Voorspellen van ritnaaldschade in aardappelen via perceelskenmerken

Een modelmatige benadering voor het inschatten van de kans op ritnaaldaantasting

Bas Allema, Janne Kool, Arjan Mager en Hilfred Huiting

WAGENINGEN
UNIVERSITY & RESEARCH

Voorspellen van ritnaaldschade in aardappelen via perceelskenmerken

Een modelmatige benadering voor het inschatten van de kans op ritnaaldaantasting

Bas Allema[1], Janne Kool[1], Arjan Mager[2] en Hilfred Huiting[1]

1 Wageningen University & Research
2 Vertify

Dit onderzoek is uitgevoerd in het kader van het PPS-project Grondige Aanpak Bodemplagen (LWV21.197). In dit project werken telers, adviespartijen, afnemers en onderzoekers samen. De ambitie is om gezamenlijk met concrete maatregelen voor een aanpak te komen of modellen te ontwikkelen die de kans op bodemplagen in percelen – of delen daarvan – voorspellen.

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

Lelystad, juli 2026

Rapport WPR-OT-1241

Bas Allema, Janne Kool, Arjan Mager en Hilfred Huiting, 2026. *Voorspellen van ritnaaldschade in aardappelen via perceelskenmerken; Een modelmatige benadering voor het inschatten van de kans op ritnaaldaantasting.* Deelrapportage bij de PPS Grondige aanpak Bodemplagen. Wageningen Research.

Dit rapport is gratis te downloaden op https://doi.org/10.18174/719895

Trefwoorden: *Agriotes*, ritnaalden, aardappel, model, voorspelling, Integrated Crop Management, ICM, perceelseigenschappen, voorvrucht

© 2026 Wageningen, Stichting Wageningen Research, Wageningen Plant Research,
Postbus 430, 8200 AK Lelystad; T 0320 29 11 11; www.wur.nl/plant-research

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

Dit werk is gelicentieerd onder de Creative Commons CC BY-NC-ND 4.0 licentie. Deze licentie stelt hergebruikers in staat om het materiaal te verveelvoudigen en te delen in elk medium of bestandsformaat, mits in onaangepaste vorm, alleen voor nietcommerciële doeleinden en alleen zolang de maker en de bron worden vermeld. Verveelvoudiging of openbaarmaking is niet toegestaan voor die gedeelten van deze uitgave waarvan duidelijk is dat de auteursrechten liggen bij derden of zijn voorbehouden.

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

Rapport WPR-OT-1241

Foto omslag: Hilfred Huiting

Inhoud

Samenvatting **5**

1 Inleiding **6**

1.1 Veranderende teeltcontext 6
1.2 Beheersing bodemplagen integraler met ICM 6
1.3 Modellen als instrument binnen ICM 7
1.4 Voorspellingsmodel voor Nederland 8

2 Aanpak **9**

2.1 Datasets ritnaaldaantasting 9
2.2 Datasets verklarende variabelen 9
2.3 Modelaanpak 9
2.4 Lineair model 11
2.5 Boosted regression trees 11

3 Resultaat **12**

3.1 Lineair model 12
3.1.1 Gecombineerde dataset (dataset 1, 2 & 3) 13
3.1.2 Dataset 1 13
3.1.3 Dataset 2 14
3.1.4 Dataset 3 15
3.1.5 Subset dataset 1 & 2 van percelen op zandgrond 16
3.1.6 Subset dataset 1 & 2 van percelen op kleigrond 17
3.2 Boosted regression trees 17
3.2.1 Gecombineerde dataset (dataset 1, 2 & 3) 17
3.2.2 Dataset 1 18
3.2.3 Dataset 2 19
3.2.4 Dataset 3 19
3.3 Vergelijken lineair model met boosted regression trees 20

4 Conclusie en aanbevelingen **21**

4.1 Vergelijking lineair model versus boosted regression tree 21
4.2 Splitsing dataset in zand- en kleipercelen 21
4.3 Interpretatie van de gevonden verklarende parameters 21
4.4 Toepassing 23

Literatuur **24**

Bijlage 1 Geteste lineaire regressiemodellen **25**

Bijlage 2 Identificatiekaart ritnaaldschade **28**

Samenvatting

In dit deelrapport binnen de PPS Grondige Aanpak Bodemplagen is onderzocht in hoeverre schade in aardappel door ritnaalden voorspeld kan woorden aan de hand van beschikbare gegevens over het perceel, inclusief teeltgeschiedenis en bodemeigenschappen. Data van ritnaaldschade in aardappelen is beschikbaar gesteld door drie in het consortium betrokken aardappelbedrijven en omvatte een totaal van 1789 bemonsterde percelen uit de teeltjaren 2021 t/m 2024. Ritnaaldschade is behandeld als een binaire variabele (0/1) voor afwezigheid of aanwezigheid van ritnaaldschade. De dataset bevatte ongeveer evenveel percelen met als zonder ritnaaldaantasting. Als modelbenadering is gekozen voor een lineair regressiemodel en boosted regression trees. De prestatie van beide modelbenaderingen zijn met elkaar vergeleken op het correct voorspellen van ritnaaldaantasting op een validatie-dataset, waarbij het lineaire regressiemodel iets beter presteerde dan de boosted regression tree. Dataset 3 bleek niet geschikt voor het doel van de studie, omdat ritnaaldaantasting niet was te onderscheiden van andere soorten aantaling; deze dataset is in de verdere behandeling van de resultaten buiten beschouwing gelaten.

Er was een groot verschil in modelpresetatie om ritnaaldaantasting juist te voorspellen tussen percelen op kleigronden (incl. zavel) enerzijds en zandgronden anderzijds. Bijna alle fout-negatieve voorspellingen waren terug te leiden tot kleipercelen. Vervolgens is de data gesplitst in percelen afkomstig van klei- of zandgrond en is voor elke subset een lineair model analyse uitgevoerd. Het meeste adequate model voor de subset van percelen op zandgrond had een fout-negatief (het model voorspelt geen ritnaaldschade terwijl die er wel is) van 4,3% en fout-positief van 28,6%. Voor de kleipercelen was 56,7% fout-negatief en 17,6% fout-positief.

Een grasland-voorgeschiedenis was de belangrijkste voorspeller voor ritnaaldaantasting in aardappelen en dit was met name zo voor zandgronden. Op kleigronden speelt een graslandvoorgeschiedenis een minder grote rol. Het aantal jaar graan en mais in de voorgeschiedenis draagt ook bij aan het verhogen van risico op ritnaaldaantasting, maar in mindere mate dan voor gras. Het modaal kalkgehalte ($CaCO_3$), uitgedrukt als massaprocent van de droge grond van de bovenste bodemlaag volgens de BOFEK2012 bodemkartering, was een andere relevante voorspeller voor ritnaaldaantasting. Op zandgronden met een relatief hoog modaal kalkgehalte, door mariene of fluviatiele afzetting, was de kans op ritnaaldaantasting beduidend lager dan op andere zandgronden.

Concluderend is voor zandgronden een goed voorspellend model gevonden. In woorden uitgedrukt komt dit er op neer dat op zandgronden zonder kalk het model ritnaaldaantasting in aardappelen voorspelt als er in voorgaande jaar gras of graan op het perceel heeft gestaan en zo niet dan voorspelt het model ritnaaldaantasting als er de afgelopen 10 jaar twee of meer keer een gras of graan op het perceel is geteeld. Voor kleigronden liet de data zich niet goed beschrijven door een model met de in deze studie gebruikte verklarende variabelen. Het vereist nadere analyse om te achterhalen wat betere verklarende parameters zijn voor ritnaaldaantasting op kleigrond. Parameters die van invloed kunnen zijn, maar in deze studie niet zijn meegenomen zijn het type en periode van grondbewerking, werkelijk gemeten hoeveelheid organische stof in de bodem en aanwezigheid van grasranden of grasland in de omgeving. Ook zou het interesant zijn om te zien of er op kleigrond andere soorten ritnaalden dominant zijn dan op zandgrond en of dat kan helpen begrijpen welke factoren mee genomen moeten worden in een voorspellend model voor kleigronden.

1 Inleiding

1.1 Veranderende teeltcontext

Een grotere inzet van groenbemesters en minder intensieve grondbewerking passen binnen de huidige ontwikkeling naar een landbouwsysteem met meer aandacht voor bodemgezondheid. Deze aanpak draagt bij aan koolstofvastlegging (CO_2-opslag) en kan het bodemleven verbeteren. Tegelijkertijd brengt dit ook risico's met zich mee, waaronder een mogelijke toename van schade door bepaalde bodemplagen. In de praktijk worden dergelijke problemen al regelmatig waargenomen, wat kan wijzen op een bredere trend.

Binnen verschillende onderzoeksagenda's worden vooral vier bodemplagen genoemd: ritnaalden, emelten, wortelduizendpoten en ondergrondse springstaarten. Het optreden van deze plagen hangt samen met bodemkwaliteit en met landhuur of landruil tussen veehouders en akkerbouw- of bollentelers. Dergelijke samenwerking past goed binnen een kringlooplandbouwsysteem, omdat nutriënten efficiënter kunnen worden benut. In de huidige teeltsystemen kan dit echter ook leiden tot een grotere kans op schade door bodemplagen.

