

WEBBOOST

ProRail

KLACHTEN OVER TRILLINGEN

ONDERZOEK NAAR INVLOED REIZIGERSMATERIEEL

COLOFON

Auteur(s) Pieter Boon
Marjolein Hordijk
Thomas Vos

Projectcode WBD2024-036

Versienummer 2.0

Datum 8 april 2025

Status Definitief

Opdrachtgever ProRail B.V.



© We-Boost 2025

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 2022 is onderzocht wat de belangrijkste zogenaamde 'drivers' zijn voor klachten over trillingen over de periode 2017 tot medio 2021. Doel hiervan we om vast te stellen of er specifieke situaties of omstandigheden aan te wijzen zijn waarbij er meer overlast is van trillingen? Het antwoord op die vraag is gebruikt om de aanpak om trillingen te verminderen te prioriteren.

Het voorliggende rapport is een uittreksel uit een groter rapport waarin ingegaan wordt op alle bepalende factoren. Het voorliggende rapport focust alleen op de invloed van vervoerder en treintype.

Net als in het eerdere onderzoek is door middel van een data-analyse met behulp van een zogenaamd Random Forest-model 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 rijsnelheid) 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.

MANAGEMENTSAMENVATTING

Wat kun je hiermee?

Op basis van dit onderzoek is te zien dat bepaalde treintypes sterker dan andere treintypes verantwoordelijk lijken te zijn voor klachten. Het gaat dan met name om goederentreinen en TRAXX+ICRmh-treinen (beide waren al bekend uit eerdere onderzoeken), maar ook treinen van het type VIRM, Flirt, DDZ en SNG lijken voor opvallend meer klachten te zorgen dan verwacht. Treinen van het type GTW lijken juist voor minder klachten te zorgen. Overigens kan hierbij meespelen dat op veel trajecten waar de GTW is geïntroduceerd, in het verleden treintypes reden die relatief vaak voor overlast zorgden, zoals de DM'90. Vanuit andere onderzoeken blijkt dat de historische trillingsbeleving (is er iets verbeterd of verslechterd?) een belangrijke rol speelt in het wel of niet melden van een klacht.

Verder kan dit rapport worden gebruikt als extra onderbouwing bij de keuze van nieuwe materieelsoorten, als trillingen hierbij een rol speelt. Blijkbaar zorgen bepaalde eigenschappen van treinen (bij sprinters SLT en GTW) voor lagere trillingen dan andere eigenschappen (bij sprinters Flirt en SNG).

De vraag is welke (combinatie) van eigenschappen deze trillingsemissie vooral beïnvloedt.

Tenslotte laat dit onderzoek (opnieuw¹⁾) 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 (die andere trillingsfrequenties aantikken). En niet eens perse door hogere trillingen zelf.

¹⁾ In eerdere onderzoeken (o.m. m.b.t. introductie TRAXX+ICRmh op Brabantroute, snelheidsverhoging Naarden-Bussum, snelheidsverhoging goederentreinen Utrecht) is te zien dat een verschuiving van de maatgevende trillingsfrequenties in sommige gebouwen tot fors hogere trillingen kan leiden (als die woningen gevoelig zijn voor de nieuwe maatgevende trillingsfrequenties). Terwijl de trillingen in andere woningen (die juist gevoelig waren voor de eerder maatgevende trillingsfrequenties) kunnen afnemen. Voor veel waargenomen effecten in dit onderzoek kan dit ook een verklaring zijn.

INHOUDSOPGAVE

Managementsamenvatting	3
1. Inleiding	5
2. Aanpak	6
3. Resultaten	9
4. Samenvatting	44
Bijlages met achtergrondinformatie	45



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 probleemlocaties zijn bekend, denk aan harde punten (overwegen, kunstwerken) en 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 2022 is hier onderzoek naar gedaan, de uitkomsten van dat onderzoek zijn mede gebruikt om projecten binnen het IBS te prioriteren. In het voorliggende rapport wordt specifiek gefocust op de bijdrage van specifieke treintypes en vervoerders aan het ontstaan van klachten.

