

WEBOOST

ProRail

KLACHTEN OVER TRILLINGEN

ONDERZOEK NAAR BEPALENDE FACTOREN

COLOFON

Auteur(s) Pieter Boon
Marjolein Hordijk
Thomas Vos

Projectcode WBD2021-054

Versienummer 2.0

Datum 20 april 2022

Status Definitief

Opdrachtgever ProRail B.V.



© We-Boost 2022

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

MANAGEMENTSAMENVATTING

Aanleiding

Binnen de Innovatie-agenda Bronaanpak Spoortrillingen (IBS) werkt ProRail aan een innovatie-agenda spoortrillingen. Binnen deze agenda worden onderzoeksprojecten uitgevoerd om de overlast van treintrillingen op de omgeving te verminderen. Veel problemen zijn bekend, denk aan problemen met harde punten (overwegen, kunstwerken) en rond wissels, maar zijn die ook vooral bepalend voor de klachten die ProRail in de loop der jaren heeft ontvangen? M.a.w., is het mogelijk om de innovatie-agenda te prioriteren o.b.v. de klachten uit het verleden, zodat de meest urgente problemen het eerst worden aangepakt?

In het voorliggende rapport hebben we onderzocht wat de belangrijkste zogenaamde 'drivers' zijn voor klachten over trillingen over de afgelopen jaren. Met andere woorden: zijn er specifieke situaties of omstandigheden aan te wijzen waarbij er meer overlast is van trillingen? Het antwoord op die vraag helpt om te prioriteren in een aanpak om de trillingen te verminderen.

Met behulp van een uitgebreide data-analyse met behulp van een zogenaamd Random Forest-model hebben we onderzocht welke parameters wel en welke geen invloed hebben op het aantal klachten.

Conclusies

Conclusie van het onderzoek is dat er inderdaad situaties zijn aan te wijzen die vaker dan verwacht tot klachten leiden. Vooral veranderingen (denk aan bovenbouwvernieuwing, een nieuwe overweg, andere treinen of een hogere rijdsnelheid) leiden vaak tot klachten. Dit kan komen doordat de trillingen toenemen, maar evengoed ook doordat de trillingen een andere frequentie-inhoud krijgen en dus anders worden ervaren. Niet onbelangrijk hierbij is ook de rol van de media en actiegroepen, deze dragen in belangrijke mate bij aan het melden van klachten, zo blijkt uit onze analyse.

Verder valt op dat ook sociologische aspecten een rol kunnen spelen in het melden van klachten: zo wordt er meer geklaagd uit wijken met veel eenpersoonshuishoudens, weinig gezinnen met kinderen en veel

MANAGEMENTSAMENVATTING

65-plussers. Ook zijn er meer klachten uit dichtbevolkte gebieden.

De belangrijkste invloedsparameters zijn hieronder weergegeven. Per parameter lichten we toe hoe sterk de invloed is op het aantal klachten.

Parameter	Invloed op klachten	Toelichting
Afstand tot spoor		Veel meer klachten op kortere afstand van het spoor. Buiten 75 meter van spoor geen sterke invloed.
Hoogteverschil spoor en maaiveld		Veel meer klachten bij afnemend hoogteverschil. Meeste klachten bij spoor op maaiveld zonder spoorloot.
Afstand tot ES-las		Veel meer klachten bij kortere afstand tot (vooral oudere) ES-las. Buiten 75 meter van ES-las geen invloed.
Percentage private huurwoningen		Veel meer klachten uit wijken met weinig private huurwoningen (en dus meer koop- en corporatiewoningen).
Gewicht (statische aslast)		Meer klachten bij zwaardere treinpassages (vooral zwaarste trein per dag/week).
Aantal passages met hoge aslast		Meer klachten bij meer zware treinpassages (vooral zwaarbeladen goederentreinen).
Pieklast (dynamische aslast)		Meer klachten bij treinpassages met hoge dynamische aslast.
Aantal treinpassages		Meer klachten bij groter aantal treinpassages, vooral bij meer goederentreinpassages in de nacht.
Gebouwhoogte		Meer klachten uit hogere gebouwen.
RMS _{low} (wielonrondheid)		Meer klachten bij treinpassages met hoge wielonrondheid.
Tonnage per week		Meer klachten van trajecten met hoog tonnage (veel en/of zware treinen).
Stijfheid toplaag bodem		Meer klachten op stijve bodems (zand, oosten en zuiden van het land).
Afstand tot overweg		Meer klachten bij kortere afstand tot (niet-STRAIL) overweg. Buiten 75 meter van overweg geen invloed.
Bouwjaar woning		Meer klachten uit oudere woningen, vooral uit periode voor 1940.

MANAGEMENTSAMENVATTING

Wat kun je hiermee?

Op basis van de analyse van de klachten is een aantal adviezen geformuleerd m.b.t. het prioriteren van een aanpak om het aantal klachten te verminderen:

1. Focus bij maatregelen op gebieden met:
 - a. Bebouwing dicht op het spoor, en vooral op stijvere bodemtypes (zand).
 - b. Druk bereden, gemengde trajecten met veel goederentreinen op volle snelheid.
2. Doe nader onderzoek naar de gemelde klachten rond (vooral oudere) ES-lassen. Komen die door de ES-lassen, of door de combinatie met wissels en/of overwegen? En hoe kun je deze problematiek verminderen?
3. Doe onderzoek naar mogelijkheden om de dynamische pieklast en de RMS_{low} (indicator van wielonrondheid) te verminderen.
4. Koester spoorsloten als relatief goedkope maatregel tegen trillingen: er wordt significant minder geklaagd als er een spoorsloot tussen spoor en bebouwing ligt.

Tenslotte laat dit onderzoek (opnieuw) zien dat trillingen eigenlijk altijd frequentie-afhankelijk moeten worden beschouwd: veel klachten lijken te zijn gerelateerd aan verschuivingen in het frequentiespectrum van de trillingen, bijvoorbeeld door een hogere rijnsnelheid, andere treintypes of wijzigingen in de spoorconstructie. En niet eens perse door hogere trillingen zelf.

INHOUDSOPGAVE

Managementsamenvatting	3
1. Inleiding	7
2. Aanpak	8
3. Resultaten	10
4. Samenvatting	34
Bijlages met achtergrondinformatie	35



Rob Dammers, [CC BY 2.0](#), via Wikimedia Commons

1. INLEIDING

1.1 Aanleiding

Binnen de Innovatie-agenda Bronaanpak Spoortrillingen (IBS) werkt ProRail aan een innovatie-agenda spoortrillingen. Binnen deze agenda worden onderzoeksprojecten uitgevoerd om de overlast van treintrillingen op de omgeving te verminderen. Veel problemen zijn bekend, denk aan problemen met harde punten (overwegen, kunstwerken) en rond wissels, maar zijn die ook vooral bepalend voor de klachten die ProRail in de loop der jaren heeft ontvangen? M.a.w., is het mogelijk om de innovatie-agenda te prioriteren o.b.v. de klachten uit het verleden, zodat de meest urgente problemen het eerst worden aangepakt?

1.2 Doel

Doel van dit onderzoek is om te onderzoeken wat de belangrijkste zogenaamde ‘drivers’ zijn voor klachten over trillingen over de afgelopen jaren. Met andere woorden: zijn er specifieke situaties of omstandigheden aan te wijzen waarbij er meer overlast is van trillingen? Het antwoord op die vraag helpt om te prioriteren in een aanpak om de trillingen te verminderen.

1.3 Leeswijzer

Dit rapport is als volgt opgebouwd:

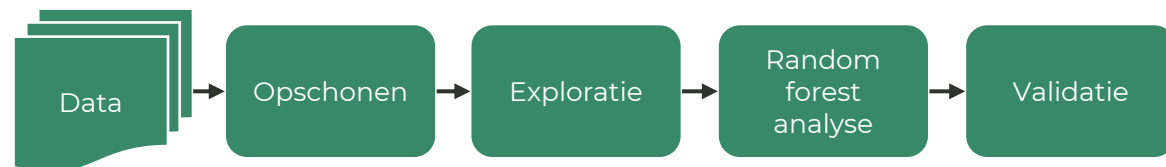
- In hoofdstuk 2 lichten we de aanpak toe
- In hoofdstuk 3 geven we de resultaten van het onderzoek
- We sluiten af met een beknopte samenvatting in hoofdstuk 4

Het rapport is zo grafisch mogelijk opgezet. Elke figuur is voorzien van een toelichting om het rapport zelfstandig leesbaar te houden. Bij het opzetten van dit rapport is beperkte voorkennis op het gebied van trillingen verondersteld.

2. AANPAK

2.1 Inleiding

In dit hoofdstuk beschrijven we onze aanpak. Omdat we onderbouwd uitspraken willen doen over wat nu wel en geen invloed heeft op klachten over trillingen, hebben we een data-analyse uitgevoerd waarbij we informatie hebben gekoppeld aan klachten uit de database van Publieksvoorlichting (PuVo) van ProRail. Vervolgens hebben we deze data opgeschoond, verrijkt en geanalyseerd op parameters die wel en geen invloed hebben op het ontstaan van de klachten. Deze aanpak is hieronder schematisch weergegeven.



2.2 Dataverzameling

Voor ons onderzoek hebben we zowel interne (ProRail) als externe openbare databronnen gebruikt. Op basis van eerdere onderzoeken en expert judgment hebben we alle mogelijk relevante gegevens verzameld en gekoppeld aan de klachten.

Hierbij hebben we gebruik gemaakt van een zogenaamde geocoder om de opgegeven adressen aan panden uit de Basisregistratie Adressen en Gebouwen (BAG) te koppelen. Op basis van de gevonden locatie is vervolgens de data van de klachten verrijkt met informatie over de omgeving. Dit betreft onder andere de bodemopbouw, gegevens over het spoor en de treinen over de buurtsamenstelling (van het CBS), zie het overzicht hieronder. Hierdoor kan voor elke klacht een breed beeld gevormd van mogelijke oorzaken.

Vervolgens is die aanpak herhaald voor ongeveer 40.000 random geselecteerde panden, verspreid over heel Nederland. Deze referentiedataset is gebruikt om in de random forest analyse te bepalen welke factoren van belang zijn bij het wel of niet ontstaan van klachten.

Klachten

- Bron: ProRail
- Omschrijving
- Adres

Gebouwen

- Bron: BAG3D
- Locatie
- Bouwjaar
- Afmetingen

Bodem

- Bron: BRO, AHN3, Openstreetmap
- Water
- Type bodem
- Stijfheid bodem
- Hoogteverschil

Treinbeeld

- Bron: ProRail BBK, Geluidregister
- Sporen
- Wissels
- ES-lassen
- Kunstwerken
- Snelheid
- Spoorconstructie
- Treinaantallen
- Treineigenschappen

Buurtcijfers

- Bron: CBS
- Leeftijden
- Huur en koop
- Huishoudens
- Bevolking

2. AANPAK

2.3 Exploratie

Na het opschonen en koppelen van de data, hebben we gebruik gemaakt van statistieken, correlaties en grafieken om grove trends in data te vangen, en bijvoorbeeld kaarten te maken die de geografische spreiding van de klachten weergeven. Deze fase helpt bij het achterhalen van globale trends. Resultaten hiervan hebben we opgenomen in Bijlage I van dit rapport.

2.4 Random Forest-analyse

Vervolgens hebben wij een data-analyse uitgevoerd op basis van een Random Forest (RF) model. Hiermee hebben wij bepaald welke parameters wel of niet relevant zijn bij het ontstaan van een klacht. Een RF-model is een machine learning techniek waarbij de invloed van parameters op elkaar wordt onderzocht met behulp van zogenaamde beslisbomen. Het mooie van een RF-model is dat de techniek in principe ongestuurd is (het RF-model gaat zelf op zoek naar relaties) en dat de resultaten goed interpreteerbaar zijn (geen black box uitkomsten). Zo zijn uitkomsten ook goed te valideren. Meer details zijn te vinden in Bijlage II.

2.5 Validatie

Tenslotte hebben we de uitkomsten van het RF-model gevalideerd door deze naast de uitkomsten uit de exploratiefase (zie paragraaf 2.3) te leggen en te beoordelen op expert judgment.

