruiken in plaats van de Q-index. In Figuur 20 is de verdeling van het aantal percelen per grondsoort en per Shannon-index te zien.



Figuur 20 Verdeling van de aantal percelen met bruikbare foto's, verdeeld per grondsoort en per Shannon-index.

2.7.3 Herkennen PKG-percelen (spoor 2b)

Er is gekozen om voor het maken van het foto algoritme gebruik te maken van een neurale netwerk. Omdat het aantal foto's beperkt is, is het noodzakelijk een al voorgetraind netwerk te gebruiken. Hiervoor is het huidige *state-of-the-art* netwerk yolov8x-cls gebruikt. Voor de selectie van regulier Engels raaigras percelen zijn de foto's van de graslanden gebruikt die een Q-index hebben van kleiner dan 1,4. Voor de selectie van PKG-percelen zijn de foto's gebruikt waarvan aangegeven is dat ze ingezaaid zijn met een grasklaver mengsel of met een kruidenrijk voedermengsel. In Tabel 7 wordt een overzicht gegeven van het aantal percelen waar foto's van beschikbaar zijn, het aantal detailfoto's over die percelen en het aantal augmented detailfoto's die gebruikt zijn voor het classificatie algoritme met daarbij ook de onderverdeling naar het type grasland.

Tabel 7 Een overzicht van het aantal foto's dat gebruikt is voor het classificatie algoritme.

Grasland type	Aantal percelen	Aantal detailfoto's	Aantal augmented detailfoto's
Regulier Engels raaigras	197	1785	2415
Productief Kruidenrijk Gras (PKG)	174	631	2233

2.8 Ontwikkeling database (en ontsluiting resultaten)

2.8.1 RS-data

Bij de start van fase 1 van dit project is besloten om voor alle grasland percelen van Nederland RS-data te verzamelen, en niet alleen voor percelen met vegetatieopnames. Hierbij is voortgebouwd op de logica van de AgroDataCube (ADC). Het doel van de ADC was om diverse soorten open data (Geo, GIS en Remote Sensing) op het gebied van landbouw overzichtelijk samen te brengen op één plek. In de ADC wordt veel informatie verzameld op BRP-perceelniveau, het landelijke registratiesysteem van de overheid. De beschikbare (open)data omvat o.a. jaarlijkse BRP-informatie (zoals gewastype, en perceelgrenzen) sinds 2009, bodemkaarten, KNMI-weerdata, maar ook RS-data zoals NDVI-tijdseries. Data wordt continue aangevuld en/of geüpdatet, waaronder jaarlijks de BRP. Daarbij is de database zo opgezet dat perceel informatie van verschillende jaren opvraagbaar is waarbij rekening is gehouden met de jaarlijkse veranderingen die plaatsvinden in de BRP (zoals het verschuiven/splitsen/samenvoegen van percelen). Data uit de ADC is via een API op te vragen, en kan zelfs gekoppeld worden aan websites met viewers (zoals de Groenmonitor, de Open Bodem Index en de Waterbehoefte Viewer).

Gezien de goede ervaringen met het gebruik van de ADC is in dit project besloten om de structuur van de ADC te volgen. Alle verzamelde RS-data van de jaren 2018 t/m 2022, zowel de ruwe RS-data als de afgeleide markers zijn allemaal beschikbaar op BRP-perceelniveau. In theorie is het daardoor redelijk eenvoudig om zowel de ruwe RS-data als de afgeleide RS-markers én model uitkomsten van het RS-algoritme(s) toe te voegen en beschikbaar te maken middels de ADC.

2.8.2 Foto's

De foto's zijn op 3 manieren verzameld; met de Ellipsis Drive app, de ODK app en los aangeleverd. De eventuele vegetatie opnamen zijn via andere bronnen gekomen. De foto's van de ODK app, en de Ellipsis Drive app staan op de respectievelijke servers. Ze zijn allemaal gedownload naar de sharepoint folder van het project.

Op basis van herkomst, de vragen en de locaties zijn de foto's opgedeeld in twee aparte datasets. In de eerste dataset staan de foto's met grasland type, raaigras, klaver, productief kruidenrijk of natuurlijk. Als er op basis van de locatie een vegetatie opname bekend was dan zijn foto's opgenomen in de tweede dataset, inclusief de Q-index, kruidenrijkdom en Shannon-index.

3 Resultaten

In sectie 3.1 zijn de resultaten van de getrainde RS-algoritmes gepresenteerd (spoor 1), gebaseerd op de model- en data selectie beschreven in Ontwikkeling RS-algoritme. Deze resultaten van de RS-algoritmes zijn deels gebaseerd op de onderzoeksrapporten van Ron Wehrens (RR, RF, RFC en SOM modellen) en NEO (CNN-model). Deze rapporten zijn terug te vinden in de Appendices D en E. **NB:** Het rapport van NEO bevat ook een analyse van de RR- en RF-modellen. Hier worden alleen de CNN-resultaten van het rapport van NEO besproken.

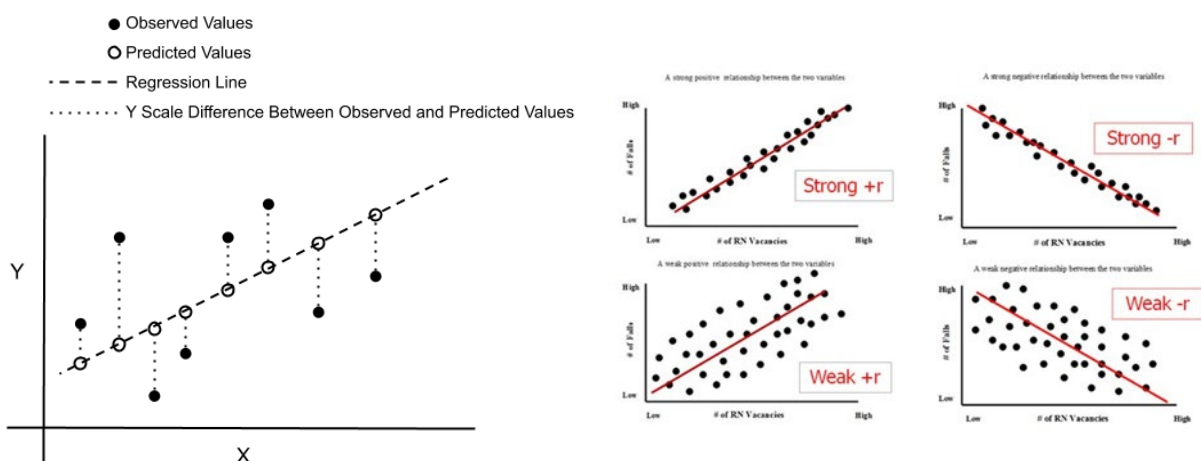
Sectie 3.2 bevat de resultaten van het foto-algoritme (spoor 2a+b), gebaseerd op de model- en data selectie beschreven in Ontwikkeling foto algoritme.

In Sectie 3.3 worden spoor 1 en spoor 2 naast elkaar gelegd, waarbij wordt gekeken naar de herkenning van PKG-percelen met behulp van het foto-algoritme (spoor 2b) en wat de uitspraken van het RS-algoritme zijn. Tot slot worden de voorspelde Q-index waarden van beide sporen met elkaar vergeleken alsmede met de vegetatieopnames (spoor 1 versus spoor 2a).

3.1 Resultaten Remote Sensing algoritmes

3.1.1 Resultaten RS-algoritmes

Bij de evaluatie van de model-trainingen is gekeken naar de gemiddelde voorspelfout van de modellen (d.w.z. de Root Mean Square Error ofwel RMSE) en de correlatiecoëfficiënt (r^2) tussen de gemodelleerde- en gemeten Q-index. De RMSE is het gemiddelde verschil tussen de voorspelde Q-index en de werkelijke Q-index (ook wel error of residu, Figuur 21 links). Hoe lager de RMSE des te nauwkeuriger het model kan voorspellen. De RMSE is te relateren aan het 95%-betrouwbaarheidsinterval, welke aangeeft dat als een model 100x een voorspelling doet, het 95 keer een goede uitspraak doet en slechts 5 keer een foute. Het 95% betrouwbaarheidsinterval is nagenoeg gelijk aan twee keer de RMSE. De correlatiecoëfficiënt is een statistische maatstaf voor de sterkte van een lineair verband tussen model resultaten en de werkelijke waarden. Hoe hoger r^2 des te beter het verband tussen model en waarnemingen (Figuur 21, rechts).



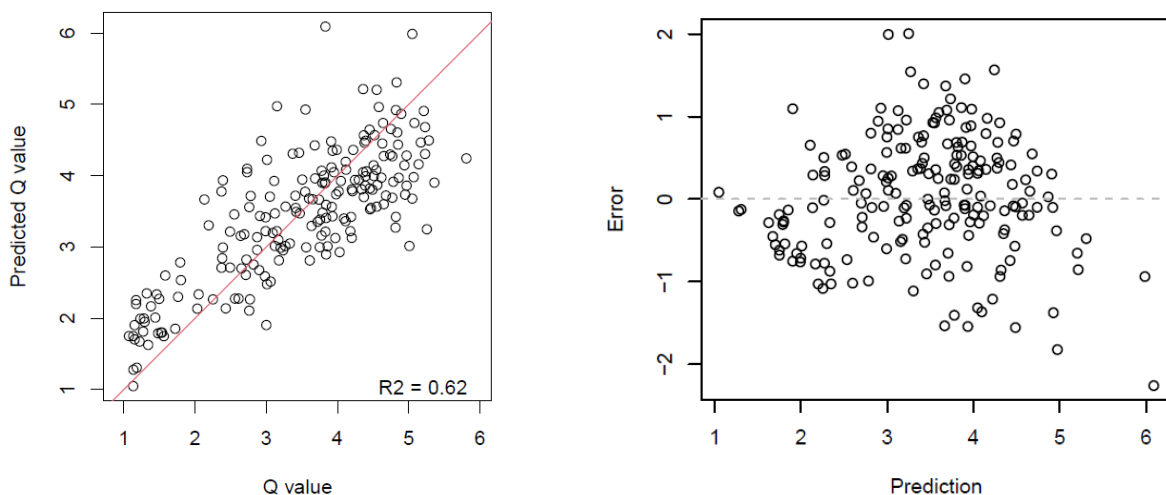
Figuur 21 Model evaluatie door te kijken naar de RMSE (links), het gemiddelde verschil tussen model (witte-rondjes) en de werkelijke waarden (zwarte-rondjes), en de correlatiecoëfficiënt (r^2 , rechts), als maat voor het lineaire verband tussen model resultaten en de werkelijke waarden.

In Tabel 8 wordt een overzicht gegeven van de trainingsresultaten voor alle drie de modellen voor zowel regressie als classificatie analyse (op basis van de validatie dataset). In het geval van de regressieanalyse zijn de RMSE en r^2 direct af te leiden. Bij de classificatie-analyse is gekozen voor een bin-grootte van 0.5 (delta Q), d.w.z. de Q-waarnemingen zijn gegroepeerd in de volgende subgroepen Q=1-1,5, Q=1,5-2, Q=2-2,5 t/m Q=5,5-6. De resultaten van een classificatie-model worden meestal uitgedrukt in een confusion-matrix, waaruit vervolgens de nauwkeurigheid (in %) van het model kan worden afgeleid. Om de resultaten van de classificatiemodellen te kunnen vergelijken met de regressie-resultaten is een RMSE afgeleid door voor elke bin het middel punt te nemen.

Tabel 8 Overzicht model trainingsresultaten op basis van de validatie dataset (bron: Bijlage D en E).

Model	RMSE	Nauwkeurigheid [%]	r^2
<i>Regressie</i>			
Ridge Regression (RR)	0.72	-	0.62
Random Forests regressie (RF)	0.63	-	0.72
Convolutional Neural Network (CNN)	0.65 ³	-	-
Self Organizing Maps (SOM)	0.82	-	-
<i>Classificatie</i>			
Random Forests classificatie (RFC)	0.73	36.5	-
Convolutional Neural Network (CNN)	-	37.4	-

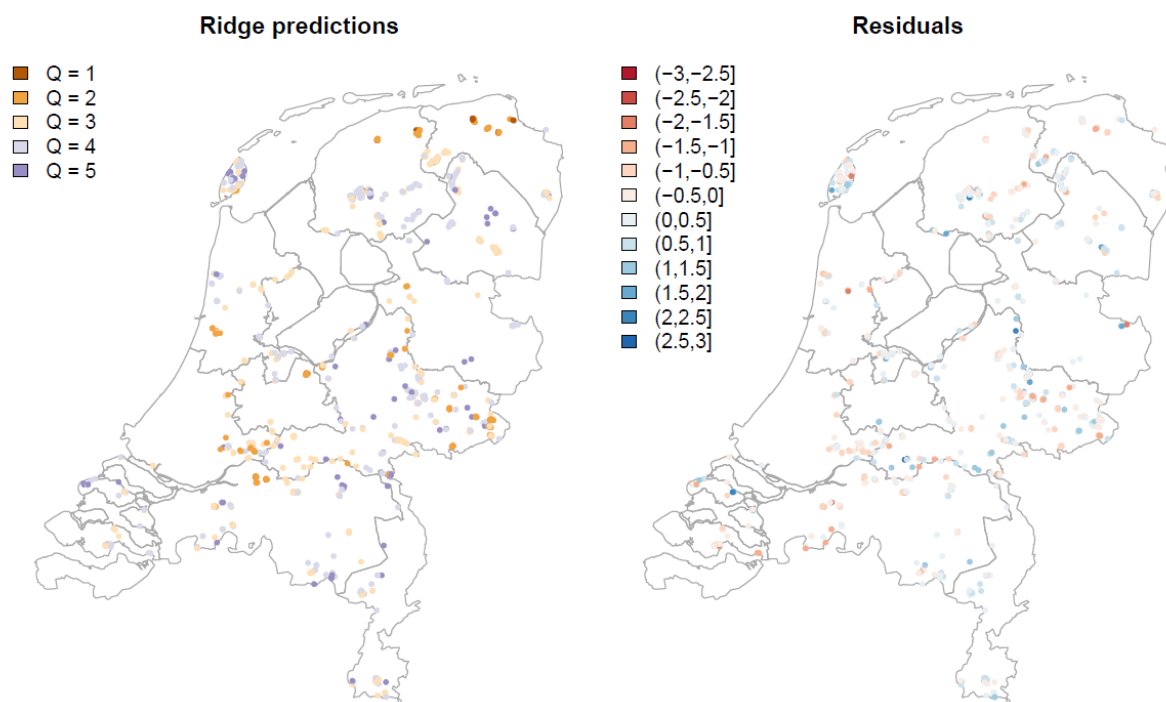
In Figuur 22 worden de resultaten van het RR-model weergegeven als scatter-plot (links) waarbij de voorspelde Q-index (y-as) is geplot tegen de werkelijke Q-index (x-as). Het rechter deel van Figuur 22 is een residuen-plot waarbij de modelfout (d.w.z. het verschil tussen gemodelleerde en werkelijke Q) is uitgezet tegen de gemodelleerde Q. NB: Dit zijn model resultaten waarbij de 'penalty parameter' van de RR-methode is geoptimaliseerd. Zowel de linker als rechter figuur laten zien dat het RR-model bij hoge Q-index de neiging heeft om te onderschatten. Dit gedrag ziet men vaker bij lineaire modellen zoals het RR-model, bijvoorbeeld wanneer het echte verband tussen de variabelen niet helemaal lineair is.



Figuur 22 Resultaten RR model op basis van de validatie dataset, links als scatter-plot (Q-model versus werkelijke Q) en rechts als residuen-plot waarbij het verschil tussen model en werkelijke Q is uitgezet tegen de model-Q (bron: Bijlage D).

Hoe de RR-modelresultaten van zowel de voorspelde Q-index als de meetfout (residu) ruimtelijke verdeeld zijn over Nederland is te zien in Figuur 23. De figuur laat ook zien hoe de ruimtelijke verdeling van de trainings- en validatiedata is over Nederland. De figuren geven de indruk dat zowel de voorspelde Q-index als de model-fout geen 'rare' ruimtelijke patronen laat zien.

³ Gebaseerd op de jaarlijkse tijdserie (Januari - December). Wanneer gekozen wordt voor de periode Februari – November neemt de RMSE af naar 0.6

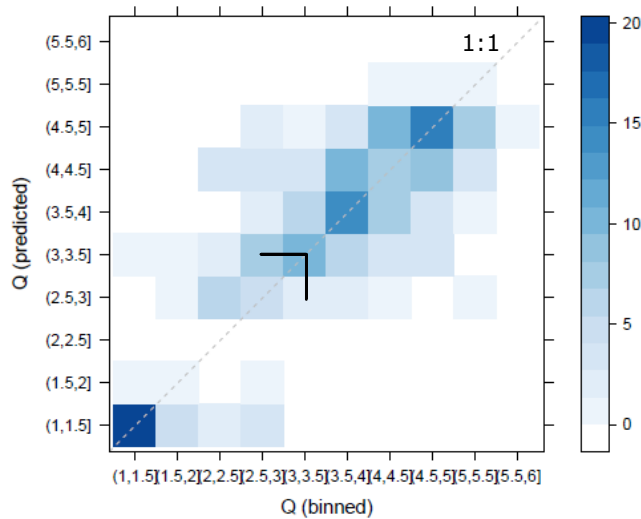


Figuur 23 RR-model resultaten: links de voorspelde Q-index en rechts de residuen voor zowel de training dataset en de validatie dataset (bron: Bijlage D).

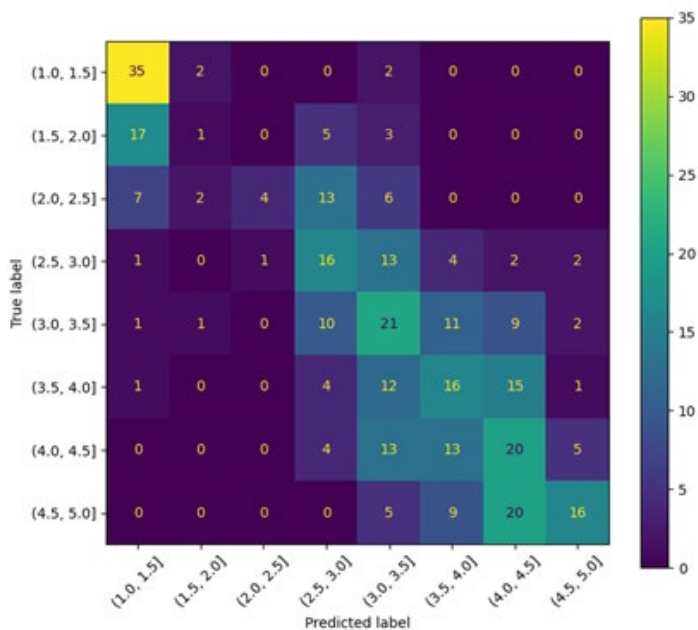
Figuur 24 laat de classificatie resultaten zien voor het RF-model, waarbij op de x-as de gecategoriseerde werkelijke Q-index en op de y-as de gecategoriseerde model Q-index. Elk 'blokje' is dus $0,5Q$ breed en hoog. De blauwtint van ieder blokje geeft aan hoe vaak deze combinatie voorkomt in de validatie dataset. Net als bij het RR-model zijn er niet veel 'punten' rond $Q=2$, en is de spreiding tussen model en werkelijk met name tussen $Q=2,5$ en 5 het grootst. De model nauwkeurigheid is 36,5%, wat betekent dat 36.5% van de percelen correct zijn geclassificeerd. De afgeleide RMSE is 0,73 (Tabel 8), en is nagenoeg gelijk met de regressie resultaten.

De classificatie resultaten van het CNN-model zijn te zien in Figuur 25. NB: In dit geval zijn de x- en y-as verwisseld in vergelijking met Figuur 24. Ook hier is een kleur gebruikt om aan te geven hoe vaak bepaalde combinaties voorkomen. Ondanks het feit dat de grootte van de trainings- en validatie dataset bij het CNN-model iets afwijkt t.o.v. het RR en RF-model, lijken de twee figuren behoorlijk veel op elkaar. De nauwkeurigheid van het model is 37,4% (Tabel 8). Wat niet zichtbaar is in Tabel 8 maar wel in de figuren is dat er weinig punten/combinaties voorkomen rond $Q \sim 2$.

Zowel de classificatie resultaten van het RF als het CNN model laten een model nauwkeurigheid zien van $\sim 37\%$. Dat is wanneer alleen de bins geaccepteerd worden die op de 1:1 lijn liggen in beide figuren. Wanneer de naast gelegen bins ook geaccepteerd worden als goed (zie Figuur 24), resulteert dat in een nauwkeurigheid van 76%, d.w.z. dat 36,5% van de percelen correct zijn geclassificeerd.



Figuur 24 Resultaten RF-classificatie model op basis van de validatie dataset (bron: Bijlage D).



Figuur 25 Resultaten CNN-classificatie model op basis van de validatie dataset (bron: Bijlage E).

3.1.2 Bijdrage RS-markers

Zoals uitgelegd in sectie Dataseselectie en voorverwerking, model strategie zijn er uiteindelijk 113 variabelen (RS-markers) overgebleven die zijn gebruikt voor het trainen van het RR- en RF-model. Een belangrijke vraag is om te weten in welke mate de verschillende RS-variabelen bijdragen aan de voorspelling van het algoritme. D.w.z. wat is de voorspellende waarde van elke variabele? In Bijlage D staat een uitgebreide analyse. Hieronder volgt een korte samenvatting.

De bijdrage van elke variabele is op 3 manieren onderzocht:

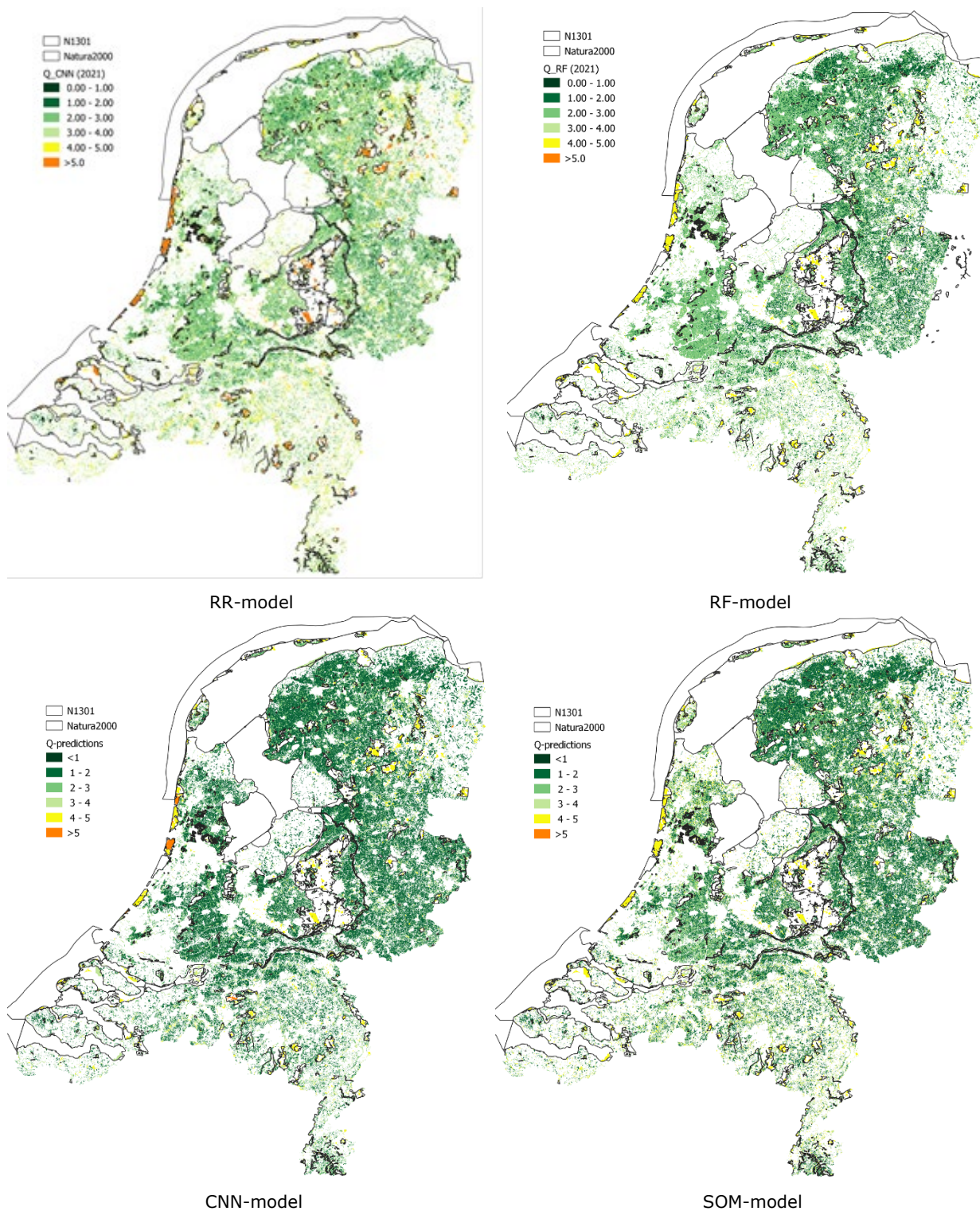
- De eenvoudigste manier om te kijken naar de invloed van iedere variabele is door te kijken naar de correlatie van elke variabele met de Q-index. Hoe hoger de correlatie des te invloedrijker de variabele.
- Een andere manier om elke variabele afzonderlijk te beoordelen is door een model te bouwen dat alleen die variabele gebruikt, en vervolgens de kwaliteit van het model te beoordelen. Dat geeft enige informatie, maar negeert de complementariteit van variabelen (d.w.z. twee variabelen die beide apart voorspellend lijken, laten eenmaal samen geen verbetering zien t.o.v. de individuele voorspellingen, of kan andere informatie toevoegen en in dat geval tot een veel beter model leiden).

-
- Laatstgenoemde aanpak kan ook in omgekeerde vorm worden toegepast: d.w.z. neem het gehele model en kijk vervolgens naar de impact als een variabele wordt weg te laten, en beoordeel de verslechtering van het model. Als die verslechtering groot is, wordt de variabele belangrijk genoemd. Echter als de verslechtering klein is betekent dat niet het tegenovergestelde: als een variabele een ogenschijnlijk klein belang heeft, dan kan dit ook te wijten zijn aan de aanwezigheid van een andere, gecorreleerde variabele, die de taak van de geëlimineerde variabele “overneemt”.
 - Uiteindelijk is de analyse niet per variabele gedaan maar op basis van groepen bestaande uit meerdere variabelen, zgn. clusters, waarbij gebruik is gemaakt van hiërarchische clustering. Dat resulteerde in 32 clusters.

De vraag ‘welke variabelen (clusters) belangrijk zijn’ bleek moeilijk te beantwoorden vanwege de hoge correlaties tussen de variabelen en het grote aantal redelijk grote correlaties. In alle gevallen bleek dat het weglaten van één cluster het model niet veel verslechterde, en dat modellen gebaseerd op slechts één cluster behoorlijk slechter presteerden dan het volledige model. Dit impliceert dat de informatie vrij wijdverspreid is over alle variabelen in de dataset (zie Bijlage D).

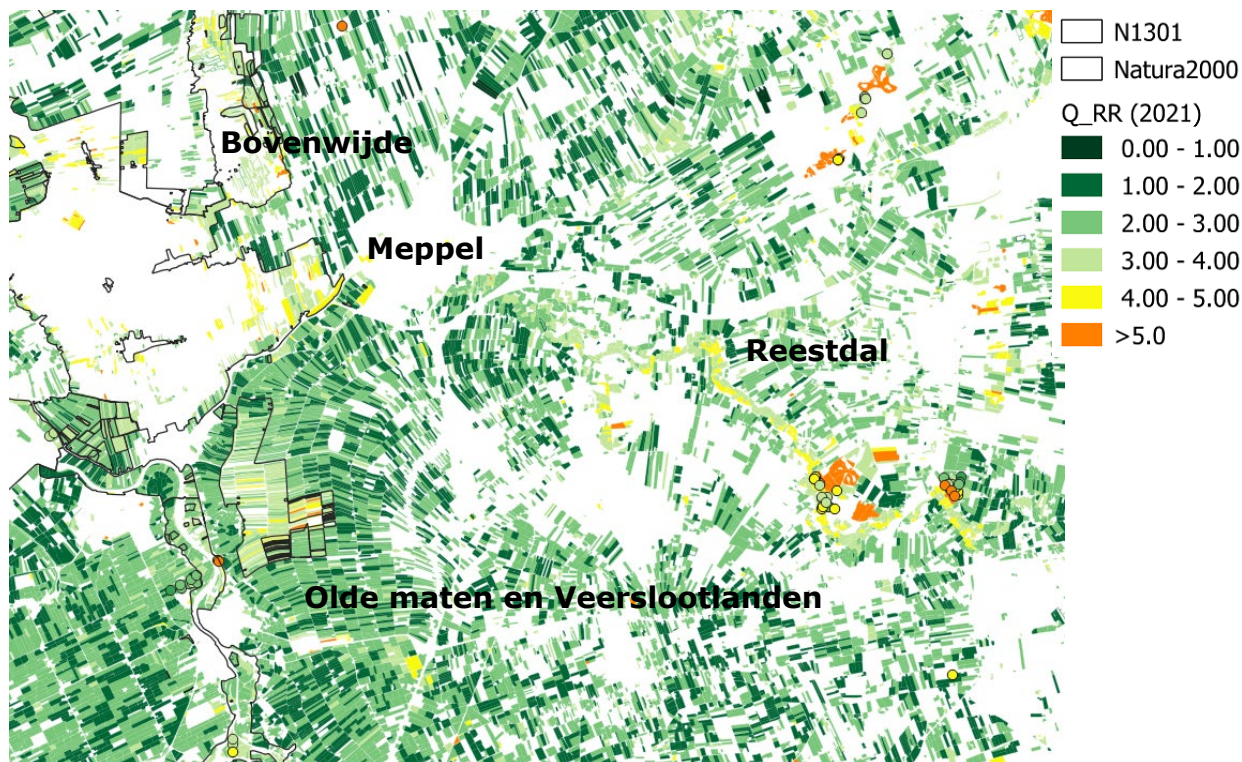
3.1.3 Resultaten voor heel NL en opeenvolgende jaren

Hoe de modellen (RR, RF, SOM en CNN) presteren voor heel Nederland en voor opeenvolgende jaren is te zien in Figuur 26 (2021) en Figuur 29 (2022), in dit geval de regressie resultaten voor alle permanente/natuurlijke grasland percelen (hoofdfunctie landbouw/natuur) en natuurterreinen (volgens de BRP). De weidevogel rust- en Natura2000-gebieden zijn aangegeven met zwarte polygonen. Alle drie de modellen tonen voor deze gebieden een hoge Q-index (> 4) vergeleken met de rest van Nederland. Het Ridge Regressie (RR) model en het Random Forest (RF) model lijken redelijk overeen te komen qua ruimtelijke patronen, al laat het RR-model wat hogere Q-index in natuurgebieden. Het RF-model laat geen gebieden zien met een $Q \geq 5$ (oranje). Verder geeft het RR-model wat lagere Q's in Friesland en Groningen (donkerder groen). Het CNN-model daarentegen lijkt consistent hogere Q-indices te voorspellen dan het RR- en RF-model. NB: Het meest oostelijke deel van Nederland ontbreekt in het RR en RF-model doordat een van de RS-markers daar geen data heeft. Het CNN-model maakt gebruik van de ruwe RS-data.

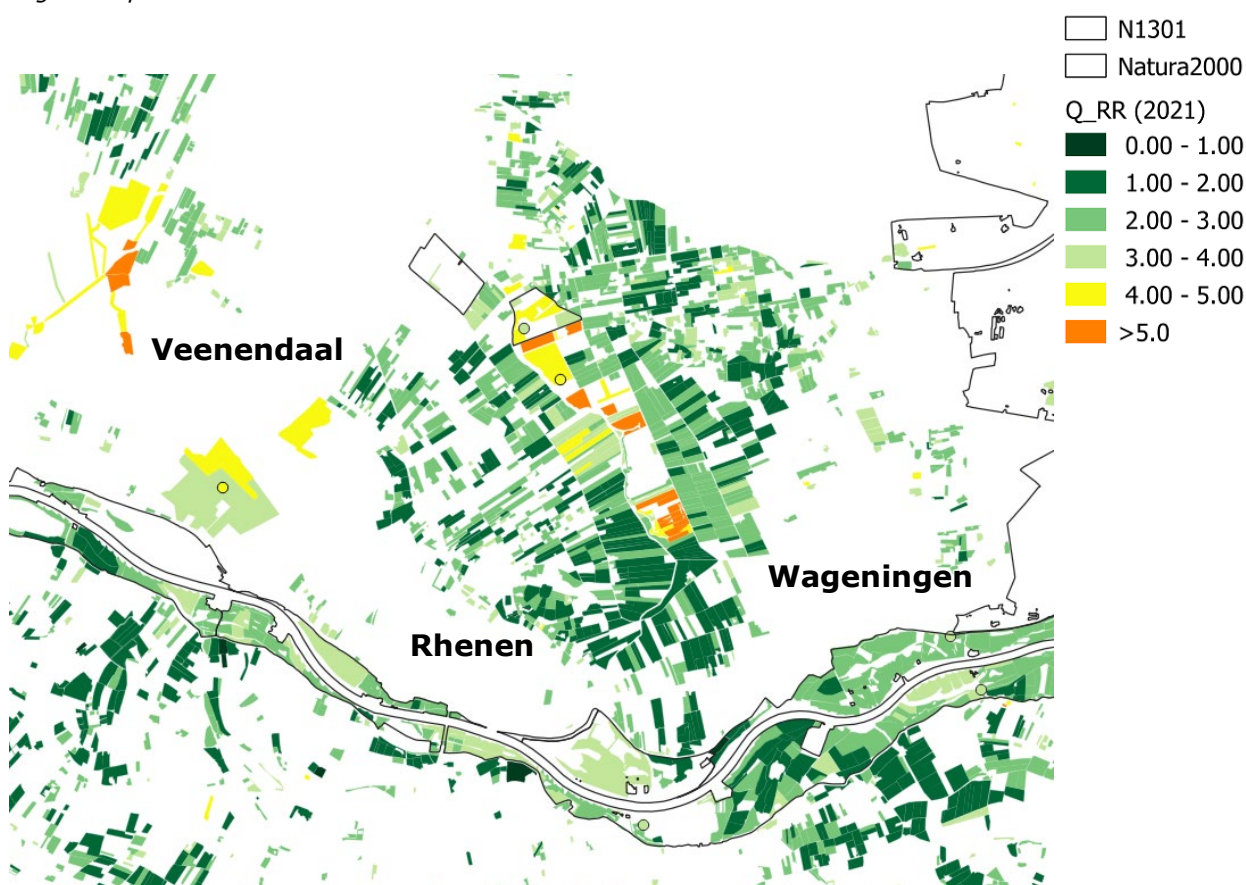


Figuur 26 Q-voorspellingen op basis van RR, RF, CNN en SOM-model voor alle permanente grasland percelen en natuurterreinen in 2021. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.

Figuur 27 toont het *Reestdal* dat ligt op de grens van provincies Overijssel en Drenthe, ten westen van Meppel. Het dal is duidelijk zichtbaar als strook met hogere Q-index (RR-model, 2021) dan de omgeving. Ook de natuurgebieden *Bovenwijde* en *Olde maten en Veerslootlanden* springen eruit in vergelijking met de omgeving. Natuurgebied *De Hooilanden*, tussen Veenendaal, Rhenen en Wageningen is te zien in Figuur 28. Ook dit natuurgebied springt er duidelijk uit.

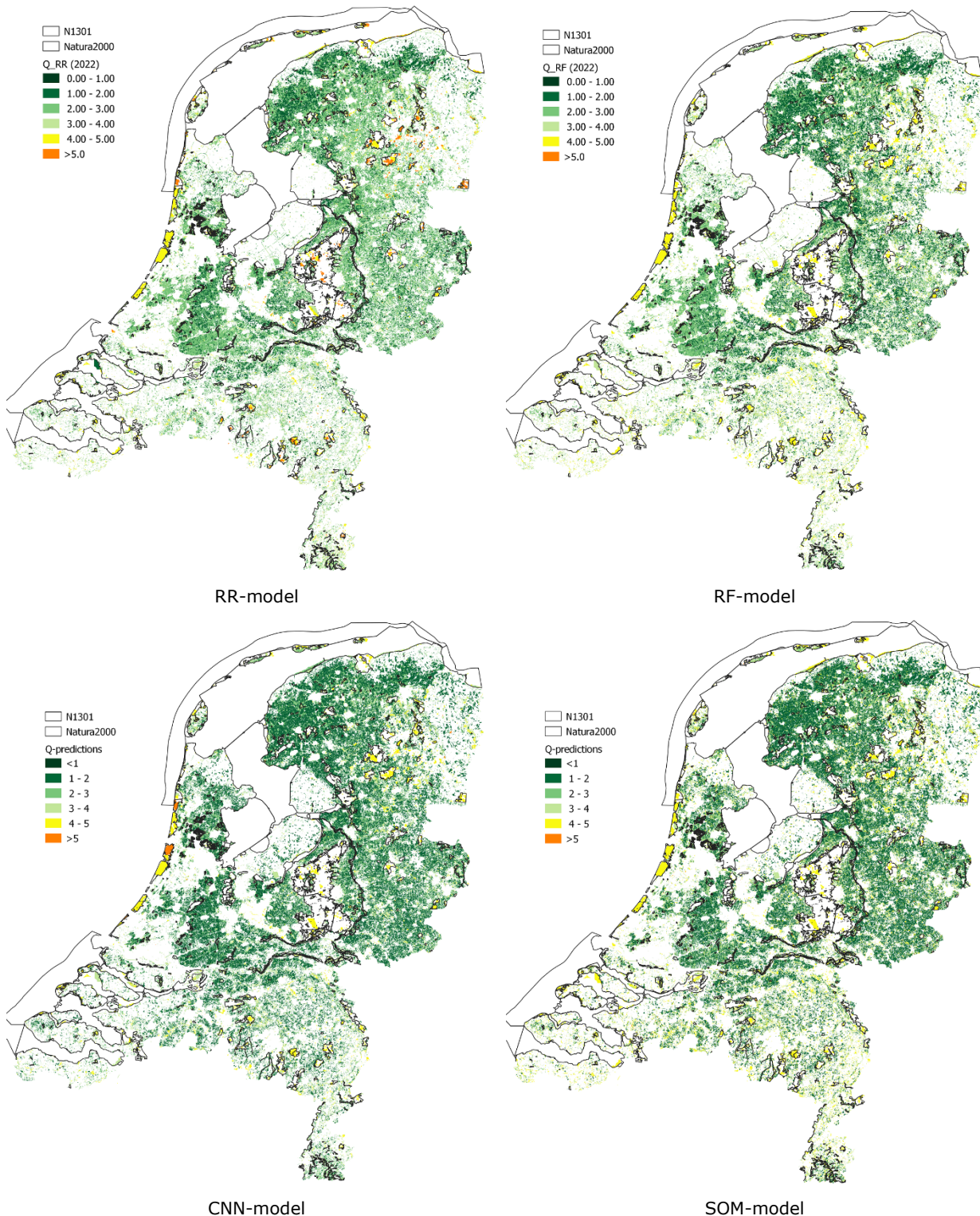


Figuur 27 Q-voorspellingen RR-model voor 2021 in de omgeving van Meppel. De gekleurde punten zijn locaties met vegetatieopnames. De kleuren legenda is van toepassing op zowel de percelen als de vegetatieopnames.



Figuur 28 Q-voorspellingen RR-model voor de hooilanden, tussen Rhenen, Veenendaal en Wageningen (2021). De gekleurde punten zijn locaties met vegetatieopnames. De kleuren legenda is van toepassing op zowel de percelen als de vegetatieopnames.

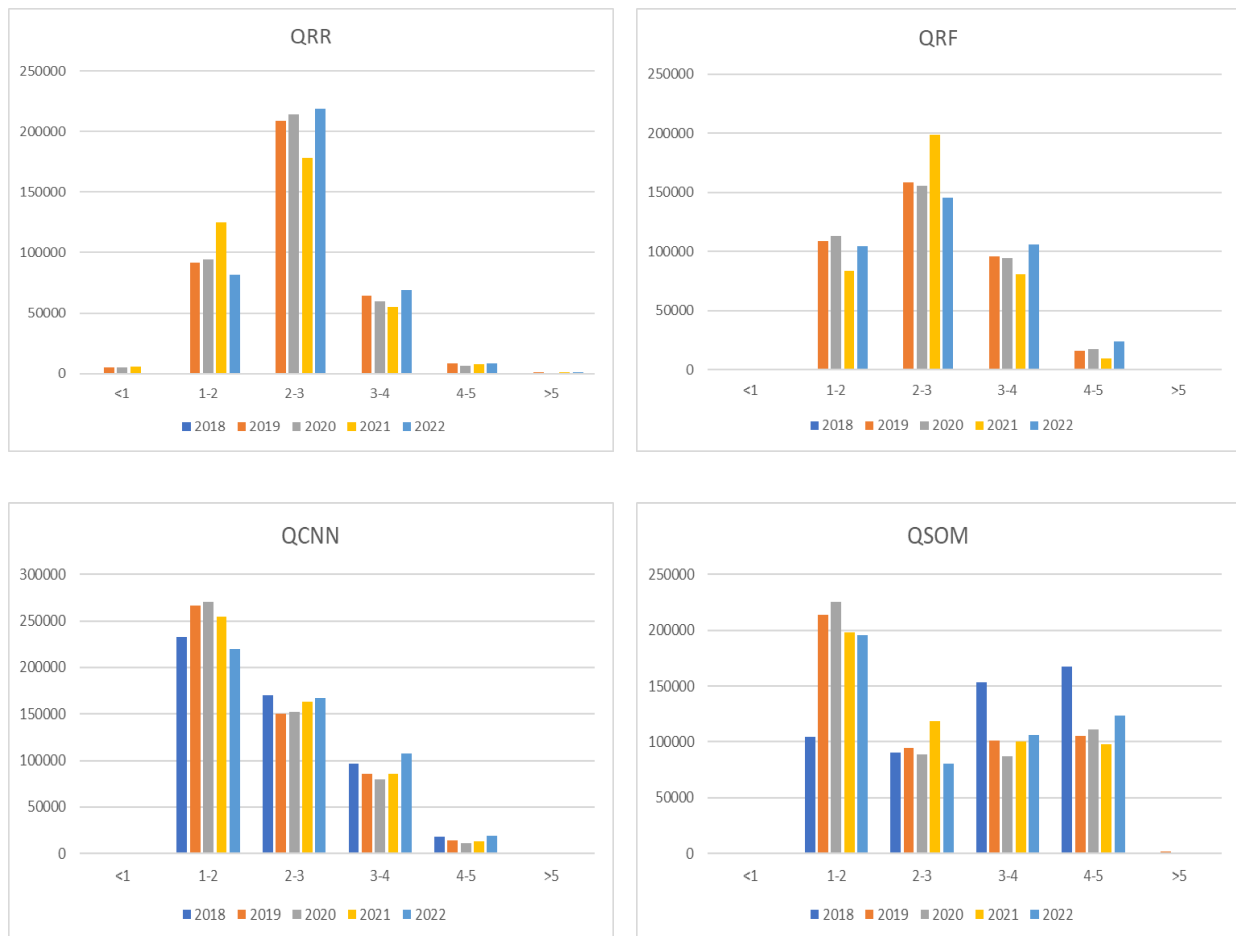
Figuur 29 toont de resultaten voor 2022 voor dezelfde modellen (RR, RF, CNN en SOM).



Figuur 29 Q-voorspellingen op basis van RR, RF, CNN en SOM-model voor alle permanente grasland percelen en natuurterreinen in 2022. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.

De histogrammen (frequentie verdeling) van de regressie resultaten van het RR-, RF-, CNN- en SOM-model voor de jaren 2018 t/m 2022 zijn te zien in de afbeeldingen in Figuur 30. Het eerste wat opvalt is dat zowel het RR- als RF-model de meeste percelen classificeren tussen Q-index 2 en 3. Het SOM- en CNN-model daarentegen classificeren de meeste percelen tussen Q-index 1 en 2. Het SOM- en CNN-model daarentegen classificeren de meeste percelen tussen Q-index 1 en 2. Bij het CNN model neemt het aantal percelen bij

toenemende Q-index geleidelijk af, waarbij de groep met de hoogste Q-index (4-5) het kleinst is. Bij het SOM model daarentegen zijn de groepen met Q-index tussen 2-3, 3-4 en 4-5 ongeveer even groot. Tot slot tonen de histogrammen de variatie tussen de verschillende jaren. Zo zijn de histogrammen van het RR- en RF-model redelijk stabiel door de jaren, maar wijkt 2021 behoorlijk af. Dit is ook te zien bij het SOM-model, waar in dit geval 2018 behoorlijk afwijkt t.o.v. de andere jaren. Het CCN-model is in vergelijking met de andere modellen, behoorlijk stabiel door de jaren heen.



Figuur 30 Histogram Q-voorspellingen op basis van regressie resultaten van het RR-model (links boven), RF-model (rechts boven), CNN-model (links onder) en SOM model (rechts onder) voor de jaren 2018 t/m 2022.

3.1.4 Twee-modellen en twee-klassen aanpak

Twee-modellen: model voor lage ($Q \leq 3$) & model voor hoge ($Q > 3$) kruidenrijkdom

Om te kijken of de model prestaties verbeteren wanneer twee modellen apart worden getraind voor graslanden met lage kruidenrijkdom ($Q\text{-index} \leq 3$) en percelen met hoge kruidenrijkdom ($Q\text{-index} > 3$), is het CNN model apart getraind voor beide groepen. Om de resultaten eerlijk te kunnen beoordelen is niet gekeken naar de gemiddelde modelfout (RMSE) maar ook naar de gemiddelde procentuele fout (MAPE). De resultaten staan in Tabel 9.

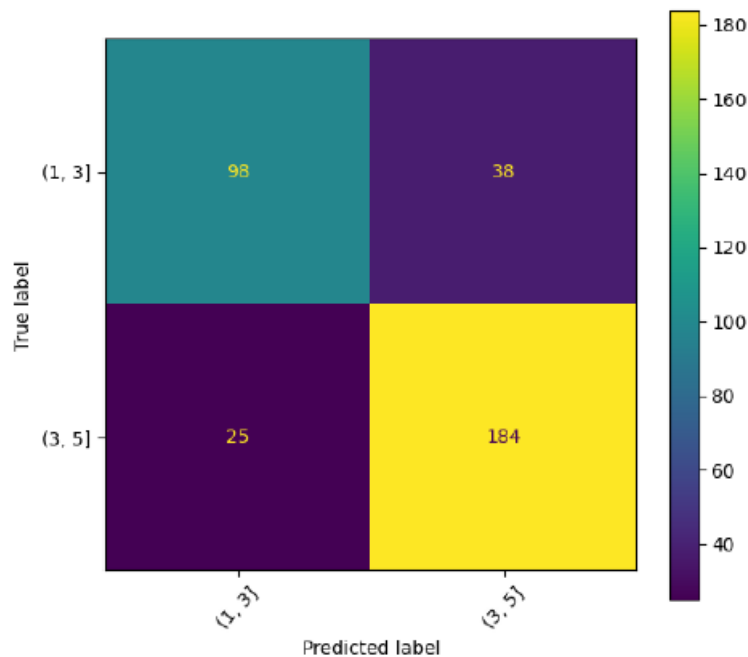
Tabel 9 Model resultaten twee-modellen aanpak (bron: Bijlage E).

Q-range	RMSE	MAPE
1-5	0,6	0,16
1-3	0,37	0,15
>3-5	0,52	0,12

Alhoewel de modelfout (MAPE) iets afneemt, roept het de vraag op of deze hogere prestaties in de praktijk zwaarder wegen dan het hebben van twee modellen, wat betekent dat er op een of andere manier gekozen moet worden welk model per perceel gebruikt moet worden. In plaats van te kiezen voor twee modellen, is het wellicht interessanter om te kiezen voor twee klassen.

Twee-klasse: lage ($Q \leq 3$) en hoge kruidenrijkdom ($Q > 3$)

Bij de classificatie analyse van het RF- en CNN-model is gebruik gemaakt van een bin-grootte van 0,5, d.w.z. de vegetatie opnames zijn gegroepeerd in klasse van $Q=1-1,5$, $1,5-2$, $2-2,5$, etc. (~ 12 klassen). Zowel bij het RF- als het CNN-model resulteerde dat in een model-nauwkeurigheid van 37%. Wanneer het aantal klassen terug wordt gebracht naar slechts twee, namelijk graslanden met een 'lage' kruidenrijkdom (Q -index ≤ 3) en met een 'hoge' kruidenrijkdom (Q -index > 3), dan resulteert dat bij het CNN model in een nauwkeurigheid van 82% (Figuur 31).

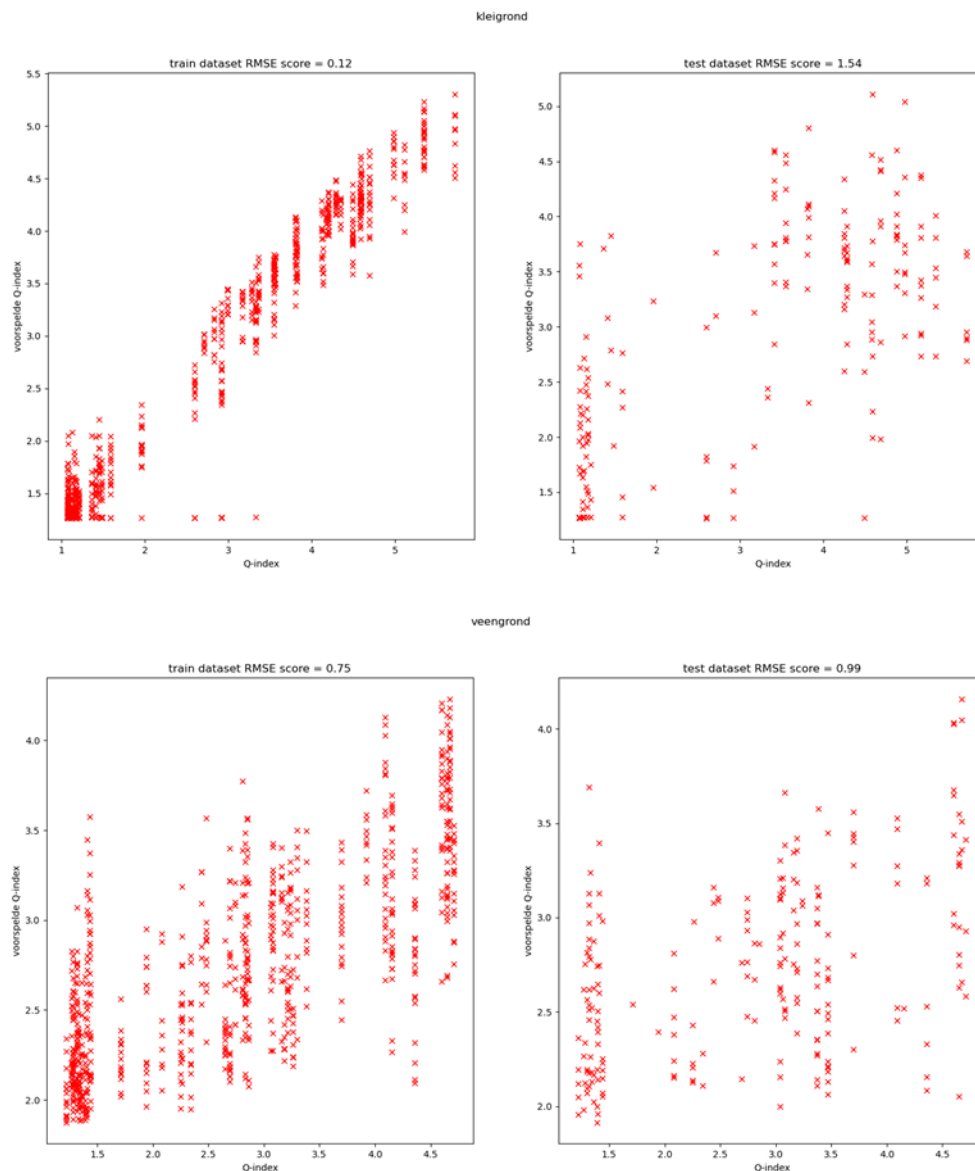


Figuur 31 Resultaten CNN-classificatie model voor slechts twee klasse (bron: Bijlage E).

3.2 Resultaten foto-algoritme

3.2.1 Voorspellen kruidenrijkdom (spoor 2a)

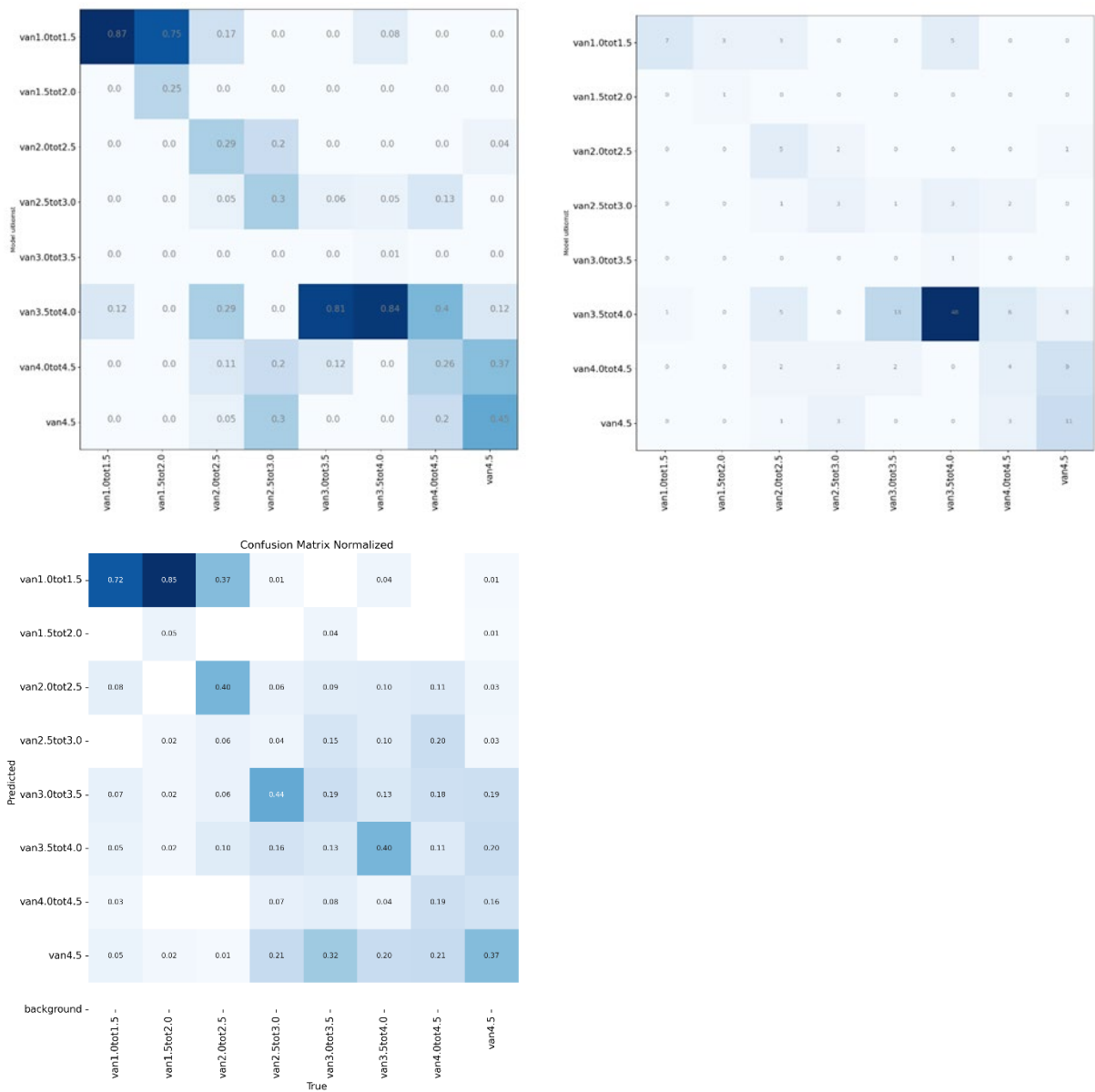
Deze sectie bevat de resultaten van het foto-algoritme waarmee de kruidenrijkdom kan worden geschat (spoor 2a). Het resultaten voor het regressie algoritme zijn geëvalueerd op basis van de RSME tussen de gemeten Q -index (vegetatie opnames) en de voorspelde Q -index door het foto-algoritme, zoals beschreven in Resultaten RS-algoritme. Voor de klei- en veengronden is de RSME bepaald voor de train- en test dataset na parameter optimalisatie (Figuur 42). Omdat ook na optimalisatie de scores op de test dataset abominabel waren, is de RSME voor de validatie dataset niet bepaald. Waar het voor de veengronden gelukt is om in ieder geval op de train dataset een redelijk model te maken, is dat voor de zandgronden niet gelukt.



Figuur 32 Regressie fotomodel resultaten voor klei-(boven) en veengronden (onder). Links op basis van de training set, rechts op basis van de validatie set.