1.2 Doel

Doel van dit onderzoek is om te onderzoeken of bepaalde treintypes (wellicht op bepaalde bodemtypes of vooral rond overwegen) voor relatief meer klachten zorgen dan andere treintypes.

1.3 Leeswijzer

Dit rapport is als volgt opgebouwd:

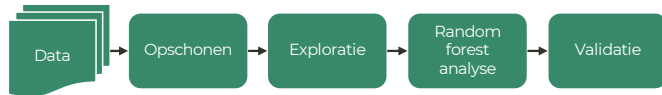
- In hoofdstuk 2 lichten we de aanpak toe
- In hoofdstuk 3 geven we de resultaten van het onderzoek
- We sluiten af met een beknopte samenvatting in hoofdstuk 4

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 per jaar 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, RIVM, QuoVadis
- Sporen
- Rijsnelheid
- Spoorconstructie
- Treinaantallen
- Treintypes
- Treinparameters
- Geluidniveaus

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.

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 I.

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.

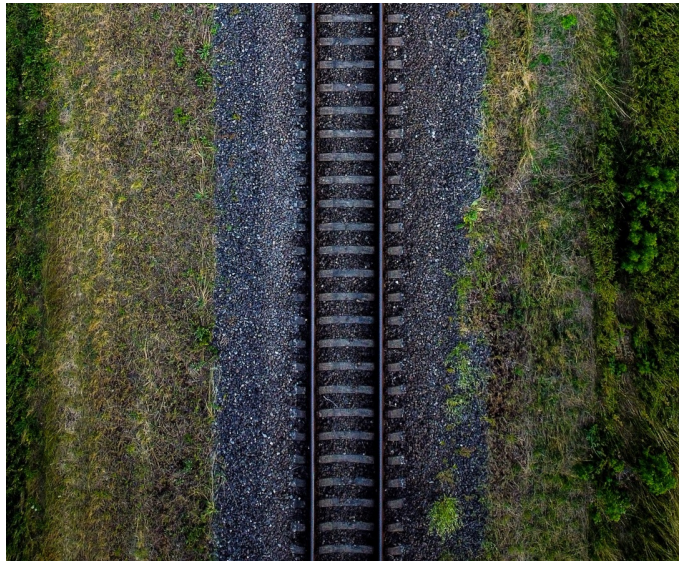
3. RESULTATEN

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 verschillen tussen treintypes en vervoerders.

