



Verspreidingskaarten van planten ten behoeve van de kwaliteitsbepaling SNL

Henk Sierdsema, Christian Kampichler &
Laurens Sparrius

In opdracht van





Colofon

© Sovon 2014

ISSN 2212-5027

Dit rapport is opgesteld in opdracht van Alterra

Wijze van citeren: Sierdsema H., Kampichler C. & Sparrius L. 2014. Verspreidingskaarten van planten ten behoeve van de kwaliteitsbepaling SNL. Sovon-rapport 2014/20. Sovon Vogelonderzoek Nederland, Nijmegen.

Niets uit dit rapport mag worden vermenigvuldigd en/of openbaar worden gemaakt d.m.v. druk, fotokopie, microfilm of op welke andere wijze dan ook, zonder voorafgaande schriftelijke toestemming van Sovon en/of de opdrachtgever.



Inhoud

1 Inleiding	4
2 Methoden	5
2.1 Plantengegevens	5
2.2 Omgevingsvariabelen	5
2.3 Regressie-analyses	13
2.4. Sensitiviteit en specificiteit van modellen	15
2.5. Invloed van specifieke variabelen op voorspellingen	16
3. Resultaten	17
3.1 Regressie-modellen	17
3.2 Combinatiekaarten	18
Literatuur	19
Bijlage 1 Inhoud excel-bestanden met validatie-statistieken	21
Bijlage 2 Inhoud excel met variabele-bijdragen	21



1 Inleiding

Ten behoeve van natuurbehoud in Nederland zijn een aantal natuurdoelen geformuleerd die zijn uitgewerkt in meerdere natuurdoeltypen en de SNL-beheertypen ('natuurtypen'). Behorende bij ieder natuurtype worden doelsoorten beschreven. Aan de hand van de aan- of afwezigheid van doelsoorten wordt bepaald of het natuurdoel is bereikt. Doelsoorten zijn soorten die in het natuurbeleid met prioriteit aandacht krijgen vanwege hun beperkte aanwezigheid en/of hun negatieve trend op internationaal en/of nationaal niveau.

In dit rapport analyseren wij de compleetheid van de plantengemeenschappen van doel-soorten (en andere soorten) van natuurtypen. Hiertoe hebben we kanskaarten gemaakt van 734 vogelsoorten. De kanskaarten geven de kans weer op het voorkomen van de soort in een gridcel van 250×250 m. Deze kansen zijn berekend op basis van waarnemingen van aan- of afwezigheid in monitoringgebieden in combinatie met gegevens van het habitat. Op basis van het voorkomen van de verschillende habitatkenmerken is vervolgens een Nederland-dekkende kanskaart geconstrueerd. Per natuurtype is een samengestelde kanskaart gemaakt.

Het doel van de analyse is om een eerste beeld te geven van de actuele kwaliteit van natuur volgens de SNL beheertype systematiek. Hiervoor wordt gebruik gemaakt van landelijke verspreidingsdata. Hierdoor kunnen uitspraken worden gedaan op niveau van gebied, provincie, land en per beheertype. Daarnaast wordt, en dat is nu voor de Balans 2014 misschien wel het belangrijkste, de trend weergegeven tussen de natuurkwaliteit (op basis van de verspreiding) van 2005 ten opzichte van de huidige situatie ('2012'). Ten slotte wordt het ook mogelijk om het verschil aan te geven in de mate van doelrealisatie volgens de natuurdoeltype systematiek en de SNL systematiek.



2 Methoden

2.1 Plantengegevens

Voor de berekeningen is gebruik gemaakt van het voorkomen per kilometerhok in de periode 2003-2007 ('2005') en 2008-2013 ('2012'). Deze gegevens zijn afkomstig uit de NDFD en aanvullingen van FLORON.

Nulwaarnemingen

Om de kans op voorkomen van een soort te berekenen wordt gebruik gemaakt van zgn. binomiale modellen voor aan-/afwezigheidsgegevens. Een probleem bij het gebruik van de gegevens in de NDFD en FLORON is dat deze maar in zeer beperkte mate informatie bevat over de afwezigheid van soorten: zogenaamde nulwaarnemingen. Dat betekent, dat deze nulwaarnemingen op een andere manier samengesteld moeten worden uit de beschikbare gegevens.

Om de kans de op voorkomen te kunnen modelleren moeten dus op een bepaalde manier nulwaarnemingen worden toegevoegd aan de dataset. Dit kan op een volledig willekeurige manier, waarbij vervolgens wordt gekeken in hoeverre de locaties van waarnemingen afwijken van deze 'achtergrond-nullen'. Deze methode wordt gebruikt in 'presence-only'-modellen zoals ENFA en Maxent. Hiervan is Maxent superieur gebleken aan ENFA om de verspreiding zo goed mogelijk weer te kunnen geven (Dudik et al. 2007; Elith and Graham 2009; Phillips et al. 2006; Phillips et al. 2009). Een nadeel van deze 'presence-only'-modellen is onder meer dat zij geen echte kans op voorkomen kunnen modelleren, maar alleen een relatieve maat, de zgn Habitat Suitability Index (HSI). Een groter nadeel, is dat deze modellen ook de wel bekende nulwaarnemingen weggooien, zodat regelmatig een positief beeld van de verspreiding ontstaat.

Een alternatieve methode is het gebruik van het programma Frescalo (Hill 2012). Met dit programma zijn nulwaarnemingen gegenereerd per atlasblok van 5x5 km. Hiervoor zijn alle beschikbare waarnemingen vanaf 1975 gebruikt. Een tussenresultaat van Frescalo, dat eigenlijk bedoeld is om trendberekeningen te maken, is een kaart met de trefkans per soort per gridcel. Atlasblokken met trefkans tot 20% zijn gebruikt als nul in de vervolgberekeningen. Voor Frescalo is de benadering gebruikt zoals beschreven door Bijlsma (2013) met de omgevingsvariabelen in de categorieën Fysische Geografische Regio, hoofdboomsoort, grondwaterstand, oppervlaktewater en sub-ecotopen zoals beschreven in de volgende paragraaf.

Naast de HSI-benadering voor het genereren van nulwaarnemingen is ook een alternatieve methode gebruikt, de zogenaamde 'slimme nullen'. In de NDFD is per soortgroep aangegeven welke kilometerhokken redelijk tot goed zijn onderzocht voor die soortgroep. Die informatie is gebruikt om nulwaarnemingen te genereren voor die kilometerhokken die goed zijn onderzocht, maar waar de soort niet is gemeld (zie ook Sierdsema en Buys in (Huizinga et al. 2010)).

Naast de gegenereerde nulwaarnemingen zijn vanzelfsprekend ook de NDFD aanwezige echte nulwaarnemingen gebruikt voor het maken van de kaarten.

2.2 Omgevingsvariabelen

Uit de beschikbare omgevingsinformatie is een selectie gemaakt van relevante omgevingsvariabelen (Tabel 2.1). Niet alle variabelen zijn gebruikt per soort. Eerst zijn de soorten geclassificeerd in habitatgroepen (bos, agrarisch en moeras). Per habitatgroep is een beperkte selectie van ca. 80 variabelen gebruikt. De gebruikte variabelen per soort zijn te vinden in Excel-bestanden met variabele-contributies.



Tabel 2.1. Overzicht van belangrijkste gebruikte variabelen in de regressie-analyses.