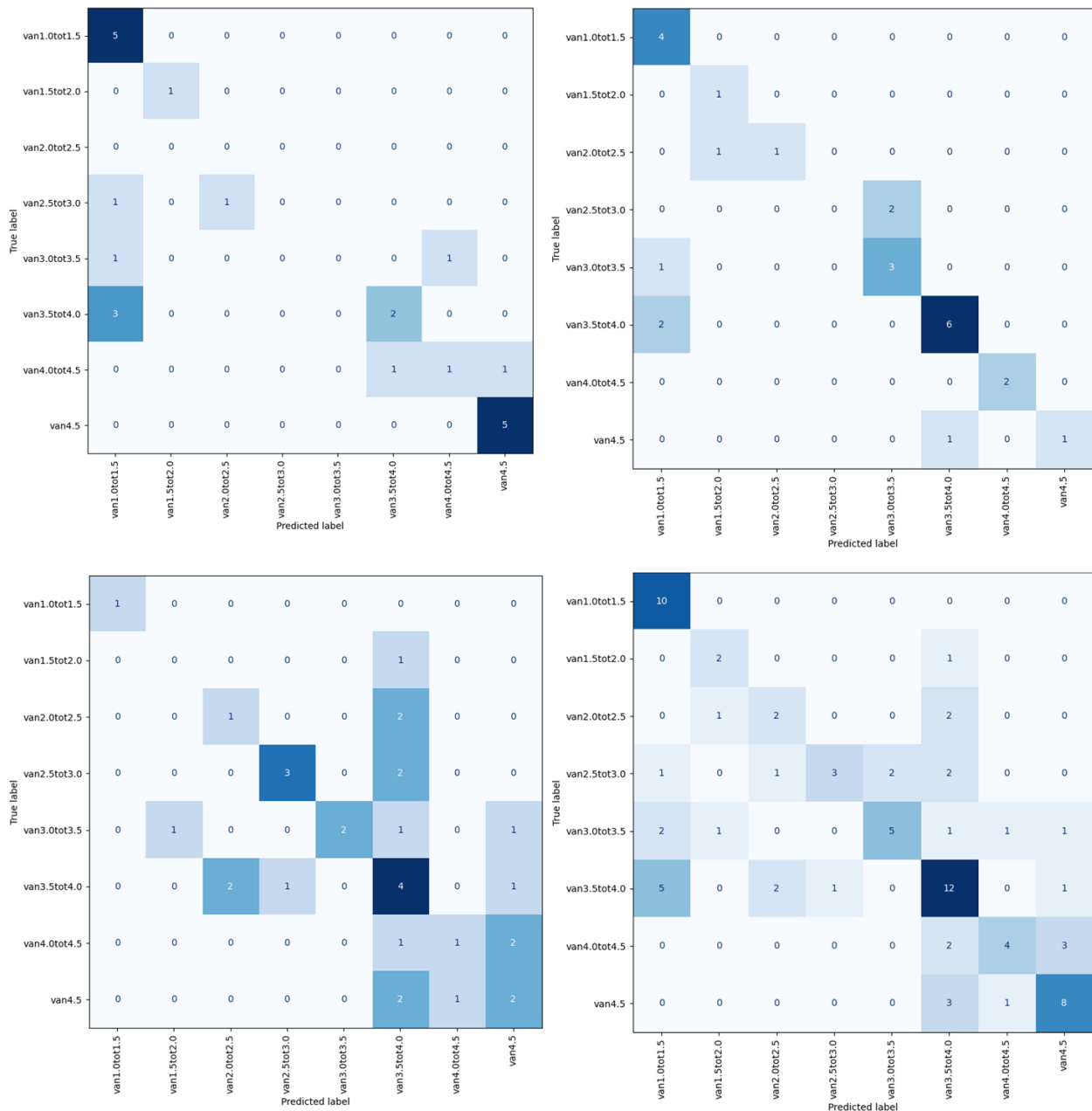
Voor het eerste classificatie algoritme, waarbij de verschillende graslanden in classes van 0,5- Q-index zijn ingedeeld, zijn de resultaten samengevat in zogenaamde confusion-matrices in Figuur 33. Omdat hier op de test dataset de resultaten redelijk waren, is het model ook geëvalueerd op basis van de validatie dataset. Daarvan werden er 52% correct geclassificeerd. Het aantal beelden in de validatie verzameling is beperkt en daarom is ook het resultaat op de test verzameling gegeven. Hiervan werd in totaal 35% van de beelden de Q-index correct geclassificeerd, als gekeken wordt naar hoeveel procent er maximaal 1 klasse naast zit, dus bijvoorbeeld 1,5 en 2 in plaats van tussen 1 en 1,5 dan zijn toch 57% van de beelden de Q-index goed geclassificeerd. Het is wel zo dat er per perceel minstens 2 foto's verzameld waren.

NB: De confusion-matrices zijn gebaseerd op de foto's, waarbij het kan voorkomen dat er meerdere foto's van een perceel zijn. En dat voor deze percelen de voorspellingen van het model niet voor alle foto's gelijk was.



Figuur 33 Confusion-matrices, genormaliseerd op de validatie verzameling (linksboven), de absolute waarden van de validatie verzameling (rechtsboven) en de genormaliseerde matrix op de verrijkte test verzameling (linksonder).

Voor het tweede classificatie algoritme, waarbij er voor klei-, veen- en zandgrond apart een model is getraind, en vervolgens per perceel de meest voorkomende voorspelling is geselecteerd staan de confusion-matrices in Figuur 34. Op 134 van de 347 percelen waren de voorspellingen van de foto's niet constant. De resultaten van alle modellen staan samengevat in Tabel 10.



Figuur 34 Confusion matrices voor de modellen waarbij de foto's gesorteerd zijn op Q-index. Linksboven kleigrond, rechtsboven veengrond, linksonder zandgrond en rechtsonder alles gecombineerd.

Tabel 10 Model resultaten foto algoritmes (spoor 2a).

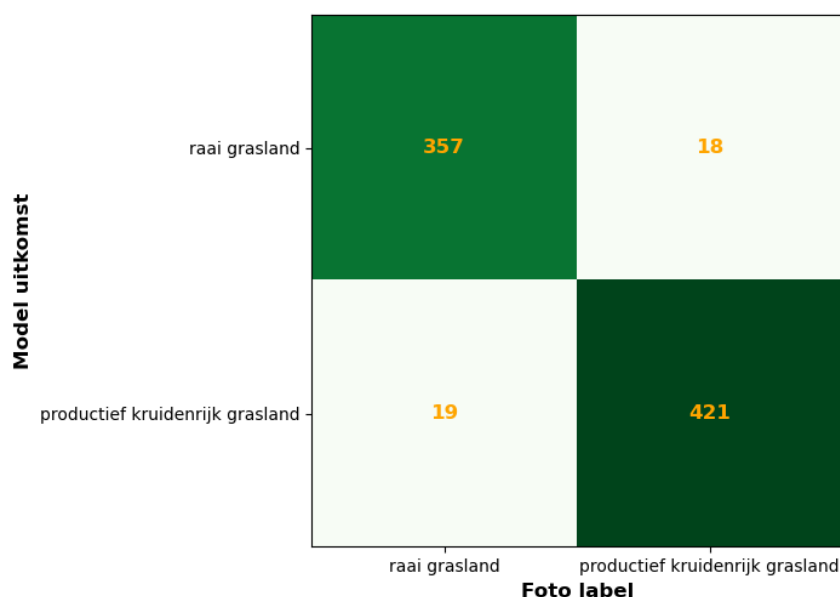
Model	RSME	Nauwkeurigheid
<i>Regressie</i>		
Klei	1,54 (op basis van test verzameling)	-
Veen	0,99 (op basis van test verzameling)	-
<i>Classificatie q-index</i>		
	0,89 (op basis van validatie verzameling)	52% (op basis van validatie verzameling)
	-	35% (op basis van test verzameling)
<i>Classificatie q-index</i>		
Klei	0,952 (op basis van validatie verzameling)	67% (op basis van validatie verzameling)
Veen	0,51 (op basis van validatie verzameling)	66% (op basis van validatie verzameling)

Zand	0,75 (op basis van validatie verzameling)	37% (op basis van validatie verzameling)
------	---	--

<i>Classificatie Shannon-index</i>		
<i>Klei</i>		
Veen	0,95 (op basis van validatie verzameling)	61% (op basis van validatie verzameling)
Zand	0,51 (op basis van validatie verzameling)	60% (op basis van validatie verzameling)
	0,75 (op basis van validatie verzameling)	46% (op basis van validatie verzameling)

3.2.2 Herkennen PKG-percelen (spoor 2b)

De resultaten van het foto-algoritme voor het onderscheiden van PKG-percelen van regulier Engels raaigras zijn te zien in Figuur 35 in de vorm van een confusion-matrix. Van de test verzameling wordt 95% van de beelden goed geïdentificeerd. In de figuur kan gezien worden dat 778 (357+421) van de in totaal 815 beelden correct geïdentificeerd worden. Slechts een kleine groep foto's zijn verkeerd geïdentificeerd (37 (18+19) foto's). In Figuur 36 een voorbeeld van geannoteerde foto's en hun model uitkomsten.



Figuur 35 De confusion-matrix waarin de prestatie van het algoritme weer gegeven wordt. Op de diagonaal staan het aantal correct geïdentificeerde afbeeldingen. Er zijn 19 beelden van het raaigrasland incorrect geïdentificeerd als productief kruidenrijk grasland en 18 beelden van productief kruidenrijk grasland die onterecht geïdentificeerd zijn als raaigraslanden.

Labels



Model uitkomsten



Figuur 36 Voorbeeld van 4 foto's met labels en 4 uitkomsten van het model. De foto rechtsboven wordt incorrect geclassificeerd.

Er is ook een validatie ronde gedaan op de originele detailfoto's uit de validatie verzameling. Deze foto's zijn qua preprocessing alleen geschaald naar 200 x 200 pixels. Hierdoor kan per perceel aangeven worden hoe goed de prestatie van het algoritme is. Van de 62 percelen met raaigras in de validatie verzameling worden er 58 (94%) correct als raaigraslanden herkend. Van de 45 percelen die ingezaaid waren met klaver of productieve voederkruiden worden er 43 (96%) percelen als zodanig herkend.

De vraag is natuurlijk of dat goed genoeg is. Met de prestaties van dit algoritme, ervan uitgaande dat er ongeveer evenveel raaigraslanden als ingezaaide graslanden zijn, zouden bijvoorbeeld 3 foto's verspreid over het perceel genomen kunnen nemen. Als dan 2 van de 3 foto's eenzelfde uitkomst hebben, dan is het 99% zeker dat dit de juiste uitkomst is.

3.3 Remote Sensing - versus foto algoritme

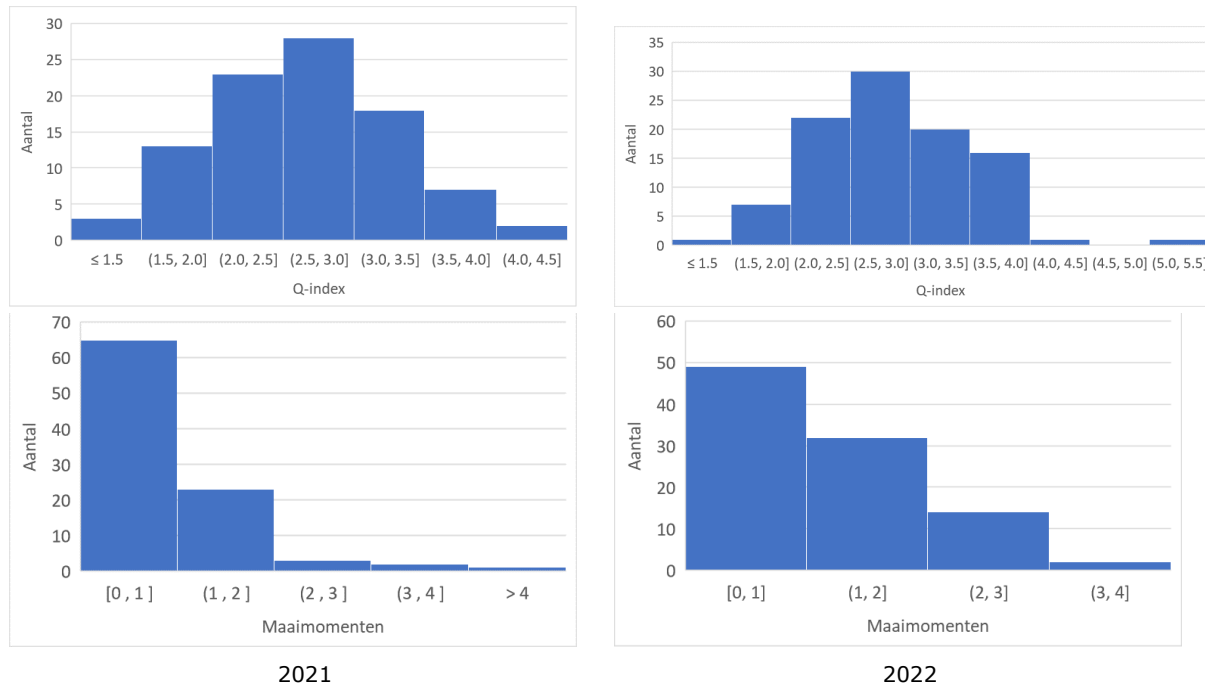
3.3.1 Productief kruidenrijke graslanden (PKG)

Zoals eerder is genoemd in hoofdstuk Methodiek, zijn ingezaaide productief kruidenrijke graslanden (PKG) een bijzondere groep omdat de soortensamenstelling wordt bepaald door zadenmengsels en niet door het creëren van ideale groeiomstandigheden voor allerlei kruiden. Mede hierdoor is het voor een RS-algoritme moeilijk om de soortenrijkdom te voorspellen. Het ontwikkelde RS-algoritme zal hoe dan ook op basis van de RS-input een voorspelling geven over de kruidenrijkdom. Echter, de vraag is wat stelt deze Q-index voor? Zoals al eerder is aangegeven, is het vermoeden dat met name het hoge intensiteit landgebruik van regulier Engels raaigras alsmede van de PKG-percelen doorwerkt in de RS-data/markers, waardoor het vermoeden is dat het RS-algoritme voor beide type percelen een lage kruidenrijkdom zal voorspellen (zoals is geschetst in Figuur 2).

De resultaten van het foto-algoritme (spoor 2b) in sectie 3.2.2 laten zien dat op basis van foto's goed onderscheid kan worden gemaakt tussen regulier Engels raaigras- en PKG-percelen. De meest logische verklaring hiervoor is dat de foto's voldoende detailinformatie bevatten om een goede uitspraak te doen, in tegenstelling tot de RS-data.

Zowel via Pure Graze als via de fotocampagne zijn foto's verzameld (met locatie gegevens) van PKG-percelen. De voorspelde Q-index (op basis van het RR-model) voor deze ~100 percelen is te zien in de histogrammen voor 2021 (linksboven) en 2022 (rechtsboven) in Figuur 37. De range van voorspelde Q is behoorlijk groot, en ligt tussen de 1,5 en 4,25. De gemiddelde Q-index in 2021 is ~2,6 (volgens het RF-model: 2,7) en in 2022 ~2,9 (RF-model: 3,1).

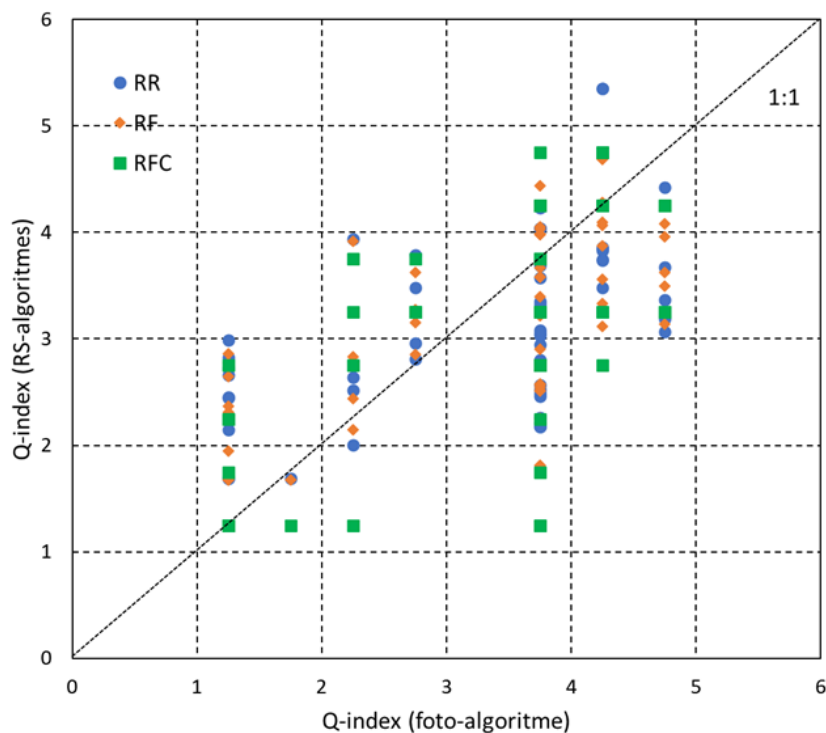
Van de PKG-percelen is niet bekend wanneer deze zijn ingezaaid met het kruidenmengsel en dus hoe oud het zaaibed is. Pure Graze geeft aan dat mengsels gemiddeld 3 tot 4 jaar meegaan, voordat moet worden doorgezaaid. Gedurende de jaren zal de samenstelling, door de omstandigheden en/of het management veranderen. Dit laatste zou een verklaring kunnen zijn voor de grote range in voorspelde Q-index. Ook is gekeken naar de intensiteit van het landgebruik van deze percelen, middels het aantal maaimomenten in 2021 en 2022 (Figuur 37, links- en rechtsonder). Interessant is dat de meeste van deze PKG-percelen helemaal niet, of slechts 1 keer worden gemaaid. Echter, het kan ook dat deze percelen voornamelijk worden beweide in plaats van gemaaid, maar dat is niet bekend.



Figuur 37 Frequentieverdeling van voorspelde Q-index (boven) volgens het RR-model voor 2021 (links) en 2022 (rechts) en bijbehorende maaimomenten (onder).

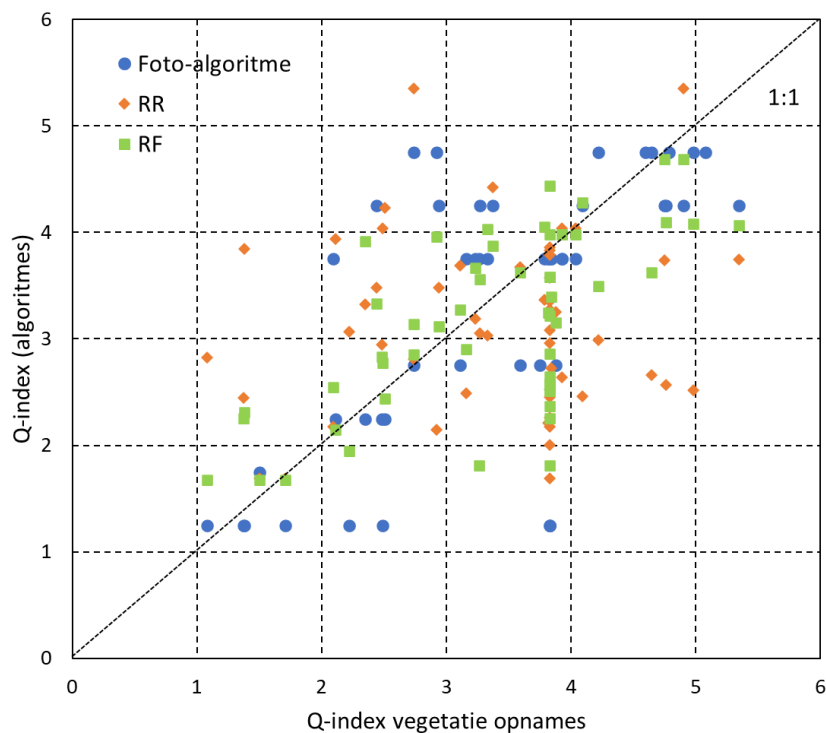
3.3.2 Kruidenrijkdom: RS- versus foto-algoritme

Hoe goed het RS- (spoor 1) en foto-algoritme (spoor 2a) overeenkomen qua voorspellen van de kruidenrijkdom (Q-index) voor dezelfde percelen is te zien in Figuur 38. In dit geval zijn ~55 percelen geselecteerd waarvoor de Q-index is voorspeld met behulp van de RS-algoritmes (RR, RF en RFC) en het foto-algoritme (spoor 2a). NB: Van deze percelen zijn meerdere foto's beschikbaar en dus ook meerdere schattingen van de kruidenrijkdom. In dit geval is er per perceel maar één foto gekozen met bijbehorende voorspelling. In Figuur 38 zijn de voorspelde Q-waarden van de RS-algoritmes geplot tegen het foto-algoritme.



Figuur 38 Voorspelde Q-index RS-algoritme (RR-, RF- en RF-classificatie model) versus Q-index foto-algoritme.

In Figuur 39 zijn beide algoritmes geplot tegen de Q-index op basis van de vegetatie opnames. De RMSE voor het RS-algoritme en foto-algoritme zijn 0,89 (RR), 0,81 (RF) en 0,89 (foto) respectievelijk. Zowel de scatter-plots als de RMSE-waarden tonen aan dat beide sporen redelijk gelijkwaardig presteren, waarbij het RF-model het net iets beter lijkt te doen. Ook is in Figuur 39 zichtbaar dat het aantal percelen met een lage Q laag is ten opzichte van percelen met een Q tussen 2 en 4, waarbij met name percelen rond $Q \sim 3,75$ oververtegenwoordigd zijn.



Figuur 39 Voorspelde Q-index RS-algoritme (RR-, RF) en foto-algoritme versus vegetatie opnames.

3.4 Stakeholder analyse workshop

Op 26 september 2024 is een stakeholder analyse workshop gehouden met als doel om de RS- en foto-algoritmes te evalueren samen met consortium partners en sponsors. Workshop deelnemers werden gevraagd om input op:

1. Stellingen formuleren over welk doel de modellen dienen.
2. Evalueren van Q-index overzicht kaarten van Nederland gemaakt met de RS-algoritmes.
3. Vergelijken van de Q_{CNN} en Q_{SOM} resultaten.
4. Inventarisatie van behoefte met betrekking tot het gebruik van de foto-app.

Ter voorbereiding op de workshop is aan de deelnemers toegang verleend tot een online kaarten viewer met de resultaten van de RS-modellen met daarbij de vraag om deze uit te proberen en alvast na te denken over de algemene impressie van de kaarten en de voorspellingen van de kruidenrijkdom van de gebruikte modellen. De vragen onder punt 2 en 3 zijn van te voren gedeeld. Deelnemers van BoerenNatuur, Vlinderstichting, Rabobank, FrieslandCampina, Staatsbosbeheer, Vogelbescherming Nederland, Wij.Land, Wageningen Research, Provincie Drenthe (Prolander), Provincie Noord-Holland, en RVO waren aanwezig bij de workshop. Deelnemers van Provincie Utrecht hebben hun bijdrage per email aangeleverd.

De groep werd gevraagd om stellingen te formuleren rondom de vraagstelling "Welk doel dienen deze kruidenrijk grasland modellen voor jou?". Het idee was om zoveel mogelijk perspectieven op te halen en als groep de verhouding tot de stelling te onderzoeken:

Stelling geformuleerd door de stakeholders	Verhouding stakeholders tot de stelling
<i>Op basis van deze data en kaartlaag kunnen we boeren belonen voor hun inzet op kruidenrijk grasland.</i>	Zowel oneens als eens, met als groep overwegend neutrale houding
<i>Als boeren en natuurbeheerders meer samenwerken is het goed voor de biodiversiteit.</i>	Iedereen mee eens
<i>Pragmatisme is belangrijker dan academische perfectie.</i>	Overwegend mee eens

<i>Het model is nooit af en je moet er energie in steken om het model draaiende te houden.</i>	Overwegend mee eens
<i>De toepassing van je model moet je aanpassen naar hoe goed je model werkt.</i>	Veel spreiding van oneens tot eens
<i>De doelgroep voor deze tool is al voldoende in beeld op dit moment.</i>	Veel spreiding van oneens tot eens
<i>Deze kaarten zijn op zich niet voldoende om een beloningssysteem te ontwikkelen voor boeren.</i>	Overwegend mee eens
<i>De kleine stapjes in de ontwikkeling van kruidenrijkdom zijn niet uit deze kaarten te halen.</i>	Overwegend mee eens

De Q-index overzicht kaarten van Nederlandse graslandpercelen zijn geëvalueerd aan de hand van drie open vragen, waarvan stakeholders zijn gevraagd om hun reacties:

Vraag	Reacties
Wat wordt <u>gewaardeerd</u> aan de kaarten?	- Op gebiedsniveau (natuur, rivieren, Natura2000, weidevogelreservaten, landbouw) en perceelniveau geeft de kaart inzicht in de mate van kruidenrijkdom gekoppeld aan satellietkaart; groot deel van Nederland is gerepresenteerd; redelijk gedetailleerd. - Actueel inzicht omgevingen waar hogere diversiteit aanwezig is. - Grofweg lijkt goed overeen te komen met kennis uit het veld en S2_REP. - Makkelijk in gebruik. - Heldere legenda. - Meerdere jaren inzichtelijk. - Centraal product, centraal interface. - Met name zandlandschappen lijken in de grote lijn te kloppen. - Onderliggende analyse van remote sensing data wordt als nuttig ervaren en het ook wel inzetbaar kan zijn. - Ik denk het vooral inbreng kan leveren op gebieden dit al door veldonderzoek als fase 3 of hoger (Schipper <i>et al.</i>, 2015) beoordeeld zijn.
Waar zijn we <u>kritisch</u> op?	- Legenda "Natura2000_kruidenrijk grasland" moet het alleen "Natura2000" zijn. - Veel variatie over de jaren, Percelen missen bij sommige jaren/algoritmes. - Kwaliteit van de data lastig in te schatten omdat de data is niet in te zien. - Zoekfunctie op locatie (GPS/perceel) werkt niet/is niet duidelijk hoe het werkt; zou fijn zijn om te kunnen filteren op perceel- en bedrijfsniveau. - Percelen niet altijd representatief; meerdere landbouw activiteiten in een geplot perceel; niet grasland (bijv. heide) krijgt soms ook een Q-index, niet consistent wat wel en niet wordt weergegeven; Onderscheid klaverpercelen, productief kruidenrijk grasland en extensief kruidenrijk grasland ontbreekt. - Onzekerheidsmarge/nauwkeurigheid zou op de kaarten zichtbaar moeten zijn. - Gradiënten binnen percelen wordt genegeerd. - Q-ranges: meer gradiënt weergegeven in plaats van een scherpe overgang in kleur groen, geel, oranje; ook veel leeg zonder kleur. - Link Q-index klassen, link met kruidenrijkdom en ANLb waardigheid. - Ontwikkeling van kruidenrijkdom in de tijd op dit moment onmogelijk.
Wat willen we er <u>minimaal</u> uithalen?	- Veldobservaties zou heel belangrijk moeten blijven: - deze tool als aanvulling om te kijken waar ontwikkelkansen liggen, te voorspellen over de tijd, en ontwikkeling te volgen - het zou niet de bedoeling (moeten) zijn om de kaarten te gebruiken voor directe beloning/afrekening; er komt altijd controle op het land - Gebruik als stimulerings- en ontwikkeling tool, bijv.: - op regio niveau in te schatten waar kruidenrijk grasland te stimuleren - ontwikkeling op perceel- en gebiedsniveau in de tijd - weergave van 5 jaar gemiddelde - trend functionaliteit per perceel - keuzehulp van geschikte pakket: bijvoorbeeld kruidenrijk ingezaaid vs kruidenrijk extensief - Gebruik als monitoring tool, bijv.: - koppeling met weidevogel monitoring - indicaties voor aandeel en kwaliteit kuikenland

- monitoren of beweiden plaatsvindt
- koppeling aan BKN [Basiskwaliteit Natuur, red.]
- effect van de ANLB pakketten
- 4. Functionaliteit om te filteren op, bijv.:
 - grasland score/fase; vooral verschil tussen Q-index (1-2) en (2-3)
 - type grasland: extensief kruidenrijk, intensief kruidenrijk
 - bij voorkeur ook tussen de verschillende types extensief: grassenmix, grassenmix-plus, en gras-kruidenmix
- 5. Kaart toegankelijk/toevoegen aan anderen, bijv.:
 - uitgebreider ID zoekfunctie voor gebruikers, match met BRP
 - open source of API om te kunnen integreren met andere tools
 - RS kaarten beschikbaar maken als shp-file
 - beheer-op-maat model van ±20 kaartlagen
- 6. Koppeling koppeling databronnen (RS, foto, veldonderzoek):
 - foto-app en RS-model resultaten weergeven
 - foto's (indien beschikbaar) ook te zien zijn in de viewer
 - veldonderzoeksresultaten (indien beschikbaar) toevoegen
- 7. Data weergeven:
 - continue Q-waarde weergeven
 - indicatie van de kwaliteit en nauwkeurigheid van de uitkomsten
 - fout/onzekerheidsmarge duidelijk maken
 - afspraken op sectorniveau over de nauwkeurigheid
- 8. Kaarten weergave, bijv.:
 - gebied kaderen
 - provinciaalijnen zichtbaar in bovenste kaartlaag
- 9. Details over de onderliggende data weergeven, bijv.:
 - is dat een combinatie van RS-data, kleur, beheer (aantal keer maaien), veldobservaties
 - maaimomenten en S2Rep weergeven
 - eo-informatie van modellen
 - discretere overgang van kleuren om genuanceerder beeld te geven
- 10. Data eigendom en gebruiksrecht waarborgen:
 - voor ontsluiting van de database is het van belang dat dit aansluit of onderdeel is van de database SCAN-ICT en UMDL-tool in Utrecht van de agrarische collectieven
 - de data blijft dan bij de boer (via SCAN-ICT) en is zichtbaar voor de agrariër
 - agrarische collectieven kunnen de data middels schouw (check in het veld) verifiëren

Vervolgens is dieper ingegaan op het vergelijken van de Q-index resultaten van het SOM-model met die van het CNN-model, en de ervaringen zijn geëvalueerd:

Opdracht	Uitkomsten
Pak een gebied wat je goed kent. Kies in dat gebied een aantal percelen. Kijkend naar de resultaten van de twee modellen, wat <u>waarden</u> we?	1. Informatief als handvat om een indruk te krijgen, bijv. ○ percelen van dezelfde veehouder hadden ~dezelfde Q-index ○ natuurgebieden komen er echt aardig uit ○ een gebied waar veel vogelaars komen laat een gevarieerd beeld aan Q-index zien, dat zou kunnen kloppen
Pak een gebied wat je goed kent. Kies in dat gebied een aantal percelen. Kijkend naar de resultaten waar zijn we <u>kritisch</u> op?	2. Ik zou er niet afhankelijk van willen zijn als enige methode; veldonderzoek blijft belangrijk. 3. Het lijkt dat de resultaten uit beide modellen niet altijd overeenkomen met de verwachting, bijv.: ○ lagere Q-index in een weidevogelgebied waar juist een hogere Q-index verwacht werd ○ boomkwekerij met twee paardenweides op één perceel met Q-index van 2.8 lijkt erg aan de hoge kant ○ perceel met bekend hoge Q-index heeft niet elk jaar een zichtbaar hoge Q ○ vijftien percelen beoordeeld, waarvan drie geen grasland. Van de elf grasland percelen lijkt bij meeste de Q-index hoger of lager ingeschat dan verwacht/bekent

4. Verschil tussen QSOM en QCNN niet helemaal duidelijk.
 - o QCNN stabiel, maar ook minder betrouwbaar?
 - o QSOM lijkt niet erg stabiel, maar ook niet erg slecht
 - o voor het ene perceel lijkt in het ene jaar CNN beter te kloppen en voor andere percelen en jaren SOM
 - o grosso modo het idee dat voor in de bekeken omgeving CNN te rooskleurig is en SOM een realistischer beeld geeft
 - o ik zou nog niet op deze modellen af durven gaan; allebei hebben voorbeelden gezien waar Engels raaigras als soortenrijk uit het model komt, terwijl dat niet klopt
5. De (jaarlijkse) variatie (van beide modellen) maakt het lastig om een goed waardeoordeel van één perceel te geven.

Tenslotte, is de foto-app besproken om inzicht te krijgen aan de gewenste functionaliteiten:

Vraag	Reactie
Welke behoefte kan de app vervullen in jouw werkveld?	1. Data verzameling verrijken: - o koppelen aan app die kan checken wat de satelliet data niet goed weergeeft - o koppelen aan indicator soorten - o teler zelf de foto laten maken, dataset achteraf checken/laten evalueren door bijv. inspecteur/adviseur - o directe terugkoppeling vanuit het veld; 2. Gebruik als stimulerings- en ontwikkeling tool, bijv.: - o geschiedenis van foto's op GPS-locatie met waardes om over de jaren heen ontwikkeling te zien - o export van data met betrekking tot ontwikkeling in de tijd 3. Gebruik als monitoring tool, bijv.: - o percentageschatting Q en hoeveelheid soorten - o met foto een Q-waarde, in elk veld koppelen met Schippers - o vlindertellingen koppelen aan kruidenrijkdom - o koppeling naar subsidie natuur en landschap (SNL) aangeven - o en app waarbij agrariërs de foto kan indienen bij het NVWA en tijdelijk geschatte waardes ontvangen; daardoor beter stroomlijnen welke percelen veldonderzoek nodig is. - o nu is alles op basis van veldonderzoek en checken op indicator soorten, de app zou dit (deels) kunnen vervangen
Wat is daarvoor nodig, en heeft de app al?	1. Publiekelijk toegankelijk beschikbaarheid app.
Wat is daarvoor nodig, en wat moet beter?	1. Helpen in de app met het maken van een juiste foto bijvoorbeeld met hoekkanteling en hulplijnen. 2. Upload foto en direct feedback ontvangen. 3. Uitbreiden met advies, bijvoorbeeld voor maaibeheer. 4. De app kunnen gebruiken om de ontwikkeling van het grasland over de jaren heen te monitoren. 5. Koppelen aan Schippers index.

4 Discussie

Spoor 1

In fase 1 van dit project zijn op grond van zoveel mogelijk beschikbare RS-informatie vier modellen gecreëerd om de Q-index (kruidenrijkdom) te voorspellen: een Ridge Regressie, een Random Forest regressie, een Random Forest classificatie, een SOM regressie en een deep learning (CNN) classificatie. Bij een regressie wordt de Q-index op een continue schaal voorspeld, en bij een classificatie wordt de Q-index onderverdeeld in "blokjes" van een halve eenheid (0,5Q), en voorspelt het model het meest waarschijnlijke blokje. Alle modellen zijn gefit op 80% van de dataset, en hun voorspellende kracht is getoetst op de overige 20% van de dataset.

Het opvallende resultaat is dat alle vier de modellen ongeveer dezelfde nauwkeurigheid geven: een geschatte voorspelfout (RMSE) van 0,7 eenheden (0,7Q). Dat is opvallend vanwege twee redenen. Allereerst zijn de modellen zeer verschillend van aard. Random Forest, Ridge Regressie, SOM's en deep learning (CNN) zijn modelleertechnieken die compleet anders werken. Er is te verwachten dat dit ook tot verschillende nauwkeurigheden zal leiden. Ten tweede zijn de data ook op een verschillend aggregatieniveau bekeken: deep learning (CNN) gebruikt de ruwe RS-tijdseries (zgn. RS-data), en de andere technieken (RR, RF en SOM) gebruiken parameters die van de ruwe RS-tijdseries zijn afgeleid (de zgn. RS-markers). Het lijkt erop dat deze aanpak (die samengevat kan worden als "alle data die aanwezig is wordt tegelijk aan de modellen gepresenteerd en die moeten er dan maar wat van maken") hier aan het plafond van de mogelijkheden zit. Praktisch betekent dit dat een voorspelling door deze modellen een zekere marge van onzekerheid heeft. De voorspelde waarde is steeds de meest waarschijnlijke waarde volgens het model, maar het kan zijn dat de werkelijke waarde hoger of lager ligt. Het is gebruikelijk om het 95% betrouwbaarheidsintervallen te rapporteren, die grofweg overeenkomen met plus of minus twee keer de geschatte voorspelfout (RMSE), dus hier plus of minus 1,4 eenheden (1,4Q). Dit zal voor de meeste toepassingen niet nauwkeurig genoeg zijn.

Er zijn een aantal mogelijke oorzaken voor deze onnauwkeurige voorspellingen. Ten eerste zijn de percelen waarvoor de Q-index bepaald zijn in de loop der tijd om specifieke redenen gekozen, redenen die niet noodzakelijk leiden tot een goede afspiegeling van de complete populatie van percelen (naar verhouding minder percelen met een lage Q-index ($Q < 2,5$) in de training dataset). Zo kunnen bepaalde bodemtypes, bepaalde gebieden in Nederland of percelen met specifieke karakteristieken onder- of oververtegenwoordigd zijn. Deze onbalans leidt tot modellen die op de "verkeerde" elementen de nadruk leggen. Overigens betekent dit ook dat de schatting van een voorspelfout van 0,7 "gekleurd" is: het kan zo zijn dat het overgrote deel van de Nederlandse percelen makkelijker te voorspellen is en dus tot een kleinere voorspelfout leidt – maar het kan ook de andere kant op vallen. Daar is met de huidige data weinig over te zeggen. Ten tweede bevat de data set weliswaar duizenden percelen met een Q-index, maar is er een enorme variatie in lokale omstandigheden binnen die groep aan percelen. Denk aan bodemtype, klimaatinvloeden, management - elke combinatie van omstandigheden zal een andere relatie tussen Remote Sensing gegevens en Q-waarden met zich meebrengen, en het is zeer onwaarschijnlijk dat een enkel model flexibel genoeg zal zijn om dat te kunnen beschrijven. AI-technieken zoals deep-learning netwerken zouden dat in principe kunnen, maar daarvoor zouden hoeveelheden voorbeelden nodig zijn die de grootte van de huidige training data vele ordes van grootte overstijgen. Alleen al het feit dat er de afgelopen periode een aantal zeer warme en droge jaren zijn geweest (zie Figuur 8), maar ook een jaar met een min of meer "normaal" groeiseizoen zal voor variatie zorgen, die tot uitdrukking komt in de onzekerheid van de modelvoorspellingen.

Tegelijkertijd wordt geconstateerd dat voorspellingen met deze modellen die voor alle percelen in heel Nederland zijn uitgevoerd, wel degelijk zeer adequate voorspellingen op een relatieve schaal kunnen doen: zo zijn voorspelde Q-index waarden in natuurgebieden duidelijk hoger dan de voorspellingen voor aanpalende graslanden (Figuur 29).

Spoor 2

Er zijn in fase 1 veel foto's verzameld van graslanden doormiddel van verschillende apps. Hiervoor zijn zowel detail en overzicht foto's gemaakt. Voor de foto algoritmen zijn uiteindelijk alleen de detail foto's gebruikt. Met deze foto's zijn twee sporen verkent. Het algoritme ontwikkeld in spoor 2b kan PKG-graslanden en Engelse raaigraslanden onderscheiden met een zekerheid van 94%. De hypothese was dat het RS-algoritme de kruidenrijkdom van ingezaaide productieve graslanden zou onderschatten. In de door ons verzamelde locaties van PKG-graslanden komt dit niet naar voren, de ranges van de voorspelde Q-indices waren behoorlijk groot. De nauwkeurigheid van het algoritme is van dien aard dat het goed in de praktijk toe te passen is.

Het doel van het tweede algoritme (spoor 2a) was om, op basis van foto's, ook Q-indices te kunnen schatten. De eerste aanpak, waarbij een regressie gemaakt is, leidde tot een te grote foute marge in de voorspellingen. Dit kan eventueel veroorzaakt zijn door de ongelijke verdeling van de Q-indices in de opgenomen foto dataverzameling. Op klei en veen was er een relatief groot aantal beelden met een lage Q-index, terwijl het op zand homogener verdeeld was.

De tweede aanpak, waarbij de Q-indices in 'blokken' van een halve Q-index verdeeld zijn, leidde tot een algoritme waarmee de Q-waarde met een RSME van 0,89 bepaald kan worden. Hierbij moet opgemerkt worden dat de validatie verzameling vrij klein was. In de tweede aanpak zijn de percelen niet per bodemtype opgedeeld. Omdat vegetaties per bodemtype erg verschillend zijn, zou een van de oorzaken hierin kunnen liggen.

In de derde aanpak zijn de foto's wel onderverdeeld per bodemtype. Dit is grofmazig gedaan op basis van klei-, veen- en zandgrond. De resultaten zijn aanmerkelijk beter als je kijkt naar het percentage goed geclassificeerde percelen. Het aantal percelen in de validatie verzamelingen was na de opdeling per bodemtype wel nog kleiner geworden. Het maken van foto's om de kruidenrijkdom te bepalen heeft dus wel degelijk potentie.

Tot slot is er een verkenning gedaan met de Shannon-index in plaats van de q-index. Dit gaf geen verbeteringen.

Stakeholder analyse workshop

Op 26 september 2024 is een stakeholder analyse workshop gehouden met als doel om de RS- en foto-algoritmes te evalueren samen met consortium partners en sponsors. Workshop deelnemers werden gevraagd om input op:

1. Stellingen formuleren over welk doel de modellen dienen.
2. Evalueren van Q-index overzicht kaarten van Nederland gemaakt met de RS-algoritmes.
3. Vergelijken van de QCNN en QSOM resultaten.
4. Inventarisatie van behoefte met betrekking tot het gebruik van de foto app.

Uit de reacties op de stellingen zijn nieuwe vragen ontstaan, zoals wanneer is het model goed genoeg? Wie gaat de tool gebruiken? en op welke manier zal het gebruikt worden? Over het algemeen is de groep positief gestemd over de nut van deze modellen waar het gaat om stimuleren van de ontwikkeling van kruidenrijkdom en het inzichtelijk maken van de mate van kruidenrijkdom van grasland percelen. Daarbij speelt vertrouwen hebben in het model een belangrijke rol, het belang van veldonderzoek wordt benadrukt. Het zou niet de bedoeling (moeten) zijn om de kaarten te gebruiken voor directe beloning/afrekening; er is behoefte aan veldonderzoek ter controle op het land. De datapositie van de eigenaar van het perceel zou ook gewaarborgd moeten zijn. Tot slot, het (kunnen) integreren van dit model met bestaande systemen werd ook als belangrijk ervaren.

5 Conclusies en aanbevelingen

Het doel van fase 1 van dit PPS-project is het 'Ontwikkeling van een monitoringssysteem dat jaarlijks een uitspraak doet over kruidenrijkdom van alle graslandpercelen in Nederland'. Om dit doel te realiseren zijn twee-sporen opgezet:

- Spoor 1: Ontwikkelen van een RS-algoritme dat op basis van verzamelde RS-data en/of afgeleide RS-markers een uitspraak kan doen over de kruidenrijkdom (van alle grasland percelen in Nederland).
- Spoor 2a: Ontwikkelen van een foto-algoritme dat op basis van foto's (detail + overzichtsfoto) van een perceel een uitspraak kan doen over de kruidenrijkdom.

Met als belangrijkste onderzoeksvragen:

- Kan op basis van een RS-algoritme met RS-data en/of RS-markers (spoor 1) de kruidenrijkdom van grasland op jaarbasis worden geschat voor heel Nederland?
 - Hoe nauwkeurig zijn de voorspellingen van het RS-algoritme?
 - Hoe consistent zijn de voorspelling jaar-op-jaar?
- Kan op basis van een foto-algoritme met foto's (spoor 2a) van een perceel de kruidenrijkdom van het perceel nauwkeurig worden geschat?
 - Hoe nauwkeurig zijn de voorspellingen van het foto-algoritme?
- Zijn de voorspellingen van beide sporen (1 en 2a) consistent in het voorspellen van de kruidenrijkdom?

Specifiek voor ingezaaid productief kruidenrijkgrasland (PKG):

- Kan het foto-algoritme (spoor 2b) onderscheid maken tussen regulier Engels raaigras en zgn. productief kruidenrijke graslanden (PKG)?
- Hoe gaat het RS-algoritme om met PKG-percelen?
 - *Hypothese: Het RS-algoritme zal de kruidenrijkdom van ingezaaide productieve kruidenrijke grasland (PKG) onderschatten.*

Conclusies – Spoor 1

Met alle onderzochte RS-algoritmes, waaronder het RR-, RF- en SOM-model op basis van de afgeleide RS-markers en het CNN-model op basis van de ruwe RS-data, kan de kruidenrijkdom (Q-index) worden voorspeld. Echter, de voorspelfout (RMSE) van alle RS-modellen, voor zowel regressie als classificatie, ligt rond de 0,7Q. Dat komt overeen met een 95%-betrouwbaarheidsinterval van $\pm 1,4Q$. Dit zal wellicht voor een aantal toepassingen (die voorzien worden in fase 2) niet nauwkeurig genoeg zijn. Het verminderen van het aantal klasse naar twee klasse (lage kruidenrijkdom $Q \leq 3$ en hoge kruidenrijkdom $Q > 3$) of het gebruik maken van twee modellen (een voor percelen met een lage kruidenrijkdom en een voor percelen met hoge kruidenrijkdom) biedt geen verbetering.

Als we kijken op landelijke schaal naar de voorspelde kruidenrijkdom van alle graslanden (Figuur 26 (2021) en Figuur 29 (2022)), dan kan geconcludeerd worden dat de ruimtelijke patronen alsmede de relatieve verschillen er logisch uit zien. Dat wil zeggen de modellen geven een hoge kruidenrijkdom aan in Natura2000, weidevogel rustgebieden, en andere natuurterreinen vergeleken met de 'rest' van de graslanden in Nederland. Kijken we vervolgens naar de frequentie verdelingen van elk model en voor de verschillende jaren (Figuur 30), dan blijkt dat met name het RR- en RF-model de kruidenrijkdom van de productieve graslanden overschatten (dat wil zeggen de grootste groep heeft 'n Q-index tussen 2 en 3). Bij het SOM- en CNN-model daarentegen, ligt het zwaartepunt bij de groep met een Q-index tussen 1 en 2. Dat lijkt beter overeen te komen met de realiteit, aangezien de meerderheid van de graslanden in Nederland voornamelijk productieve graslanden zijn. Het SOM model heeft echter wel de neiging om de groep met Q-index 4-5 te overschatten in tegenstelling tot het CNN model (en RR- en RF-).

Kijken we tot slot naar de jaarlijkse variabiliteit van de voorspelde kruidenrijkdom door de jaren heen (dat wil zeggen de periode 2018 t/m 2022), dan is met name het CNN-model behoorlijk stabiel, in tegenstelling tot de andere modellen (Figuur 30).

Conclusies – Spoor 2a

Waar in spoor 1 zowel de regressie aanpak als de classificatie aanpak gelijkwaardige resultaten opleveren voor de verschillende RS-algoritmes, is dit bij het foto-algoritme niet het geval. De regressie aanpak resulteert in bedroevende resultaten ($RMSE > 1Q$). De classificatie aanpak daarentegen leverde betere resultaten op, namelijk een RSME van 0,89Q (en model nauwkeurigheid van 52%). Net als het RS-algoritme zal dit voor de meeste toepassingen (in fase 2) niet nauwkeurig genoeg zijn.

Conclusies - PKG-percelen (Spoor 2b)

Het herkennen van PKG-percelen ten opzichte van regulier productief Engels raaigras met het foto-algoritme (spoor 2b) laat zeer goede resultaten zien. Het algoritme is in staat om 95% van de percelen correct te classificeren, dat wil zeggen het model kan succesvol onderscheid maken tussen Engels raaigras- en ingezaaide PKG-percelen.

Wordt vervolgens gekeken naar de Q-voorspelling van deze PKG-percelen met het RS-algoritme, dan wordt een behoorlijke range aan Q-waarden waargenomen, variërend van Q is 1,5 tot 4,5 (soms zelfs 5). Van PKG-percelen is bekend dat het aantal kruiden met de jaren afneemt. Dit zou een verklaring kunnen zijn voor de grote spreiding in voorspelde Q-index, echter een Q-index van 4,5 lijkt wel aan de hoge kant. Een kleine kanttekening dient nog wel te worden geplaatst, de intensiteit landgebruik (maaifrequentie) van de geanalyseerde PKG-percelen lijkt wel aan de lage kant en suggereert dat het geen intensief gebruikte percelen zijn.

Conclusies – Stakeholder analyse workshop

Over het algemeen is de groep positief gestemd over de nut van deze modellen waar het gaat om stimuleren van de ontwikkeling van kruidenrijkdom, en het inzichtelijk maken van de mate van kruidenrijkdom van grasland percelen. Echter, het belang van veldonderzoek wordt benadrukt, waarbij de kaarten lijken in de huidige vorm ongeschikt om te gebruiken voor directe beloning/afrekening. De datapositie van de eigenaar van het perceel en het (kunnen) integreren van dit model met bestaande systemen zijn ook benoemd als belangrijke aspecten voor de stakeholders. Het is zaak om rekening te houden met deze ethisch, juridisch en sociale aspecten bij het verdere ontwerp van de applicatie van de modellen.

6 Literatuur

- Bax, I. & W. Schippers (1998). Ontwikkeling van botanisch waardevol grasland: veldgids. Dienst Landelijk Gebied, Utrecht, Publicatie/ IKC Natuurbeheer (nr. C-18), 88 pags. <https://edepot.wur.nl/336189>
- Bekkema, M. E., & Eleveld, M. (2018). Mapping grassland management intensity using sentinel-2 satellite data. *GI Forum*, 1, 194-213.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- De Vroey, M., Radoux, J., & Defourny, P. (2021). Grassland Mowing Detection Using Sentinel-1 Time Series: Potential and Limitations. *Remote Sens.* 2021, 13, 348.
- Fauvel, M., Lopes, M., Dubo, T., Rivers-Moore, J., Frison, P. L., Gross, N., & Quin, A. (2020). Prediction of plant diversity in grasslands using Sentinel-1 and-2 satellite image time series. *Remote Sensing of Environment*, 237, 111536.
- Groenmonitor.nl (staat op bladzijde 108; Bijlage C)
- He, K., Zhang, X., Ren, S. and Sun, J. Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- Hoerl, A. and Kennard, R. (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 55-67. <https://doi.org/10.1080/00401706.1970.10488634>
- Howison, R. A., Piersma, T., Kentie, R., Hooijmeijer, J. C., & Olff, H. (2018). Quantifying landscape-level land-use intensity patterns through radar-based Remote Sensing. *Journal of Applied Ecology*, 55(3), 1276-1287.
- Kingma, D. P. and Welling, Max (2013). Auto-Encoding Variational Bayes. *ICLR*.
- Kohonen, T. (2021) "Self-organising maps". Springer.
- Kolecka, N., Ginzler, C., Pazur, R., Price, B., & Verburg, P. H. (2018). Regional scale mapping of grassland mowing frequency with Sentinel-2 time series. *Remote Sensing*, 10(8), 1221.
- Li, Y., Piersma, T., Hooijmeijer, J. C., & Howison, R. A. (2023). Land-use intensity impacts habitat selection of ground-nesting farmland birds in The Netherlands. *Ecological Solutions and Evidence*, 4(1), e12201.
- Montaldo, Nicola, Laura Fois, and Roberto Corona (2021). "Soil moisture estimates in a grass field using Sentinel-1 radar data and an assimilation approach." *Remote Sensing* 13.16 (2021): 3293.
- Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F.A., Schwarz, L.M., Ghazaryan, G., Schultz, J., Malenovský, Z. and Dubovyk, O. (2022). Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with Remote Sensing and deep learning. *Remote Sensing of Environment*, 282, p.113262.
- Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., & Reppucci, A. J. R. S. O. E. (2013). Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sensing of Environment*, 134, 234-248.
- Pirtti, A., Hoşbaş, R. G., & Yücel, M. A. (2023). Examination of the Earthquake (Samos Island) in Izmir (30.10. 2020) by Using Cors-Tr GNSS Observations and InSAR Data. *KSCE Journal of Civil Engineering*, 27(1), 135-144.
- Reinermann, S., Gessner, U., Asam, S., Ullmann, T., Schucknecht, A., & Kuenzer, C. (2022). Detection of grassland mowing events for Germany by combining Sentinel-1 and Sentinel-2 time series. *Remote Sensing*, 14(7), 1647.
- Schaminée, J.H.J. & Jansen, A.J.M. (1998). Wegen naar natuurdoeltypen; ontwikkelingsreeksen en hun indicatoren ten behoeve van herstelbeheer en natuurontwikkeling (sporen A en B). IBN-DLO Wageningen, 320 pags.
- Schaminée, J.H.J., Jansen, A., Aggenbach, C., Haveman, R., Sierdsema, H., Smits, N.A.C., R. van 't Veer (2000). Wegen naar natuurdoeltypen 2; ontwikkelingsreeksen en hun indicatoren voor herstelbeheer en natuurontwikkeling (sporen B en C. EC-LNV, Wageningen (Rapp. Dir. Natuurbeheer 46), 364 pags.
- Schippers, W., Bax, I., & Gardeniers, M. (2015). Ontwikkelen van kruidenrijk grasland (1^e editie). Samenwerkende Uitgevers VOF.

-
- Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., & Hostert, P. (2022). Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sensing of Environment*, 269, p.112795.
- Specht, D. F. (1996). Probabilistic neural networks and general regression neural networks. In *Fuzzy Logic and Neural Network Handbook* (pp. 301-344). USA: McGraw-Hill, Inc.
- Stumpf, F., Schneider, M.K., Keller, A., Mayr, A., Rentschler, T., Meuli, R.G., Schaepman, M. and Liebisch, F. (2020). Spatial monitoring of grassland management using multi-temporal satellite imagery. *Ecological Indicators*, 113, p.106201.
- Tamm, T., Zalite, K., Voormansik, K., & Talgre, L. (2016). Relating Sentinel-1 interferometric coherence to mowing events on grasslands. *Remote Sensing*, 8(10), 802.
- Taravat, A., Wagner, M. P., & Oppelt, N. (2019). Automatic grassland cutting status detection in the context of spatiotemporal Sentinel-1 imagery analysis and artificial neural networks. *Remote Sensing*, 11(6), 711.
- Van 't Veer, R. (2020). Kruidenrijke graslanden in de Enge Wormer. Biodiversiteit, graslandherkenning en beheer. Van 't Veer & De Boer Ecologisch Advies & Onderzoeksbureau, Jisp, 80 pag.s + bijlagen.
- Vreugdenhil, M., Navacchi, C., Bauer-Marschallinger, B., Hahn, S., Steele-Dunne, S., Pfeil, I., ... & Wagner, W. (2020). Sentinel-1 cross ratio and vegetation optical depth: A comparison over Europe. *Remote Sensing*, 12(20), 3404.
- Wehrens, R., & Kruisselbrink, J. (2018). Flexible self-organizing maps in kohonen 3.0. *Journal of Statistical Software*, 87, 1-18. <https://doi.org/10.18637/jss.v087.i07>
- Wosten, J. H. M., de Vries, F., Hoogland, T., Massop, H. T. L., Veldhuizen, A. A., Vroon, H. R. J. & Bolman, A. (2013). BOFEK2012, de nieuwe bodemfysische schematisatie van Nederland (No. 2387). Alterra.
- Yehoshua, Roi (Mar 25, 2023). Random Forests. Medium, geraadpleegd dd-mm-jjjj. <https://medium.com/@roiyeho/random-forests-98892261dc49>
- Zhang, B., Wdowski, S., Gann, D., Hong, S. H., & Sah, J. (2022). Spatiotemporal variations of wetland backscatter: The role of water depth and vegetation characteristics in Sentinel-1 dual-polarization SAR observations. *Remote Sensing of Environment*, 270, 112864.
- Zou, B., Zhang, Y., Cao, N., & Minh, N. P. (2015). A four-component decomposition model for PolSAR data using asymmetric scattering component. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(3), 1051-1061.

Bijlage A Methodiek Q-index, voederwaarde, kruiden- en grassenindex

Hoe is de Q-index berekend? Bij het maken van een vegetatieopname wordt een soortenlijst gemaakt. Voor iedere soort is bekend met welk percentage de soort het telvlak bedekt. Aan iedere soort hangt een uniek Q-index (zie tabel voor de Q-index van de grassen en kruiden die in dit onderzoek zijn aangetroffen). Door de abundantie te vermenigvuldigen met het Q-index, kan vervolgens de gewogen Q-index voor een vegetatieopname worden berekend. Zie Tabel 11 voor een fictief voorbeeld.

Tabel 11 Voorbeeld van het berekenen van de Q-index voor een fictieve vegetatieopname.