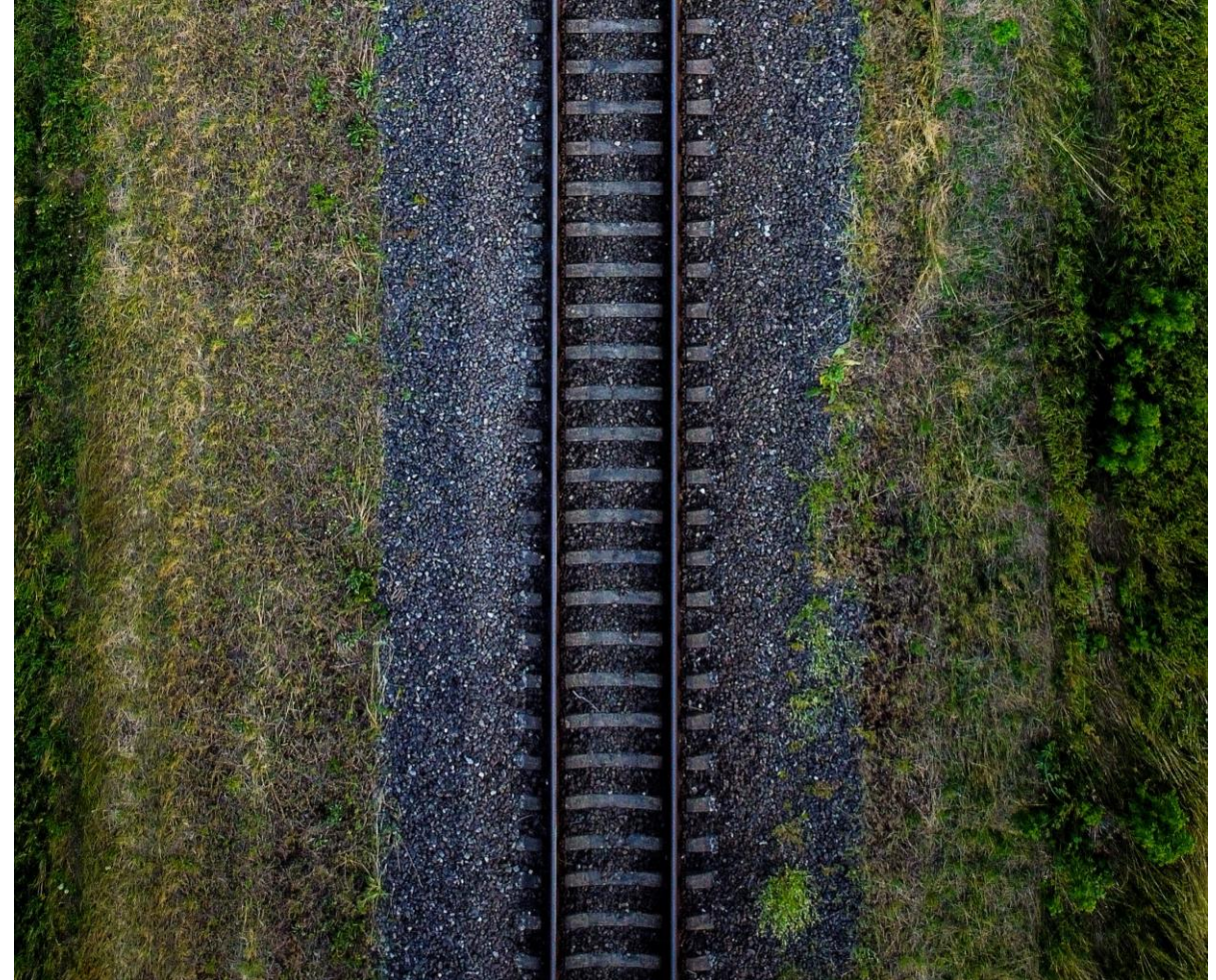
3. RESULTATEN

3.1 Inleiding

In dit hoofdstuk presenteren we de resultaten van ons onderzoek. We geven daarbij eerst algemene informatie over de klachten (denk aan geografische verdeling over Nederland en ontwikkeling over de jaren heen), daarna zoomen we in op de bepalende factoren of 'drivers' voor de klachten.

3.2 Exploratiefase

De volgende slides bevatten resultaten van de exploratiefase, waarin we algemene trends en uitkomsten uit de klachtenanalyse geven.



3. RESULTATEN

WAAR WORDT HET VAAKST OVER GEKLAAGD?



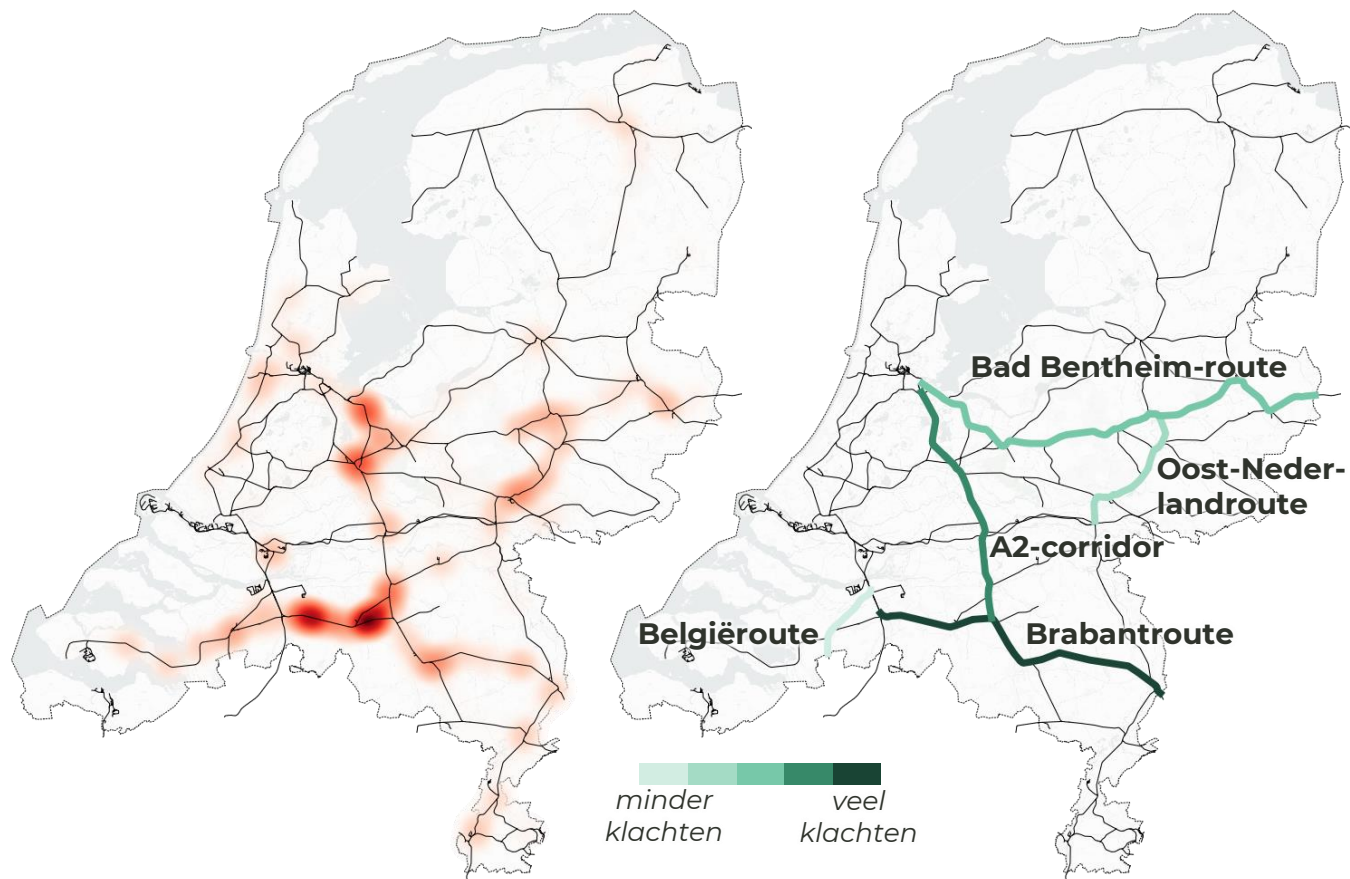
3. RESULTATEN

WANNEER WORDT HET MEEST GEKLAAGD?



3. RESULTATEN

VERDELING OVER HET LAND



Gemeente	Locatie	Klachten	Unieke klachten
Oisterwijk	Brabantroute	103	77
Rheden	Oost-Nederlandroute	858	49
Oosterhout	Brabantroute	95	48
Gooise Meren	Bad Bentheim-route	57	42
Tilburg	Brabantroute	81	42
Utrecht	A2-corridor	71	41
Gilze en Rijen	Brabantroute	43	30
Eindhoven	Brabantroute	59	29
Vught	A2-corridor	49	26
Roosendaal	Belgiëroute	35	25
Deventer	Bad Bentheim-route	20	20
's Hertogenbosch	A2-corridor	193	19

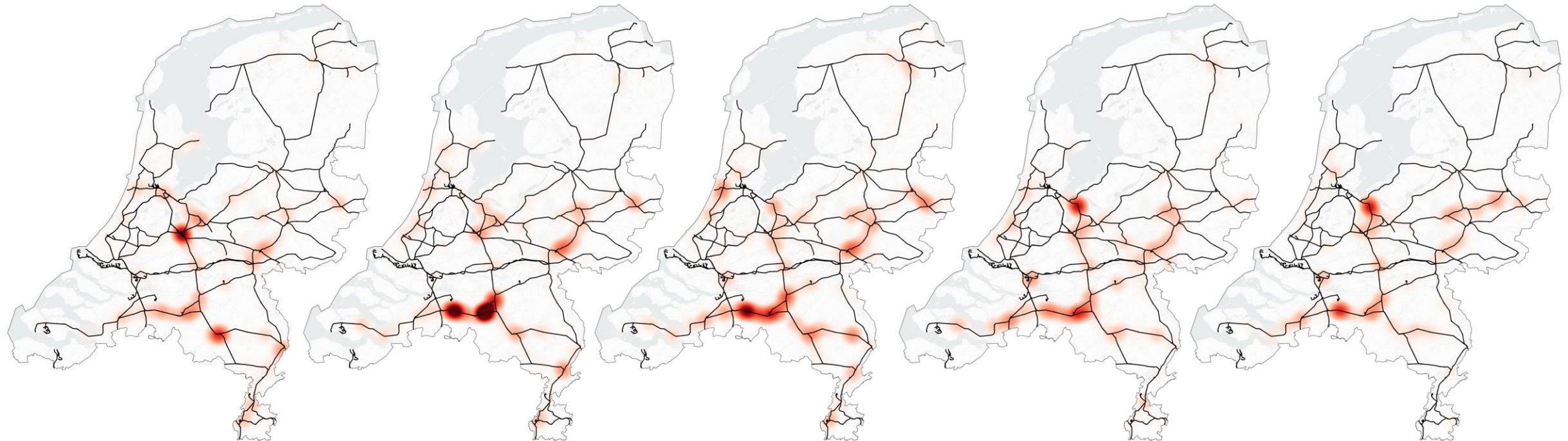
GECONCENTREERDE KLACHTEN

In 12 gemeenten:

- 36% van unieke klachten
- 60% van alle klachten

3. RESULTATEN

HOTSPOTS PER JAAR

**2017**

Hogere snelheid
goederentreinen Utrecht
Centraal

2018

Introductie TRAXX-locomotief
op Brabantroute

2019

Veel media-aandacht voor
TRAXX-locomotief op
Brabantroute

2020

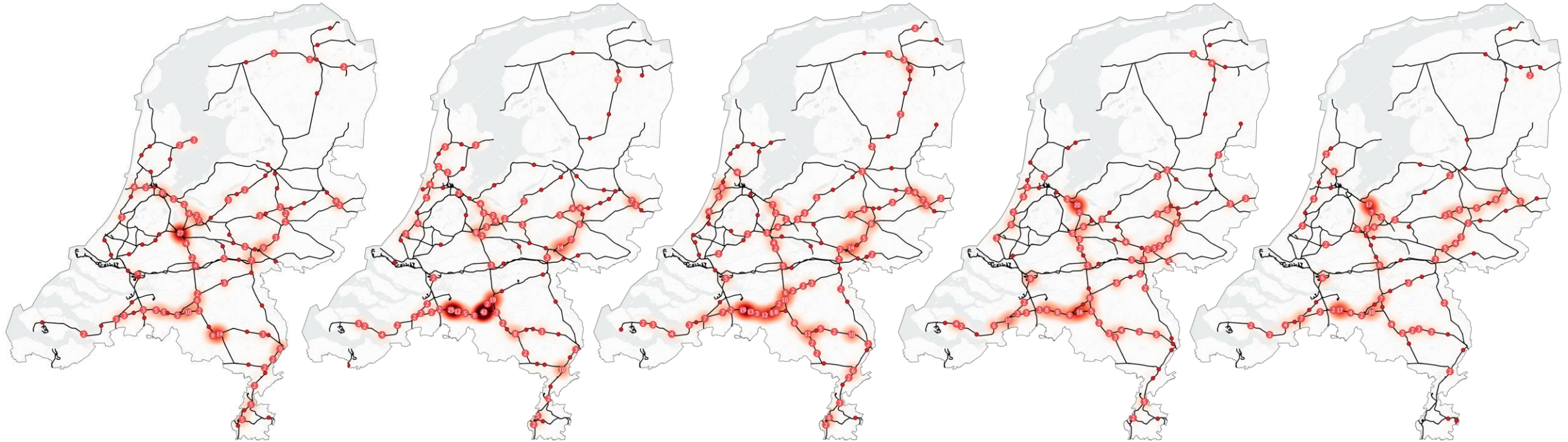
Snelheidsverhoging Naarden-
Bussum, minder klachten op
Brabantroute

2021

Media-aandacht voor
snelheidsverhoging Naarden-
Bussum, nog minder klachten
op Brabantroute

3. RESULTATEN

HOTSPOTS PER JAAR – MET AANTAL KLACHTEN

**2017**

Hogere snelheid
goederentreinen Utrecht
Centraal

2018

Introductie TRAXX-locomotief
op Brabantroute

2019

Veel media-aandacht voor
TRAXX-locomotief op
Brabantroute

2020

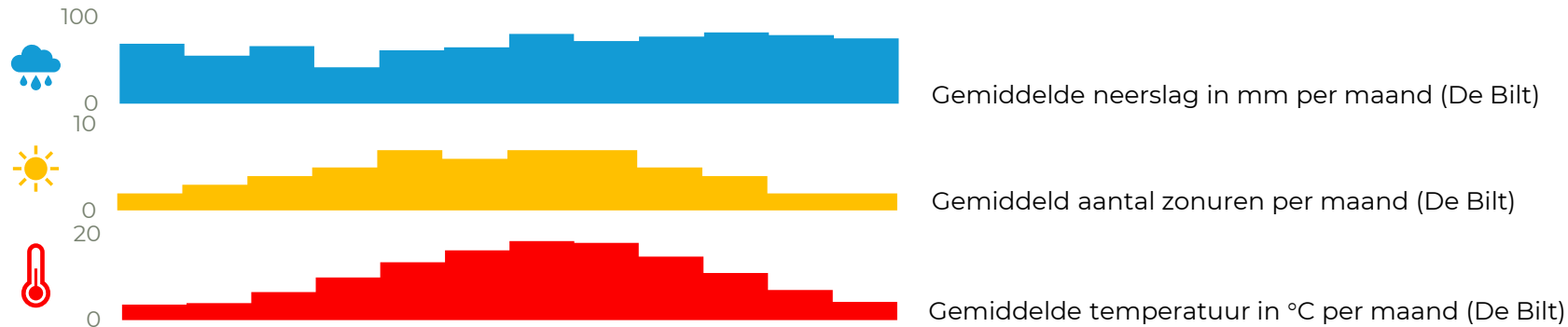
Snelheidsverhoging Naarden-
Bussum, minder klachten op
Brabantroute

2021

Media-aandacht voor
snelheidsverhoging Naarden-
Bussum, nog minder klachten
op Brabantroute

3. RESULTATEN

VERDELING OVER JAREN EN MAANDEN



ALLE KLACHTEN

- Stabilisatie na toename, maar laatste jaren zijn niet compleet.
- Duidelijk verschil tussen **winter en zomer** ontbreekt.

UNIEKE KLACHTEN

- In aantal unieke klachten is geen afname of stabilisatie zichtbaar.
- Geen duidelijke invloed van **temperatuur**, **neerslag** of andere **seizoenspatronen** zichtbaar.

3. RESULTATEN

3.3 Random Forest-analyse

Hierna volgen de resultaten van het Random Forest-model. We starten met een totaaloverzicht, waarin de meest dominante parameters worden weergegeven.

Vervolgens geven we per parameter met behulp van grafieken meer inzicht in de invloed van die parameter op het ontstaan van klachten. Die grafieken hebben we ook gebruikt om de resultaten te valideren en te duiden.



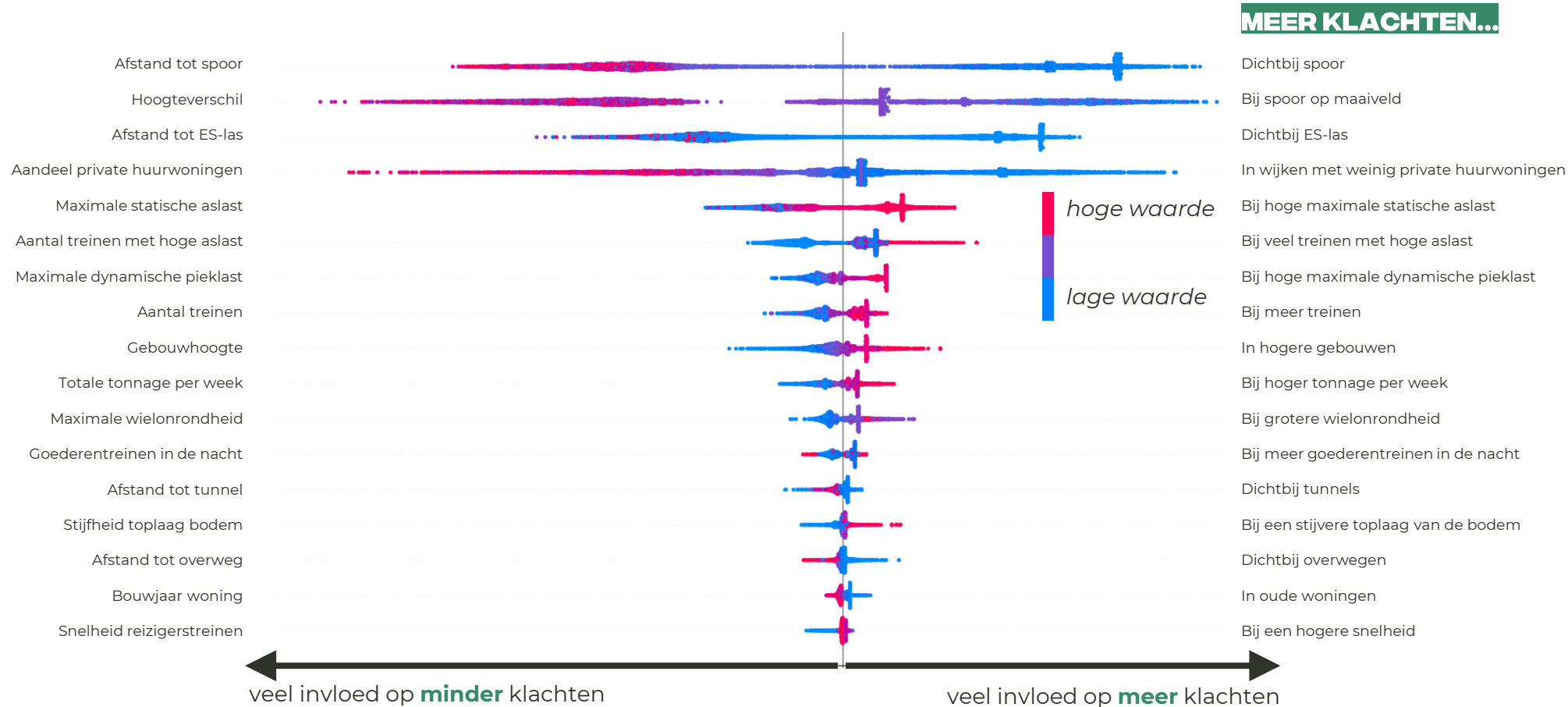
Steven Lek, [CC BY-SA 4.0](#), via Wikimedia Commons

3. RESULTATEN

RANDOM FOREST PARAMETER DOMINANTIE

Deze SHAP-plot geeft aan hoeveel invloed een parameter heeft op het ontstaan van klachten, en of de klachten toe- of afnemen als de parameter toe- of afneemt.

Vooraf **afstand tot het spoor**, **hoogte** van het spoor (t.o.v. maaiveld) en afstand tot **ES-las** zijn onderscheidende parameters die veel invloed hebben op het wel of niet ontstaan van klachten. Dit zijn de meest dominante parameters.