Variabele	Variabele groep	Toelichting
Bodemhfd_Kleiopveen	Bodem (hoofdingeling)	Klei_op_veen
boshfd_berk	Hoofdboomsoort	berk
boshfd_beuk	Hoofdboomsoort	beuk
boshfd_douglas	Hoofdboomsoort	douglas
boshfd_es	Hoofdboomsoort	es
boshfd_fijnspar	Hoofdboomsoort	fijnspar
boshfd_gewoneesdoorn	Hoofdboomsoort	gewoone esdoorn en Spaanse aak
boshfd_groveden	Hoofdboomsoort	grove den
boshfd_inlandseeik	Hoofdboomsoort	inlandse eik
boshfd_Japlariks	Hoofdboomsoort	Japanse lariks
boshfd_populier	Hoofdboomsoort	populier
boshfd_wilg	Hoofdboomsoort	wilg
boshfd_zwarteels	Hoofdboomsoort	zwarte els
Bosmeng_..	Idem als memgboomsoort	
boskiemper_1_Vr1900	Kiemjaarklasse	Voor 1900
boskiemper_2_1900	Kiemjaarklasse	1900-1930
boskiemper_3_1930	Kiemjaarklasse	3_1930-1960
boskiemper_4_Na1960	Kiemjaarklasse	4_Na 1960
Eco_akker	Sub-ecotopen	akker
Eco_bebouwing_agra	Sub-ecotopen	bebouwing-agrarisch
Eco_bebouwing_buiten	Sub-ecotopen	bebouwing-buiten
Eco_bebouwing_stad	Sub-ecotopen	bebouwing-stad
Eco_bos_gemengd	Sub-ecotopen	bos-gemengd
Eco_bos_griend	Sub-ecotopen	bos-griend
Eco_bos_loof	Sub-ecotopen	bos-loof
Eco_bos_naald	Sub-ecotopen	bos-naald
Eco_bos_nat	Sub-ecotopen	bos-nat
Eco_bos_populier	Sub-ecotopen	bos-populier
Eco_duinheide	Sub-ecotopen	duinheide
Eco_grasland	Sub-ecotopen	grasland
Eco_heide_overig	Sub-ecotopen	heide-overig
Eco_hoogveen	Sub-ecotopen	hoogveen
Eco_kwelder	Sub-ecotopen	kwelder
Eco_moeras_overig	Sub-ecotopen	moeras-overig
Eco_moeras_riet	Sub-ecotopen	moeras-riet
Eco_moeras_ruigte	Sub-ecotopen	moeras-ruigte
Eco_onbekend	Sub-ecotopen	onbekend
Eco_water	Sub-ecotopen	water
Eco_wegen	Sub-ecotopen	wegen
Ecoh_bos	Hoofd-ecotopen	bos
ecoregio	Ecoregio	
FGR_AFZ	Fysisch Geografische Regio	AFZ



FGR_DUO	Fysisch Geografische Regio	DUO
FGR_DUW	Fysisch Geografische Regio	DUW
FGR_GTW	Fysisch Geografische Regio	GTW
FGR_GTZ	Fysisch Geografische Regio	GTZ
FGR_HLL	Fysisch Geografische Regio	HLL
FGR_HZN	Fysisch Geografische Regio	HZN
FGR_HZO	Fysisch Geografische Regio	HZO
FGR_HZW	Fysisch Geografische Regio	HZW
FGR_HZZ	Fysisch Geografische Regio	HZZ
FGR_LVH	Fysisch Geografische Regio	LVH
FGR_LVN	Fysisch Geografische Regio	LVN
FGR_NZN	Fysisch Geografische Regio	NZN
FGR_NZZ	Fysisch Geografische Regio	NZZ
FGR_RIV	Fysisch Geografische Regio	RIV
FGR_YSS	Fysisch Geografische Regio	YSS
FGR_ZKM	Fysisch Geografische Regio	ZKM
FGR_ZKN	Fysisch Geografische Regio	ZKN
FGR_ZKW	Fysisch Geografische Regio	ZKW
FGR_ZKZ	Fysisch Geografische Regio	ZKZ
Gewas_Aardappelen	Gewas	Aardappelen
Gewas_Bieten	Gewas	Bieten
Gewas_Bloemen	Gewas	Bloemen
Gewas_Braak	Gewas	Braak
Gewas_Fruit	Gewas	Fruit
Gewas_Gras_blijvend	Gewas	Gras_blijvend
Gewas_Gras_tijdelijk	Gewas	Gras_tijdelijk
Gewas_Groenten	Gewas	Groenten
Gewas_Handelsgewas	Gewas	Handelsgewas
Gewas_Luzerne	Gewas	Luzerne
Gewas_Mais	Gewas	Mais
Gewas_Natuurl_gras	Gewas	Natuurlijk_gras
Gewas_Overig	Gewas	Overig
Gewas_Peulvruchten	Gewas	Peulvruchten
Gewas_Uien	Gewas	Uien
Gewas_Wintergranen	Gewas	Wintergranen
Gewas_Zomergranen	Gewas	Zomergranen
GT0_water	Grondwaterstand	0-water
GT1_nat	Grondwaterstand	1-nat
GT2_vrij_nat	Grondwaterstand	2-vrij_nat
GT3_vochtig	Grondwaterstand	3-vochtig
GT5_wisselvochtig	Grondwaterstand	5-wisselvochtig
GT6_vrij_droog	Grondwaterstand	6-vrij_droog
GT7_droog	Grondwaterstand	7-droog
Hoogte_mean		Gemiddelde hoogte ten op zichte van NAP

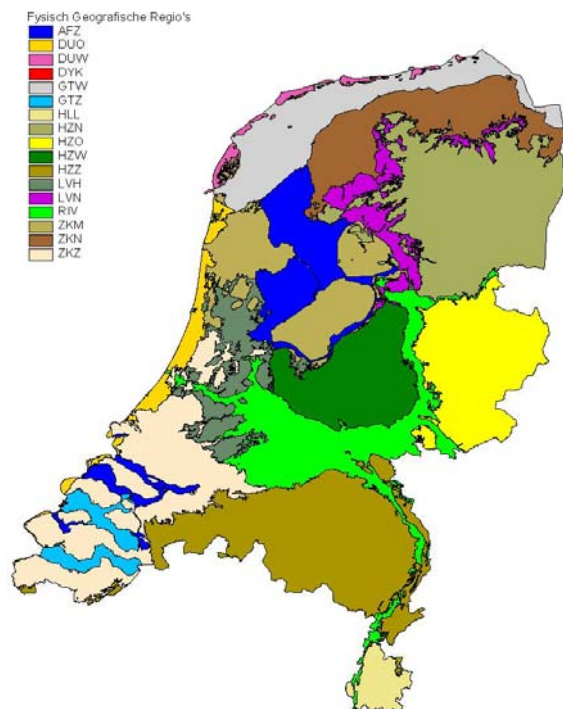


Hoogte_range		Hoogteverschil
lynheg	Landgebruik top10-vector - lijnen	lynheg
lynsloot03	Landgebruik top10-vector - lijnen	lynsloot03
lynsloot36	Landgebruik top10-vector - lijnen	lynsloot36
openheid2009_mean		Openheid van het landschap
Opp_ha		Oppervlakte in ha
Riet_area_perc	Aanwezigheid van riet	Oppervlakte riet
SANSN_Gras	Beheerovereenkomsten SAN-SN-SBB	SAN-Gras
SANSN_Laat_maaaien	Beheerovereenkomsten SAN-SN-SBB	SAN-Laat_maaaien
SBB_Natuurgras	Beheerovereenkomsten SAN-SN-SBB	SBB-Natuurgras
SBB_Weidevogels	Beheerovereenkomsten SAN-SN-SBB	SBB-Weidevogels
Water_klein_diep	Oppervlaktewateren	Kleine diepe wateren
Water_ondiep_veen	Oppervlaktewateren	Ondiep water in het laagveengebied
Water_vennen	Oppervlaktewateren	Vennen
Top10_2006_gebouwdh	Landgebruik top10-vector - huizen	dichtheid aan gebouwen

Enkele variabelen worden hier nader omschreven.