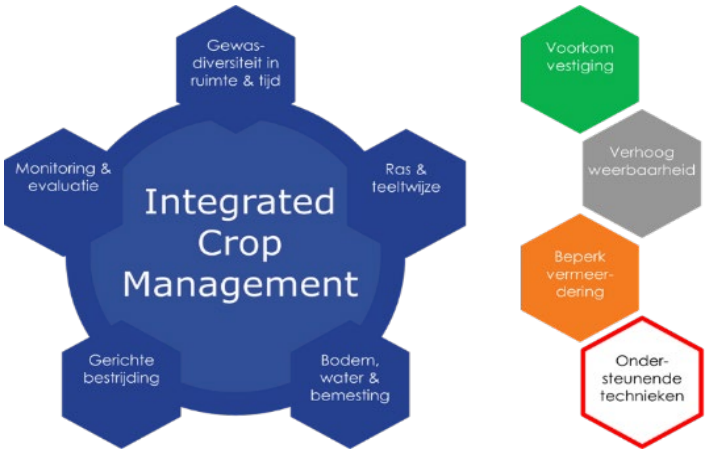
De mogelijkheden om schade door bodemplagen te beheersen zijn momenteel beperkt. De beschikbare en betrouwbare beheersingsmaatregelen waren tot voor kort bovendien voor een belangrijk deel afhankelijk van het gebruik van insecticiden.

1.2 Beheersing bodemplagen integraler met ICM

Voor toekomstige beheersing is een andere aanpak nodig, die meer kennis van soorten en gedrag vraagt dan in het verleden. Ook bestaat de oplossing minder dan in het verleden uit één enkele afdoende maatregel, maar is het combineren van maatregelen de sleutel tot beheersing. Naast kennis over de plaag is daarin specifieke informatie over het teeltperceel belangrijk, omdat het effect van (combinaties van) maatregelen kan wisselen al naargelang de situatie. Idealiter heeft een teler een gereedschapskist met maatregelen ter beschikking waaruit afhankelijk van de situatie geput kan worden. Op die manier kan Integrated Crop Management (ICM) gerealiseerd worden.
ICM is een raamwerk dat is gebaseerd op vier mechanismen en vijf pijlers (Figuur 1).

De vier mechanismen:
- Voorkom vestiging;
- Verhoog weerbaarheid;
- Beperk vermeerdering;
- Ondersteunende technieken.

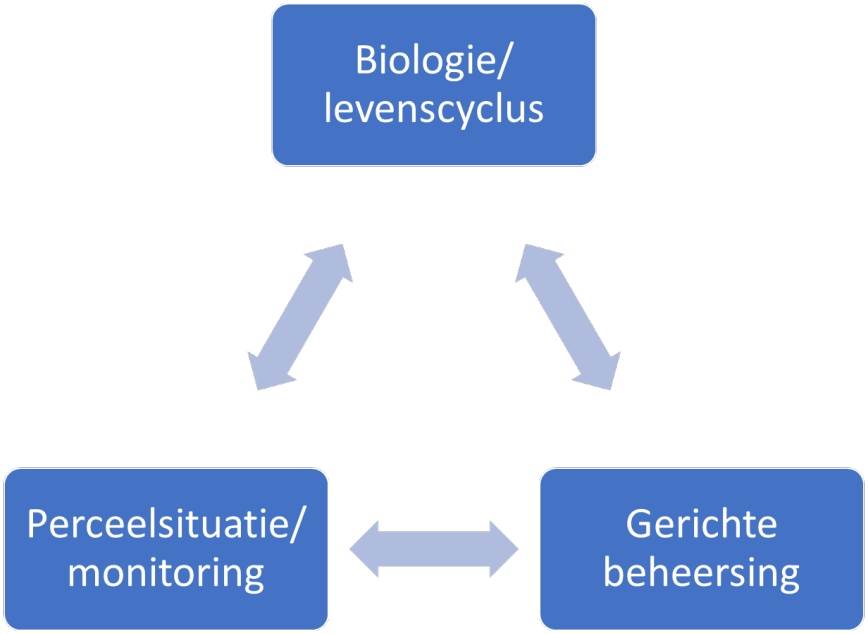
De vijf pijlers:
- Gewasdiversiteit in ruimte & tijd;
- Ras & teeltwijze;
- Bodem, water & bemesting;
- Gerichte bestrijding;
- Monitoring & evaluatie.



Figuur 1 *Schematische samenvatting van de ICM-benadering, met 5 pijlers (links) en 4 mechanismen (rechts).*

Een toekomstbestendige oplossing bestaat in beginsel uit een combinatie van componenten uit meerdere pijlers én is gebaseerd op meerdere mechanismen. Het gebruik van ondersteunende technieken maakt het mogelijk om gebaseerd op data besluiten te nemen én zo gericht mogelijke oplossingen te formuleren.

Voor dit alles is goed begrip nodig van de gewassituatie en de wederzijdse samenhang tussen (1) kennis van biologie en levenscyclus van de bodemplaag, (2) perceelsituatie en monitoring, en (3) gerichte beheersingsmaatregelen (Figuur 2). Inzicht in deze gewas-plaaginteracties maakt het mogelijk om binnen een ICM-benadering te komen tot een integrale aanpak van het kernprobleem. Daarnaast ondersteunen deze inzichten een zorgvuldige en objectieve afweging van effecten op ander ziek en zeer, neveneffecten en het totale teeltrendement.



Figuur 2 *Weergave van de interactie tussen plaagbiologie, perceelsituatie en gerichte beheersing binnen een ICM-benadering.*

Binnen het Publieke Private Samenwerkingsproject 'Grondige Aanpak Bodemplagen' is in de periode 2022-2026 gezocht naar een stapeling van beheersingsstrategieën voor de hier boven genoemde bodemplagen. Eén onderdeel hiervan is het ontwikkelen van een model dat schade door deze bodemplagen kan voorspellen op basis van perceelskarakeristieken. In deze deelrapportage wordt verslag gedaan van deze activiteit voor het voorspellen van ritnaaldaantasting in aardappel.

1.3 Modellen als instrument binnen ICM

Modellen kunnen helpen om op basis van perceels- en/of omgevingskenmerken een inschatting te maken van populaties ritnaalden en/of de kans dat schade optreedt. Deze modellen kunnen telers helpen in de beslissing welke gewassen op welk perceel te telen, waarbij percelen met hoog risico op ritnaaldaantasting beter vermeden kunnen worden voor ritnaaldgevoelige teelten. De modellen kunnen ook helpen om te besluiten welke maatregelen te nemen om een eventueel risico op ritnaaldaantasting te verlagen. Een andere toepassing van modellen die ritnaaldaantasting voorspellen komt van de Po-vlakte in Italië. Al sinds 2005 wordt een risico op ritnaaldschade in mais voorspeld op basis van de risicofactoren (percentage organische stof en doorlopende vegetatiebedekking) en dichtheidskaarten van volwassen kniptorren die met feromoonvallen worden gemonitord (Ferrari *et al.*, 2015). Bij aanwezigheid van hoog-risicofactoren en hoge dichtheden van kniptorren (volwassen stadium) worden ritnaalden bemonsterd in een toekomstig maisperceel. Bij een drempelwaarde van aantal ritnaalden per val, en dit is soortafhankelijk, is het advies om uit te wijken naar een ander perceel. Telers in deze regio kunnen zich aansluiten bij een collectiviteitsfonds (verzekering) dat uitkeert bij ritnaaldschade in mais als de teler heeft gehandeld volgens de richtlijnen van geïntegreerde gewasbescherming.

Het ene model voor ritnaaldaantasting is de andere niet en ze zijn vaak context-specifiek. SIMAGRIO-W is ontwikkeld door Jung *et al.* (2014) voor het westen van Duitsland. Dit prognosemodel gebruikt de verticale distributie van ritnaalden gebaseerd op bodemvocht, bodemtemperatuur en bodemtype om *Agriotes* spp. activiteit te voorspellen. Meer specifiek voorspelt dit model het percentage van de populatie ritnaalden in de bodem dat zich in de bovenste 15 cm van de bodem bevindt. Dit is de zone waar schade aan ondergrondse plantendelen kan optreden en het model geeft dus een inschatting van het risico op schade. Dit model is succesvol toegepast in Duitsland, maar in Oostenrijk bleek dit model de lokale situatie slecht te voorspellen.

In Quebec (Canada) is in 2017 het VFF-QC model ontwikkeld dat een risiconiveau voor aanwezigheid van ritnaalden voorspelt en bepaalt of een specifiek veld een drempelwaarde overschrijdt dat ingrijpen rechtvaardigt. Belangrijke parameters die in dit model van invloed zijn op de risico-inschatting op ritnaaldschade zijn de regio, bodemtextuur, teeltgeschiedenis, bodemhelling, percentage organische stof en grondbewerkingen. Dit model is voor telers die een account hebben te raadplegen via de website https://cerom.qc.ca/vffqc/modules.html (geraadpleegd oktober 2025).

Poggi *et al.* (2018) gebruikt een met VFF-QC vergelijkbare statististische benadering voor maispercelen in Frankrijk, waarbij ze de relatieve invloed van bepalende verklarende variabelen op ritnaaldschade in maisvelden onderzochten en daarmee een voorspelling konden maken voor het risico op ritnaaldschade. Parameters die het meest van invloed waren op aanwezigheid van ritnaalden op het moment van monitoren waren de gemiddelde bodemtemperatuur tussen zaaien en 10 dagen na zaaien, bodem pH, gemiddelde regenval gedurende 50 dagen voor zaaien, gecumuleerde regenval tussen zaaien en het derde bladstadium, type groenbemester, gemiddelde bodemtemperatuur tussen 80 en 70 dagen voor zaaien en gemiddelde bodemtemperatuur 10 dagen voor het zaaien.