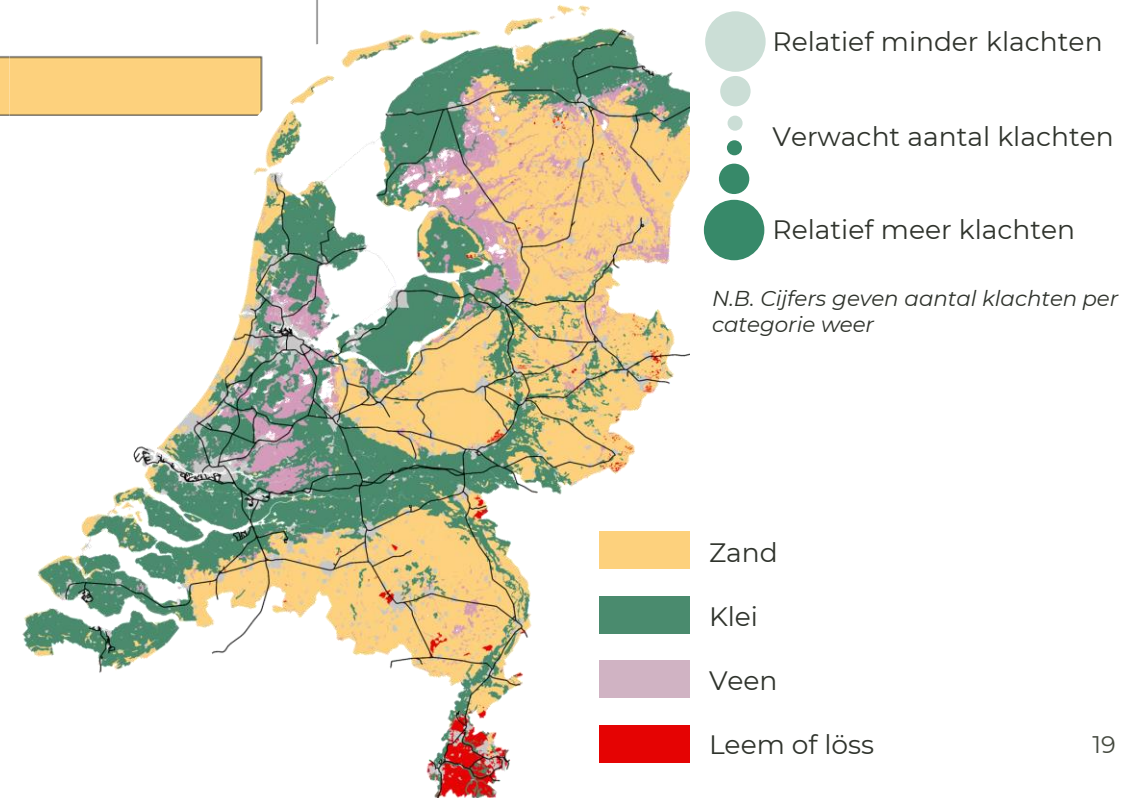
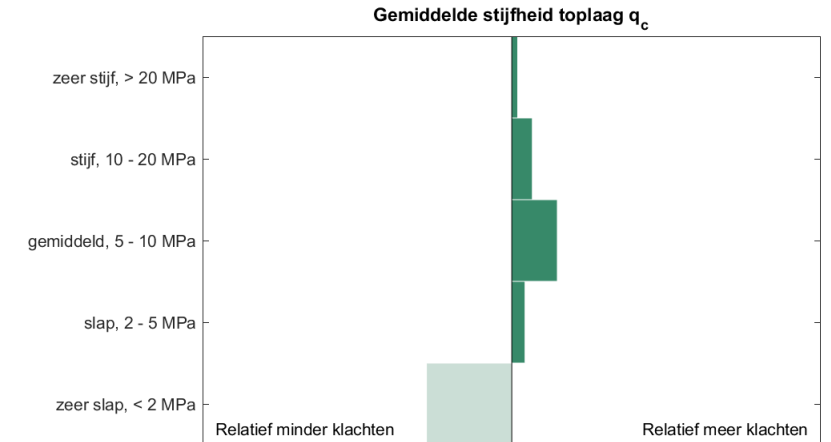
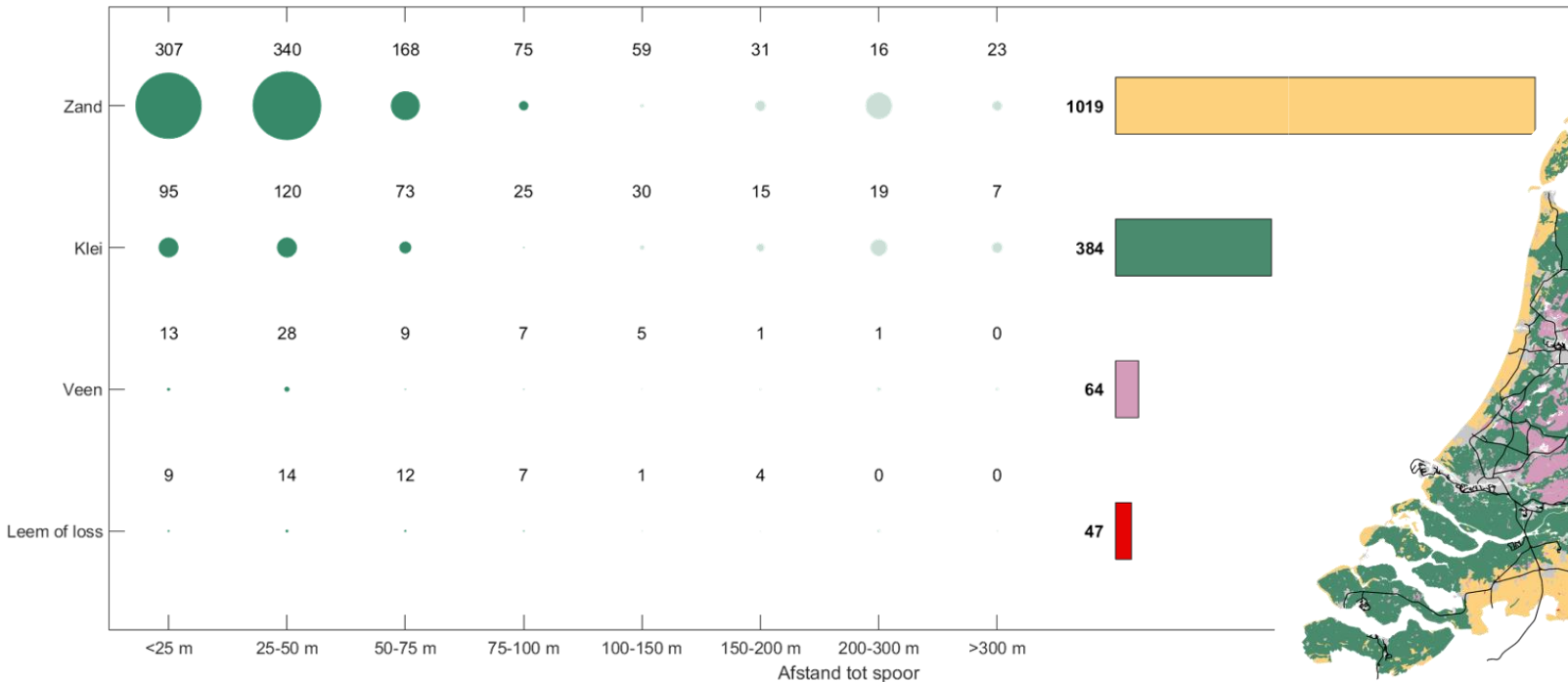


3. RESULTATEN

INVLOED VAN AFSTAND EN BODEMTYPE

Belangrijkste bevindingen

- Relatief meer klachten op **korte afstand** van het spoor, minder op grote afstand van het spoor.
- Relatief meer klachten op **zandbodems**, hier komen ook absoluut de meeste klachten vandaan.
- Relatief minder klachten op slappe bodems.
- Ook klachten op meer dan 2 km gemeld.

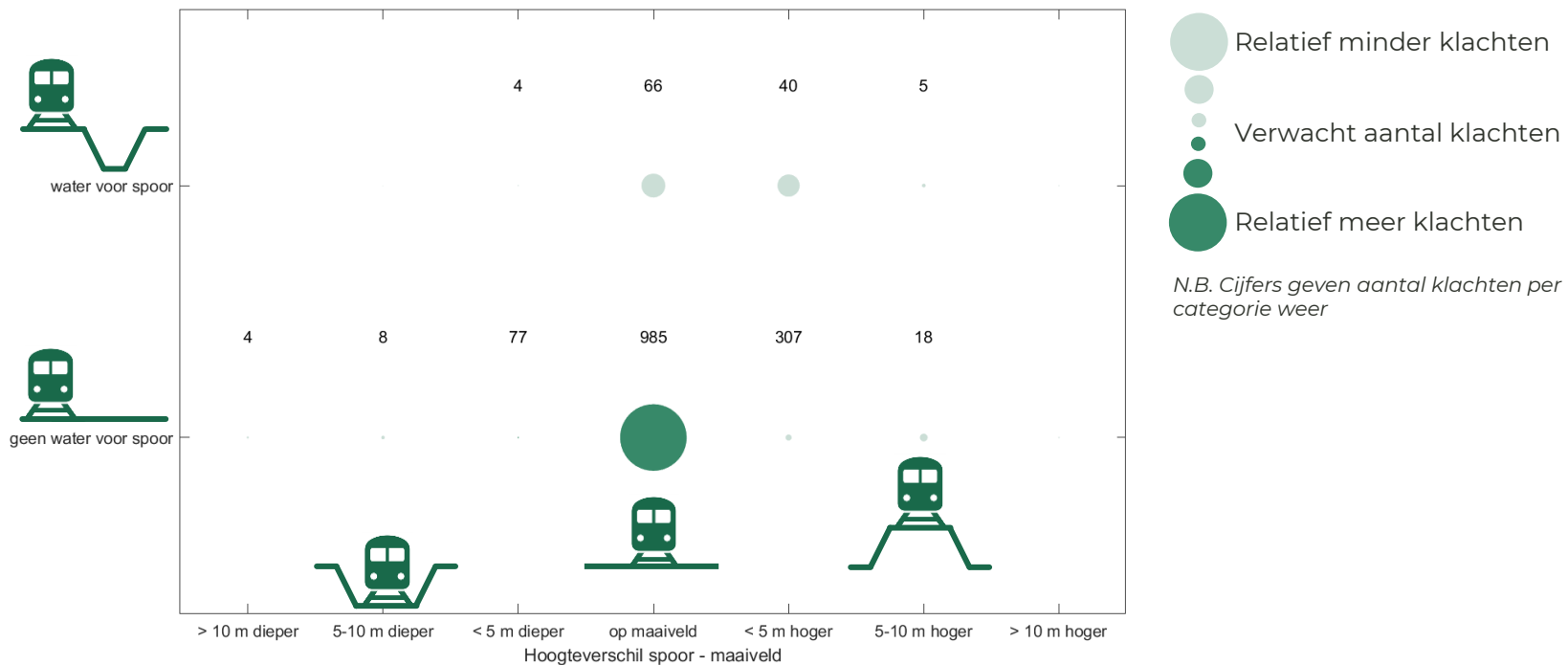


3. RESULTATEN

INVLOED VAN HOOGTELIKKING EN WATER

Belangrijkste bevindingen

- Relatief meer klachten bij **spoor op maaiveld**, als er **geen sloot** is tussen spoor en bebouwing.
- Relatief minder klachten bij spoor op maaiveld als er wel een sloot is tussen spoor en bebouwing.
- Relatief minder klachten bij spoor op verhoogd talud met sloot.

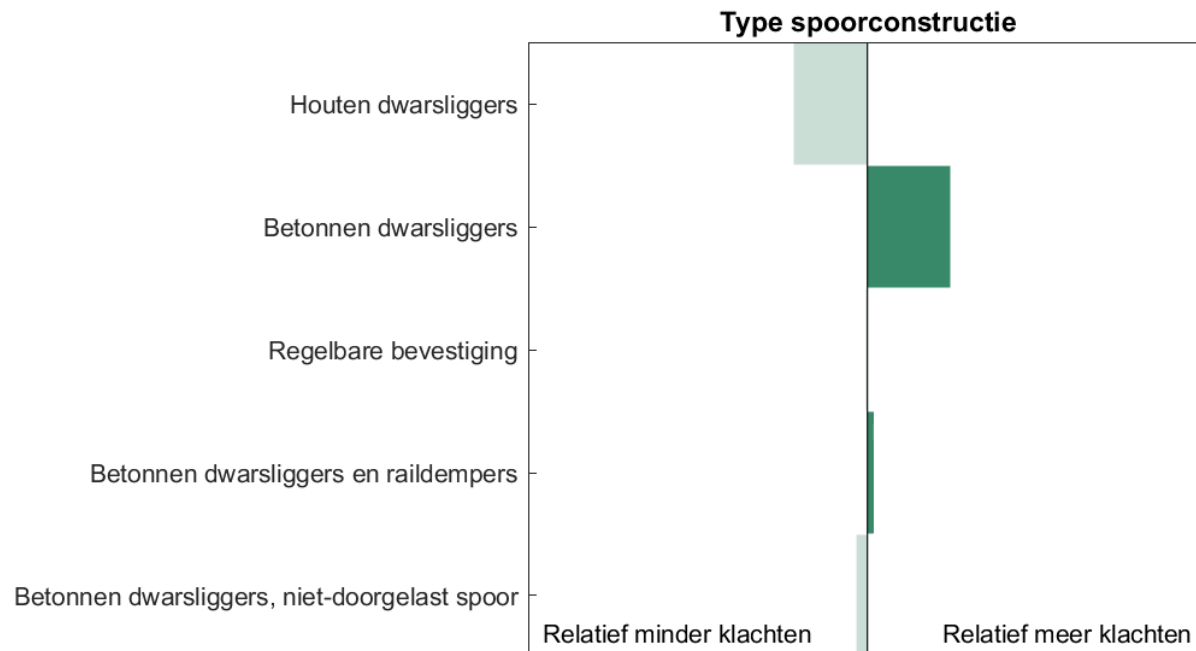


3. RESULTATEN

INVLOED VAN SPOORCONSTRUCTIE

Belangrijkste bevindingen

- Relatief meer klachten bij **betonnen dwarsliggers**.
- Relatief minder klachten bij houten dwarsliggers.