Fysische Geografische Regio's (subeenheden)

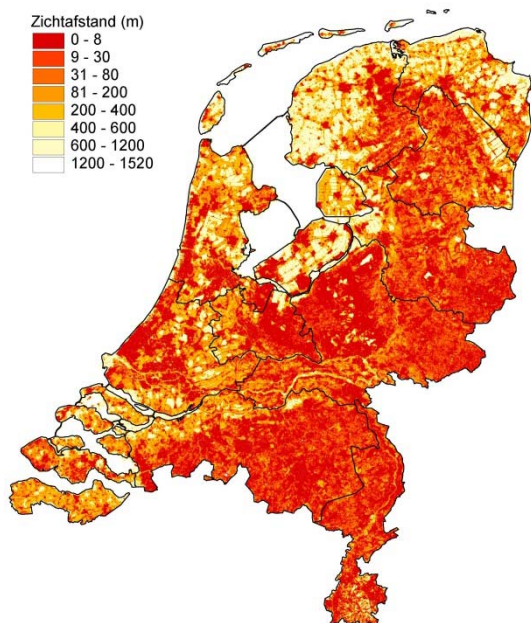
Nederland is verdeeld in regio's (FGR's) die overeenkomen in bodemsamenstelling en geomorfologie/ontstaansgeschiedenis (Natuurbeleidsplan, LNV 1990). Deze zijn op basis van de ligging weer onderverdeeld in subregio's. De Nederlandse kaart is gemaakt door het voormalige ministerie van LNV (IKC-Natuurbeheer) en wordt onder meer gebruikt door het Centraal Bureau voor de Statistiek voor het berekenen van regionale trends. De originele kaart van 1990 is later verfijnd en beschikbaar gekomen als GIS-bestand. De Fysisch Geografische Regio's (FGR's) zijn verder opgedeeld in sub-FGR's (Figuur 2.1). Zo zijn de meeste regio's opgedeeld in noord, west, midden en zuid. Hiermee sluiten de sub-FGR's beter aan bij regionale verschillen als gevolg van bijvoorbeeld klimaat, dan de hoofd-FGR's.



Figuur 2.1. Sub-Fysisch Geografische Regio's (GTW = Getijdengebied Wadden, DUW = Duinen Waddengebied, ZKN = Zeekleigebied Noord, HZN = Hogere Zandgronden Noord, LVN Laagveengebied Noord, ZKW = Zeekleigebied West, ZKM Zeekleigebied Midden, DUO = Duinen Holland en Zeeland, RIV Rivierengebied, LVH = Laagveengebied Holland, ZKZ = Zeekleigebied Zuid, HZZ = Hogere Zandgronden Zuid, HZO = Hogere Zandgronden Oost, GTZ = Getijdengebied Zuid, HLL = Heuvelland, HZW = Hogere Zandgronden West (Utrechtse Heuvelrug en Veluwe), AFZ = Afgesloten Zeearmen.)

Openheid van het landschap

Recent is een kaart met de zichtbare openheid van het landschap beschikbaar gekomen (Meeuwsen & Jochem 2011); zie ook Figuur 2.2. Deze kaart is weliswaar gemaakt voor de menselijke beleving van het landschap, maar heeft een veel hogere resolutie dan de kaarten met schaalkenmerken van het landschap van Dijkstra en Lith-Kranendonk (2000) en leent zich daarom beter voor de analyses die in dit rapport worden beoogd. De resolutie van de openheidskaart is 100 meter, die van de schaal van het landschap was 2 kilometer.



Figuur 2.2. Openheid van het landschap. Weergegeven is de gemiddelde zichtafstand. Deze is berekend door voor elke punt de zichtafstand (in meters) in alle richtingen te bepalen en daarover het gemiddelde te nemen. Duidelijk is dat de zichtafstand laag is in de beboste gebieden in het zuidoosten van het land en hoog in de weidegebieden in Friesland. Deze kaart is met het model ViewScape vervaardigd door Meeuwsen & Jochem (2011).

Beheerstatus

Voor de beheerstatus is gebruik gemaakt van informatie uit 2006. Voor dit jaar is gekozen omdat het midden in de onderzochte periode ligt en bovendien lang genoeg geleden is om invloed te kunnen hebben op het voorkomen en de ontwikkeling van weidevogels. Uit de bestanden van SAN (Subsidieregeling Agrarisch Natuurbeheer), SN (Subsidieregeling Natuur) en SBB is de volgende informatie afgeleid: aandeel reservaat, aandeel grasland met een beheerovereenkomst met een uitgestelde maaidatum tot 15 juni ('vroeg maaien') of met een uitgestelde maaidatum na 15 juni ('laat maaien') en aandeel grasland met beheerovereenkomsten voor verschillende graslandrelevante doelstellingen. Samenvoeging van de vele verschillende beheerspakketten was nodig om de enorme diversiteit aan beheerspakketten overzichtelijk en analyseerbaar te maken.

Drooglegging en gemiddelde voorjaars grondwaterstand

Uit eerder onderzoek in Noord-Holland (Van 't Veer *et al.* 2008) is gebleken dat de drooglegging in de winter een belangrijke relatie heeft met de trend van grondwatergebonden vogels. Om de drooglegging te kunnen bepalen werd gebruik gemaakt van de peilbesluiten van de waterschappen. Deze waren beschikbaar voor alle waterschappen met digitale peilbesluiten met uitzondering van het Waterschap Veluwe.

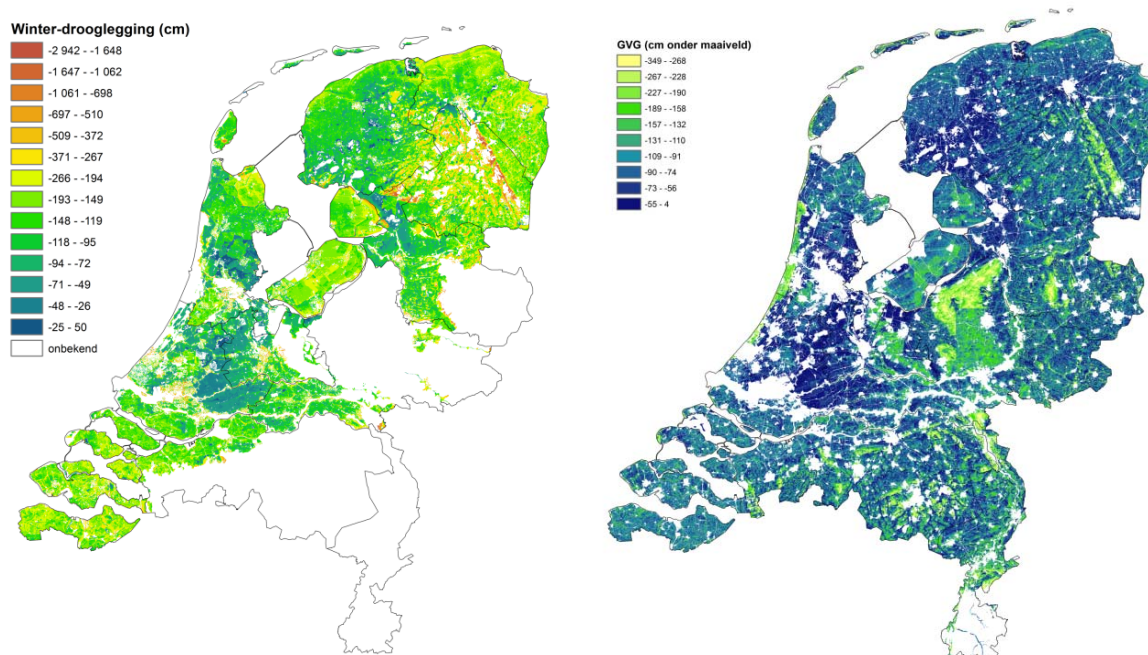
Een probleem bij de bewerking van de peilbesluitbestanden is dat niet altijd duidelijk is of er géén peilbesluit is óf dat het peilbesluit 0 cm NAP is. De controle daarop moest handmatig gebeuren door te kijken of er binnen een gebied met peilbesluiten polders waren die hoogstwaarschijnlijk een peilbesluit van 0 cm hadden. In de voorliggende analyse zijn peilbesluiten van 0 cm NAP deels buiten beschouwing gelaten waardoor lokaal dus omissies in de kaarten kunnen voorkomen. Het bestand met



peilbesluiten is omgezet naar een 100m-grid bestand door per gridcel minimum, gemiddelde en maximum peil te berekenen.