Agriorisk is een Belgisch regressiemodel dat ontwikkeld is binnen het Europese ElatPro project (2016-2019) (Wechselberger et al., 2018). Het model voorspelt de aan- of afwezigheid van schadelijke ritnaalden (*Agriotes obscurus* of *Agriotes sputator*) in een perceel op basis van bodemparameters en teeltgeschiedenis. Het voorspellingsmodel was vanaf 2019 een paar jaar online beschikbaar. De belangrijkste factoren waren bodemtype, zuurtegraad (pH), het organische stofgehalte, het aantal jaren granen/gras/mais als hoofdteelt in de afgelopen vijf jaar. De aanwezigheid van een groenbemester in de winter voorafgaand aan de teelt en het type grondbewerking, (kerend of niet-kerend) die één tot vier jaar geleden werden uitgevoerd, kunnen in bepaalde omstandigheden ook een impact hebben op de aanwezigheid of het optreden van ritnaaldschade. Vanaf 2026 wordt er bij het IVLO gewerkt aan een update van het Agriorisk-model met het doel dit weer online beschikbaar te maken voor telers.

Het WW INFEST-RISK (Roche et al., 2023) is een Frans beslissingsondersteunend model dat als een Shiny web applicatie uitgevoerd kan worden (te vinden via doi:10.57745/CBVYQ1). Het model voorspelt het aantal ritnaalden in de bodem (voor 3 spadesteken) op basis van omgevingskarakteristieken, teeltgeschiedenis, bodemeigenschappen, klimaatgegevens en grondbewerkingen. Het WW INFEST-RISK model is anders dan de voorgaande modellen omdat het niet alleen zoekt naar statistische verbanden tussen omgevingsfactoren en plaagdruk, maar de onderliggende biologische processen van ritnaaldverspreiding daadwerkelijk beschrijft.

1.4 Voorspellingsmodel voor Nederland

Geïnspireerd op de benadering van het Agriorisk-model in België is een plan opgesteld om een voorspellingsmodel te maken voor ritnaaldschade in aardappel voor Nederlandse percelen. Een belangrijk uitgangspunt was dat de verklarende factoren die in het model meegenomen worden online beschikbaar moeten zijn, zodat dit bij een latere toepassing van het model geen belemmering hoeft te zijn. Het doel was om ritnaaldschade in aardappelen te voorspellen en hiervoor is dankbaar gebruikgemaakt van datasets van drie in het projectconsortium betrokken aardappelbedrijven, waarin ritnaaldaantasting bij aardappelen is beoordeeld en terug te herleiden tot een perceel waarvan de GPS-coördinaten bekend zijn. Via de GPS-coördinaten van het perceel kon de ritnaaldaantasting gelinkt worden aan de potentieel verklarende factoren van de hiervoor beschreven datasets.

2 Aanpak

2.1 Datasets ritnaaldaantasting

Data over ritnaaldaantasting van aardappelknollen is verkregen van drie in het consortium betrokken aardappelbedrijven, over verschillende jaren in de periode 2021 t/m 2024. De aantasting van knollen is in de drie datasets op verschillende manieren bepaald, elk met een eigen klasse-indeling. Daarom is ervoor gekozen om de mate van ritnaaldschade te reduceren tot een binaire variabele die aangeeft of er wel of geen schade is waargenomen (ritnaald_01). Vooranalyses toonden bovendien aan dat een model met ritnaaldschade als binaire variabele betere voorspellende prestaties had dan modellen waarin ritnaaldschade als nominale of categorische variabele werd opgenomen. Voor elke subset van data, per aardappelbedrijf en per jaar, is gezorgd voor een ongeveer gelijk aantal percelen met en zonder aantasting. In Tabel 1 staan de aantallen percelen per dataset voor zand, klei en overige bodems.

Ritnaaldaantasting is te herkennen als smalle gangetjes met meestal een kleine diameter in de aardappelknol en is soms moeilijk te onderscheiden van Rhizoctonia-aantasting of andere dierlijke vreterij. In de Bijlage 2 is een foto-overzicht toegevoegd om ritnaaldschade te kunnen herkennen. In dataset 1 is specifiek beoordeeld op aantasting door ritnaalden; in dataset 2 is onderscheid gemaakt tussen knollen met gaatjes en zonder gaatjes en in dataset 3 op knollen met niet nader gespecificeerde aantasting of zonder aantasting.

Tabel 1 *Aantal datapunten in de verschillende datasets per grondsoort, zand, klei of anders.*

Grondsoort	Dataset 1	Dataset 2	Dataset 3	Totaal
Zand	299	44	492	835
Klei	395	382	60	837
Anders	50	1	66	117

2.2 Datasets verklarende variabelen

Op basis van het rapport 'Beheersing van bodemplaaginsecten via bodemgezondheidsmaatregelen' (Van Rozen et al. 2021) is een screening gemaakt van mogelijk verklarende variabelen voor verschillen in ritnaaldaantasting in aardappel tussen percelen en is gezocht naar openbare datasets om deze variabelen uit op te halen op basis van een GPS-locatie van het betreffende perceel. Dit heeft uiteindelijk geleid tot het overzicht van mogelijk verklarende variabelen in Tabel 2.

2.3 Modelaanpak

Er zijn twee benaderingen gekozen om een model te maken dat ritnaaldaantasting in aardappelknollen kan verklaren en voorspellen (lineair model en boosted regression trees) die hieronder verder worden uitgewerkt. Bij het lineair regressiemodel is een aantal modellen met verschillende verzamelingen van verklarende parameters met elkaar vergeleken om tot het meest adequaat model te komen; dat is een model dat met zo min mogelijk parameters de aanwezigheid van ritnaaldschaden nog steeds adequaat beschrijft. In het regression trees model is het aan het algoritme overgelaten om tot een set van meest verklarende factoren te komen. De uitkomsten van beide benaderingen zijn met elkaar vergeleken.

Tabel 2 *Overzicht van de gebruikte parameters om ritnaaldaantasting van aardappelknollen te verklaren.*

Variabele	Beschrijving	Dataset[1]
OS	Organische stof gehalte op verschillende diepten van het bodemprofiel.	BOFEK 2012
dichtheid	Volumieke massa (g/cm^3) op verschillende diepten van het bodemprofiel.	BOFEK 2012
M50	Mediaan van de korrelgrootte van de zandfractie (50-2000 µm) op verschillende diepten van het bodemprofiel.	BOFEK 2012
leem	Lutum- + siltfractie op verschillende diepten van het bodemprofiel.	BOFEK 2012
silt	Siltgehalte op basis van de fractie van bodemdeeltjes tussen dan 2-50 µm op verschillende diepten van het bodemprofiel.	BOFEK 2012
lutum	Lutumgehalte op basis van de fractie van bodemdeeltjes kleiner dan 2 µm op verschillende diepten van het bodemprofiel.	BOFEK 2012
BOFEK2012	Nummer van de bofekeenheid voor koppeling met de profielschets waarin de bodemopbouw schematisch wordt weergegeven.	BOFEK2012
CaCO3	Modaal kalkgehalte: meest representatieve calciumcarbonaatgehalte, uitgedrukt als massaprocent van de droge grond, voor verschillende dieptelagen van het bodemprofiel.	BOFEK2012
pH	Modale waarde voor de pH-KCl op verschillende diepten van het bodemprofiel.	BOFEK2012
oogstjaar	Jaar waarin de aardappelen van een bepaalde dataset zijn geoogst	Leveranciers datasets
grondwatertrap	Afgeleide parameter uit de ghg en glg	
glg	Gemiddelde laagste grondwaterstand	Grondwater basiskaart
ghg	Gemiddelde hoogste grondwaterstand	Grondwater basiskaart
groenbemester	Aantal voorgaande jaren tot vijf jaar terug waarop het perceel wintergroen was (Voorjaar_tst of Najaar_tst was groen). Deze variabele is beschikbaar vanaf 2018.	
voorjaar_tst	Classificering of het perceel bedekt (groen, niet groen of onbekend) was in het voorjaar in de vijf voorgaande jaren op het oogstjaar. Deze variabele is beschikbaar vanaf 2018.	Op basis van de NDVI-parameter uit satalietdata
najaar_tst	Classificering of het perceel bedekt (groen, niet groen of onbekend) was in het najaar in de vijf voorgaande jaren op het oogstjaar. Deze variabele is beschikbaar vanaf 2018.	de NDVI-parameter uit satalietdata
grasgraanmais10jaar	Aantal voorgaande jaren waarin er gras, graan of mais op het perceel heeft gestaan tot 10 jaar terug	RVO
grasgraan10jaar	Aantal voorgaande jaren waarin er gras of graan op het perceel heeft gestaan tot 10 jaar terug	RVO
gras10jaar	Aantal voorgaande jaren waarin er gras op het perceel heeft gestaan tot 10 jaar terug	RVO
graan10jaar	Aantal voorgaande jaren waarin er graan op het perceel heeft gestaan tot 10 jaar terug	RVO
mais10jaar	Aantal voorgaande jaren waarin er mais op het perceel heeft gestaan tot 10 jaar terug	RVO
grasgraanjaar_1	Er stond wel/niet gras of graan op het perceel in het voorgaande teeltjaar	RVO
grasgraanjaar_2	Er stond wel/niet gras of graan op het perceel twee seizoenen geleden	RVO
grasgraanjaar_3	Er stond wel/niet gras of graan op het perceel drie seizoenen geleden	RVO
grasjaar_1	Er stond wel/niet gras op het perceel in het voorgaande teeltjaar	RVO
grasjaar_2	Er stond wel/niet gras op het perceel twee seizoenen geleden	RVO
grasjaar_3	Er stond wel/niet gras op het perceel drie seizoenen geleden	RVO
graanjaar_1	Er stond wel/niet graan op het perceel in het voorgaande teeltjaar	RVO
graanjaar_2	Er stond wel/niet graan op het perceel twee seizoenen geleden	RVO
graanjaar_3	Er stond wel/niet graan op het perceel drie seizoenen geleden	RVO
maisjaar_1	Er stond wel/niet mais op het perceel in het voorgaande teeltjaar	RVO
zand-klei	Grondsoort (zand, klei of 'leeg veld')	BOFEK 2012
staring bouwsteen	Bodemtype classificering op verschillende diepten van het bodemprofiel volgens indeling uit Alterra-rapport 153 (vernieuwde uitgave 2001).	?
Specifiek voor de dataset 1		
Rascijfer initial	Cijfer dat iets zegt over de vroege ontwikkeling van de plant	Leverancier dataset 1
Rascijfer final	Cijfer dat iets zegt over de prestatie van de plant richting de oogst	Leverancier dataset 1
Rascijfer maturity	Cijfer dat iets zegt of hoe snel een gewas oogstrijp is	Leverancier dataset 1
Specifiek voor de dataset 2		
Rascijfer	Cijfer dat iets zegt of hoe snel een gewas oogstrijp is	Leverancier dataset 2