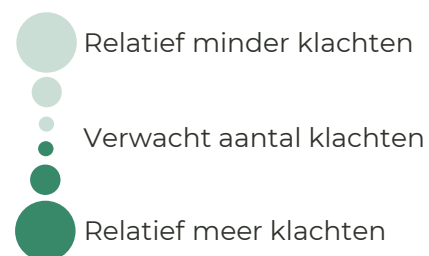
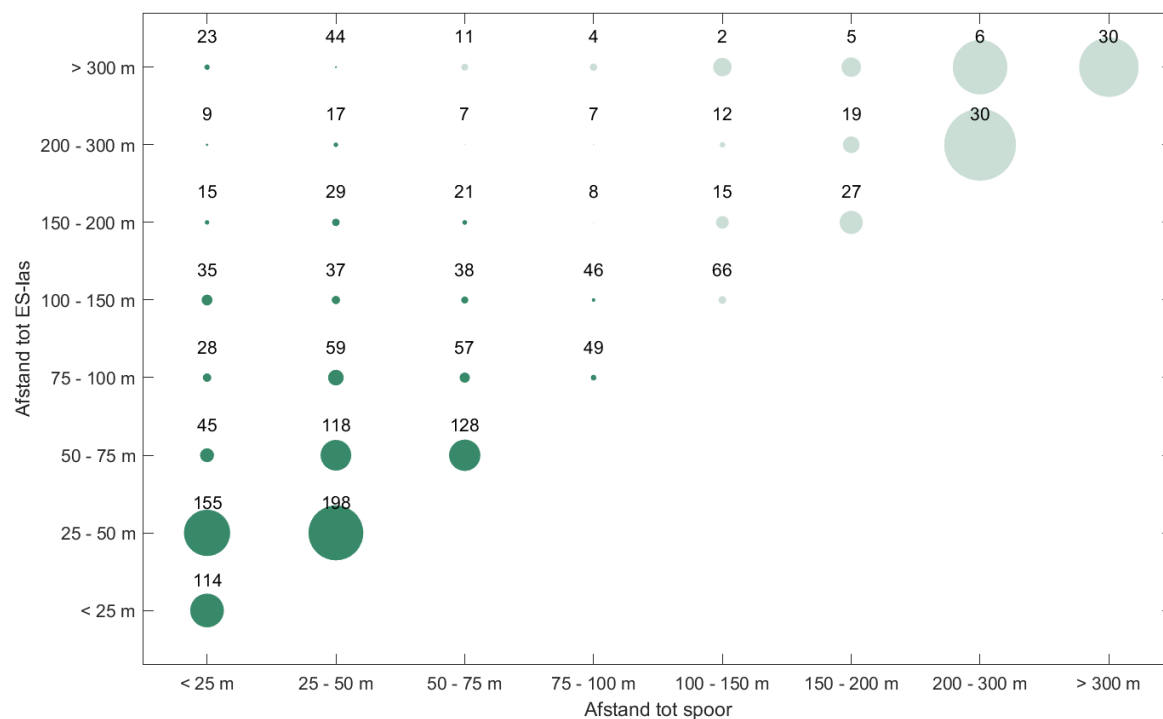


3. RESULTATEN

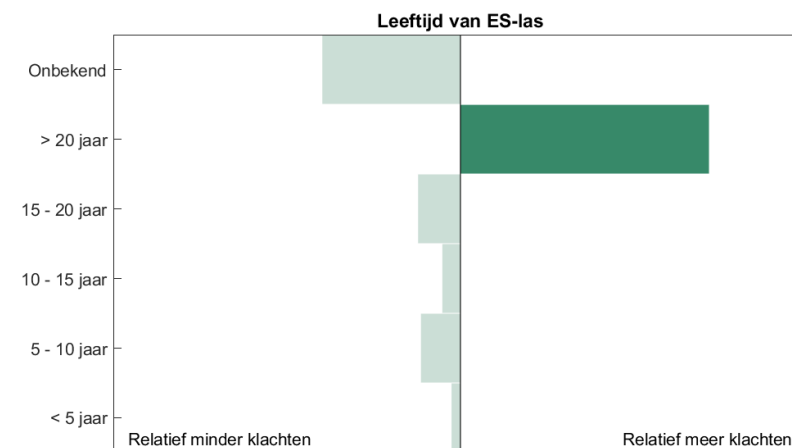
INVLOED VAN ES-LASSEN

Belangrijkste bevindingen

- Relatief meer klachten op **korte afstand van ES-lassen**.
- Invloed is beperkt tot ca. 75 meter.
- Relatief meer klachten in de buurt van **oudere ES-lassen**.



N.B. Cijfers geven aantal klachten per categorie weer

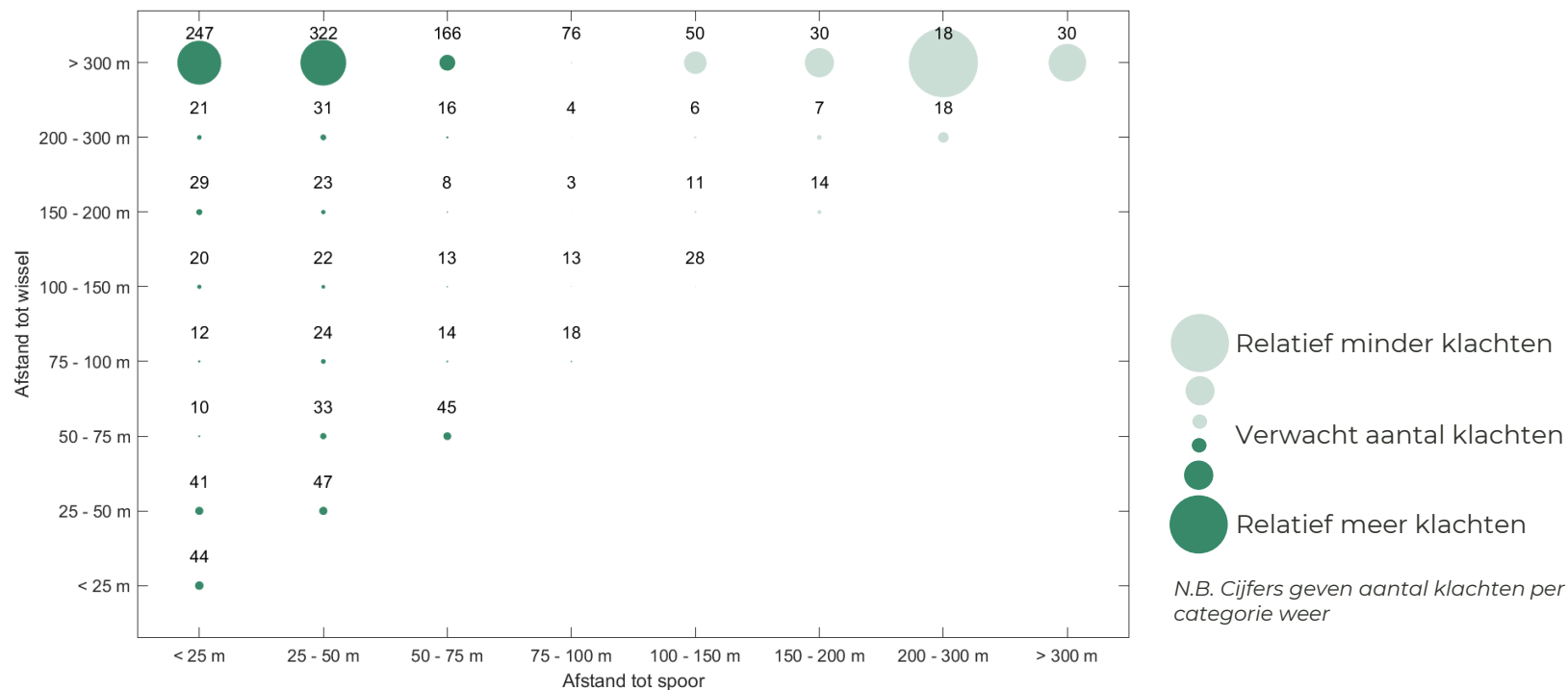


3. RESULTATEN

INVLOED VAN WISSELS

Belangrijkste bevindingen

- Relatief iets meer klachten op **korte afstand van wissels**.
- Invloed is beperkt tot ca. 75 meter.

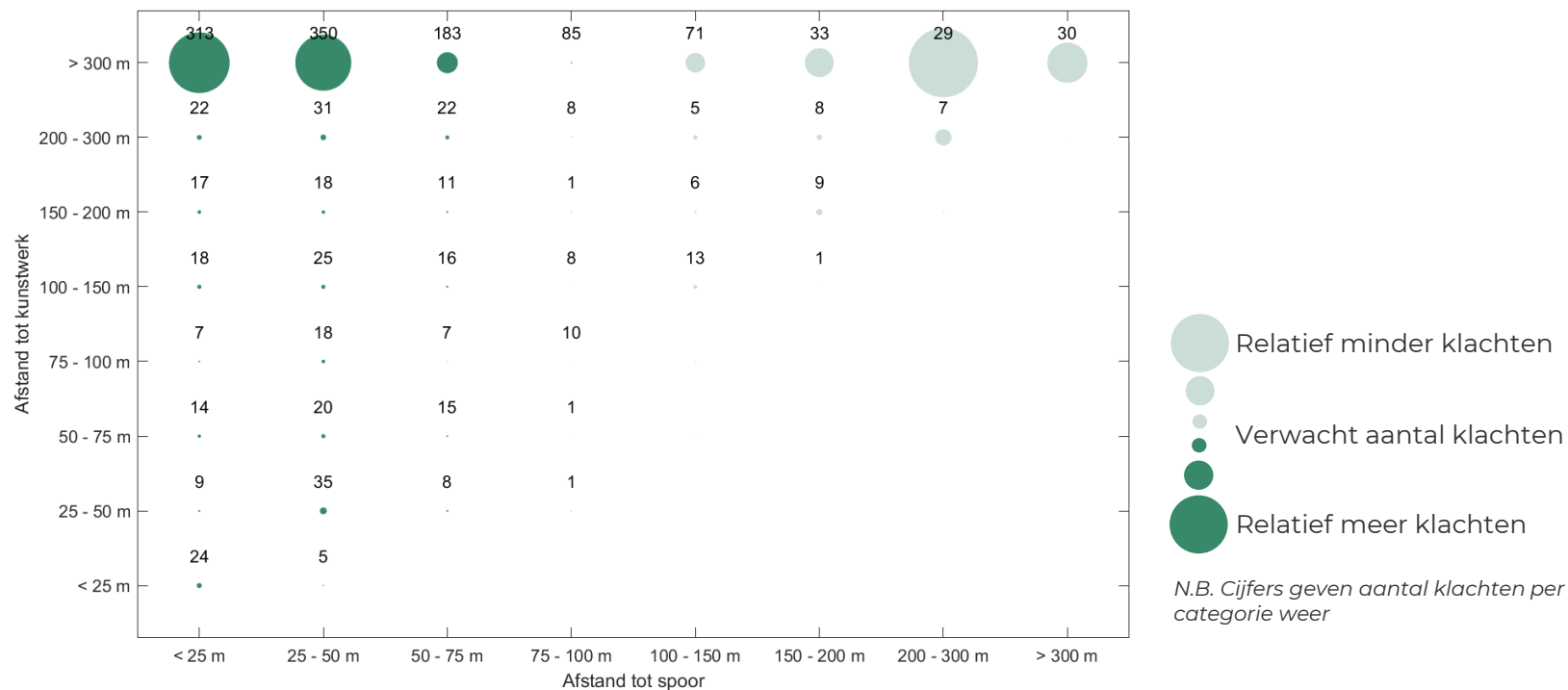


3. RESULTATEN

INVLOED VAN KUNSTWERKEN

Belangrijkste bevindingen

- Kunstwerken hebben geen significante invloed op het aantal klachten.