Soort	Abundantie (% dekking)	Q-getal	Dekking* Q-index
A	75	1	75
B	10	2	20
C	10	2	20
D	10	2	20
E	5	3	15
F	5	3	15
TOTAAL	115	-	165

Q-index = $165/115 = 1.43$

Tabel 12 Karakteristieken voor alle soorten die zijn aangetroffen in dit onderzoek: de Q-index, de voederwaarde (kv getal) en de categorie waartoe de soort behoort (gras/kruid/storingssoort). Alle waardes zijn direct ontleend aan van 't Veer (2020). De Q-index is gebaseerd op de graslandfasen van Schippers et al. (2015), Schaminee & Jansen (1998) en Schaminee et al. (2000). Lage Q-indices passen bij soorten die vooral voorkomen in soortenarme, intensief gebruikte graslanden. Voor hoge Q-indices geldt het tegenovergestelde. De kv-getallen zijn gebaseerd op werk van Kruijne & de Vries (1964). Des te hoger het getal, des te groter de voederwaarde (eiwitgehalte, verteerbaarheid).

Wetenschappelijke naam	Nederlandse naam	Q-index	Categorie (gras, kruid, storingssoort)	Voederwaarde (kv-getal)
Agrostis stolonifera	Fioringras	4	Gras	5
Alopecurus geniculatus	Geknikte vossenstaart	3	Gras	3
Alopecurus pratensis	Grote vossenstaart	3	Gras	7
Anthoxanthum odoratum	Gewoon reukgras	5	Gras	4
Arrhenatherum elatius	Glanshaver	3	Gras	8
Bellis perennis	Madeliefje	4	Kruid	1
Brachythecium rutabulum		5	Kruid	0
Bromus hordeaceus		3	Gras	3
Cardamine hirsuta	Kleine veldkers	3	Storingsplant	2
Cardamine pratensis	Pinksterbloem	4	Kruid	3
Carex disticha	Tweerijige zegge	6	Kruid	2
Carex hirta	Ruige zegge	5	Kruid	2
Carex species	Zegge (G)	4	Kruid	2
Cerastium fontanum	Gewone hoornbloem s.l.	2	Kruid	2
Cerastium glomeratum	Kluwenhoornbloem	2	Kruid	1
Cirsium arvense	Akkerdistel	2	Storingsplant	0
Crepis capillaris	Klein streepzaad	4	Kruid	3
Cynosurus cristatus	Kamgras	5	Gras	7
Dactylis glomerata	Kropaar	3	Gras	6

Wetenschappelijke naam	Nederlandse naam	Q-index	Categorie (gras, kruid, storingssoort)	Voederwaarde (kv-getal)
Dactylorhiza majalis subsp. Praetermissa	Rietorchis	6	Kruid	2
Deschampsia cespitosa	Ruwe smele	3	Gras	0
Elytrigia repens	Kweek	3	Gras	5
Equisetum palustre	Lidrus	5	Kruid	0
Festuca pratensis	Beemdlangbloem	4	Gras	9
Festuca rubra	Rood zwenkgras	5	Gras	4
Glechoma hederacea	Hondsdrif	3	Kruid	0
Glyceria fluitans	Mannagras	3	Gras	4
Glyceria maxima	Liesgras	3	Gras	4
Gnaphalium uliginosum	Moerasdroogbloem	4	Kruid	0
Heracleum sphondylium	Gewone berenklauw	3	Storingsplant	2
Holcus lanatus	Gestreepte witbol	3	Gras	5
Holcus mollis	Gladde witbol	3	Gras	3
Juncus articulatus	Zomprus	5	Kruid	2
Juncus bufonius	Greppelrus	4	Kruid	1
Leontodon saxatilis	Kleine leeuwentand	5	Kruid	2
Lolium perenne	Engels raaigras	1	Gras	10
Matricaria species		2	Storingsplant	0
Myosotis laxa/scorpioides		5	Storingsplant	2
Persicaria amphibia	Veenwortel	2	Storingsplant	2
Phalaris arundinacea	Rietgras	3	Storingsplant	6
Phleum pratense s.l.		3	Gras	9
Plantago lanceolata	Smalle weegbree	5	Kruid	4
Poa annua	Straatgras	2	Gras	4
Poa pratensis	Veldbeemdgras	1	Gras	9
Poa trivialis	Ruw beemdgras	2	Gras	8
Polygonum aviculare	Gewoon varkensgras	3	Kruid	3
Potentilla anserina	Zilverschoon	3	Kruid	1
Ranunculus acris	Scherpe boterbloem	4	Kruid	0
Ranunculus repens	Kruipende boterbloem	3	Kruid	1
Ranunculus sceleratus	Blaartrekkende boterbloem	2	Storingsplant	0
Rhinanthus angustifolius	Grote ratelaar	6	Kruid	0
Rorippa species		2	Storingsplant	1
Rumex acetosa	Veldzuring	4	Kruid	3
Rumex crispus	Kruizuring	3	Storingsplant	2
Rumex obtusifolius	Ridderzuring	2	Storingsplant	1
Stellaria media	Vogelmuur	2	Storingsplant	1
Taraxacum species	Paardenbloem	2	Kruid	4

Wetenschappelijke naam	Nederlandse naam	Q-index	Categorie (gras, kruid, storingssoort)	Voederwaarde (kv-getal)
Trifolium dubium	Kleine klaver	5	Kruid	6
Trifolium pratense	Rode klaver	4	Kruid	7
Trifolium repens	Witte klaver	3	Kruid	8
Tripleurospermum maritimum	Reukeloze kamille	2	Storingsplant	0
Urtica dioica	Grote brandnetel	2	Storingsplant	0
Veronica arvensis	Veldereprijs	4	Kruid	1

Bijlage B Toelichting RS-data

Optische satelliet data

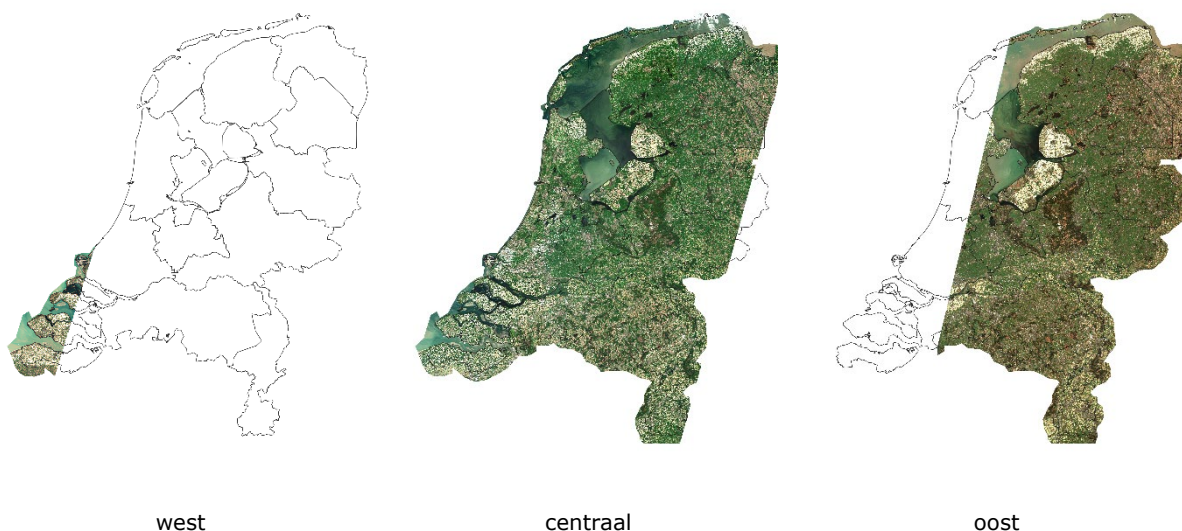
Zowel Sentinel-2 als Landsat-7 & 8 hebben een multi-spectrale sensor met meerdere spectrale banden. Deze banden liggen in verschillende delen van het ElektroMagnetisch (EM) spectrum, waaronder in het zichtbaar⁴ blauwe (B), groene (G) en rode (R) deel. Daarnaast beschikken de satellieten nog over andere spectrale banden, zoals een nabij-infrarode band (NIR), een of meerdere Red-Edge banden en shortwave-infrarode (SWIR) banden (Tabel 13). Met behulp van deze spectrale banden is het mogelijk onderscheid te maken tussen verschillende landoppervlakken en vegetaties (zoals bv water, kale bodem, gras, bos, etc.) en kunnen verschillende eigenschappen van de bodem en het gewas worden afgeleid (zoals bladoppervlak, biomassa, etc.).

Tabel 13 Overzicht van Sentinel-2 banden, en bijbehorende ruimtelijke resolutie.

Band nr.	Band	Resolutie (m)
2	Blauw (B)	10
3	Groen (G)	10
4	Rood (R)	10
5	Red-Edge (RE)	20
6	Red-Edge (RE)	20
7	Red-Edge (RE)	20
8	Near-InfraRed (NIR)	10
8A	Red-Edge (RE)	20
11	ShortWave-InfraRed (SWIR)	20
12	ShortWave-InfraRed (SWIR)	20

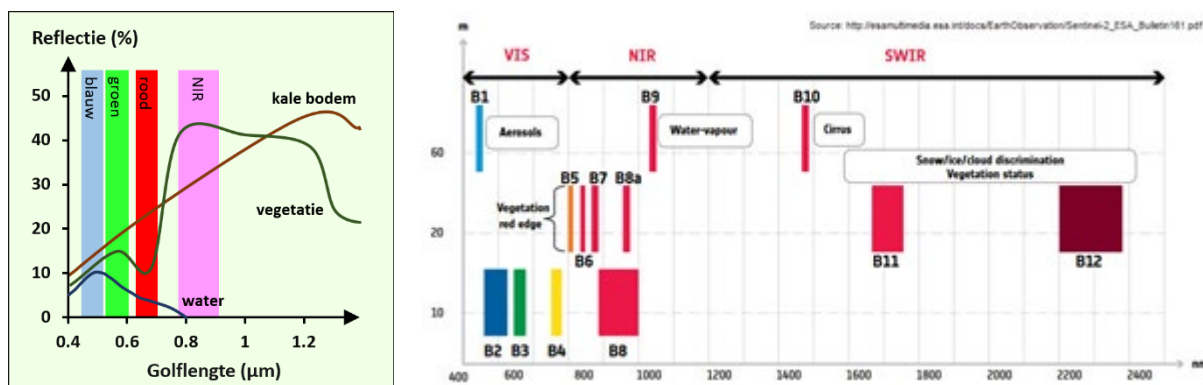
Met de rode, groene en blauwe (RGB) banden kan de satelliet, net als het menselijk oog, naar het aardoppervlak kijken en kleurenfoto's maken (Figuur 40). Vanwege de hoge ruimtelijke resolutie (10 m tot 30 m) is het daardoor mogelijk om te kijken naar kleurvariaties tussen percelen, maar ook binnen een perceel.

Sentinel-2, bestaat nu uit 2 satellieten (2A en 2B). Beide volgen dezelfde vaste baan rond de aarde waardoor ze Nederland volgens een vast patroon opnemen. Figuur 40 laat zien hoe Sentinel-2 over Nederland komt, waardoor soms alleen het meest westelijke puntje van Nederland wordt opgenomen, of bijna helemaal (90%), of alleen het centraal-oostelijke deel opneemt zonder de kuststrook.



Figuur 40 Voorbeelden van een (onbewolkte) westelijke, centrale en oostelijke Sentinel-2 overpass opname.

⁴ Vandaar de term optische satelliet data.



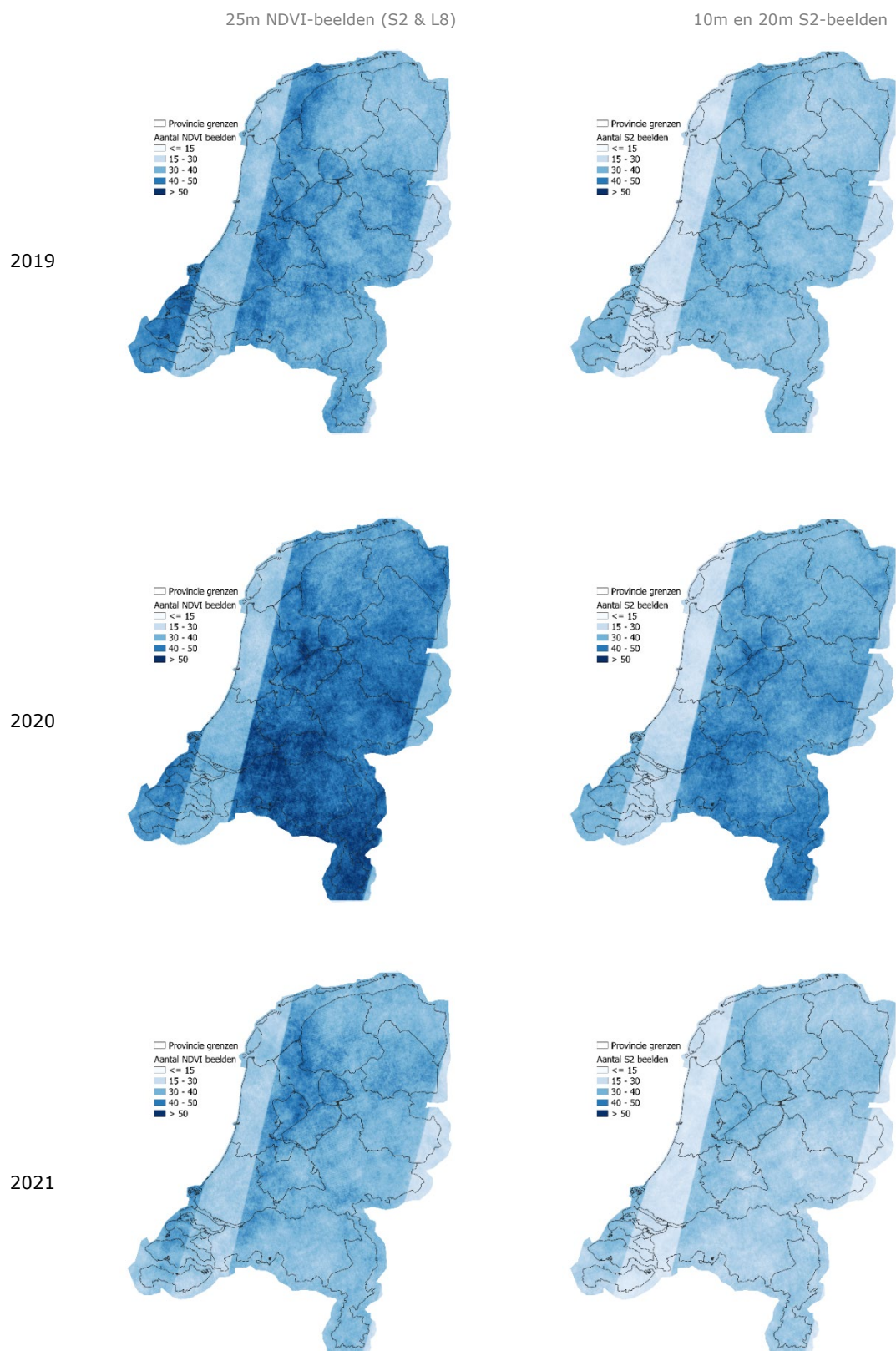
Figuur 41 Links: Spectrale reflecties van vegetatie, water en kale bodem. Rechts: Spectrale banden Sentinel-2 ([bron](#)).

Zowel de rode (R) band als de nabij-infrarode (NIR) band zijn zeer geschikt voor het monitoren van vegetatie. Planten, of beter gezegd groene biomassa, absorberen grotendeels het rode zonlicht middels fotosynthese en reflecteren het merendeel van het nabij-infrarode (NIR) licht. Bij kale bodems (rotsen, zand, geploegd land) is dit verschil veel kleiner (Figuur 41, links). Dit principe wordt gebruikt om een groenindex te berekenen. De bekendste groenindex is de Normalised Difference Vegetation Index (NDVI). De NDVI-groenindex is een ratio tussen de reflecties van het rode en het nabij-infrarode licht, en heeft waarden tussen -1 en 1:

$$NDVI = \frac{NIR - R}{NIR + R}$$

Waarbij open water een NDVI-groenindex heeft van -1 tot 0 en kale bodems een waarde tussen 0.1 en 0.2. De eerste vegetatie (bijv. opkomende grassprietjes) geeft waarden van ~0,25. Bij ~0,6 is het gewas gesloten (dat wil zeggen de bodem is volledig bedekt door de vegetatie) en een gewas met meerdere bladlagen geeft NDVI-waarden tot bijna 1. Een andere veel gebruikte vegetatie index is de WdVI (Weighted Difference Vegetation Index). Ook deze index maakt gebruik van de rode en nabij-infrarode, maar heeft in tegenstelling tot de NDVI minder last van verzadiging aan de hoge kant.

De aanvullende Red-Edge banden (RE, band 5, 6, 7 en 8A) zijn zodanig gepositioneerd op de steile reflectie curve van vegetatie (Figuur 41 en Figuur 65), waardoor aanvullende karakteristieken kunnen worden afgeleid. Zo is het mogelijk om beter te kijken naar chlorofyl en stikstof concentraties in het blad en is het mogelijk om beter onderscheid te kunnen maken tussen verschillende type vegetaties. Tot slot zijn er nog de SWIR-banden. Deze banden zijn o.a. gevoelige voor vocht in de plant. Op basis van deze set van verschillende banden zijn meerdere soorten indicatoren afgeleid, zoals de NDVI, S2REP en andere. Zoals al eerder gezegd speelt in het geval van optische satelliet data bewolking een belangrijke rol. Als het bewolkt is op het moment van de opname, dan zijn de beelden (of delen van de beelden waar het bewolkt is) niet bruikbaar. De afbeeldingen in Figuur 42 laten een overzicht zien van de opname dichtheid, dat wil zeggen het aantal bruikbare opnames per jaar voor heel Nederland. De opname dichtheid (of temporele resolutie) verschilt per jaar aangezien het weer (bewolking) de metingen beïnvloed en per locatie. Dat laatste is te relateren aan de ligging van de baan van S2 ten opzichte van Nederland (Figuur 42). Omdat de NDVI-tijdseries zijn gebaseerd op zowel Sentinel-2 (A+B) als Landsat-7 & 8, zijn er van de 25 m NDVI-opnames in het algemeen meer opnames beschikbaar dan alleen op basis van S2 (10 m en 20 m). Met name in de kuststrook van Noord- en Zuid-Holland, delen van Zeeland en het meest oostelijke deel van Nederland is de dekkingsgraad beperkt.



Figuur 42 Opname dichtheid 25 m NDVI (Sentinel-2A+B en Landsat) en 10 m+20 m Sentinel-2A+B gebaseerd op basis van het aantal beelden, waarbij bewolkte en schaduw zijn verwijderd.

Radar satelliet data

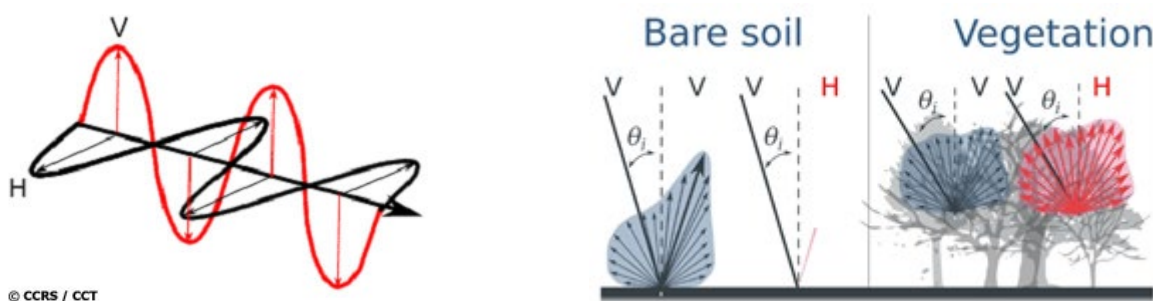
Backscatter

Sentinel-1 is een actieve radar satelliet (C-band, 5.405 GHz), en zendt RADAR (RADar Detection And Ranging) pulsen naar het aardoppervlak, en meet vervolgens de hoeveelheid radarstraling die door het aardoppervlak wordt teruggekaatst ('verstrooit'). Over het algemeen dragen verschillende verstrooiingsmechanismen bij aan het terugverstrooiingssignaal (het zgn. backscatter-signaal), namelijk: oppervlakteverstrooiing, volumeverstrooiing en interactieverstrooiing⁵ (Figuur 43). De mate van verstrooiing wordt bepaald door de diëlektrische eigenschappen van het medium, die op zijn beurt voornamelijk een functie is van het vochtgehalte in de bodem en/of de vegetatie (Vreugdenhill *et al.*, 2018, 2020). Verstrooiing van een oppervlak wordt ook beïnvloed door de ruwheid van het oppervlak. Kale bodems zijn over het algemeen niet heel vlak, waardoor de verstrooiing een combinatie is van spiegelende en diffuse verstrooiing (Figuur 43).



Figuur 43 Spiegelende en diffuse oppervlakteverstrooiing (links + midden) en volumeverstrooiing (rechts) (Zou *et al.*, 2015).

Vegetatie zorgt vooral voor volumeverstrooiing en is net als de bodem een functie van het vochtgehalte (in het blad, stengel/takken en stam), maar ook gevoelig voor het type vegetatie (dichtheid, vorm en lengte van de vegetatie). Bovendien leidt volumeverstrooiing ook tot depolarisatie effecten ten gevolge van de meervoudige verstrooiing die plaats vindt in het 'volume' (dat wil zeggen in de vegetatie). Depolarisatie in radar is de verandering in de oriëntatie van het elektrisch veld tijdens een reflectie (Figuur 44).



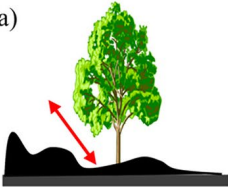
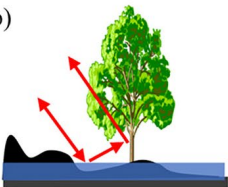
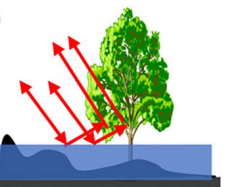
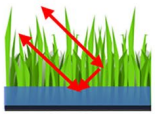
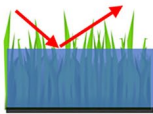
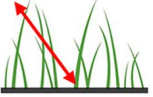
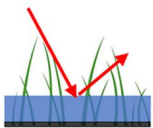
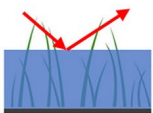
Figuur 44 Horizontale (H) en verticale (V) polarisatie (Links) (Bron), en het verstrooiings- mechanisme bij kale grond en bij vegetatie (Vreugdenhill *et al.*, 2015).

Sentinel-1 emitteert verticaal (V) gepolariseerde pulsen en detecteert zowel verticaal (V) als horizontaal (H) gepolariseerde signalen, dat wil zeggen het meet backscatter-VV én VH. Zowel de bodem als vegetatie dragen bij aan het terugverstrooiingssignaal in VV en VH.

De interpretatie van radar beelden/signalen is ingewikkeld omdat zowel de ruwheid en vochtgehalte van de bodem als van de vegetatie (indien aanwezig) invloed hebben op het scatteringsmechanisme en dus het gemeten backscatter-signaal. Om iets te zeggen of het vochtgehalte in de bodem dient het signaal gecorrigeerd te worden voor 'vegetatie-deel' en omgekeerd (Vreugdenhil *et al.*, 2018, 2020; Montaldo *et al.*, 2021). Daarbij helpt het om kennis van het oppervlak te hebben om het signaal te kunnen interpreteren zoals hoogte en dichtheid van de vegetatie, vochtcondities van de grond(soort) en de vegetatie(type), plasdras omstandigheden en (eventuele) managementactiviteiten (maaïen, beweiden, bemesten en beregenen), zoals in Figuur 45 wordt weergegeven (Zhang *et al.*, 2022).

⁵ Interactieverstrooiing is een combinatie van oppervlakte- en volumeverstrooiing.

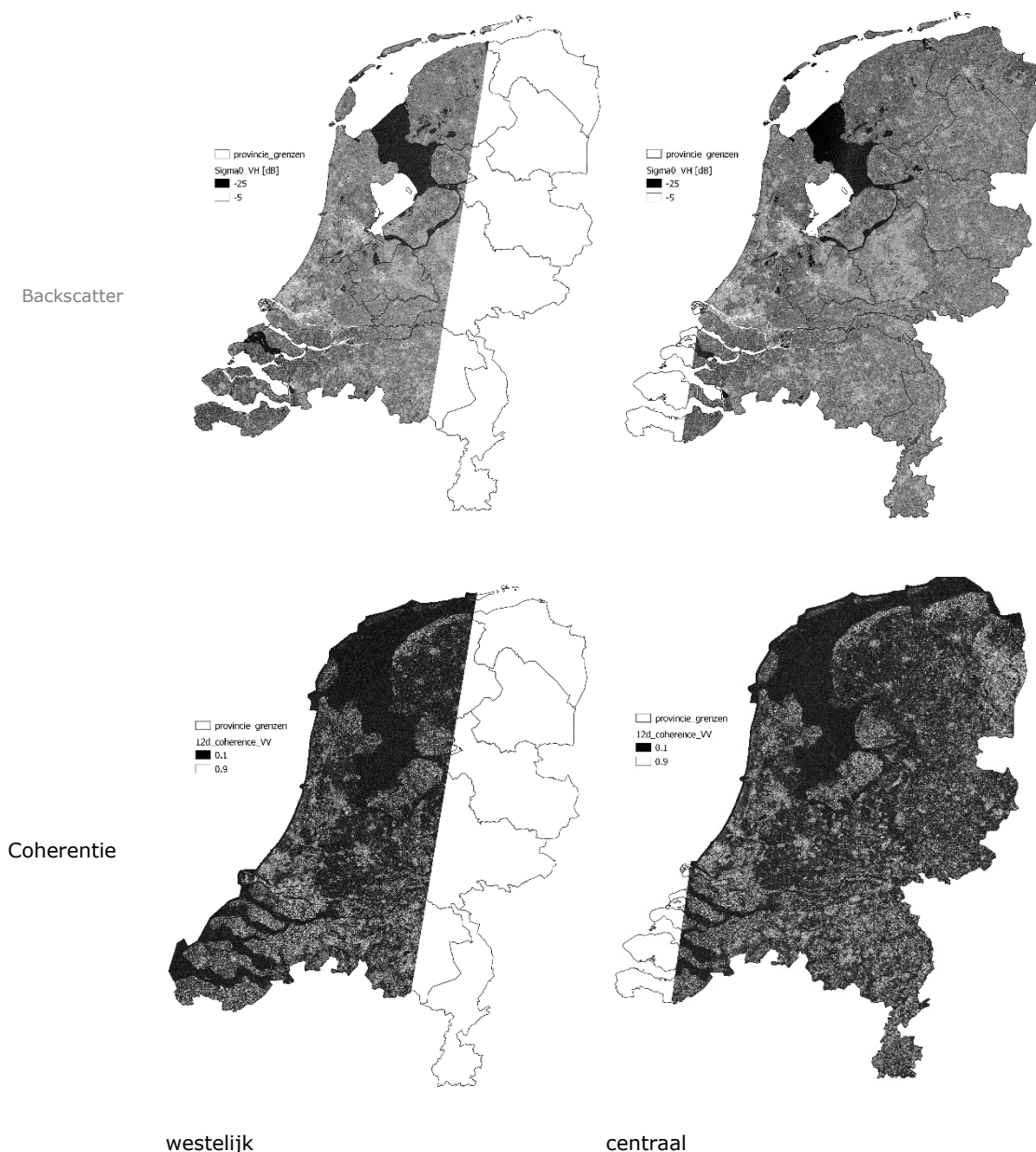
Onderzoek heeft aangetoond dat het verschil tussen het VV en VH backscatter-signaal, ook wel bekend als de depolarisatie ratio of kruisverhouding (CR), zeer gevoelig is voor veranderingen in het watergehalte en de biomassa van vegetatie (gewassen), waardoor CR zeer geschikt is voor het monitoren van vegetatie (Vreugdenhill *et al.*, 2020).

Vegetation	Unflooded	Shallow Water	Deep Water
Sparse Woody	(a)  Surface Scattering	(b)  Double-bounce and multiple-path scattering	(c)  Enhanced double-bounce and multiple-path scattering
Medium Dense Herbaceous	(e)  Surface Scattering	(f)  Double-bounce and multiple-path scattering	(g)  Specular Reflection
Sparse Herbaceous	(i)  Surface Scattering	(j)  Specular Reflection	(k)  Specular Reflection

Figuur 45 Schematische illustraties die de verstrooiing van C-band microgolfenergie vanaf grond/wateroppervlakken tonen voor drie vegetatietypen: schaars houtachtig, middelmatig dicht en schaars kruidachtig, en bij verschillende omstandigheden (droog, ondiepe en diepe waterstand) (Zhang *et al.*, 2022).

In Figuur 46 zijn twee backscatter (VV) beelden te zien van Nederland (westelijke en centrale overpass), waarbij de backscatter coëfficiënt is uitgedrukt in decibels (dB). De beelden geven hier informatie over de ruwheid van het oppervlak: hoe ruwer het oppervlak, hoe hoger de backscatter coëfficiënt (des te 'witter'). Vegetatie is over het algemeen redelijk ruw met een gemiddelde backscatter coëfficiënt en daardoor zichtbaar als donker- tot lichtgrijs. Kale grond is daarentegen redelijk wit (hoge backscatter coëfficiënt) waarbij natte gronden (zeer wit) meer reflecteren dan droge gronden (redelijk wit). Hier speelt dus naast de ruwheid ook het vochtgehalte in de bodem een rol. Water tot slot geeft juist hele lage backscatter waarden (donkergrijs tot zwart, zie Veluwemeer, Veerse meer en IJsselmeer⁶), het water fungeert namelijk als een spiegel en reflecteert het radarsignaal weg van de sensor (tenzij de wind voor een ruw wateroppervlak zorgt).

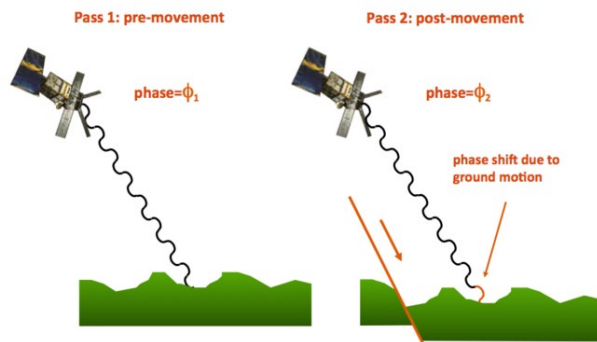
⁶ Rest van het wateroppervlak is weg gemaskeerd.



Figuur 46 Voorbeeld van backscatter-VV en coherentie-VV beelden van 27 mei (links, westelijke overpas) en 28 mei (rechts, centrale overpas) 2021.

Coherentie

Interferometrische coherentie (verder aangeduid als coherentie) kan worden gedefinieerd als de amplitude van de complexe correlatiecoëfficiënt tussen twee complexe SAR-beelden (zoals die van Sentinel-1) genomen vanuit verschillende baanposities en/of van verschillende tijdstippen. Coherentie beschrijft de gelijkenis van het gereflecteerde radarsignaal tussen twee beelden en varieert tussen 0 (incoherent) en 1 (coherent). Wanneer de twee beelden van dezelfde scène worden genomen, met dezelfde invalshoek (dezelfde baan) maar op verschillende tijdstippen, geeft de coherentie informatie over de stabiliteit van het verstrooiingsoppervlak (of doel) (Figuur 47). Onder dergelijke omstandigheden geeft lage coherentie aan dat het oppervlak is veranderd tussen de twee waarnemingsdata, terwijl hoge coherentie aangeeft dat er geen veranderingen zijn, waardoor coherentieanalyse bijzonder nuttig is voor het detecteren van veranderingen.

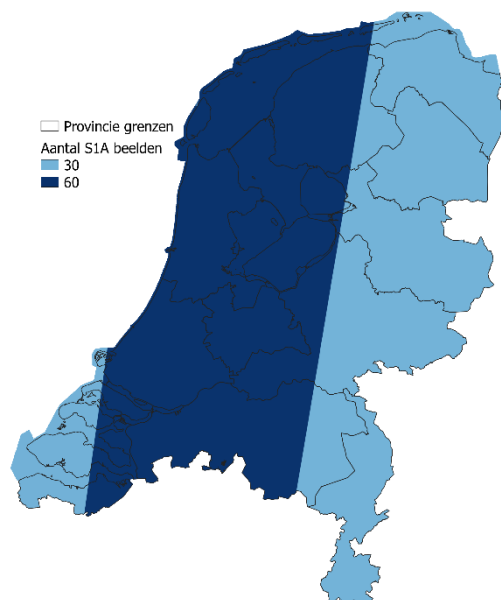


Figuur 47 Voorbeeld van hoe twee SAR beelden kunnen worden gebruikt voor het detecteren van veranderingen (Pirti, et al., 2023).

Voor het afleiden van coherentie wordt gebruik gemaakt van twee radarbeelden, die exact hetzelfde gebied bestrijken, op exact dezelfde manier zijn opgenomen (positie en kijkhoek van de satelliet), en op twee verschillende tijdstippen zijn opgenomen. In het geval van Sentinel-1A is dat 12 dagen. Als aan deze voorwaarden voldaan wordt kan uit het radarsignaal van beide opnames worden afgeleid of het object (grasland perceel), dat de radargolf heeft gereflecteerd, is veranderd. De coherentie is dus eigenlijk een genormaliseerde indicator van gelijkvormigheid tussen de twee opnames. Een lage coherentie (richting 0,2) suggereert dat het object behoorlijk is veranderd (dat kan zijn omdat de structuur/ruwheid is veranderd maar kan ook het gevolg zijn van veranderingen in vocht). Een hoge coherentie (richting 0,8) geeft aan dat het object stabiel is gebleven (onveranderd). Het blijkt dat coherentie zeer geschikt is voor het detecteren van veranderingen ten gevolge van managementactiviteiten zoals ploegen en oogsten.

Figuur 46 toont twee (12-daagse) coherentie beelden van Nederland (westelijke en centrale overpas). In dit geval is de coherentie gebaseerd op twee beelden (27/28 mei en het beeld van 12 dagen eerder). Net als het backscatter beeld betreft het een zwart-wit beeld waarbij de coherentie ligt tussen 0 (zwart) en 1 (wit). De zwarte gebieden in de figuren zijn voornamelijk gebieden met vegetatie en hebben een lage coherentie (omdat vegetatie beweegt door de wind en groeit en daardoor 'verandert'). De lichtgrijze tot witte gebieden zijn over het algemeen kale grond en bebouwde gebieden en hebben een hoge coherentie (deze oppervlakken zijn nagenoeg onveranderd in de 12 dagen).

Net als Sentinel-2 (2A & 2B) volgt Sentinel-1 (1A & 1B) een vaste baan rondom de aarde, zodat elke stukje aarde volgens een vast patroon wordt opgenomen. Om precies te zijn elke 12 dagen. Deze banen zijn anders dan die van Sentinel-2. In deze studie wordt alleen gebruik gemaakt van de westelijke en centrale overpass van S1 (de oostelijke overpass wordt niet meegenomen) (Figuur 48). Omdat S1 geen last heeft van bewolking betekent dat er elke 12-dagen een opname beschikbaar is (~30 beelden per jaar). Op de plekken waar de 2 opnames overlappen is dat elke 6 dagen (~60 beelden per jaar). Wanneer gebruik wordt gemaakt van beide S1A en S1B satellieten wordt dat verdubbeld. Echter omdat S1B eind 2021 is uitgevallen, is besloten om helemaal geen gebruik te maken van S1B zodat een consistente tijdserie beschikbaar is voor de jaren 2017 t/m 2022 op basis van alleen S1A.



Figuur 48 Opname dichtheid (temporele resolutie) sentinel-1A voor Nederland.

Door net als bij Sentinel-2 gebruik te maken van meerdere opeenvolgende beelden door de tijd kan vervolgens gekeken worden naar het verloop van de backscatter en coherentie.

Bijlage C Toelichting RS-markers

Intensiteit landgebruik

Maaien, 1^{ste} maaidatum, scheuren en beweiden

Het maaibeheer van grasland is goed waarneembaar op opeenvolgende satellietbeelden als een plotselinge neergang van de groenindex (NDVI), immers de biomassa (gras) wordt gemaaid en vervolgens van het weiland gehaald. In Figuur 49 de bovenste twee panelen toont het verloop van de NDVI-groenindex voor een aantal percelen. De bovenste twee figuren laten duidelijk een aantal abrupte dalingen in het NDVI-signaal zien (zogenaamde zaagtand). Dit zijn maaimomenten, waarbij de blauwe punt is gedefinieerd als de start van het maaimoment en de rode punt als het einde van het maaimoment. De duur zegt iets over de betrouwbaarheid van het maaimoment. Hoe korter deze duur is des te exacter is de maaisnede-datum te bepalen. Mocht, als gevolg van bewolking er geen opnames beschikbaar zijn, en de duur ('window') langer zijn dan is het moment waarop de maaiactiviteit plaats vond onzekerder.

Het grasland-monitoring algoritme kijkt nu naar maaimomenten waarop de NDVI plotseling afneemt (met minimaal 0,15 NDVI-afname) tussen 1 april (dag nummer 91) en 30 november (334), het groeiseizoen van gras.

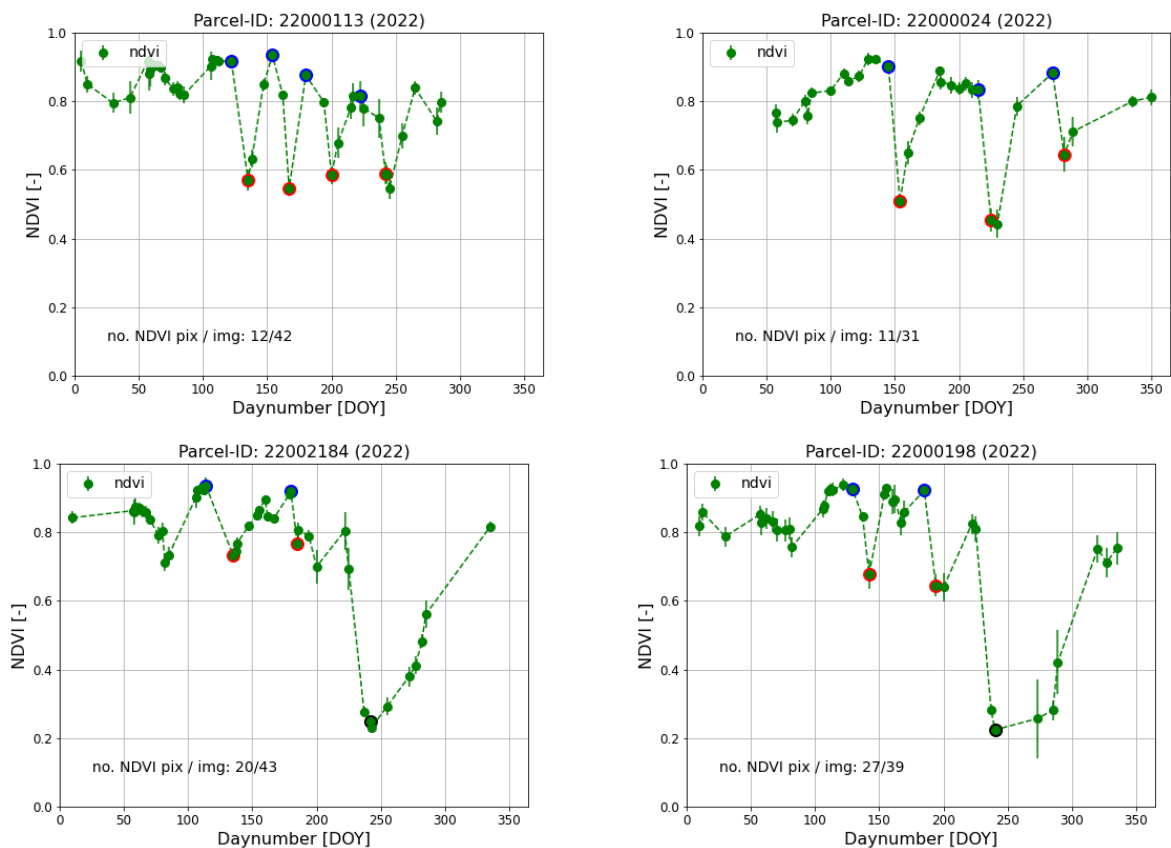
$$N_{maaien} = \sum_{t=1}^K [(NDVI_t - NDVI_{t-1}) < -0.15, NDVI_t > 0.3],$$

met:

- $NDVI_t$: NDVI op tijdstip t
- $NDVI_{t-1}$: NDVI op tijdstip $t-1$
- N_{maaien} : aantal momenten dat $\Delta NDVI < -0,15$ en $NDVI_t > 0,3$

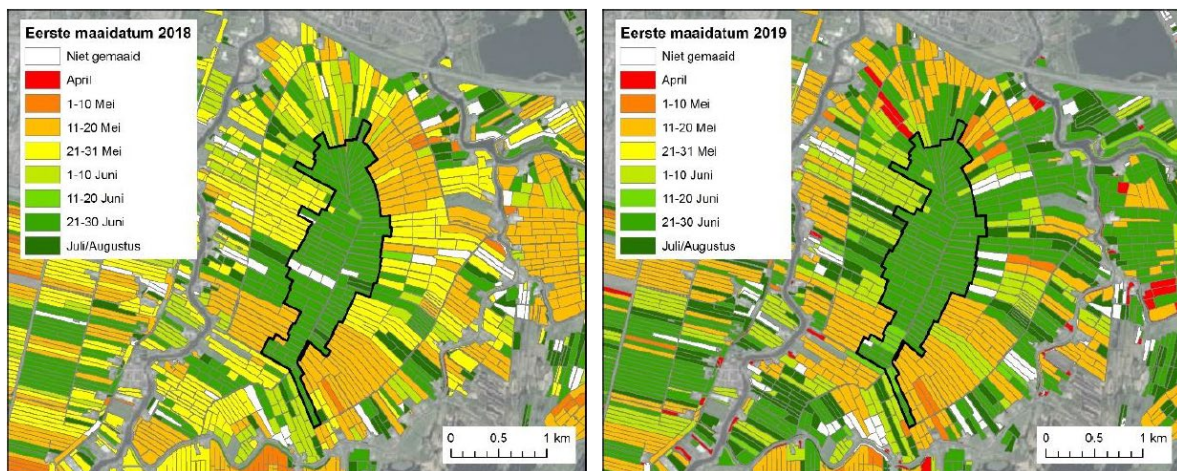
Daarnaast kijkt het ook naar de start en het eind van elk maaimoment.

NB: Voor het algoritme wordt gebruik gemaakt van de 25m NDVI-tijdseries, gebaseerd op voornamelijk Sentinel-2 aangevuld met Landsat-beelden, zodat zoveel mogelijk opnames worden gebruikt in de analyse. Dat betekent wel dat de betrouwbaarheid beperkt is voor met name kleine en smalle percelen. Bij gebruik van alleen de 10 m NDVI-beelden (Sentinel-2) neemt de betrouwbaarheid voor deze percelen toe, echter bestaat de kans dat cruciale (maai)momenten worden gemist omdat er te weinig opnames beschikbaar zijn. Naast maaien kan ook het scheuren of vernietigen van grasland worden waargenomen. Scheuren van grasland wordt gedaan om de kwaliteit van het grasland op peil te houden en te verbeteren. Bij maaien zakt het NDVI-signaal meestal tot rond de 0,5 – 0,6, om daarna weer langzaam toe te nemen. Wanneer de NDVI onder de 0,3 valt, is er sprake van 'kale grond' omstandigheden. Dan heeft er vermoedelijk scheuren of vernietigen van grasland plaats gevonden, zoals in Figuur 49 (onderste twee panelen) is te zien.

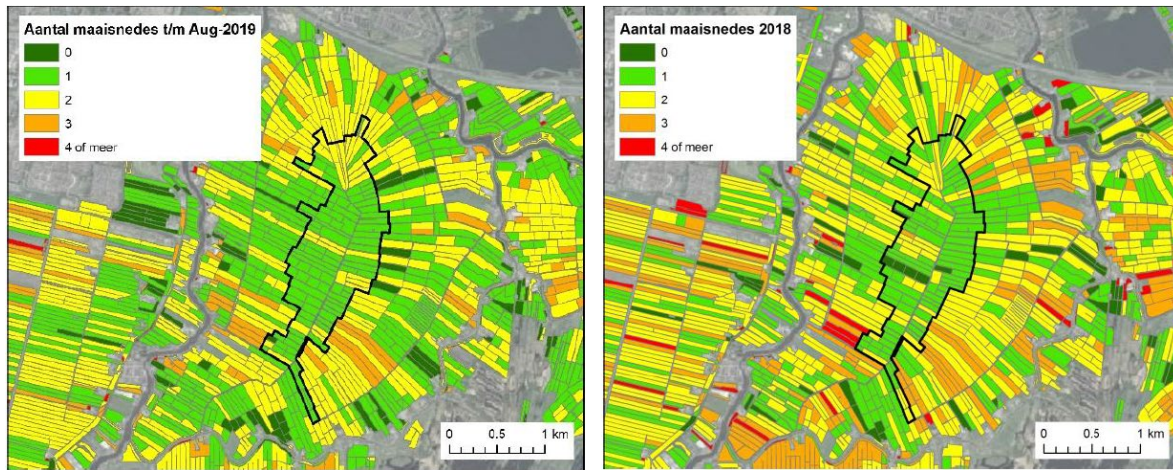


Figuur 49 Voorbeelden van grasland percelen, waarbij het verloop van de NDVI-groenindex is te zien uitgezet tegen dag nummer (DOY). In de NDVI-tijdseries zijn abrupte dalingen te zien (bovenste twee figuren). Dit zijn de maaimomenten, waarbij de blauwe punt is gedefinieerd als de start van het maaimoment en het rode punt als het einde van het maaimoment. Wanneer het NDVI-sigitaal beneden de 0,3 valt is er zeer waarschijnlijk sprake van scheuren (zwarte punt, onderste twee figuren).

Figuur 50 en Figuur 51 geven een voorbeeld van gedetecteerde maaimomenten (1^{ste} maaisnede en het aantal maaisnedes) voor de graslandpercelen in de Ronde Hoep en omgeving. In de kern van de Ronde Hoep ligt een weidevogelreservaat, waar laat wordt gemaaid. Rondom het reservaat is veel agrarisch natuurbeheer afgesloten, waaronder uitgesteld maai-beheer en extensief weiden. De kaartbeelden maken inzichtelijk dat het reservaat het meest extensieve deel van de polder is: er wordt laat en weinig gemaaid. Buiten het reservaat bestaat een divers palet van extensieve tot redelijk intensief gebruikte percelen.



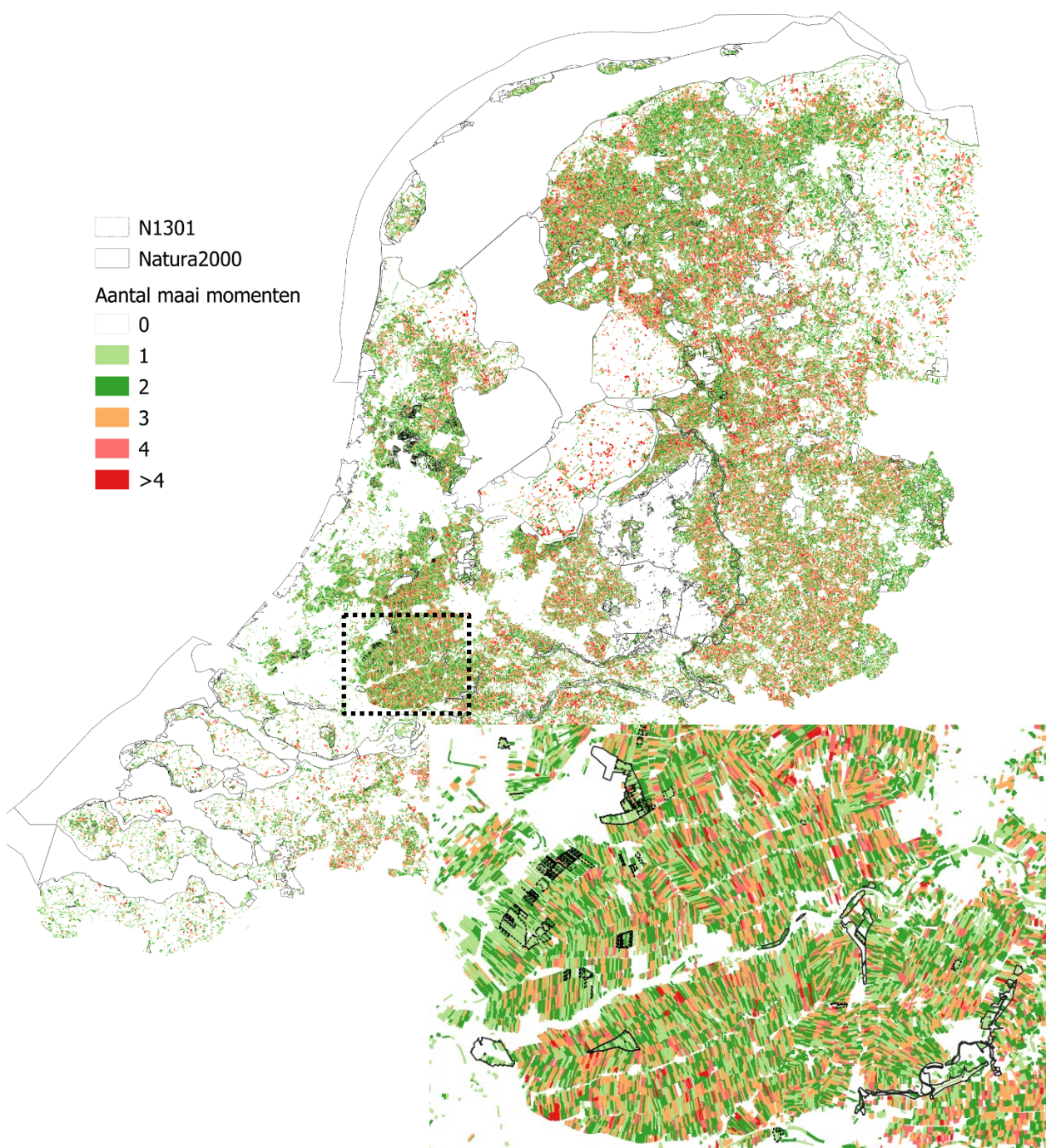
Figuur 50 Datum eerste maaisnede per grasland perceel in 2018 en 2019 voor het Amstelland net onder Ouderkerk a/d Amstel. Het zwart omliggende gebied betreft een weidevogelreservaat. Rondom het reservaat ligt agrarisch natuurbeheer.



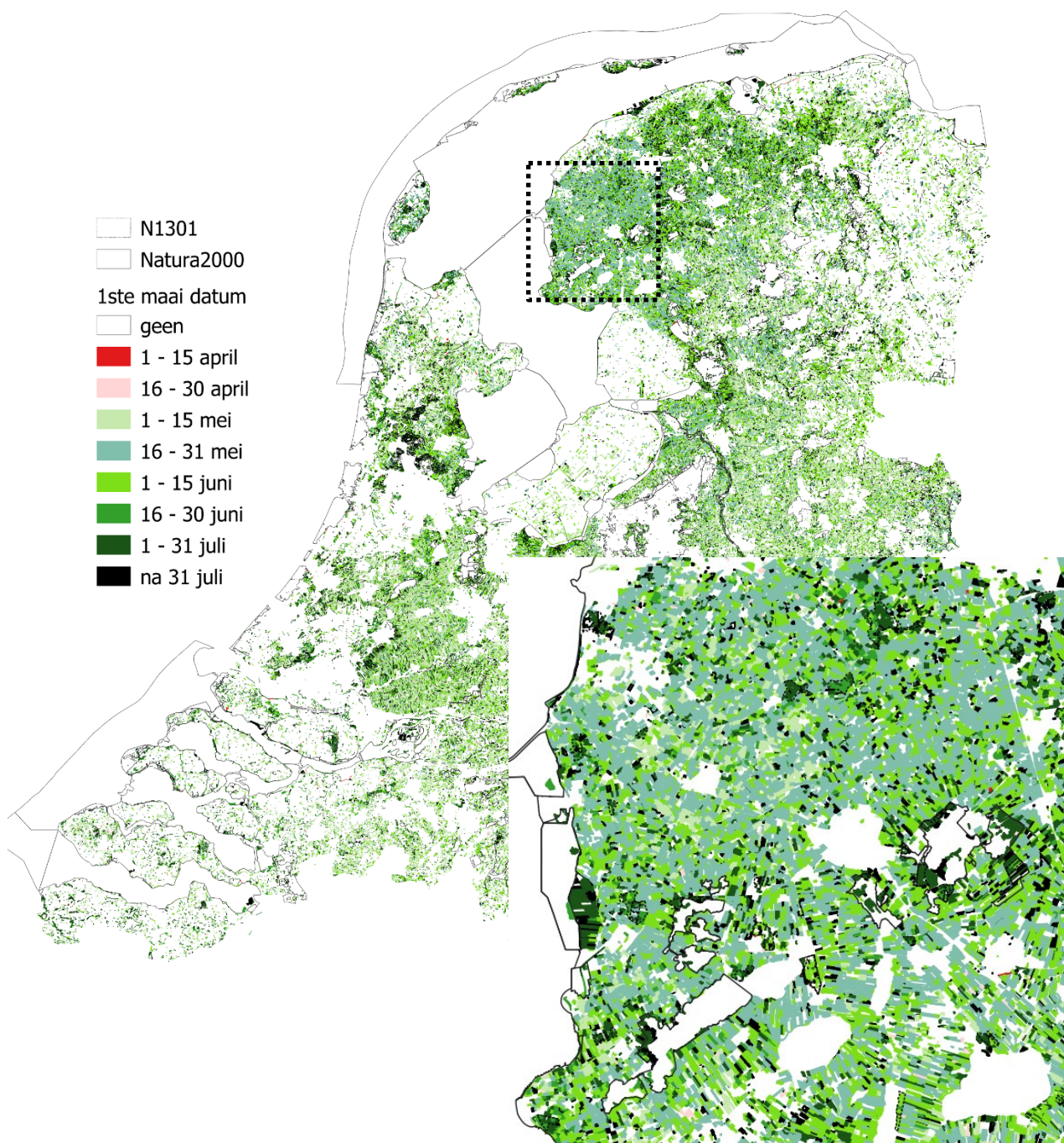
Figuur 51 Het totaal aantal maaisnedes per grasland perceel in 2018 en 2019 (t/m 31 august) voor het Amstelland net onder Ouderkerk a/d Amstel.

Het totaal aantal maaisnedes is een belangrijke indicator die aangeeft hoe intensief of extensief het grasland wordt gebruikt. In de Ronde Hoep worden buiten het reservaat de meeste graslanden één tot drie keer gemaaid in 2018 met uitschieters naar 4x maaien. De graslanden binnen het reservaat worden 1 tot 2 keer gemaaid.

Figuur 52 en Figuur 53 laten het aantal maaisnedes en eerste maaidatum zien in 2022 voor heel Nederland. Duidelijk zichtbaar is dat rustgebieden (N1301) (als de Rond Hoep) en Natura2000 gebieden een lage intensiteit laten zien en een uitgesteld maaibeheer.



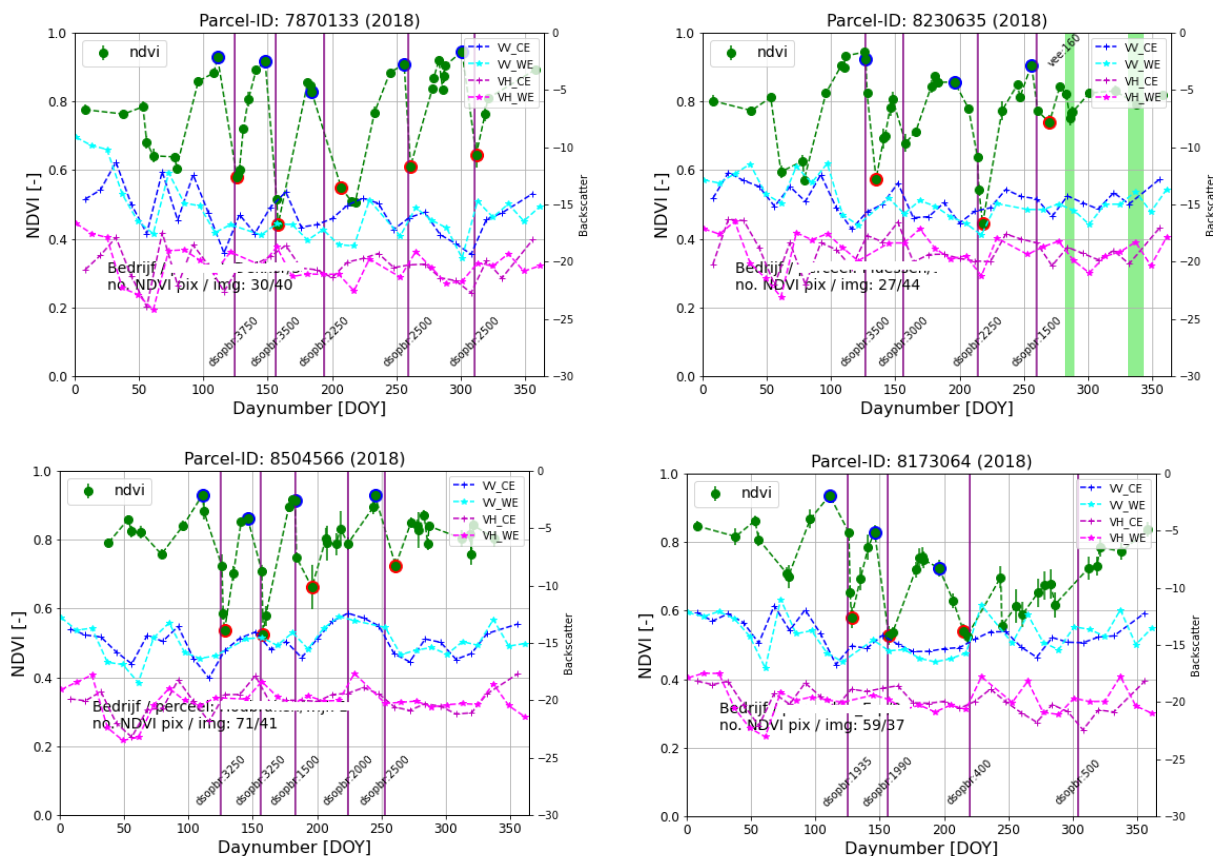
Figuur 52 Aantal maaimomenten grasland percelen in 2022. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.



