



TRILLINGSPREDICTIE INTRILLEN DAMWANDEN

Definitieve Scriptie Aanvulling

Projectomschrijving:	Trillingspredictie Intrillen Damwanden
Afstudeeratelier / Bedrijf:	Deltares (Extern)
Bedrijfsonderdeel:	Geo-engineering
Naam:	Marwan el Morabet
Studentennummer:	500735324
Maand en Jaar van afstuderen:	Januari 2020
Examinator:	R.A.J.M. Mom

Abstract

Dit document bevat een onderzoek naar de nauwkeurigheid van het CUR 166 trillingsmodel, een poging tot het verbeteren van het model middels Machine Learning en een beschrijving van de complexiteit van trillingen.

Marwan el Morabet

Marwan.el.morabet@hva.nl

Colofon

Auteur:

Naam: Marwan el Morabet
Telefoonnummer: 06-86362033
E-mail: marwanelmorabet@gmail.com
Opleiding: Built Environment/ Civiele Techniek
Afstudeerrichting: Constructie

Bedrijfsbegeleider:

Naam: Paul Hölscher
E-mail: Paul.Holscher@deltares.nl
Bedrijf: Deltares

Afstudeerbegeleider & Tweede lezer:

Naam: René Mom
E-mail: r.a.j.m.mom@hva.nl
Bedrijf: Hogeschool van Amsterdam

Naam: William van der Bult
E-mail: w.j.m.bult@hva.nl
Bedrijf: Hogeschool van Amsterdam

CONTENTS

Voorwoord.....	4
1. Samenvatting.....	5
2. Introductie	6
2.1 Probleembeschrijving	6
2.2 Onderzoeksvragen	7
2.3 Eindresultaat	7
2.4 Scope	7
2.5 Geobrain	7
2.6 Leeswijzer.....	8
3. Trillingen	9
3.1 Dynamische Belasting	9
3.2 Energieverlies	9
3.3 Weerkaatsingen	10
3.4 Trillingsoverdracht.....	10
3.5 Complexiteit.....	11
4. CUR 166.....	12
4.1 Methode CUR 166 trillingsprognose	12
4.2 Voorbeeldberekening	15
4.3 Opmerkingen CUR 166 model.....	16
5. SBR trillingsrichtlijn.....	16
5.1 SBR trillingsrichtlijn A; Schade aan Bouwwerken	16
5.2 Bepaling Grenswaarde	17
6. Evaluatie CUR 166	20
6.1 Numerieke Evaluatie CUR 166	20
6.2 Categorische Evaluatie.....	29
7. CPT averaging Techniques	32
7.1 Nauwkeurigheid vs. Simpliciteit	32
7.2 Dimensie Reductie	32
7.3 Conclusie.....	34
8. Feature Importance	35
8.1 Correlatie	35
8.2 Boruta	38
9. Machine Learning Model	39
9.1 Regressie	39
9.2 Classificatie	40

9.3 Empirisch Modelleren	41
10. Vergelijking Defintieve Model met CUR model	41
11. Conclusie	42
12. Bibliography	44
 Bijlagen:.....	45
Bijlage A: Statistiek	46
Bijlage B Feature Importance Technieken	48
Bijlage C Machine Learning	50
Bijlage D: Toetsing statistische Significantie Invloed Lange Damwandlengte op Modelperformance	57
Bijlage E: Toetsen van normaliteit van de Verdeling	59
Bijlage F: R-Scripts.....	61
Bijlage G: 2 Bladzijden CUR 166 Evaluatie	77

VOORWOORD

In het eerste semester van het studiejaar 2019-202 heb ik, student Built Environment – Civiele Techniek aan de Hogeschool van Amsterdam, onderzoek mogen doen op de Geo-engineering afdeling van Deltares.

Gedurende mijn onderzoek ben ik bezig geweest met twee van mijn interesses. Dit betreft Geotechniek en Data Science. Tijdens deze periode ben ik bezig geweest met een onderzoek over trillingspredicties ten gevolgen van het intrillen van dammen. Hierin heb ik de nauwkeurigheid van het CUR 166 model proberen te kwantificeren en een alternatief proberen te bieden middels Machine Learning. Ik wil van deze gelegenheid gebruik maken door mijn begeleiders P. Hölscher en R. Mom te bedanken voor deze enorm leerzame periode.

Delt, 14 Januari 2020

Marwan el Morabet

1. SAMENVATTING

Door bouwwerkzaamheden kunnen er trillingen ontstaan in de ondergrond. Deze trillingen hebben de potentie om schade te veroorzaken aan aangrenzende woningen. Zo ook bij het intrillen van damwanden.

Trillingen in de ondergrond is een complex fenomeen. Vooral omdat er niet één type trilling bestaat. P, S en -oppervlaktegolven vinden allemaal tegelijk plaats. Trillingen kunnen ook reflecteren wat het systeem nog onvoorspelbaarder maakt. Dit maakt het voorspellen van de trillingen lastig. Tijdens de analyse van grondtrillingen werd geconcludeerd dat grondtrillingen complex en chaotisch gedrag vertonen. Een hele kleine verandering in initiële condities kan leiden tot een totaal andere trillingspatroon.

In de afgelopen decennia zijn er veel modellen ontwikkeld die de maximale trillingsintensiteit aan een woning proberen te voorspellen. De meest gebruikte model staat vermeld in de CUR 166. Het doel van dit afstudeerwerk is tweeledig. Enerzijds gaat het om het evalueren van de nauwkeurigheid van de CUR 166. Anderzijds gaat om het creëren van een model door gebruik te maken van machine learning.

De CUR 166 kent drie waarden. De 50%, 95% en 99% waarden. Dit betreffen de overschrijdingskansen. De 95%-waarde dient dus in slechts 5% van de gevallen overschreden te worden. De prognoses op basis van het CUR 166 model zijn bepaald voor 150 use cases en vergeleken met de gemeten waardes in de beschikbare database. De CUR 166 (50%-waarde) blijkt onnauwkeurig in het voorspellen van de maximale trilsnelheid. De CUR 166 model zit er gemiddeld 1.45 mm/s naast. Ook lijkt er weinig verband te bestaan tussen de voorspelling van de CUR en de gemeten topsnelheid.

De 95%-waarde is te conservatief. Het wordt in minder dan 1% van de gevallen overschreden. De 95%-waarde is dus eigenlijk een 99%-waarde. Om de 99%-waarde te kunnen transformeren naar de werkelijke 95%-waarde dient de waarde met een factor 2 gereduceerd te worden. De CUR 166 is daarmee betrouwbaarder dan het model zelf suggereert. Het overschatten van de trilsnelheid betekent dat er onnodig dure en trage installatiemethoden gebruikt worden.

Met behulp van machine learning zijn er verschillende modellen gemaakt. De fout van het model is nooit onder 1.40 mm/s gekomen. Dit is een verbetering, maar niet nauwkeurig genoeg om tot een verandering van de CUR 166 te resulteren. De machine learning maakt echter gebruik van veel minder parameters. Met slechts de afstandsinvloed kan het CUR 166 model evenaart worden in nauwkeurigheid.

2. INTRODUCTIE

2.1 PROBLEEMBESCHRIJVING

INSTALLATIEWIJZEN

In de bouwwereld zijn vaak horizontale grond- en/of waterkerende constructies nodig. Dit om te voorkomen dat grond horizontaal gaat verplaatsen. Hierdoor is het mogelijk om grond af te graven zonder verlies van stabiliteit. Deze constructies worden vaak gebruikt bij bouwkuipen, oevers en dijken. Een populaire grond- en/of waterkerende constructie is een damwand.

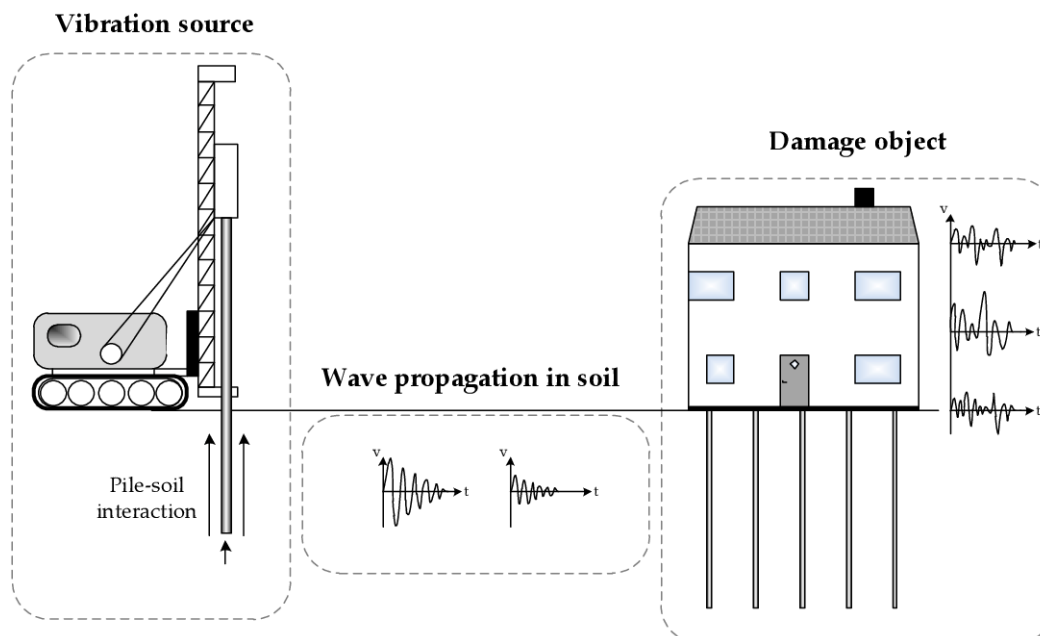
Een damwand kan op drie wijzen worden geïnstalleerd.

1. Heien
2. Trillen
3. Drukken

Trillen is de goedkoopste methode, gevolgd door heien en daarna drukken. Met heien kunnen echter grotere dieptes bereikt worden. De derde belangrijke overwegingsfactor zijn trillingen. Drukken zorgt voor de minste trillingen. Het drukken van damwanden wordt ook wel trillingsvrij genoemd (in praktijk trillingsarm). Het heien geeft herhaalde trillingen en het intrillen geeft continue trillingen. In praktijk krijgt het intrillen zoveel mogelijk de voorkeur.

TRILLINGEN

Bij het heien en trillen ontstaan er trillingen. In figuur 1 is een afbeelding te zien waarin het probleem wordt uitgebeeld.



Figuur 1; Intrillen van een damwand (Deckner, 2012)

Het is van uitermate belang voor een aannemer om onderzoek te laten doen naar het mogelijk ontstaan van schade aan omliggende bebouwing. Een dergelijk onderzoek wordt

ook wel een trillingspredictie genoemd. Het ontstaan van schade kan zorgen voor schadeclaims en langdurige vertragingen aan bouwprojecten.

Door het voorspellen van de kans op schade kan tijdig in het bouwproces geanticipeerd worden. Zo kan er gekozen worden voor een andere installatiewijze of kunnen er grondverbeteringstechnieken toegepast worden. Dit komt de voortgang van het bouwproces ten goede.

TRILLINGSPREDICTIE

Een trillingspredictie is een rekenkundige analyse naar de trillingsintensiteit veroorzaakt door bouwwerkzaamheden. Een veelgebruikte methode voor het berekenen van de maatgevende trillingsintensiteit van het installeren van damwanden staat vermeld in de CUR 166. De berekende trillingsintensiteit wordt vervolgens getoetst aan de hand van de grenswaarden vermeld in de SBR-richtlijn A: schade aan gebouwen.

De methode voor het berekenen van de trillingsintensiteit volgens de CUR 166 wordt echter door velen gezien als onbetrouwbaar. Het doel van het afstuderen is dan ook het vinden van een betrouwbaarder alternatief.

2.2 ONDERZOEKSVRAGEN

HOOFDVRAAG

- Kan de trillingspredictie voor het intrillen van damwanden verbeterd worden ten opzichte van de huidige technieken (CUR 166), met behulp van Machine Learning?