De droogleggingskaart is vervolgens gemaakt door de peilbesluiten te combineren met het AHN (Actuele Hoogtekaart Nederland)-bestand versie 1. Hiervoor zijn eerst alle afzonderlijke 25m-gridbestanden samengevoegd tot één groot bestand.

Op basis van de peilbesluiten en de hoogtekaart is een schatting gemaakt van de grondwaterstand in de winter. Deze waterstand betreft feitelijk de berekende drooglegging van een gebied ten opzichte van het maaiveld. Om de maaiveldhoogte te kunnen bepalen zijn uit de hoogtekaart alleen gemeten oppervlakte-eenheden ('cellen') geselecteerd die volgens de top10-vector van 2006 grasland (tdn-code 5213) of bouwland (tdn-code 5203) zijn. Voorts werden de bestanden met peilbesluiten en het maaiveldhoogtebestand omgewerkt naar een gridbestand dat uit cellen van 25 meter bestond. Hierna is de maaiveldhoogte afgetrokken van het peilbesluit in cm ten opzichte van NAP. Dit levert de geschatte grondwaterstand (drooglegging in cm beneden maaiveld) in de winter op in gridcellen van 25meter (Figuur 2.4). De zomerstanden zijn niet berekend omdat is aangenomen dat de waterstanden aan het begin van het broedseizoen van doorslaggevend belang zijn.



Figuur 2.4. Links: drooglegging in de winter in cm onder het maaiveld. Voor uitleg zie tekst. Rechts: De gemiddelde grondwaterstand in het voorjaar (GVG) in cm t.o.v. het maaiveld.

De gemiddelde grondwaterstand in het voorjaar (GVG) is bepaald door de grondwatertrappen (GWT) uit de bodemkaart en het AHN-hoogtebestand met elkaar te combineren. Uit de grondwatertrappen is de GVG afgeleid voor de eenheden van de bodemkaart. Vervolgens is deze informatie neergeschaald door combinatie met de hoogtekaart. Hierdoor ontstaat een veel fijnmaziger patroon van de ingeschatte GVG (Figuur 2.6).

De informatie uit de droogleggingskaart en de GVG-kaarten zijn samengevoegd om tot een zo goed mogelijk beeld te verkrijgen van de grondwaterstanden in Nederland.



Landgebruik

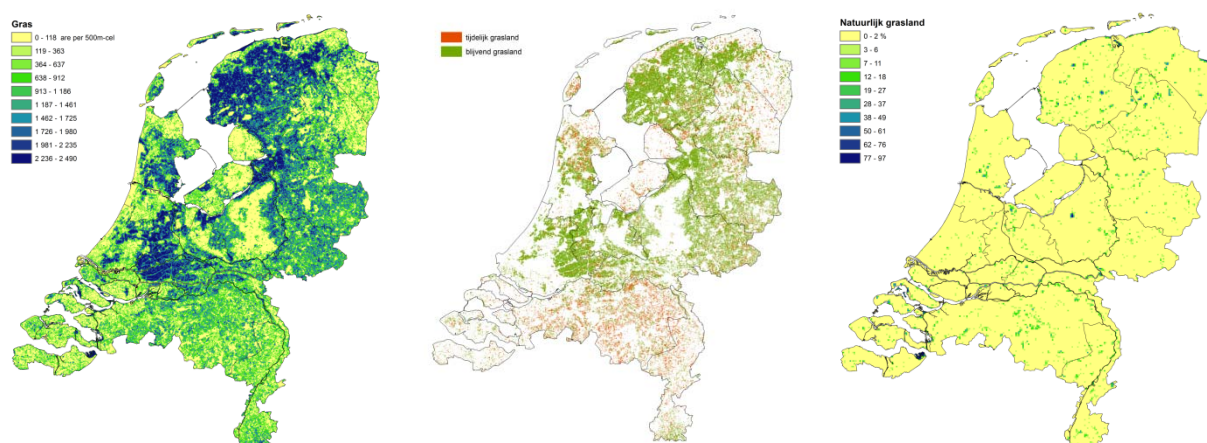
Het landgebruik is afgeleid uit de top10-vector kaart, versie 2006 (Topografische Dienst). De informatie over gewassen, zoals blijvend, tijdelijk en natuurlijk grasland en over winter- en zomergranen en maïs werd afgeleid uit de Gewassenkaart van Dienst Regelingen. Informatie over boomsoortensamenstelling en kiemjaar komt uit de vierde bodemstatistiek, aangevuld met informatie van de CBS-bodemstatistiek

Onder blijvend grasland (Figuur 2.6) wordt verstaan: gras dat voor ten minste vijf jaar niet in de vruchtwisseling is meegenomen. Het gewas bestaat uit een natuurlijke of ingezaaide vegetatie van grassen of andere kruidachtige voedergewassen.

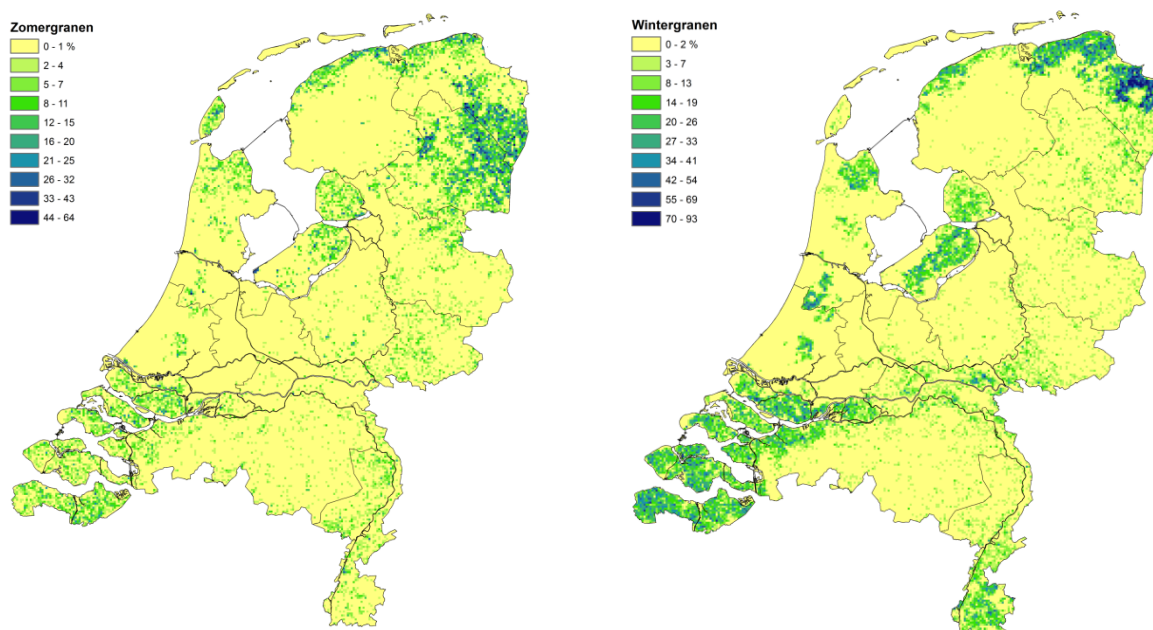
Onder tijdelijk grasland (Figuur 2.6) wordt verstaan: gras dat in de vruchtwisseling is opgenomen. Het gras wordt niet langer dan vijf jaar aaneengesloten geteeld. Het gewas bestaat uit een natuurlijke of ingezaaide vegetatie van grassen of andere kruidachtige voedergewassen.