Figuur 53 Eerste maaidatum grasland percelen in 2022. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.

Validatie maaien en beweiding

Dankzij de medewerking van het Koeien & Kansen (K&K) project, kon voor een aantal percelen in Nederland (2018-2020) het maai-algoritme gevalideerd worden. De afbeeldingen in Figuur 54 (2018) tonen enkele voorbeelden van gemaaide K&K percelen (datum maaisnedes zijn aangeduid met de verticale paarse lijnen) samen met de maairesultaten van het RS-algoritme. Wanneer de K&K maaisnede valt in het maaimoment van het algoritme (tussen blauwe en rode punt) dan heeft het algoritme de maaisnede goed gedetecteerd.

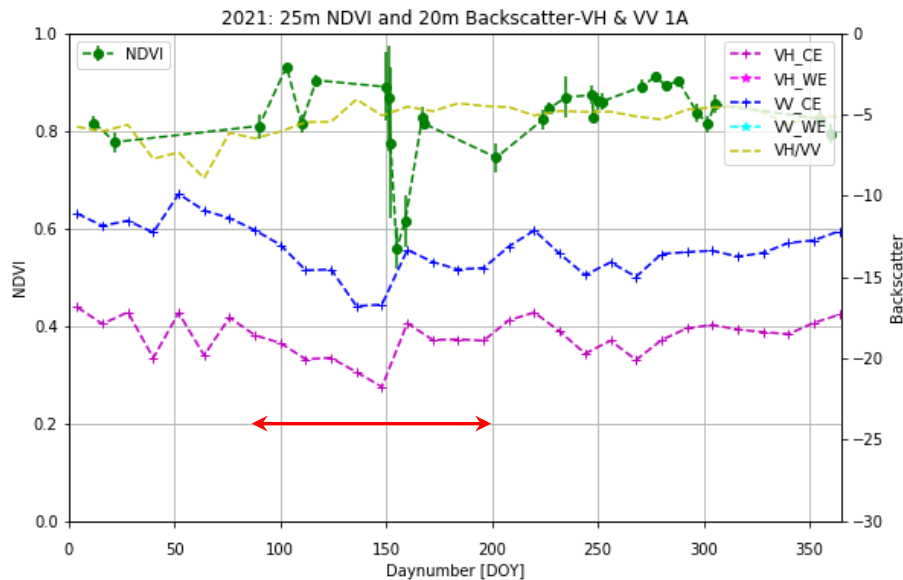


Figuur 54 Voorbeelden van grasland percelen van het Koeien & Kansen project waarbij de maaimomenten van het RS-algoritme zijn vergeleken met maai- en beweidingsdata van K&K.

Binnen de PPS is ook gekeken naar de detectie van beweiding. Echter de variatie in beweiding is groot, in zowel de duur van beweiding (enkele dagen tot weken) als het aantal dieren. Daarnaast worden percelen vaak opgesplitst in sub-percelen waarover het vee rouleert. In de meeste gevallen is de impact van beweiding op het NDVI-signaal onvoldoende om door het algoritme te worden opgemerkt. Wanneer de beweiding zeer intensief is (veel vee in verhouding tot de grootte van het perceel en over een relatief korte periode) dan neemt het NDVI-signaal voldoende af en wordt het opgepikt door het RS-algoritme (als maaisnede).

Management intensiteit op basis van Radar

In 2018 publiceerde Howison *et al.* (2018) een methodiek waarmee op basis van Sentinel-1 radar data vogelvriendelijke weiden kunnen worden gedetecteerd. In een intensief gebruikt grasland perceel vinden voortdurend veranderingen plaats t.g.v. beweiding, maaien en het weer (snel) aangroeien van gras tgv bemesting en goede waterhuishouding. In een extensief gebruikt perceel zullen (veel) minder veranderingen plaats vinden. Een voorbeeld daarvan is te zien in Figuur 55, waar zowel het NDVI-, VH- als VV-tijdseries zijn te zien voor een grasland perceel. Duidelijk is te zien dat rond dag nummer 150 een abrupte afname is te zien van het NDVI-sigitaal. Ook het radar VH- en VV-sigitaal laten dat zien. Dit is het gevolg van een maai moment. Kort daarop is te zien dat de NDVI langzaam herstelt naar de oude waarden.



Figuur 55 Tijdseries van NDVI, radar backscatter VH en VV voor een grasland perceel (2021). Het weidevogel broedseizoen is aangegeven met de rode lijn (1 april – 15 juli).

Howison *et al.* (2018) vonden dat de gebruiksintensiteit van percelen kon worden geschat m.b.v. het Sentinel-1 radarsigitaal. Daarbij werd gebruik gemaakt van backscatter VH-tijdseries (geaggregeerd naar perceel niveau). Door te kijken naar de veranderingen in het backscatter-sigitaal tussen twee opeenvolgende beelden ($\Delta_{\text{backscatter}}$) over het gehele broedseizoen (1 april – 15 juli) en vervolgens de standaarddeviatie te nemen, konden ze het Intensiteit landgebruik van graslanden, en daarmee de kruidenrijkdom/vegetatie type, in Friesland op perceel-niveau in kaart brengen (Figuur 56).



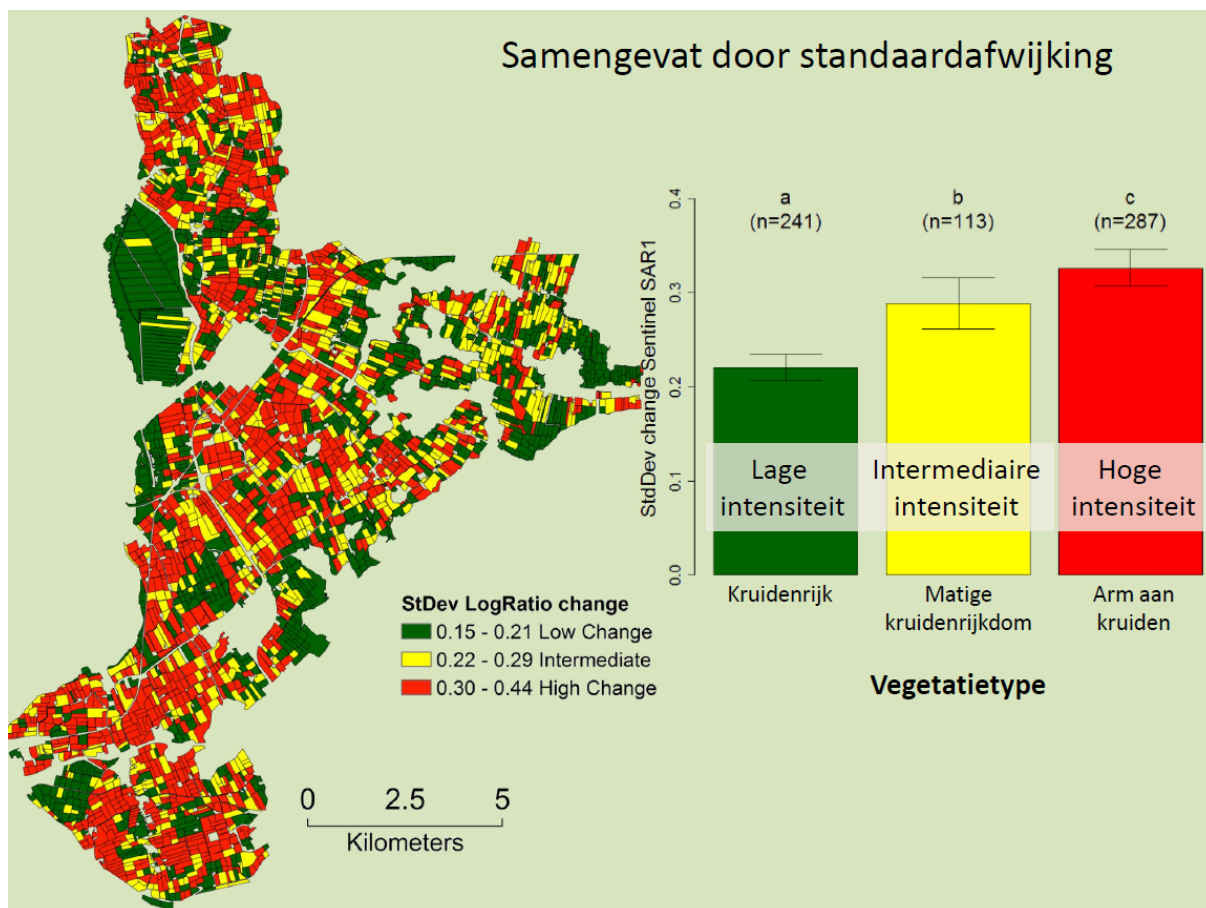
Lage intensiteit (kruidenrijk)



Intermediair (matige kruidenrijkdom)



Hoge intensiteit (arm aan kruiden)



Figuur 56 Landbouw percelen in zuidwest Friesland op basis van Sentinel-1 backscatter signalen in de periode 1 april – 15 juli 2016 (bron: Howison et al., 2018).

In deze studie is de methodiek van Howison toegepast, op basis van de backscatter VH-tijdseries en voor heel Nederland. De reden dat er gebruik wordt gemaakt van het backscatter-VH signaal is dat deze gevoeliger is voor vegetatie dan het VV-signaal.

Standaarddeviatie van de 12-daagse veranderingen in het σ_{VH}^0 signaal over de periode 1 april (dag nummer 91) – 15 juli (196) wat overeenkomt met het weidevogel broedseizoen, gedefinieerd als (Howison et al., 2018):

$$std(\text{delta} - VH) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2},$$

met:

N : het aantal radarbeelden

$x_i = \ln\left(\frac{\sigma_{VH,t+12}^0}{\sigma_{VH,t}^0}\right)$, VH-verandering (delta-VH) tussen tijdstip t en $t+12$ dagen

$\bar{x}$: gemiddelde VH-verandering

$\sigma_{VH,t}^0$: backscatter VH (lineair) op tijdstip t

$\sigma_{VH,t+12}^0$: backscatter VH (lineair) op tijdstip $t+12$ dagen

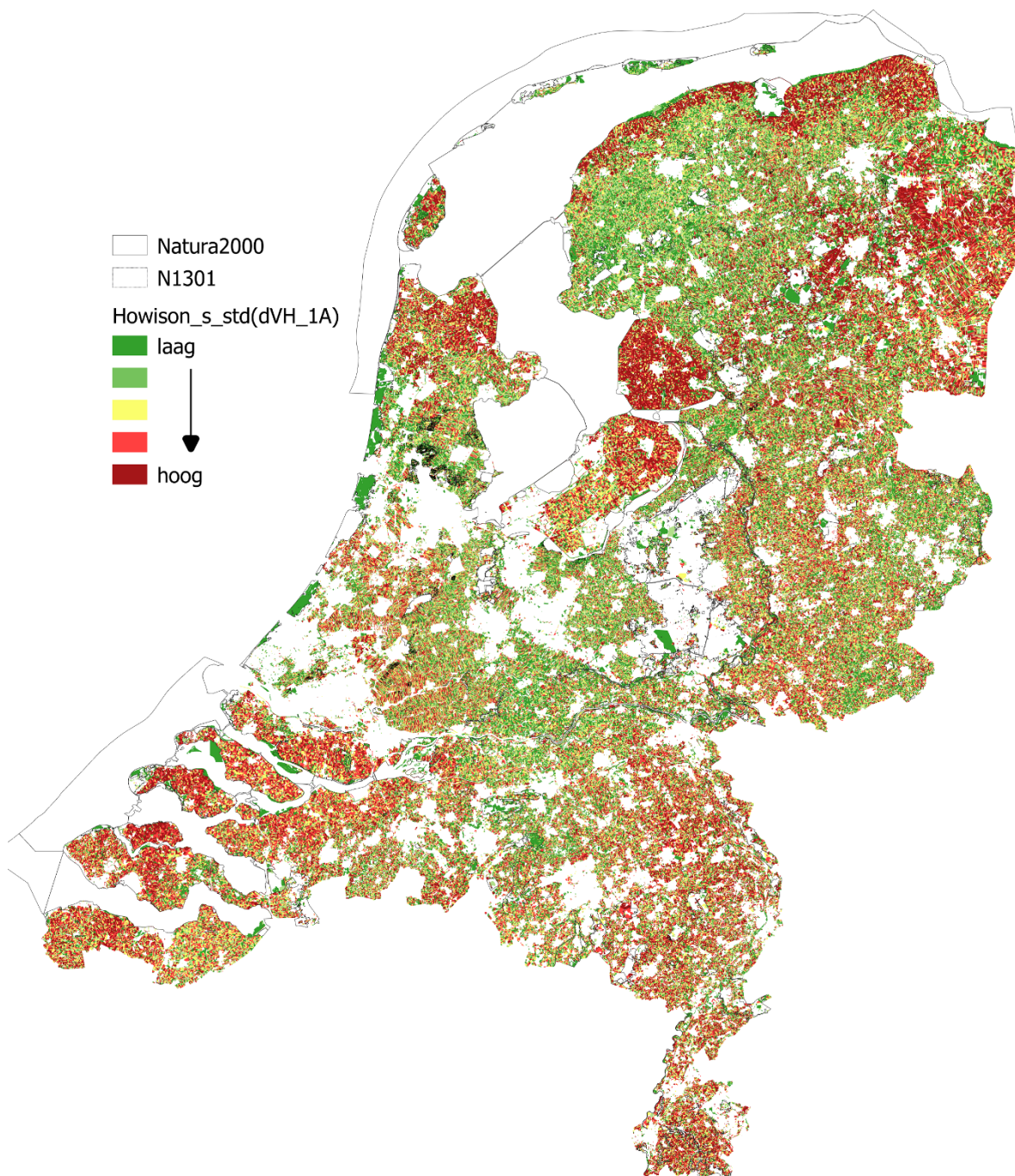
Uiteindelijk is de std(delta-VH) afgeleid voor de volgende perioden:

- *Weidevogel seizoen (1 april – 15 juli)*
- *Groeiseizoen (1 april – 30 november)*
- *Hele jaar (1 januari – 31 december)*

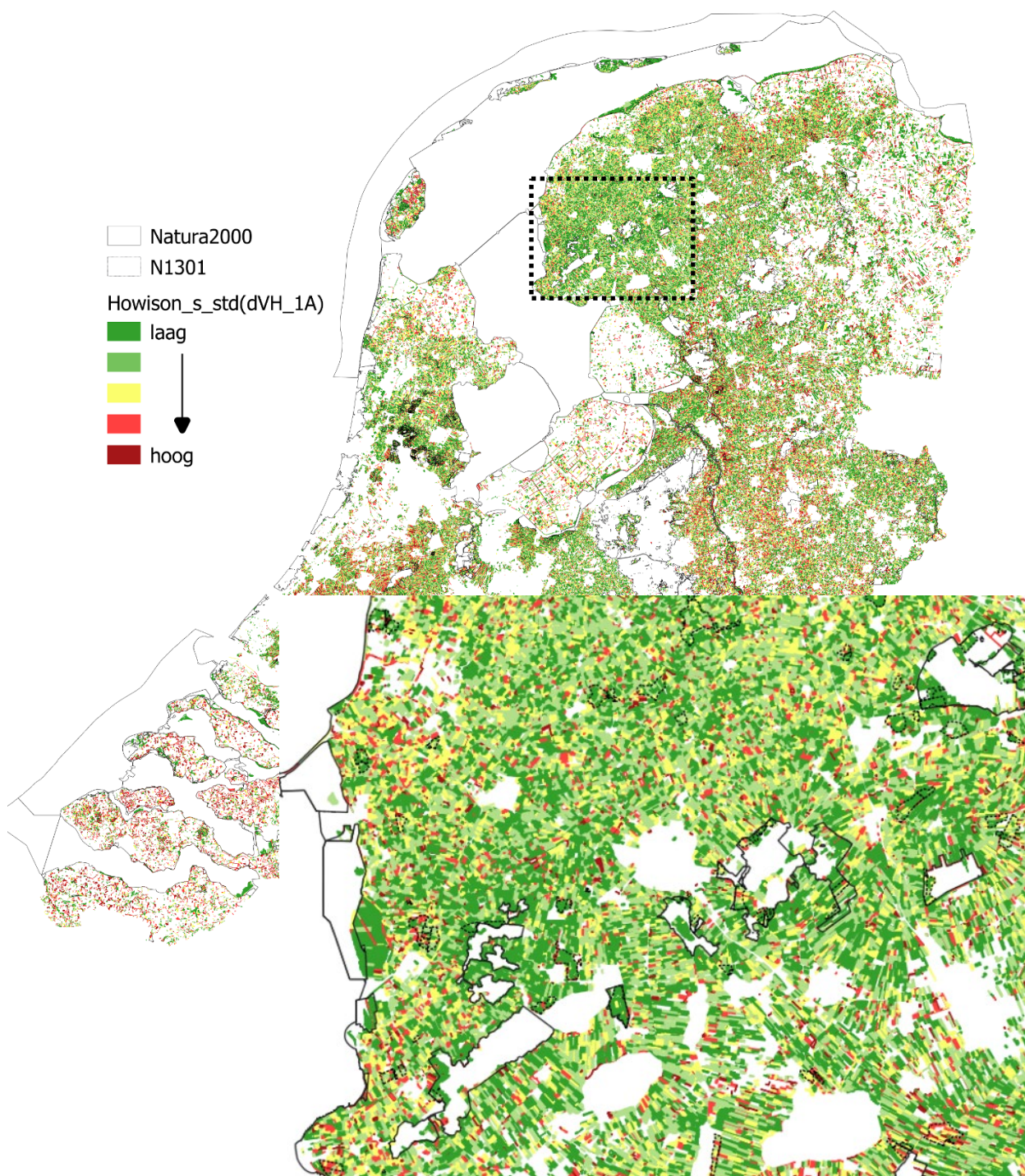
De geselecteerde Sentinel-1A radarbeelden bestaan uit een westelijke (WE) en een centrale (CE) overpass (Figuur 47). Voor beide overpasses is std(delta-VH) apart afgeleid, en zijn als volgt samengevoegd om tot een landelijk dekkende std(delta-VH) te komen:

$$std(\text{delta} - VH) = \sqrt{\frac{std(\text{delta}-VH)_{CE}^2 + std(\text{delta}-VH)_{WE}^2}{2}}$$

Figuur 57 laat de intensiteit van het landgebruik zien van alle geregistreerde BRP-percelen, d.w.z. bouwland, grasland en natuurgebieden, waarbij de intensiteit (2021, weidevogel broedseizoen) is uitgedrukt tussen hoog (rode kleur) en laag (groene kleur). Duidelijk is te zien dat met name bouwland percelen een rode kleur hebben (Flevopolder, Noordoostpolder, Zeeland en noordelijk kust strook in Friesland en Groningen). De patronen in Figuur 57 komen goed overeen met Li *et al.* (2023), al zijn die gebaseerd op 2016. Figuur 58 is identiek aan Figuur 57, maar laat alleen de grasland en natuurgebieden zien. Hierin is te zien dat met name rustgebieden (N1301) en Natura2000 gebieden een lage intensiteit zien in vergelijking met de andere graslanden.



Figuur 57 Landgebruiksintensiteit op basis van waargenomen (12-daagse) veranderingen in VH gedurende het weidevogelbroedseizoen (1 april – 15 juli) van 2021 voor ALLE BRP-percelen (d.w.z. grasland, bouwland en natuurterreinen). Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.



Figuur 58 Als Figuur 57 maar dan alleen voor BRP-grasland percelen en BRP-natuurterreinen. Zwart omliggende gebieden zijn weidevogel rust- en Natura2000 gebieden.

Naast de Howison-indicator voor Intensiteit landgebruik zijn ook de volgende radar-gebaseerde intensiteitsindicatoren afgeleid:

- *std(delta-VV), à la de Howison-indicator (VH-sigitaal) maar dan op basis van het VV-sigitaal*

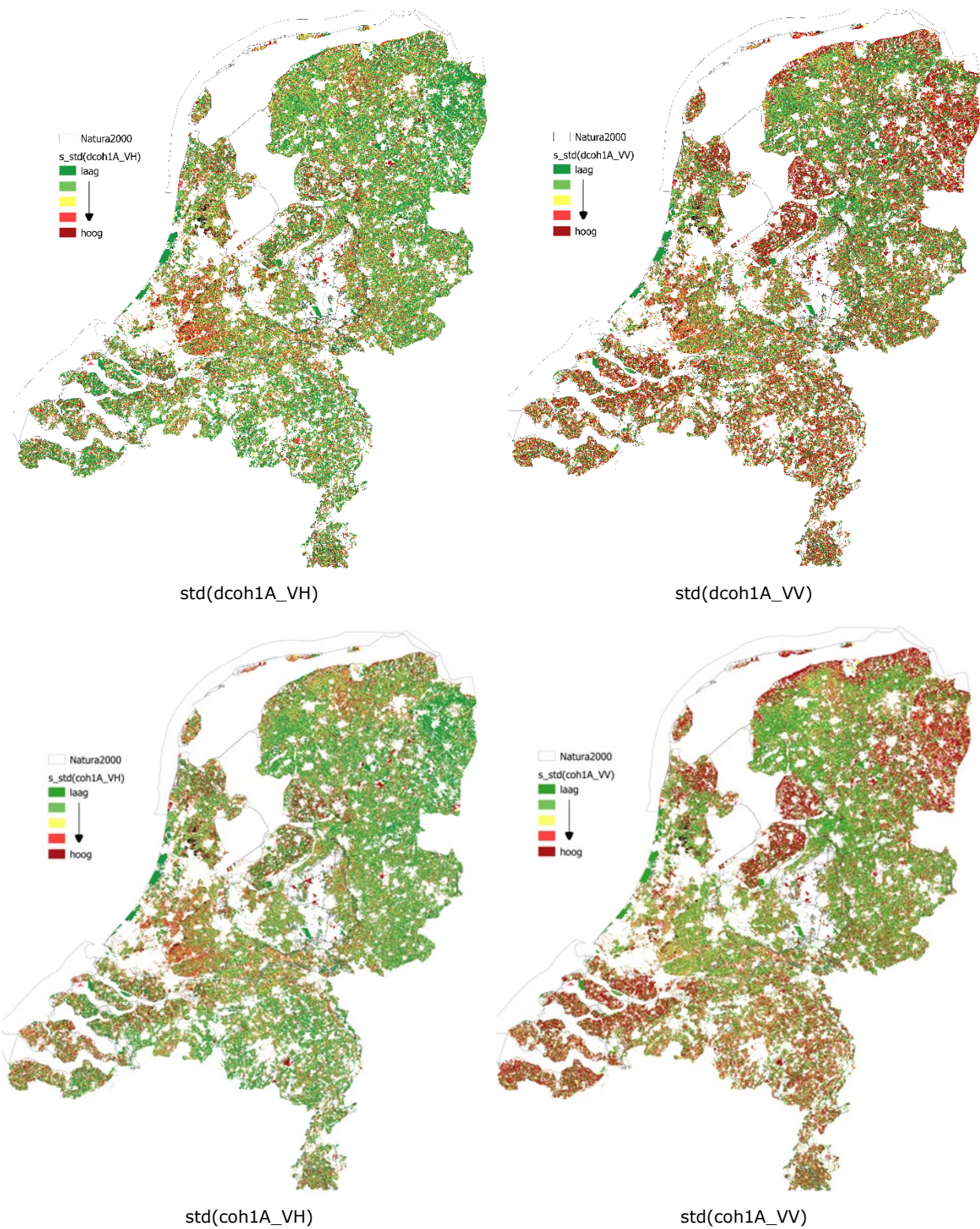
Het VV-sigitaal is minder gevoelig voor vegetatie, waardoor delta-VV meer gevoelig is voor verandering in het vochtgehalte in de bodem (en plant).

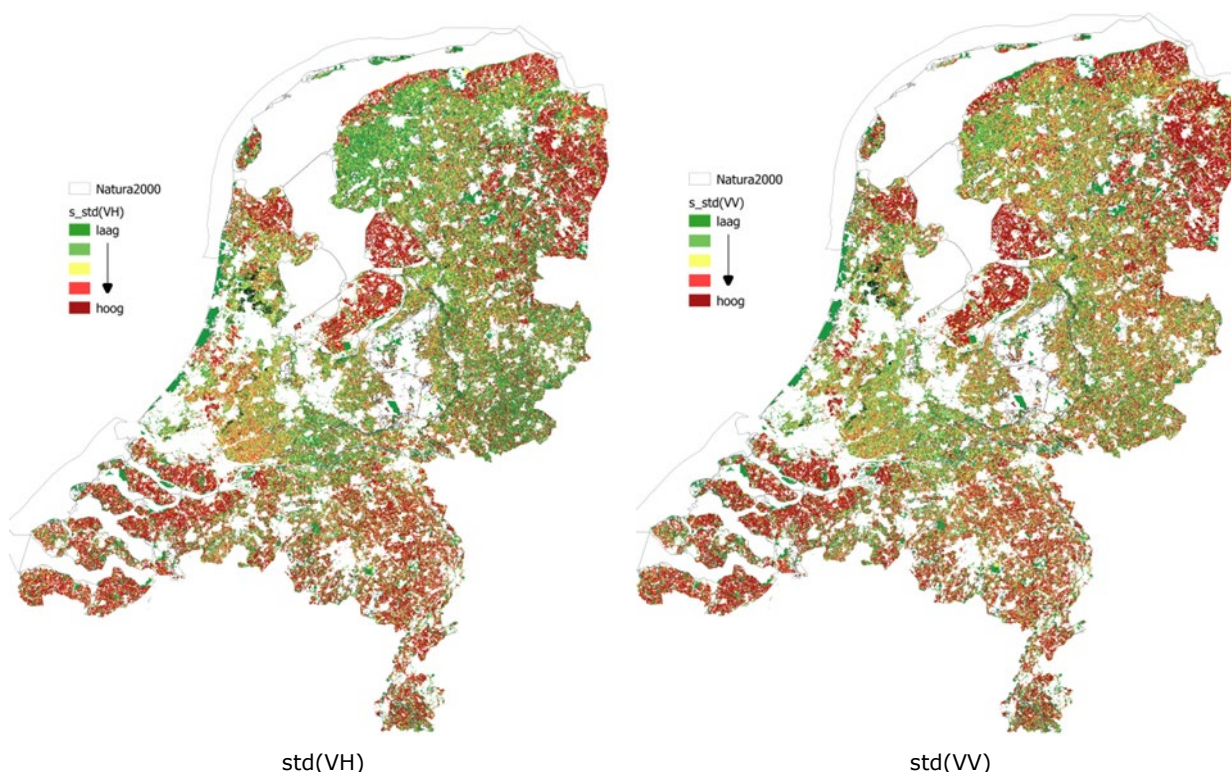
- Standaarddeviatie van het backscatter VH- en VV-sigitaal, std(VV), std(VH).

Het coherentie sigitaal:

- Standaarddeviatie van het coherentie VH- en VV-sigitaal, std(coh_VH), std(coh_VV).
- Standaarddeviatie van veranderingen in het coherentie VH- en VV-sigitaal, std(dcoh_VH), std(dcoh_VV).

Voorbeelden zijn te zien in Figuur 59.





Figuur 59 Standaarddeviatie van $dcoh1A_VH$, $dcoh1A_VV$, $coh1A_VH$, $coh1A_VV$, VH en VV gedurende het weidevogelbroedseizoen (1 april – 15 juli) van 2022, voor ALLE BRP percelen (d.w.z. grasland, bouwland en natuurterreinen). Zwart omliggende gebieden zijn weidevogel rust- en Natura2000 gebieden.

MASD

De MASD (Mean Absolute Spectral Dynamics) geeft de spectrale veranderingen aan tussen twee opeenvolgende beelden, en kan net als de Howison-methode worden gebruikt als indicator voor Intensiteit landgebruik. In het geval van Sentinel-2 kan de MASD worden afgeleid voor alle tien de spectrale banden (4x10m en 6x20m banden) op basis van verzamelde opnames binnen een bepaalde periode. MASD beschrijft de spectrale dynamiek van het land oppervlak door de tijd, en is gedefinieerd als (Franke *et al.*, 2012; Bekkema en Eleveld, 2018):

$$MASD = \frac{1}{N} \sum_{t=1}^N \left(\frac{1}{m} \sum_{i=1}^m |\rho_i^t - \rho_i^{t+1}| \right),$$

Met:

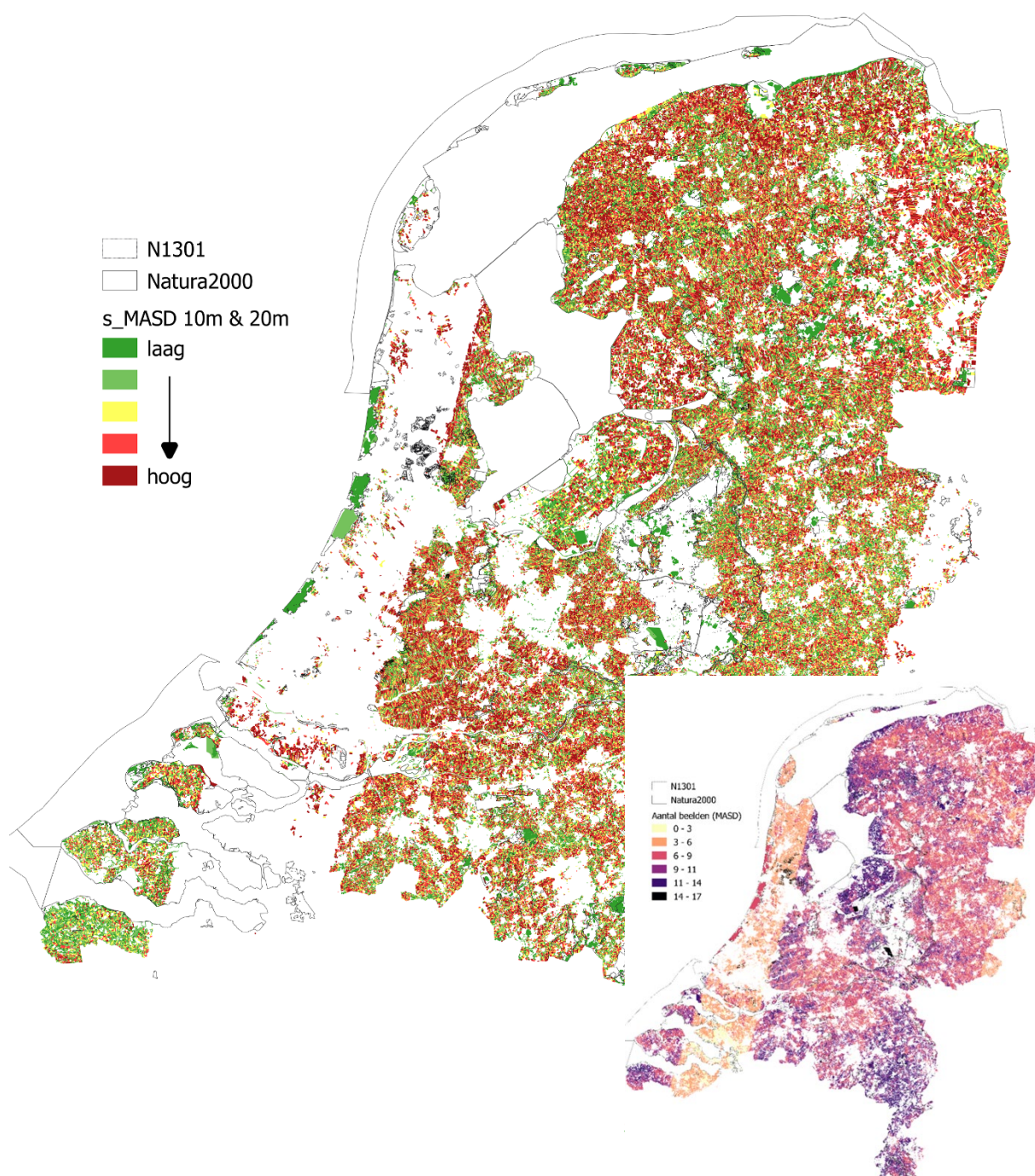
ρ_i^t : reflectie band i oppervlak op tijdstip t

ρ_i^{t+1} : reflectie band i oppervlak op tijdstip $t+1$

N : het aantal beelden

m : aantal banden (10: 4x10m (R,G,B,NIR) en 6x20m (1xNIR, 3xRed-Edge, 2xSWIR))

Figuur 60 laat de afgeleide MASD zien voor 2022 (weidevogel broedseizoen). Een lage MASD (groene kleur) suggereert weinig veranderingen door de tijd en is een indicatie voor een lage landgebruiksintensiteit. Een hoge MASD (rode percelen) is suggereert juist veel veranderingen door de tijd tgv management (begrazing, maaien) en is dus een indicatie voor een hoge landgebruiksintensiteit. Figuur 60 is alleen maar gebaseerd op S2-beelden en laat duidelijk zien dat de dekingsgraad niet voldoende is voor heel NL. Het oostelijke deel, de kuststrook van Noord- en Zuid-Holland en delen van Zeeland hebben onvoldoende beelden om een betrouwbare MASD te bepalen.



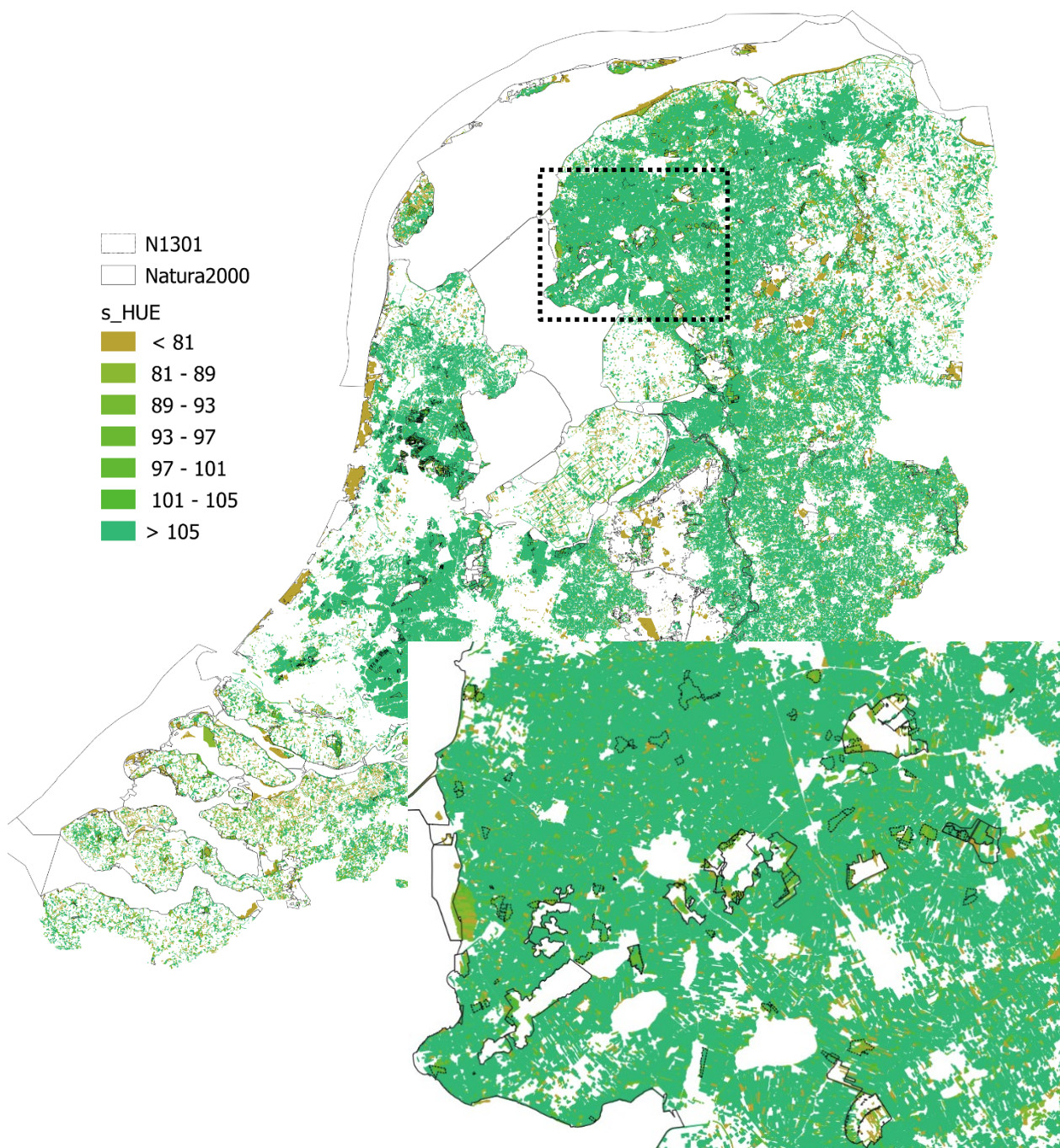
Figuur 60 MASD-landgebruiksintensiteit op basis van waargenomen veranderingen in de 10m en 20m spectrale reflecties gedurende het weidevogelbroedseizoen (1 april – 15 juli) van 2022, voor alle BRP-percelen. Figuur rechtsonder toont het aantal beelden waarop MASD is gebaseerd. Zwart omliggende gebieden zijn weidevogel rust- en Natura2000 gebieden.

Visuele kenmerken

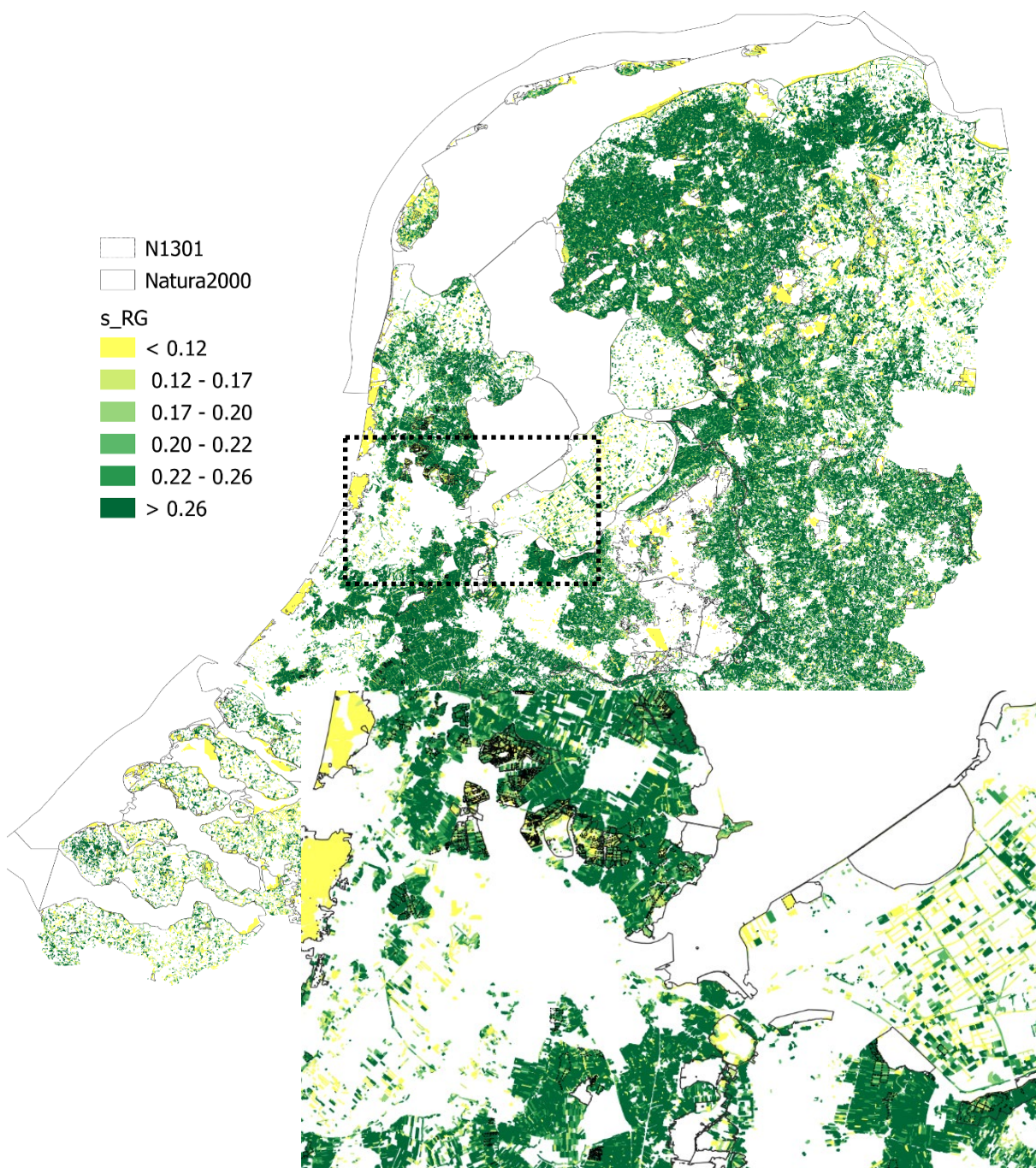
Kleur

Kruidenrijke graslanden hebben vaak een afwijkende kleur ten opzichte van de productieve graslanden. Deze kleurafwijking is het gevolg van veranderingen in de samenstelling van de vegetatie door de aanwezigheid van kruiden, bloemen en andere grassoorten. Op basis van de gemeten (Rood, Groen, Blauw) reflecties kan de kleur van het gras worden bepaald. Om de analyse te vereenvoudigen zijn de RGB-waarden vertaald naar een andere kleurenschaal, de zgn. Hue-angle, waarbij de kleur (kleurtint) wordt uitgedrukt als één waarde tussen 0 en 360 graden. Er is zowel gekeken naar de gemiddelde kleur van een perceel op jaarbasis als specifiek in het voorjaar. Om te voorkomen dat maai activiteiten de kleur van het perceel verstoren worden uitsluitend beelden tussen 1 maart en 31 mei meegenomen. Een voorbeeld is te zien in Figuur 61.

Naast de Hue-angle is ook gekeken naar de verhouding tussen de rode en groene reflectie (R/G-ratio). Voor de gangbare graslanden ligt deze R/G-ratio rond $\sim 0,7$ (< 1 , omdat de groen-reflectie dominant is). De R/G-ratio wijkt af zodra grasland percelen minder intens groen zijn ten gevolge van de aanwezigheid van andere vegetatie (bloemen, kruiden en andere grassoorten). Net als bij de Hue-angle is gekeken naar de gemiddelde ratio op jaarbasis en voor de periode 1 maart – 31 mei (zie Figuur 62). Pluspunt van de rode en groene reflecties is dat deze minder gevoelig zijn voor atmosferische variaties (en de atmosfeercorrecties die worden gebruikt bij het verwerken van de beelden). Tot slot is ook de B/G-ratio is bepaald, echter omdat met name de blauwe band erg gevoelig is voor atmosferische variaties is het vermoeden dat deze minder bruikbaar zal zijn.



Figuur 61 Gemiddelde Hue-angle (graden) van BRP-grasland percelen en natuurterreinen, gedurende het weidevogelbroedseizoen (1 april – 15 juli) van 2021. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.



Figuur 62 Gemiddelde verhouding rood-groen reflectie van BRP-grasland percelen en natuurterreinen gedurende het weidevogelbroedseizoen (1 april – 15 juli) van 2021. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.

Spectrale diversiteit/ heterogeniteit

Spectrale diversiteit (of optische diversiteit) wordt vaak gebruikt als indicator voor soortenrijkdom en is gerelateerd aan de ruimtelijke variaties van gemeten reflecties van een set RS-pixels, in dit geval binnen een grasland perceel. Studies die de spectrale diversiteit hebben vergeleken met vegetatieopnames van soortenrijkdom laten zien dat over het algemeen hoge CV-waarden corresponderen met hoge soortenrijkdom (Wang *et al.*, 2016; Wang *et al.*, 2018a, b; Gholizadeh *et al.*, 2018; Xu *et al.*, 2022). Figuur 63 is een voorbeeld van een kruidenrijk grasland (links) en een productief grasland (rechts). Duidelijk is te zien dat het kruidenrijke perceel meer spectrale/kleur variatie (heterogeniteit) vertoont binnen het perceel ten gevolge van de verscheidenheid aan kruiden en planten, terwijl het productief grasland perceel strak homogeen (gemaaid) “grasgroen” van kleur is.



Figuur 63 Voorbeeld van een kruidenrijk grasland versus een productie grasland bestaande uit vnl. Engels raaigras.

Een goede indicator voor de spectrale diversiteit van een perceel is de som van de variatiecoëfficiënt (coefficient of variation, CV) van elk S2-band, in dit geval 10 banden. Uiteindelijk is de gemiddelde CV bepaald per jaar en voor de periode 1 april – 15 juli. De (gemiddelde) spectrale diversiteit of heterogeniteit van een perceel kan m.b.v. de coëfficiënt van variatie (CV) (Wang *et al.*, 2016; Xu *et al.*, 2022) worden bepaald:

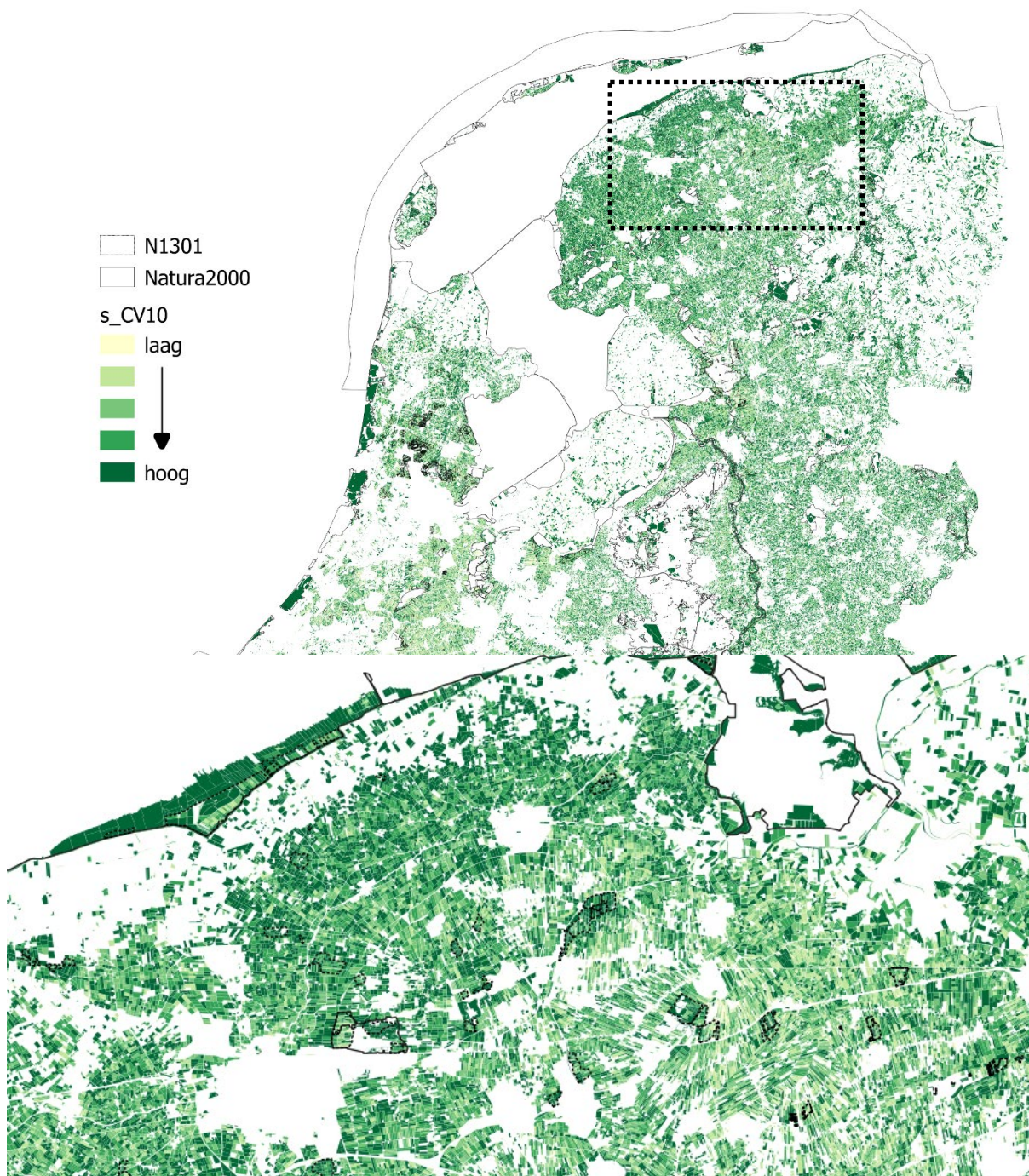
$$\overline{CV}_{10} = \frac{1}{N} \sum_{t=1}^n \frac{1}{m} \sum_{i=1}^m \left(\frac{std_i}{mean_i} \right),$$

met:

- *std_i*: standaardafwijking reflectie band *i* van alle pixels binnen het grasland perceel op tijdstip *t*.
- *mean_i*: gemiddelde reflectie band *i* van alle pixels binnen het grasland perceel op tijdstip *t*.
- *m*: aantal banden (= 4x10m (R, G, B, NIR) + 6x20m (1xNIR, 3xRed-Edge, 2xSWIR)).
- *N*: het aantal beelden over een bepaalde periode: jaar, weidevogel broedseizoen, groeiseizoen.
- $\overline{CV}_{10}$: gemiddelde CV op basis van de tien S2-band.

Daarnaast kan de spectrale diversiteit ook worden afgeleid uit de NDVI-tijdseries (CV_{NDVI}).

NB: probleem van gefaseerd maaien in één BRP-grasland perceel. In Figuur 64 wordt de gemiddelde CV_{10} voor de weidevogel periode (1 april – 15 juli, 2021) voor alle graslanden en natuurterreinen in Nederland weergegeven.

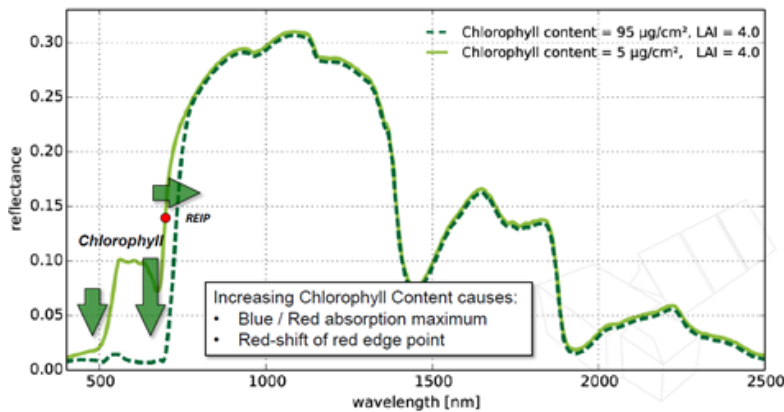


Figuur 64 Gemiddelde coëfficiënt van variatie CV (10 S2-banden) gedurende de weidevogel broedperiode (1 april – 15 juli) van 2021 voor alle BRP-grasland percelen en natuurterreinen. Zwart omliggende gebieden zijn weidevogel rust- en Natura2000 gebieden.

Productiviteit

S2REP

S2REP (Sentinel-2 Red-Edge Position) maakt gebruik van de unieke eigenschappen van de Red-Edge banden van Sentinel-2. Deze banden zijn gevoelig voor chlorofyl en stikstof concentraties in het blad en de conditie van het gewas (Scheffler *et al.*, 2014; Bekkema en Eleveld, 2018). S2REP is het punt waarop de spectrale response curve van het gewas het steilst is. De exacte locatie (uitgedrukt in nm) van dit punt varieert per gewas maar is ook gevoelig voor chlorofyl en stikstof concentraties. Een verschuiving van lage naar hoge chlorofyl concentraties komt overeen met een verschuiving van lage naar hoge S2REP-waarden.



Figuur 65 Verschuiving Red-Edge positie als gevolg van chlorofyl gehalte in de plant (Scheffler et al., 2014).

S2REP biedt daardoor de mogelijkheid om onderscheid te kunnen maken tussen bemeste en niet-bemeste graslanden. Daarnaast is S2REP minder gevoelig voor verzanding dan de NDVI, echter omdat de ruimtelijk resolutie van de Red-Edge banden 20 m is, zijn hele kleine/smalle percelen minder nauwkeurig te monitoren. Bekkema en Eleveld (2018) vonden dat met name S2REP-beelden opgenomen in het voorjaar de beste resultaten opleverden, dat is voordat het maaiseizoen begint. In dit onderzoek is de S2REP bepaald voor de maanden maart t/m juli.

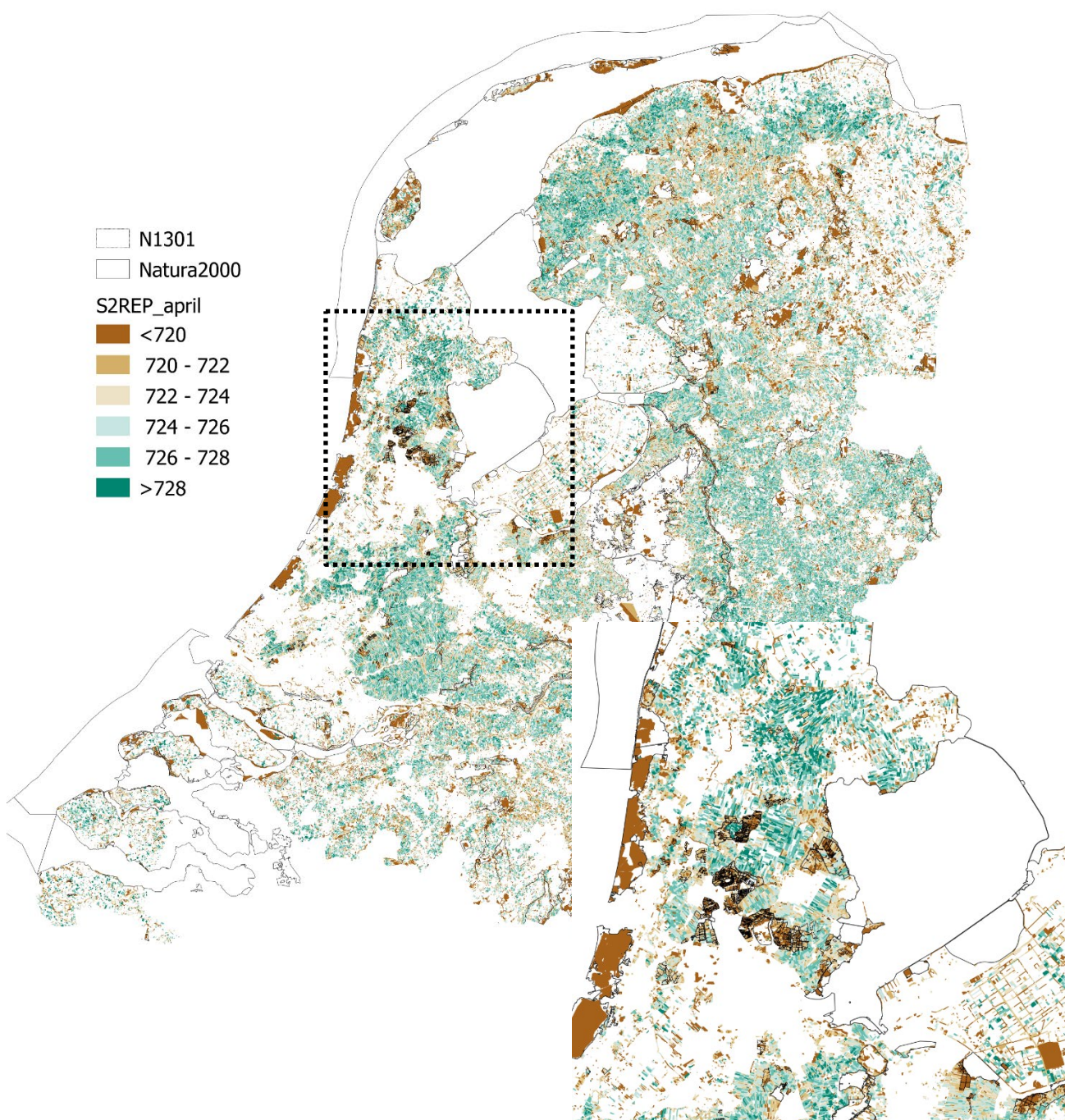
Red-Edge positie (ook wel bekend als Red-Edge inflection point, uitgedrukt in nm) kan als volgt worden bepaald:

$$S2REP = 705 + 35 \left(\frac{\frac{(B7+B4)}{2} - B5}{(B6-B5)} \right),$$

met:

- B4: Rood reflectie 10m (665 nm)
- B5: Red-Edge 1 reflectie 20 m (705 nm)
- B6: Red-Edge 2 reflectie 20 m (740 nm)
- B7: Red-Edge 3 reflectie 20 m (783 nm)

Figuur 66 laat de S2REP-waarde zien voor april 2022 voor alle grasland percelen en natuurterreinen. Duidelijk is te zien dat veel natuurterreinen (Natura2000) en weidevogelrustgebieden (N1301) een lagere S2REP-waarde hebben dan de productieve graslanden.