DEELVRAGEN

- Wat is de betrouwbaarheid en de nauwkeurigheid van de methode vermeld in de CUR 166 met betrekking tot het voorspellen van de trillingsintensiteit? (hfdst. 6)
- Hoe kan een sondering het best worden samengevat om als input te dienen voor mijn model? (hfdst. 7)
- Wat is de betrouwbaarheid en de nauwkeurigheid van het zelf opgestelde model ten opzichte van het CUR-model? (hfdst. 9 en 10)

2.3 EINDRESULTAAT

Het eindresultaat is een model die in staat is om de trillingsintensiteit aan een woning ten gevolge van het intrillen van damwanden binnen een zekere foutmarge te voorspellen. Deze voorspelling dient vervolgens vergeleken te worden met de CUR 166.

Het model zal gemaakt worden met behulp van een modelleringstechniek genaamd 'Machine Learning'.

2.4 SCOPE

Er zal alleen gekeken worden naar stalen damwanden die trillend zijn ingebracht met een frequentie tussen de 30 en 40 hertz. Omdat het type fundering niet vermeld stond in de Geobrain dataset zal deze parameter buiten beschouwing gelaten worden.

2.5 GEOBRAIN

Geobrain bouwtrillingen is een langdurig project van 13 aannemers en ingenieursbureaus. Deze groep heeft een ervaringsdatabase gefinancierd. Het doel van deze ervaringsdatabase is dat de betrokken partijen deze database zouden vullen met trillingsmetingen. Aan de hand van deze metingen zouden er empirische modellen gemaakt kunnen worden die zou kunnen leiden tot een verbeterde trillingsvoorspelling. Na een aantal jaren en een lege database heeft Deltares ervoor gekozen om deze database zelf te vullen. In de jaren 2009 tot en met 2011 heeft Deltares 153 trillingsmetingen uitgevoerd bij dijkswerkzaamheden te Nederlek.

Bij het intrillen van de damwanden worden er bij nabijgelegen woningen trillingsmeters opgehangen. Na afloop van de installatie worden bij deze trillingsmeters de maximale trilsnelheid genoteerd. Deze dataset bevat dus de maximale trilsnelheid van 153 woningen.

De dataset bevat de volgende elementen.

- Informatie over de trillingsbron (damwandprofiel, installeringsdiepte)
- Informatie over het medium (492 Sonderingen uit Dinoloket, afstand)
- Informatie over de constructie (SBR Categorie, Trillingsintensiteit)

Met behulp van deze dataset zal getracht worden om op een empirische wijze en met machine learning een model te maken die de maatgevende trillingsintensiteit aan een woning kan voorspellen. Ook zal de dataset gebruikt worden om de nauwkeurigheid van het CUR 166 model vast te stellen. Er zal met behulp van de CUR 166 een voorspelling gemaakt worden van de trillingsintensiteit. Deze zal vervolgens vergeleken worden met de waarde in de dataset.

2.6 LEESWIJZER

PROGRAMMEREN

Alle grafieken zijn gemaakt met de R programmeertaal. Dit is een programmeertaal die uitblinkt in het maken van datavisualisaties en geavanceerde statistische berekeningen. Het bestand waar code in zit heet een script. Mijn R-scripts bevinden zich in bijlage F.

STATISTIEK

In dit verslag zal er veel gebruik gemaakt worden van begrippen en technieken afkomstig uit de statistiek. In bijlage A zijn de meest gebruikte statistische begrippen uitgelegd.

OPZET VERSLAG

In hoofdstuk 3 zal ingegaan worden op de theoretische achtergrond van trillingen. Hoe ontstaan die en hoe planten die zich voort? In hoofdstuk 4 en 5 zal het CUR 166 model en de SBR trillingsrichtlijn besproken worden.

In hoofdstuk 6 zal de nauwkeurigheid van het CUR 166 model geanalyseerd worden. Hoofdstuk 7 en 8 zal gaan over het kiezen voor de juiste parameters voor het model. In hoofdstuk 9 zal het gaan over een zelfgemaakt model met behulp van Machine Learning. Hoofdstuk 10 zal het definitieve model vergeleken worden met het CUR-model. Het verslag zal beëindigd worden met een conclusie en een advies voor verder onderzoek.

3. TRILLINGEN

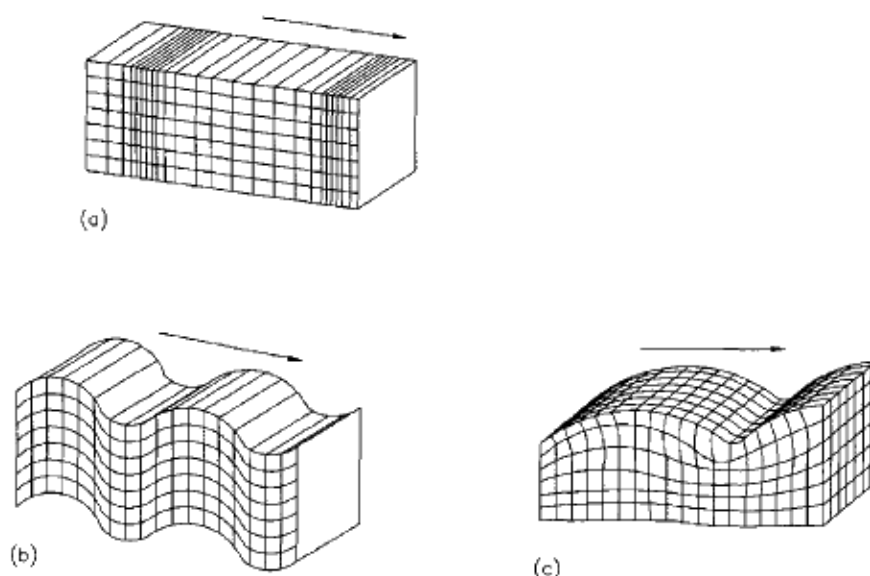
In dit hoofdstuk zal de theoretische achtergrond van trillingen besproken worden. We zullen het eerst hebben over de hoe een trilling op gang komt. Daarna zullen we hebben over hoe trillingen afzwakken en weerkaatsen. Daarnaast zal de overgang van de trillingen naar de woning besproken worden. Dit eindigt met een uitleg over de complexiteit van trillingen.

3.1 DYNAMISCHE BELASTING

Het trilblok voert een dynamische belasting uit op de damwand. Er ontstaan vervolgens interactiekrachten tussen de damwand en de ondergrond. Daardoor gaat de omliggende bodem meetrillen. Deze trillingen breiden zich uit. In het geval van trillen komen er drie soorten trillingsgolven voor;

- Compressiegolven (P-golven); De bewegingsrichting van de golf komt overeen met de richting waarin de golf zich uitbreidt.
- Afschuifgolven (S-golven); De bewegingsrichting van de golf staat loodrecht op de richting waarin de golf zich uitbreidt.
- Oppervlaktegolven (Rayleigh-golven); De golven die zich uitbreiden langs het oppervlak, waarbij het oppervlak een bolvormige beweging laat zien.

Voor een visualisatie van de drie typen trillingen verwijs ik u naar figuur 2.



Figuur 2; a. Compressiegolven. B. Afschuifgolven. C. Oppervlaktegolven (SBRCURnet, 2016)

Deze trillingen vinden alle drie tegelijk plaats. Van deze drie golven zorgen de oppervlaktegolven voor de meeste trillingen aan de woning. Deze trillingen hebben echter wel een aantal meter nodig om gang te komen.

3.2 ENERGIEVERLIES

Naarmate de trillingen verder van de woning verwijderd raken zwakken de trillingen af. Deze energieverlies heeft twee oorzaken.

- Geometrische verzwakking
- Materiaaldemping

- Geometrische verzwakking

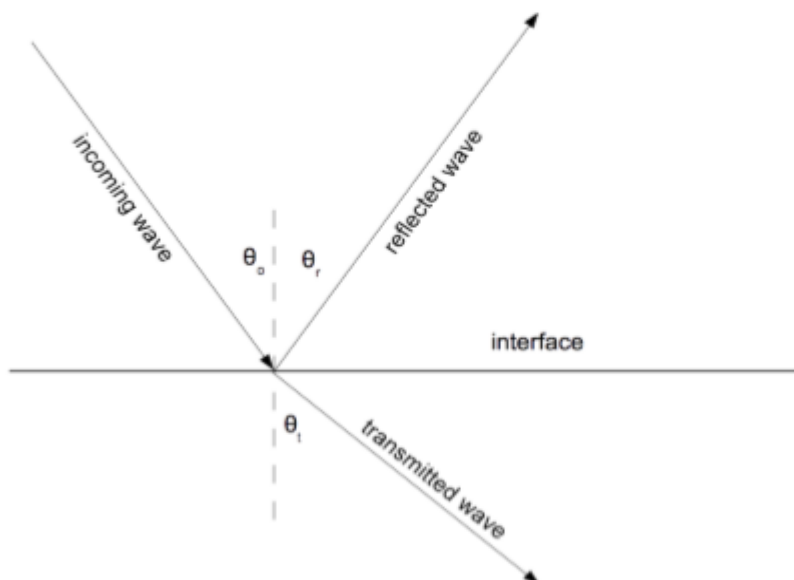
Dit heeft simpelweg te maken met het feit dat hoe verder je van de damwand verwijderd raakt, des te groter is de omtrek waarover de trillingsenergie zich zal verspreiden. Volgens de literatuur zwakt de trilling af evenredig aan $1/\sqrt{r}$

- Materiaaldemping

Dit heeft te maken met energieverlies ten gevolge van interne wrijving van het korrelskelet. De grootte van deze demping is grondsoortafhankelijk.

3.3 WEERKAATSINGEN

De ondergrond is bijna altijd heterogeen. De grondeigenschappen veranderen in zowel verticale als horizontale richting. In verticale richting kom je vaak dichter gepakte lagen met hogere conusweerstand tegen. In horizontale richting is er vaak sprake van horizontale lagen. Bodemprofielen kunnen echter ook in horizontale richting variëren door de aanwezigheid van stoorlagen of de aanwezigheid van een gedroogde rivier. Wanneer de trilling een verandering in grondeigenschap tegenkomt zal de trilling gaan reflecteren. Een voorbeeld van zo een weerkaatsing is te zien in figuur 3.



Figuur 3; Gedrag van een trilling bij het tegenkomen van een grondlaag (Hölscher, 2016)

In dit geval komt de golf een horizontale grondlaag tegen. Deze golf wordt gedeeltelijk doorgelaten en gedeeltelijk gereflecteerd.

3.4 TRILLINGSOVERDRACHT

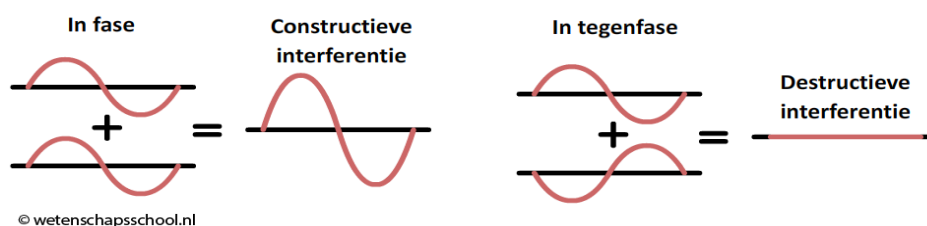
Vervolgens zullen de trillingen de woning bereiken. De overgang van trillingen van de ondergrond naar het gebouw is afhankelijk van de soort fundering en draagconstructie. De exacte hoeveelheid energieverlies die hier plaatsvindt is niet exact te bepalen.

Afhankelijk van de afmetingen en het materiaal van het gebouw zal het bij een bepaalde frequentie (eigenfrequentie) gaan resoneren. Deze trillingen hebben de potentie om schade te veroorzaken.

3.5 COMPLEXITEIT

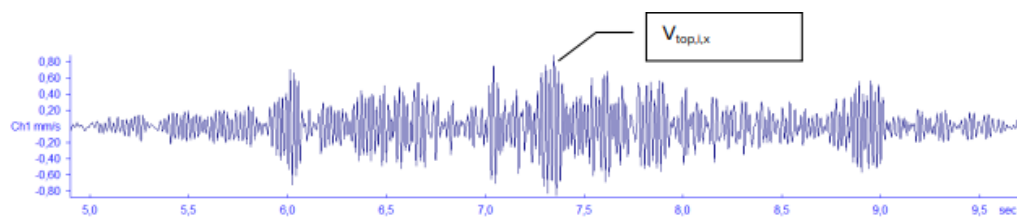
Een trillend ingebrachte damwand genereert een dynamische belasting op de omliggende ondergrond. Tussen de damwand en de ondergrond ontstaan interactiekrachten die lokaal enorm kunnen verschillen. De omliggende ondergrond gaat resoneren met de damwand en deze golven zullen zich uitbreiden in iedere richting.