Onder natuurlijk grasland (Figuur 2.6) wordt verstaan: gewas dat bestaat uit een natuurlijke of ingezaaide vegetatie van grassen of andere kruidachtige voedergewassen. De opbrengst per ha mag niet meer zijn dan vijf ton droge stof per hectare. Het beheer mag gedurende meerdere jaren op geen enkele manier de landbouwkundige productie verhogen of in stand houden (bijv. bemesting, drainage en onkruidbestrijding).

Onder zomergranen (Figuur 2.7) worden tarwe, gerst, haver, graansorgho en gierst verstaan, die aan het einde van de winter worden gezaaid. Onder wintergranen (Figuur 2.7) worden tarwe, gerst, rogge en triticale verstaan, die in in het najaar worden gezaaid.



*Figuur 2.6. Verschillende vormen van grasland in Nederland. Links: totale bedekking met gras in are per cel van 500 * 500 m. Midden: bedekking met tijdelijk en blijvend grasland . Rechts: bedekking met natuurlijk grasland.*



Figuur 2.7. Links: teelt van zomergranen. Rechts: teelt van wintergranen.

2.3 Regressie-analyses

Voor de analyse van het verband tussen verklarende omgevingsvariabelen en het voorkomen per kilometerhok per periode (resp. 2003-2007 en 2008-2013) is gebruik gemaakt van een state-of-the-art type regressiemodellen, namelijk boosted regression trees (BRTs) (Elith et al. 2008). Met BRTs kunnen op een robuuste wijze niet-lineaire verbanden worden beschreven. BRTs combineren de sterke punten van twee algorithmen: regression trees (modellen die het verband tussen afhankelijke en verklarende variabelen tot stand brengen middels recursieve binaire splitsing) en boosting (een adaptieve methode om veel simpele modellen te combineren en hun voorspellende kracht te verbeteren). Hawkins (2012) beveelt expliciet aan voor ruimtelijke analyses regression trees of ervan afgeleide methoden te gebruiken. BRTs schatten een groot aantal vrij eenvoudige modellen waarna de modelschattingen worden gecombineerd hetgeen in robuustere uitkomsten resulteert. Elk model bestaat uit een classificatieboom die regels construeert voor de onafhankelijke variabelen waarmee de responsevariabele (aan-/afwezigheid in ons geval) kan worden opgedeeld in zo homogeen mogelijke groepen. De classificatieboom wordt gevormd door de data herhaaldelijk in tweeën te splitsen volgens een regel gebaseerd op enkele enkele habitatvariabelen. Bij elke splitsing wordt de data in twee zo homogeen mogelijke groepen gesplitst.

Kruisvalidaties en drempelwaardes

De betrouwbaarheid van de modellen worden getoetst met behulp van een evaluatie-dataset. Hiertoe zetten we 20% van onze gegevens opzij en vergelijken het voorkomen in deze 20% met schattingen van voorkomen gebaseerd op de analyse die is uitgevoerd met de resterende 80%. Dit wordt vijf maal herhaald met wisselende, willekeurige selectie van 20% van de data. Vervolgens wordt op verschillende manieren de betrouwbaarheid van de schattingen berekenen: gebruik makend van de AUC-ROC (Area Under ROC-Curve, waar ROC staat voor Receiver Operating Characteristic) en



kruisvalidaties. Evaluatie met behulp van AUC maakt geen gebruik van een drempelwaarde (zie hieronder), de twee andere methodes wel. De drempelwaarde bepaalt bij welke geschatte kanswaarde het verschil tussen aan- en afwezigheid optimaal is. Voor het maken van een voorspelling van het voorkomen van een soort is het schatten van een drempelwaarde meestal gewenst.

De uitkomsten van de kruisvalidaties geven weer hoe goed een model voorspellingen kan maken. Echter het blijft arbitrair waar de grens wordt gelegd tussen een goed en een slecht model. De AUC varieert tussen 0.5 en 1. Hoe hoger de waarde hoe beter het model presteert (tabel 2.3). We gebruiken de classificatie zoals weergegeven in Tabel 3 die algemeen wordt gebruikt in de ecologische literatuur. De beslissing welke nauwkeurigheid nog aanvaardbaar is is tot op zekere hoogte arbitrair. Wij leggen de drempel bij een AUC-waarde van 0.7 (tabel 2.4). Een meer conservatieve drempelwaarde kan worden geprefereerd afhankelijk van de toepassing.

Een AUC-ROC waarde van (bijna) 1 duidt op overfitting: het model beschrijft de waarnemingen dan heel nauwkeurig, maar is niet geschikt voor predicties voor andere locaties. Dit komt vooral voor bij soorten met heel weinig waarnemingen.

Een groot verschil tussen de minimum en maximum AUC-ROC waarde duidt op instabiele modellen: de ene keer is het model wel goed, de andere keer niet.

Area Under Curve van de Receiver-Operator Characteristic (AUC-ROC)

De AUC-ROC is een maat voor de kwaliteit waarmee de binomiale waarnemingen worden voorspeld door het model (Fielding and Bell 1997). Zie ook

http://en.wikipedia.org/wiki/Receiver_operating_characteristic en

<http://nl.wikipedia.org/wiki/ROC-curve> voor achtergronden over AUC-ROC. Een AUC van 0.5 betekent dat het model het niet beter doet dan random, een AUC van 1 betekent dat het model de waarnemingen perfect voorspelt.

De AUC-ROC is zowel bepaald voor het volledige model met alle waarnemingen als voor een 5-voudige kruisvalidatie.

Voor de beoordeling is de volgende indeling gebruikt in tabel 2.2 en 2.3.

Tabel 2.2. Index voor classificatie van modeluitkomsten op basis van de verklaarde deviance. Deze waarden zijn slechts ter indicatie en hangen af van onder meer de grootte van de steekproefgebieden

Percentage verklaarde deviance:

> 80 : uitmuntend

65-80: zeer goed

45-65: goed

30-45: redelijk

15-30: matig

<15: slecht

Tabel 2.3. Index voor classificatie van modeluitkomsten op basis van verschillende validatiemethodes.