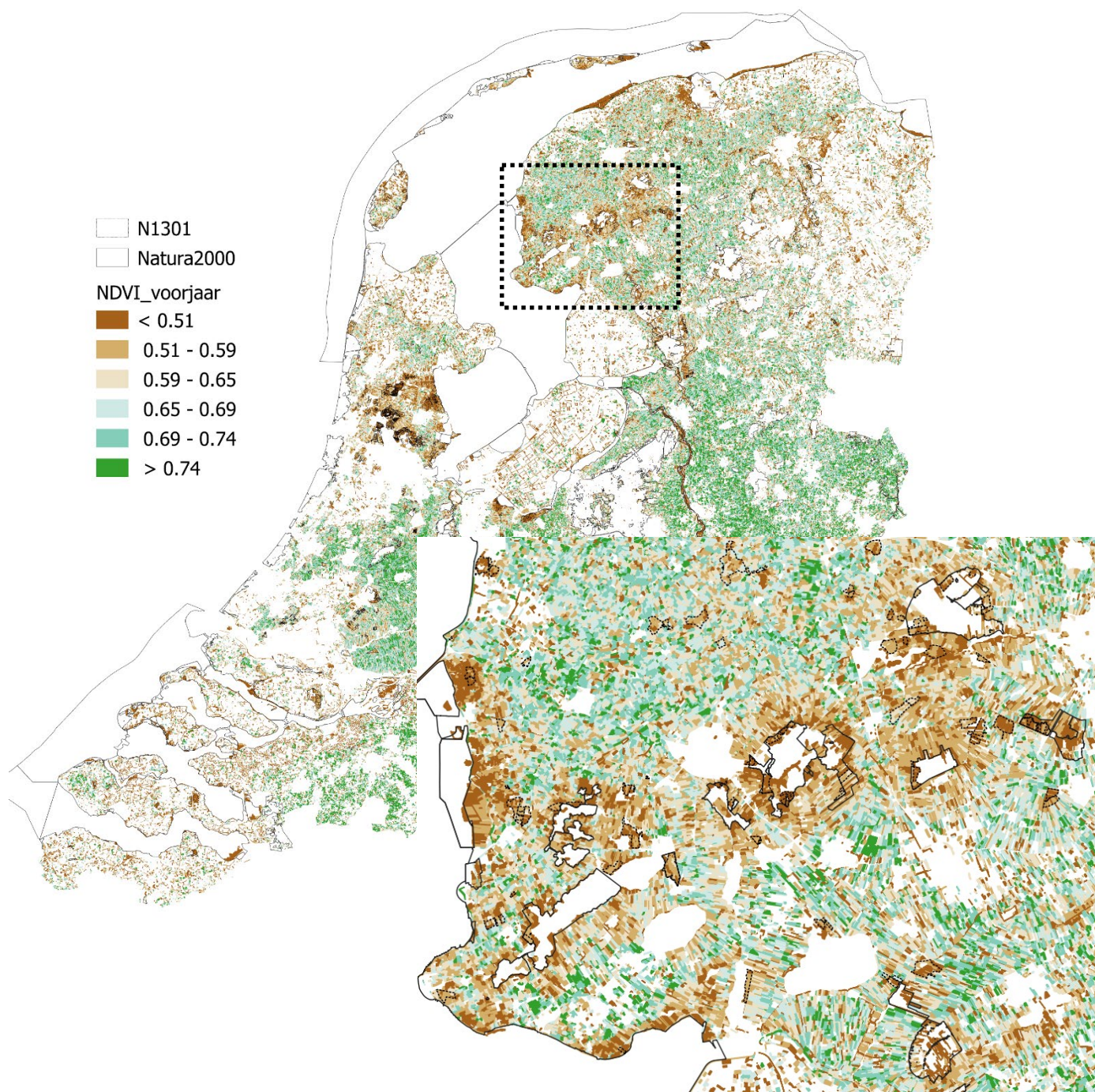
Figuur 66 S2REP van BRP-grasland percelen en natuurterreinen voor april 2022. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.

Voorjaar NDVI

Tijdens de winterperiode staat de groei van grasland over het algemeen stil. Soms is er zelfs een afname zichtbaar als de winter streng is. Zodra het voorjaar begint komt de groei langzaam weer op gang.

Uitgebreide analyse van de Groenmonitor beelden (van Groenmonitor.nl) laat zien dat productieve/gangbare graslanden aan het eind van de winter begin voorjaar (eind maart), over het algemeen een hogere start NDVI-waarden hebben dan natuurlijke grasland (voorjaar NDVI). Dit heeft o.a. te maken met de betere groeicondities (betere nutriënten- en vochthuishouding), waardoor deze graslanden beter door de winter komen en sneller van start kunnen, resulterend in een hogere biomassa (NDVI) waarden.

Figuur 67 laat de NDVI-waarde zien van eind maart 2018. Duidelijk zichtbaar zijn de lage NDVI-waarden van natuurterreinen (Natura2000) en weidevogelgebieden (N1301) in tegen stelling tot de hoge NDVI-waarde van productieve graslanden. De ruimtelijke patronen komen ook goed overeen met het voorafgaande S2REP-beeld (Figuur 66).



Figuur 67 NDVI-eind maart 2018. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.

Cumulatieve NDVI in combinatie met maaien

Op basis van het aantal maaimomenten kan ook een schatting worden gemaakt van de grasproductie door te kijken naar de cumulatieve waarde van de NDVI (als proxy), dat wil zeggen door de NDVI-afnames bij elk maaimoment te sommeren (grasscheuren wordt genegeerd). Naast de NDVI-fluctuaties ten gevolge van maaien ($dNDVI < -0,15$) bevat het NDVI-sigitaal ook kleine fluctuaties ($-0,025^7 > dNDVI > -0,15$), ook voor deze fluctuaties is de cumulatieve NDVI bepaald. Deze fluctuaties zijn gelabeld onder beweiding, al moet worden benadrukt dat de detectie van beweiding moeilijk is (zie Sectie Validatie maaien en beweiding) (Stumpf *et al.*, 2020).

$$CUM_{NDVI-maaien} = \sum_{t=1}^K |(NDVI_t - NDVI_{t-1})[(NDVI_t - NDVI_{t-1}) < -0,15, NDVI_t > 0,3]|,$$

$$CUM_{NDVI-beweiding} = \sum_{t=1}^K |(NDVI_t - NDVI_{t-1})[-0,025 > (NDVI_t - NDVI_{t-1}) > -0,15]|,$$

$$CUM_{NDVI} = CUM_{NDVI-maaien} + CUM_{NDVI-beweiding}$$

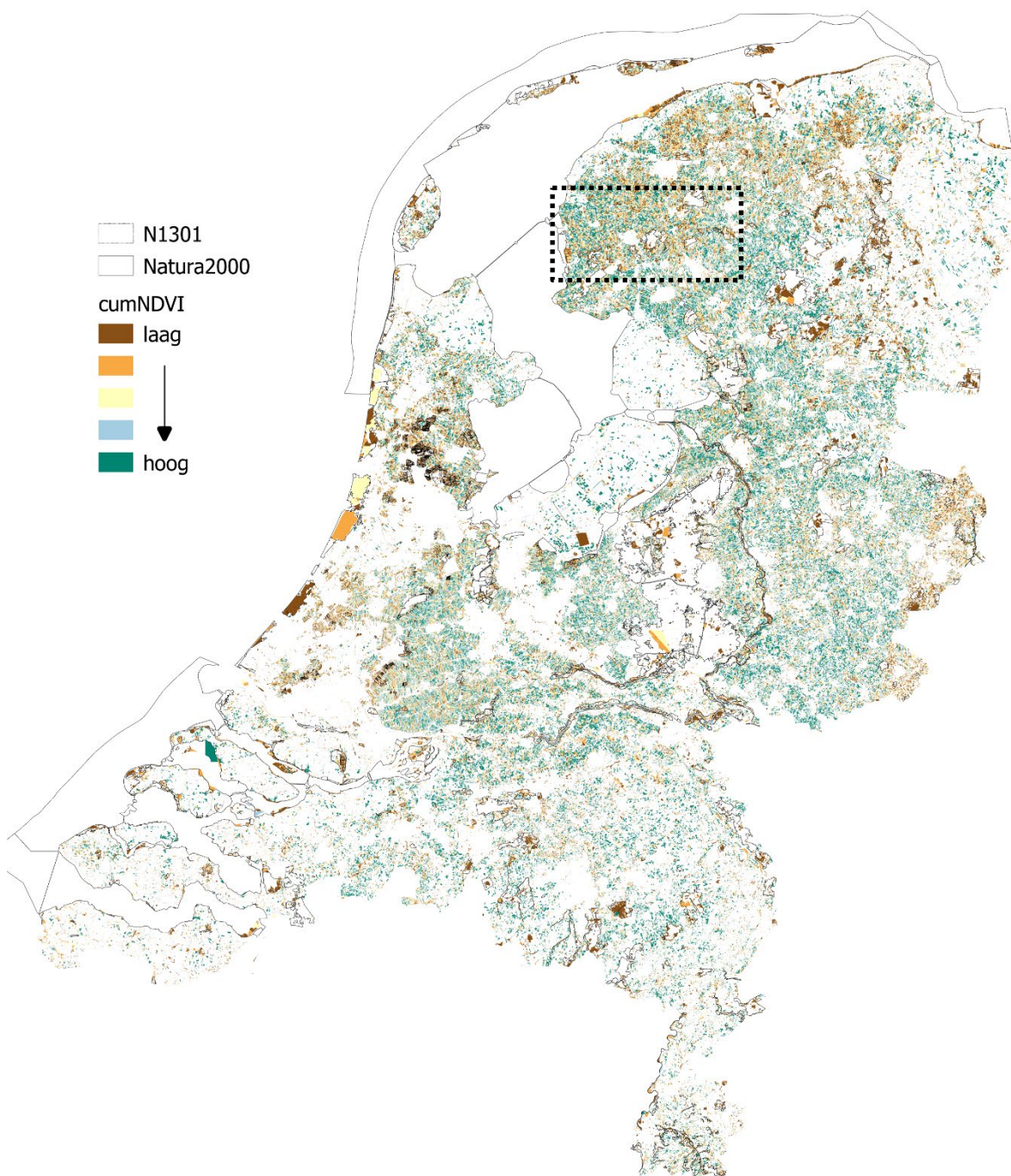
met:

- $NDVI_t$: NDVI op tijdstip t
- $NDVI_{t-1}$: NDVI op tijdstip $t-1$
- $CUM_{NDVI-maaien}$: totale NDVI-afname voor momenten dat $dNDVI < -0,15$ en $NDVI_t > 0,3$
- $CUM_{NDVI-grazen}$: totale NDVI-afname voor momenten dat $dNDVI > -0,15$ en $dNDVI < -0,025$
- CUM_{NDVI} : som van $CUM_{NDVI-maaien}$ en $CUM_{NDVI-grazen}$

Wanneer sprake is van scheuren in plaats van maaien, worden deze niet meegenomen.

In Figuur 68 wordt de totale CUM_{NDVI} voor 2022 weergegeven. De cumulatieve NDVI toont veel overeenkomsten met het aantal maaimomenten (Figuur 68). Percelen met veel maaimomenten laten ook hoge CUM_{NDVI} -waarden zien. Ook hier zijn duidelijk zichtbaar de lage CUM_{NDVI} -waarden van natuurterreinen (Natura2000) en weidevogelgebieden (N1301) in tegenstelling tot de hoge CUM_{NDVI} -waarden van productieve graslanden.

⁷ Hele kleine fluctuaties tussen $-0,025$ en 0 NDVI worden genegeerd, aangezien deze vnl. worden veroorzaakt door atmosferische effecten.

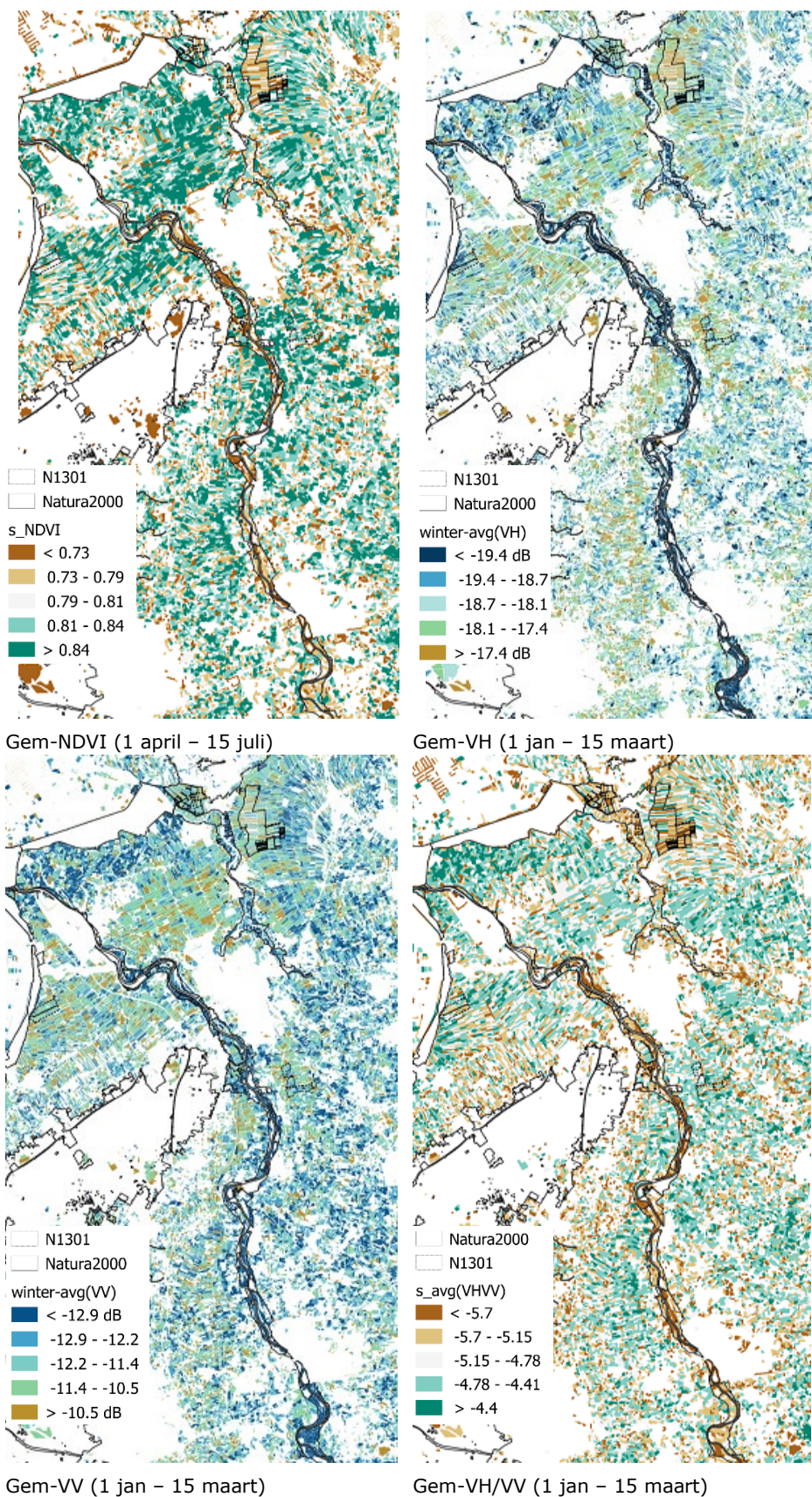


Figuur 68 Cumulatieve NDVI (gekoppeld aan maaimomenten) 2022. Zwart omlijnde gebieden zijn weidevogel rust- en Natura2000 gebieden.

Ruwheid/biomassa/vocht

Zoals toegelicht in Sectie Radar satelliet data wordt de mate van backscatter (of wel verstrooiing van het radarsignaal) bepaald door de ruwheid van het oppervlak, het vochtgehalte in zowel de bodem en, indien aanwezig, vegetatie. Daarom wordt er gesproken van zogenaamde ruwheid/biomassa/vocht beelden. Ondanks het feit dat het radar verstrooiingsmechanisme wordt beïnvloed door meerdere factoren zijn de ruwe backscatter-beelden wel bruikbaar voor het onderscheiden van kruidenrijke en productieve grasland.

Dat wordt gedemonstreerd in Figuur 69, waarin de winter gemiddelde backscatter-VH en VH, en de verhouding VH/VV (verschil tussen VH en VV) langs de IJssel en omstreken. Ter verduidelijking is ook de gemiddelde NDVI te zien (voor het weidevogelbroedseizoen (1 april – 15 juli)). En tot slot zijn ook de weidevogel rustgebieden en Natura2000 gebieden zichtbaar. Wat als eerste opvalt is dat het NDVI-beeld (linksboven) veel overeenkomsten vertoont met het VH/VV-beeld (rechtsonder). Lage NDVI-waarden in de uiterwaarde als gevolg van weinig vegetatie en/of natte omstandigheden komen overeen met lage VH/VV-waarden. Dat de omstandigheden nat zijn in de uiterwaarde (met name in de winterperiode) is te zien aan de zeer lage backscatter waarden in zowel VV als VH. Verder is te zien dat het VV-sigitaal wat sterker is dan het VH-sigitaal, waarbij de VH-verstrooiing (polarisatie) wordt veroorzaakt door vegetatie.



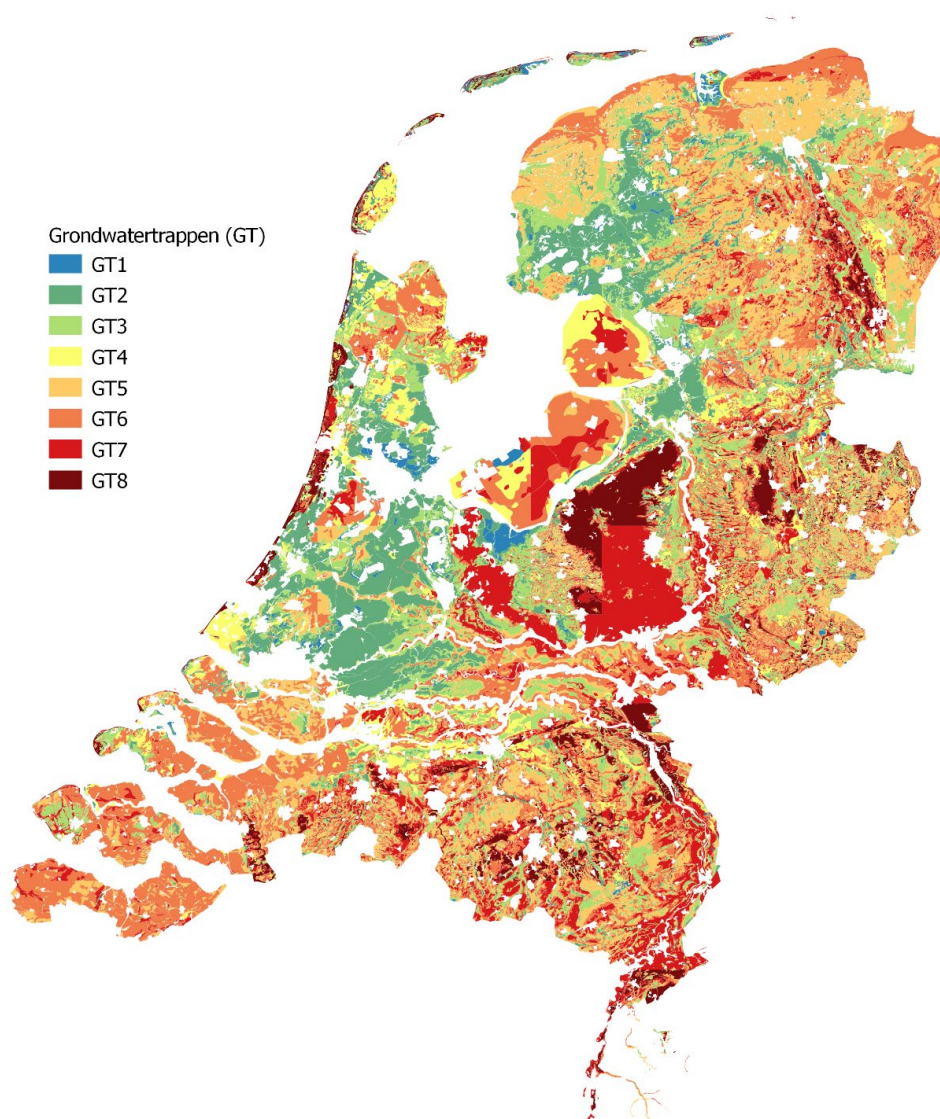
Figuur 69 Gemiddelde NDVI voor het weidevogelbroedseizoen (1 april – 15 juli, linksboven), winter gemiddelde backscatter-VH (rechtsboven), backscatter-VV (linksonder) en de verhouding VH/VV (verschil tussen VH en VV, rechtsonder) langs de IJssel en omstreken.

Co-variabelen

Grondwatertrappen (GT)

Tabel 14 GT-klasse (GHG: Gemiddeld Hoogste Grondwaterstand, GLG: Gemiddeld Laagste Grondwaterstand, centimeters t.o.v. maaiveld).

GT-klasse	GHG 'winter' [cm]	GLG 'zomer' [cm]
1 (I)	<20	<50
2 (II)	<40	50-80
3 (III)	<40	80-120
4 (IV)	>40	80-120
5 (V)	<40	>120
6 (VI)	40-80	>120
7 (VII)	>80	-
8 (VIII)	>140	-

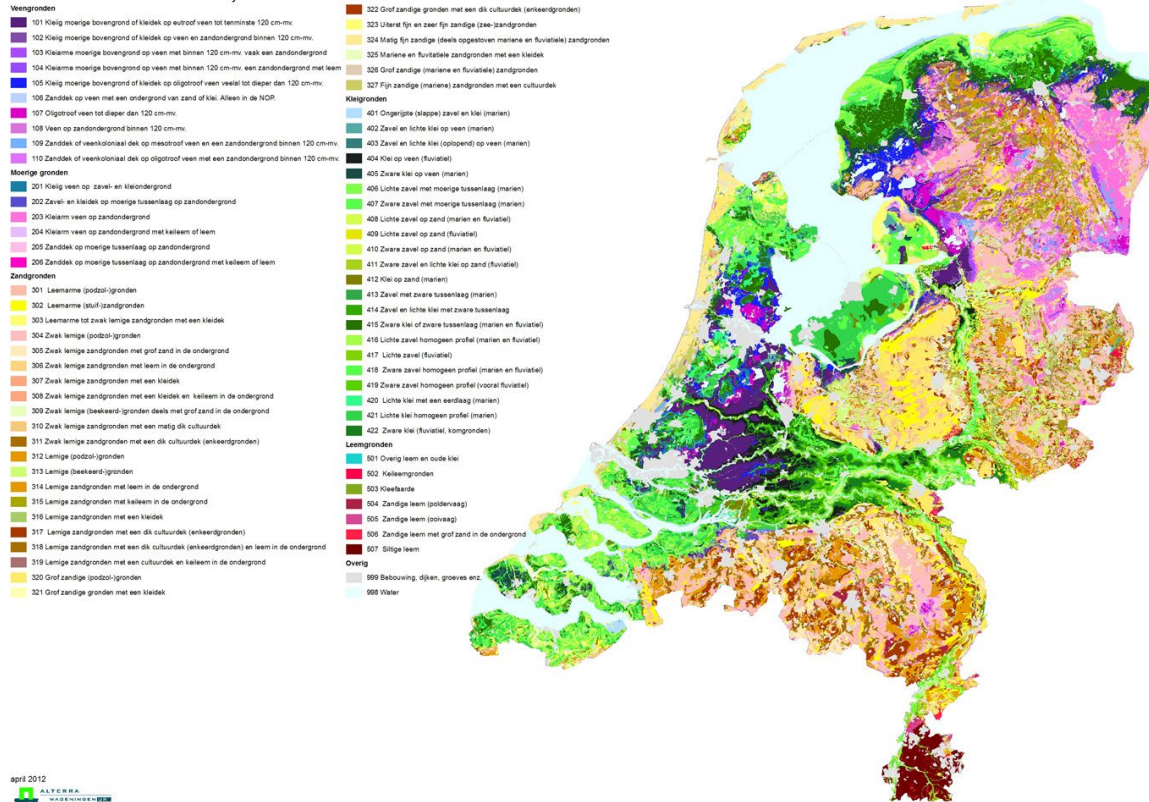


Figuur 70 Grondwatertrappen kaart Nederland.

BOFEK schematisatie

Het BOFEK (Bodem Fysische Eenheden Kaart) 2012-bestand geeft de bodem fysische schematisatie weer van Nederland. Dit bestand is gecreëerd door aan de 315 bodemeenheden behorende bij de Bodemkaart van Nederland, schaal 1:50.000, waterretentie- en doorlatendheidskarakteristieken uit de Staringreeks toe te kennen. Met een model zijn voor deze eenheden vervolgens functionele kenmerken berekend. Op grond van verwantschap in functionele kenmerken zijn de 315 bodemeenheden geclusterd in 72 bodem fysische eenheden en afgebeeld in de nieuwe Bodem Fysische Eenheden Kaart (BOFEK2012).

BOFEK2012: De "nieuwe" Bodem Fysische Eenheden Kaart



Figuur 71 Bodem fysische eenheden kaart (BOFEK2012), 72 eenheden.

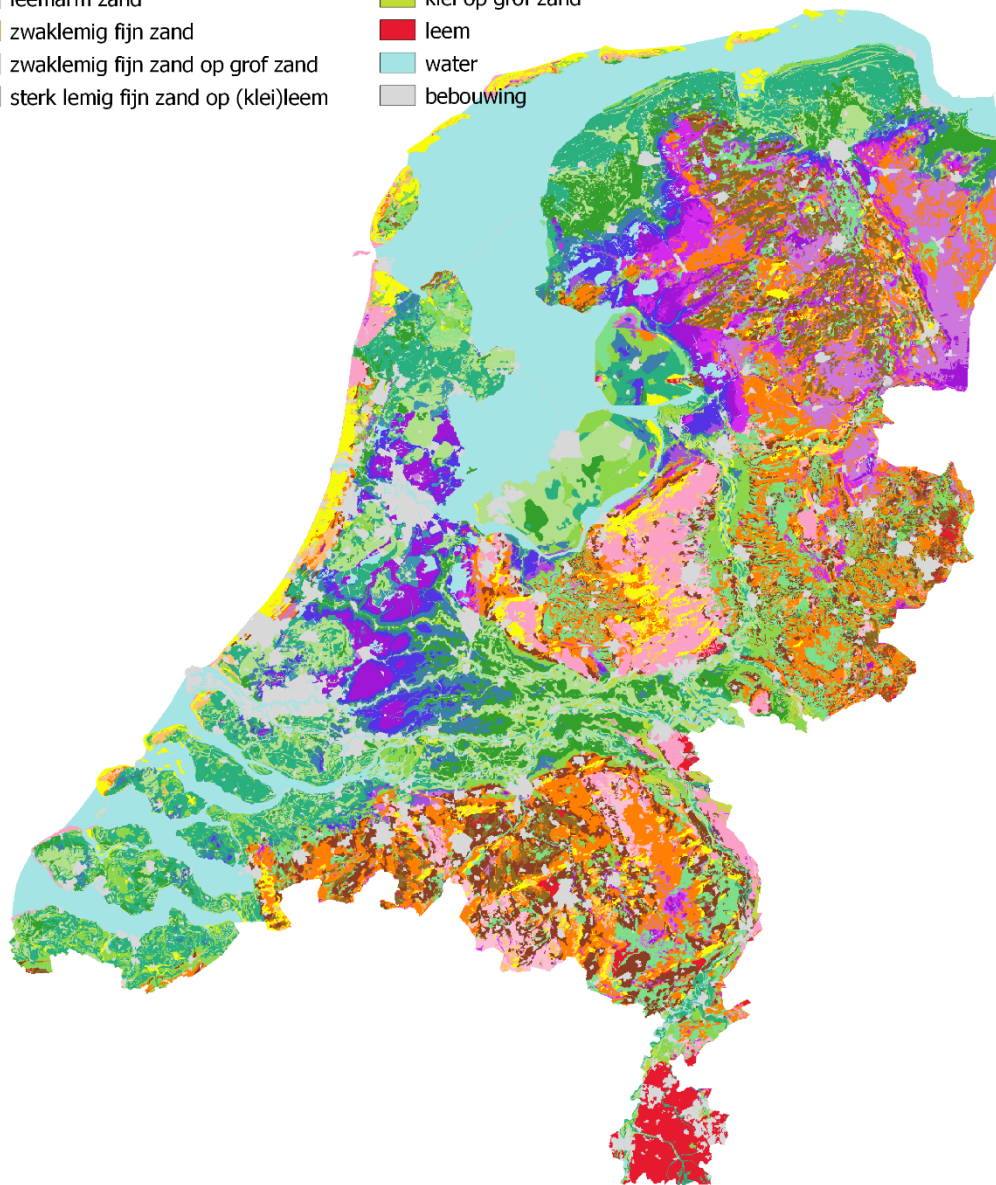
PAWN schematisatie

Een ander veel gebruikte indeling voor het clusteren van de eenheden van de bodemkaart is de indeling naar bodemopbouw. Deze indeling is in 1988 opgesteld in het kader van de PAWN-studie (Policy Analysis for the Watermanagement of the Netherlands) en wordt daarom ook wel PAWN-indeling genoemd (Wösten *et al.*, 1988). De indeling onderscheidt naar bodemopbouw 21 verschillende eenheden (Tabel 1). Door middel van een zeer schematische profielschets wordt per eenheid de laagopvolging beschreven.

BOFEK_2012_PAWN

veraarde bovengrond op diep veen
 veraarde bovengrond op veen op zand
 kleidek op veen
 kleidek op veen op zand
 zanddek op veen op zand
 veen op ongerijpte klei
 stuifzand
 leemarm zand
 zwaklemig fijn zand
 zwaklemig fijn zand op grof zand
 sterk lemig fijn zand op (klei)leem

enkeerdgronden (fijn zand)
 sterk lemig zand
 grof zand
 zavel met homogeen profiel
 lichte klei met homogeen profiel
 klei met zware tussenlaag of ondergrond
 klei op veen
 klei op zand
 klei op grof zand
 leem
 water
 bebouwing



Figuur 72 PAWN-indeling Nederlands bodem op basis van bodemopbouw, 21 eenheden.

Tabel 15 PAWN-indeling, 23 eenheden.

Klasse	Beschrijving
1	veraarde bovengrond op diep veen
2	veraarde bovengrond op veen op zand
3	kleidek op veen
4	kleidek op veen op zand
5	zanddek op veen op zand
6	veen op ongerijpte klei
7	stuifzand
8	leemarm zand
9	zwaklemig fijn zand
10	zwaklemig fijn zand op grof zand
11	sterk lemig fijn zand op (klei)leem
12	enkeerdgronden (fijn zand)
13	sterk lemig zand
14	grof zand
15	Zavel met homogeen profiel
16	lichte klei met homogeen profiel
17	klei met zware tussenlaag of ondergrond
18	klei op veen
19	klei op zand
20	klei op grof zand
21	leem
22	water
23	bebouwing

Bijlage D Estimating biodiversity of grasslands from marker data and RS data

Estimating biodiversity of grasslands from marker data and RS data

Ron Wehrens

February 28, 2024

1 Introduction

In the PPS “Herkenning kruidenrijk grasland” the goal is to get an idea of biodiversity, focusing on herbs especially, of grasslands on the basis of remote-sensing data. This is summarized in a so-called Q value, consisting of values ranging from 1 (very poor) to 5 (very rich) and in some cases even beyond. In this report I assess the use of ridge regression for RS-derived indices, and maybe will briefly cover some alternatives. Obtaining a reliable measure for prediction accuracy will be important, since predictions may be used in the future for policy making.

In the first instance we will focus on predicting Q values on a continuous scale, using the remote-sensing markers supplied by WEnR, and in addition summaries of time series of raw remote sensing data for the areas of interest. Here, I am concentrating first on mean values and standard deviations of these time series, but minima, maxima, ranges and potentially other summary statistics can be used, too.

2 Exploratory analysis

To get an idea of the data it is important to do a quick exploratory analysis of the independent data (this paragraph) as well as the dependent data (next). Tools used here include principal component analysis, clustering and correlation analysis. The total number of records is 1022, and the total number of columns of the independent data is 113. Of these, 27 are variables based on raw remote-sensing time series, and the other 87 columns are marker variables, or other descriptors of the plots.

Just as a check, here are the three groups of columns corresponding to plot/field characteristics, RS markers and raw RS summary data, respectively:

Plot/field characteristics:

[1] "Jaar"	"fieldid"	"q_20xx"
[4] "std_q_20xx"	"richness_2018"	"std_richness_2018"
[7] "shannon_2018"	"std_shannon_2018"	"simpson_2018"


```
[10] "std_simpson_2018" "Selectie_Tim" "Longitude"
[13] "Latitude"
```

RS markers or plot descriptors:

```
[1] "ahn2" "jr_mean_NDWI_10m"
[3] "jr_mean_NREVI_20m" "sz_mean_NDWI_10m"
[5] "sz_mean_NREVI_20m" "jr_mean_VHdB"
[7] "jr_mean_VVdB" "range_VHdB"
[9] "range_VVdB" "winter_mean_VHdB"
[11] "winter_mean_VVdB" "sz_std_ln(dVV)"
[13] "sz_std_ln(dVH)" "jr_std_ln(dVV)"
[15] "jr_std_ln(dVH)" "summer_std_ln(dVV)"
[17] "summer_std_ln(dVH)" "sz_mean_ratio_VH/VV"
[19] "sz_std_ratio_VH/VV" "jr_mean_ratio_VH/VV"
[21] "jr_std_ratio_VH/VV" "sz_std_coh1A_VV"
[23] "sz_std_dcoh1A_VV" "jr_std_coh1A_VV"
[25] "jr_std_dcoh1A_VV" "sz_std_coh1A_VH"
[27] "sz_std_dcoh1A_VH" "jr_std_coh1A_VH"
[29] "jr_std_dcoh1A_VH" "jr_std_VHdB"
[31] "jr_std_VVdB" "sz_std_VHdB"
[33] "sz_std_VVdB" "10m_NDVI_avg_std"
[35] "no_maai*" "cumNDVI_mow*"
[37] "cumNDVI_gra*" "cumNDVI_tot*"
[39] "cumNDVI*" "jr_CV_NDVI"
[41] "sz_CV_NDVI" "eind_m1"
[43] "eind_m2" "eind_m3"
[45] "eind_m4" "eind_m5"
[47] "scheuren*" "10m_NDVI_vjaar"
[49] "10m_NDVI_std_vjaar" "jr_MASD_10m*"
[51] "jr_MASD_10+20m*" "jr_mean_ratio_Red/Green"
[53] "jr_mean_ratio_Blue/Green" "jr_mean_HUE"
[55] "jr_mean_CV_10bands" "jr_std_ln(dNDVI)*"
[57] "sz_MASD_10m*" "sz_MASD_10+20m*"
[59] "sz_mean_ratio_Red/Green" "sz_mean_ratio_Blue/Green"
[61] "winter_mean_ratio_Red/Green*" "winter_mean_ratio_Blue/Green*"
[63] "sz_mean_HUE" "winter_mean_HUE*"
[65] "winter_mean_NDMI_20m*" "sz_mean_NDMI_20m"
[67] "sz_mean_CV_10bands" "sz_std_ln(dNDVI)*"
[69] "summer_std_ln(dNDVI)*" "S2REP_mar*"
[71] "S2REP_apr*" "S2REP_may*"
[73] "S2REP_jun*" "S2REP_jul*"
[75] "gwt1" "gwt2"
[77] "gwt3" "gwt4"
[79] "gwt5" "gwt6"
[81] "gwt7" "gwt8"
[83] "bf1" "bf2"
```

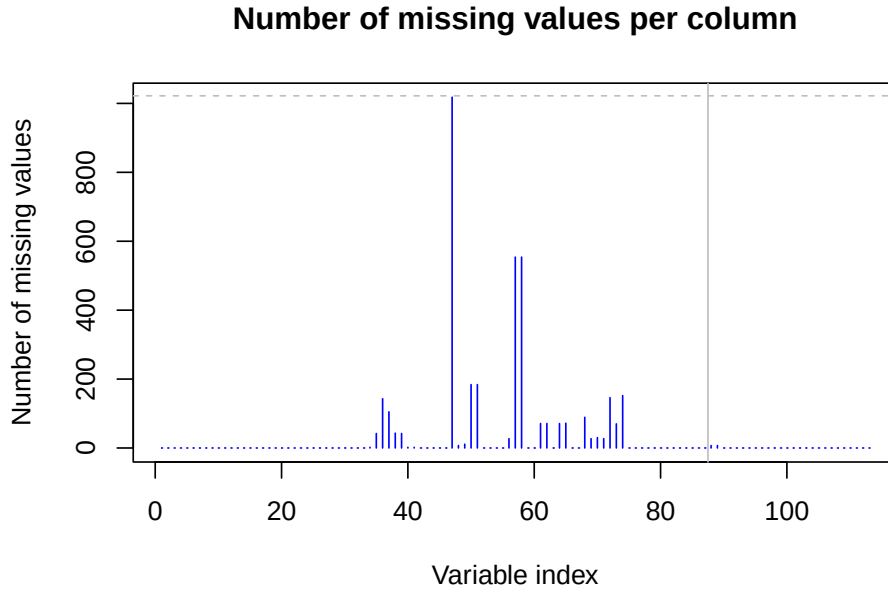


Figure 1: Number of missing values in each of the columns. The vertical gray line separates the marker variables (left) from the remote-sensing variables.

```
[85] "bf3"          "bf4"
[87] "bf5"
```

Summaries of raw RS data:

```
[1] "mean...3" "std...7" "mean...9" "std...13" "mean...15" "std...19"
[7] "mean...21" "std...25" "mean...27" "std...31" "mean...33" "std...37"
[13] "mean...39" "std...43" "mean...45" "std...49" "mean...51" "std...55"
[19] "mean...57" "std...61" "mean...63" "std...67" "mean...69" "std...73"
[25] "mean...75" "std...79"
```

Variable names with a asterix postfix (only present in the second group) are eliminated in the final selection of the data set since they contain too many missing values.

The original dataset also contains information on how often Q values have been determined in a particular year. If that is more than once, then the average Q value is used. This is probably not the best way of handling this information: for such cases the variability of the Q value will be reduced. It would have been better to include each Q measurement as a separate record.

First, we assess the number of missing values in each column – Figure 1 shows the result. A few columns have quite a lot of missing values, some only have a few, most have zero. The one variable consisting mostly of missing values is **scheuren**. The distribution of missing values across rows of the matrix shows a different pattern, with fewer than twenty missing values in each row. We'll remove the missing values in a multi-stage

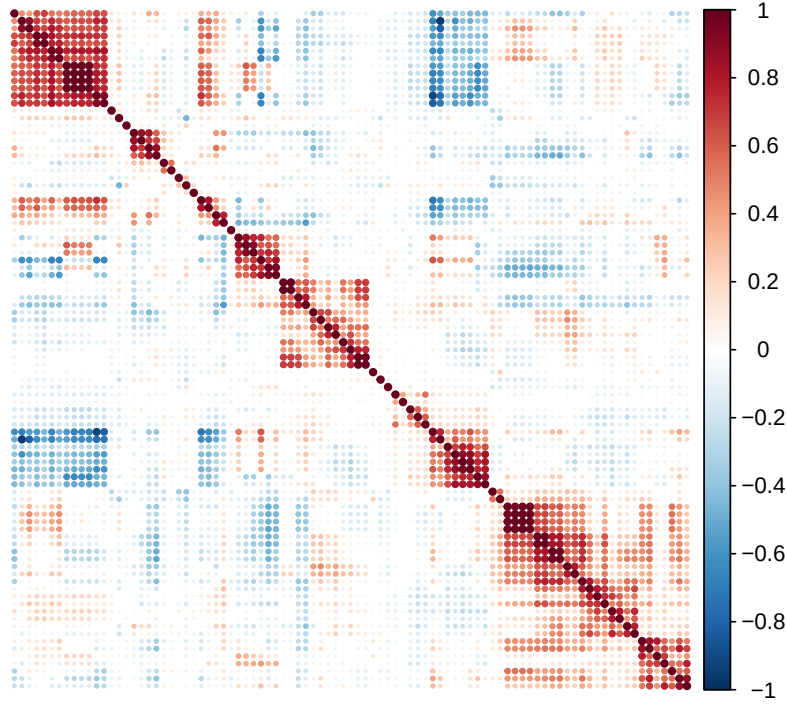


Figure 2: Correlations between the markers – clearly, blocks of highly correlated (or anti-correlated) variables are present. Variables in rows and columns are ordered according to a hierarchical clustering criterion.

process, first eliminating all columns with more than 25 missing values, then eliminating any remaining rows containing NAs. This leads to a data matrix of 1004 rows and 91 columns.

Figure 2 shows the correlations between the variables. Variables are ordered so that highly correlated variables are grouped. It is clear that there is a lot of structure in the data: several blocks of variables can be identified that are highly correlated. This is no surprise, of course, but it does mean that the information in the data can be represented in a much more efficient way.

Another way to visualize correlations between variables and relationships with database records is a PCA biplot, as shown in Figure 3. The small circles are the database records, projected in the two-dimensional subspace of the two most informative linear combinations of marker variables. They are coloured according to the Q value (discretized into integer categories). It is clear that most low Q values are located in the right side of the plot, and higher Q values in the left. In other words: the most important variation in the independent data is related to Q values. This makes it very likely that it should be possible to set up a predictive model. Other colourings for the PCA scores are possible, too – year may be relevant. The small group of samples at the top of the figure seem

to possess some general characteristics that set them apart from the bulk of the data; it could be interesting to assess this in more detail.

The blue arrows in this figure point to the triangles. These are the so-called loadings, corresponding to the axes at the top and right of the figure. The loadings indicate the weights of the original variables on the new axis, PC1 and PC2. The arrows pointing left or right from the origin, e.g., have large (absolute) loadings on PC1 but small loadings on PC2; the reverse is true for the arrows pointing up- and downward. The group of arrows pointing East can be interpreted as a set of variables that, when showing high values, indicate low Q values. Conversely, the arrows pointing in the opposite direction correspond to variables that are positively correlated with Q value. Almost 35% of the information in the complete data set is contained in only these two variables, confirming a large overlap between the information content of the marker variables.

Note that there is a small group of records that is forming a separate cluster in the top of the figure; these correspond to records with high values for the variables (blue arrows) pointing upwards. We did not investigate this further, since it is only a small fraction of the data set. All records that have a PC2 value larger than 10 will be marked in a separate column in the output file – their locations are shown in the small map on the right in Figure 3. The area between Sneek and Joure (?) seems to be well represented.

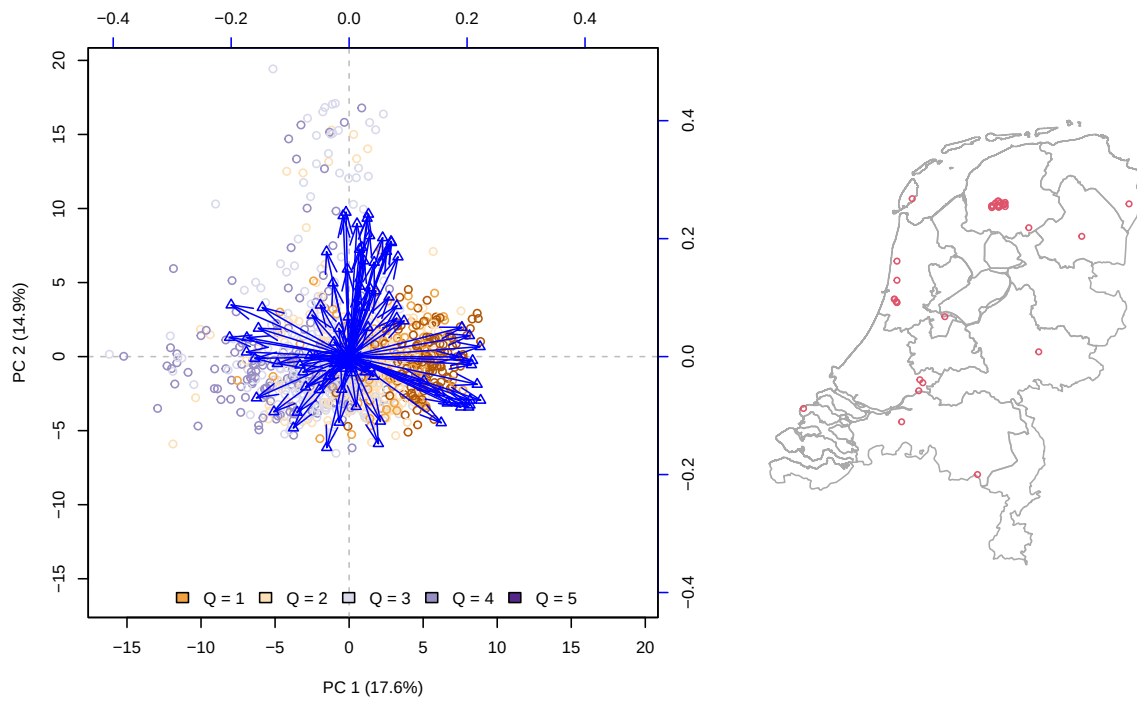


Figure 3: PCA biplot of the normalized grassland data. Each record is a dot in the plot, coloured according to the Q value. Contributions of the original variables to the newly defined axes are indicated with arrows. The map on the right shows the location of records with a PC2 score larger than 10.

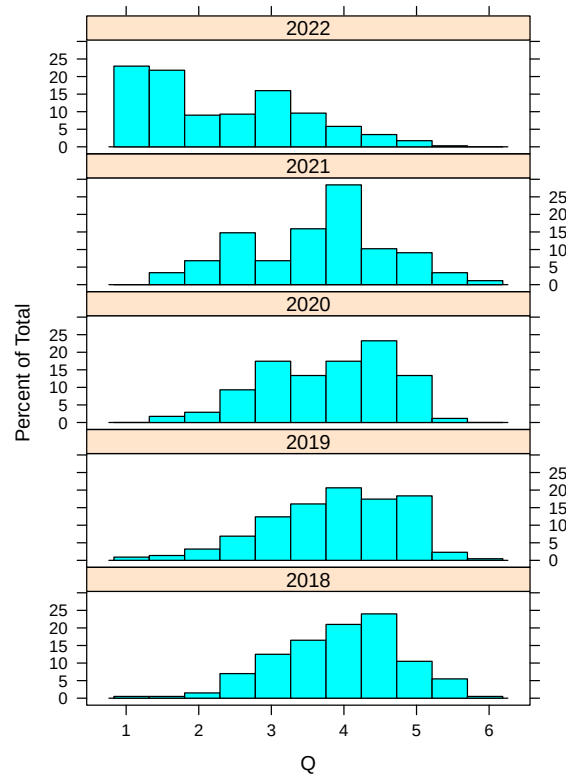


Figure 4: Histograms of the Q values in the data set without missing values.

Before trying to predict Q from the RS markers, it is interesting to explore this variable in its own right. Yearly histograms are shown in Figure 4. Clearly, the data from 2022 focus more on low- Q fields – the other years are more similar to each other.

3 Modelling Q values

The general strategy for a multivariate regression model can be summarized as follows:

- an exploration of the independent variables: which ones are correlated? What ranges do they have? The aim is to get an idea of the information content of each variable, both in an absolute and in a relative sense. This should also help in deciding whether particular forms of scaling may be useful. Some variables are categorical – these should be replaced by indicator variables.
- assess the distribution of the independent variable. It has been discussed that the variability of the Q value estimates in the data may depend on the value, with larger variabilities possibly associated with larger values. This usually implies a log transformation, or, less drastic, a square-root transformation.
- divide the data into separate training and test sets. This needs some thought: the data are coming from different years, and the question is whether it is useful to keep years separate (so that we can do forecasting: predictions for a complete, unseen, year) or whether it is better to mix years so that the training sets are really representative for the test sets. Another issue is that fields may be sampled in multiple years – if that is the case, then the sampling for dividing the data in training and test sets should be done on the field level rather than the record level.
- fit a model, using crossvalidation to select the optimal penalty weight, and investigate the residuals to see if there are patterns pointing to missed sources of information or violations of assumptions.
- finally, present the results for the test sets, including expected prediction errors, and investigate whether the distribution of the residuals is in line with that in the test set.

Some of those aspects are less relevant for random forests: there, scaling of the data usually does not influence the results, and crossvalidation is not necessary since no method parameter is optimized. Rather, the random forest uses an internal sampling approach to assess prediction error. Nevertheless, for comparison purposes it is useful to use the same training/test set division as is used in the ridge regression.

3.1 Independent variables

Categorical variables have been identified and replaced by indicator variables, each corresponding to a number of columns equal to the number of classes, and ones in locations where a record belongs to a certain class. Histograms of individual variables by and large show patterns similar to normal or log-normal distributions, with the exception of the `eind_m2` – `eind_m5` parameters, which consist almost exclusively of the imputed value of 365 – these have been removed from the data. Furthermore, the data matrix of independent variables has been cleared of any missing values, as described above.

3.2 Division in training and test sets

Here, we will use a randomly sampled test set of size 200, with the remaining 804 records used for training in all models. Linear models like ridge regression will apply crossvalidation on the data in the training set only, just like random forests will do their internal out-of-bag validation using the data of the training set only. Using the optimised metaparameter (for ridge regression and friends) we'll obtain predictions for the test set, which will be summarized by R^2 and RMS values.

3.3 Ridge regression

Setting up a RR model involves crossvalidation, to select the optimal value for the penalty parameter λ . Figure 5 shows the result. It is clear that smaller values of λ (a low penalization) correspond with lower RMS errors – the optimal value of λ is 0.07. Very often, not the minimum of λ is taken, but the last value within one standard deviation from the optimum (here: 0.19). This reduces the chance of overfitting, and this is what we'll use here, too. The numbers above the plot are indicating the number of non-zero coefficients for each value of λ – since ridge regression will (almost) never lead to coefficients with size zero this number equals the total number of coefficients in all cases. For other methods like the lasso or the elastic net, we would see a gradual decrease in the number of non-zero coefficients for increasing penalty sizes.

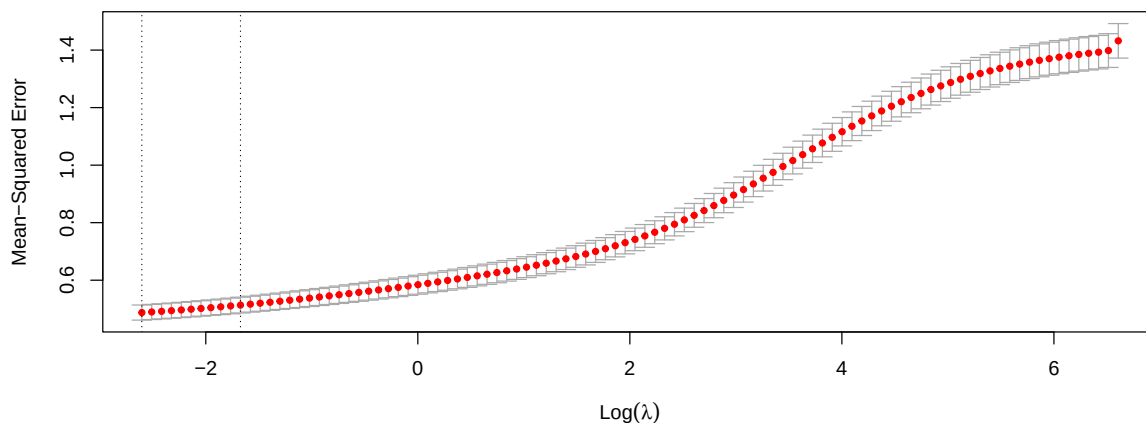


Figure 5: Crossvalidation results for ridge regression, relating RS markers to Q . The “optimal” value for λ we will use in the following is 0.188.

Having made the choice for λ , here 0.188, we can proceed by inspecting the residuals of the regression. Predictions are plotted against true values in Figure 6. These figures give an immediate sense of the quality of the predictions. The R^2 values in the plots

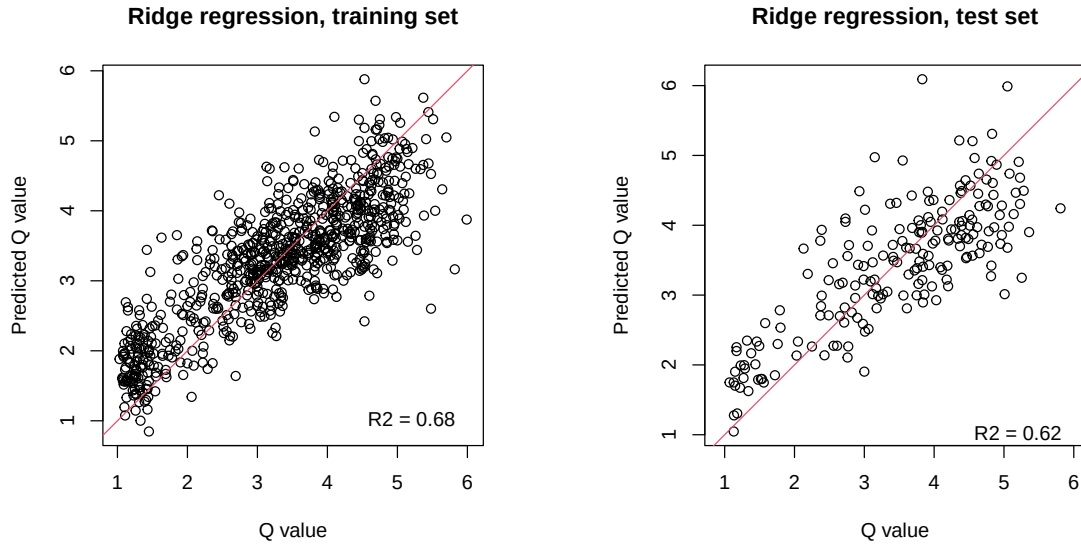


Figure 6: True Q values versus predictions for the ridge regression model.

indicate that the training predictions are slightly closer to the true data, as should be expected. These numbers correspond to RMS values of 0.671 for the training data, and 0.723 for the test data.

A more detailed look at the residuals is obtained by the diagnostic plots in Figure 7, which shows the analysis of the residuals of the training data in the top row, and of the test data in the bottom row. The histograms, shown in the left panels, look reasonably normal, which is confirmed by the Q-Q plots in the middle column.

The two rightmost plots show the prediction errors for different predicted values. It is clear that at the very low end of the scale, say, below 2, predictions are quite accurate – at the top end of especially the test data we see quite a few negative residuals, *i.e.* cases where the prediction is smaller than the true value. This is quite usual: often extreme values are underestimated – the linear model naturally dampens extremes.

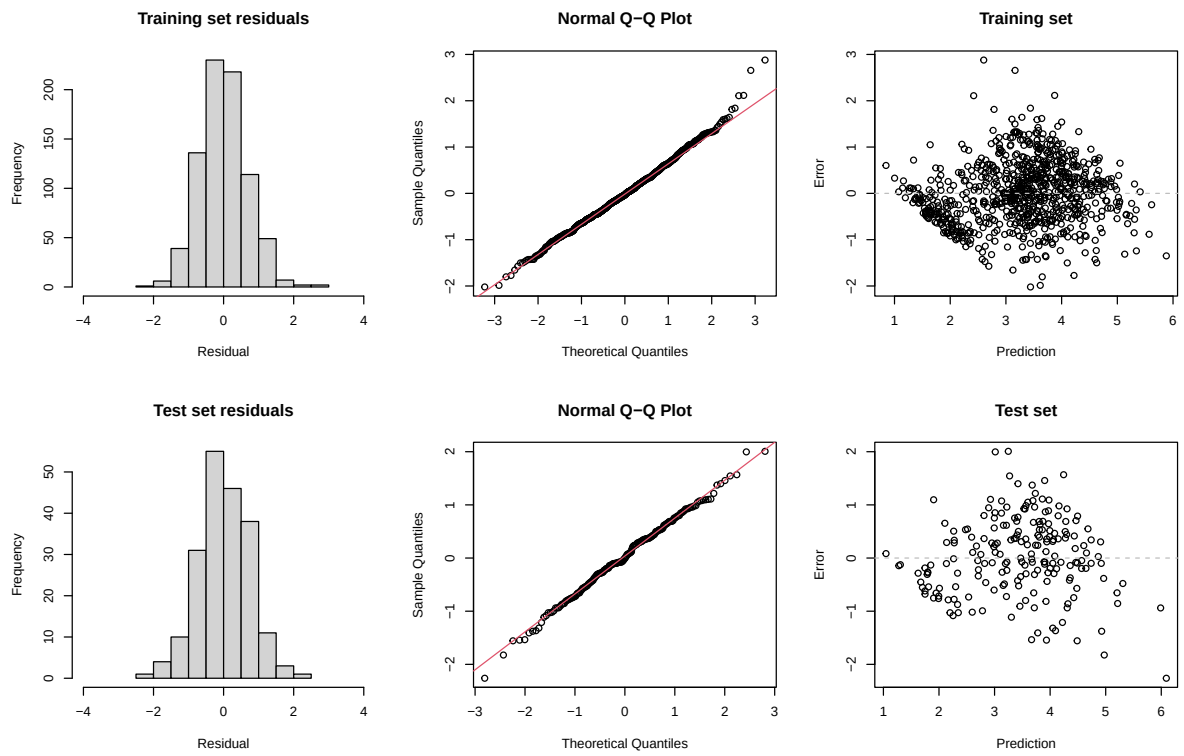


Figure 7: Analysis of residuals from the ridge regression with penalty parameter $\lambda = 0.19$. Top line: training set residuals; bottom line: test set residuals.