[1] Met uitzondering van de glg en ghg en de datasetpecifieke zijn alle parameters ontsloten via Agrodatacube. Voor glg en ghg is www.bodemdata.nl geraadpleegd.

De prestaties van meeste geschikte modellen van elke benadering zijn vervolgens met elkaar vergeleken door middel van een verwarringsmatrix. Hiervoor is de dataset willekeurig verdeeld in een set om de modellen op te fitten dan wel trainen (80% van de datapunten) en een set om het model op te testen (20% van de datapunten).

2.4 Lineair model

Het lineaire model is gebouwd in R met de functie lm. Er zijn verschillende modellen opgesteld en met elkaar vergeleken op basis van de AIC-waarde (Akaike Information Criterion). Modellen die meer dan 2 AIC-punten verschillen, worden beschouwd als betekenisvol verschillend in modelondersteuning. Het model met de laagste AIC-waarde is geselecteerd als het best ondersteunde model. De modelselectie is uitgevoerd voor de combinatie van de drie datasets en voor elke dataset apart, en voor een subset van kleipercelen en een subset van zandpercelen van de gecombineerde dataset 1 en 2. Voor een overzicht van de geteste modellen zie Bijlage 1.

2.5 Boosted regression trees

Er zijn verschillende boosted regression trees gebouwd zoals beschreven in (Elith, Leathwick, & Hastie, 2008). Hierin worden veel relatief eenvoudige modellen gemaakt, die samen een voorspelling geven. De modellen zijn geïmplementeerd in Python 3.9 met behulp van de GradientBoostingRegressor uit de scikit-learn-bibliotheek. Tijdens de modelontwikkeling zijn de boomdiepte, het aantal bomen en de set van voorspellende variabelen gevarieerd. Omdat voor boosted regression trees geen AIC beschikbaar is, zijn de uiteindelijke modellen geselecteerd op basis van een combinatie van voorspellende prestaties en modelparsimonie; daarbij wordt de voorkeur werd gegeven aan modellen met een hoge voorspellende nauwkeurigheid en een beperkt aantal voorspellende variabelen.

3 Resultaat

3.1 Lineair model

De modelselecties uitgevoerd op de gecombineerde en afzonderlijke datasets resulteerden in verschillende modelvergelijkingen die hieronder in de deelparagrafen staan beschreven. De prestaties van het meest adequate model dat voor elk van de datasets was geselecteerd verschilde behoorlijk tussen de datasets (Tabel 3). Het meest adequate model bij dataset 3 gaf een gecorrigeerde R^2 score van nagenoeg 0 en een balanced accuracy van 0,5, wat erop duidt dat de parameters in het model geen toegevoegde waarde hadden ten opzichte van een willekeurige voorspelling en het model dus geen verklarende waarde had.

Tabel 3 *Samenvatting van de prestaties van de geselecteerde meeste adequate modellen bij de verschillende datasets. De aangepaste R-squared is weergegeven voor de fit van het model op de subset van trainingsdata. De andere parameters zijn bepaald voor de fit van het model op de subset van testdata (modelvalidatie).*

	Meest adequate model[1]	Aangepaste R-squared[2]	Balanced accuracy[3]	Precision[4]	Sensitivity[5]	F1-score[6]
Gecombineerde datasets	model 90	0,21	0,701	0,749	0,733	0,741
Dataset 1	model 89	0,41	0,836	0,892	0,773	0,828
Dataset 2	model 106	0,18	0,579	0,6	0,558	0,578
Dataset 3	model 84	0,03	0,502	0,689	0,973	0,807

[1]Meest adequate model: het eenvoudigste model dat de data nog adequaat beschrijft. Zie de volgende paragrafen voor de modelbeschrijving.

[2]Aangepaste R-squared: het aandeel verklaarde variantie, gecorrigeerd voor het aantal parameters in het model, waarbij onnodige variabelen worden bestraft. De R^2 en de volgende parameters zijn weergegeven voor de prestaties van het model op de subset van de test data.

[3]Balance accuracy: gemiddelde van het correct herkennen van positieve én negatieve gevallen, gecorrigeerd voor ongebalanceerde klassen.

[4]Precision: welk deel daadwerkelijk ritnaaldaantasting heeft van alle percelen die als "positief" worden voorspeld.

[5]Sensitivity: welk deel correct wordt geïdentificeerd door het model van alle percelen met ritnaaldaantasting. 1 minus de sensitiviteit is het aantal fout-negatieve voorspellingen.

[6]F1-score: samengestelde maat die precision en sensitivity combineert tot één getal.

In een nadere analyse van de fout-positieve en fout-negatieve uitslagen bleek er een groot onderscheid te zijn tussen de percelen op klei- en zandgrond. Dit waren tevens de grootste twee groepen grondsoorten in de dataset (percelen op leem- en andere gronden waren (vrijwel) absent). Vervolgens is er een modelselectie uitgevoerd op een subset van zand- en een subset van kleipercelen van de gecombineerde dataset 1 & 2. De subset van zand- en subset van kleipercelen van dataset 3 was in eerste instantie ook meegenomen, maar verlaagde de modelprestatie enorm en is verder buiten beschouwing gelaten.

Tabel 4 *Samenvatting van de prestaties van de geselecteerde meeste adequate modellen bij subset van zandpercelen en de subset van kleipercelen in de gecombineerde dataset 1 en 2. De aangepaste R-squared is weergegeven voor de fit van het model op de subset van trainingsdata. De andere parameters zijn bepaald voor de fit van het model op de subset van test data (model validatie).*

Subsets uit dataset 1 & 2		Aangepaste R-squared	Balanced accuracy	Precision	Sensitivity	F1-score
Percelen op zandgrond	Model 98	0,384	0,836	0,882	0,957	0,918
Percelen op kleigrond	Model 101	0,173	0,629	0,659	0,433	0,522

De voorspellende waarde van het model doet het beduidend beter voor percelen op zandgrond dan voor percelen op kleigrond (Tabel 4). Het percentage fout-negatieven (het model voorspelt geen schade terwijl deze er wel is) is 4,3% (sensitivity 0,957) voor percelen op zandgrond tegen 56,7% voor percelen op kleigrond.

3.1.1 Gecombineerde dataset (dataset 1, 2 & 3)

Het geselecteerde meest adequate model was model 90 (zie Bijlage 1):

> ritnaald_01 ~ **$CaCO_3$** + **xgras10jaar** + **xmais10jaar** + **xgraan10jaar:mais10jaar** + **grasgraan_1TRUE**

Hierbij waren alleen bovenstaande vetgedrukte verklarende parameters significant (Tabel 5). In dit model wordt de binaire variabele van aan- of afwezigheid van ritnaaldaantasting in knollen van een perceel het best verklaard door het modaal kalkgehalte uitgedrukt als massaprocent van de droge grond, het aantal voorgaande jaren waarop er gras is geteeld (xgras10jaar), het aantal voorgaande jaren waarop er mais is geteeld (xmais10jaar) een interactie tussen het aantal jaar dat er graan en het aantal jaar dat er mais is geteeld (xgras10jaar:mais10jaar) en de binaire parameter voor aan- of afwezigheid van gras of graan in het voorgaande teeltjaar.

Er zijn weliswaar andere modellen met een tot 4 punten lagere AIC score (Bijlage 1) dan model 90, maar deze deze bevatten aanzienlijk meer onderling gecorreleerde bodemparameters. Hoewel dit statistisch kan leiden tot een betere fit, is dit moeilijk te rechtvaardigen op basis van de onderliggende biologische en bodemkundige mechanismen.

Tabel 5 *Samenvatting van het meest adequate model (model 90) bij de gecombineerde dataset.*

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Signif.**
(Intercept)	0.39964	0.03175	12.59	< 2e-16	***
$CaCO_3$	-0.03011	0.00495	-6.09	1.5e-09	***
xgras10jaar	0.03949	0.00403	9.80	< 2e-16	***
grasgraan_1TRUE	0.14105	0.02757	5.12	3.6e-07	***
xmais10jaar	0.04570	0.00649	7.04	3.0e-12	***
xgraan10jaar:xmais10jaar	-0.01606	0.00657	-2.44	0.015	*

Significantiecodes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

De verwarringsmatrices voor de trainings- en testdata staan weergegeven in Tabel 6.