Wanneer de grondeigenschappen van de ondergrond veranderen zullen de golven reflecteren. Naarmate de trilling verder van de damwand verwijderd raakt zal deze afzwakken. Vervolgens zullen de golven de woning bereiken vanuit allerlei invalshoeken. Deze trillingen kunnen elkaar versterken, maar ook verzwakken. (interferentie-effect)



Figuur 4; Schematische weergave van een interferentie effect (Dinkgreve, 2018)

Bij de overgang tussen de ondergrond en de woning vindt een onbekende hoeveelheid aan energieverlies plaats. Vervolgens gaat een constructieonderdeel resoneren met een bepaalde eigenfrequentie. Deze trilling wordt gedecomposeerd in drie dimensies. Een voorbeeld van deze trilling is te zien in figuur 5.



Figuur 5; Voorbeeld Trillingsmeting (net, 2017)

En bij het beoordelen van de trilling is alleen de topsnelheid van invloed. De volgende 4 redenen maakt het voorspellen van grondtrillingen complex.

1. Een grondtrilling is een non-lineair dynamisch systeem waarbij een heel kleine verandering in initiële condities, bijvoorbeeld door de aanwezigheid van een stoorlaag, ervoor kan zorgen dat golven op een andere wijze reflecteren. Deze trillingen kunnen elkaar bij het gebouw versterken en voor grote trillingen zorgen. Ook kunnen deze trillingen elkaar verzwakken. Een kleine verandering van de grondeigenschappen kan dus voor grote veranderingen zorgen bij de trillingen.
2. De grondeigenschappen zijn lastig te meten. Door een heterogene ondergrond en dure kosten voor het verrichten van grondonderzoek zijn de grondeigenschappen tussen de damwand en de woning met slechts grote onzekerheid vast te stellen.
3. De werking van bepaalde gedeelten van het systeem dient nog nader onderzocht te worden. Zoals de invloed van de fundering op trillingen of de impact van mate van homogeniteit van de ondergrond op de nauwkeurigheid van de voorspelling.

Dit laat zien dat het exact voorspellen van de maximale trillingsintensiteit onmogelijk is.

4. CUR 166

De CUR 166 is een handboek waarin een overzicht wordt gegeven van de methoden die in Nederland van toepassing zijn bij het ontwerp, de uitvoering en het onderhoud van damwandconstructies. In de CUR 166 deel 2 hoofdstuk 5.8 wordt een paragraaf gewijd aan trillingen, het toetsen van trillingen en het verrichten van een trillingsprognose bij het installeren van damwanden. De methode die hierin wordt vermeld is de meest gebruikte methode in Nederland. Velen gebruiken dit model, maar twijfelen aan de nauwkeurigheid hiervan. In dit hoofdstuk zullen we de werking van dit model bespreken. We zullen dit hoofdstuk beëindigen met een voorbeeldberekening.

4.1 METHODE CUR 166 TRILLINGSPROGNOSE

Zoals eerder aangegeven dempt een trilling af om 2 redenen;

1. Geometrische verzwakking

$$A(r) = A_0 \sqrt{\frac{r_0}{r}}$$

Waarin:

r is de afstand tot aan de trillingsbron in meters

r_0 is een referentieafstand in meters. Bij het intrillen geldt hier een referentieafstand van 5 meters.

A_0 is de amplitude op $r = r_0$

2. Materiaaldemping

$$A(r) = A_0 e^{-\alpha(r-r_0)}$$

α is de dempingswaarde die rekening houdt met de grondomstandigheden

De twee effecten gecombineerd geeft de volgende formule

$$A(r) = A_0 \sqrt{\frac{r_0}{r}} e^{-\alpha(r-r_0)}$$

α is een empirische waarde die bepaald zijn aan de hand van proeven. De dempingswaarde staat vermeld in tabel 5.21 van de CUR 166 deel 2. De dempingswaarde ligt tussen de 0 en 0.1 en is afhankelijk van het type bodem. Het invullen van de referentieafstanden geeft de volgende formule.

$$A(r) = A_0 \sqrt{\frac{5}{r}} e^{-\alpha(r-5)}$$

Dempingswaarde bij het intrillen van damwanden staat in de CUR 166 gelijk aan 0. Dit zorgt voor een versimpeling van de formule.

$$A(r) = A_0 \sqrt{\frac{5}{r}}$$

Volgens deze formule is de amplitude dan ook evenredig aan $1/\sqrt{r}$.

Als het gaat om het toetsen van de trillingen ten gevolge van het installeren van damwanden volgens de SBR-richtlijn dien je drie variabelen te kennen;

1. V_{max} : Dit is een trilsnelheid [mm/s] gecombineerd met een overschrijdingskans.
2. f : Bij het trillend installeren is de frequentie gelijk aan de draaifrequentie van het trilblok
3. SBR-Categorie: Dit zegt iets over de type woning. De SBR kent 3 categorieën. Meer informatie is te vinden in hoofdstuk 5.

V_{max} wordt bepaald aan de hand van de volgende formule:

$$V_{max} = V_{0,cor} C_0 f$$

Waarin $V_{0,cor}$ iets zegt over de bronsterkte op een referentieafstand. (5 meter bij het intrillen)
Het bepalen van deze bronsterkte wordt gedaan aan de hand van de volgende formule;

$$V_{0,cor} = V_0 + C_{vel}(F - 350)$$

Waarin: F de slagkracht in kN.

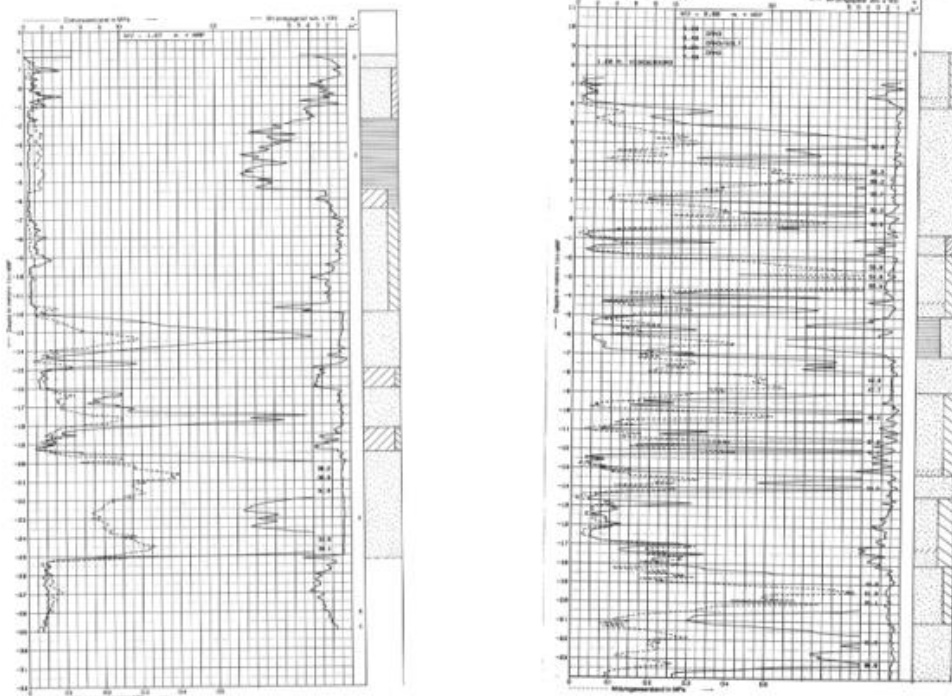
Deze formule is geldig bij een slagkracht groter dan 350 kN. Bij een lagere slagkracht dan 350 is $V_{0,cor}$ gelijk aan V_0 . V_0 en C_{vel} zijn afhankelijk van de ondergrond en worden daarom ook bodemparameters genoemd. De bodemparameters staan vermeld in tabel 5.21 van de CUR 166 deel 2.

Tabel 1; bodemparameters volgens tabel 5.21 van CUR 166 (SBRCURnet, 2016)

Bodemprofiel	50%		95%		99%		α (m ⁻¹)
	V_0	C_{vel}	V_0	C_{vel}	V_0	C_{vel}	
Amsterdam	1,6	0.002	9,0	0.011	18.3	0.023	0
Eindhoven	1,9	0.002	6,7	0.007	11.3	0.012	0
Groningen	1,7	0.002	13,4	0.016	31.6	0.037	0
Den Haag	1,9	0.002	6,7	0.007	11.3	0.012	0
Maasvlakte							
Rotterdam	1,6	0.002	9,0	0.011	18.3	0.023	0
Tiel	1,6	0.002	9,0	0.011	18.3	0.023	0

De bodemparameters zijn afhankelijk van de onderschrijdingsgrens. Een onderschrijdingsgrens van 95% betekent dat de kans dat de voorspelde trilsnelheid overschreden wordt 5% is. Welke onderschrijdingsgrens de geotechnische adviseur gebruikt mag zelf bepaald worden. Meestal wordt de 95%-waarde gebruikt.

De bodemparameters zijn ook afhankelijk van de maatgevende bodemprofiel. De CUR 166 kent 11 verschillende bodemprofielen, Bij de trillingsproeven zijn er echter maar 6 verschillende soorten bodemprofielen getoetst. Hieruit kwamen 3 unieke groepen bodemparameters. Aan de hand van sonderingen vermeld in de CUR 166 deel 1 bijlage C kan de maatgevende bodemprofiel gekozen worden.



Figuur 6; Maatgevende Bodemprofielen Amsterdam en Eindhoven (SBRCURnet, 2016)

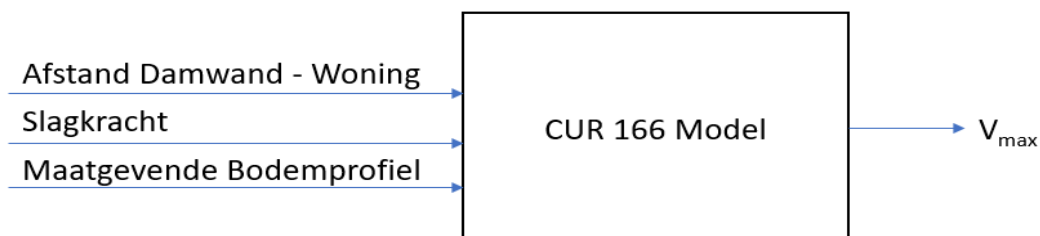
Aan de hand van een maatgevend bodemprofiel kan de bronsterkte bepaald worden.

$C_{0,f}$ zegt iets over de damping van de trillingen naarmate de trillingen verder van de trillingsbron verwijderd raken. Ook zegt het iets over het energieverlies die optreedt bij de overgang van de ondergrond naar de fundering. De formule is als volgt:

$$C_{0,f} = 0.7 \sqrt{\frac{5}{r}}$$

In de bovenstaande berekening is aangenomen dat de fundering de trillingen met 30% reduceert.

Interessanter is om naar het model te kijken vanuit een zeker abstractieniveau. Bij een empirisch model dien je te begrijpen welke variabelen de werking van je model beïnvloeden. In figuur 7 is te zien welke variabelen je dient te weten voordat je een trillingspredictie kunt verrichten.



Figuur 7; CUR 166 model vanuit abstractieniveau

4.2 VOORBEELDBEREKENING

Een 10 meter lange damwand zal trillend geïnstalleerd worden. Op 6 meter afstand bevindt zich een woning. Het intrillen zou mogelijk schade kunnen veroorzaken. De damwand wordt ingetrild met een slagkracht van 1800 kN. Bereken met behulp van het CUR 166 trillingsmodel de trilsnelheid met de grootste kans van voorkomen.