4 Random Forests

Random forests are collections of trees (yes!), where each tree is a series of questions to the data leading to a series of splits. Examples are: is the value of the first variable larger than 5? And the second larger than 3? And the first again larger then 10? In this way the solution space is partitioned into sections, each associated to a particular value of the outcome variable. Training the tree corresponds to finding the optimal series of splits. Since individual trees are highly variable in their predictions, an ensemble of trees is used (the forest) so that predictions are averaging out. To avoid all trees to be very similar, each tree gets to see a different random subset of the data (both in terms of rows and of columns), hence the term random forest. They are known to be highly robust, and optimization of meta-parameters (like the penalty term in ridge regression) is usually not necessary.

Random forests can be used both for continuous outputs and for class outputs – here we’ll apply both types to modelling Q values. For the classification output we’ll use Q bins with a width of 0.5. This provides a natural comparison to the neural network classification results that are being generated by NEO.

4.1 RF regression

For random forests we use the exact same data set as in ridge regression, that is: we use the training data to set up the model, which also will give us estimates of prediction error. These can be compared with the performance of the random forest on the test set. The table below summarizes the prediction results in terms of RMS values:

	RMS	R ²
Training	0.273	0.948
OOB	0.686	0.671
Test	0.629	0.715

The first line gives the RMS and R^2 numbers for the prediction of the training data. These seem to be much better than observed for the ridge regression, but this is due to the extreme flexibility of non-parametric models such as the random forest – obviously, if we are forced to fit a high-dimensional plane through a data set, as is done in ridge regression, we cannot accomodate deviating records. In a random forest this can easily be done. That is why this number is never considered in evaluating a random forest model. Rather, one focuses on the so-called out-of-bag (OOB) estimates: these are predictions by trees in the forest that have not seen a particular record (remember that each tree is presented only a subset of the data). The idea is very similar to crossvalidation but a bit less structured, and more relying on brute force. Clearly, it works: the OOB estimates for RMS and R^2 are very similar to the values found for the test set.

Overall, the quality of the predictions is very similar to that of ridge regression. Even though the test set seems to be predicted better than in the ridge regression, I would be cautious to assign much importance to that since the OOB estimate is still close to the RMS value of 0.7 we saw earlier.

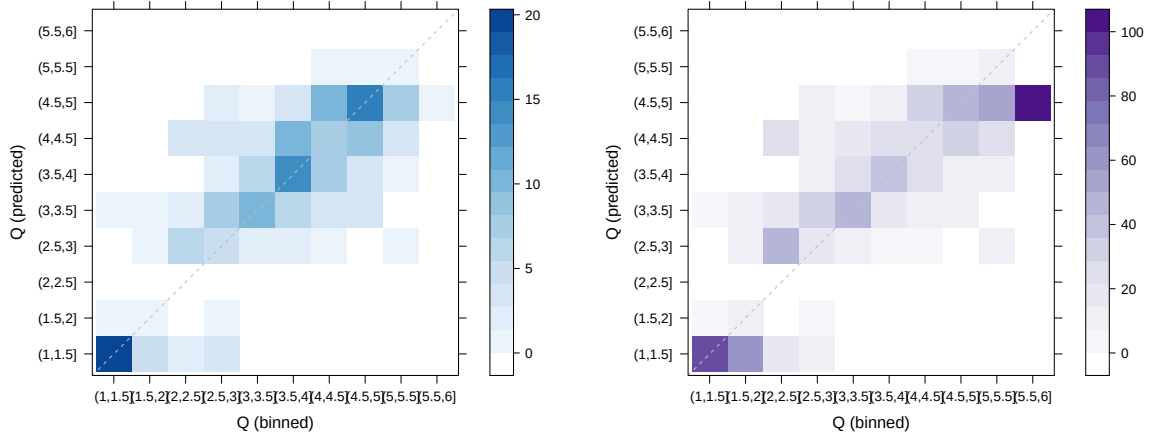


Figure 8: Results of RF classification, expressed as number of cases (left) or bin percentage (right). As an example, the left plot indicates that 19 cases are correctly predicted to fall in the first bin. The right plot shows that this is approximately 90% of all cases that should be in the first bin. Second example: there is only one Q value in the highest bin, and this one is predicted incorrectly, leading to very light blue color in the left figure, but an intensity of 100% in the right figure.

4.2 RF classification

A very simple quality criterion for any classification method is the fraction of cases in which the “true” class equals the predicted class. Focusing again on the OOB estimate for the training data, we find a success rate of 34.6% – quite low. For the test data, this number equals 36.5%. These numbers are in quite good agreement, as one would hope. If we would replace each predicted bin by its centre value and again calculate RMS values we would end up at an RMS value of 0.73 for the test set, very similar to the earlier results.

Not all errors are equally costly: classifying a record in a neighbouring category is of course less bad than classifying it in a remote Q bin. The spread of the classifications (test set only) is shown in Figure 8. Here, the columns indicate the true binned Q value, and the rows indicate the predictions. In the left plot, colors indicate the number of cases in each combination: an ideal prediction would see a blue diagonal from the bottom left to the top right (indicated with the gray dashed line), and white everywhere else. Here, we do see some correlation, indicating that most errors are “close” to the true value, although we see that the bin of 4–4.5 contains predictions ranging from 2.5 to 5.5. Note that there are no predictions falling in bins 2–2.5 and 5.5–6. Since this plot cannot really be interpreted directly without taking into account the number of true cases in each bin, the right plot in the figure shows the same information, but now normalized for each bin separately. This is done by dividing each true bin by the number of elements in the bin, and then multiplying by 100.

5 Variable importance

5.1 Background

The term variable importance can be interpreted in different ways. In general, we associate it with the predictive power of a particular variable with respect to the outcome of interest. In multivariate data, such as the data in this case, this cannot be decoupled from the other variables in the model, which complicates matters. If data are orthogonal, their estimates are independent of each other, and then the size of, say, a regression coefficient in a ridge regression model can be interpreted as the importance of a variable (provided the data have been normalized). However, as soon as correlations are present – and in real data they always are: here the correlations are even quite extreme, as shown in Figure 2 – this no longer is true.

Two approaches are often used. The first is to assess each variable in isolation, build a predictive model using only that variable, and then assess the quality of the model. That gives some information, but ignores the complementarity of variables: two variables that both seem predictive might together show no improvement over the individual predictions, or may add different information and in that case lead to a much better model. A procedure like stepwise variable selection is based on such a strategy, trying to build parsimonious models with high predictive power.

The second approach is to leave out individual variables and assess the deterioration in model quality (as assessed by, e.g., crossvalidation or OOB strategies). If that deterioration is large, the variable is said to be important. This is the standard approach in random forest research, and while a large importance does mean that the variable is influential, a small importance does not mean the opposite: if a variable has an apparently small importance this can also be due to the presence of another, correlated variable, that “takes over” the task of the eliminated one.

5.2 One-by-one approach

A very simple approach to quickly get an idea of the relationship between individual variables and the target variable is to plot the (absolute) size of correlations: larger correlations indicate greater predictive power. This is shown in Figure 9. There is quite a lot of variation: 14 variables show correlations larger than 0.5 (in absolute size), and 27 have a correlation smaller than 0.1. The five variables with the highest correlation to Q are:

```
[1] "jr_mean_ratio_Red/Green" "sz_mean_ratio_Red/Green"  
[3] "mean...45"              "mean...39"  
[5] "jr_mean_NREVI_20m"
```

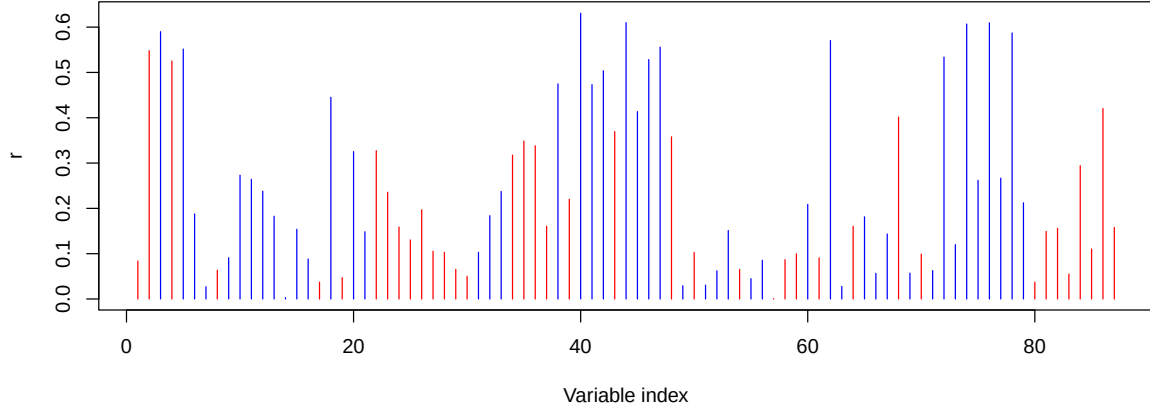


Figure 9: Absolute correlations between the target variable (here: Q) and individual variables. Positive correlations are shown in red; negative ones in blue.

5.3 Variable groups

Rather than focusing on individual variables, one could ask how good a ridge regression model would be if only one type of information would be presented. For this, we can divide the variables in groups, where each group is either a highly correlated set (e.g., obtained from an analysis like shown in Figure 2), or a group of variables from a common origin or with related meaning.

5.3.1 Groupings based on correlation

Using a hierarchical clustering on the transposed data matrix we can find groupings in the variables. Note that there are many different clustering methods, each leading to different results, so the current implementation should be seen as just one of many possibilities. The dendrogram obtained from a average-linkage clustering is shown in Figure 10. Cutting the dendrogram at a particular height leads to a number of clusters; in the figure, a line is drawn at a height of 30, leading to 32 clusters. Many of these consist of singletons, but there are a few with more than five elements:

Cluster 3 :

[1] "jr_mean_NREVI_20m"	"sz_mean_NREVI_20m"
[3] "10m_NDVI_vjaar"	"jr_mean_ratio_Red/Green"
[5] "jr_mean_HUE"	"sz_mean_ratio_Red/Green"
[7] "sz_mean_HUE"	"sz_mean_NDMI_20m"
[9] "mean...3"	"mean...33"
[11] "mean...39"	"mean...45"
[13] "mean...51"	

We can fit ridge regression models for these 6 groups only, and then assess how well these small models are predicting the test set. The results are shown in the table below:

	R2	RMS
Cluster 3	0.498	0.84
Cluster 5	0.297	0.99
Cluster 11	-0.022	1.19
Cluster 13	0.090	1.12
Cluster 30	0.401	0.91
Cluster 31	0.111	1.11

Clearly, the full ridge regression model fares better than either of these single-cluster models. This is not surprising since each of these clusters basically contain one particular kind of information.

We can also do the reverse: rather than including only one cluster, we can exclude each cluster. This can potentially be more informative, since we created the clusters on the basis of correlation – excluding one cluster then should eliminate all highly correlated variables. The results are shown below:

	R2	RMS
Cluster 3	0.57	0.77
Cluster 5	0.62	0.73
Cluster 11	0.62	0.72
Cluster 13	0.62	0.72
Cluster 30	0.61	0.74
Cluster 31	0.62	0.73

The interpretation again is quite straightforward: if leaving out a cluster leads to a deterioration of quality (indicated by a low R2 or high RMS value), then that cluster is important. Here, only eliminating cluster 3 would seem to lead to a small deterioration. This is not unexpected: cluster 3 is also the largest cluster. We probably left in a lot of lower-level correlation, and correlation between multiple parameters is to be expected, too.

5.3.2 Groupings based on physical meaning

Since I have no feeling for the meaning of individual coefficients, I am going to create some groups based on the names of the coefficients and some information from the explanatory excel file. My first attempt is to formulate the following groups:

Group abiotic:

```
[1] "ahn2" "gwt1" "gwt2" "gwt3" "gwt4" "gwt5" "gwt6" "gwt7" "gwt8" "bf1"
[11] "bf2" "bf3" "bf4" "bf5"
```

Group year:

```
[1] "jr_mean_NDWI_10m" "jr_mean_NREVI_20m"
```



```

[3] "jr_mean_VHdB"          "jr_mean_VVdB"
[5] "jr_std_ln(dVV)"        "jr_std_ln(dVH)"
[7] "jr_mean_ratio_VH/VV"   "jr_std_ratio_VH/VV"
[9] "jr_std_coh1A_VV"       "jr_std_dcoh1A_VV"
[11] "jr_std_coh1A_VH"       "jr_std_dcoh1A_VH"
[13] "jr_std_VHdB"           "jr_std_VVdB"
[15] "jr_CV_NDVI"            "jr_mean_ratio_Red/Green"
[17] "jr_mean_ratio_Blue/Green" "jr_mean_HUE"
[19] "jr_mean_CV_10bands"

```

Group season:

```

[1] "sz_mean_NDWI_10m"      "sz_mean_NREVI_20m"
[3] "sz_std_ln(dVV)"        "sz_std_ln(dVH)"
[5] "sz_mean_ratio_VH/VV"   "sz_std_ratio_VH/VV"
[7] "sz_std_coh1A_VV"       "sz_std_dcoh1A_VV"
[9] "sz_std_coh1A_VH"       "sz_std_dcoh1A_VH"
[11] "sz_std_VHdB"           "sz_std_VVdB"
[13] "sz_CV_NDVI"            "sz_mean_ratio_Red/Green"
[15] "sz_mean_ratio_Blue/Green" "sz_mean_HUE"
[17] "sz_mean_NDMI_20m"      "sz_mean_CV_10bands"

```

Group raw:

```

[1] "mean...3" "std...7" "mean...9" "std...13" "mean...15" "std...19"
[7] "mean...21" "std...25" "mean...27" "std...31" "mean...33" "std...37"
[13] "mean...39" "std...43" "mean...45" "std...49" "mean...51" "std...55"
[19] "mean...57" "std...61" "mean...63" "std...67" "mean...69" "std...73"
[25] "mean...75" "std...79"

```

For each of these groups we fit ridge regression models including only variables from that group – the R2 and RMS results are shown below.

	R2	RMS
abiotic	0.068	1.14
year	0.515	0.82
season	0.426	0.89
raw	0.593	0.75

From the four groups, the raw parameters clearly lead to the best predictions. Again, we can also observe how the models fare when leaving out one of these groups:

	R2	RMS
abiotic	0.62	0.73
year	0.61	0.73
season	0.63	0.72
raw	0.57	0.78

Again, it seems that leaving out the variables related to the raw RS data lead to the biggest deterioration, whereas excluding season or abiotic variables does not deteriorate the model at all.

5.4 Special requests

The S2REP variables were removed since they contained too many missing values. Yet, according to Wouter and Tim, they could be promising. Here we'll repeat the process of eliminating rows and columns containing missing values, now making sure at least some of the S2REP variables are in. March and April are highly correlated – they are also the columns with the lowest number of missing values. June is remarkably uncorrelated with the others. This is probably due to mowing events, which may completely disturb values for this type of variable. We'll keep these three R2SEP markers in our data. This has consequences for the selection of other rows and columns; that means our data are not going to be a simple subset of the training and test sets used earlier. Rather, we'll define new training and test sets (with approximately the same ratio in terms of numbers of samples) and assess the results. If the influence of the three S2REP markers is quite large, we may have to rethink our original approach.

We fit the three models assessed in this report: ridge regression, random forest regression and random forest classification. The RMS value for the ridge regression test set is 0.69, and for the random forest regression the corresponding value is 0.63. Both are slightly smaller than the values found in the earlier results, not including any S2REP variables. The percentage of correct classifications in the RFC approach now is 0.4%, again slightly better than what we found earlier.

While it is tempting to conclude from this that S2REP is important, the different records in training and test sets may have something to do with this, too. A more complete picture emerges when we assess the contribution of the S2REP variables alone, and the effect of excluding the S2REP variables. This, again, will be done for the ridge regression models only. Here are the results:

	R2	RMS
Only	0.43	0.97
Excluded	0.70	0.70

These numbers show that a regression based only on these three S2REP variables leads to results that are comparable to the other groups we have seen – in any case not spectacularly better. The second line shows that for this particular data set, slightly better results are obtained anyway: even though no S2REP variables are present, the R2 seems a bit higher than what we saw earlier, and the RMS seems a bit lower. It seems fair to say that the S2REP variables may be included but at the moment their relatively large number of missing values is not compensated for by a better performance.

5.5 Principal Components

Figure 3 showed that a principal component analysis leads to a strong negative association of the first PC with Q values: high PC1 coordinates correspond to low Q values, and vice versa. Here are the top and bottom 10 variables with respect to their loadings on PC 1:

Top ten:

[1]	"jr_mean_NREVI_20m"	"jr_mean_ratio_Red/Green"
[3]	"mean...3"	"sz_mean_ratio_Red/Green"
[5]	"sz_mean_NREVI_20m"	"mean...39"
[7]	"mean...45"	"sz_mean_NDMI_20m"
[9]	"mean...51"	"jr_mean_HUE"

Bottom ten:

[1]	"sz_mean_NDWI_10m"	"jr_mean_NDWI_10m"	"sz_CV_NDVI"
[4]	"jr_CV_NDVI"	"10m_NDVI_avg_std"	"mean...21"
[7]	"sz_mean_CV_10bands"	"jr_mean_CV_10bands"	"mean...75"
[10]	"10m_NDVI_std_vjaar"		

So records with high values in the top ten variables will have low Q values, and records with high values in the bottom ten will have high Q values.

Note that an alternative to ridge regression is formed by a regression on the principal components from Figure 3 – this is (unsurprisingly) known as principal component regression (PCR). In most practical applications the performance of the methods is very similar, something that is also true for another cousin in the multivariate family, Partial Least Squares regression (PLS). The (possible) advantage of PCR and PLS is that they define a limited number of latent variables, which sometimes can be interpreted more easily than a regression coefficient vector with many numbers, as in ridge regression.

The group of samples in the top of the figure seems to stand out a bit from the bulk of the data. Simply applying a cutoff of 10 for PC2 allows us to find the records associated with these samples. They will be highlighted in the output file in a separate column.

5.6 Lasso models

The Lasso is a model that is very similar to ridge regression. The difference is that where ridge regression penalizes the sum of squared coefficients ($\sum \beta_i^2$), the lasso penalizes the sum of the absolute coefficient values $\sum |\beta_i|$. This has the effect that in a lasso model many of the coefficients are exactly zero – the ones that are not zero are then the important variables in the model. Note that this does not imply causality, and also does not imply that other subsets are inferior: probably many equivalent sets can be constructed in data sets with high correlations, such as in this case. It does give an indication of how many variables may be required to “cover” the information content in the current data set.

In this case, the lasso crossvalidation procedure (executed exactly like the one for ridge regression) leads to a value for the penalty term λ of 0, corresponding to a model

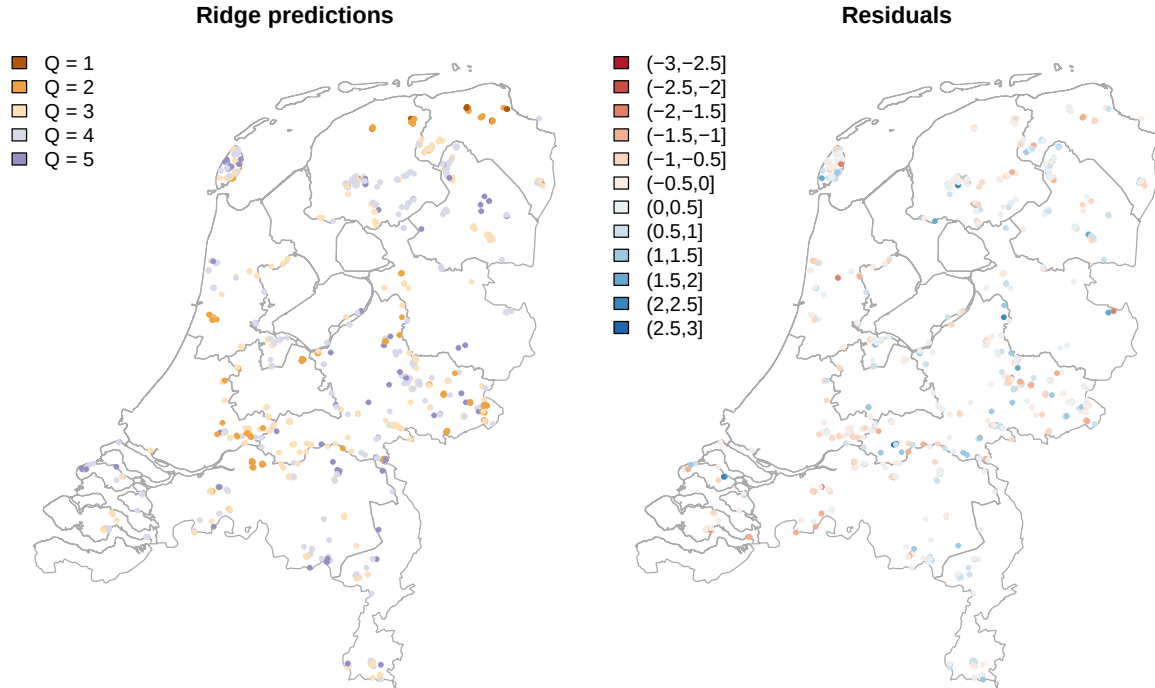


Figure 11: Ridge regression outcomes: predictions (left) and residuals (right) for training and test sets combined, visualized in maps of the Netherlands.

in which there are 55 non-zero coefficients (out of 87). This corresponds to a relatively small reduction, and indicates that despite the rather large correlations the variables do contain complementary information. The corresponding R^2 and RMS values for the test set are 0.639 and 0.707, respectively.

6 Plotting results in a Netherlands map

To see if we can observe any regional trends, it can be interesting to plot predictions and residuals in a map of the Netherlands. Later, we can also predict values for all fields for which independent data (but not necessarily Q values) are available and significantly increase the density of the predictions. A first example is shown in Figure 11.

7 Year influence

In the discussion of Nov 28, 2023, it was concluded that conditions in different years would very likely lead to systematic differences in RS markers and descriptors, whereas Q values are assumed to remain more or less constant. This introduces another source of variance, which may be captured in our models. The simplest possible approach is to include year as a variable in the three different models. That should correct for systematic and constant differences between years, and could easily lead to an improvement of 0.1 units in our expected accuracy. More detailed information would consist of specific descriptors for years (e.g., total amount of sun hours, precipitation, cumulative temperature and similar descriptors) – the “final” level would, for each record in the database, include weather information and the like.

7.1 Separate factors for individual years

Systematic differences between years can be captured by including the year as a factor in the equation. For Random Forests this is easy, and year can directly be included as a factor – for ridge regression, again, one has to convert it to several indicator columns. Here we use 2018 as the default, and the coefficients for the years 2019–2022 are then containing the offsets. Indeed, RMS and R2 values for the test set show improvements, now having values of 0.69 and 0.65, respectively. The Jaar coefficients of the model are:

2019	2020	2021	2022
0.042	0.056	0.065	-0.245

indicating that especially the year 2022 deviates from 2018, with in general much lower Q values. Now we know that 2022 also contains many more parcels that actually have a low Q value, so this is to some extent cheating. Rather than including year as a factor, we should include independent descriptors of the conditions in individual years (see the next section).

In addition, we can apply random forest regression and random forest classification to these data. Here are the results, also including the ridge regression outcomes:

	RMS	R2
Ridge regression	0.69	0.65
Random Forest regression	0.69	0.66
Random forest classification	0.69	0.66

Clearly, all three methods lead to roughly the same result. Note that it is not possible to extract coefficients for “Year” from the random forest models – the variable is taken into account, but in inexplicable ways.

7.2 Descriptive variables for years

We can try to capture relevant year information (equal for all parcels) in a small number of variables, and use these in the regression rather than a factor Year. Tim has provided

the following variables: the start of the growing season (expressed in a day number), and amounts of precipitation in four seasons (the preceding winter, the spring, summer and autumn). Note that this will, almost by definition, lead to an improvement that is very similar to the completely free adjustment for year means in the previous section: this is basically doing the same (the fact that we have multiple variables is irrelevant – these variables are 100% correlated so do not imply additional information) apart from the fact that the numerical values impose a constraint on the shape of the correction. In addition, the precipitation numbers for the four seasons add up to the grand total, so the latter is not taken into account (without any loss of information).

The results are shown below:

	RMS	R2
Ridge regression	0.69	0.66
Random Forest regression	0.69	0.66
Random forest classification	0.69	0.66

As expected, these results do not deviate from the models with a year factor. What can be interesting is the coefficients for the new variables in the ridge models:

	dayNr	Spring	Summer	Autumn	Winter
	0.061	0.079	0.202	-0.089	-0.036
previousWinter					
	-0.140				

As per usual, one should be very cautious in interpreting these coefficients, but in general a positive coefficient would indicate an increase in Q values with higher values for that coefficient, so a positive correlation, and a negative coefficient indicates the opposite. The biggest coefficient is `previousWinter`, with a negative coefficient: so a wet winter in the year before would decrease Q values.

8 Conclusions

We have fitted models to predict Q values from RS-derived variables. Two broad model classes have been considered: linear models like ridge regression and the lasso on the one hand, and non-parametric models in the form of random forest on the other. In all cases similar prediction errors were found: the RMS estimate was always around 0.7 (random forest regression performs best: 0.63; ridge regression 0.72; and random forest classification, after back-transformation to real numbers, 0.73), indicating that the prediction interval for a new measurement would have a 95% confidence interval of $\pm 2 \times 0.7 = \pm 1.4$.

The question of what variables are important is hard to answer because of the high correlations in the data, and the large number of reasonably large correlations. Several groups have been defined – in all cases it appeared that leaving out one group did not deteriorate the model much, and models based on one group performed quite a lot worse than the complete model. This implies that the information is quite widely dispersed over all variables in the data set. No exhaustive analysis has been done, given the lack of time.

Technical details

R version 4.2.1 (2022-06-23)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 22.04.4 LTS

Matrix products: default
BLAS: /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p-r0.3.20.so

locale:
[1] LC_CTYPE=en_GB.UTF-8 LC_NUMERIC=C
[3] LC_TIME=nl_NL.UTF-8 LC_COLLATE=en_GB.UTF-8
[5] LC_MONETARY=nl_NL.UTF-8 LC_MESSAGES=en_GB.UTF-8
[7] LC_PAPER=nl_NL.UTF-8 LC_NAME=C
[9] LC_ADDRESS=C LC_TELEPHONE=C
[11] LC_MEASUREMENT=nl_NL.UTF-8 LC_IDENTIFICATION=C

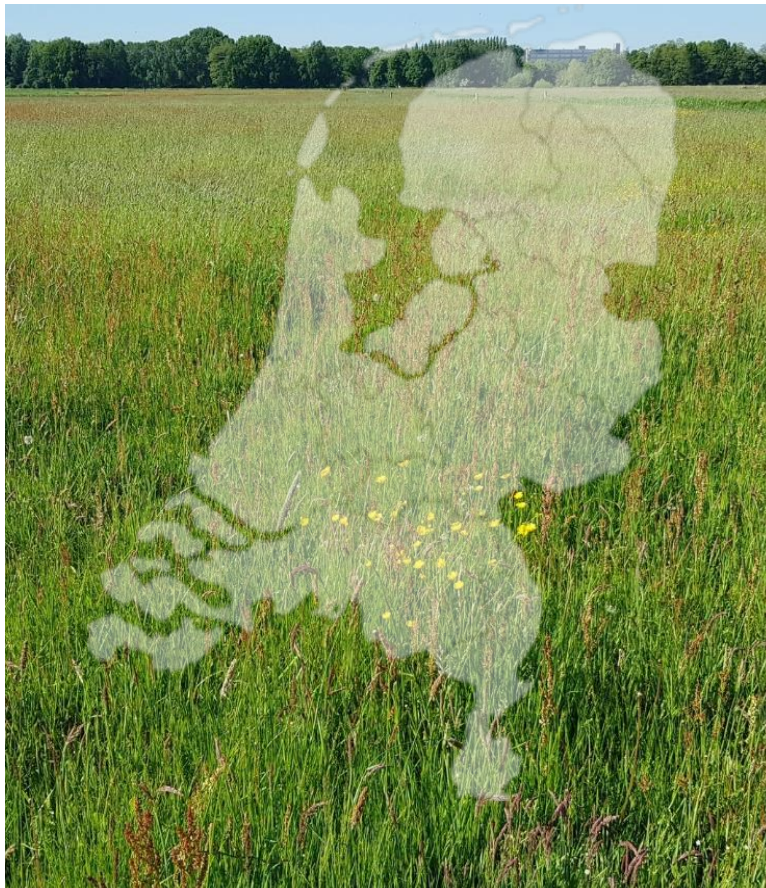
attached base packages:
[1] stats graphics grDevices utils datasets methods base

other attached packages:
[1] randomForest_4.7-1.1 Hmisc_4.8-0 ggplot2_3.4.0
[4] Formula_1.2-4 survival_3.2-13 readxl_1.4.1
[7] glmnet_4.1-6 Matrix_1.5-3 kohonen_3.0.12
[10] PCA_0.0.4 stringr_1.5.0 corrplot_0.92
[13] sp_1.6-0 reshape2_1.4.4 RColorBrewer_1.1-3
[16] gridExtra_2.3 latticeExtra_0.6-30 lattice_0.20-45
[19] knitr_1.41

loaded via a namespace (and not attached):
[1] Rcpp_1.0.10 deldir_1.0-9 png_0.1-8 digest_0.6.32
[5] foreach_1.5.2 utf8_1.2.3 R6_2.5.1 cellranger_1.1.0
[9] plyr_1.8.8 backports_1.4.1 evaluate_0.21 pillar_1.9.0
[13] rlang_1.1.1 rstudioapi_0.14 data.table_1.14.6 rpart_4.1.16
[17] checkmate_2.1.0 splines_4.2.1 foreign_0.8-82 htmlwidgets_1.6.1
[21] munsell_0.5.0 compiler_4.2.1 xfun_0.36 pkgconfig_2.0.3
[25] base64enc_0.1-3 shape_1.4.6 htmltools_0.5.4 nnet_7.3-17
[29] tidysselect_1.2.0 tibble_3.2.1 htmlTable_2.4.1 codetools_0.2-18
[33] fansi_1.0.4 dplyr_1.1.2 withr_2.5.0 grid_4.2.1
[37] gtable_0.3.3 lifecycle_1.0.3 magrittr_2.0.3 scales_1.2.1
[41] cli_3.6.1 stringi_1.7.12 generics_0.1.3 vctrs_0.6.3
[45] iterators_1.0.14 tools_4.2.1 interp_1.1-4 glue_1.6.2
[49] jpeg_0.1-10 fastmap_1.1.0 colorspace_2.1-0 cluster_2.1.3

Bijlage E Prediction of Herb Richness in grasslands in the Netherlands with Remote Sensing

Prediction of Herb Richness in grasslands in the Netherlands with Remote Sensing



- Regression Analysis Report and DL Model Candidates
 - CNN Model: Nation-wide results

Colophon

Version control

Version	Authors	Organization	Date	Description
0.1	Itsaso Guembe Cia Aaron Bakx	NEO	21/09/2023	Results from the regression analysis and suggestion of DL approaches. Includes answers to the questions formalised by WUR on 29/08/2023
0.2	Itsaso Guembe Cia Aaron Bakx Laurens Hagendoorn	NEO	13/02/2024	Addition of technical details and results for the CNN model
0.3	Itsaso Guembe Cia Aaron Bakx Laurens Hagendoorn Henk Janssen	NEO	20/08/2024	Inclusion of research as part of 4c+, aiming to optimize CNN results

Table of Contents

Table of Contents	2
1. Introduction	4
2. Data description	5
2.1. Dataset	5
2.2. Target variables	5
2.2.1. Q-index	5
2.2.2. Richness value	7
2.3. Zonal statistics	8
2.4. Remote-Sensing based Markers	8
2.5. Filter for unrepresentative parcels	8
3. Method: Ridge Regression Model	10
3.1. Ridge Regression Estimator	10
3.2. Data pre-processing	10
3.3. Model training	11
3.3.1. Spatial K-fold cross-validation	11
3.3.2. Statistical accuracy assessment	11
4. Results and discussion: Ridge Regression Model	12
4.1. Scenario 1: Q value	12
4.1.1. Parcel filters	12
4.1.2. Residual behaviour	12
4.1.3. Relative importance of features	14
4.1.4. Annual variability	18
4.2. Scenario 2: Richness value	22
5. Method: Convolutional Neural Network	24
5.1. Data pre-processing	24
5.2. Model architecture	25
5.3. Model training	26
6. Results and discussion: Convolutional Neural Network	27
6.1. Experiment I: Regression versus Classification	27
6.2. Experiment II: Seasonal effects	29
6.3. Experiment III: Q-index range effects	29
6.4. Nation-wide prediction	30

7.	Evaluation of potential AI approaches	33
7.1.	Models	33
7.1.1.	Recurrent Neural Network (GRU)	33
7.1.2.	Convolutional Neural Network.....	34
7.1.3.	Random Forest	34
7.1.4.	Dense Neural Network.....	35
7.2.	Suitability	36
7.2.1.	Data Preparation	36
7.2.2.	Results	36
8.	Optimization of CNN model	37
8.1.	Map Q values vs training data to double check whether south areas values are indeed lower than the predictions.....	37
8.2.	Check if multiyear consistency assumption is valid, based on GT parcels with Q observations for several years.....	38
8.3.	Compare ZonalStats 2022 to provided values by WUR.....	39
8.4.	Visualize feature space for different Q values to understand similarity between parcels with similar Q values	40
8.5.	Normalize signals across the whole country, per year.	43
8.6.	Data augmentation by moving time series windows.....	45
8.7.	Inclusion of weather data new features	46
8.8.	Data augmentation by temporal interpolation	47
8.9.	Alternative DL model: transformers.....	49
8.10.	Comparison of photo model and CNN results	50
9.	References.....	52

1. Introduction

On January 2023, the legislation supporting the new CAP entered into force, meaning a fairer, greener and more performance-based set of rules to ensure a sustainable future for European farmers, provide more targeted support to smaller farms, and allow greater flexibility for EU countries to adapt measures to local conditions.¹

In the Netherlands, grasslands play an important role as biomass producers for dairy farmers as well as key ecosystems for biodiversity development. The new CAP aims to integrate these biodiversity services in the eco-schemes as a way of supporting farmers in the implementation of sustainable practices.

Grasslands herb-richness can be used as a KPI for biodiversity development. The Q index is one of the indicators that represent the ecological maturity of grasslands, linked to the presence and abundance of herbaceous species.

The aim of this project is to develop an automated AI-based system with which herb-rich grassland can be cost-efficient recognized based on the combination of satellite observations (spectral and radar) and photos from the field.

NEO has the goal of developing a Deep Learning- based model trained on remote sensing data and other biophysical parameters. The outcome of it should be combined with results from the photo-derived model.

The ground truth data (Q values, target of the prediction) and features to be included in the model have been provided by WUR for different parcels and years in the country. The characteristics and distribution of these features will shape and limit the performance of the model.

The data used in this work is presented in Chapter 2. The methods are described in Chapter 3. Chapter 4 presents the results and discussion. Finally, Chapter 5 evaluates the potential AI approaches.

¹ https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-2023-27_en

2. Data description

This chapter describes the data used in this work. In 2.1 a full description of the dataset is presented along with record counts. The target variables will be described in 2.2. The features of the dataset are presented in 2.3 and 2.4. Finally, 2.5 presents a filter for unrepresentative parcels.

2.1. Dataset

The dataset consists of unique records of parcel and year combinations, accompanied by measured biodiversity indicators, remote sensing (RS) markers, and temporal zonal-statistics (zonal-stats). There are also some geophysical features, such as ground-water table depth, elevation and soil type included as attributes. The total dataset spans from 2017 to 2022 and contains a total of 2020 records.

2.2. Target variables

Biodiversity can be described as the amount of diversity within an ecosystem, e.g. grasslands. It can be measured and explained in various ways. In this work, two biodiversity measurements are investigated; the Quality index and the Richness value, as described in 2.2.1 and 2.2.2, respectively.

2.2.1. Q-index

The Quality index Q is an indication value based on the phase of the development of low-in-species agricultural grassland to an herbal-rich natural grassland. The higher the Q-index, the more an intensively managed grassland with few species (index 1-2) has developed to an extensively managed herbal-rich grassland (index 3-6) (van 't Veer, 2020). The Q value of a grassland plot is determined by a weighted average of the measured species' abundance (coverage ratio) times the Q value attached to that species (Visser, Meijninger, Roerink, Kleyheeg, & Goedhart, 2022).

The corresponding Q-value for a parcel is the in-situ measured Q-value for that parcel and year². If multiple plots in the same parcel have been analysed, the mean of these measurements is taken. The resulting Q-value distribution per year is presented in Figure 1. High Q-value corresponds to herb-rich grasslands, whereas low Q-value corresponds to herb-poor grasslands.

From Figure 1, the Q value distribution varies per year, as also expressed by the yearly mean and standard deviation values in Table 1. The year 2022 has a bias towards lower Q values, while the rest of the years approach more to a normal distribution, although skewed towards higher Q values.

² [Herb rich grasslands & remote sensing.pdf](#)

Table 1 Yearly number of records and statistics for the Q-index in the dataset.

Year	Number of records	Mean Q value	Q value Standard Deviation
2017	367	3.80	0.96
2018	353	4.00	0.84
2019	355	3.87	0.97
2020	267	3.70	0.95
2021	125	3.54	1.02
2022	553	2.29	1.05
All	2020	3.40	1.19

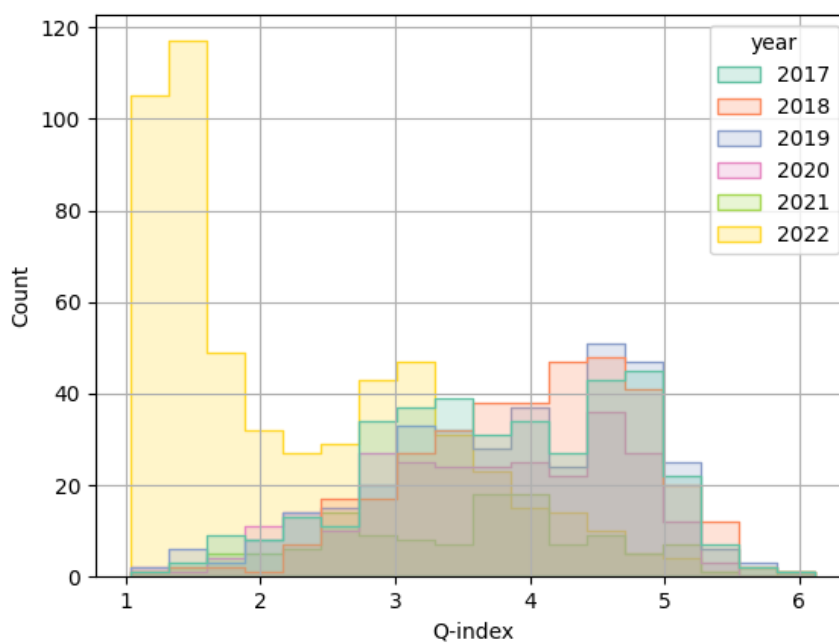


Figure 1. Q-value distribution in the dataset, separated per year.

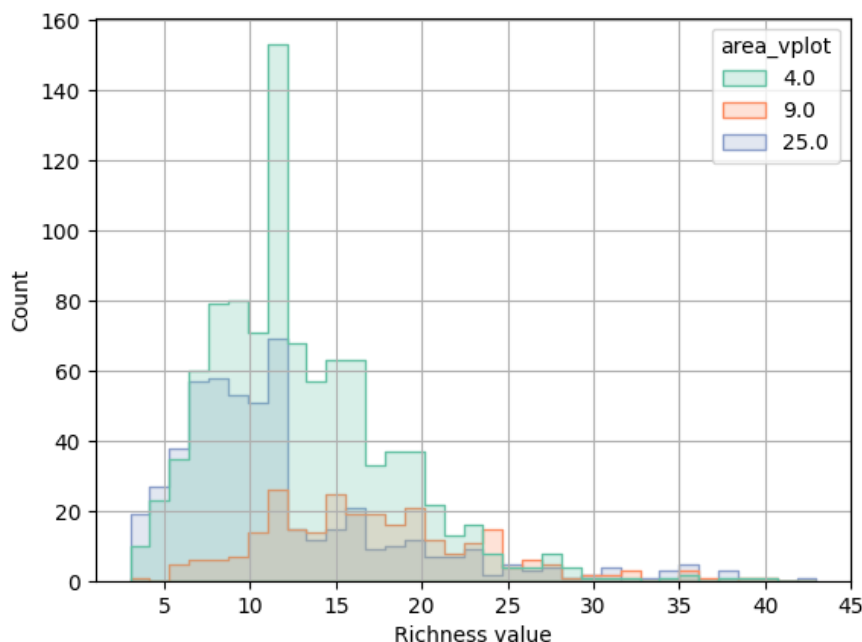


Figure 2. Distribution of the richness values for the three largest subsets based on vegetation survey plot area.

2.2.2. Richness value

The richness value is a measurement of the total number of species that exist within a defined area. It is thus area dependent (Magurran, 2021). Therefore, the data will be divided into subsets based on the area of the vegetation survey plot. The three largest subsets will be chosen for subsequent analysis, since the remaining subsets do not contain enough data to be modelled.

The number of records per plot area subset (*area_vplot*) and corresponding richness value statistics are presented in Table 2. It is observed that 4m², 9m² and 25m² are the three largest subsets. The distribution of the richness values for each of these subsets is displayed in Figure 2. Richness values vary from 3 to ~45. Most of the richness values are approximately between 5 and 20, confirmed by the mean richness values in Table 2.

Table 2 Number of records and statistics for richness value per vegetation survey plot area subset.

Area plot [m ²]	Number of records	Richness value Mean	Richness value Standard Deviation
4	957	12.91	5.39
9	269	17.21	6.47
16	86	15.23	7.58
25	523	11.69	6.77
100	39	17.60	9.86

2.3. Zonal statistics

The zonal statistics consist of temporal mean and standard-deviation satellite signal statistics for a given parcel in such a way that each record in the dataset is accompanied by a time-series of zonal-stats. Outliers of the time-series are considered to be outside a 3-standard deviation threshold window of the value distribution and are subsequently discarded. An example is the mean NDVI and is presented in Figure 3.

2.4. Remote-Sensing based Markers

The RS-markers consist of yearly and seasonal descriptive statistics of Sentinel-1 radar signals and Sentinel-2 optical signals, and often relate to agricultural management practices. Examples include the yearly mean Blue reflectance, the seasonal standard deviation of the NDVI, and the start day number (day-of-year) of the first mowing event. In total there are 171 RS-markers present in the dataset.

2.5. Filter for unrepresentative parcels

Furthermore, during data inspection it was observed that some records included in the dataset are probably not representative grasslands. This led to the development of a parcel filter as presented in

Table 3. The data models will be subjected to these parcel filters to determine which set will score best for that model. 174 records are excluded from both filters, having an irrelevant nature type and cover of trees/water > 25%.

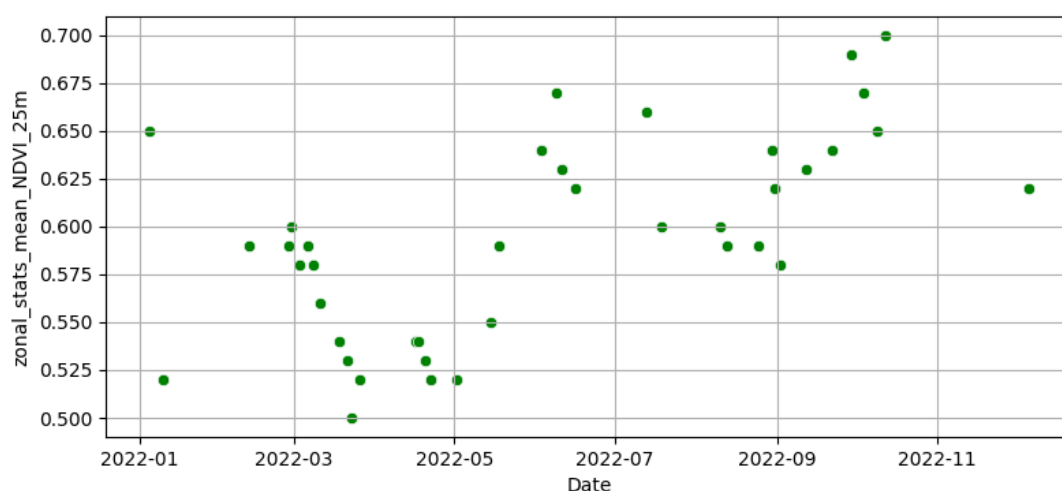


Figure 3. Zonal-stats time-series of the mean NDVI for field_id 22005512.

Table 3 Filter for unrepresentative parcels with corresponding number of parcels.

Parcels	Filter name	Number of records
Relevant nature type, cover of trees/water <= 25%	"Lenient"	1846
Relevant nature type and cover of trees/water <= 10%	"Strict"	1820

3. Method: Ridge Regression Model

The Ridge Regression model is a linear regression method that mitigates multicollinearity by adjusting the regression coefficients with a penalty term. This penalty term makes the model more robust and better suited for datasets with high inter-feature correlations (Gruber, 1998). Ridge Regression and its benefits are briefly presented in 3.1, in 3.2 additional data pre-processing is discussed, and 3.3 covers the model training methodology.

The Ridge Regression model is applied using the sci-kit learn python package (Pedregosa, 2011). Data (pre-processing is applied using the pandas python package (McKinney, 2010).

3.1. Ridge Regression Estimator

Analogous to the Ordinary Least Squares (OLS) estimator, the coefficients in the Ridge Regression model are given by

$$\hat{\beta}(\alpha) = (X^T X + \alpha I)^{-1} X^T y$$

where $\hat{\beta}(\alpha)$ are the estimated coefficients for each input variable, α is the ridge parameter and helps stabilizing the coefficient estimates in the presence of multicollinearity, X is the design matrix containing the values of the input features, I is the identity matrix and y is the dependent variable (Hoerl & Kennard, 2012). $\alpha = 0$ will result in the traditional OLS estimator equation. The predicted target variable $\hat{y}$ can then be calculated as

$$\hat{y} = X \hat{\beta}$$

This model has been selected for its capacity to manage multicollinearity, its pronounced interpretability due to its linear nature, and the accessibility to directly access the coefficients for each input feature. The latter allows for a direct feature importance analysis.

Multicollinearity is important to consider as the dataset contains correlated features, e.g. the max, min and difference of the zonal-statistics are inherently correlated (see 2.3). In the case of features being correlated some columns of X will be (almost) linear combinations, $X^T X$ will be close to singular (close to zero determinant), thus $(X^T X)^{-1}$ will be large. Therefore, the addition of αI will result in more stable coefficients; it balances small coefficients with small residuals.

3.2. Data pre-processing

In addition to the data preparation steps described earlier in 2.3 and 2.4, some additional preprocessing will be performed to facilitate the regression analysis.

For the Ridge Regression model the RS-markers are included in the features of the dataset. Next to the RS-markers, the *time-aggregated* zonal-stats features are also included. This has been achieved by taking the mean, standard deviation, minimum, maximum and the difference between the maximum and minimum of each time-series feature. All features are normalized towards mean and standard deviation.

Empty features will be omitted. This is the case for the features concerning the 6th mowing event (*m6*), and in the richness analysis also for *m4* and *m5*. Partly missing marker values are filled with that marker's mean value.

3.3. Model training

3.3.1. Spatial K-fold cross-validation

The Ridge Regression model will be trained using 5-folded cross-validation. One fold will act as test set, while the remaining make up the training set. This way, the (negative) impact due to having a limited number of records in the dataset will be reduced.

As mentioned in 2.1, each record in the dataset is a unique parcel-year combination. This means that multiple records can correspond to the same parcel but in a different year. Therefore, records are grouped by parcel before the fold construction of the cross validation. This way, potential overestimation due to spatial autocorrelation is reduced (Fauvel, 2020).

3.3.2. Statistical accuracy assessment

The performance of the models is evaluated and compared through the Root-Mean-Square Error ($RMSE$) and the coefficient of determination (R^2), as given below. y_i denotes the actual target variable value of the i^{th} record, $\hat{y}_i$ is the corresponding predicted value, $\bar{y}$ is the mean value of the validation sample and n denotes the number of records in the fold.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

The $RMSE$ and R^2 metrics of the model are calculated by taking the average of the 5 resulting $RMSE$ and R^2 metrics due to folding. Predicted Q values will be clipped with a lower boundary of 1 (predictions below 1 are not allowed).

4. Results and discussion: Ridge Regression Model

Using the methods described in Chapter 3, results are gathered which are split into two scenarios. The first scenario covers estimation of the Q-index (see 2.2.1) and is presented in 4.1. The second scenario presented in 0 will cover the estimation of the richness value (see 2.2.2).

4.1. Scenario 1: Q value

In Scenario 1, the performance of the Ridge Regression model will be analysed in predicting Q value. The parcel filter and possible excluding unrepresentative parcels will be analysed in 4.1.1. The best performing model is picked for subsequent analysis. In 4.1.2 the behaviour of the residuals is analysed and discussed to assess the model's validity. In 4.1.3 an assessment will be made on the relative importance of the features in predicting Q value. A recommendation will be made about which features can be removed for subsequent analyses. Finally, the annual variability of the Ridge Regression model will be analysed in 4.1.4. For all subsequent results, a ridge parameter of $\alpha = 5$ is used, as this led to the best performing models.

4.1.1. Parcel filters

Using the three parcel filters as described in 2.5, the following results as given in Table 4 are obtained. A Ridge Regression model on the whole dataset results in an RMSE of 0.716. By filtering the unrepresentative parcel filter, a Q RMSE of 0.676 is obtained, which is a 5.6% improvement by using the "lenient" parcel filter compared to no parcel filter, e.g. only keeping parcels with relative nature type and cover of water/trees $\leq 25\%$. It is noted that using the "Strict" parcel filter results in a similar, but slightly larger, RMSE metric of 0.682. The R^2 value is also highest for the "lenient" parcel filter model. Therefore, the "lenient" parcel filter is chosen for subsequent analyses.

Table 4 RMSE and R^2 metric values for a Ridge Regression model when predicting Q value using various parcel filters.

Parcel filter	RMSE	R^2
Keep all parcels	0.716	0.637
"Lenient"	0.676	0.674
"Strict"	0.682	0.661

4.1.2. Residual behaviour

To assess whether the model is valid or not, the model's residuals are analysed. The residuals are calculated by taking the difference between the actual Q value and the predicted Q value. Negative residuals thus represent a predicted Q value larger than the actual value, and vice versa.

In Figure 4, the residuals are plotted in $Q_{\text{pred}} - Q_{\text{actual}}$ space. The residuals approximately follow the line $Q_{\text{actual}} = Q_{\text{pred}}$. The distribution of the residuals are presented in Figure 5, where it can be seen that the residuals are normally distributed with mean (close to) zero, satisfying the assumption of normality for linear regression (Schreiber-Gregory & Bader, 2018). Looking at the residuals plotted against the area of the parcel in Figure 6, it can be seen that no clear dependency exists between residual and area.

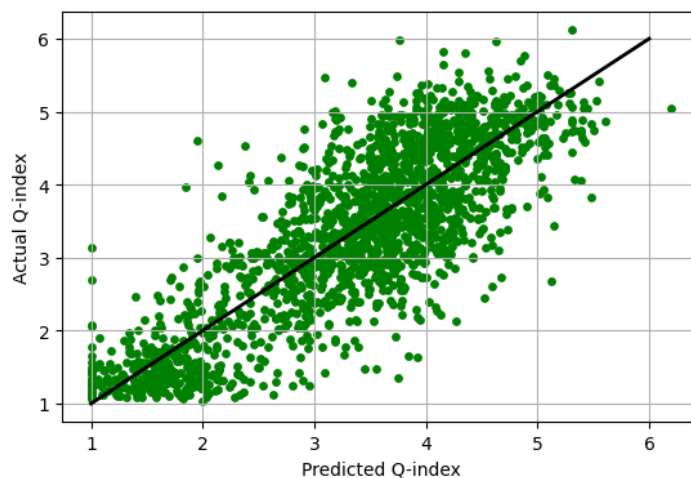


Figure 4 Residuals in Q-Q space for the Ridge Regression model.

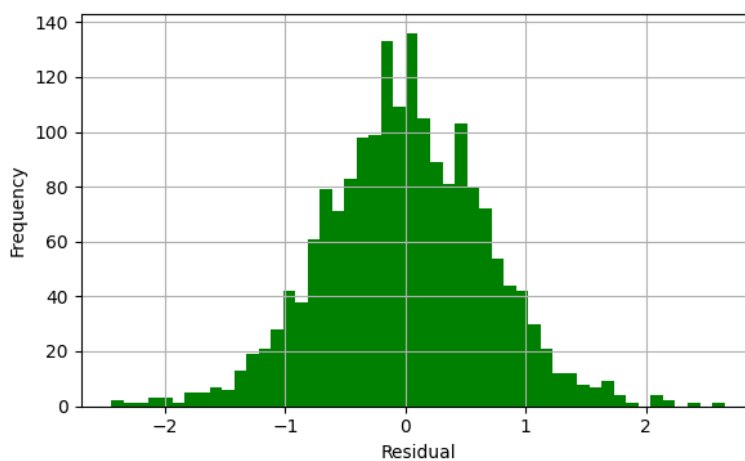


Figure 5. Residual distribution for the Ridge Regression model.

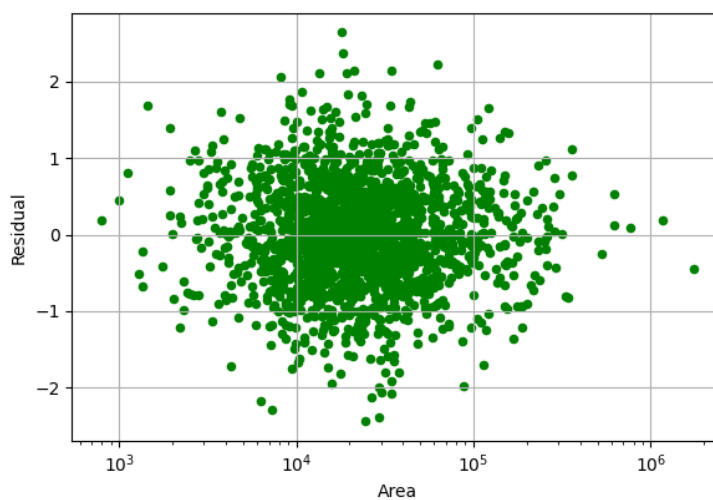


Figure 6. Residuals as a function of parcel area for the Ridge Regression model.

4.1.3. Relative importance of features

Approach

In order to determine the relative importance of the features, the magnitude of the coefficients of the trained Ridge Regression model is analysed. These coefficients represent the estimated impact of each feature on the target variable. The normalization of features (see 3.2) allow for direct comparison of the coefficients; features with larger coefficients are therefore considered more important. Additionally, the sign of the coefficients indicates the direction of the relationship between each feature and the target variable; a positive sign means that an increase in the feature's value is associated with an increase in the target variable, and vice versa for a negative sign.

As mentioned in 3.3, the model undergoes a 5-fold cross-validation process, resulting in the creation of five distinct trained models. Consequently, for each feature, there are five coefficient values. Therefore, the average of these five coefficients will be taken.

Furthermore, the features are grouped by signal, e.g. *zonal_stats_B02_mean_max*, *zonal_stats_B02_std_mean*, *zonal_stats_B02_mean_diff*, *zonal_stats_B02-_mean_mean*, etc. are grouped to *B02*. This is done to improve interpretability, as analysing the relative importance of 300+ features may not provide clear insight. Secondly, the intent is to provide the time series of features to a future model. For example, instead of providing the mean, max, min, diff and std of the *B02* band as input, the *B02* time series is provided as input. Therefore, grouping the *B02* band in this context may make more sense. Finally, since the processing of the data to generate the input features varies by data source, identifying the features with low importance could support a decision on whether it is cost-efficient to process all data (e.g. Sentinel 1 features are not the most important for the model performance and could be excluded as input data if the processing cost of them is too high).