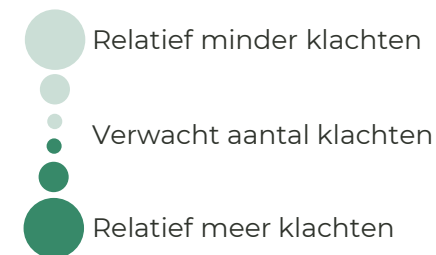


3. RESULTATEN

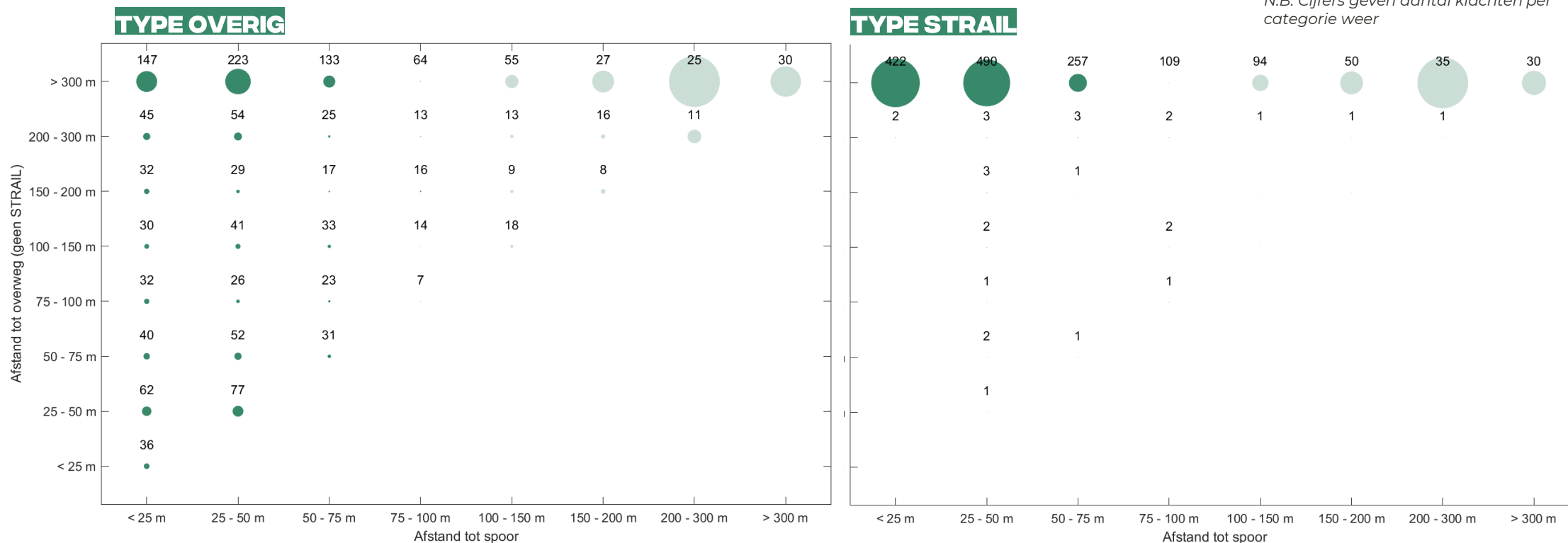
INVLOED VAN OVERWEGEN

Belangrijkste bevindingen

- Relatief meer klachten **dichtbij overwegen**, maar niet bij overwegen van type STRAIL.
- Invloed is beperkt tot ca. 75 meter.



N.B. Cijfers geven aantal klachten per categorie weer

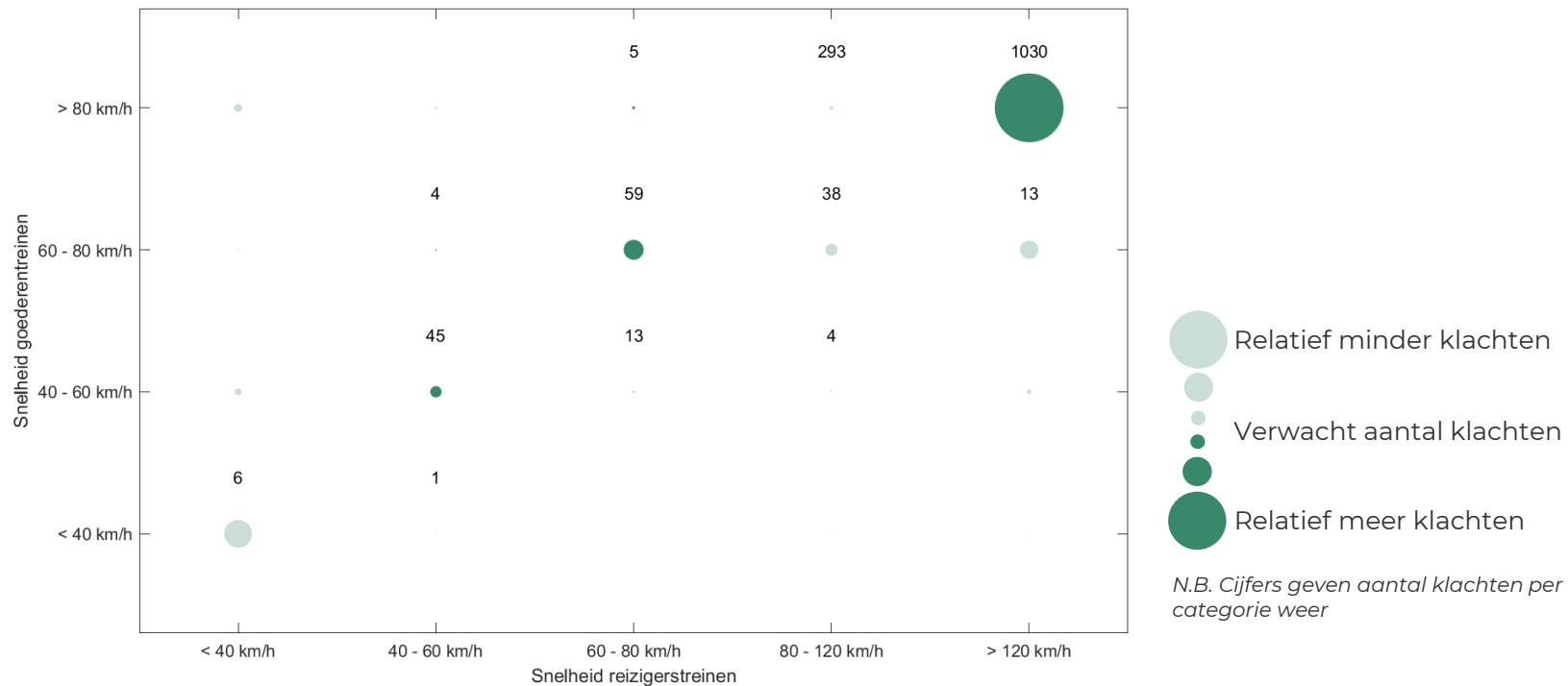


3. RESULTATEN

INVLOED VAN RIJSNELHEID

Belangrijkste bevindingen

- Relatief meer klachten bij treinen op **hoge rijsnelheid**.

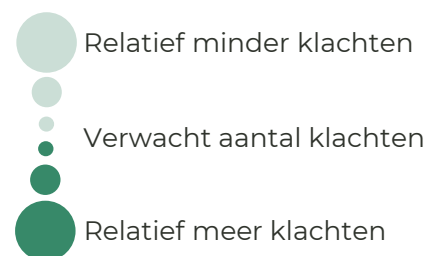
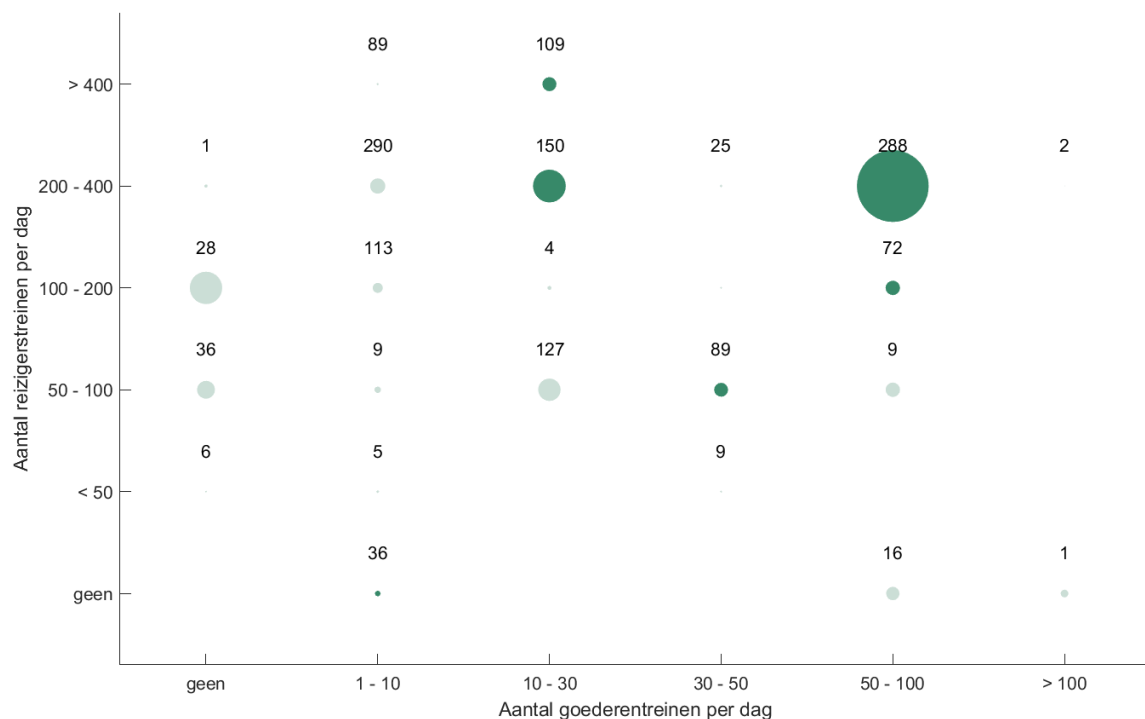


3. RESULTATEN

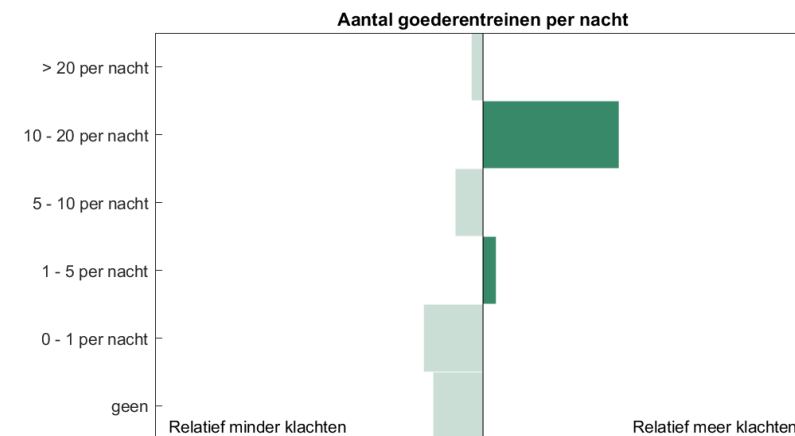
INVLOED VAN AANTAL TREINEN

Belangrijkste bevindingen

- Relatief meer klachten op **drukke, gemengde trajecten**.
- Relatief minder klachten op trajecten zonder goederentreinen.
- Relatief meer klachten bij veel **goederentreinen in de nacht**.



N.B. Cijfers geven aantal klachten per categorie weer

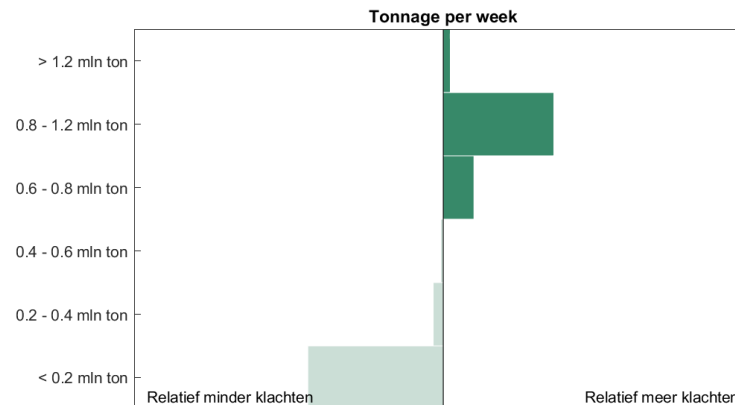
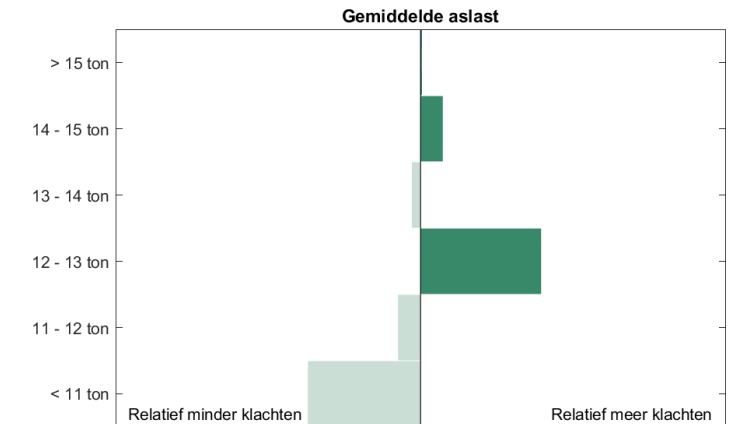
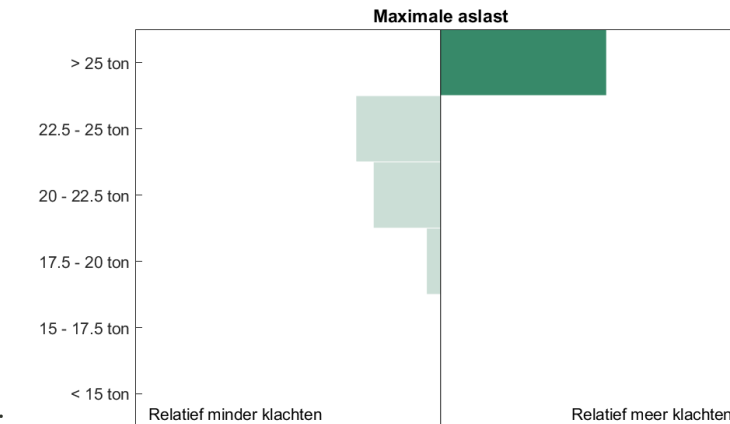


3. RESULTATEN

INVLOED VAN ASLAST

Belangrijkste bevindingen

- Relatief meer klachten van trajecten met **hoge maximale aslast** en lagere gemiddelde aslast: gemengde lijnen met afwijkende treinen.
- Relatief minder klachten van trajecten met lichtere treintypes.
- Relatief meer klachten bij **zwaar bereden trajecten**.