Stap 1: Bepaal maatgevende bodemprofiel

Na de sonderingen uit de omgeving te hebben bestudeerd kunnen we concluderen dat er sprake is van een typerend Amsterdams bodemprofiel. Dit hebben we gedaan door een maatgevende bodemprofiel te kiezen en die te vergelijken met de maatgevende bodemprofielen die te vinden zijn in de CUR 166 deel 1.

Stap 2: Bepaal overschrijdingskans

In de CUR 166 staan 3 overschrijdingskansen vermeld. Een 50, 95 en een 99% waarde. Bij een 95%-waarde is de kans dat deze waarde in praktijk wordt overschreden 5%. ($= 100 - 95$) In de vraagstelling wordt er gesproken over de trilsnelheid met de meeste kans van voorkomen. Dit is de 50%-waarde.

Stap 3: Bepaal bodemparameters

De juiste bodemparameters dienen gehaald te worden uit tabel 5.21 van de CUR 166 deel 2. Dit zijn de volgende parameters:

$$V_0 = 1,6$$

$$C_{vel} = 0,002$$

Stap 4: Bereken de gecorrigeerde brontrilsnelheid

Dit dient gedaan te worden met de volgende formule.

$$v_{0;cor} = v_0 + C_{vel} * (F - 350)$$

Deze formule is alleen geldig bij slagkrachten groter dan 350 kN. Dit is bijna altijd het geval.

$$v_{0;cor} = 1,6 + 0,002 * (1800 - 350) = 4,5 \text{ mm/s}$$

Stap 5: Bereken maximale trilsnelheid aan de woning

Om de gecorrigeerde brontrilsheid om te rekenen naar de maximale trilsnelheid is de volgende formule nodig.

$$v_{max} = v_{0;cor} * C_{0,f}$$

Waarbij $C_{0,f}$ gelijk is aan:

$$C_{0,f} = 0,7 * \sqrt{\frac{5}{r}}$$

$$C_{0,f} = 0,7 * \sqrt{\frac{5}{6}} = 0.64$$

V_{\max} is dus gelijk aan:

$$v_{\max} = 4,5 * 0,64 = 2,88 \text{ mm/s}$$

De verwachtte trilsnelheid is dus gelijk aan 2,88 mm/s.

4.3 OPMERKINGEN CUR 166 MODEL

De CUR 166 Model is simpel en makkelijk uit te voeren. Echter valt er ook veel aan op te merken:

- De achterliggende data waarop deze model is gebaseerd bevat alleen stalen damwanden.
- Niet alle soorten bodemprofielen zijn vertegenwoordigd in de CUR 166.
- Door het zoeken naar een maatgevende sondering en door niet gebruik te maken van de daadwerkelijke conusweerstand wordt er veel informatie weggegooid. Onder andere dit gegeven leidt er tot dat onder dezelfde maatgevende bodemprofiel er een grote variatie is aan gemeten trilsnelheden.
- Het type fundering word niet meegenomen in het bepalen van de trillingsintensiteit. In plaats daarvan hebben woningen met een trillingsgevoelige fundering een strengere grenswaarde.
- De CUR 166 trillingsprognose voor het intrillen van damwanden is gebaseerd op damwandproeven tot aan 14 meter. In praktijk wordt de CUR 166 ook voor langere damwanden gebruikt.
- De CUR 166 gaat ervan uit dat er geen materiaaldemping plaatsvindt. Dit is een uiterst conservatieve aanname.

5. SBR TRILLINGSRICHTLIJN

In Nederland bestaan er richtlijnen over trillingen. Dit wordt de SBR-trillingsrichtlijn genoemd. Deze is uitgebracht door de SBR (Stichting BouwResearch) en bestaat al 28 jaar. De richtlijn bestaat uit drie delen;

- A. Schade aan bouwwerken
- B. Hinder voor personen
- C. Trillingsgevoelige apparatuur

Voor het verrichtten van trillingspredicties voor het voorkomen van schade aan gebouwen dient er dus gebruik gemaakt te worden van richtlijn A. deze richtlijn is in december 2017 gereviseerd.

5.1 SBR TRILLINGSRICHTLIJN A; SCHADE AAN BOUWWERKEN

De SBR Trillingsrichtlijn A bevat de volgende aspecten

- Een procedure voor het meten van trillingen

- Een methode voor het verwerken van meetresultaten
- Een methode voor het beoordelen van de invloed van trillingen met het oog op mogelijke schade aan het bouwwerk of onderdelen daarvan. Hierin maakt de SBR-richtlijn geen onderscheid tussen constructieve en niet constructieve schade.

De trillingen die voortkomen uit het installeren van een damwand wordt door de richtlijn gezien als een stochastische trillingen. De SBR definieert een stochastische trilling als volgt;

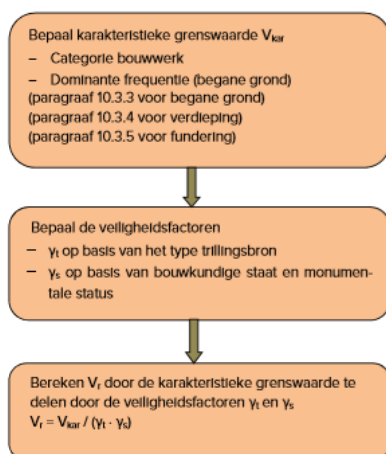
"bij een stochastische trilling wordt de waarde van de amplitude bepaald door het toeval" - (SBR Trillingsrichtlijn A; schade aan bouwwerken, 2016, p. 21)

Als een geotechnische adviseur gebaseerd op de CUR 166 een trillingsvoorspelling heeft verricht dient deze getoetst te worden aan een grenswaarde. Deze grenswaarde dient vanuit de SBR-richtlijn bepaald te worden.

5.2 BEPALING GRENSWAARDE

Uit de trillingsprognose uit de CUR 166 komt een V_{max} naar voren. Dit is een berekende trillingsnelheid. Vervolgens wordt deze getoetst aan een grenswaarde. In deze paragraaf zal ik kort uitleggen hoe deze toetsing in zijn werk gaat.

Een overzicht van de methode om tot de rekenwaarde van de grenswaarde V_r te komen is te zien in figuur 8:



Figuur 8; Flowchart voor het bepalen van de grenswaarde van de trilling (net, 2017)

De SBR kent twee typen gebouw categorieën. (De vro SBR 2 en 3 categorie zijn in de meest recente uitgave in 2017 samengevoegd)

- SBR-Categorie 1

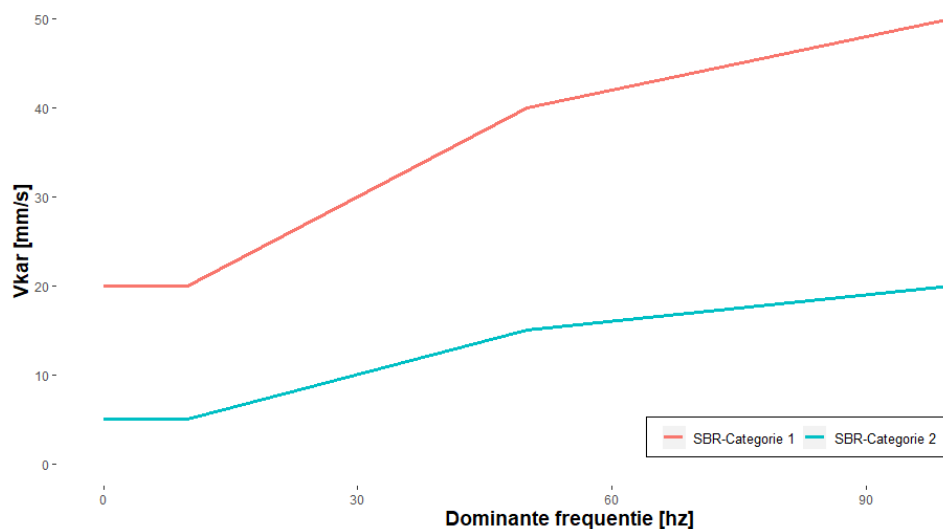
Dit betreft constructies die bestaan uit gewapend beton of hout

- SBR-Categorie 2

Dit betreft constructies die bestaan uit niet-gewapend beton, metselwerk of brosse steenachtige materialen.

Als de SBR-categorie bekend is dan kan de karakteristieke grenswaarde van de trilsnelheid bepaald worden aan de hand van de volgende grafiek.

Grenswaarden van de karakteristieke Grenswaarden ten opzichte van de dominante frequentie



Figuur 9; karakteristieke grenswaarden afhankelijk per SBR-categorie

Volgens de CUR 166 is de dominante frequentie bij het intrillen van een damwand gelijk aan die van het trilblok. De rekenwaarde van de toelaatbare trilsnelheid dient eerst gevonden te worden alvorens de trilsnelheid getoetst kan worden. De rekenwaarde van de trilsnelheid wordt als volgt berekend:

$$V_d = \frac{V_{kar}}{\gamma_t * \gamma_s}$$

γ_t heeft te maken met de type trilling. Is het een kortdurende trilling (explosies), een herhaald kortdurend trilling (heiwerkzaamheden) of een Continue trilling (Intrillen van damwanden). Hierbij horen de volgende veiligheidsfactoren.

Type Trilling	Veiligheidsfactor γ_t
Kortdurend	1
Herhaald Kortdurend	1.5
Continu	2.5

γ_s heeft te maken met de bouwkundige staat en/of de monumentale status van een gebouw. Hierbij horen de volgende veiligheidsfactoren.

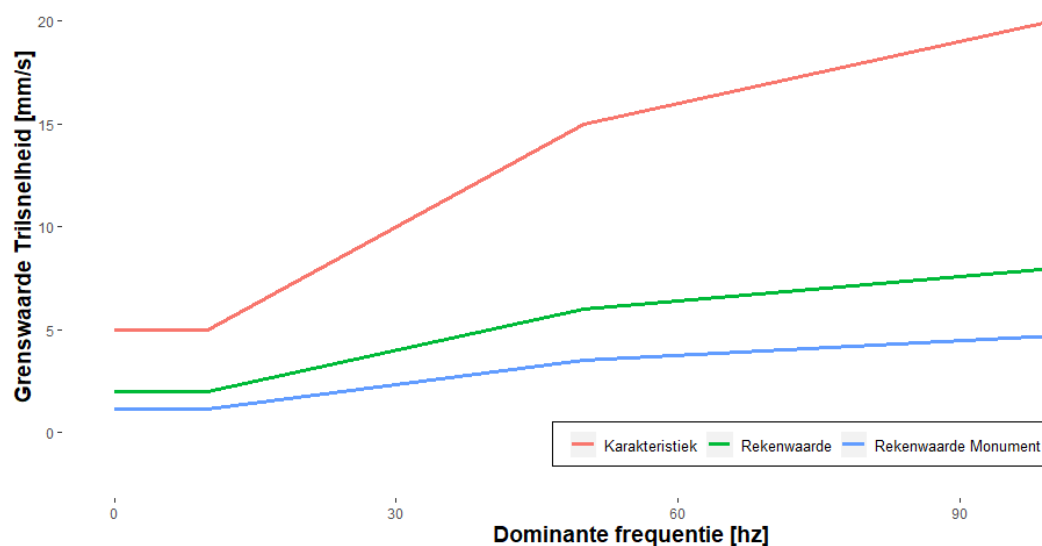
Type Trilling	Veiligheidsfactor γ_s
Bouwkundige Staat: Normaal	1
Monumentale Status: Geen	
Bouwkundige Staat: Gevoelig en/of	1.7
Monumentale Status: Monument	

Binnen onze dataset zijn het SBR-categorie 2 gebouwen. We dienen dan ook naar de blauwe lijn in grafiek 10 te kijken. Bij het intrillen van damwanden is er spraken van een continue trilling. Een veiligheidsfactor γ_t van 2.5 is dan ook van toepassing.

Bij de overgrote meerderheid van de gevallen is de bouwkundige staat normaal en is er geen spraken van een monumentale status. In 9 gevallen van de 153 is er spraken van een monumentale pand.

Cluster Woning	Veiligheidsfactor γ_{totaal}
Normaal	2.5
Monumentale Status	4.25

Daardoor is er uiteindelijk spraken van de volgende grenswaarden:



Figuur 10; Rekenwaarde grenswaarde

Aangezien de monumentale status geen invloed heeft op de gemeten trillingsintensiteit, wordt bij het beoordelen van de nauwkeurigheid van de CUR 166 trillingsprognose de monumentale status buiten beschouwing gelaten.