Nauwkeurigheid AUC-ROC

Excellent of hoog 0.9 – 1

Goed 0.8 – 0.9

Redelijk 0.7 – 0.8

Matig 0.6 – 0.7

Geen/slecht 0.5 – 0.6



Residue-interpolatie

De residuen vertellen ons waar het model blijkbaar nog niet helemaal goed zit. Vooral als we gebieden zien met overwegend positieve residuen (het voorkomen wordt onderschat) of negatieve residuen (het voorkomen wordt overschat), is er blijkbaar sprake van lokale omstandigheden die niet goed worden beschreven door de variabelen die zijn opgenomen in het regressiemodel. Een vervolgstap kan dan zijn om op zoek te gaan naar variabelen die het gevonden patroon in de residuen kunnen verklaren. Dit zijn dan zgn. 'taylor-made'-modellen: voor elke soort afzonderlijk wordt zo goed mogelijk de meest relevante set aan omgevingsvariabelen bij elkaar gezocht en gemodelleerd. Voor een aantal soorten zal zelfs dat geen soelaas bieden: de relevante informatie is simpelweg niet beschikbaar voor elke locatie in Nederland (denk aan zoiets als de lengte aan bramenwallen of de PH van de bodem) (zie ook (Van Kleunen et al. 2007)). Voor de hier gepresenteerde kanskaarten zijn geen 'taylor-made'-modellen gemaakt omdat die per soort (zeer) veel tijd kosten om te maken. Er is echter nog een andere oplossing om de voorspelde verspreiding te verbeteren: interpolatie van de residuen.

Door de residuen te interpoleren naar een vlakdekkend kaartbeeld ontstaat een kaart met gebieden die overwegend onderschat of overschat worden. Voor interpolatie van de residuen kan gebruik worden gemaakt van (block-) Inverse Distance Weighting (IDW) en Kriging. De laatste methode is veel rekenintensiever dan de eerste en levert veelal globalere patronen op: in deze versie van de kanskaarten is daarom gebruik gemaakt van IDW. (voor een beschrijving van de twee bovengenoemde interpolatie methodieken zie Bivand et al 2008).

De modelvoorspellingen per kilometerhok en de geïnterpoleerde residuen worden tenslotte bij elkaar opgeteld. De predicties en residuen bij elkaar opgeteld op de response-schaal (= de niet getransformeerde schaal). Dit heeft al nadeel dat de in de uiteindelijke predicties getallen kleiner dan 0 en groter dan 1 kunnen optreden. Aangezien kansen altijd tussen 0 en 1 moeten liggen zijn de finale predicties daarom afgebroken tussen 0 en 1.

Technische uitvoering

De berekeningen voor de kanskaarten zijn uitgevoerd met het statistische programma R (R_Development_Core_Team 2004), versie 2.12.0 (64-bits versie). Voor de analyses is het programma 'TRIMmaps' gemaakt. TRIMmaps is een verzameling van R-functies die zorg draagt voor het inlezen van de waarnemingen, samenvoegen met ruimtelijke data en uitvoering van de ruimtelijke modellen. De BRT-modellen zijn gemaakt met het script van Elith et al. (2008) gebaseerd op het package gbm (Ridgeway 2012) en de bij Sovon ontwikkelde verzameling scripts 'TRIMmaps' (Hallmann & Sierdsema 2011). De interpolaties zijn uitgevoerd met R package 'gstat' (Pebesma and Wesseling 1998), binnen de context van TRIMmaps.

2.4. Sensitiviteit en specificiteit van modellen

De sensitiviteit van een model is de kans dat het model een positieve uitslag geeft als dat terecht is. Het is de verhouding tussen het aantal plots met een positieve voorspelling en waarbij de onderzochte soort daadwerkelijk aanwezig is, en het totaal van alle onderzochte plots waar de soort aanwezig is (inclusief het aantal plots dat negatief scoort maar waar de soort toch wel aanwezig is). Het is dus een maat voor de gevoeligheid van het model voor de aanwezigheid van de onderzochte soort.

De specificiteit van een model is een maat voor de kans dat bij afwezigheid van de soort die het model moet voorspellen het resultaat negatief is. De specificiteit van een model is de verhouding tussen het aantal terecht negatieve uitslagen (niet aanwezig, negatieve uitslag) en het totaal van alle plots waarbij de soort afwezig is. Het totaal van alle plots waar de soort afwezig is bestaat uit een som van de plots waar een vals positieve uitslag (wel aanwezig maar niet als zodanig voorspeld) geldt en de gevallen die



een terechte negatieve uitslag kregen.

Een model kan een hoge sensitiviteit (gevoeligheid) hebben, maar vaak vals alarm slaan. Het model moet ook specifiek zijn, dat wil zeggen zoveel mogelijk positieve voorspellingen doen bij de door het model onderzochte soort, en zo weinig mogelijk bij afwezigheid van de soort. Een ideaal model zou een sensitiviteit van 100% moeten hebben (bij alle voorkomens is de voorspelling positief) en ook een specificiteit van 100% (als de soort afwezig is, is de voor-spelling negatief). In werkelijkheid komt dit niet voor. Meestal daalt het ene als het andere stijgt.

De kanskaarten met een waarde tussen 0 en 1 zijn omgezet naar een 0 (verwachte afwezigheid) en 1 (verwachte aanwezigheid) door zowel de specificiteit een twee maal zo hoog belang te geven als de sensitiviteit. Dit heeft tot gevolg, dat voor elke soort een andere 'cutoff'-waarde wordt gebruikt om de kans om te zetten naar een 0 of een 1. Het hogere gewicht van specificiteit zorgt er voor dat op de locaties die de waarde 1 krijgen de soort met een grote kans ook daadwerkelijk voorkomt. Het nadeel van deze benadering is dat meer locaties waar de soort daadwerkelijk voorkomt, toch een 0 als waarde krijgen.

2.5. Invloed van specifieke variabelen op voorspellingen

Een relatieve maat voor het de grote van het effect van elke variabele op de voorspelling van het voorkomen wordt geschat met behulp van een permutatieprocedure. De specifieke variabele wordt vervangen door toevalsgetallen en de uitkomst wordt vergeleken met het model inclusief de originele variabele. Het belang van elke variabele ('variable contribution') wordt weergegeven als 1 min de correlatiecoëfficiënt tussen de originele voorspelling en de voorspelling met de toevalsgetallen: 100 is zeer belangrijk, 0 is geheel onbelangrijk.

De relatieve bijdrage van een variabele

De relatieve invloed (bijdrage) van een verklarende variabele wordt bepaald aan de hand van het aantal keren dat deze variabele wordt geselecteerd voor splitsing, gewogen met het kwadraat van de verbetering van het model veroorzaakt door de splitsing en gemiddeld over alle dendrogrammen (Friedman & Meulman 2003). De bijdrage van elke variabele wordt geschaald, zodat de som van alle bijdragen gelijk is aan 100.

Soortselectie en toevoegen van originele waarnemingen aan de kaarten

Op basis van de validatiestatistieken en cutoff-waardeberekeningen is eerst op geautomatiseerde manier bepaald welke modellen kwalitatief goed zijn en waar de cutoff-waarde moet komen te liggen. Deze zijn vervolgens door een expert beoordeeld door visuele vergelijking van de modelkaarten en presence-absencekaarten. Alleen van soorten waarvan het kaartbeeld als betrouwbaar werd beoordeeld door de expert zijn gebruikt in de vervolgstappen. Indien een andere cutoff-waarde volgens de expert tot een betere kaart zou leiden is deze waarde gebruikt.