Tabel 6 *Verwarringsmatrices voor trainings- en testdata; gecombineerde dataset.*

TRAINING			**TEST**		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling ↓	**0**	**1**
0	410	205	**0**	95	51
1	198	565	**1**	47	140

3.1.2 Dataset 1

Het geselecteerde meest adequate model was model 89 (zie Bijlage 1):

> ritnaald_01 ~ M50 + **$CaCO_3$** + **xgras10jaar** + **xmais10jaar** + **xgraan10jaar:mais10jaar** + **grasgraan_1TRUE**

Hierbij waren alleen bovenstaande vetgedrukte verklarende parameters significant (Tabel 7). In dit model wordt de binaire variabele van aan- of afwezigheid van ritnaaldaantasting in knollen van een perceel het best verklaard door de mediaan van de korrelgrootte van de zandfractie (M50), het modaal kalkgehalte uitgedrukt als massaprocent van de droge grond, het aantal voorgaande jaren waarop er gras is geteeld (xgras10jaar), het aantal voorgaande jaren waarop er mais is geteeld (xmais10jaar), een interactie tussen het aantal jaar waarop er graan en aantal jaar waarop er mais is geteeld (xgraan10jaar:mais10jaar) en de binaire parameter voor aan- of afwezigheid van gras of graan in het voorgaande teeltjaar.

Het gekozen model 89 was equivalent of 1 AIC punt hoger dan drie andere modellen die in parameterisering niet veel verschilden van model 89. De reden dat voor model 89 is gekozen is dat deze een interactieterm had voor aantal jaar graan en aantal jaar mais in de voorgeschiedenis los van de hoofdeffecten, wat het model beter interpreteerbaar maakt.

Tabel 7 *Samenvatting van het meest adequate model (model 89) bij dataset 1.*

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Signif.**
(Intercept)	0.160809	0.132636	1.21	0.226	
M50	0.001749	0.000923	1.90	0.059	.
$CaCO_3$	-0.036145	0.009056	-3.99	7.4e-05	***
xgras10jaar	0.044681	0.005535	8.07	4.0e-15	***
grasgraan_1TRUE	0.163835	0.035390	4.63	4.5e-06	***
xmais10jaar	0.046227	0.010658	4.34	1.7e-05	***
xgraan10jaar:xmais10jaar	-0.019854	0.009115	-2.18	0.030	*

Significantiecodes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

De verwarringsmatrices voor de trainings- en testdata staan weergegeven in Tabel 8.

Tabel 8 *Verwarringsmatrices voor trainings- en testdata; dataset 1.*

TRAINING			**TEST**		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling ↓	**0**	**1**
0	246	76	**0**	62	17
1	45	220	**1**	7	58

3.1.3 Dataset 2

Het geselecteerde meest adequate model was model 106 (zie Bijlage 1):

ritnaald_01 ~ **xgras10jaar** + **grasgraan_1TRUE** + **grasgraan_2TRUE** + **grasgraan_3TRUE**

Hierbij waren alleen bovenstaande vetgedrukte verklarende parameters significant (Tabel 9). In dit model wordt de binaire variabele van aan- of afwezigheid van ritnaaldaantasting in knollen van een perceel het best verklaard door het aantal voorgaande jaren waarop er gras (xgras10jaar) is geteeld en de binaire parameters voor aan- of afwezigheid van gras of graan in het voorgaande teeltjaar (grasgraan_1TRUE), twee jaar terug (grasgraan_2TRUE) en drie jaar terug (grasgraan_3TRUE).

Tabel 9 *Samenvatting van het meest adequate model (model 106) bij dataset 2.*

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	Signif.
(Intercept)	0.3526	0.0471	7.48	6.5e-13	***
xgras10jaar	0.0643	0.0090	7.15	5.7e-12	***
grasgraan_1TRUE	0.1561	0.0538	2.90	0.0039	**
grasgraan_2TRUE	-0.1235	0.0546	-2.26	0.0244	*
grasgraan_3TRUE	-0.1447	0.0629	-2.30	0.0220	*

Significantiecodes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

De verwarringsmatrices voor de training en testdata staan weergegeven in Tabel 10.

Tabel 10 *Verwarringsmatrices voor trainings- en testdata; dataset 2.*

TRAINING			TEST		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling ↓	**0**	**1**
0	113	63	**0**	24	19
1	53	108	**1**	16	24

3.1.4 Dataset 3

Het geselecteerde meest adequate model was model 84 (zie Bijlage 1):

ritnaald_01 ~ **M50** + **pH** + **xgraan10jaar**

Hierbij waren bovenstaande vetgedrukte verklarende parameters significant (Tabel 11). Dit model had echter een aangepaste R^2 van 0,03 en weinig voorspellende waarde, waardoor interpretatie van de gevonden verklarende parameters eerder een statistische uitvloeisel zijn van een grote dataset dan gebaseerd op een verklarend mechanisme.

Tabel 11 *Samenvatting van het meest adequate model (model 84) bij dataset 3.*

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	Signif.
(Intercept)	-0.58735	0.42805	-1.37	0.17070	
M50	0.00313	0.00144	2.17	0.03070	*
pH	0.16212	0.04810	3.37	0.00082	***
xgraan10jaar	-0.05951	0.02159	-2.76	0.00609	**

Significantiecodes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

De verwarringsmatrices voor de trainings- en testdata (Tabel 12) laten zien dat het model eigenlijk altijd ritnaaldaantasting voorspelt en dus weinig voorspellende waarde heeft.

Tabel 12 *Verwarringsmatrices voor trainings- en testdata; dataset 3.*

TRAINING			TEST		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling ↓	**0**	**1**
0	9	4	**0**	1	2
1	142	299	**1**	32	71

3.1.5 Subset dataset 1 & 2 van percelen op zandgrond

Het geselecteerde meest adequate model was model 98 (zie Bijlage 1):

ritnaald_01 ~ **$CaCO_3$** + **xgrasgraanmais10jaar** + **grasgraan_1**

Hierbij waren alleen bovenstaande vetgedrukte verklarende parameters significant (Tabel 13). In dit model wordt de binaire variabele van aan- of afwezigheid van ritnaaldaantasting in knollen van een perceel het best verklaard door het modaal kalkgehalte ($CaCO_3$), uitgedrukt als massaprocent van de droge grond, het aantal voorgaande jaren waarop er gras, graan of mais is geteeld (xgrasgraanmais10jaar) en de binaire parameter voor aan- of afwezigheid van gras of graan in het voorgaande teeltjaar.

Er waren twee andere modellen met gelijke aantal AIC punten als het gekozen model 98 (Bijlage 1). Model 98 is van deze drie gekozen als meest adequate model, omdat het de minste parameters had.

Tabel 13 *Samenvatting van het meest adequate model (model 98) bij de subset uit dataset 1 & 2 van percelen op zandgrond.*

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Signif.**
(Intercept)	0.40868	0.07939	5.15	5.2e-07	***
xgrasgraanmais10jaar	0.04492	0.01005	4.47	1.2e-05	***
grasgraan_1TRUE	0.13993	0.04829	2.90	0.0041	**
$CaCO_3$	-0.04736	0.00988	-4.79	2.7e-06	***

Significantiecodes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

De verwarringsmatrices voor de trainings- en testdata staan weergegeven in Tabel 14.

Tabel 14 *Verwarringsmatrices voor trainings- en testdata; subset zandgrond van dataset 1 & 2.*

TRAINING			TEST		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling ↓	**0**	**1**
0	52	15	**0**	15	2
1	29	173	**1**	6	45

3.1.6 Subset dataset 1 & 2 van percelen op kleigrond

Het geselecteerde meest adequate model was model 101 (zie Bijlage 1):

ritnaald_01 ~ **xgras10jaar** + **xmais10jaar** + x**graan10jaar:xmais10jaar** + **grasgraan_1TRUE**

Hierbij waren alleen bovenstaande vetgedrukte verklarende parameters significant (Tabel 15). In dit model wordt de binaire variabele van aan- of afwezigheid van ritnaaldaantasting in knollen van een perceel het best verklaard door, het aantal voorgaande jaren waarop er gras is geteeld, het aantal voorgaande jaren waarop er mais is geteeld, een interactie tussen het aantal voorgaande jaren waarop er graan en het aantal jaar waarop er mais is geteeld en de binaire parameter voor aan- of afwezigheid van gras of graan in het voorgaande teeltjaar.

Tabel 15 *Samenvatting van het meest adequate model (model 101) bij de subset uit dataset 1 & 2 van percelen op kleigrond.*

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Signif.**
(Intercept)	0.18745	0.02965	6.32	5.0e-10	***
xgras10jaar	0.05029	0.00602	8.36	4.4e-16	***
grasgraan_1TRUE	0.16786	0.03836	4.38	1.4e-05	***
xmais10jaar	0.04208	0.02012	2.09	0.037	*
xgraan10jaar:xmais10jaar	-0.02880	0.01275	-2.26	0.024	*

De verwarringsmatrices voor de training en testdata staan weergegeven in Tabel 16.