6. EVALUATIE CUR 166

In hoofdstuk 2 zijn trillingen behandeld. In hoofdstuk 3 is de CUR 166 trillingsprognose behandeld. Aan de hand van deze methode kunnen trillingsprognoses gedaan worden. In hoofdstuk 4 kwam de SBR trillingsrichtlijn A aan bod. Aan de hand van deze richtlijn kan de trillingsprognose getoetst worden.

De CUR 166 trillingsprognose is het meest gebruikte model. Veel ingenieurs twijfelen echter aan de nauwkeurigheid van dit model. In dit hoofdstuk wordt de nauwkeurigheid van het model onderzocht. De CUR 166 kan op twee manieren geëvalueerd worden.

- Numerieke Evaluatie

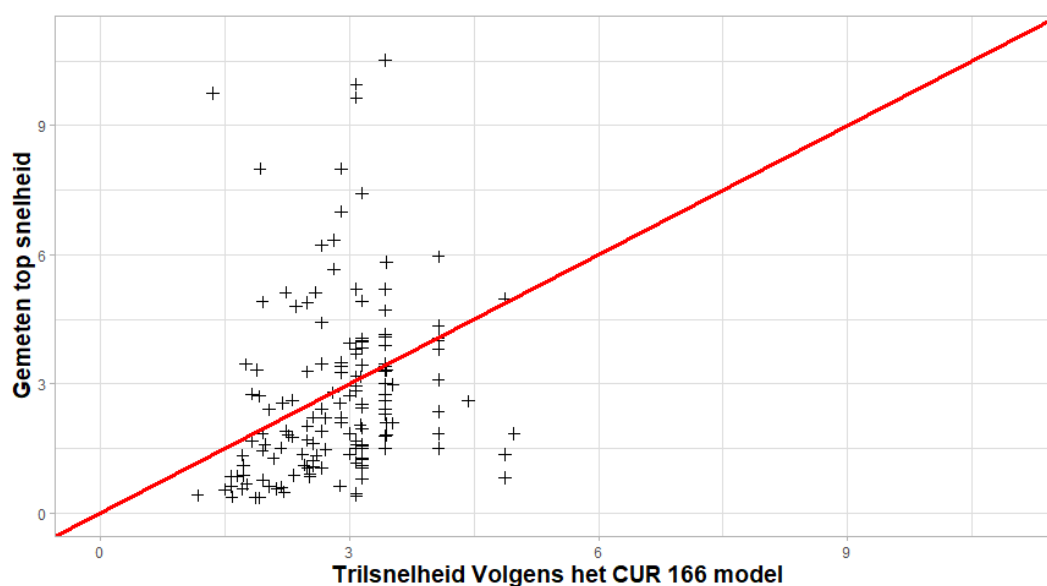
Dit houdt in dat we de gemeten trillingsintensiteit gaan vergelijken met de berekende trillingsintensiteit. We gebruiken hierbij de 50% overschrijdingskans omdat dit het dichtst bij de werkelijkheid zit. Ook kunnen we gaan kijken of het aantal overschrijdingsgevallen wel overeenkomen met de overschrijdingskansen van 50%. Hetzelfde kunnen we doen met de overschrijdingskansen van 95% waar de trillingsprognose vaak op gebaseerd is.

- Categorische Evaluatie

Bij de categorische evaluatie zullen we gaan onderzoeken of met behulp van de CUR 166 betrouwbaar kan worden voorspeld of de SBR-richtlijn wordt overschreden.

6.1 NUMERIEKE EVALUATIE CUR 166

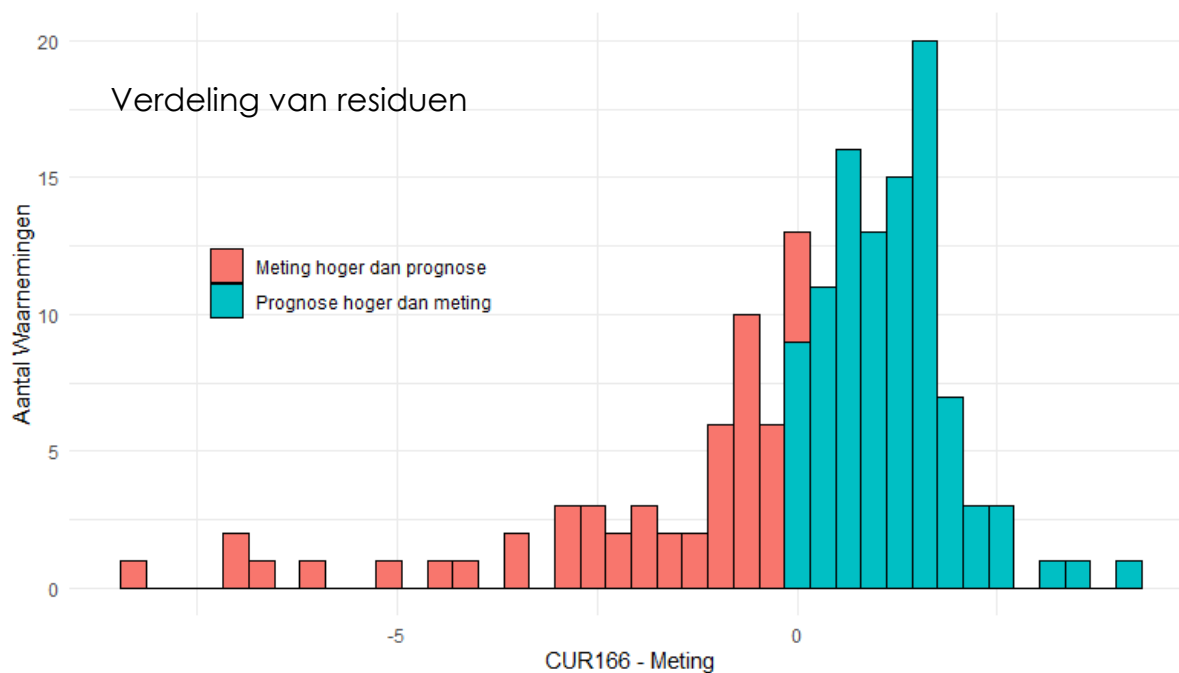
Voor 150 cases is de 50%-waarde van de CUR bepaald. Als we gaan kijken naar de relatie tussen de voorspelling van de CUR 166 en de gemeten trillingsintensiteit zien we het volgende.



Figuur 11; Spreidingsdiagram V_{gemeten} vs. V_{CUR}

In de bovenstaande grafiek is een spreidingsdiagram te zien met de voorspelde trilsnelheid op de horizontale as en de gemeten topsnelheid op de verticale as. Als een voorspelling perfect is dan is de voorspelling gelijk aan de gemeten waarde. In dat geval ligt de punt dan exact op de rode diagonale lijn.

Het verschil tussen het CUR-model en de meting is een indicatie van de nauwkeurigheid van het model. Deze verschil is te zien in de volgende afbeelding.



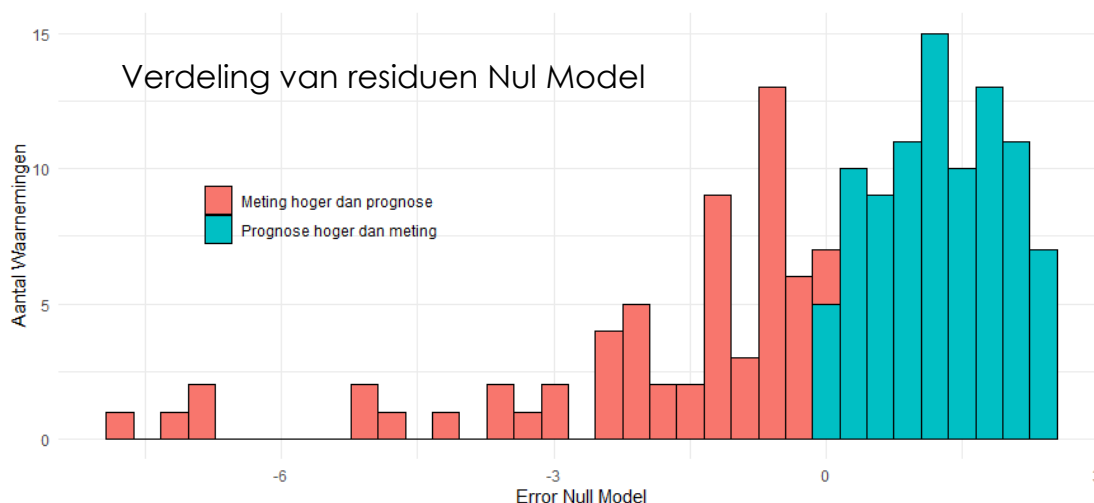
Figuur 12; Verdeling van residuen CUR 166 model

Mean Absolute Error¹ is gelijk aan 1.45 mm/s. Dit betekent dat de CUR 166 er met gemiddeld 1.45 mm/s naast zit.

De R^2 is gelijk aan 0,13. Dit betekent dat 13% van de variantie door het model wordt verklaard. Om verder te illustreren hoe onnauwkeurig de punt voorspelling is kunnen we de nauwkeurigheid gaan vergelijken met het nul-model.

Als we het nul model gebruiken voor het maken van voorspelling krijgen we de volgende residuen. Een uitleg van een nul model wordt gegeven in bijlage A van dit rapport.

¹ Meer info over MAE, R^2 en NUIS te vinden in bijlage A



Figuur 13; Verdeling van residuen van NULL model

In dit geval is de Mean Absolute Error gelijk aan 1,53 mm/s. Dit komt in de buurt van de CUR 166 model waarbij de MAE gelijk is aan 1.45 mm/s is. De hele CUR berekening kan vervangen door altijd de waarde 2,76 mm/s te voorspellen en hierbij verlies je weinig nauwkeurigheid.



Figuur 14; Visualisatie Nauwkeurigheid CUR 166 model ten opzichte van NULL model

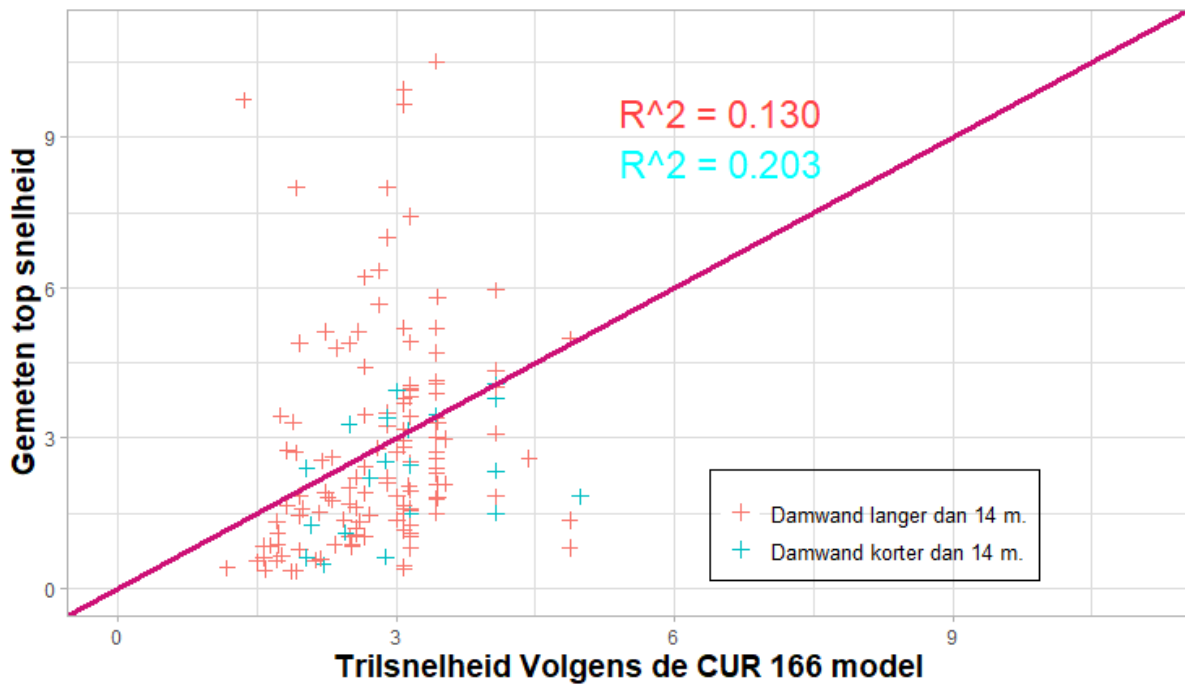
In de bovenstaande afbeelding wordt dit principe gevisualiseerd. We zien een getallenlijn met links het nul-model met een MAE gelijk aan 1,53. Rechts zien we een perfect model met een MAE gelijk aan 0. Het CUR 166 model laat nauwelijks verbetering zien ten opzichte van het nul-model.