3.2 Exploratiefase

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



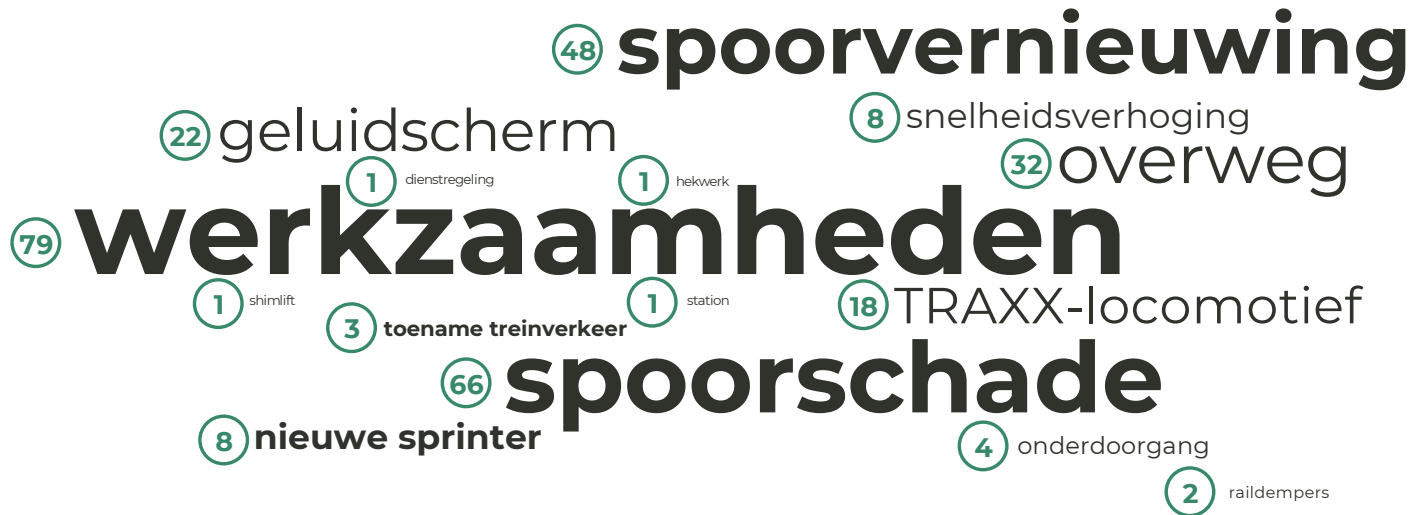
3. RESULTATEN

WAAR WORDT HET VAAKST OVER GEKLAAGD?



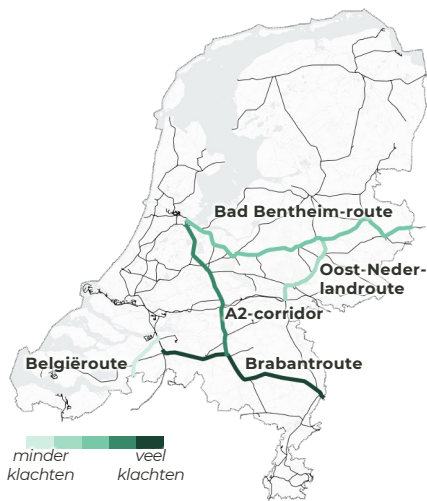
3. RESULTATEN

WANNEER WORDT HET MEEST GEKLAAGD?



3. RESULTATEN

VERDELING OVER HET LAND - MAATGEVENDE CORRIDORS



CORRIDORS

Vooral klachten op Brabantroute

Gemeente	Vorige positie	Route	Klachten	Unieke klachten
Oisterwijk	1 (0)	Brabantroute	135	87
Tilburg	5 (+4 ↑)	Brabantroute	129	73
Rheden	2 (-1 ↓)	Oost-Nederlandroute	892	70
Utrecht	6 (+2 ↑)	A2-corridor	93	55
Oosterhout	3 (-2 ↓)	Brabantroute	99	50
Gooise Meren	4 (-2 ↓)	Bad Bentheim-route	75	45
Roosendaal	10 (+3 ↑)	Belgiëroute	48	36
Breda	17 (+9 ↑)	Brabantroute	58	34
Eindhoven	8 (-1 ↓)	Brabantroute	87	34
Gilze en Rijen	7 (-2 ↓)	Brabantroute	46	33
Vught	9 (-2 ↓)	A2-corridor	61	32
Helmond	18 (+6 ↑)	Brabantroute	41	31