Tabel 16 *Verwarringsmatrices voor trainings- en testdata; subset kleigrond van dataset 1 & 2.*

TRAINING			**TEST**		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling. ↓	**0**	**1**
0	319	138	**0**	70	38
1	49	108	**1**	15	29

3.2 Boosted regression trees

3.2.1 Gecombineerde dataset (dataset 1, 2 & 3)

Het beste model voor de gecombineerde dataset had een nauwkeurigheid van 0.73 op de training dataset en 0.69 op de testdataset. De verwarringsmatrices staan in Tabel 17, de meegenomen variabelen staan in Tabel *18.* De diepte van de gebruikte bomen was 5.

Tabel 17 *De verwarringmatrices voor het regression trees model met de hoogste nauwkeurigheid op de gecombineerde dataset.*

TRAINING			**TEST**		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling. ↓	**0**	**1**
0	336	92	**0**	72	31
1	84	693	**1**	75	164

Tabel 18 *Meegenomen variabelen voor de gecombineerde dataset en de volgorde van belangrijkheid.*

Rang	Variabele
1	xgroenbemester
2	xgraan10jaar
3	Xgrasgraanmais10jaar
4	CaC03(0, 5)BOFEK2012
5	BOFEK2012
6	Gras_1

3.2.2 Dataset 1

Het beste model voor de gecombineerde dataset had een nauwkeurigheid van 0.83 op de training dataset en 0.81 op de testdataset. De verwarringsmatrices voor de trainings- en testdata voor het beste presterende boosted regression trees staan weergegeven in Tabel 19.

Tabel 19 *Verwarringsmatrices voor trainings- en testdata voor dataset 1.*

TRAINING			**TEST**		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling ↓	**0**	**1**
0	263	65	**0**	61	18
1	33	233	**1**	9	57

Bij het maken van op bomen gebaseerde modellen kan worden vastgesteld hoe belangrijk een parameter is door de nauwkeurigheid te vergelijken als je eenzelfde model zonder de desbetreffende parameter mee te nemen. Hoe groter het verschil in nauwkeurigheid tussen beide modellen hoe belangrijker de parameter. In scikit-learn wordt deze bijdrage uitgedrukt als de *feature importance* en deze waarde wordt genormaliseerd, zodat de som van alle waarden gelijk is aan 1. Een hogere waarde duidt erop dat een variabele een grotere bijdrage levert aan de voorspellingen van het model. Hierbij dient te worden opgemerkt dat de *feature importance* geen uitspraken doet over causaliteit of de richting van het effect van een variabele. De meegenomen parameters en hun feature importances staan in Tabel 20.

Tabel 20 *De meegenomen parameters voor dataset 1, de volgorde van belangrijkheid en relatieve bijdrage (Feature importance) van de voorspellende variabelen.*

Rang	Variabele	Feature importance
1	xgrasgraanmais10jaar	0.56
2	pH-KCl(15, 30)BOFEK2012	0.16
3	BOFEK2012	0.08
4	xgraan10jaar	0.05
5	Xgroenbemester	0.04
6	CaC03(0, 5)BOFEK2012	0.03
7	pH-KCl(0, 5)BOFEK2012	0.03
8	pH-KCl(5, 15)BOFEK2012	0.02
9	M50(0, 5)BOFEK2012	0.01
10	gras_1	0.01

3.2.3 Dataset 2

Het beste model voor de gecombineerde dataset had een nauwkeurigheid van 0.76 op de training dataset en 0.69 op de testdataset. De verwarringsmatrices staan in Tabel 21, de meegenomen variabelen staan in Tabel 22. De diepte van de gebruikte bomen was 5.

Tabel 21 *De verwarringmatrices voor het regression trees model met de hoogste nauwkeurigheid op dataset 2.*

TRAINING			**TEST**		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling. ↓	**0**	**1**
0	148	58	**0**	30	11
1	20	131	**1**	19	24

Tabel 22 *Meegenomen variabelen voor dataset 2 en de volgorde van belangrijkheid.*

Rang	**Variabele**
1	xgroenbemester
2	xgraan10jaar
3	Xgrasgraanmais10jaar
4	CaC03(0, 5)BOFEK2012
5	BOFEK2012
6	Gras_1

3.2.4 Dataset 3

Het beste model voor dataset 3 had een nauwkeurigheid van 0.69, en er werd vrijwel overal wel schade voorspelt. De verwarringsmatrices staan in Tabel 23, de meegenomen variabelen staan in Tabel 24. De diepte van de gebruikte bomen was 5.

Tabel 23 *De verwarringmatrices voor het regression trees model met de hoogste nauwkeurigheid op de dataset 3.*

TRAINING			**TEST**		
	Observatie →			Observatie →	
Voorspelling ↓	**0**	**1**	Voorspelling. ↓	**0**	**1**
0	15	5	**0**	0	3
1	5	311	**1**	36	73

Tabel 24 *Meegenomen variabelen voor dataset 3 en de volgorde van belangrijkheid.*

Rang	**Variabele**
1	xgroenbemester
2	xgraan10jaar
3	Xgrasgraanmais10jaar
4	CaC03(0, 5)BOFEK2012
5	BOFEK2012
6	Gras_1

3.3 Vergelijken lineair model met boosted regression trees

De vergelijking tussen beide model benaderingen is uitgevoerd voor dataset 1, omdat deze de beste kwaliteit data had van ritnaaldaantasting. In Tabel 25 staan de prestaties van het meest adequate model van de lineaire modelbenadering en het beste presterende model van de random-forest benadering.

Tabel 25 *Overzicht van de prestaties van beide model benaderingen op de testdata van dataset 1.*

	Balanced accuracy[1]	**Precision[2]**	**Sensitivity[3]**	**F1-score[4]**
Lineair model (model 89)	0,836	0,892	0,773	0,828
Boosted regression trees	0,816	0,864	0,760	0,809

[1]Balance accuracy: gemiddelde van het correct herkennen van positieve én negatieve gevallen, gecorrigeerd voor ongebalanceerde klassen.

[2]Precision: van alle percelen die als "positief" worden voorspeld, welk deel daadwerkelijk ritnaaldaantasting heeft.

[3]Sensitivity: van alle percelen met ritnaaldaantasting, welk deel correct wordt geïdentificeerd door het model. 1 minus de sensitiviteit is het aantal fout-negatieve voorspellingen.

[4]F1-score: samengestelde maat die precision en sensitivity combineert tot één getal.

Het lineaire model scoort op alle prestatie-indicatoren iets hoger dan het boosted regression tree en heeft daarmee een iets betere voorspellende waarde. Het percentage fout-negatieven (het model voorspelt geen schade terwijl deze er wel is) was echter nog erg hoog: 22,7% (sensitivity 0,773). Het lineaire model is daarmee het minst slechte model in een vergelijking tussen deze twee benaderingen.

4 Conclusie en aanbevelingen

4.1 Vergelijking lineair model versus boosted regression tree

Uit een eerste analyse van de drie datasets kwam al snel naar voren dat er grote verschillen waren in de mate waarmee de data door een model (lineair of random-forest) beschreven kon worden, wat kon worden toegeschreven aan een verschil in de manier van beoordelen van ritnaaldaantasting aan de knollen. Dataset 1 liet zich het beste door een model beschrijven en werd daarom gebruikt voor een vergelijking tussen het lineaire model en boosted regression tree.

Beide modelbenaderingen zijn eerst gefit of getraind op dezelfde subset van trainingsdata en vervolgens gevalideerd op dezelfde testdata. De modelprestatie van beide modelbenaderingen lag erg dicht bij elkaar met een gering betere prestatie van het lineaire model. In het best presterende boosted regression tree kwamen dezelfde parameters terug als in het meest adequate lineaire model, maar ook een aantal extra parameters zoals pH op verschillende bodemdiepten, aantal jaar dat er een groenbemester heeft gestaan en het Bofek getal voor bodemtype. In de modelselectie van het lineaire model gaf het aantal voorgaande jaren met groenbemester ook een lichte modelverbetering ten op zichtte van een model zonder deze parameter (vergelijk AIC model 22 en 27, in Bijlage 1). Deze verbetering was echter marginaal ten opzichte van de effecten van andere parameters en hield uiteindelijk geen stand in de modelselectie. Dat wil zeggen dat het aantal voorgaande jaren met groenbemester niet significant bijdroeg aan verbetering van de modelprestatie.

4.2 Splitsing dataset in zand- en kleipercelen

Het meest adequate lineaire model had nog een percentage foutnegatief van 22,7% op de dataset 1, terwijl we vanuit landbouwkundig oogpunt pas tevreden zouden zijn met een percentage fout negatief onder de 5%. Bij nadere inspectie viel op dat met name de voorspellingen van ritnaaldaantasting voor percelen op kleigrond niet overeenkwamen met de werkelijk gevonden aantasting. Vervolgens zijn alle afzonderlijke datasets gesplitst in percelen op zandgrond en percelen op kleigrond. Een deel van de percelen die zich niet in zand of kleigrond lieten classificeren werd hierbij buiten beschouwing gelaten. De modelprestaties voor de subset van zandpercelen van dataset 1 & 2 lieten zich veel beter door een model beschrijven dan de subset van kleipercelen (resultaten niet weergegeven). De modelprestaties op de subset van zandpercelen van dataset 3 waren net als op de totale dataset 3 slecht, met een aangepaste R^2 rond de nul. Vervolgens is er een modelselectie uitgevoerd op de subset van zandpercelen en de subset van kleipercelen voor de gecombineerde datasets 1 & 2.