Lengte Damwand

In de CUR 166 staat vermeld dat voorspellingen van trilsnelheden alleen gedaan kunnen worden op damwanden met een lengte korter dan 14 meter. De reden hiervoor is dat ten tijde van het opstellen van de empirische formules er geen of weinig trillingsproeven aanwezig waren van langer dan 14 meter. Dit was de juiste beslissing omdat bij empirisch onderzoek extrapoleren kan leiden tot groffe fouten. Maar wel moet vermeld worden dat deze restrictie afkomstig is uit tekort aan kennis. Het zou dus kunnen zijn dat het model net zo nauwkeurig is bij damwanden korter dan 14 meter als damwanden langer dan 14 meter.

Ook zien we in de praktijk dat de CUR 166 trillingsprognose ook gebruikt wordt bij damwanden langer dan 14 meter. De vraag die we ons zelf moeten stellen is of dit wel terecht is. Gebruiken de praktijk een model op een situatie waar het niet voor bedoeld is? Of kunnen we vaststellen dat de CUR 166 even nauwkeurig is bij damwanden langer dan 14 meter? In het laatste geval kan de beperkingen die de CUR 166 oplegt genegeerd worden.

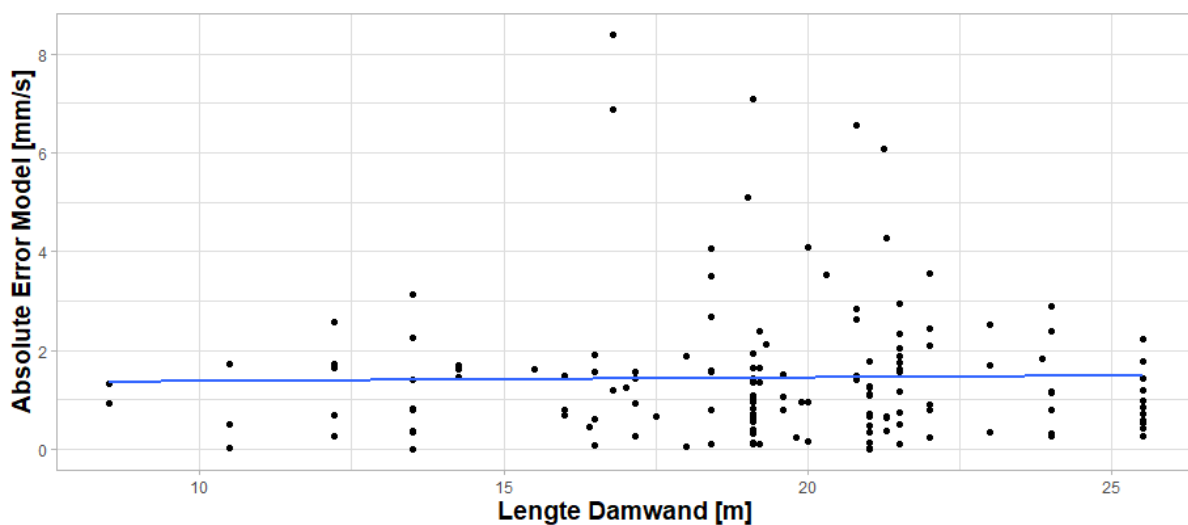
In de Geobrain dataset zijn er twintig damwandproeven die kleiner zijn dan 14 meter. De overgrote meerderheid (130) is dat niet. Wat als we gaan kijken naar de analyse van de nauwkeurigheid opgesplitst in twee delen?



Figuur 15; Spreidingsdiagram V_{gemeten} vs. V_{CUR} gekleurd naar damwandlengte

In eerste instantie lijkt het dat damwanden die kleiner zijn beter voorspeld worden. Bij damwanden kleiner dan 14 meter zien we een R^2 van 20,3%. Dit is een verbetering als we kijken naar damwanden langer dan 14 meter. (13%)

Ook lijkt het dat grote groffe fouten alleen plaatsvinden bij damwanden langer dan 14 meter. Deze hypothese zullen we in de volgende afbeelding onderzoeken.

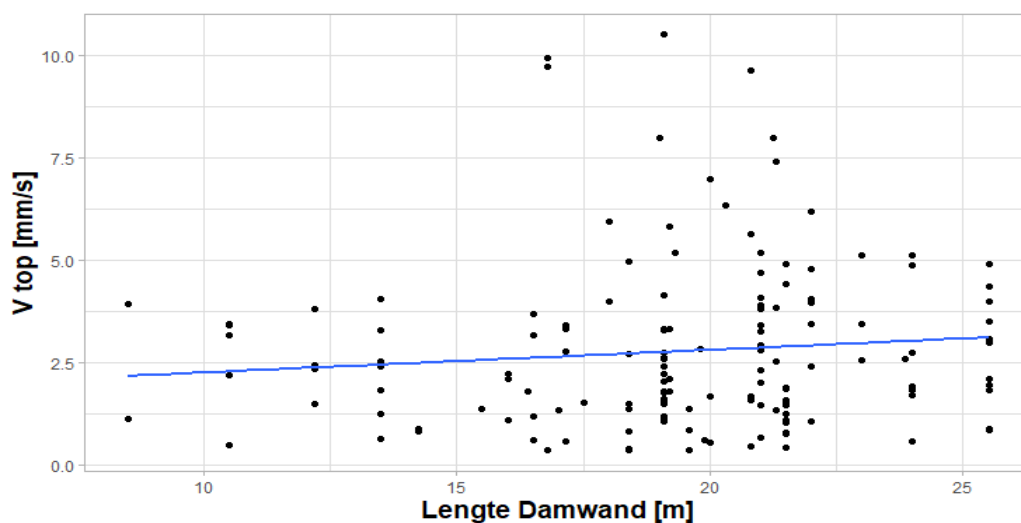


Figuur 16; Spreidingsdiagram Error CUR model vs. Damwandlengte

Deze aanname wordt ondersteund door de data. De grote uitschieters vinden alleen plaats bij damwanden langer dan 15 meter. In gemiddelde zin lijkt hier echter geen verband in te zitten. Dit is aan de blauwe trendlijn te zien.

Ook dient er opgemerkt te worden dat er slechts 20 damwandproeven aanwezig zijn met een damwandlengte korter dan 20 meter. Het aantal uitschieters kan simpelweg te maken hebben met het verschil in aantal waarnemingen. In bijlage A is er een statistische toetsing uitgevoerd. Hieruit bleek dat het geobserveerde verschil in nauwkeurigheid tussen damwanden korter dan 14 meter en langer dan 14 meter niet statistisch significant is.

De volgende afbeelding geeft een andere zienswijze op het probleem;



Figuur 17; Spreidingsdiagram V_{top} vs. Damwandlengte

Een langere damwand zorgt voor vaker voor meer trillingen. En zoals we in figuur 17 zagen zijn het vooral de grote trillingen die slecht voorspeld worden. Ook kan een langere damwand duiden op een langere installatieduur.

Ondanks de beperkingen die vermeld staan in de CUR 166 wordt deze methode in de praktijk ook gebruikt voor damwanden langer dan 14 meter. Het is een beperking zonder praktische en toegankelijk alternatief.

Uit deze analyse kunnen we twee kanten op.

1. De lengte van een damwand heeft geen invloed op de nauwkeurigheid van het CUR model.

In dit geval laat dit hoofdstuk zien hoe onnauwkeurig het CUR model is.

2. De lengte van een damwand heeft wel invloed op de nauwkeurigheid van het CUR model.

Zelfs al heeft de lengte van een damwand invloed, dan dient er alsnog een model gemaakt te worden die ook gebruikt kan worden voor damwanden die langer zijn dan 14 meter.

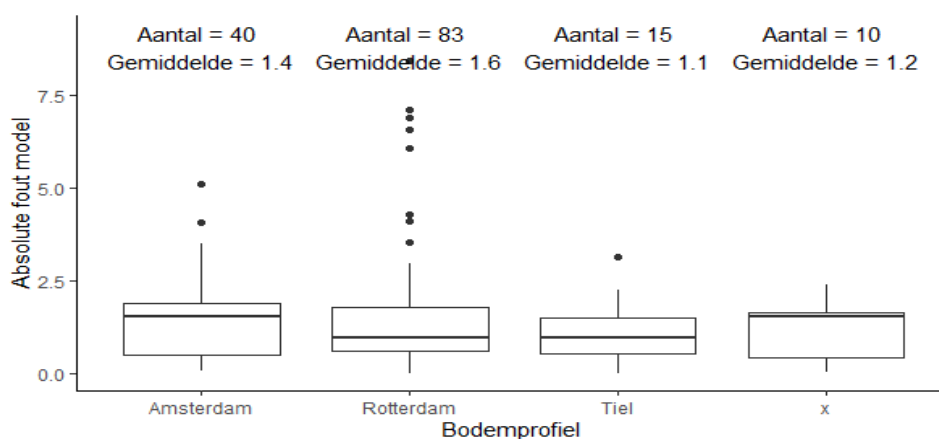
ERROR CORRELATIE ANALYSE

Een aantal interessante vragen die gesteld kunnen worden zijn;

- Is het type damwand gecorreleerd aan de fout van het model?
- Is het model vaker onnauwkeurig bij sommige bodemprofielen?

Een dergelijke wordt een foutenanalyse genoemd. Het zou zo kunnen zijn dat het model in bepaalde scenario's onnauwkeurig is. Zo kan het model verbeterd worden door het simpelweg veranderen van de restricties.

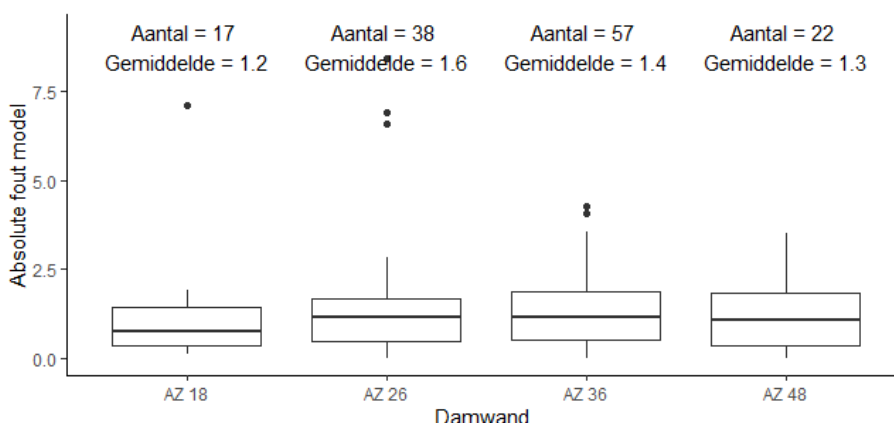
Bodemprofiel



Figuur 18; Boxplot absolute error model per bodemprofiel

Als we kijken naar de boxplot² hierboven zien we geen verband tussen de bodemprofiel en de absolute fout van het model. Hierboven zijn de boxplots te zien van alle bodemprofielen die meer dan 9 voorkomen in de dataset. Bodemprofiel x betekent een onbekend bodemprofiel. Dit is een bodemprofiel die bij geen enkele bodemprofiel past. De werking van een boxplot wordt in bijlage A verder toegelicht.