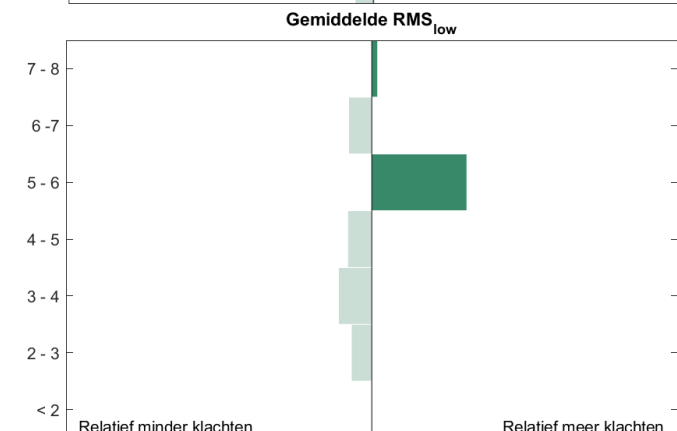
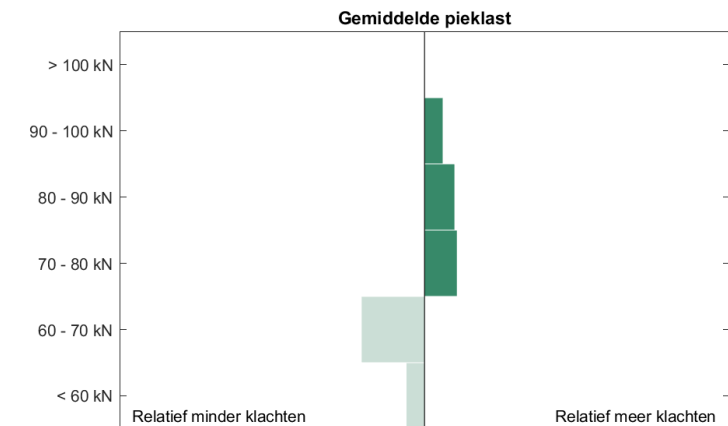
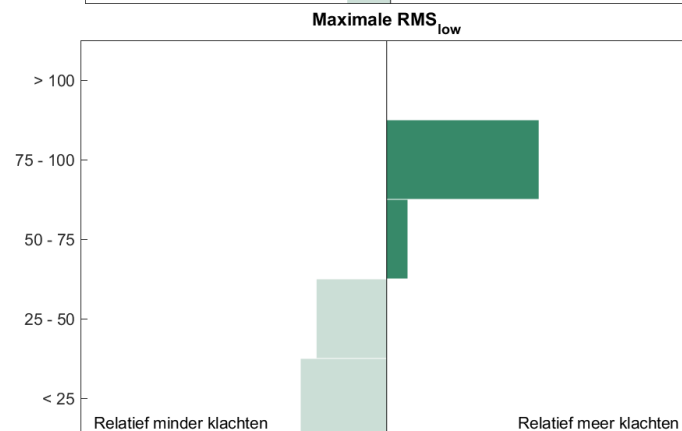
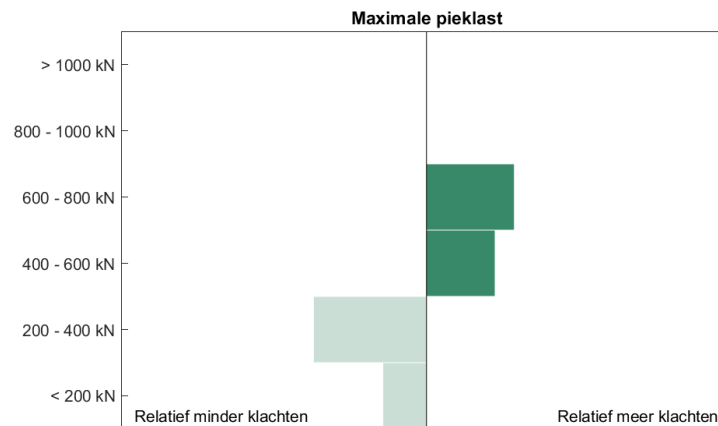


3. RESULTATEN

INVLOED VAN WIELKWALITEIT

Belangrijkste bevindingen

- Relatief meer klachten bij **hoge maximale (dynamische) pieklast**, indicator van dynamische aslast (trillingen van voertuig op baan).
- Relatief meer klachten van trajecten met **hoge RMS_{low}** , indicator van wielonrondheid.

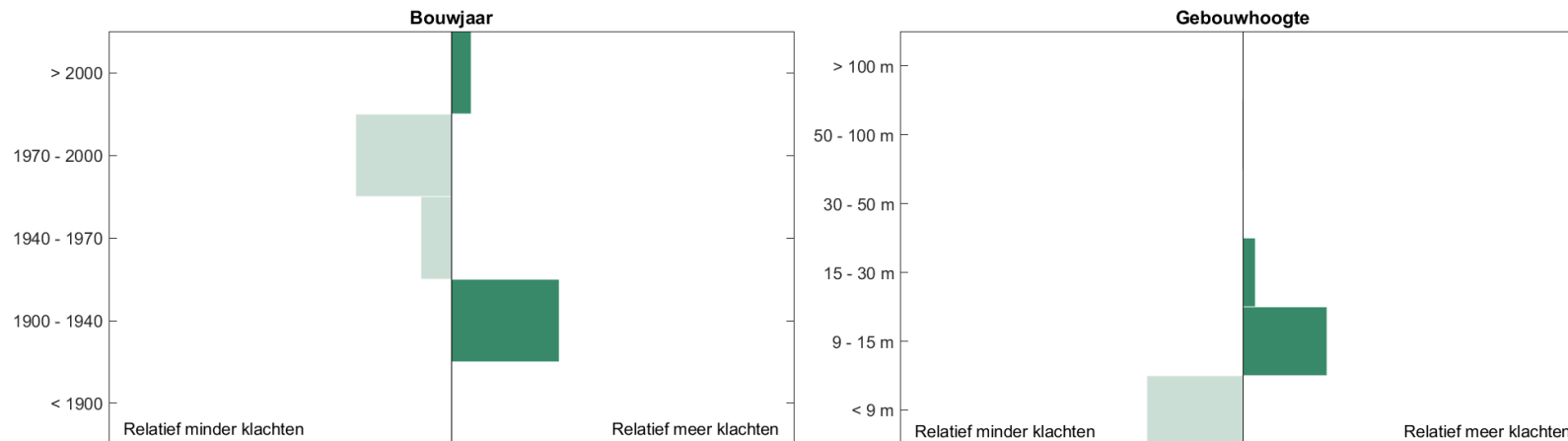


3. RESULTATEN

INVLOED VAN GEBOUWEIGENSCHAPPEN

Belangrijkste bevindingen

- Relatief meer klachten bij **vooorlogse woningen** (houten vloeren), minder bij nieuwere woningen (betonnen vloeren).
- Relatief meer klachten bij **appartementengebouwen**, minder bij grondgebonden woningen. Maar: ook meer potentiële klagers per gebouw in hoogbouw (meer adressen).

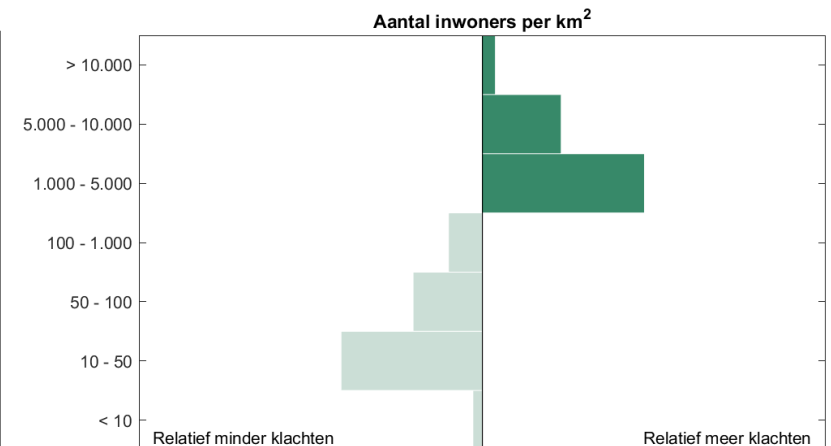
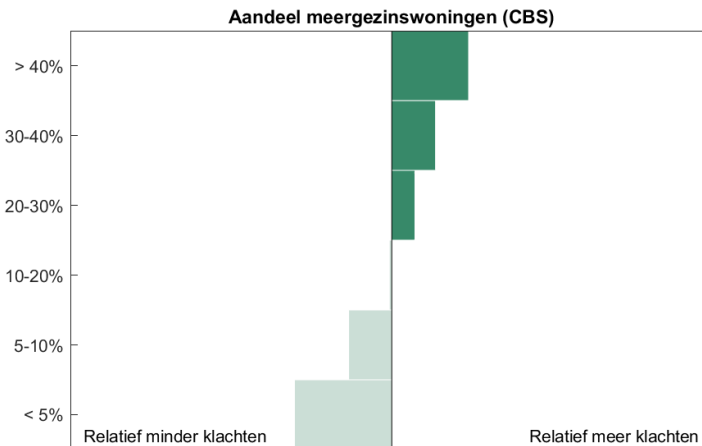
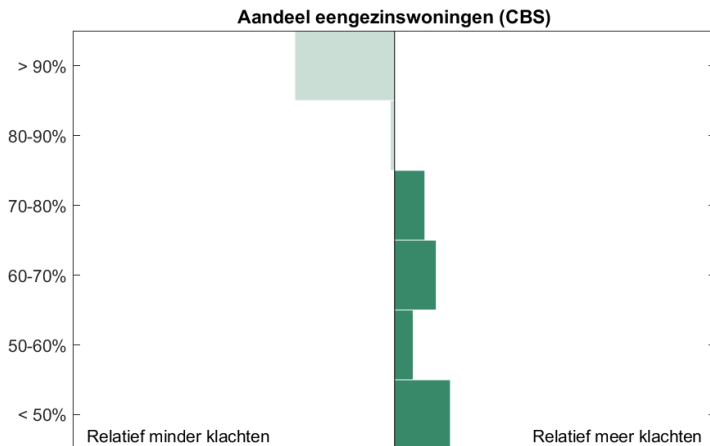


3. RESULTATEN

INVLOED VAN BUURTGEGEVENS

Belangrijkste bevindingen

- Relatief meer klachten uit wijken met **hoge bevolkingsdichtheid** (stedelijke milieus).
- Relatief meer klachten uit wijken met **meer appartementengebouwen**, en minder grondgebonden woningen (eengezinswoningen).