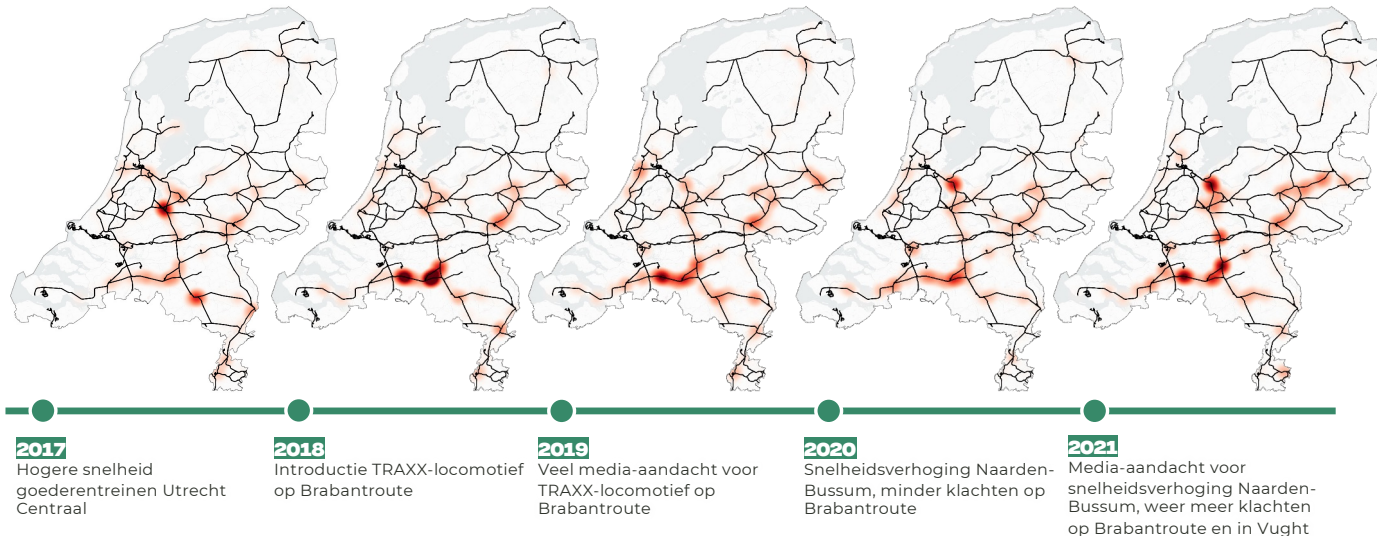
GECONCENTREERDE KLACHTEN

In 12 gemeenten:

- 33% van unieke klachten (was 36% over 2017-2021)
 - 46% van alle klachten (was 60% over 2017-2021)
- Meer spreiding, minder hotspots dan in periode 2017-2021

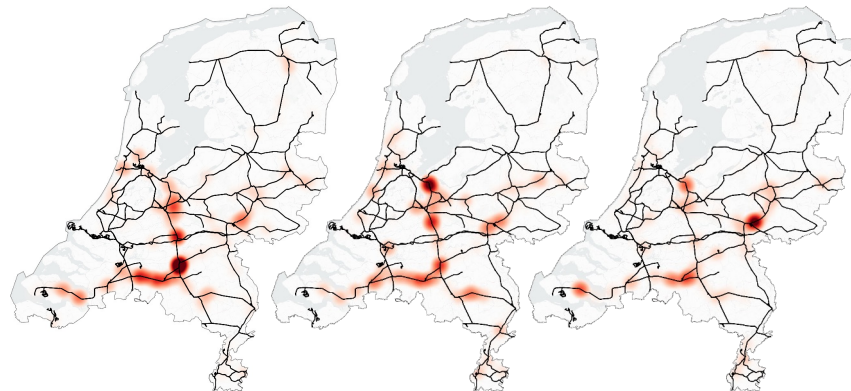
3. RESULTATEN

HOTSPOTS PER JAAR



3. RESULTATEN

HOTSPOTS PER JAAR

**2022**

Werkzaamheden in
Geldermalsen en Vught

2023

Weer meer klachten in
Naarden-Bussum, afname in
Vught en Geldermalsen

2024

Meer klachten in Goes en Velp

3. RESULTATEN

VERDELING OVER JAREN EN MAANDEN



ALLE KLACHTEN

- Afname na eerdere toename.
- Duidelijk verschil tussen **winter en zomer** ontbreekt.

UNIEKE KLACHTEN

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

3. RESULTATEN

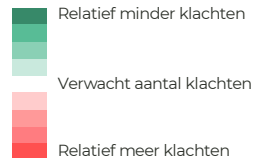
INVLOED VAN TREINTYPE

Belangrijkste bevindingen

- Relatief meer klachten bij elk uur **goederentreinen** en treinen van type **Flirt**, **VIRM** en **TRAXX+ICR**¹⁾.
- Treintypes GTW, WINK en Protos zorgen juist voor relatief minder klachten.