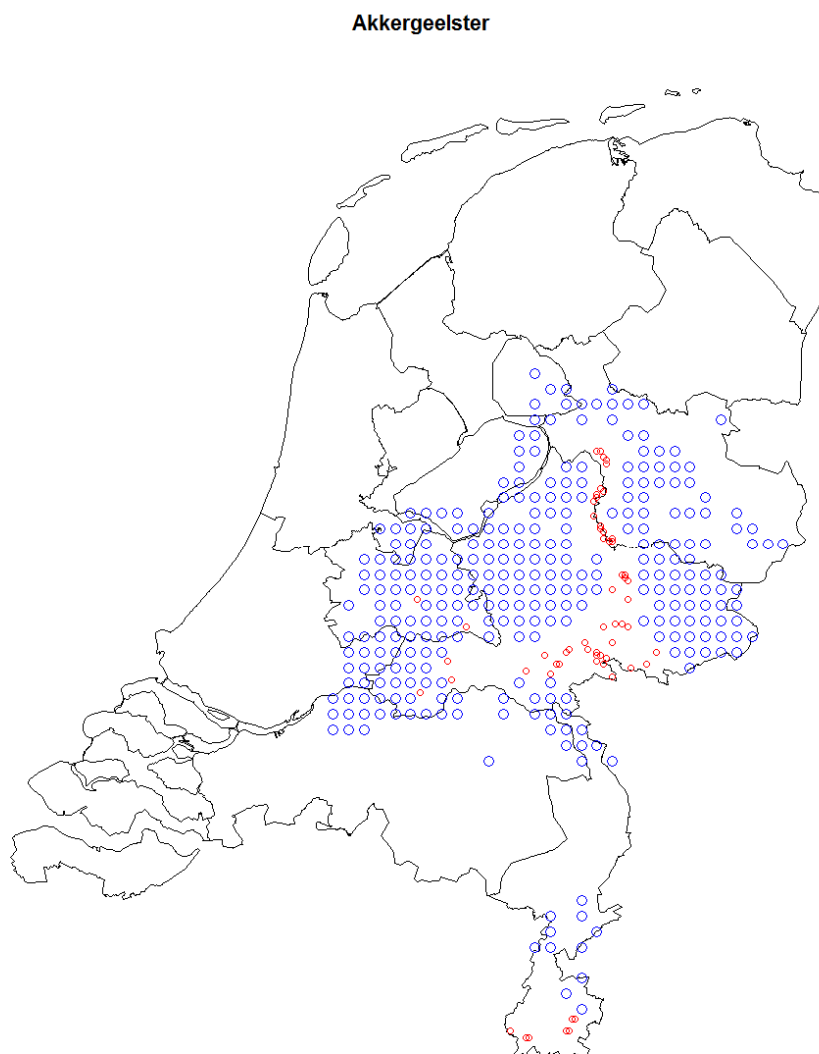
Aan de tot presence-absence kaarten omgezette modelkaarten zijn tenslotte nog de kilometerhokken toegevoegd waar de soort is gemeld. Voor zeer zeldzame en afgekeurde soorten zijn alleen maar de waarnemingen gebruikt. Dit betekent, dat met voor wat ruimer verspreide soorten dit tot een onderschatting van het ruimtelijk voorkomen kan leiden.



3. Resultaten

3.1 Frescalo-kaarten

Voor 802 soorten zijn met behulp van Frescalo nulwaarnemingen per atlasblok van 5x5 km gegeneerd. De nulwaarnemingen hebben veelal niet betrekking op totale gebied waar de soort waarschijnlijk niet voorkomt, maar alleen op regio's aangrenzend aan gebieden met waarnemingen (figuur 3.1)

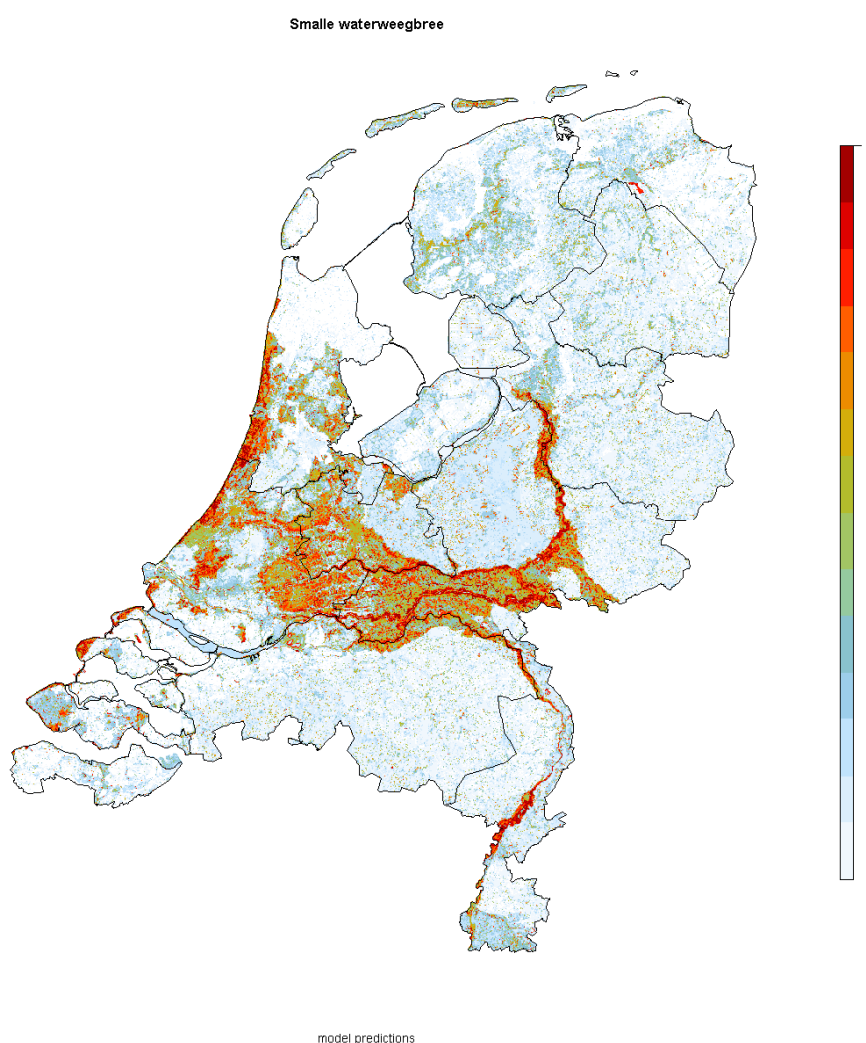


Figuur 3.1 Voorbeeld van de bekende verspreiding van een plantensoort (in rood), de Akkergeelster, en de door Frescalo gegeneerde nulwaarnemingen (blauw) voor deze soort.



3.1 Regressie-modellen

Validaties van de BRT-modellen laten zien dat deze voor de 2005-modellen van 669 van de 734 soorten redelijke tot goede modellen oplevert. De expert beoordeelde 542 als kwalitatief voldoende. De resultaten van 2012 zijn vergelijkbaar met die van 2005.



Figuur 3.2 Voorbeeld van een kanskaart, in dit geval van de Smalle waterweegbree.

3.2 Combinatiekaarten

Voor 100 natuurdoeltypen, 43 SNL-typen en 61 Natura2000-habitats zijn vervolgens



combinatiekaarten gemaakt van het verwachte aantal soorten per 250m-cel.