By grouping the features by signal, the coefficient of the group will be analysed. This grouped coefficient will be determined using two distinct methods.

First, the sum of all coefficients in the group is taken. Summing the coefficients places more weight on the collective impact of the features within that group. As Ridge regression penalizes intercorrelated features (see 3.1), summing will not lead to an overestimated impact of a feature by double-scoring. The downside of this approach is that it has a bias towards feature group size; if a feature group contains more features it can result in a higher group coefficient than a smaller feature group. Taking the mean would solve this, however, it could also downplay the collective impact of the group if some coefficients are near-zero.

Second, the maximum coefficient (positive or negative) of a feature in the group is taken. By taking the maximum coefficient within each feature group, the most influential feature within each group is emphasized. This approach therefore pinpoints the most critical feature within each group. However, it may be sensitive to outliers. For example, if one feature within a group has an unusually large coefficient, it will dominate the assessment of group performance.

It is believed that a tandem assessment of the above two methods will result in a robust feature importance ranking and recommendation.

Results

Using the two methods presented above, the feature rankings are obtained and displayed in Figure 7 and Figure 8.

These figures show that similar feature rankings are obtained. The following general observations can be made:

- Sentinel-2 Red Edge (RE), Near Infra-Red (NIR) and Short-Wave Infra-Red (SWIR) bands are considered the most important features,
- Sentinel-1 Backscatter is more important than Sentinel-1 Coherence,
- The RS-markers are the least important features, the number of mowing events and/or the BRP crop code (*gws_gewasc*) being the most important. Under this category there are also some biophysical parameters included such as the elevation and ground water table.

Both rankings indicate the B07 (Sentinel-2 RE 3) and B06 (RE 2) bands as the most important features. The B08 (NIR), B11 (SWIR 1), B12 (SWIR 2), Sentinel-1 Backscatter VV, B8A (Narrow NIR) are also important features in both rankings, followed by Sentinel-2 indices such as NDVI, NDMI, NREVI, WdVI. Further analysis of the most important feature groups reveals that the mean signal values contribute the most to the feature group compared to the standard deviation values.

Differences between the two rankings in Figure 7 and Figure 8 also exist. For example, the WdVI index is the 5th most important feature when looking at the maximum coefficient in the feature group but drops to the 20th most important feature when looking at the sum of the coefficients. Similarly for Sentinel-1 backscatter VH; 9th most important when looking at the maximum, and 18th most important when looking at the sum. Probably one feature within this feature group dominates the group performance. Nevertheless, these feature groups remain important.

Finally, there is no clear cut-off point for which features to remove or not; the coefficients slowly decrease. It could be that for future models some features should be removed, e.g., to improve model convergence. In that case, based on the Ridge Regression analysis, it is recommended to keep the subsets in the order presented in Table 5. It should be noted that these recommendations follow from the use of a linear regression model. It could be that the feature importance will differ when using non-linear models.

Similar feature importance rankings have been found in literature, for predicting biodiversity or species richness. (Fauvel, 2020) found that S2 Red and Infra-Red bands are more important than S1 signals, where the latter was not found to significantly improve the prediction accuracy. (Muro, 2022) found Sentinel-2 Red-Edge and Near Infra-Red bands (from April-October) to be the most important.

Table 5 Recommendation for which features to keep based on the Ridge Regression analysis.

Order to keep	Signals	Observation
1	Sentinel-2 RE, NIR and SWIR	Most important features
2	Sentinel-1 backscatter	VV more important than VH
2	Sentinel-2 Indices	WdVI, NREVI and NDVI being the most important
3	Sentinel-2 visible bands	RGB and/or HSV representation
4	Sentinel-1 coherence	VH more important than VV
5	RS-Markers	possibly keep number of mowing events and/or BRP crop code

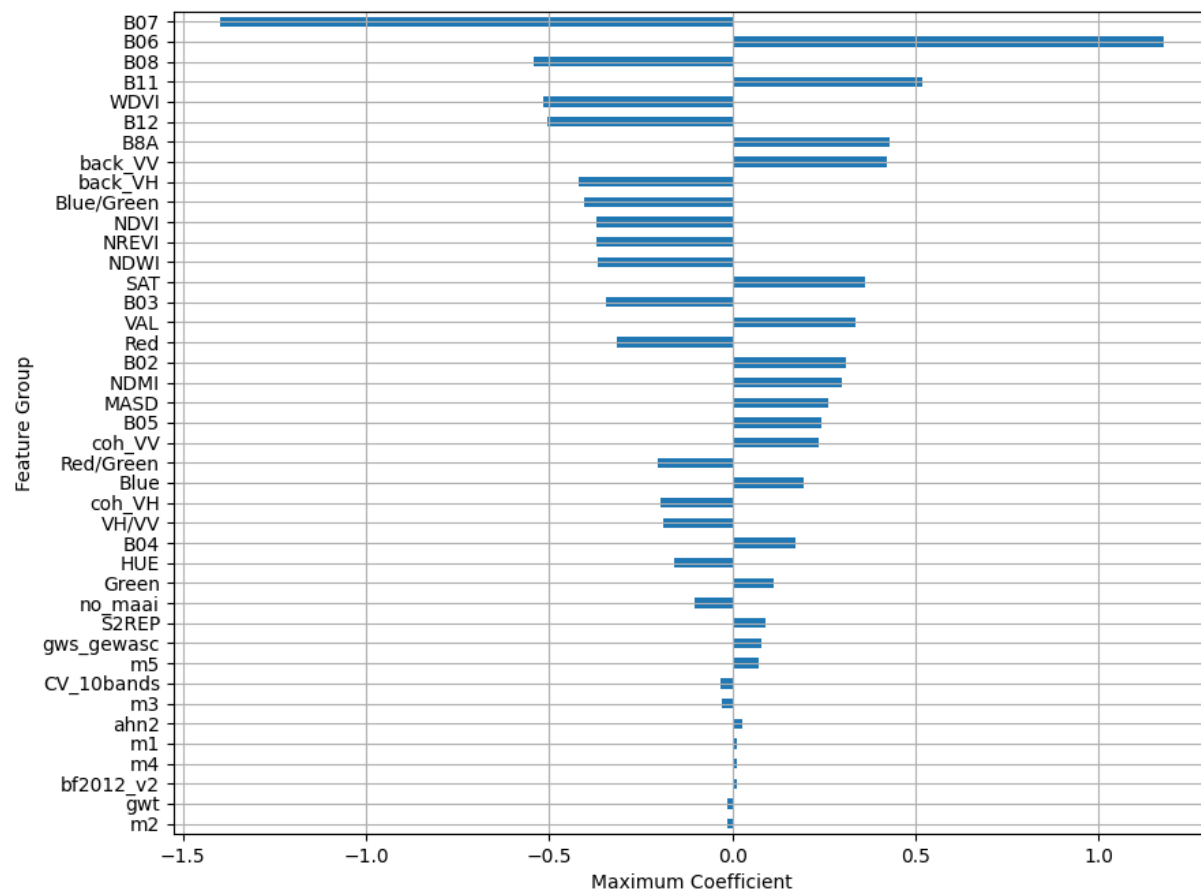


Figure 7. Grouped feature ranking determined by the maximum coefficient in the feature group.

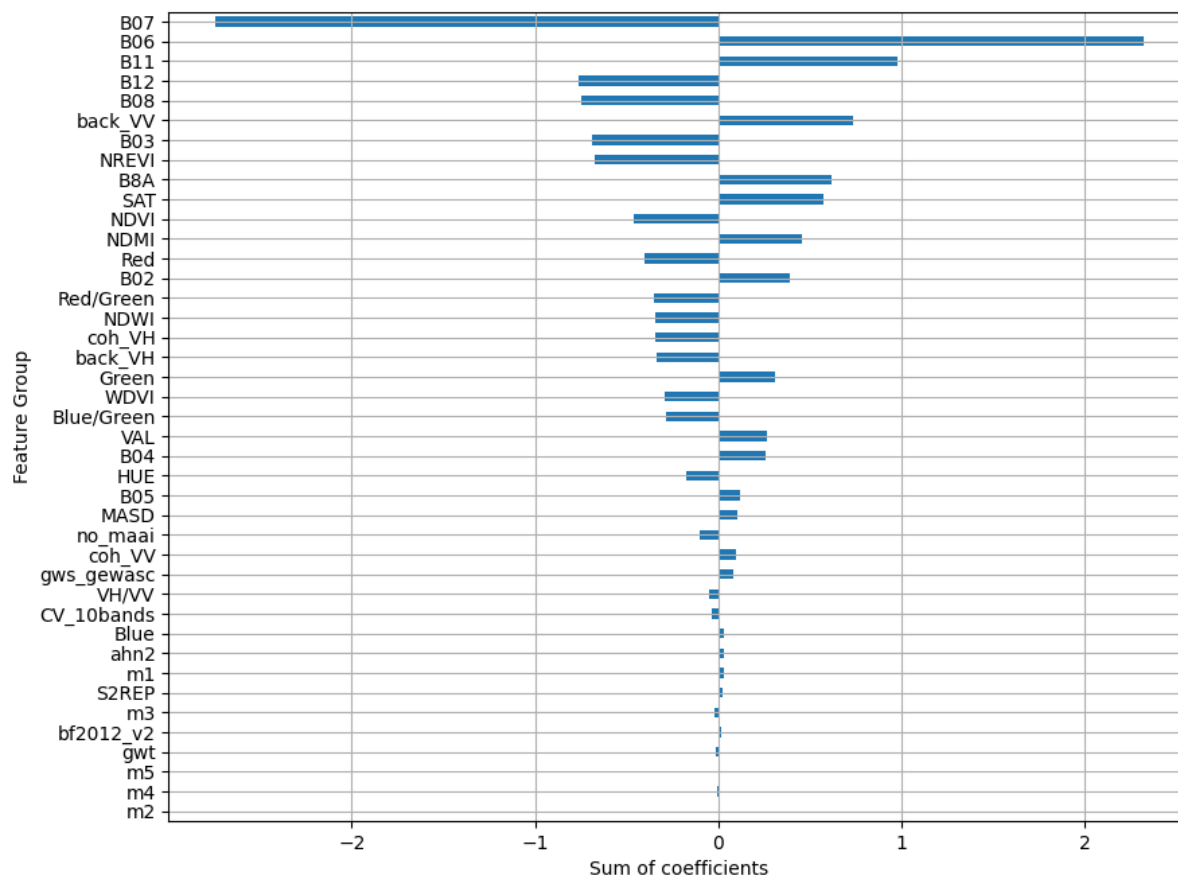


Figure 8. Grouped feature ranking determined by the sum of coefficients in the feature group.

4.1.4. Annual variability

Approach

The annual variability of the Ridge Regression model will be analysed via two distinct methods.

First, the model's RMSE metric will be calculated for each year the dataset spans. The ground-truth data, grouped by year, will be used to get the annual variability in the model's RMSE metric for all 2020 labelled records in the dataset. For each fold, the corresponding model is used for prediction on the fold's test set to prevent training data leakage into the predictions.

Second, the annual variability of the model will be assessed by prediction on additional unlabelled records. Besides the 2020 labelled records in the dataset described previously, 11864 records without a measured Q value are also present in the dataset. These records cannot be used for training the model but can be used as prediction set.

As mentioned in 2.1, each record is a unique parcel-year combination. By using the provided parcel ID history table, we can use these parcel IDs to group the predictions of the same parcel over different years. This way the temporal consistency of the model can be assessed. The subsequent grouping of the predictions results in predictions for 2017-2022 for 2576 unique parcels. It should be noted that these parcels have not been visually checked for relevant nature type and cover of water/trees $\leq 10\%$.

Both methods will be executed using the best-performing model from Table 4, which is the model with "Lenient" parcel filter.

Results: Yearly RMSE

The model's yearly RMSE and R^2 values are presented in Table 6. In general, the RMSE values are decreasing and R^2 values are increasing over the years.

Especially for the year 2022, the model performs best. The ground-truth Q distribution (Figure 1) for this year has a bias towards lower Q values. This may suggest a performance bias of the model towards lower Q values. To test this, the RMSE as a function of Q value is calculated and presented in Table 7. There is no trend for the RMSE as a function of Q value; lower Q values do not necessarily mean a lower RMSE. On the contrary, the lowest RMSE is found for $3 \leq Q < 4$, near the total mean Q value of 3.40 (Table 1).

Furthermore, the difference in yearly RMSE may be due to environmental variations over the years currently not being included in the model, e.g. the temperature or precipitation varying over the years, possibly influencing the prediction of the Q value.

Table 6 Yearly RMSE values for the Ridge Regression model in predicting Q -index.

Year	2017	2018	2019	2020	2021	2022
RMSE	0.764	0.754	0.724	0.704	0.612	0.542
R^2	0.334	0.188	0.421	0.440	0.644	0.721

Table 7 RMSE values as a function of Q value.

Q values	$1 \leq Q < 2$	$2 \leq Q < 3$	$3 \leq Q < 4$	$4 \leq Q < 5$	$5 \leq Q < 6$	$Q \geq 6$
RMSE	0.641	0.741	0.583	0.678	0.967	0.809

Results: Yearly Q prediction of the same parcel

The temporal behaviour of the model's predictions is presented in Figure 9, Figure 10 and Figure 11. Figure 9 shows the distribution of the difference between the first and last predictions (in time) of the model for a given parcel. A positive difference therefore means an increasing predicted Q value in time for the same parcel, and vice versa. The median (0.006) is very close to zero and the box and whiskers are well centred, indicating no overall temporal increasing or decreasing tendency in the predicted Q value for the Ridge Regression model for the prediction set. This means that no significant increase or decrease of the parcel's predicted herb-richness from the first to the last years is observed.

This observation is further demonstrated in Figure 10, where the distribution of the predicted Q values for each year is shown. Apart from yearly variations, there is no significant trend in time, with median Q values oscillating between 3 and 4 for all years.

The yearly variabilities are further expressed in Figure 11 which shows the distribution of the standard deviation of the predictions over the years for each parcel. The model's prediction varies on average 0.34 in Q value over the years. Outliers range from 0.43 to even 2.97 of standard deviation in Q -value's prediction. Such extreme cases could be, for example, due to unrepresentative parcels in the dataset or for changes in land use over time.

The last effect is further explored by visualizing three parcels over the years. In Figure 12, 3 different cases are displayed: a parcel with a predicted negative progression of the Q index over the years, with a high standard deviation of those predictions. Another parcel with a gain of biodiversity indicated by an increase of the Q index between the first and last prediction of more than 2 points (with a considerable standard deviation). Finally, a third case where the difference of the predicted Q is below 0.5, with a low standard deviation, which points towards a subtle improvement of biodiversity over the years.

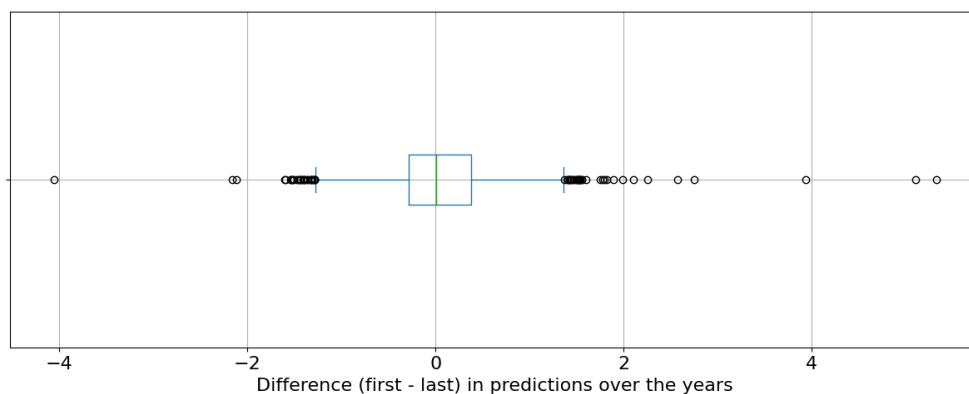


Figure 9. Distribution of the difference in first and last predictions in Q value over the years.

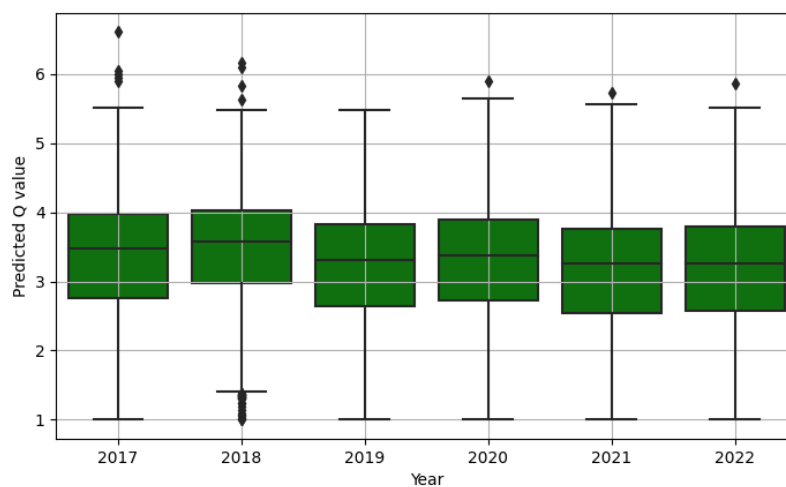


Figure 10. Distribution of predicted Q value for each year.

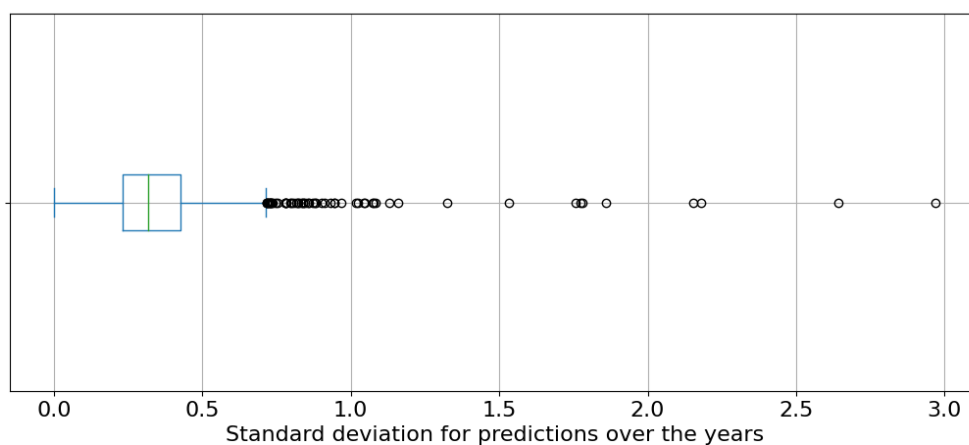
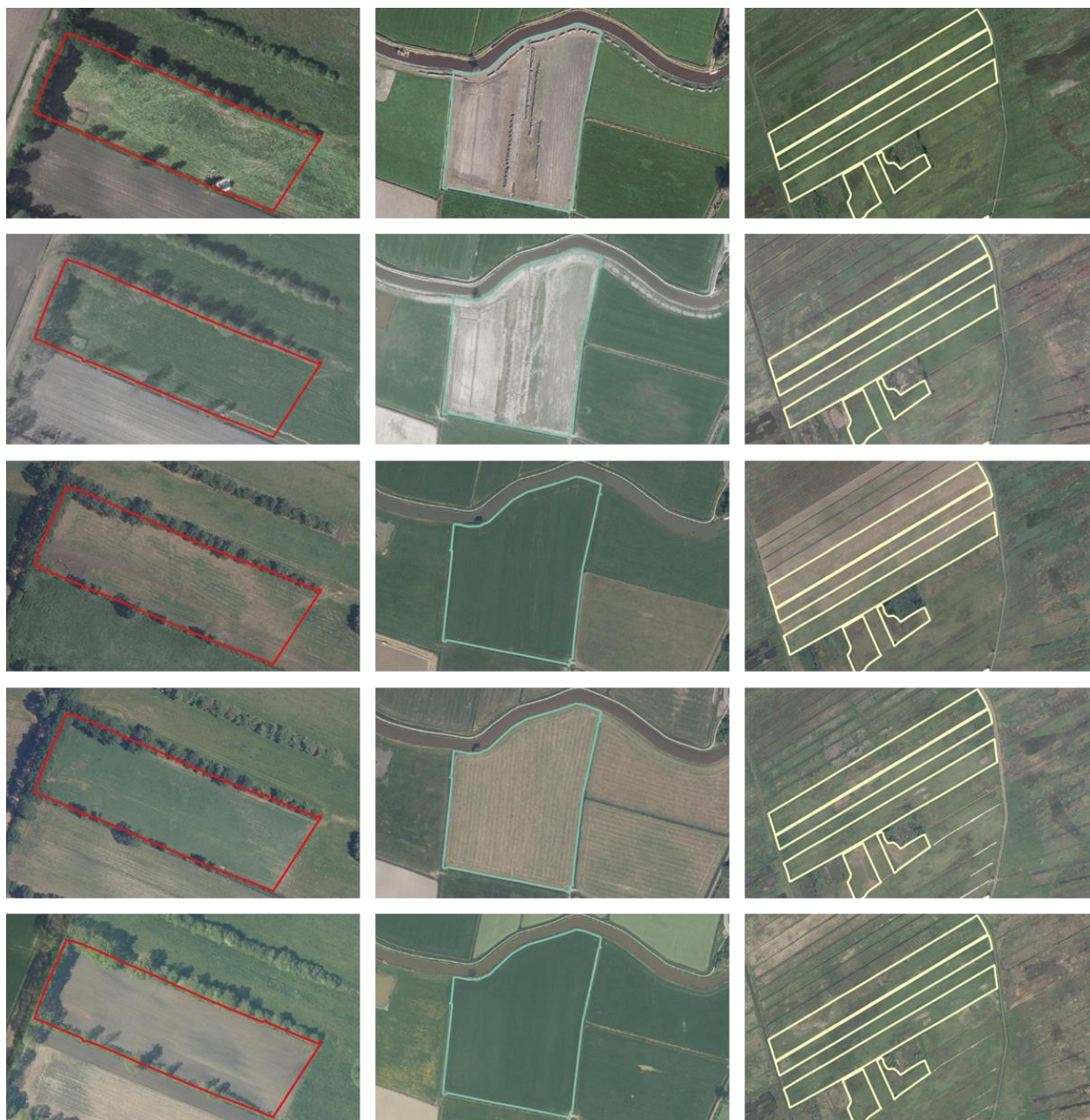


Figure 11. Distribution of the standard deviation of the Q-value predictions over the years for each parcel.



Field ID	A	B	C
Difference between first and last year	-4.1	2.7	0.4
Standard deviation in prediction over the years	1.8	1.0	0.3

Figure 12. Aerial images of parcels (field_id = A in the left column, field_id = B in the center, and field_id = C in the right column). On the top row, images of 2017, 2018 in the second row, 2019 in the third row, 2020 in the fourth, and 2022 in the bottom row.

4.2. Scenario 2: Richness value

In this scenario, the performance of the Ridge Regression model will be analysed for predicting the richness value, in similar fashion to Scenario 1.

As previously mentioned in 2.2.2, the data will be split based on the area of the vegetation survey plot. Plot areas of 4m², 9m² and 25m² are chosen as they contain the largest number of records. The Ridge Regression results for these subsets for each parcel filter (see 0) are presented in Table 8.

Table 8 RMSE and R^2 metric values per area subset for a Ridge Regression model when predicting richness value using various parcel filters.

Parcel filter	Area subplot [m ²]	RMSE	R^2
Keep all parcels	4	5.914	-0.234
	9	9.688	-1.399
	25	6.251	0.048
“Lenient”	4	5.678	-0.111
	9	8.140	-0.626
	25	6.235	0.019
“Strict”	4	5.665	-0.101
	9	8.396	-0.709
	25	6.132	0.043

As can be seen from Table 8, the best models are found for a plot area of 4 m², indicated by the lowest RMSE values of the model for that data subset. However, the R^2 values are close to zero or even negative. This means that the model predicts worse than a model that just predicts the average richness value (see 3.3.2). Clearly, the Ridge Regression model fits the richness value data poorly.

Several attempts have been made to improve the model’s performance:

1. Multiple feature sets have been tried and tested, e.g. only Sentinel-1 (S1) and Sentinel-2 (S2) data including and excluding S2 indices, only S1, only S2, including or excluding the RS-markers and using the most important features resulting from the Q-index analysis (see 4.1.3). All combinations have resulted in similar RMSE and R^2 metric values.
2. All above combinations have been analysed for various values of the Ridge parameter α (see 3.1). α -values of 0.1, 0.5, 1.0, 2.0, 5.0, 10.0 and 15.0 have been tried. Again, no improvements in the model’s performance have been observed.
3. The model has been trained and evaluated with and without normalized richness values. No improvement in R^2 is observed. Therefore, the model’s performance is not defined by the target variable range. This is confirmed by scaling the Q-index to the same range as the richness values and repeating Scenario 1: similar performance as in 4.1.1 is observed.

The model’s bad performance can be due to (a combination of) several factors, such as:

1. The Ridge model could be too simple to predict richness values from remote sensing data. Non-linear models such as Random Forests or Deep Learning may be better suitable for predicting richness value, as demonstrated by (Fauvel, 2020) and (Muro, 2022).

2. The reduced dataset due to sampling of the vegetation survey plot could have a negative effect on the model's performance, as less data is available for training and validation. However, repeating Scenario 1 for the same plot area subsets results in $R^2 \approx 0.4$ values, lower than reported in 4.1.1 but remains positive.

Feature importance analysis is therefore inapplicable as no suitable model has been found.

5. Method: Convolutional Neural Network

5.1. Data pre-processing

In addition to the data preprocessing steps described earlier in 3.2, some additional steps are taken to use the signals' temporal component.

The time-series in the dataset have variations in observation timestamp across different years which leads to a temporal misalignment in the dataset. To prevent this, the time-series are resampled to weekly intervals. Any missing values are linearly interpolated and if needed backwards- and forwards filled.

Subsequently, images are created from the timeseries by means of horizontal concatenation, resulting in images having dimensions (number of weeks, number of signal bands).

In accordance with WUR, a decision has been made to include Sentinel-2 + NDVI mean and standard deviation and Sentinel-1 mean zonal stats timeseries in the images.

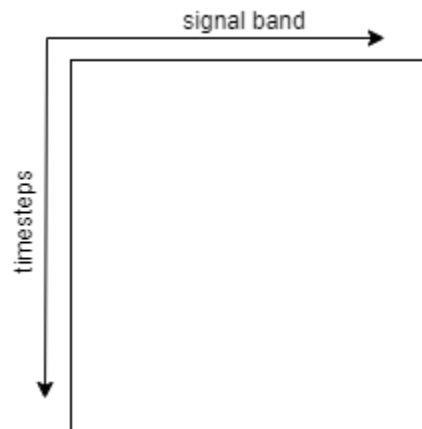


Figure 13. Dimensions of the timeseries "image".

5.2. Model architecture

The Ridge Regression model is only able to learn a linear relationship between the features and labels. Therefore, it is hard to learn patterns in the more complex time series data. Still, the time series likely contains much more information, so the expectation is that a model that would be able to capture the change in the different signals over time could outperform the Ridge Regression model.

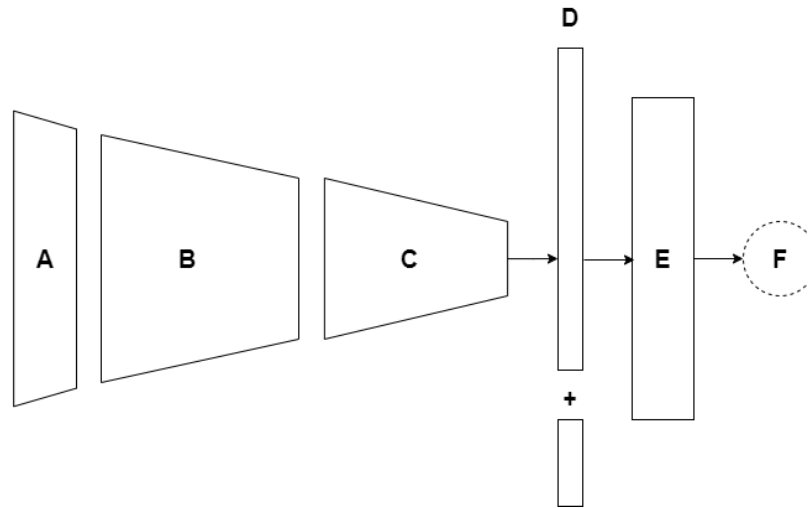


Figure 14. Schematic overview of the final CNN architecture used.

For this task, we have designed a Convolutional Neural Network (CNN) that looks at the temporal relations in the “image” created as described in paragraph 5.1. Figure 14 shows an overview of this architecture. The network consists of two different steps, in the first part (A, B and C), the temporal data is processed and condensed into a feature vector that describes how the features relate to each other and how they change over time. To this flattened vector, we concatenate the categorical features, soil type (*bf2012_v2*), ground water levels (*gwt*) and the parcels height (*ahn2*) in *part D* after which several dense layers come to a final feature vector in *step E*, combining the temporal information with these categorical values. This is then finally converted into either a class or a regressed Q value in *F*.

To further specify the temporal part of the network: in part A, a convolution of $1 \times N_{\text{bands}}$ (timesteps x bands) and 32 kernels is applied over the bands. Thus, independent of the time component, the input features are condensed and combined. This step is followed by several 1×1 convolutions in part B that learn combinations of the channels created in the previous step. Finally, in step C three convolutions of size 4×1 are applied, which captures the temporal relations within a spread of 4 timesteps at each stage. Resulting in a field of vision of 4, 16 and 64 “weeks” respectively.

The final model architecture was the result of several small iterations. In these small experiments we varied the kernel sizes and the number of layers in parts A, B, C and E.

The regression and classification variants of this model primarily differ in F, where for regression this was a single output node with linear activation, and for classification this was a layer with N_{classes} nodes, and SoftMax activation.

For the classification variant of the model, we also experimented with sigmoid activation, also in combination with soft labels (Nguyen, 2014). For the soft labels, the classes left and right of the correct

class were assigned a value of 0.2, and the ground truth class was given a value of 0.6. This way, the model would be punished less for a small mistake than for a big mistake, with the hypothesis that this would help the model overcome some of the inherent uncertainty in the ground truth. In the end this approach was not chosen, as it did not increase the final performance.

5.3. Model training

The model was trained using the Adam optimizer, with a learning rate of 0.001.

For regression, the target Q value was capped between 1 and 5, and the final output node had a linear activation function. The predicted value was only clipped to be ≥ 1 . For this task the Mean Squared Error was chosen as the loss function.

For classification, the target Q value was binned into 8 bins of 0.5 Q size. The network was then given 8 output nodes with a SoftMax activation function. For this task the Categorical Cross Entropy was optimized. For classification, the addition of focal loss to the loss function was tried, but likely due to some uncertainty in the ground truth and the relatively limited number of training samples this did not improve performance.

6. Results and discussion: Convolutional Neural Network

Using the approach defined in Chapter 5, the results will be reported and discussed grouped by experiments described below.

- I. Regression versus classification performance (6.1)
- II. Seasonal effect of timeseries on model performance (6.2)
- III. Effect of Q-index range on model performance (6.3)

6.1. Experiment I: Regression versus Classification

A CNN regression approach results in an RMSE of 0.647 Q, outperforming the Ridge Regression model having an RMSE of 0.676.

Binning the Q-index with bin-size 0.5 Q results in a confusion matrix presented in

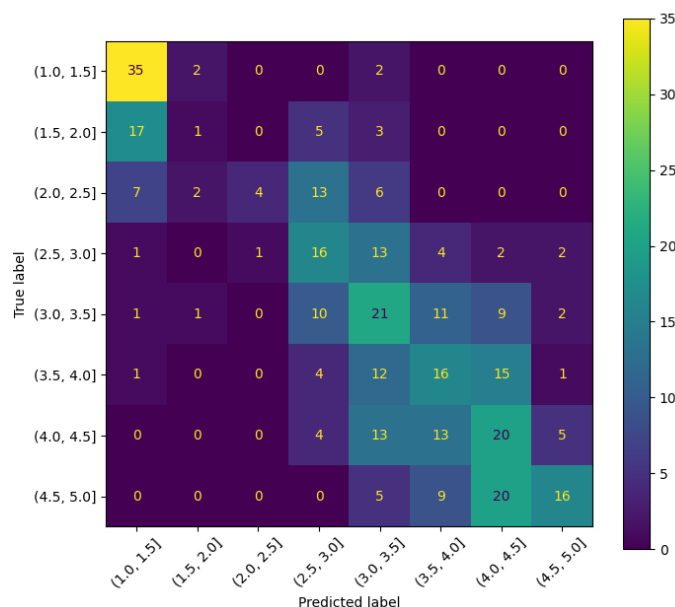


Figure 15, having an accuracy over 1-bin of 37.4% and over 3-bins of 76.2%. The model is more accurate for lower Q values ($Q < 3$) than for higher ($Q > 3$) Q values, likely due to a parcel being more heterogeneous for higher Q-index values, which also results in a less representative ground truth plot.

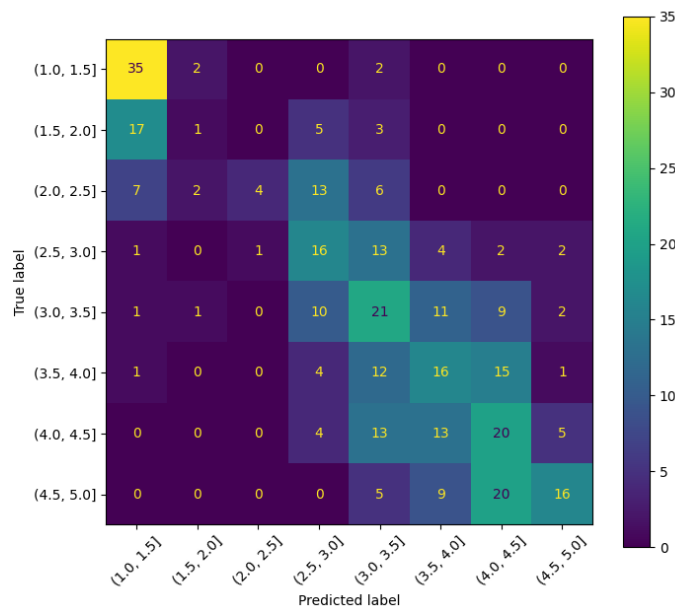


Figure 15. Confusion matrix of a CNN classification approach.

6.2. Experiment II: Seasonal effects

Previously, the whole date range (January – December) of the time-series is included in the image. This experiment is about varying this date range, thereby removing the (cloudy) winter months. This may improve the model's performance as there will be missing data points.

The CNN regression RMSE metrics for various time-series date ranges are presented in Table 9. The date range February – November results in a slightly better RMSE than the other date ranges, likely due to less missing data points due to cloudy winter conditions.

Table 9 CNN Regression RMSE for different input satellite time-series date range.

Date Range	RMSE
January – December	0.622
April – October (Muro, 2022)	0.630
May – October (Fauvel, 2020)	0.676
February – November	0.598
March – October	0.607
April - August	0.729

6.3. Experiment III: Q-index range effects

The performance metrics for a CNN regression model for the full Q 1-5 range and the low Q 1-3 / high Q 3-5 models are presented in Table 10. A date range of February – November is used for the time-series as determined by 6.2. The MAPE refers to the Mean Average Percentage Error, which is relevant when directly comparing the different models. The RMSE will automatically decrease as the target range decreases, therefore, a relative error metric is more suitable.

From Table 10 the average model error decreases from 16.2% to 15.3% and 11.5% for a low Q 1-3 and a high Q 3-5 model, respectively. It begs the question whether this increased performance outweighs having two models in practice, meaning that there should be some way of choosing between which model to use for a parcel. A CNN classification approach with bins [1, 3] and (3, 5] may be suitable for this purpose. Therefore, this model has been trained for and the confusion matrix is presented in Figure 16. This model reaches an accuracy of 81.7%, which will add to the regression errors presented in Table 10. Therefore, a single Q 1-5 model may be more suitable for production purposes.

Table 10 CNN Regression RMSE and MAPE metrics for different Q ranges.

Q-range	RMSE	MAPE	Number of training samples
[1, 5]	0.598	0.162	1486
[1, 3]	0.372	0.153	629
(3, 5]	0.524	0.115	857

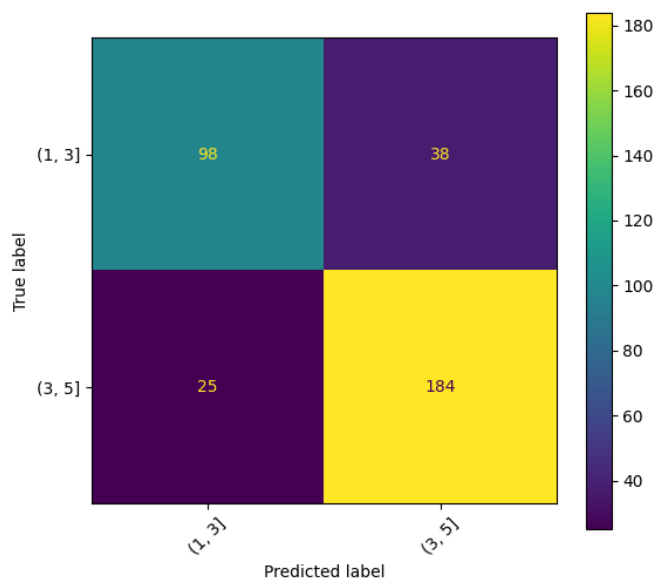


Figure 16. CNN Classification confusion matrix for a two-binned low/high Q model.

6.4. Nation-wide prediction

For the nation-wide prediction we use the CNN regression configuration trained on timeseries ranging from February – November on the full 1-5 Q range. We predict for the years 2021 and 2022 separately and plot the distributions in Figure 17. As can be seen, Q values predictions are generally lower in 2022.

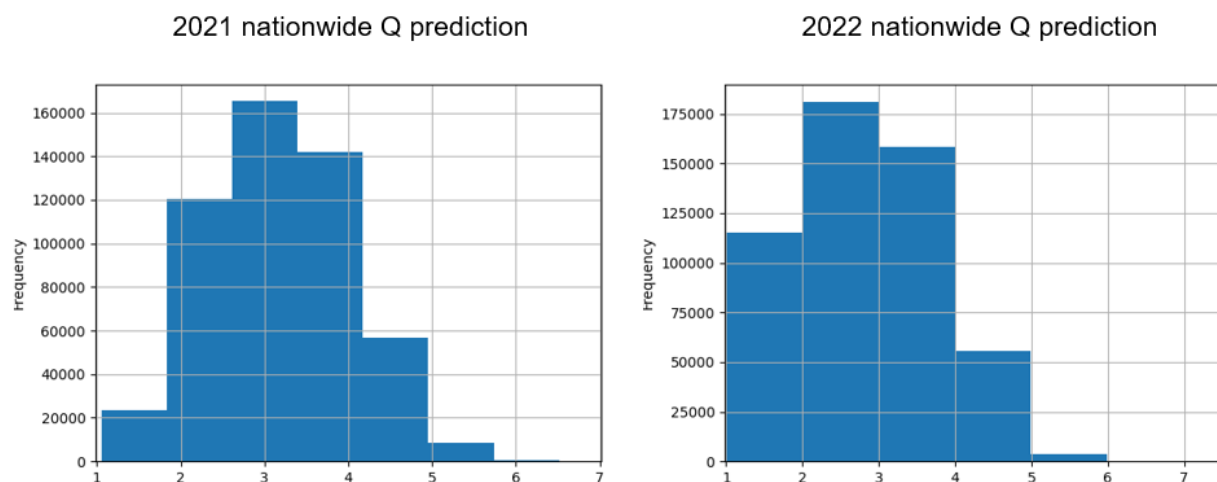


Figure 17. Nationwide prediction distribution for 2021 and 2022.

On order to compare the Q estimates for the same parcel for different years, a spatial overlay was applied, and parcel intersects were used for the comparison (Figure 18).

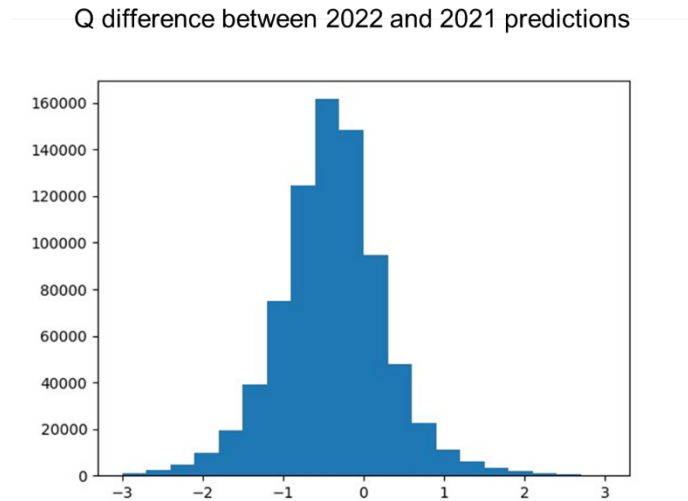


Figure 18. Distribution of the differences of predictions for each parcel between 2021 and 2022. For most of the parcels, 2022 Q values are around 0.5 lower than for 2021.

In figure 19 an example of Q predictions for parcels in Flevoland, for year 2021 and year 2022 is displayed.

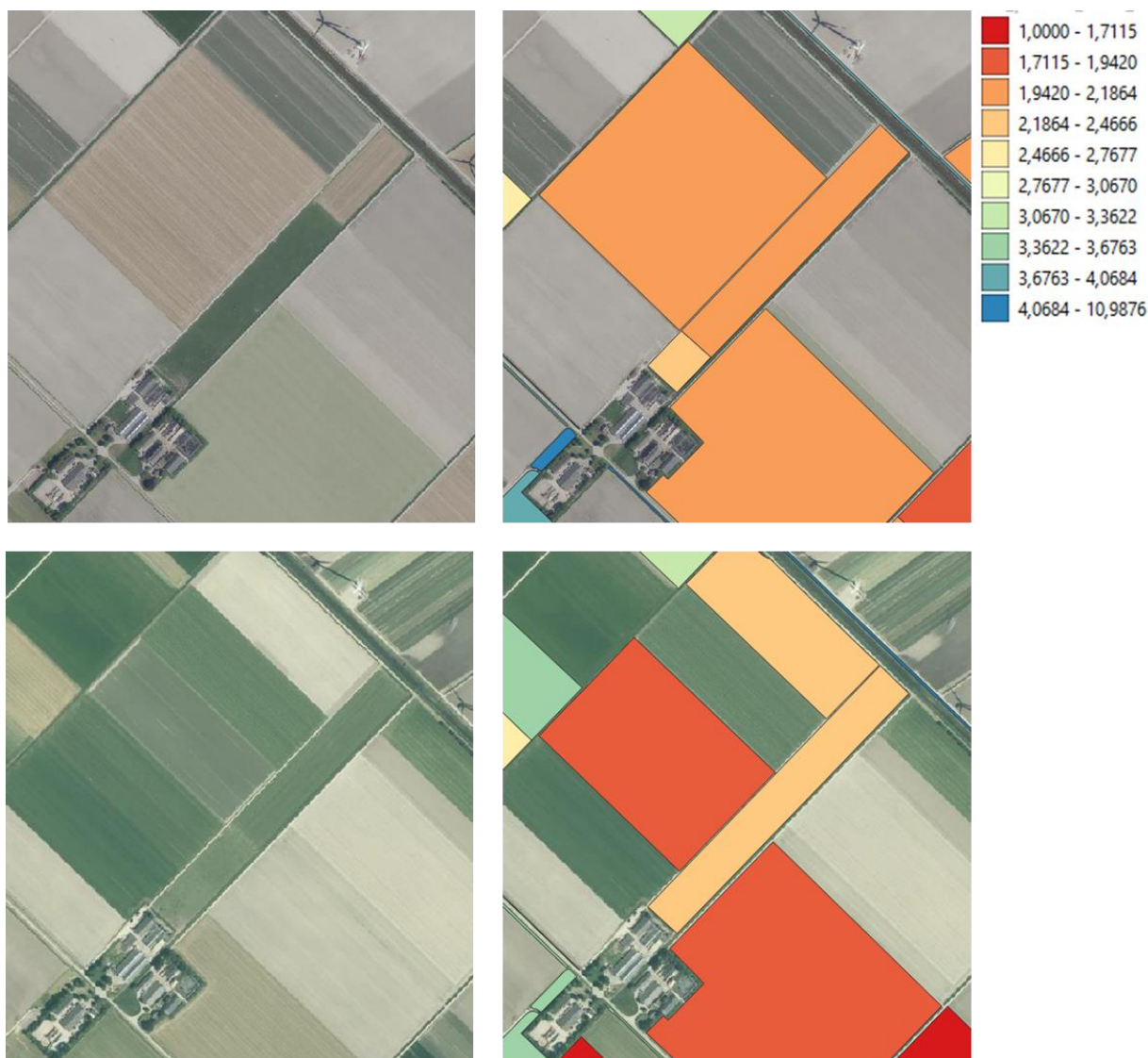


Figure 19. Detailed view of the Q value mapping for parcels in Flevoland. On top, aerial photography and Q values for 2021; in the bottom, aerial photography and Q values for 2022.

7. Evaluation of potential AI approaches

This work has focused on the Ridge Regression model, which is a linear regression model. There is a possibility that a pattern or relationship between a biodiversity indicator and remote-sensing data is not fully captured by the current Ridge Regression model. Limitations of the Ridge Regression model have been observed while predicting the richness value (see 4.2); it could therefore be that other models may be more suitable. Even in the case of Q value predictions, a better performance could be obtained using other (non-linear) models.

Additionally, more complex models could be more suitable for incorporating the temporal component of the data by directly providing the time series of the satellite signals as input (vs yearly aggregates).

This chapter evaluates potential AI approaches in predicting the previously mentioned biodiversity indicator. Proposed models are described in 7.1, together with their advantages and disadvantages. In 7.2 the suitability of the models is assessed.

This chapter addresses the request made by WUR.

7.1. Models

In the following sections, the proposed models are described in their description, advantages, disadvantages and a possible approach using the data provided in this work.

7.1.1. Recurrent Neural Network (GRU)

Description

The Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN) designed for sequential data. The recurrent layers in the network allow information to persist from one timestep to the other, meaning it can have a long-term memory (Chung, Gulcehre, Cho, & Bengio, 2014).

Advantages

- Able to capture long-term sequential patterns in the data.
- Can have variable-length sequences.

Disadvantages

- Computationally- and data-intensive.
- Prone to *vanishing gradients*; the gradients represent the rate of change of the loss (error) with respect to the network's parameters. During training, the network adjusts its parameters based on these gradients. These gradients can become extremely small as they are propagated through the layers of the network and may cause the training to stall or slow down (Pascanu, Mikolov, & Bengio, 2013).

Approach

The RNN model is constructed by three GRU layers, to learn from the timeseries, followed by some fully connected layers to get a final regressed value. The GRU layers learn a feature vector for each timestep, remembering useful information from previous timesteps as they go. The first two layers return such a feature vector for each timestep, but the last returns only a single final feature vector (that of the last

timestep) which is fed to the final fully connected layers. This way the model learns to represent the entire timeseries in this final feature vector.

7.1.2. Convolutional Neural Network

Description

The Convolutional Neural Network (CNN) is a type of neural network that is primarily used for image analysis but can be adapted to time series data. The convolutional layers in the network can capture patterns and relationships within the data and automatically learns the relevant features (Zhao, Chen, Liu, & Wu, 2017).

Advantages

- Robust to local variations in data.
- Automatically learns the relevant features.

Disadvantages

- Limited ability to capture long-range dependencies.
- May not perform well on non-spatial data.

Approach

An 'image' of the time series can be constructed as input for the CNN. This image has dimensions: number of features x timesteps. First, meaningful combinations between the features can be found by using a kernel size of the number of features by 1, creating a single feature vector per timestep. Several 1x1 kernels can be used to further compute combinations of these features. Subsequently, the features will be combined over time, by using a 3x1 kernel, looking at three consecutive timesteps each time. The model is finalized by a few fully connected layers to reach a final prediction. RS-markers can either be included as a constant feature (in time) in the image, or via a concatenation layer after the convolution layers.

7.1.3. Random Forest

Description

The Random Forest (RF) is an ensemble learning method that combines multiple decision trees to make predictions. It can handle both numerical and categorical data and is known for its versatility and simplicity (Amini, Saber, Rabiei-Dastjerdi, & Homayouni, 2022). The final prediction of the random forest is obtained by taking the mean of the predictions of all individual trees in the forest.

Advantages

- Provides feature importance scores.
- Can perform well for shorter time series and/or smaller datasets.

Disadvantages

- May not capture more complex relationships as effectively as other models.
- Not explicitly designed for sequential data, which can limit its performance for handling time series data.

Approach

The same dataset as the Ridge Regression model will be used for an initial metric of the Random Forest. An initial optimization is based on modifying the number of decision trees. Each tree within the ensemble

will be constructed by drawing a sample with replacement from the training dataset. All features from the sample will be used for training. No maximum depth of the trees is provided.

7.1.4. Dense Neural Network

Description

The Dense Neural Network (DNN) is an artificial neural network composed of multiple layers of interconnected neurons. It is highly versatile and designed to model complex relationships and patterns in the data (Fawaz, Forestier, Weber, Idoumghar, & Muller, 2019).

Advantages

- Highly flexible; domain-specific knowledge can be incorporated through architecture design.
- Able to model complex non-linear relationships.

Disadvantages

- May require extensive hyperparameter tuning.
- Typically requires a lot of data to extract its potential.

Approach

The RS-markers and RS-data time series are flattened into a single vector and provided as input to the network. A similar architecture as in literature will be used for an initial model metric, e.g. as in (Muro, 2022).

7.2. Suitability

An initial model architecture for the above models has been constructed to get a global indication of the suitability for these models. These models allow to directly incorporate the temporal component of the remote sensing signals, achieved by including the time series in the dataset, instead of time-aggregated features (see 3.2).

7.2.1. Data Preparation

Some additional steps with respect to Chapter 2 are taken to make use of the temporal component of the signals. The time-series in the dataset have variations in observation timestamp across different years which lead to a temporal misalignment in the dataset. Therefore, the time-series are resampled to weekly intervals. Any missing values are linearly interpolated or else forwards- or backwards-filled.

7.2.2. Results

All models predict the Q value (2.2.1) on the dataset filtered using the “Lenient” parcel filter 4.1.1). Making use of the approach for each model described in 7.1, the results in Table 11 are obtained. From Table 11 the initial performance of the Random Forest is the best among the new models. However, the Ridge Regression model performs better. Nevertheless, the new models could be further optimized for better performance.

Table 11 Initial model metrics in predicting Q value to assess the suitability of potential AI approaches.

Model	RMSE	R^2	Notes
Ridge Regression	0.676	0.674	Baseline model, see Chapter 4
RNN (GRU)	0.732	0.619	3 GRU layers (64-128-64 units) followed by 3 Dense layers
CNN	0.707	0.644	7 Convolutional layers followed by 4 Dense layers
RF	0.678	0.672	600 decision trees
DNN	0.776	0.571	3 layers x 64 neurons followed by 1 l x 1 neuron, same as (Muro, 2022)

8. Optimization of CNN model

After the exploration of different models' performance in previous sections, the CNN approach seems the most appropriate for our problem. Still, the results show some trends that are not expected from an agro-ecological perspective. In this chapter, some of extra analyses performed on the ground truth data and model tweaking attempts are explained.

8.1. Map Q values vs training data to double check whether south areas values are indeed lower than the predictions

The Q predictions, for all models, show north-south trends, difficult to explain in ecological terms. NEO wants to analyse the hypothesis of an irregular ground truth (GT) data spatial distribution as the cause of such an imbalance in the nation-wide predictions.

When we investigate whether the ground truth data could explain the latitudinal trend in the nationwide prediction maps, we observe:

- There is a clear latitudinal trend in the nationwide predictions: southern NL clearly has higher Q predictions than the rest of the country, which we saw already in the maps
- 2022 predictions are lower compared to 2021
- The ground truth data does **not** reveal a clear latitudinal trend

Therefore, we can conclude that the latitudinal trend is not caused by any latitudinal trend in the ground truth data

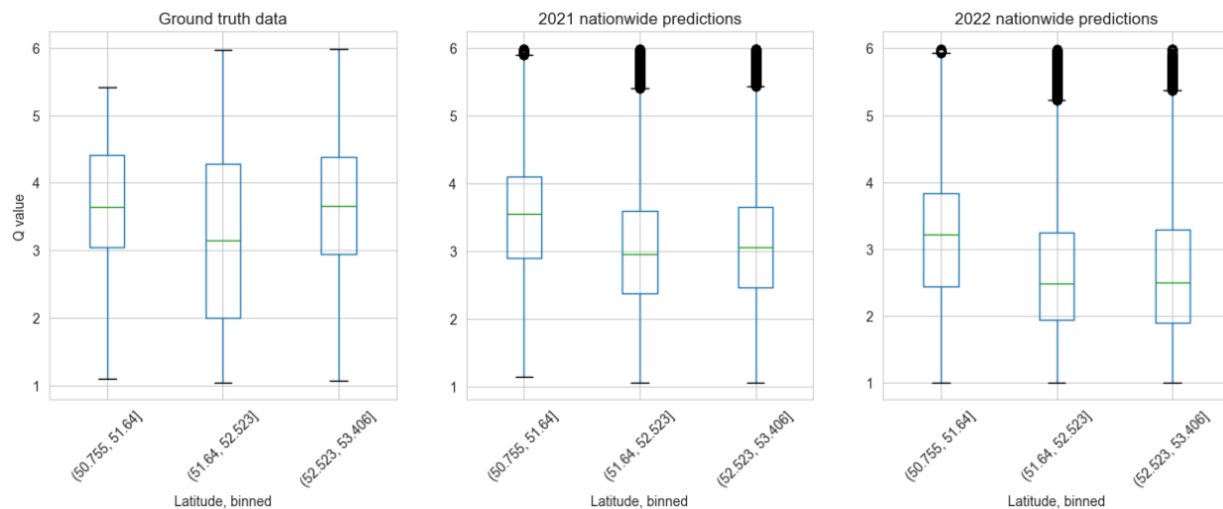


Figure 20. Distribution of Q values for 3 different latitudinal bands, for the ground truth dataset and predictions for 2021 and 2022. Southern NL is up to around 51.6 degrees latitude (left bar); the central latitudes go up to 52.5 degrees (central bar).

8.2. Check if multiyear consistency assumption is valid, based on GT parcels with Q observations for several years

In comparison with 2020-2021, predictions for 2021 show a descending trend, which is not expected from an ecological perspective; relatively stable Q values are expected for the same parcel across years. However, this assumption has not been tested in the training data: how do GT Q values, for the same parcel, evolve across years?

Our key findings are:

- there are only 21 parcels having a GT value in multiple years, so it is hard to draw conclusions based on that sample size. On average the Q differs $\pm 0.3Q$ across years
- for the whole population the Q GT values of 2022 are significantly lower than in 2021. This may have introduced a model bias towards lower Q values in 2022, as we see in the nationwide predictions.
- the mean Q value in the total training dataset is relatively high, and may have caused the models to bias towards higher Q values (this was one of the issues found in the nationwide prediction maps)

In summary, the training dataset is imbalanced leading to model biases that are hard to explain from an ecological perspective.

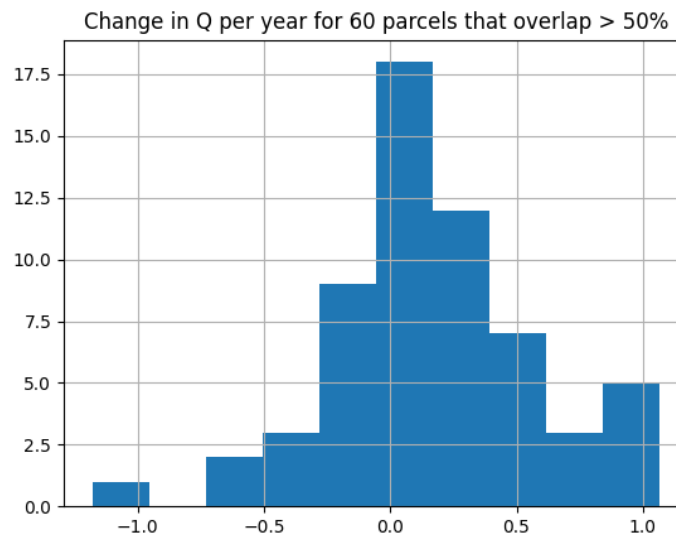


Figure 21. Change in Q per year for 60 parcels that overlap >50%.

8.3. Compare ZonalStats 2022 to provided values by WUR

For year 2022, NEO uses Zonal Stats machine to create vegetation indices, from March till December, and compared them with the values provided by WUR, to identify whether there are significant deviations. This way, we want to rule out the presence of incorrect RS-based values in the training and validation features.

After a Euclidian distance analysis, and visual inspection of some of the time series, we can say that signals created by NEO and signals provided by WUR are negligibly different (see Figure 22). We understand that the main disparity between observations may come from different cloud masking methods. With these results, we are confident that the RS features used for the model development do not suffer from any structural issues, at least for year 2022.