	geen	incidenteel	dagelijks	elk uur	
Goederen	122	110	1683	781	
Overig reizigers	1197	1021	284	194	
ICE	2017	305	374		
Thalys	2354	270	72		
Velaro	2615	76	5		
Protos	2538	137		21	
WINK	2677		4	15	
GTW	1490	750	25	431	
Flirt	97	183	1083	1333	
SNG	1262	664	308	462	
SLT	1132	610	392	562	
VIRM	271	40	803	1582	
DDZ	584	934	576	602	
ICM	796	597	573	730	
ICNG	2218	400	78		
ICR	971	611	620	494	
	Aantal treinen				

¹⁾ Uit correlatiematrix (zie Bijlage II) zien we dat treinen van het type goederen, TRAXX+ICR en Flirt vaak op dezelfde trajecten rijden, waardoor klachten van het ene type ook door een ander van deze types kunnen komen op een deel van de trajecten. Dit kan het beeld vertekenen in dit overzicht. Op de volgende slide is dit terug te zien: uit de RF-analyse komt de Flirt namelijk niet als belangrijk naar voren, maar uit een vergelijking van klachten en niet-klachten wel. Blijkbaar is die laatste vergelijking vertekend doordat er vaak ook goederentreinen en TRAXX+ICR-treinen rijden als er ook Flirt-treinen rijden. De RF-analyse haalt dit eruit, de vergelijking alleen niet.



N.B. Cijfers geven aantal klachten per categorie weer

3. RESULTATEN

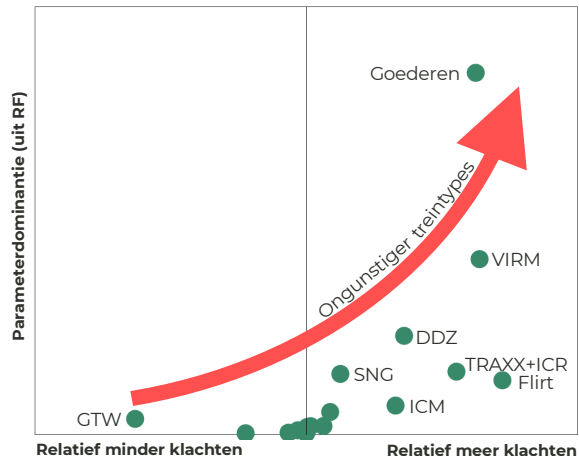
INVLOED VAN TREINTYPE

Belangrijkste bevindingen

- Treintypes die relatief meer klachten veroorzaken, hebben waarschijnlijk slechtere trillingseigenschappen: **Flirt, VIRM, TRAXX+ICR, DDZ, goederen**
- Treintypes die relatief minder klachten dan verwacht veroorzaken, hebben waarschijnlijk gunstiger trillingseigenschappen (**GTW**)

Treintype	Invloed op klachten	Toelichting
Goederentrein		Meer klachten bij meer goederentreinen
VIRM		Vooraf minder klachten als deze <u>niet</u> rijdt
DDZ		Diffuus beeld
ICR		Vooraf meer klachten als deze <u>wel</u> rijdt
SNG		Vooraf meer klachten als deze <u>wel</u> rijdt
Flirt		Vooraf minder klachten als deze <u>niet</u> rijdt
ICM		Geen duidelijke invloed
SLT		Geen duidelijke invloed
GTW		Vooraf minder klachten als deze <u>wel</u> rijdt
ICE		Geen duidelijke invloed
ICNG		Geen duidelijke invloed
Thalys		Geen duidelijke invloed
Protos		Geen duidelijke invloed
Wink		Geen duidelijke invloed

Uit RF-analyse (parameterdominantie)



3. RESULTATEN

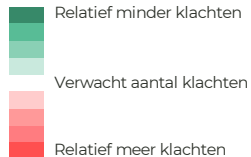
INVLOED VAN SNELHEID EN TREINTYPE

Belangrijkste bevindingen

- Relatief meer klachten bij **hoge snelheid** bij lagere snelheden is aantal klachten zoals verwacht.