Type Damwand



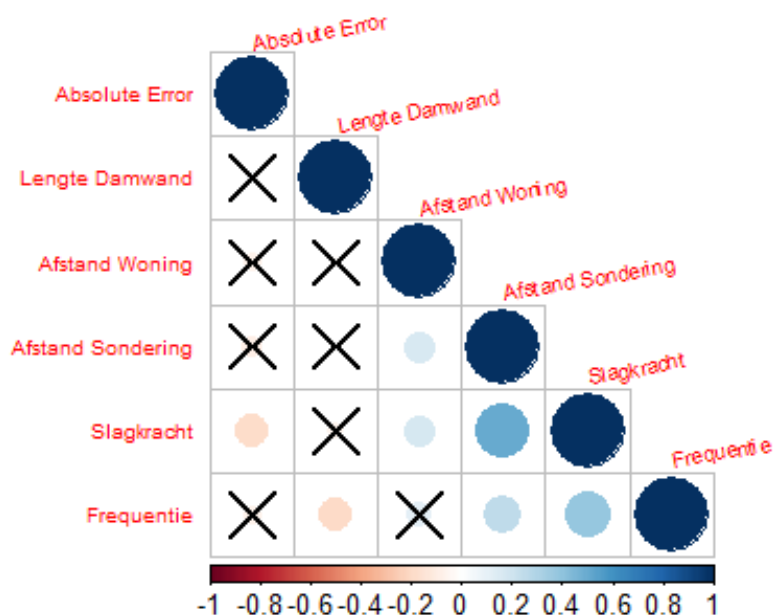
Figuur 19; Boxplot absolute error model per type damwand

² Informatie over Boxplots is te vinden in bijlage A

Als we kijken naar het type damwand zien we dat de mediaan van de error van het model gelijk blijft bij alle typen damwanden. Ook hier zijn damwandtypen die niet vaak voorkomen in de dataset verwijderd.

Numerieke Variabelen

Daarnaast kunnen we kijken welke numerieke variabelen invloed hebben op de error van het model. Dit doen we met een correlatieplot.



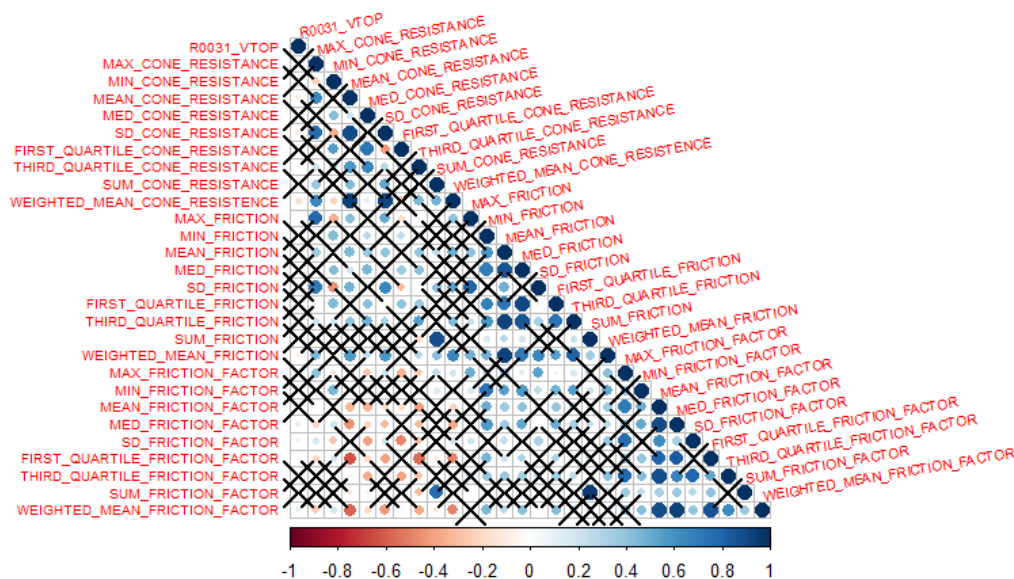
Figuur 20; Correlatieplot tussen fout CUR 166 systeemparemeters

De correlatieplot laat veel informatie zien. Waaronder;

- De grootte van de relatie wordt aangegeven met de grootte van de stip. In dit geval wordt er gebruik gemaakt van een Spearman-rank correlatie.
- De richting van de relatie wordt aangegeven met de kleur. Rood is negatief en blauw is positief.
- Statistisch insignificant relatie zijn doorgekruist.

Bij het analyseren van de correlatieplot dient er in dit geval alleen gekeken te worden naar de eerste kolom. Dit betreft namelijk de relaties van de variabelen met de fout van het model. Uit deze correlatieplot kunnen we opmerken dat weinig variabelen noemenswaardig correleren met de fout van het model. De slagkracht laat een correlatiecoëfficiënt van -0,29 zien. Hoe hoger de slagkracht des te nauwkeuriger wordt het model. Ook heeft het een hoge statistische significantie. (Een p-waarde kleiner dan 0,001)

Ook kunnen we dezelfde grafiek maken maar dan de absolute fout van het model proberen te correleren met informatie uit de sondering.

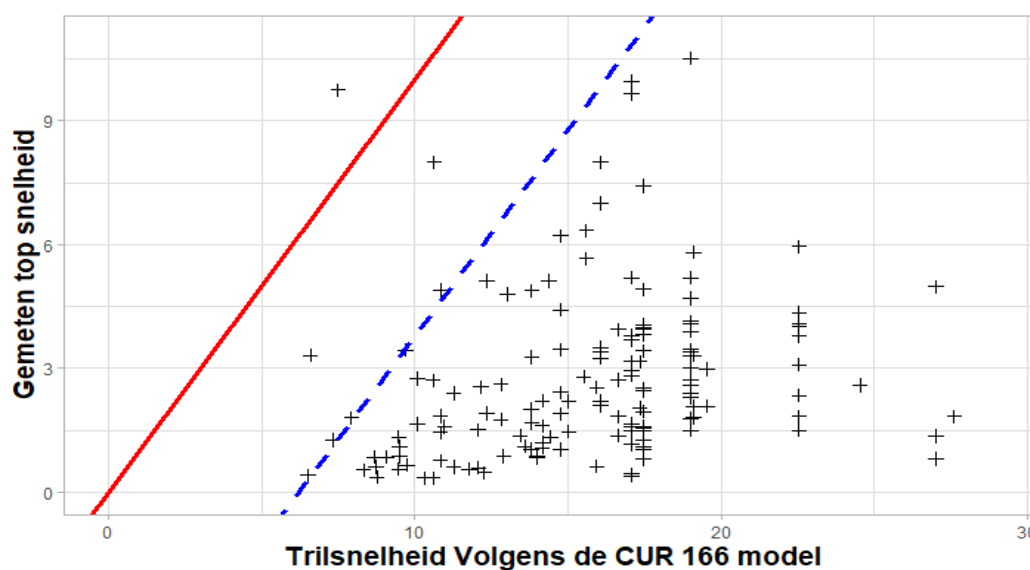


Figuur 21; Correlatie tussen error model met bodemtypen

Ook hier dient er alleen gekeken te worden naar de eerste kolom. Hier zien we geen enkele significante correlatie. Het model voorspelt dus niet beter in bepaalde ondergrondtypen. We zien geen sterke correlatie met een bodemparameter. We kunnen dus niet stellen dat het CUR 166 model slechter presteert in bepaalde ondergrondtypen.

5% OVERSCHRIJDINGSKANS

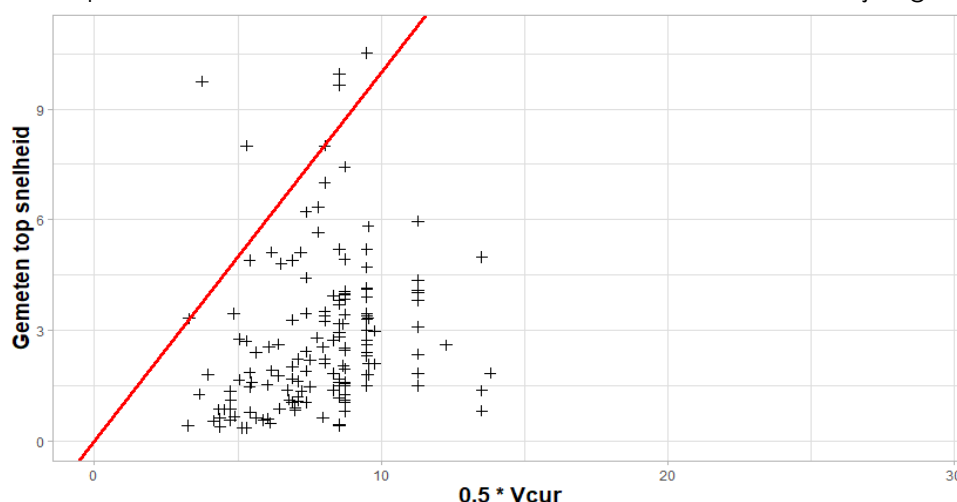
Aan het begin van deze analyse is er gekeken naar de relatie tussen de voorspelde en de gemeten topsnelheid. Bij de voorspelde trilsnelheid zijn we gaan kijken naar de 50% overschrijdingskans omdat dit het dichtst bij de waarheid ligt. Echter bij een trillingspredictie wordt niet gebruik gemaakt van een overschrijdingskans van 50%, maar een hogere waarde zoals de 95%-waarde. Dit wordt gedaan om veiligheid in te bouwen in het model.



Figuur 22; 95% onderschrijdingskans CUR vs. Gemeten topsnelheid

Als we kijken naar de 5% overschrijdingskans van de CUR 166 zien we een interessant verhaal. Als de punten boven de rode lijn zouden zitten dan zou dat betekenen dat voorspelde trilsnelheid overschreden is. Bij een 5% overschrijdingskans zou je verwachten dat rond de 7 waarnemingen boven de rode lijn zouden liggen. Dit is echter niet het geval. Bij slechts een meting is de CUR 166 trillingsprognose overschreden. Dit laat zien dat de trillingsprognose volgens de CUR 166 conservatief is. De ruimte tussen de rode en de blauw lijn is ruimte voor een mogelijke efficiëntieslag.

De vraag die we onszelf zouden kunnen stellen is, hoever kunnen we de CUR 166 versoepelen zodat het ook echt overeenkomt met de 5% overschrijdingskans.



Figuur 23; 50% van de 95%-waarde

In de bovenstaande afbeelding wordt de gemeten topsnelheid vergeleken met 50% van de CUR 166 prognose. In 5 gevallen is er sprake van een overschrijding. Door de 95%-waarde met een factor 2 te verlagen krijgen we een waarde die een benadering is voor de 95%-waarde.

CONCLUSIE: DE CUR 166 IS ONNAUWKEURIG MAAR VEILIG

Uit de numerieke evaluatie konden we concluderen dat de trillingsprognose uit de CUR 166 onnauwkeurig is. Dit is echter geen geheim. Hier wordt in de CUR 166 het volgende over gezegd:

“De onzekerheid van de trillingspredictie kan in kaart worden gebracht door deze te valideren met behulp van metingen. Uit de praktijk blijkt dat de spreiding van verschillende metingen onder vrijwel identieke omstandigheden groot is. Dit heeft als consequentie dat een voorspelling van een trillingsniveau nooit exact kan zijn. In het algemeen zal daarom een bepaalde bandbreedte worden gegeven. Een veiligheidsfactor van 2 is daarbij niet uitzonderlijk.” (CUR 166 deel 2 hoofdstuk 5.8.3, p. 572)

De wijzen waarop trillingen zich door de grond verspreiden is een complexe materie. Zoals in veel gevallen in de geotechniek is het gedrag van een systeem afhankelijk van de materiaaleigenschappen. Het zijn echter deze materiaaleigenschappen die onbekend zijn door grote heterogeniteit in de ondergrond.

Zoals de CUR 166 ook al benadrukt, dient de voorspelling van de trillingsintensiteit niet gezien te worden als een betrouwbaar getal. Sterker nog we zien in de analyse iets dat meer lijkt op een puntenwolk. Het model wordt dusdanig onnauwkeurig dat er weinig wijsheid aan te ontleiden valt. De wijzen waarop de CUR dat oplost is door strenge veiligheidsfactoren toe te passen. Een onnauwkeurig model oplossen met hoge veiligheidsfactoren brengt echter de nodige problemen met zich mee;

- Situaties waar het intrillen van damwanden niet voor problemen zou zorgen, wordt onnodig geclassificeerd als een risicogebied. Dit zorgt ervoor dat aannemers onnodig moeten grijpen naar duurdere technieken zoals drukken.
- Risicogebieden worden niet geclassificeerd wat kan leiden tot onnodig schade.