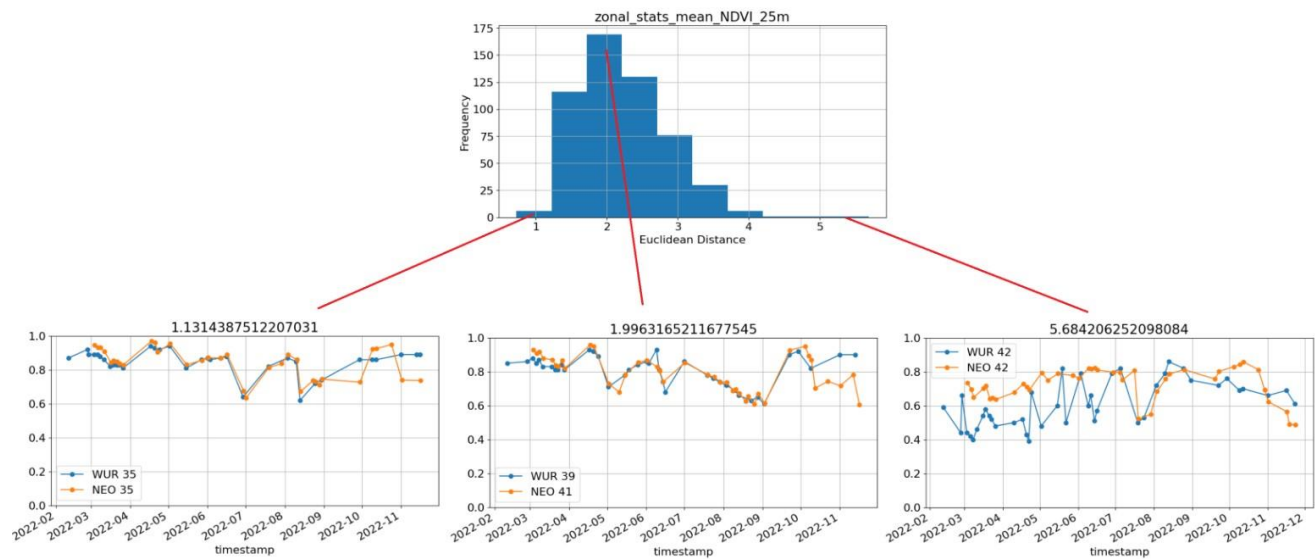


Figure 22. Euclidian distance distribution for NDVI values generated by NEO and provided by WUR, together with 3 time-series signals graphs. Plot in the right graph corresponding to field_id 22722414.

8.4. Visualize feature space for different Q values to understand similarity between parcels with similar Q values

NEO uses TSNE (t-distributed stochastic neighbour embedding) to understand if parcels with similar Q values do have similar features.

t-SNE³ is a tool to visualize high-dimensional data. It converts similarities between data points to joint probabilities and tries to minimize the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high dimensional data (Pedregosa, 2011).

i.e. A t-SNE projection takes multi-dimensional data and tries to map it to a 2D space, where it tries to place similar points closer together. It is primarily a tool to explore the data a bit.

The features used in the current analysis are the same as were used as input for the ridge (i.e. aggregated zonal statistics & markers)

The labels are Q values unless mentioned otherwise, they are on the “turbo” colour scale, so similar colours have similar Q values



In this case, we see that the t-SNE projection with data from all years does clearly show a pattern where similar Q values are indeed closer together in feature space (Figure 23). This would suggest that building a model to predict this is possible (as we have indeed seen). It also shows, however, that the parcels with similar Q values are quite diverse, and despite showing a clear pattern, the clusters are very spread. This suggests it could be complicated to create a very accurate model based on this data.

Furthermore, when we re-label the points to years instead of q-value, we see that the features are generally much more predictive of the year of the measurement, than of the Q-value.

³ <https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>

Total TSNE

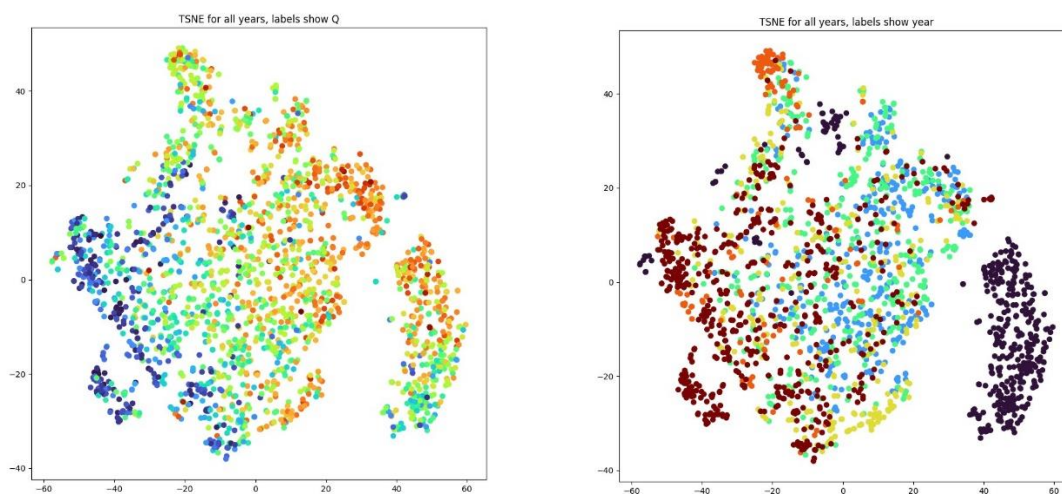
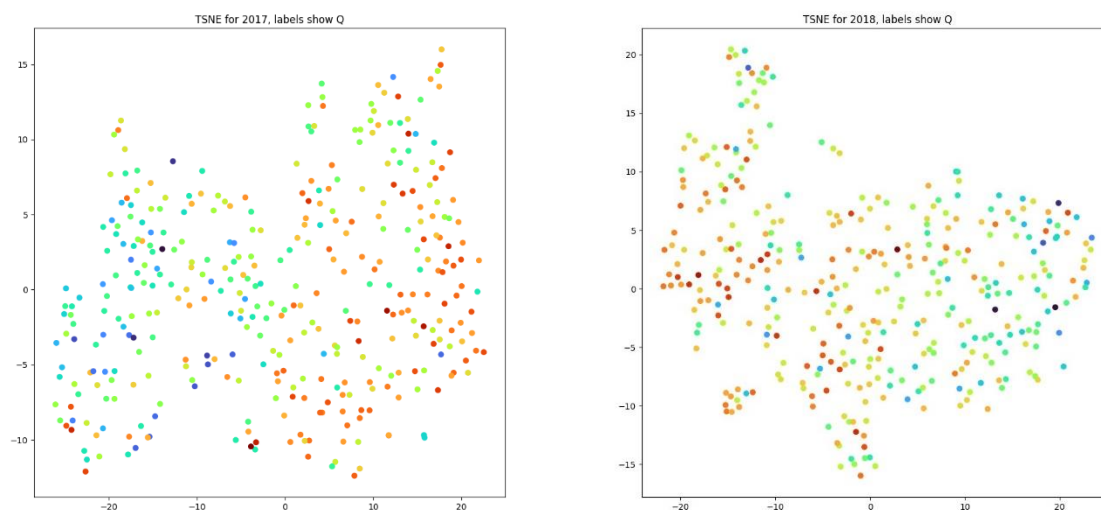


Figure 23. 2D representation of similarity between features, visualized per Q value (left) and year (right).

TSNE within each year



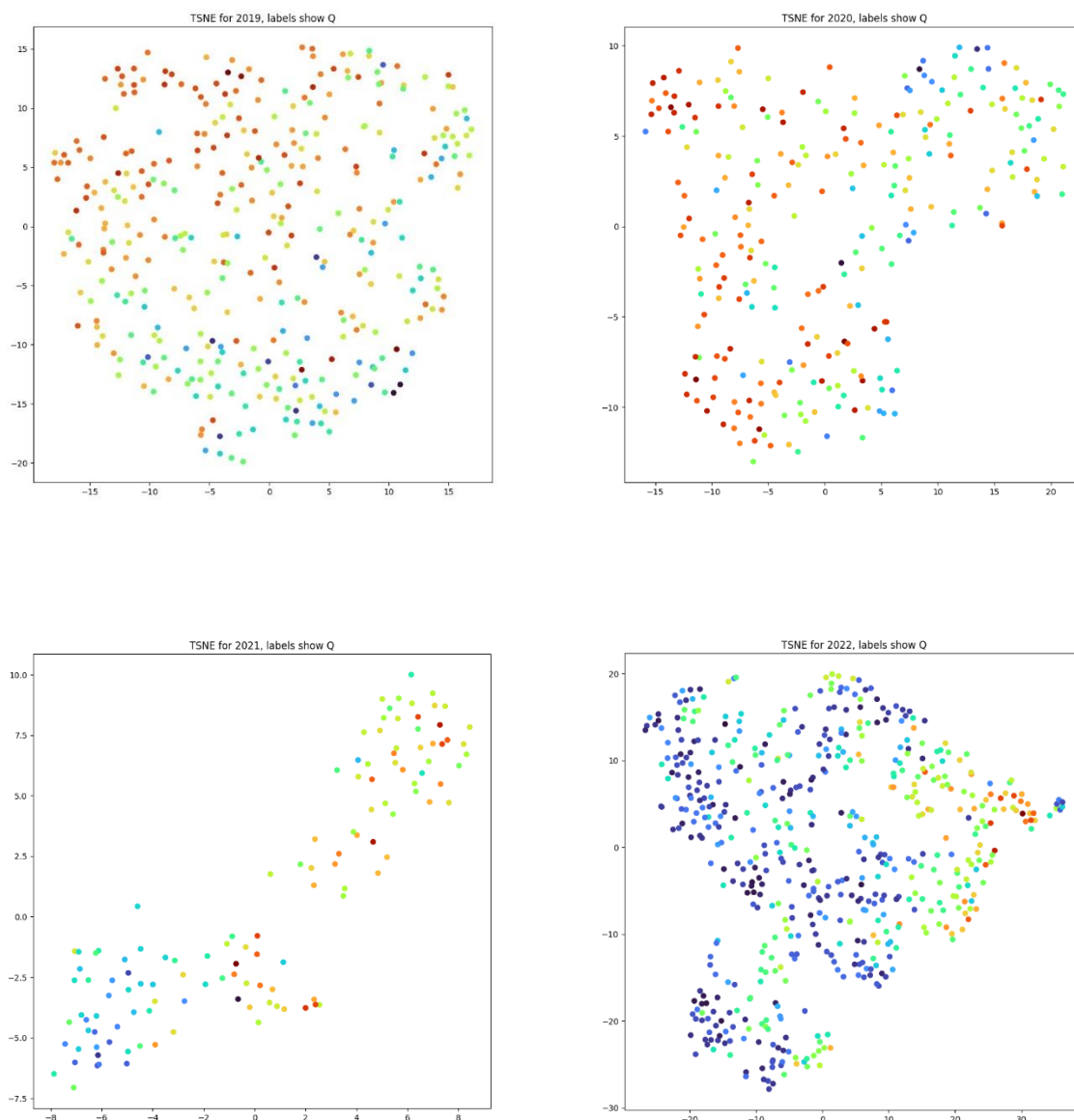


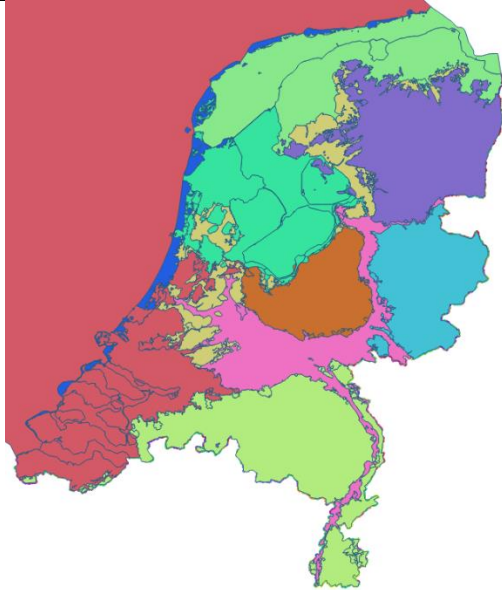
Figure 24. 2D representation of similarity between features, visualized per Q value, for each individual year (2017 till 2022).

8.5. Normalize signals across the whole country, per year.

To reduce the north-south prediction trends, NEO wants to explore averaging signals for an area and year and subtract from the parcel signal. This way, we only use the deviation to the mean as feature value

We have explored the use of 2 features to do this grouping, that relate to soil categories present in the Netherland: bf12_pawn and bf2012_v2. Table 12 displays the grouping.

Table 12. Features used for the grouping and normalization.

Grouping feature	Array of values	Number of groups
bf12_pawn	10., 14., 1., 12., 13., 3., 16., 2., 19., 11., 17., 9., 4., 7., 5., 20., 15., 18., 8., 23., 6., 22., 21	23
bf2012_v2	305., 326., 101., 311., 312., 317., 421., 203., 320., 107., 411., 313., 315., 413., 304., 102., 417., 307., 416., 323., 108., 206., 325., 401., 309., 302., 408., 407., 410., 418., 205., 202., 412., 419., 999., 105., 415., 409., 405., 403., 310., 422., 324., 103., 998., 507., 201., 204., 414., 404., 504., 104., 109., 502., 314., 322., 301., 501., 316., 327., 303., 319., 308	64
Agroclimatic zones (used in AMS)		10

The results we use as baseline show RMSE= 0.63276 and r^2 = 0.68082.

If normalization is done only with the temporal component, without geographical grouping, results show RMSE= 0.71271 and r^2 = 0.62156.

When normalization is done per bf2012_v2 and per year: RMSE= 0.85599 and r^2 = 0.4541. while using bf12_pawn leads to RMSE= 0.75633 and r^2 = 0.54399.

These results show that the normalization and grouping do not improve the results of the model. In fact, it seems that the addition of groups affects the model negatively.

An additional analysis has been done applying the normalization on the agroclimatic zones used in the AMS projects as depicted in Table 12. These areas represent environments with similar soil and weather conditions, so their use to normalize the features use for training and predicting Q value should cancel out their influence in the model. These results are characterized by a RMSE= 0.7693, and $r^2 = 0.52822$. The boxplots shown below, distributing results in 3 latitude categories, do not indicate a better uniformity between north and south results. They display a similar trend as the baseline model, where predicted Q values are overall higher in southern latitudes.

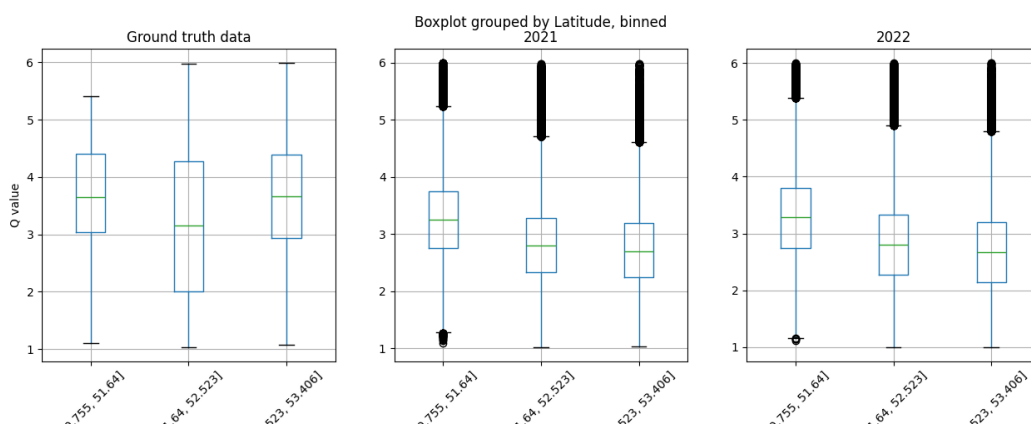


Figure 25. Distribution of Q values for 3 different latitudinal bands, for the ground truth dataset and predictions for 2021 and 2022 after normalization has been applied following the agroclimatic zones used in AMS NL. Southern NL is up to around 51.6 degrees latitude (left bar); the central latitudes go up to 52.5 degrees (central bar).

A summary of the performance of the different models and normalizations is displayed in Table 13.

Table 13. Performance statistics of the models with different normalizations.

	RMSE	r^2
Baseline	0.63276	0.68082
Grouped per year	0.71271	0.62156
bf2012_v2	0.85599	0.4541
bf12_pawn	0.75633	0.54399
AMS zones	0.7693	0.52822

8.6. Data augmentation by moving time series windows

The limited amount of training data is limiting the performance of our model, since this might cause the model to value certain patterns too strongly. We can make the model invariant to these differences by adding augmented data.

There are several data augmentations we should be able to apply safely:

- a. Shift the date range by a few days to the left of the right (before interpolation): When exactly the warmer weather starts, or the blooming season, should not matter too much to the model
For now, training and predictions are done using the time series signals from February to November, given that this time range provides the most accurate predictions. However, to increase the amount of training data, and in this way minimize the effect of weather and early/late growing seasons, different approaches are tested, where different time intervals are used as training data, linked to the same labelled Q values (e.g. Jan-Oct, Mar-Dec...)
- b. Add a small amount of noise to the input signal: This would make the model more robust against small differences

The results of this approach, shown in Table 14.

Table 14. Performance statistics of the models after several data augmentation approaches.

	RMSE	r^2
Baseline	0.63276	0.68082
4 weeks shift	0.64507	0.66829
4 weeks shift * 5	0.65628	0.65666
2 weeks shift	0.66518	0.64728
2 weeks shift * 5	0.63753	0.67599
2 weeks shift * 10 + random gaussian noise	0.67559	0.63615

There is not a clear improvement of model performance after multiplying training data; none of the iterations outperform the baseline results. This can be due to the fact that both data augmentation approaches are actually adding noise so relevant that the model gets confused with it and decreases its performance.

8.7. Inclusion of weather data new features

Weather conditions can affect the phenology of flora as well as their spectral signature in remote sensing data. By adding weather data as features we intend the model to be able to better learn the relationship between those features and the target Q value.

NEO collects the daily gridded datasets ⁴available at KNMI for temperature, precipitation and evaporation and calculates weekly averaged parcel statistics from them. They are then added to the list of features that the CNN model is trained on.

Results show that the addition of these features do not improve the performance of the model. This can be due to the information of these features being already included in the rest of variables already included in the baseline CNN, or to the addition of too much complexity in relation to the relatively scarce number of observations and hidden layers in the model architecture.

Table 15. Performance statistics of the models after adding weather data and a combined approach with weather data, spatial normalization and temporal interpolation.

	RMSE	r ²
Baseline	0.63276	0.68082
With KNMI data	0.66673	0.64564
KNMI data + normalization with agrizone + shift 4 week * 5	0.74209	0.561

⁴

https://dataplatform.knmi.nl/organization/knmi?q=gridded+daily&sort=score+desc%2C+metadata_modified+desc

8.8. Data augmentation by temporal interpolation

With this method we intend to improve the overall accuracy of the model and also improve the consistency of the predictions across years. This approach works under the assumption that the Q value does not change too much over time, that it is consistent across years.

Based on that, we linearly interpolate the observed Q values of one year (or multiple years, if it is the case) for the rest of the years. This way, 10213 training samples are obtained, against the original 1846.

Results show that, by altering our training regime and linking an observed Q value with features from past and future years, the model performance improves slightly in overall accuracy terms, especially when the new observations are given the same weight as the original ones (see Table 16).

Table 16. Performance statistics of the models after augmenting the training set adding interpolated observations.

	RMSE	r^2
Baseline	0.63276	0.68082
Interpolated	0.62687	0.68674
Interpolated weighing the interpolated labels *0.5	0.63057	0.68303

When analysing the North-South predictions consistency, a latitudinal trend is still distinguishable (Figure 26), but less persistent than what is was observed in the baseline CNN model and after the cluster normalization (Figure 25)

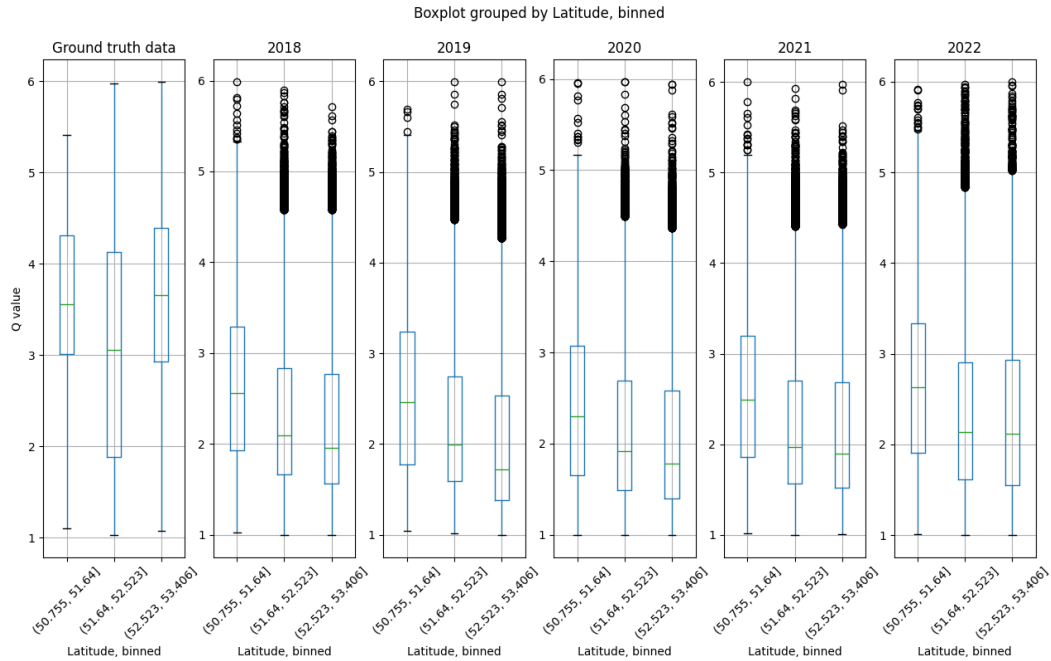


Figure 26. Distribution of Q values for 3 different latitudinal bands, for the ground truth dataset and predictions from 2018 till 2022 after augmenting the training set adding interpolated observations. Southern NL is up to around 51.6 degrees latitude (left bar); the central latitudes go up to 52.5 degrees (central bar).

Lastly, this new approach does show improvements in temporal consistency. Predictions for the same parcel across year display less variability than in the baseline CNN results, and even more consistency than the ground truth data itself. As seen in Figure 27, the difference between predictions per year are more centered around the value 0 (vs -0.5 in the baseline CNN model), with shorter tails in their distribution (maximum of ± 2 , vs ± 3 in the baseline CNN model)

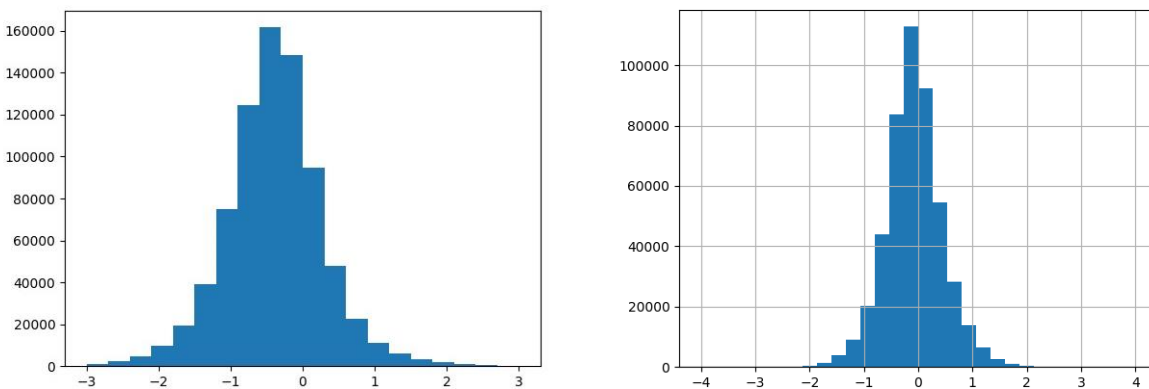


Figure 27. Change in predicted Q value per year for the baseline CNN model (left) and for the new model with interpolated training data (right).

8.9. Alternative DL model: transformers

To improve the accuracy of model predictions, NEO has explored the use of other model architectures different from the CNN.

Transformers are a family of neural network architectures that learn context and thus meaning by tracking relationships in sequential data. We find this user case, where we input a time series of data, a fitting candidate for this type of models to perform appropriately.

Presto (Pretrained Remote Sensing Transformer) is a model pre-trained on remote sensing pixel-timeseries data (Tseng, et al., 2024). Because Presto is designed and trained for remote sensing data (Sentinel 1 and Sentinel 2, among others), even in a larger volume than the one available for this project, we expect a reasonable performance, without having to replicate or perform the whole model training.

For this case, we only retrained some layers after Presto, also enriching the features by appending our markers.

Results are close to the ones for the CNN model, but it does not overperform them.

Table 17. Performance statistics of the transformer models after several optimizations. CNN included for reference.

	RMSE	r^2
CNN with interpolated training set	0.62687	0.68674
Presto: Interpolated to 23 biweekly intervals, including interpolated labels	0.65844	0.65439
Presto: Interpolated to 23 biweekly intervals, including interpolated labels, including months	0.66476	0.64773
Presto: Interpolated to 23 biweekly intervals, including interpolated labels, adding RS markers	0.66953	0.64265

8.10. Comparison of photo model and CNN results

The model based on geotagged photographs, developed by WENR does also predict Q values for the photographed fields. To assess the consistency of both models to predict Q values a comparative analysis has been performed. Q predictions from the photo-model and Q predictions of the CNN model (with temporal interpolation of training data) are subtracted for each plot to visualize how they differ from each other.

The distribution of the difference follows a similar pattern between year 2012 and 2022, with most of differences centering about 0 (Figure 28). In some cases, photo-based predictions are higher than the CNN-based ones, but in the same number of cases we encounter the opposite situation. We can rule out then any hypothesis of one of the models systematically yielding higher predictions than the other.

One of the most extreme cases affects a plot next to a golf course (or part of it), where the CNN prediction $Q=4.31$ and photo prediction $Q=1.25$ (see Figure 29). In this case, the CNN prediction seems too high given that the field is probably intensively managed to reduce to the maximum the presence of elements different than grass.

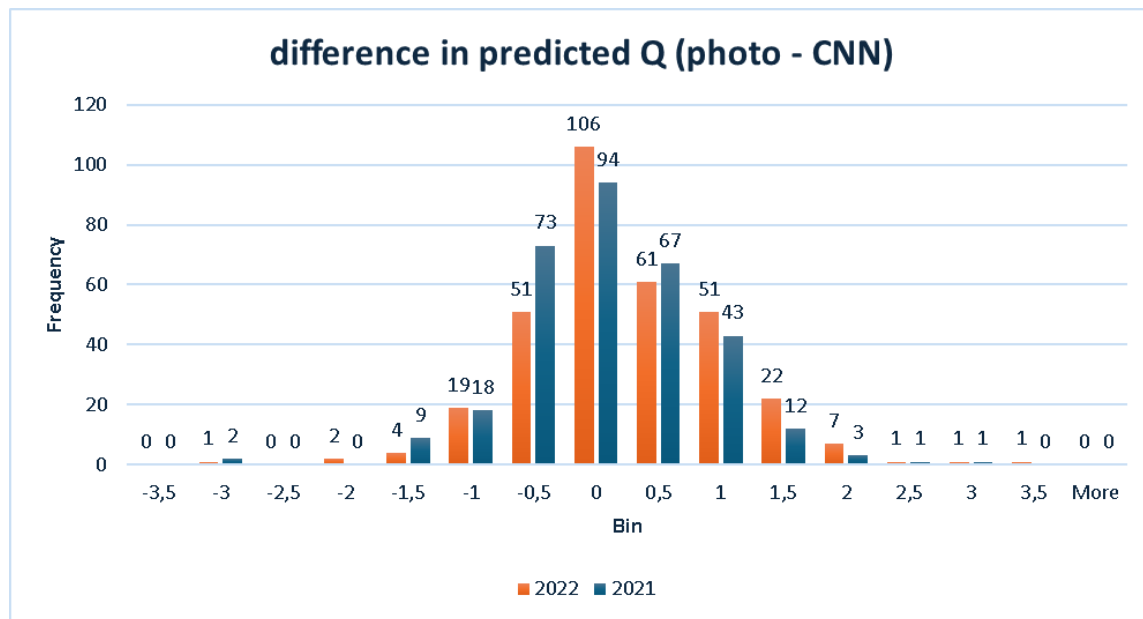


Figure 28. Distribution of the difference between Q predictions of the photo-model and the CNN model, for year 2021 and 2022.



Figure 29. Example of discrepancy between predictions.

The relation between prediction differences and the size of the plot has been also assessed. Results are visible in the figure below.

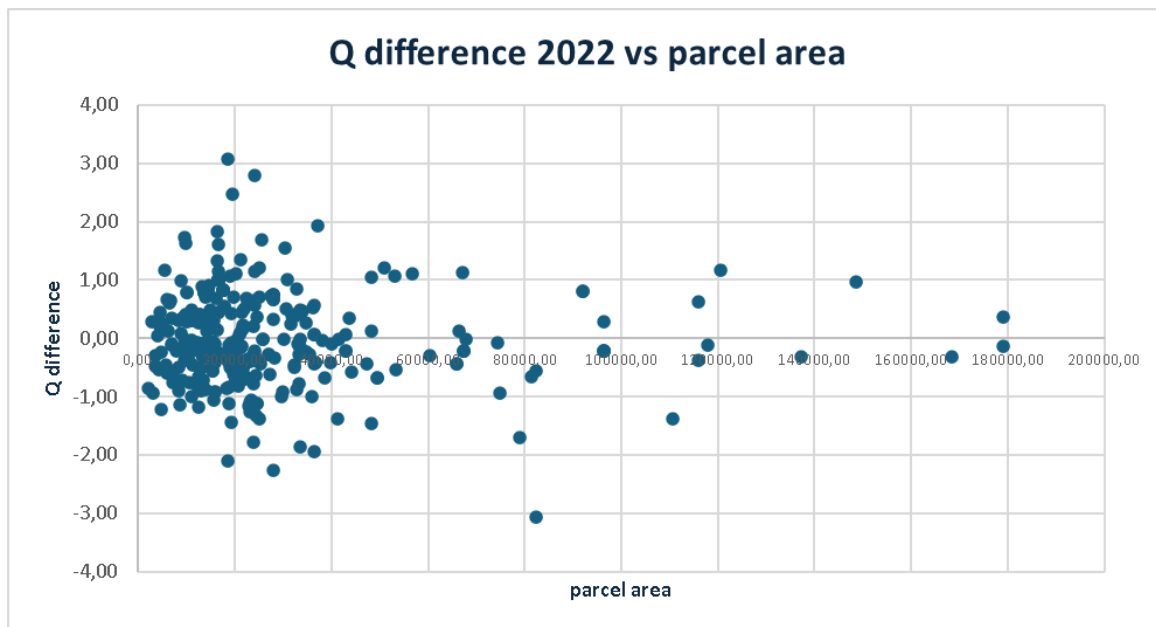


Figure 30. Relation between predictions differences and the size of the plot.

There is no clear trend between those variables. Bigger differences are observed for the smallest parcels, but they are also more frequent. We cannot conclude that the size of the parcel under study affects the level of discrepancy between the models' predictions.

9. References

- Amini, S., Saber, M., Rabiei-Dastjerdi, H., & Homayouni, S. (2022). Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series. *Remote Sensing*, 14(11).
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv*.
- Fauvel, M. (2020). Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series. *Remote Sensing of Environment*(237).
- Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 917-963.
- Gruber, M. (1998). *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators*. CRC Press.
- Hoerl, A., & Kennard, R. (2012). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67.
- Magurran, A. (2021). *Measuring Biological Diversity*. Elsevier Inc.
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, (pp. 51-56).
- Muro, J. (2022). Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. *Remote Sensing of Environment*(282).
- Nguyen, Q. V. (2014). Learning classification models with soft-label information. *Journal of the American Medical Informatics Association*, 21(3), 501-508.
- Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training Recurrent Neural Networks. *arXiv*.
- Pedregosa, F. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825-2830.
- Schreiber-Gregory, D., & Bader, K. (2018). Logistic and Linear Regression Assumptions: Violation Recognition and Control. *SESUG, WUSS, SCSUG, PharmaSUG, MWSUG*.
- Tseng, G., Cartuyvels, R., Zvonkov, I., Purohit, M., Rolnick, D., & Kerner, H. (2024). Lightweight, Pre-trained Transformers for Remote Sensing Timeseries. *ICLR*.
- van 't Veer, R. (2020). *Kruidenrijke Graslanden in de Enge Wormer: Biodiversiteit, graslandherkenning en beheer*. Jisp.
- Visser, T., Meijninger, W., Roerink, G., Kleyheeg, E., & Goedhart, P. (2022). *Ontwikkeling systematiek voor vaststellen en monitoren van habitatkwaliteit agrarisch gebied*. Wageningen: Wageningen Environmental Research.

Zhao, B., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification.
Journal of Systems Engineering and Electronics, 28(1), 162-169.

Bijlage F Rapportage PPS herkenning kruidenrijk grasland



ELLIPSIS DRIVE

Nieuwe Kazernelaan 2 D41, suite B22
6711 JC, Ede, Nederland
KvK nummer: 75011964
BTW nummer: NL860108168B01
info@ellipsis-drive.com

Rapportage PPS herkenning kruidenrijk grasland

Datum: 15-02-2024
Daniel van der Maas
Rosalie van der Maas





1. Inleiding

Het doel van het project PPS herkenning kruidenrijk grasland is om een geautomatiseerd systeem te ontwikkelen waarmee kostenefficiënt kruidenrijk grasland kan worden herkend op basis van de combinatie van satellietwaarnemingen (spectraal en radar) en foto's vanuit het veld. Er wordt beoogt om de uitkomst van het systeem - kruidenrijkdom van alle Nederlandse graslandpercelen - jaarlijks te ontsluiten naar de gebruikers met een aansluiting op de ICT-systemen die voor deze partijen relevant zijn.

Het eindproduct van dit project wordt een centrale database met - op gebruikers afgestemde - interfaces en een doorontwikkelde versie van de foto-app waarmee niet alleen gestructureerd gegevens uit het veld kunnen worden verzameld, maar welke ook een typering van het gefotografeerde perceel en een beheersadvies aan de gebruiker kan terug geven.

Gezien de omvang van het project, wordt dit in meerdere fases uitgevoerd. Gedurende Fase 1 is de haalbaarheid van de behoefte onderzocht en is geëvalueerd of het systeem technisch haalbaar is en op een kostenefficiënte manier operationeel kan functioneren. Als onderdeel van deze fase zijn twee digitale hulpmiddelen ontwikkeld:

- 1) Een Foto-app om foto's en andere gegevens uit het veld gestructureerd te verzamelen.
- 2) Een centrale database met passende interfaces om vergaarde gegevens centraal in op te slaan en via relevante interfaces/endpoints beschikbaar te stellen aan stakeholders.

Deze twee digitale hulpmiddelen zijn ook tussentijdse producten op de weg naar het uiteindelijke eindproduct zoals hierboven beschreven.



2. Omschrijving van de uitvoering: Fase 1

Voor het ontwikkelen van de Foto-app heeft Ellipsis capaciteit beschikbaar gesteld om in zes sprints van een week de nieuwe (mobile first) webapp op te leveren in lijn met de functionaliteiten zoals in sectie 3 beschreven. Vier van deze sprints zijn uitgevoerd in maart-mei 2022 en twee sprints in februari-maart 2023. De sprints in 2023 zijn toegevoegd om de app op basis van ervaringen in 2022 verder te verbeteren. De doelstelling was om via deze app van ~2500 graslandpercelen verspreid over Nederland data in te winnen. Er is uiteindelijk van ruim 5000 percelen data ingewonnen.

Na iedere sprint is overlegd met de consortiumpartners over updates, veranderingen en testervaring. Na de eerste 3 sprints was er sprake van een Minimal Viable Product en is het door de stakeholders in het veld getest. De gevonden issues tijdens deze tests zijn in de laatste sprint van 2022 verwerkt. Na de ontwikkeling van de app zijn op basis van feedback van eindgebruikers nog enkele nieuwe features en verduidelijkingen doorgevoerd, ook in 2023.

De app is ontwikkeld als een single page REACT app en gebruikt de open source mapping library leaflet. De app is gebouwd op de Ellipsis Drive standaard REST-API, er wordt gebruikgemaakt van het posten van locaties en het daaraan verbinden van antwoorden en foto's van gebruikers.

2.1 UX afwegingen

De app is zo eenvoudig mogelijk gehouden. De doelstelling was om het zo simpel mogelijk te maken om de juiste foto's te nemen, vragen over het desbetreffende grasland perceel te beantwoorden en deze gegevens vervolgens op te sturen naar de database. Er is getracht om andere eventuele afleidingen zo veel mogelijk achterwege te laten zodat de gebruiker deze flow frictieloos kan doorlopen.

Er is hier een uitzondering op gemaakt voor de perceel selectie. Het zou optimaal zijn als een gebruiker direct foto's kan maken en vragen kan beantwoorden zonder een precieze locatie te selecteren (idealiter zou dat automatisch gebeuren). In de praktijk is de locatiebepaling van mobiele telefoons echter niet altijd voldoende betrouwbaar. Daarom vraagt de app aan de gebruiker in het veld om het juiste perceel te selecteren uit zijn/haar directe omgeving. Hoewel dit een extra stap voor de gebruiker betekent, bleek dit in de praktijk de moeite waard om te kunnen garanderen dat verzamelde data aan het juiste perceel wordt toegekend. Voor het slagen van het project is dat immers essentieel.

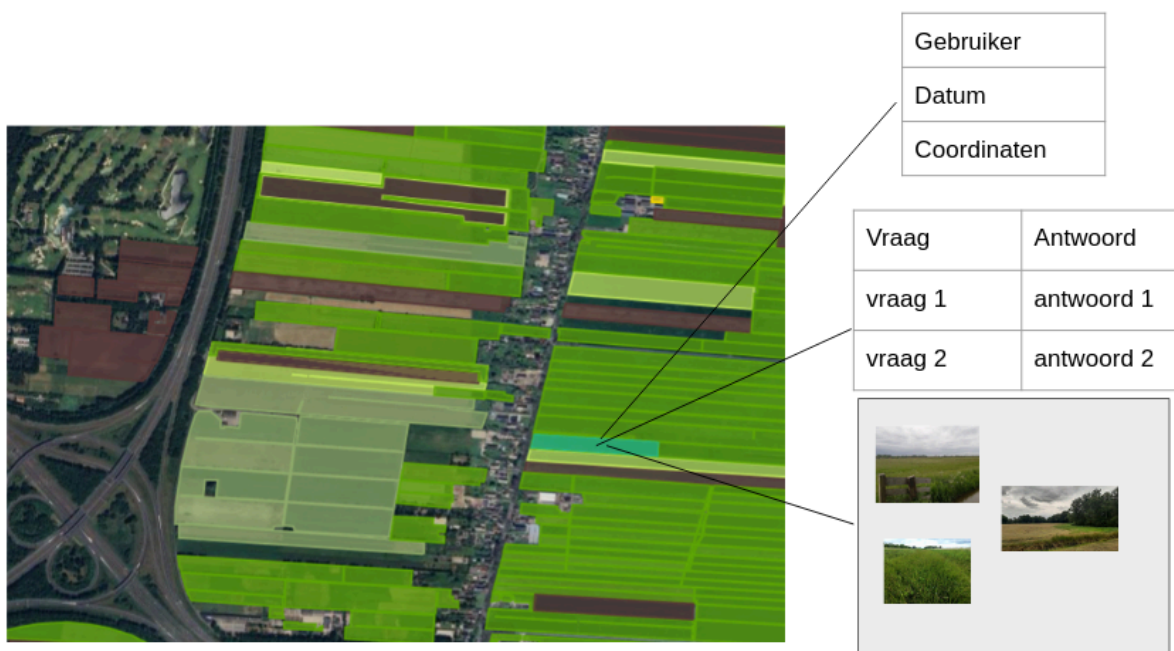
Er is ook voor gekozen om registratie en login voor gebruikers verplicht te stellen. Dit creëert extra frictie, maar hier is toch voor gekozen om data herleidbaar te houden en bij eventueel misbruik in te kunnen grijpen.

2.2 Internetconnectie uitdagingen

Het internet in het veld is niet heel betrouwbaar gebleken, hierdoor is het noodzakelijk om bij gebruik van de app data transfer via het internet tot een minimum te beperken. In sommige gevallen bleek er zelfs geen internet te zijn. Dit maakt de webapp uiteraard onbruikbaar. Deze situatie kwam in de praktijk weinig voor (op het moment dat het voorkomt was het uiteraard wel zeer hinderend), maar zal in een opgeschaald productie scenario gemitigeerd moeten worden met een mobiele app die offline bruikbaar is (zie ook aanbevelingen in sectie 4).

2.3 Data structuur

De structuur waarin de data wordt ingewonnen en opgeslagen staat hieronder afgebeeld. Iedere submittie via de app voegt een locatie toe met daaraan de gebruiker, de datum, de coördinaten, de foto's, en de antwoorden op de vragen. Iedere submittie geeft een nieuwe locatie. Submissies worden dus niet direct aan een perceel gekoppeld. Dit wordt pas tijdens analyse gedaan door middel van intersectie. Dit geeft in de analyse stap meer flexibiliteit en houdt rekening met het feit dat perceelsgrenzen kunnen veranderen over de tijd.



Figuur 1

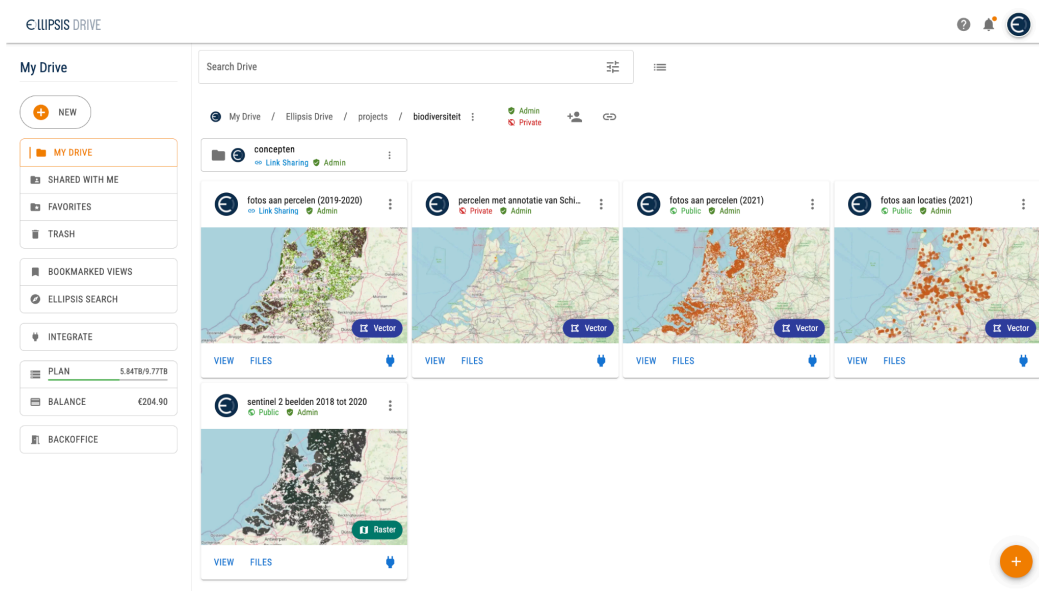
Naast minimaal twee overzichtsfoto's en twee detailfoto's per grasland perceel, worden er aan de gebruiker ook een set aan vragen gesteld over kruidenrijkheid en de productiviteit van grasland. In Annex A zijn de vragen en mogelijke antwoorden onder elkaar gezet.



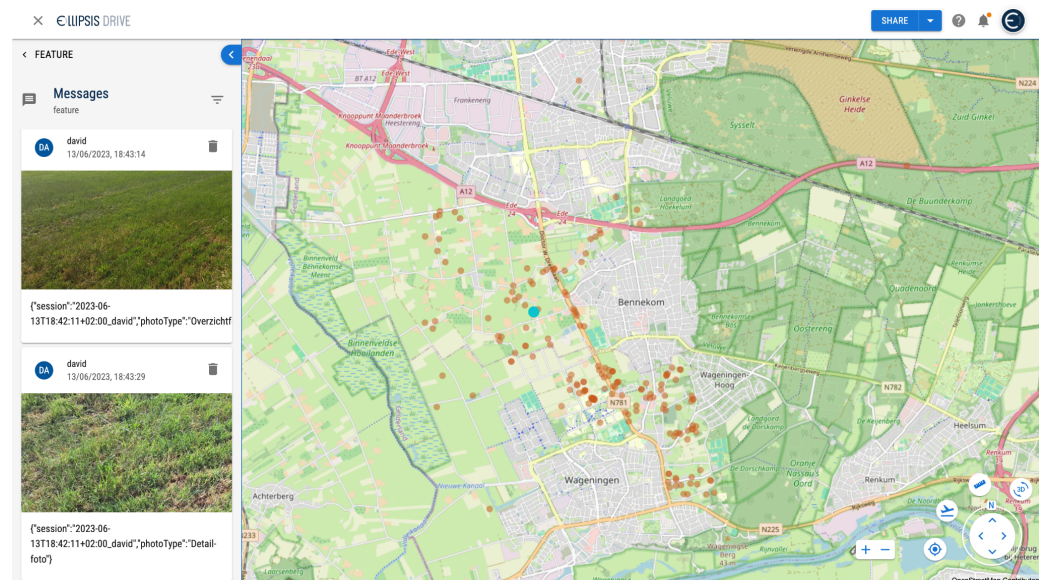
2.4 WUR admin op Ellipsis Drive Data Base

Het tweede digitale hulpmiddel onder de eerste fase was het beschikbaar maken van een centrale database met passende interfaces om vergaarde gegevens centraal in op te slaan en via relevante interfaces/endpoints beschikbaar te stellen aan stakeholders. Hiertoe is gebruikgemaakt van de Ellipsis Drive standaard API. Binnen Ellipsis Drive is een layer aangemaakt waar naar de app direct de gegevens wegschrijft. Deze layer wordt beheerd door een WUR account, maar anderen hebben permissie om gegevens (foto's, locaties en vragen) naar deze layer weg te schrijven.

Individuele gebruikers kunnen via hun eigen account toegang krijgen tot deze data. Deze toegangsrechten worden vanuit het admin account ingesteld en beheerd. Het Ellipsis team biedt support aan de gebruikers van deze database.



Figuur 2



Figuur 3



3. Deliverables: Fase 1

De volgende 2 concrete deliverables zijn in fase 1 gerealiseerd:

3.1 Centrale database en interfaces

- Ellipsis heeft een centrale database opgeleverd waarin de data van ruim 5000 grasland percelen is opgenomen. De database kan in Fase 2 (als hier groen licht voor is) worden opgeschaald voor het opslaan en beschikbaar stellen alle graslandpercelen in Nederland.
- De database voldoet aan de volgende eisen:
 - o De database kan per perceel de volgende gegevens bevatten:
 - Remote Sensing data (RS-data: optisch-spectraal en/of radar, zoals perceelgemiddelde tijdseries van NDVI, droge stof, spectraal, backscatter, coherence, etc.)
 - Remote Sensing markers (RS-markers, zoals datum eerste maaisnede, aantal maaisnedes per jaar etc.)
 - Geo-informatie zoals bodemeigenschappen, grondwatertrappen, etc.
 - Andere perceel relevante informatie zoals foto's (via de Foto-App, Ellipsis Drive UI of API), verzamelde typologie informatie verkregen uit veldbezoeken of andere bronnen, management gerelateerde informatie (zoals bv. mestgiftten, etc.), leeftijd gras, etc.
 - Typologie van het graslandperceel en bijbehorende kwaliteitschecks
 - o Er kunnen op uniforme wijze raster en geometrie bestanden worden toegevoegd aan de database.
 - o Er kunnen op uniforme wijze foto's en andere attachments worden toegevoegd aan geometrieën.
 - o Er kunnen op uniforme wijze tijdseries worden toegevoegd aan geometrieën.
 - o Toegangsrechten/permissions zijn per gebruiker en systeem flexibel te managen en aan te passen.
 - o De database heeft de volgende endpoints/interfaces:
 - REST-API
 - OGC standards (WMTS, WMS, WCS, WFS, WFS-t, XYZ, Vector Tiles, COGs)
 - Python package & R package
 - Web app



- o Database content is via alle endpoints/interfaces gelijktijdig bewerkbaar en opvraagbaar (real time operational picture voor alle gerechtigde gebruikers).
- o *Change history* wordt bewaard en is eenvoudig naar individuele gebruikers te herleiden en zo nodig terug te draaien.
- o Alle opgeslagen informatie in de database is via REST-API (als één van de endpoints/interfaces) door derde partijen op te vragen (per perceel en per boundingbox).

3.2 Foto-app

- Iedere gebruiker kan onder zijn/haar eigen gebruikersnaam foto's van, en opmerkingen over, percelen kunnen insturen.
- De App geeft in het geval van een landbouwperceel automatisch een perceel (BRP) suggestie op basis van locatie, welke door de gebruiker bevestigd dient te worden. In geval van een natuurterrein wordt de informatie onder de coördinaten van het standpunt van de fotograaf vastgelegd, tenzij er een perceelsbegrenzing beschikbaar is gesteld, dan geldt hetzelfde als voor landbouwpercelen
- De app laat zien welke foto's al zijn ingestuurd en voor welke percelen dit is gebeurd.
- Foto's die niet gematched zijn aan een perceel kunnen alsnog worden geupload.
- De app geeft instructies over hoe gebruikers het beste bruikbare foto's kunnen maken.
- Extra gegevens die van het perceel moeten worden ingevoerd, worden vanuit de app begeleid, via een voorkeuzemenu voor de typologie en het beheer etc.
- Foto's ingestuurd via de app komen automatisch in de database terecht waar ze als attachment per perceel worden opgeslagen onder vermelding van de datum.
- De ingestuurde foto's zijn via REST-API (als één van de endpoints/interfaces) door derde partijen op te vragen (per perceel en per boundingbox) samen met de opgeslagen attributen.



4. Aanbevelingen

We hebben de volgende aanbevelingen voor het verbeteren van de gebruikerservaring en de doelmatigheid van de app. We doen deze aanbevelingen op basis van onze ervaringen en de ondervonden uitdagingen gedurende fase 1.

4.1 mobile native app

Het belangrijkste is het overstappen van een webapp naar een mobile native app. Dit brengt extra kosten met zich mee, maar is de enige manier om de app offline bruikbaar te maken. Onze inschatting is dat offline gebruik in de pilotfase niet noodzakelijk is, maar in de productiefase wel.

4.2 Less is more

We merken dat het verleidelijk is om extra elementen toe te voegen in de app. Dit om meer informatie op te halen of tegemoet te komen aan stakeholders. We merken echter ook dat de eindgebruiker van de app in het veld gebaat is bij een zo licht- en eenvoudig mogelijke applicatie. We raden daarom met klem aan om de huidige flow niet verder uit te bouwen, maar in plaats daarvan te richten op het lichter en makkelijker maken van wat al kan.

4.3 Fit for purpose

Om dezelfde reden raden we aan om te blijven werken met een applicatie die speciaal voor dit doeleinde wordt ontwikkeld. Er zijn algemene data inwin tools, maar op het moment dat je externen (zoals boeren etc.) gaat vragen data te verzamelen via een app, dan moet die app qua branding, flow en gebruik echt volledig toegespitst zijn op het specifieke doeleinde.

4.4 Data structuur en locatie

Momenteel zijn de antwoorden op de vragen gebonden aan de locaties als aparte berichten. We raden echter aan om dit om te zetten in een vast stramien. Dat wil zeggen dat iedere locatie de vragen als properties heeft, die vervolgens een waarde hebben. Dit zorgt ervoor dat antwoorden makkelijker geüpdatet kunnen worden en het filteren op basis van antwoorden makkelijker wordt.

In de initiële fase van maart tot mei 2022 werden waarnemingen aan percelen gebonden. Dit bleek echter niet handig. Over de jaren heen kunnen percelen immers veranderen. Om deze reden is ervoor gekozen om deze informatie aan locaties te binden (x,y coördinaten). We raden aan om dit in de toekomst ook te blijven doen.



4.5 Kies voor één data flow en één database

Door het gebruik van twee verschillende apps om data mee te verzamelen, zijn binnen het project ook twee aparte datasets met foto's, antwoorden en locaties ontstaan. Hierdoor wordt het lastiger voor stakeholders om een overzicht te krijgen en analyses uit te voeren. We raden aan om in de toekomst gebruik te maken van één app. Als hier toch van wordt afgeweken, moet er in ieder geval voor worden gezorgd dat de data stromen weer samen worden gebracht naar één uniforme structuur en locatie.

4.5 gebruikers uitnodigen

Voor een volwaardige productie situatie zal er moeten worden nagedacht over hoe gebruikers kunnen worden uitgenodigd om deel te nemen aan het initiatief, en of er gebruikers met verschillende rollen moeten zijn. Ook moet er beheer komen op het verwijderen van accounts die niet langer toegang moeten hebben. Kortom, een proces moet worden vastgesteld dat beschrijft hoe gebruikers het systeem in komen en hoe deze weer kunnen worden verwijderd.



Annex A

Vragen en mogelijke antwoorden over kruidenrijkheid classificatie en productiviteit van grasland

SATREvolution

Stap 3a: Classificatie

Bent u in staat om voor dit perceel het volgende onderscheid te maken? Zo ja vink dan aan, kies anders nee

0-A. Engels Raaigrasland
glanzend groene productieve grassoorten; Type 0 in Schippers et al., 2012

SELECT 0-A

0-B. Productieve voederkruiden
Ingezaaide productieve voederkruiden (productieve grassen met o.a. smalle weegbree, cichorei en div. klavers)

SELECT 0-B

0-C. Productief klavergrasland
Productief klavergrasland (productieve grassen met minimaal 10% klaver)

SELECT 0-C

Geen van de typen 0-A, 0-B of 0-C (kies dit alleen als u het onderscheid tussen typen 1-5 niet kent)
Het grasland voldoet aan geen van de genoemde typen

SELECT GEEN TYPE

Nee
Ik ben niet in staat deze typering goed te maken.

SELECT NEE

> SLA OVER

SATREvolution

Stap 3b: Classificatie

Bent u bekend met de indeling van Wim Schippers' veldgids uit 2012? Zo ja vink dan het juiste type aan, kies anders nee

1. Grassen-mix (Type 1 in Schippers, 2012)

SELECT 1

2. Dominant stadium (Type 2 in Schippers, 2012)

SELECT 2

3. Gras-kruiden-mix (Type 3 in Schippers, 2012)

SELECT 3

4. Bloemrijk grasland (Type 4 in Schippers, 2012)

SELECT 4

5. Schraalland (Type 5 in Schippers, 2012)

SELECT 5

6. Nee ik ben hier niet mee bekend

SELECT 5

> SLA OVER

Stap 4: Productiviteit van grasland



Bent u bekend met het beheer of gebruik van dit grasland?

> NEE

> JA

Stap 4: Productiviteit van grasland



Fill:

Wordt dit perceel begraasd?

Is dit grasland recentelijk (afgelopen 5 jaar) ingezaaid of doorgezaaid?

☒ Weet ik niet

☐ Nee

☐ Ja, met productieve grassen en/of klavers

☐ Ja, met productieve overige kruiden

☐ Ja, met biodiverse kruiden

Stap 4: Productiviteit van grasland



Fill:

Wordt dit perceel begraasd?

Is dit grasland recentelijk (afgelopen 5 jaar) ingezaaid of doorgezaaid?

Hoe productief schat u dit grasland in?

☒ Weet ik niet

☐ hoog

☐ middel

☐ laag

Stap 4: Productiviteit van grasland



Fill:

Wordt dit perceel begraasd?

☒ Weet ik niet

☐ Er loopt nu graasvee

☐ Er zijn tekenen van begrazing (mestflaten, urineplekken)

☐ Het is mij bekend dat het perceel soms begraasd wordt

☐ Nee

✓ KLAAR MET DEZE STAP

> SLA OVER

Correspondentie adres voor dit rapport:
Postbus 16
6700 AA Wageningen
T 0317 48 07 00
www.wur.nl/plant-research

Rapport WPR-1309



De missie van Wageningen University & Research is 'To explore the potential of nature to improve the quality of life'. Binnen Wageningen University & Research bundelen Wageningen University en gespecialiseerde onderzoeksinstituten van Stichting Wageningen Research hun krachten om bij te dragen aan de oplossing van belangrijke vragen in het domein van gezonde voeding en leefomgeving. Met ongeveer 30 vestigingen, 7.600 medewerkers (6.700 fte) en 13.100 studenten en ruim 150.000 Leven Lang Leren-deelnemers behoort Wageningen University & Research wereldwijd tot de aansprekende kennisinstellingen binnen haar domein. De integrale benadering van de vraagstukken en de samenwerking tussen verschillende disciplines vormen het hart van de unieke Wageningen aanpak.