	< 40 km/h	40 - 60 km/h	60 - 80 km/h	80 - 120 km/h	> 120 km/h
Goederen	6	36	85	2045	
Overig reizigers	1	12	9	63	104
ICE		1	3	29	102
Thalys					1
Velaro					
Protos			4	7	10
WINK		6	1	2	6
GTW	1	15	32	117	257
Flirt	1	7	28	197	1090
SNG		8	10	101	341
SLT		18	24	121	391
VIRM		11	43	313	1200
DDZ		2	7	76	511
ICM		12	12	85	621
ICNG					
ICR		2	5	30	454
	< 40 km/h	40 - 60 km/h	60 - 80 km/h	80 - 120 km/h	> 120 km/h

Rijdsnelheid



N.B. Cijfers geven aantal klachten per categorie weer voor treinen die elk uur passeren (bij goederen en internationaal minimaal 5 per dag, beide richtingen samen)

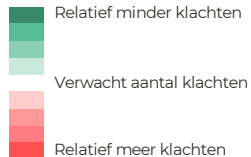
3. RESULTATEN

INVLOED VAN SPOORCONSTRUCTIE EN TREINTYPE

Belangrijkste bevindingen

- Invloed van treintype op klachten is grootst **ES-lassen**, in mindere mate ook bij **overwegen** en **doorgaand spoor**
- Bij wissels en kunstwerken geen duidelijke invloed van treintype op klachten, bij wissels wordt vaak ook langzamer gereden (zie vorige slide)

	Doorgaand spoor	Overweg	ES-las Binnen 75 meter van	Wissel	Kunstwerk
Goederen	472	334	1110	296	153
Overig reizigers	46	41	103	34	27
ICE	17	10	87	17	4
Thalys					1
Velaro					
Protos	3	3	12	5	2
WINK	3	2	11	10	
GTW	96	93	204	120	38
Flirt	292	201	692	144	84
SNG	101	88	242	59	29
SLT	98	80	315	127	53
VIRM	363	275	848	210	129
DDZ	124	100	325	68	36
ICM	157	132	369	119	45
ICNG					
ICR	128	55	193	57	30



N.B. Cijfers geven aantal klachten per categorie weer voor treinen die elk uur passeren (bij goederen en internationaal minimaal 5 per dag, beide richtingen samen)

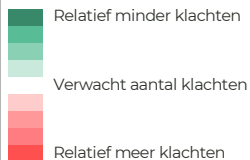
3. RESULTATEN

INVLOED VAN BODEM EN TREINTYPE

Belangrijkste bevindingen

- Relatief meer klachten op **zandbodem**, in mindere mate op kleibodems
- Bij overige bodemtypes (veen, leem of loss) is aantal klachten zoals verwacht

	Zand	Klei	Veen	Leem of loss
Goederen	1483	576	73	42
Overig reizigers	105	70	15	4
ICE	88	52	2	
Thalys		2		
Velaro				
Protos	19	2		
WINK	4	11		
GTW	128	257	17	29
Flirt	1059	210	6	58
SNG	345	85	32	
SLT	253	228	81	
VIRM	873	559	80	70
DDZ	513	67	21	1
ICM	630	66	32	2
ICNG				
ICR	457	28	6	3



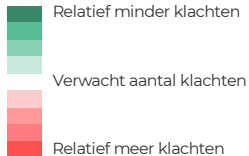
N.B. Cijfers geven aantal klachten per categorie weer voor treinen die elk uur passeren (bij goederen en internationaal minimaal 5 per dag, beide richtingen samen)