4.3 Interpretatie van de gevonden verklarende parameters

Het meest adequate model voor de subset van zandpercelen in de gecombineerde dataset 1 & 2 (Model 98) had nog drie parameters: $CaCO_3$-gehalte, aantal voorgaande jaren met gras, graan of mais en een parameter voor de aan- of afwezigheid van gras of graan in de voorvrucht. Dat $CaCO_3$ overeind bleef in de modelvergelijking is opmerking, omdat in de meeste zandgronden geen koolzurekalk aanwezig is. Wat hier echter meespeelt is dat op grondtypen met een BOFEK2012 classificatie 323, 324, 325 wel kalk in de bodem zit in de range van 1-8. Dit betreft de uiterst fijn en zeer fijn zandige (zee-)zandgronden, matige fijn zandige (deels opgestoven mariene en fluviatiele) zandgronden en mariene en fluviatiele zandgronden met een kleidek. Op percelen met deze bodemtypen werden nagenoeg geen ritnaaldaantasting aangetroffen, wat de negatieve coëfficiënt voor deze parameter verklaart.

Dat een voorgeschiedenis met veel gras in het bouwplan de kans op ritnaaldaantasting vergroot was – op basis van kennis over de levenscyclus van kniptorren/ritnaalden – bekend en komt duidelijk uit de modelanalyses in deze studie. Het aantal jaar graan en mais in de voorgeschiedenis draagt ook bij aan het verhogen van risico op ritnaaldaantasting, maar in mindere mate dan dat voor gras (vergelijk model 4, 5 en 6 of 16, 17 en 18in Bijlage 1). Voor percelen op zandgrond was de coëfficiënt bij deze parameter (grasgraanmais10jaar) bijna 3 keer zo groot als voor percelen op kleigrond, wat laat zien dat een graslandvoorgeschiedenis op zandgrond belangrijker is om rekening mee te houden dan op kleigrond. De positieve coëfficiënt voor de aanwezigheid van gras of graan als voorvrucht geeft aan dat het risico op ritnaaldaantasting het grootst is in een teelt direct volgend op gras of graan.

In het meest adequate model voor de subset van kleipercelen voor de gecombineerde dataset 1 & 2 (model 101) zit ook een significante parameter voor het aantal jaar dat er mais op het perceel heeft gestaan (xmais10jaar) en een significante interactieterm voor het aantal jaar dat er mais en het aantal jaar dat er graan op het perceel heeft gestaan (xgraan10jaar:xmais10jaar). Een biologische interpretatie van deze effecten ligt niet volledig voor de hand. Van graan is bekend dat het aantrekkelijk is voor eiafzet en dus een stimulans is, maar voor mais is dat minder duidelijk. In dat opzicht valt het op dat het aantal jaren graan op het perceel (xgraan10jaar) niet voorkomt in de lijst significante parameters. Mogelijk zijn sommige of al deze factoren een indirecte indicator van een extensiever akkerbouwrotatie, of van uitruilen van grond met veehouders inclusief grasland, aspecten die op zichzelf invloed kunnen hebben op ritnaaldaantasting.

De M50 parameter (textuur) in het meest adequate model voor dataset 1 (model 89) lijkt met name van invloed te zijn geweest op het onderscheiden van klei- van zandpercelen. Op zandpercelen is de M50 hoger dan op kleipercelen. Bij het verder optimaliseren van de subsets van zand- en kleipercelen verdwijnt deze parameter uit de selectie en doet er dus niet meer toe.

In het meeste adequate model voor dataset 2 (model 106) heeft de teelt van gras of graan 2 of 3 jaar geleden een risicoverlagend effect (negatieve coëfficiënten voor grasgraan_2 en grasgraan_3) terwijl de teelt van gras of graan in het voorgaande jaar (positieve coëfficiënt voor grasgraan_1) een risicoverhogend effect heeft (Tabel 7). Dit resultaat is tegengesteld aan de verwachting dat een geschiedenis met een monocotyl het risico op ritnaaldaantasting vergroot. Nadat echter de dataset is gesplitst in klei- en zandpercelen zijn deze effecten van grasgraan_2 en grasgraan_3 niet meer significant (resultaten niet weergegeven) en is er hoogstwaarschijnlijk geen biologische verklaring voor dit gevonden resultaat.

Het meest adequate model voor dataset 3 had een voorspellende waarde van nul en een interpretatie van de gevonden parameters is daarmee irrelevant. Dat het model geen voorspellende waarde heeft is toe te schrijven aan de wijze van beoordelen; daarbij is niet alleen ritnaaldaantasting gescoord, maar ook andere aantastingen die uitval in de verwerking opleveren.

In deze modelbenadering is gewerkt met de binaire variabele wel of geen ritnaaldaantasting per partij (dus zodra er aantasting werd gevonden werd de partij op 100% aangetast gesteld), terwijl in de dataset ook de mate van aantasting was opgenomen als percentage knollen met ritnaaldaantasting. Bij eerste verkenningen om de data met een lineair model of boosted regression tree te beschrijven werd al snel duidelijk dat mate van aantasting veel slechter te voorspellen was dan de binaire aan- of afwezigheid van ritnaaldaantasting.

4.4 Toepassing

Met de gekozen modelbenadering in deze studie konden we voor zandpercelen een adequaat model opstellen met 4,3% fout-negatieven en 28,6% fout-positieven. De voorgeschiedenis met grassen, graan of mais op het perceel was de belangrijkste voorspeller, samen met of er in voorgaande jaar gras of graan op het perceel stond. Zandpercelen waren geclassificeerd op basis van de BOFEK2012 bodemkartering. Zandbodems op fluviele of maritieme afzettingen gaven minder risico op ritnaaldaantasting dan bodems op andere zandgronden. Het model corrigeert hiervoor door middel van het $CaCO_3$-gehalte op deze bodems. De gevonden coëfficiënten bij deze parameters (Tabel 13) kunnen worden gebruikt om kans op ritnaaldaantasting in aardappelen te voorspellen volgens:

$0{,}409 - 0{,}047*\alpha + 0{,}045*\beta + 0{,}14*\gamma \geq 0{,}5$ → ritnaaldaantasting in aardappel op zandgrond
$0{,}409 - 0{,}047*\alpha + 0{,}045*\beta + 0{,}14*\gamma < 0{,}5$ → geen ritnaaldaantasting in aardappel op zandgrond

Waarbij α is het modaal kalkgehalte ($CaCO_3$), uitgedrukt als massaprocent van de droge grond op te halen uit de BOFEK2012 dataset, β het aantal voorgaande jaren dat er een gras, graan of mais op het perceel heeft gestaan tot 10 jaar terug en γ met de waarde 1 als de voorvrucht een gras of graan was en 0 voor alle andere voorvruchten.

In woorden uitgedrukt komt dit er op neer dat op zandgronden zonder kalk het model ritnaaldaantasting in aardappelen voorspelt als er in voorgaande jaar gras of graan op het perceel heeft gestaan en zo niet dan voorspelt het model ritnaaldaantasting als er de afgelopen 10 jaar twee of meer keer een gras of graan op het perceel is geteeld.

Voor kleigronden liet de data zich niet goed beschrijven door een model met de in deze studie gebruikte verklarende variabelen. Het vereist nadere analyse om te achterhalen wat betere verklarende parameters zijn voor ritnaaldaantasting op kleigrond. Parameters die van invloed kunnen zijn, maar in deze studie niet zijn meegenomen zijn het type en periode van grondbewerking, werkelijk gemeten hoeveelheid organische stof in de bodem en aanwezigheid van grasranden of grasland in de omgeving. Ook zou het interesant zijn om te zien of er op kleigrond andere soorten ritnaalden dominant zijn dan op zandgrond, of dat het gedrag anders is op verschillende grondsoorten, en of dat kan helpen begrijpen welke factoren mee genomen moeten worden in een voorspellend model voor kleigronden.

Literatuur

Elith, J., Leathwick, J., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*, 802-813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Ferrari, R., Tassini, C., Furlan, L., Fracasso, F., Sartori, E., Codato, F., Oddino, B. (2015). La gestione degli elateridi con i fondi mutualistici. Terra e Vita 15, 42-45.

Jung, J., Racca, P., Schmitt, J. and Kleinhenz, B. (2014), SIMAGRIO-W: Development of a prediction model for wireworms in relation to soil moisture, temperature and type. J. Appl. Entomol., 138: 183-194. https://doi.org/10.1111/jen.12021