Literatuur

- BIJLSMA, R.J. 2013 The estimation of species richness of Dutch bryophytes between 1900 and 2011. BLWG-rapport 15, Gouda.
- BIVAND R.S., PEBESMA E.J. & GÓMEZ-RUBIO V. 2008. Applied Spatial Data Analysis with R. Springer, New York.
- DIAMOND J.M. & MAY R.M. 1977. Species turnover rates on islands: dependence on census interval. *SCIENCE* 197:266-270.
- DIJKSTRA H. & VAN LITH-KRANENDONK J. 2000. Schaalkenmerken van het landschap in Nederland. Alterra, Wageningen.
- DRAY S., PÉLISSIER R., COUTERON P., FORTIN M.-J., LEGENDRE P., PERES-NETO P.R., BELLIER E., BIVAND R., BLANCHET F.G., DE CÁCERES M., DUFOUR A.-B., HEEGAARD E., JOMBART T., MUNOZ F., OKSANEN J., THIOULOUSE & WAGNER H.H. 2012. Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs* 82:257-275.
- ELITH J., LEATHWICK J.R. & HASTIE T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 2008, 77:802–813.
- ENS B.J., AARTS B., HALLMANN C., OOSTERBEEK K., SIERDSEMA H., SLATERUS R., TROOST G., VAN TURNHOUT C., WIERSMA P. & VAN WINDEN E. 2011. Scholeksters in de knel: onderzoek naar de oorzaken van de dramatische achteruitgang van de Scholekster in Nederland. SOVON-Onderzoeksrapport 2011/13. SOVON Vogelonderzoek Nederland, Nijmegen.
- ENS B.J., AARTS B., OOSTERBEEK K., ROODBERGEN M., SIERDSEMA H., SLATERUS R. & TEUNISSEN W. 2009. Onderzoek naar de oorzaken van de dramatische achteruitgang van de Scholekster in Nederland. *Limosa* 89:83-92.
- ENS B.J., BERREVOETS C.M., BRUINZEEL L., BULT T., HAANSTRA L., HULSCHER J.B., KOKS B., VAN DE POL M., RAPPOLDT C., TEUNISSEN W.A. & VERHULST S. 2003. Synthese: wat veroorzaakt de huidige achteruitgang van Scholeksters in Nederland? *Limosa* 76:34-38.
- FRIEDMAN J.H. & MEULMAN J.J. 2003. Multiple additive regression trees with application in epidemiology. *Statistics in Medicine* 22:1365–1381.
- GIRAUDOUX P. 2012. *pgirmess*: Data analysis in ecology. R package version 1.5.4. URL <http://CRAN.R-project.org/package=pgirmess>.
- GUERRERO I., MORALES M.B., OÑATE J.J., GEIGER F., BERENDSE F., DE SNOO G., EGGERS S., PÄRT T., BENGTTSSON J., CLEMENT L.W., WEISSER W.W., OLSZEWSKI A., CERYNGIER P., HAWRO V., LIIRA J., AAVIK T., FISCHER C., FLOEHRE A., THIES C. & TSCHARNTKE T. 2012. Response of ground-nesting farmland birds to agricultural intensification across Europe: Landscape and field level management factors. *Biological Conservation* 152:74-80.
- HALLMANN C. & SIERDSEMA H. 2012. TRIMmaps: a R package for the analysis of species abundance and distribution data. Manual, Sovon Vogelonderzoek Nederland.
- HAWKINS B.A. 2012. Eight (and a half) deadly sins of spatial analysis. *Journal of Biogeography* 39:1-9



- HILL, M.O. 2012. Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution* 3: 195-205.
- HULSCHER J.B. & VERHULST S. 2003. Opkomst en neergang van de Scholekster *Haematopus ostralegus* in Friesland in 1966-2000. *Limosa* 76:11-22.
- KLEIJN D., BERENDSE F., SMIT R. & GILISSEN N. 2001. Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. *Nature* 413:723-725.
- KLEIJN D., BERENDSE F., SMIT R., GILISSEN N., SMIT J., BRAK B. & GROENEVELD R. 2004. Ecological effectiveness of agri-environment schemes in different agricultural landscapes in The Netherlands. *Conservation Biology* 18, 775-786.
- KLEIJN D. 2013. De effectiviteit van Agrarisch Natuurbeheer. Rapport voor RLI, 23 bladzijden. Alterra, Wageningen.
- LIPS M. 2011. Detection of grassland management intensity using satellite imagery to support the meadow bird protection. Thesis Report GIRS-2011-21. WUR, Wageningen.
- LNV 1990. Natuurbeleidsplan. Regeringsbeslissing. Ministerie van Landbouw, Natuurbeheer en Visserij. Den Haag.
- MEEUWSEN H.A.M. & JOCHEM R. 2011. Openheid van het landschap; Berekeningen met het model ViewScape. WOt-werkdocument 281. Wettelijke Onderzoekstaken Natuur & Milieu, Wageningen.
- NEWTON I. 2004. The recent declines of farmland bird populations in Britain: an appraisal of causal factors and conservation actions. *Ibis* 146:579-600.
- R DEVELOPMENT CORE TEAM 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Wenen, Oostenrijk. URL <http://www.R-project.org/>.
- RIDGEWAY G. 2012. gbm: Generalized Boosted Regression Models. R package version 1.6-3.2. URL <http://CRAN.R-project.org/package=gbm>.
- SCHOTMAN A.G.M., KIERS M.A. & MELMAN T.C.P. 2007. Onderbouwing Grutto-geschiktheidkaart; Ten behoeve van Grutto-mozaïkmodel en voor identificatie van weidevogelgebieden in Nederland. Alterra, Wageningen.
- TEUNISSEN W.A., ALTENBURG W. & SIERDSEMA H. 2005. Toelichting op de Gruttokaart van Nederland 2004. SOVON Vogelonderzoek Nederland & Altenburg & Wymenga ecologisch onderzoek bv., Beek-Ubbergen.
- TEUNISSEN W.A., VAN PAASSEN A., NIENHUIS J. & SIERDSEMA H. 2013. Weidevogelbalans 2013. Sovon Vogelonderzoek Nederland, Nijmegen.
- TEUNISSEN W.A., SCHOTMAN A.G.M., BRUINZEEL L.W., TEN HOLT H., OOSTERVELD E.O., SIERDSEMA H., WYMENGA E. & MELMAN T.C.P. 2012. Op naar kerngebieden voor weidevogels in Nederland. Werkdocument met randvoorwaarden en handreiking. Wageningen, Alterra, Alterra-rapport 2344. Nijmegen Sovon Vogelonderzoek Nederland, Sovon-rapport 2012/21, Feanwâlden, Altenburg & Wymenga ecologisch onderzoek, A&W- rapport 1799. 144 blz.; 63 fig.; 22tab.; 76 ref
- VAN 'T VEER R., SIERDSEMA H., MUSTERS C.J.M., GROEN N. & TEUNISSEN W. 2008. Weidevogels op landschapsschaal, ruimtelijke en temporele veranderingen. Ministerie van Landbouw, Natuur en Voedselkwaliteit; directie kennis Ede.
- WILLEMS F, BREEUWER A, FOPPEN R, TEUNISSEN W, SCHEKKERMAN H, GOEDHART P, KLEIJN D, BERENDSE F. 2004. Evaluatie Agrarisch Natuurbeheer: effecten op weidevogeldichtheden. Sovon Onderzoekrapport 2004/02. Sovon, Beek-Ubbergen.



Bijlage 1 Inhoud excel-bestanden met validatie-statistieken

In de bestanden 'cvstats2005_edt.xlsx' en 'cvstats2012_edt.xlsx' is te zien welke soorten zijn gemodelleerd, wat de validatiestatistieken opleveren en van welke soorten uiteindelijk de regressiemodellen zijn gebruikt. Deze zijn in de kolom 'OK' aangegeven met een '1'. Van de overige soorten zijn dus alleen de daadwerkelijke waarnemingen per kilometerhok gebruikt. Tevens is per soort vermeld in de kolom 'Cutoff' welke cutoff-waarde is gebruikt om de kanskaarten om te zetten in presence-absencekaarten.

Bijlage 2 Inhoud excel met variabele-bijdragen

In de excel-bestanden 'gbm_contribs_2005.xlsx' en 'gbm_contribs_2012.xlsx' is voor elke soort en elke variabele vermeld wat deze variabele in relatieve zin (percentages) bijdraagt aan de verklarende waarde van het model. De getallen per soort tellen dus op tot 100. De variabelen zijn per soort gesorteerd op afnemend belang.