3. RESULTATEN

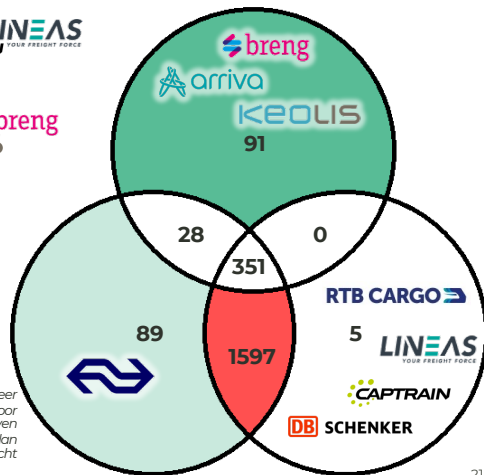
INVLOED VAN VERVOERDER EN TREINTYPE

Belangrijkste bevindingen

- Relatief meer klachten als treinen van **NS** elk uur passeren en **goederentreinen** minimaal 1 keer per dag passeren.
- Echter: relatief minder klachten als alleen reizigerstreinen passeren. Effect is bij treinen van NS kleiner dan bij treinen van andere vervoerders.



N.B. Cijfers geven aantal klachten per categorie weer



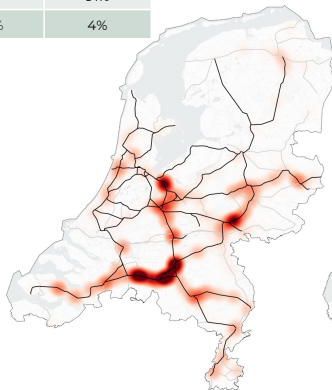
N.B. Cijfers geven aantal klachten per categorie weer als sprake is van structureel vervoer (> 1 per uur voor reizigers, > 3 per dag voor goederen); kleuren geven aan of er relatief meer of minder klachten zijn dan verwacht

3. RESULTATEN

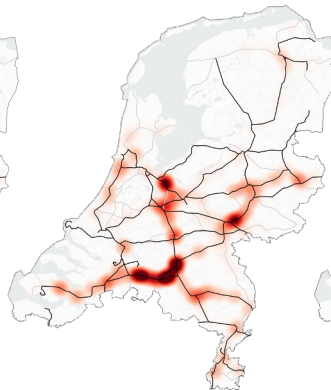
INVLOED VAN VERVOERDER EN TREINTYPE

Dit beeld (meer klachten als NS en andere vervoerder rijdt, minder klachten als NS niet rijdt) is ook terug te zien in onderstaande tabel en kaarten

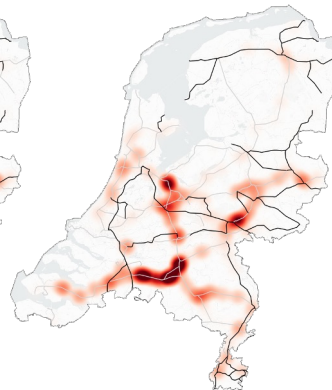
Type traject	Verwacht percentage klachten	Werkelijk percentage klachten
Trajecten met alleen NS	14%	4%
Trajecten met NS en andere vervoerder	63%	91%
Trajecten zonder NS	23%	4%

**TRAJECTEN NS**

Vooral klachten waar NS rijdt ...

**TRAJECTEN GOEDEREN**

... en waar goederentreinen rijden

**TRAJECTEN ANDERE VERVOERDERS**

Weinig klachten bij andere reizigersvervoerders

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 (handtekeningenacties) 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 zijn wel een aantal conclusies te trekken, deze zijn hiernaast weergegeven.

4.2 Belangrijkste treintypes voor grotere kans op klachten

Er is een grotere kans op klachten als er **goederentreinen** passeren, vooral als deze op zandgrond en op hoge snelheid passeren. Het effect is het sterkst in de nabijheid van ES-lassen.

Specifiek voor reizigerstreinen zorgen vooral de **VIRM**, de **DDZ**, de **SNC**, **TRAXX+ICR**, **Flirt**, en de **SNC** voor hogere trillingen. Treinen van het type **GTW** zorgen juist voor lagere trillingen.

BIJLAGE I. OPZET RANDOM FOREST-MODEL

I.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.

I.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 I. 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.