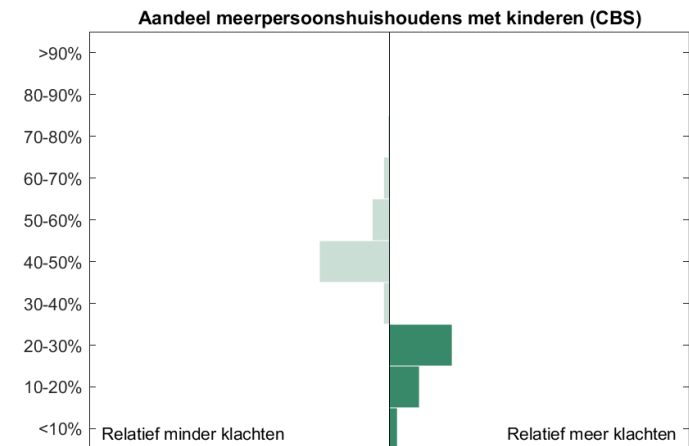
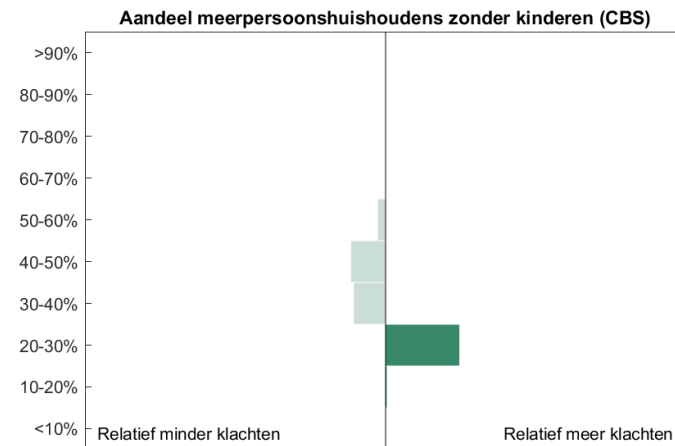
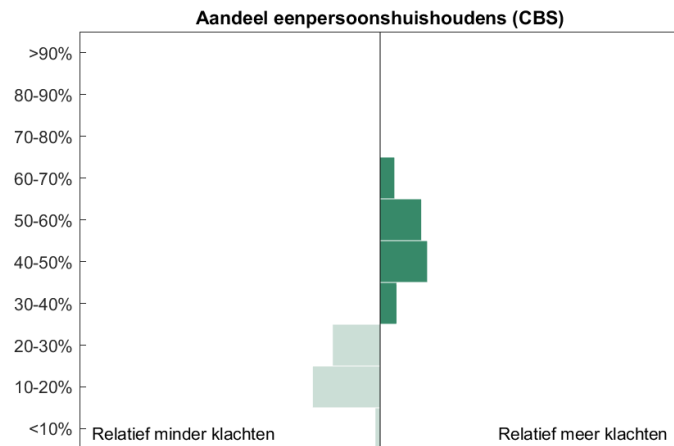


3. RESULTATEN

INVLOED VAN BUURTGEGEVENS

Belangrijkste bevindingen

- Relatief meer klachten uit buurten met **weinig kinderen** en **veel eenpersoonshuishoudens**.

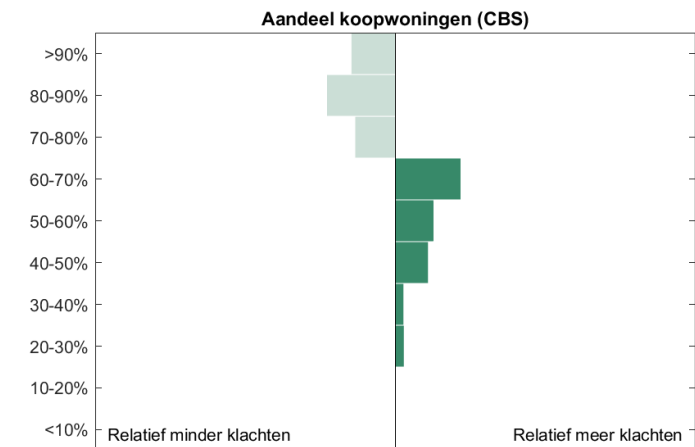
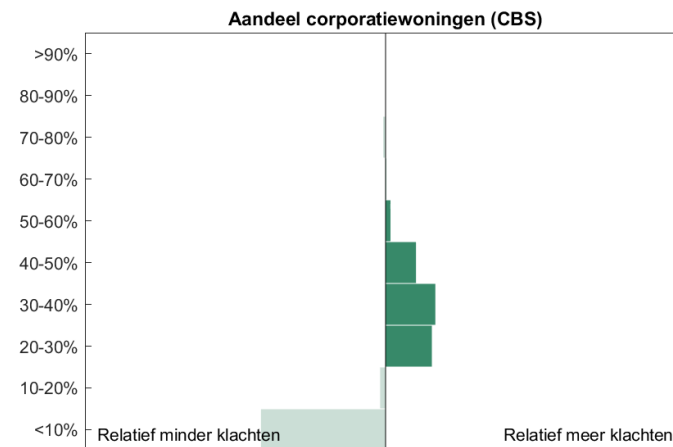
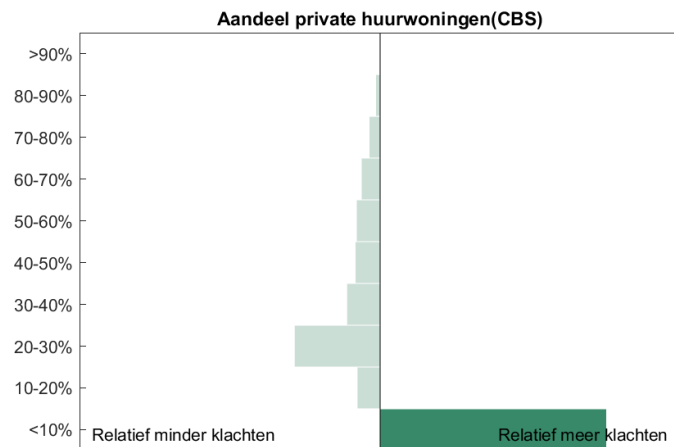


3. RESULTATEN

INVLOED VAN BUURTGEGEVENS

Belangrijkste bevindingen

- Relatief meer klachten uit buurten met **weinig private huur woningen**, verder vrij diffuus beeld.



4. SAMENVATTING

4.1 Afsluitende opmerkingen

In dit rapport hebben we beschreven wat de bepalende parameters zijn voor het ontstaan van klachten over trillingen door treinen. Bij de resultaten zijn een aantal nuances te plaatsen:

1. Veel klachten ontstaan als gevolg van bewonersacties (handtekeningacties) of media-aandacht, en kunnen daarmee het beeld vertekenen (bias). Waar bekend hebben we dit ook aangegeven in dit rapport.
2. Veel parameters hangen met elkaar samen: zo zijn bijvoorbeeld waterpartijen langs het spoor vaak aanwezig op slappe bodems, en komen appartementengebouwen vaker voor in dichtbevolkte gebieden. Het Random Forest-model houdt rekening met dit soort kruiscorrelaties, maar de uitwisseling van kruisgecorrleerde parameters is niet helemaal te voorkomen in de overige resultaten.

Op basis van dit onderzoek is rechts aangegeven wat condities zijn voor het vaker voorkomen van klachten.

4.2 Belangrijkste condities voor grotere kans op klachten

1. Woning staat **binnen 75 meter van spoor**
2. Spoor ligt **op maaiveld**, **zonder spoorloot**
3. Woning staat **binnen 75 meter van (oudere) ES-las**
4. Woning staat in **wijk met weinig private huurwoningen**
5. (Veel) **zware treinen** (hoge aslast)
6. Woning staat in **dichtbevolkte wijk**
7. Veel treinen met **hoge dynamische aslast** en **grote wielonrondheid**
8. Veel treinen, vooral **veel goederentreinen in de nacht**
9. Wekelijks **hoog tonnage** aan verkeer
10. Woning bevindt zich in **appartementengebouw**
11. Woning staat **binnen 75 meter van overweg**
12. Woning is **gebouwd tussen 1900 en 1940**

BIJLAGE I. CORRELATIEMATRIX

I.1 Over correlaties

Correlaties geven aan wat de samenhang is tussen twee variabelen. Daarmee zijn ze een indicatie of er een relatie is tussen twee parameters. Een sterke positieve relatie ligt dichtbij +1, een sterke negatieve relatie dichtbij -1. Nuance hierbij is dat correlaties vooral iets zeggen over de lineaire samenhang tussen parameters. Parameters kunnen ook bij de afwezigheid van een lineaire correlatie wel degelijk met elkaar samenhangen.

Om deze correlaties te bepalen, hebben we de dataset met klachten samengevoegd met de referentiedataset met niet-klachten. Vervolgens hebben we de correlatie berekend tussen het wel/niet aanwezig zijn van een klacht en de kenmerken van de locatie van de klacht.

De correlatiematrix is weergegeven op de volgende pagina. De belangrijkste bevindingen hieruit hebben we opgenomen in het overzicht rechts, en komen relatief goed overeen met de uitkomsten uit het Random Forest-model.

I.2 De kans op een klacht lijkt kleiner als...

1. De **afstand tot het spoor groot** is
2. De **afstand tot een niet-STRAIL overweg, ES-las of wissel groot** is
3. Er **water langs het spoor** ligt, de breedte hiervan maakt niet uit
4. De woning **nieuw** is, en **laag**
5. De **maximale aslast en pieklast van treinen laag** zijn
6. Het **tonnage per week laag** is
7. Het **aantal treinpassages laag** is
8. Er **geen TRAXX** locomotief passeert
9. De woning in een wijk met **lage bevolkingsdichtheid** ligt
10. De woning in een wijk met **weinig senioren** en **veel 45-65 jarigen** ligt
11. De woning in een wijk met **grote huishoudens** ligt
12. De woning in een wijk met **veel private huurwoningen** ligt

[illegible]

BIJLAGE II. OPZET RANDOM FOREST-MODEL

II.1 Model uitgangspunten

Voor het ontwikkelen van het model zijn de volgende uitgangspunten gebruikt:

- Er is gebruik gemaakt van repeated stratified k-fold met 10 splits.
- Er is een test dataset aangemaakt van 25% van de totale dataset. Deze zogeheten “hold-out sample” is gebruikt om de kwaliteit van het model te testen.
- Er is gebruik gemaakt van Synthetic Minority Oversampling Technique (SMOTE), gezien het aantal klachten ten opzicht van geen-klachten relatief laag lag en de dataset dus niet de benodigde balans had om een RF analyse uit te kunnen voeren.
- Om overfitting te voorkomen is de maximale diepte van de ‘trees’ en het maximaal aantal variabelen per ‘tree’ beperkt tot 5. Het model is robuust voor deze settings.

- De afstand tot het spoor is begrensd op 250 meter. Dit komt overeen met de afstand waarbinnen de meeste klachten voorkomen. Hierdoor wordt het effect van afstand tot het spoor niet overschat.
- De kwaliteit van de modellen hebben wij gemeten op basis van de F1 score, recall, precision en een confusion matrix.

II.2 Gebruikte methode voor parameterdominantie

Voor het bepalen van de parameter dominantie is gekozen voor de SHAP methode. Bij deze methode wordt de waarde van elke variabele vervangen door een baseline waarde. Vervolgens wordt de performance van het model vergeleken met het model met de originele waarden. Op basis hiervan wordt het effect van een variabele vastgesteld. Deze methode is daardoor niet gevoelig voor de ranges van de variabelen of sterke onderlinge correlaties.

BIJLAGE II. OPZET RANDOM FOREST-MODEL

Er zijn ook alternatieve methodes mogelijk om de parameterdominantie vast te stellen:

- Standaard methode, gebaseerd op Gini. Deze methode is niet gebruikt, omdat in dit onderzoek sprake is van combinaties van numerieke variabelen met grote en kleine ranges, en dummy variabelen (range 0 tot 1). De Gini methode is gevoelig voor deze verschillen.
- Permutation methode. Deze methode is gevoelig voor sterke correlaties tussen variabelen. Bij een sterke correlatie kunnen individuele effecten worden onderschat, deze methode is vanwege deze gevoeligheid niet gebruikt.