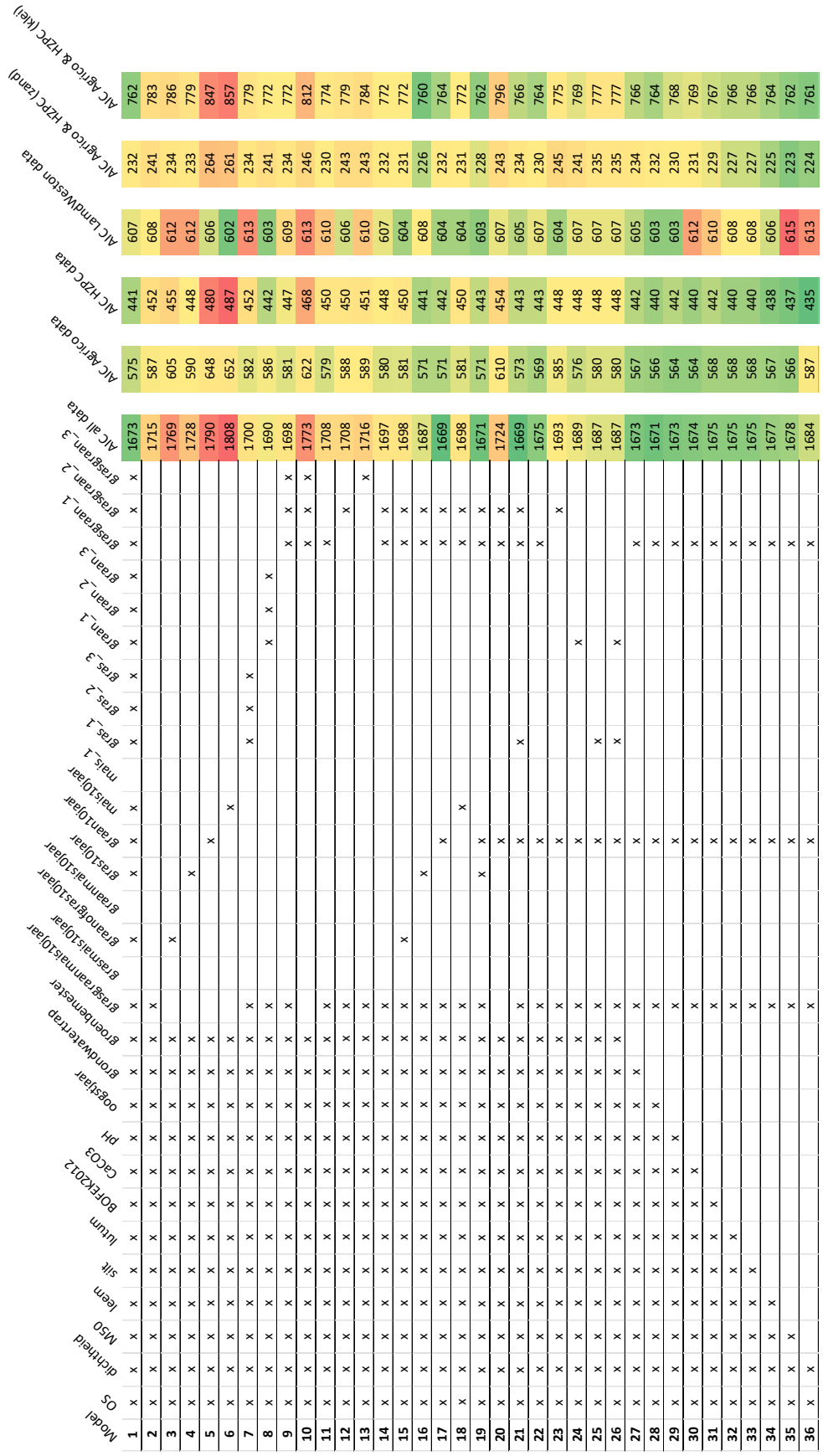
Poggi, S., Le Cointe, R., Riou, J., Larroudé, P., Thibord, J. et al.. Relative influence of climate and agroenvironmental factors on wireworm damage risk in maize crops. Journal of Pest Science, 2018, 91 (2), pp.585-599. ⟨10.1007/s10340-018-0951-7⟩. ⟨hal-01840983⟩

Roche, J., Plantegenest, M., Larroudé, P., Thibord, J., Le Cointe, R., Poggi, S. A decision support system based on Bayesian modelling for pest management: Application to wireworm risk assessment in maize fields. Smart Agricultural Technology (4) 2023, 100162, ISSN 2772-3755, https://doi.org/10.1016/j.atech.2022.100162.

Van Rozen, K., Huiting, H. F., Allema, A. B., van Tol, R. W. H. M., & Postma, J. (2021). Beheersing van bodemplaaginsecten via bodemgezondheidsmaatregelen: Een overzicht van de beschikbare kennis voor een selectie van akkerbouwgewassen met hun bijbehorende bodemplaaginsecten. Stichting Wageningen Research, Wageningen Plant Research, Business unit Open Teelten; Rapport No. WPR-1061. https://doi.org/10.18174/538253

Wechselberger, K., Racca, P., & Hann, P. (2018). ElatPro - Spotting the needle in a haystack: Predicting wireworm activity in top soil for integrated pest management in arable crops - ERA-NET C-IPM. *Impact, 6*, 45–47. https://doi.org/10.21820/23987073.2018.6.45

Bijlage 1 Geteste lineaire regressiemodellen



Model	OS	dichtheid	M50	leem	silt	lutum	BOFEK2012	CaCO3	pH	oogstjaar	grondwatertrap	groenbemester	grasgraanmais10jaar	grasmais10jaar	graanofgras10jaar	graanmais10jaar	gras10jaar	graan10jaar	mais10jaar	mais_1	gras_1	gras_2	gras_3	graan_1	graan_2	graan_3	grasgraan_1	grasgraan_2	grasgraan_3	AIC all data	AIC Agrico data	AIC HZPC data	AIC LamdWeston data	AIC Agrico & HZPC (zand)	AIC Agrico & HZPC (klei)
37	x		x									x					x								x				1684	565	439	613	223	765
38			x									x					x								x				1687	567	438	611	223	763
39		x	x									x					x								x				1688	567	435	613	224	763
40			x	x								x					x								x				1688	568	437	607	223	761
41			x		x							x					x								x				1689	568	438	606	222	762
42			x			x						x					x								x				1687	568	436	609	225	761
43			x				x					x					x								x				1689	567	440	609	225	765
44			x					x				x					x								x				1677	556	440	612	223	765
45			x						x			x					x								x				1687	556	439	601	221	764
46			x					x								x	x	x											1700	568	447	613	229	773
47	x	x	x	x	x	x	x	x	x	x	x		x														x		1721	590	458	606	241	793
48	x	x	x	x	x	x	x	x	x	x			x														x		1719	588	456	604	239	792
49	x	x	x	x	x	x	x	x	x				x														x		1719	588	456	604	239	792
50			x										x														x		1732	589	454	612	229	792
51								x					x														x		1725	587	454	612	229	793
52			x					x					x														x		1724	580	456	614	230	794
53			x						x				x														x		1733	579	454	606	228	793
54			x				x						x														x		1734	588	456	611	231	794
55		x	x										x														x		1734	588	451	614	230	792
56	x		x										x														x		1728	587	454	614	229	794
57	x	x											x														x		1736	621	451	613	232	791
58	x												x														x		1753	627	453	612	234	794
59		x											x														x		1773	640	449	612	241	792
60		x					x						x														x		1750	601	451	610	238	793
61		x						x					x														x		1727	589	451	614	231	793
62			x				x	x	x			x	x				x			x									1692	570	448	602	224	776
63																													1693	572	451	605	226	776
64																													1679	557	442	613	225	764
65			x					x						x			x								x				1677	556	440	612	223	765
66			x					x								x	x	x							x				1679	557	442	613	225	764
67			x					x					x												x				1712	569	456	614	221	791
68			x					x								x	x	x							x				1681	559	444	614	227	766
69			x					x								x	x	x							x				1676	553	441	614	226	761
70			x					x								x	x	x							x				1679	559	444	614	227	766
71			x					x								x	x	x							x				1679	558	447	618	230	765
72			x					x								x	x	x							x				1718	601	446	613	242	795

Model
OS
dichtheid
M50
leem
slib
lutum
BOFEK2012
CaCO3
pH
oogsteaar
bronduwetertrap
broendbemeester
braSgraanmais10jaar
braSmais10jaar
braanofgras10jaar
braanmais10jaar
braS10jaar
braan10jaar
mais10jaar
mais_1
bras_1
bras_2
bras_3
braan_1
braan_2
braan_3
braSgraan_1
braSgraan_2
braSgraan_3
AIC all data
AIC Agrico data
AIC HZPC data
AIC landWestion data
AIC Agrico & HZPC (rand)
AIC Agrico & HZPC (kel)

Bijlage 2 Identificatiekaart ritnaaldschade

Ritnaaldaantasting in aardappel – herkennen en onderscheiden

GRONDIGE AANPAK Bodemplagen

Recent

Ouder, eerder aangetast

Rhizoc. & ritnaalden
Drycore
Paar mm diep
Ø 3-6 mm
Vliesje aan de rand
Soms een propje

Typische ritnaaldschade: smalle gangetjes, meestal kleine diameter.

Soms is het beeld anders en dan is een indruk over een partij nodig.

Symptomen variëren:
- Vroege of late aantasting
- Raseffecten
- Veel of weinig ritnaalden
- Weer tijdens de teelt
- Veel gangen bij elkaar worden een holte

Ritnaaldaantasting niet verwarren

Aardrupsen

Engerlingen

Meestal grotere diameter schil

Grotere, holle ruimtes

Naaktslakken

To explore the potential of nature to improve the quality of life

Correspondentieadres voor dit rapport:
Postbus 430
8200 AK Lelystad
T 0320 291423
wur.nl/plant-research

Rapport WPR-OT-1241

De missie van Wageningen University & Research (WUR) is 'To explore the potential of nature to improve the quality of life'. WUR brengt onderwijs, fundamenteel en toegepast onderzoek en innovatie samen op het snijvlak van natuur, mens en voedsel.

WUR werkt aan oplossingen voor mondiale vraagstukken rond biodiversiteit, klimaat, duurzame voedselsystemen en gezondheid. De kracht van WUR ligt in de integrale aanpak: het verbinden van disciplines, het combineren van wetenschappelijke diepgang met praktische relevantie en het samenbrengen van wetenschap, beleid, bedrijfsleven en samenleving. Door de samenwerking tussen Wageningen University en Wageningen Research draagt WUR bij aan kennis, innovatie en impact, en aan de transities die nodig zijn voor een wereld in balans met de natuur.