BIJLAGE II. CORRELATIEMATRIX TREINTYPES

	Stijfheid bodem	Treinen met aslast >20 ton	DDZ	FLIRT	Goederentreinen	Goederentreinen nacht	GTW	ICE	ICM	TRAXX+ICR	Protos	SLT	Thalys	VIRM	SNG	Velaro	ICNG	WINK
Stijfheid bodem	1.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.0	0.1	-0.3	-0.1	0.0	0.0	0.0	0.0	-0.1
Treinen met aslast >20 ton	0.0	1.0	0.1	0.6	1.0	0.9	-0.2	0.1	0.0	0.7	0.0	-0.1	0.2	0.2	-0.1	0.1	0.2	-0.1
DDZ	0.1	0.1	1.0	0.1	0.1	0.1	0.0	0.1	0.4	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.0	-0.1
FLIRT	0.1	0.6	0.1	1.0	0.6	0.5	-0.3	0.1	-0.1	0.5	0.1	-0.1	0.0	0.2	-0.1	0.0	0.2	-0.1
Goederentreinen	0.0	1.0	0.1	0.6	1.0	1.0	-0.3	0.1	0.0	0.7	0.0	-0.1	0.1	0.2	-0.1	0.1	0.2	-0.1
Goederentreinen nacht	0.0	0.9	0.1	0.5	1.0	1.0	-0.3	0.1	0.0	0.5	0.0	-0.1	0.0	0.1	-0.1	0.0	0.1	-0.1
GTW	0.1	-0.2	0.0	-0.3	-0.3	-0.3	1.0	0.0	-0.2	-0.1	0.0	-0.1	-0.3	0.0	-0.1	-0.1	0.0	0.0
ICE	0.1	0.1	-0.1	0.1	0.1	0.1	0.0	1.0	-0.1	0.0	0.0	0.1	0.0	0.2	-0.1	0.0	0.0	0.0
ICM	0.1	0.0	0.4	-0.1	0.0	0.0	-0.2	-0.1	1.0	0.1	0.0	0.2	0.0	0.0	0.4	0.0	0.0	-0.1
TRAXX+ICR	0.0	0.7	0.0	0.5	0.7	0.5	0.0	0.1	0.0	1.0	0.0	0.0	0.3	0.2	0.0	0.2	0.2	0.0
Protos	0.1	0.0	0.0	0.1	0.0	0.0	-0.1	0.0	0.0	0.0	1.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0
SLT	-0.3	-0.1	0.2	-0.1	-0.1	-0.1	0.1	0.2	0.0	0.0	1.0	0.1	0.4	0.1	0.0	0.0	0.0	-0.1
Thalys	-0.1	0.2	0.0	0.0	0.1	0.0	-0.1	0.0	0.0	0.3	0.0	1.0	0.1	0.1	0.1	0.6	0.1	0.0
VIRM	-0.1	0.2	0.0	0.2	0.2	0.1	-0.3	0.2	0.0	0.2	-0.1	0.4	0.1	1.0	0.1	0.1	0.2	-0.1
SNG	0.0	-0.1	0.1	-0.1	-0.1	-0.1	-0.1	0.4	0.0	0.0	0.1	0.1	0.1	1.0	0.2	0.1	-0.1	0.0
Velaro	0.0	0.1	0.0	0.0	0.1	0.0	-0.1	0.0	0.0	0.2	0.0	0.0	0.6	0.1	0.2	1.0	0.2	0.0
ICNG	0.0	0.2	0.0	0.2	0.2	0.1	-0.1	0.0	0.0	0.2	0.0	0.0	0.1	0.2	0.1	0.2	1.0	0.0
WINK	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.0	1.0

De correlatiematrix laat zien dat sommige combinaties van treintypes vaker op trajecten rijden dan andere combinaties.

Opvallende zaken:

- Treinen van het type SLT rijden relatief meer op slappe bodems
- De TRAXX+ICR rijdt relatief vaak samen met de Flirt en goederentreinen, waardoor het lastig is om klachten van deze types uit elkaar te halen. Treinen van het type VIRM en SLT rijden ook relatief vaak op dezelfde trajecten. De RF-analyse corrigeert hier wel voor, maar de analyses naar oververtegenwoordiging van klachten niet.