Als we namelijk gaan kijken naar de 95% overschrijdingsgrens van de trilsnelheid zien we dat in slechts 1 geval de gemeten trilsnelheid hoger ligt dan de voorspelde trilsnelheid. Bij 95% overschrijdingsgrens zouden er ongeveer 7 (5% van 153) overschrijdingen moeten hebben plaatsgevonden. Dit laat zien dat de 95% overschrijdingsgrens een conservatief gekozen waarde is. De 95%-waarde komt eerder overeen met de 99% waarde. In het onderzoek is er laten zien dat er een grote efficiëntieslag te halen valt bij het versoepelen van het CUR-model. De 95%-waarde kan met een factor 2 worden gereduceerd om te komen tot de empirisch bepaalde 95%-waarde.

6.2 CATEGORISCHE EVALUATIE

De numerieke evaluatie is interessant om de nauwkeurigheid van het model te analyseren. Vanuit praktisch perspectief is het belangrijker om te weten of voldaan zal worden aan de SBR-richtlijn. Kan aan de hand van de CUR voorspelling iets gezegd worden over of de SBR-richtlijn overschreden wordt?

De CUR 166 doet een puntpredictie van de verwachte maximale trilsnelheid. Vervolgens wordt deze trilsnelheid getoetst aan de grenswaarden volgens de SBR-richtlijn. Bij het categorisch evalueren van een model gaat het om de volgende vier scenario's.

1. CUR 166 voorspelt dat de SBR-richtlijn overschreden wordt en in praktijk klopt dat.
(TRUE POSTIVE)
2. CUR 166 voorspelt dat de SBR-richtlijn overschreden wordt en in praktijk klopt dat niet.
(FALSE POSITIVE)
3. CUR 166 voorspelt dat de SBR-richtlijn niet overschreden wordt en in praktijk klopt dat.
(TRUE NEGATIVE)
4. CUR 166 voorspelt dat de SBR-richtlijn niet overschreden wordt en in praktijk klopt dat niet.
(FALSE NEGATIVE)

Als we dat in tabel form visualiseren ziet het er als volgt uit:

Tabel 2; Voorbeeld Confusion Matrix

We moeten ons ook inzien dat de twee fouten niet gelijkwaardig zijn:

	Predicted Positive	Predicted Negative
Actual Positive	Correct	False Negative
Actual Negative	False Positive	Correct

False Negative: Als de CUR 166 zegt dat de SBR-richtlijn niet overschreden zal worden maar in praktijk gebeurt dat toch dan is er sprake van een False Negative. Dit kan de volgende gevolgen hebben.

- Dit kan in de praktijk ervoor zorgen dat de aannemer zijn werk stil moet leggen omdat er een te hoge trilsnelheid gemeten wordt. Eerst moeten er trillingsarme materieel naar de bouwplaats komen. Het drukken van damwanden duurt ook langer dan trillen. Projecten kunnen daardoor lange tijd vertraagd raken.
- Grenswaarden worden overschreden zonder dat de aannemer dat door heeft. Een gebouw wordt dan lange tijd blootgesteld aan een onverantwoorde hoeveelheid aan trillingen. Bij schade is een aannemer verantwoordelijk.
- Ook wordt de kans op het bezwijken van een constructie ten gevolge van constructieve schade groter. Dit is echter zeer onwaarschijnlijk. Meestal komt dit tot uiting in cosmetische (niet constructieve schade) schade. Denk hierbij aan kleine scheuren in het metselwerk of gebroken tegels.

False Positive: Als de CUR 166 zegt dat de SBR-richtlijn niet overschreden zal worden maar in praktijk gebeurt dat toch dan is er sprake van een False Negative.

- Damwand wordt gedrukt wanneer dat niet nodig was. Een damwand drukken is duurder en langzamer dan trillen. Dit kan zorgen voor een onnodig lange bouwduur en hoge kosten.

De CUR 166 vindt een False Negative erger dan een False Positive. Vandaar dat de CUR 166 alleen opties heeft voor hoge onderschrijdingskansen. De geotechnische ingenieur doet een trillingsprognose meestal gebaseerd op de 95%-waarde. De kans is 95% dat de numerieke puntvoorspelling van de CUR niet overschreden wordt. Volgens deze overschrijdingskansen is een False Negative schaars.

De CUR voorspellingen zijn gebruikt om een SBR-toetsing op uit te voeren. In de volgende tabel zijn de frequentie van de vier mogelijkheden aangegeven.

Tabel 2; Confusion Matrix 50%-waarde CUR

50%-waarde		Werkelijk	
		Overschrijding	Geen Overschrijding
Voorspeld	Overschrijding	5	8
	Geen Overschrijding	22	115

In de bovenstaande tabel is laten zien dat in de 27 gevallen waarbij de SBR-overschreden werd dit in slechts 5 gevallen (19%) als zodanig door de CUR is voorspeld. Ook kunnen we

zien dat wanneer de CUR voorspeld dat de SBR-richtlijn niet overschreden wordt dit in 16% van de gevallen niet klopt. Figuur 5 toont dezelfde analyse, maar dan met de 95%-waarde.

Tabel 3: Confusion Matrix

95%-waarde		Werkelijk	
		Overschrijding	Geen Overschrijding
Voorspeld	Overschrijding	27	123
	Geen Overschrijding	0	0

De 95%-waarde voorspelt in alle gevallen een overschrijding van de SBR-richtlijn, terwijl de overschrijding in slechts 18% van de gevallen optreedt.

CONCLUSIE; CUR 166 ZEGT WEINIG OVER EEN OVERSCHREIDING VAN DE GRENSWAARDE

De analyse van de nauwkeurigheid leert ons de volgende drie dingen.

- De voorspelling van de CUR laat geen sterk verband zien met de gemeten topsnelheid.
- De 95%-waarde wordt in minder dan 1% van de gevallen overschreden en is een factor 2 te conservatief.
- Het CUR-model is niet in staat om een overschrijding van de SBR-norm te voorspellen.

Als ingenieur die bezig is met trillingspredicties dien je te beseffen dat de CUR-waarden niet beschouwd moeten worden als de ultieme werkelijkheid, maar als een indicatie die iets nauwkeuriger is dan het gemiddelde.

De onnauwkeurigheid van het CUR-model laat het volgende zien.

- De noodzaak van monitoring tijdens de bouw.
- De noodzaak naar meer onderzoek over trillingen.
- De noodzaak van het empirisch valideren van modellen en indirect daarmee ook het verzamelen van data.

Notitie:

Voor een mogelijke publicatie is er gekozen om de gehele CUR 166 evaluatie samen te vatten in twee bladzijden. Deze samenvatting is te vinden in bijlage G.

7. CPT AVERAGING TECHNIQUES

Iedere sondering kent minimaal drie parameters. Dit zijn:

- Conusweerstand q_c
- Schachtwrijving f_s
- Wrijvingsgetal $R_f = f_s/q_c$

Deze drie waarden worden iedere twee centimeter gemeten. Bij een sondering van 30 meter praat je dan al gauw over 4500 getallen. We kunnen geen modellen maken die afhankelijk zijn van 4500 inputwaarden. De kunst is om 4500 dimensies te reduceren tot een paar nieuwe waarden die voorspellend zijn voor de trilsnelheid. Dit zullen we doen met statistische manipulaties. Vervolgens zullen de meest invloedrijke parameters vinden met behulp van correlatiecoëfficiënten. Meer informatie over correlatiecoëfficiënten is te vinden in bijlage B.

7.1 NAUWKEURIGHEID VS. SIMPLICITEIT

Bij het creëren van deze parameters is het belangrijk om deze parameters simpel te houden. Een maximale of minimale conusweerstand kan makkelijk afgelezen worden van een grafiek. De standaarddeviatie van de conusweerstand is praktisch onmogelijk om nauwkeurig af te lezen uit een sondering. Daarom zal bij het kiezen van de juiste parameter niet alleen gekeken worden naar het belang van parameters maar ook naar de simpliciteit. Een reeks getallen kan met de volgende statistische manipulaties worden samengevat:

1. Maximum
2. Minimum
3. Gemiddelde
4. Mediaan
5. Standaard Deviatie
6. Eerste Kwartiel
7. Derde Kwartiel
8. Som

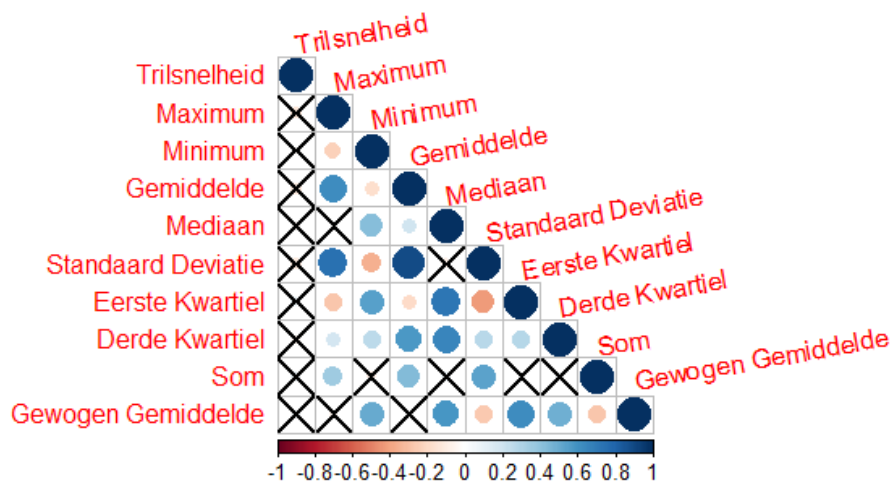
Al deze manipulaties worden berekend over de drie grondparameters. Dit zorgt voor 24 parameters. Daarnaast is er nog een gewogen gemiddelde.

De gewogen gemiddelde zal ook toegepast worden op Conusweerstand, Schachtwrijving, en Wrijvingsgetal. De gewogen gemiddelde zal de afstand van de onderkant van de sondering als weging nemen. Dit betekent dat hoe dieper je naar beneden gaat in de sondering, des te minder invloed het zal hebben. Dit heeft te maken met het feit dat hoe dieper je in de sondering komt, des te verder het van het gebouw af ligt. Met drie extra parameters komt het totaal aantal parameters uit op 27 parameters. Ook 27 parameters is echter te veel. We dienen dus een dimensiereductie toe te passen.

7.2 DIMENSIE REDUCTIE

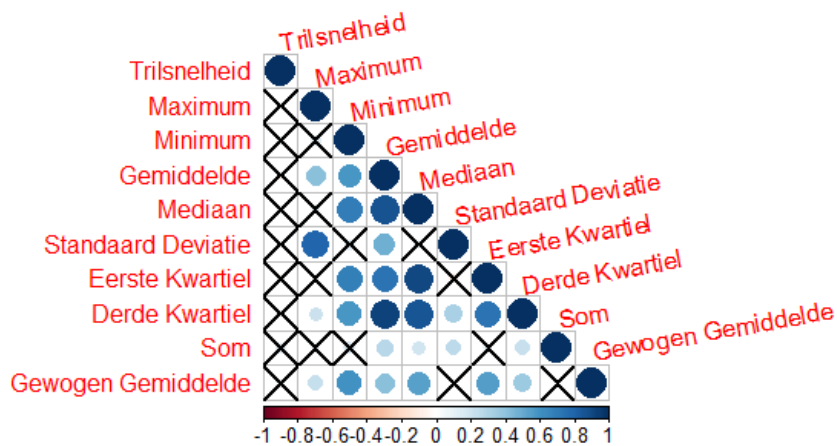
We dienen dus te bepalen welke van de 27 grondparameters invloedrijk zijn op de trilsnelheid. Dit doen we door de correlatie tussen de parameter en de trilsnelheid te bekijken. Dit zullen we doen met dezelfde correlatieplotten als het vorige hoofdstuk. We zullen de correlatieplotten opdelen in drie delen om zo het overzicht te bevorderen.

1. Conusweerstand



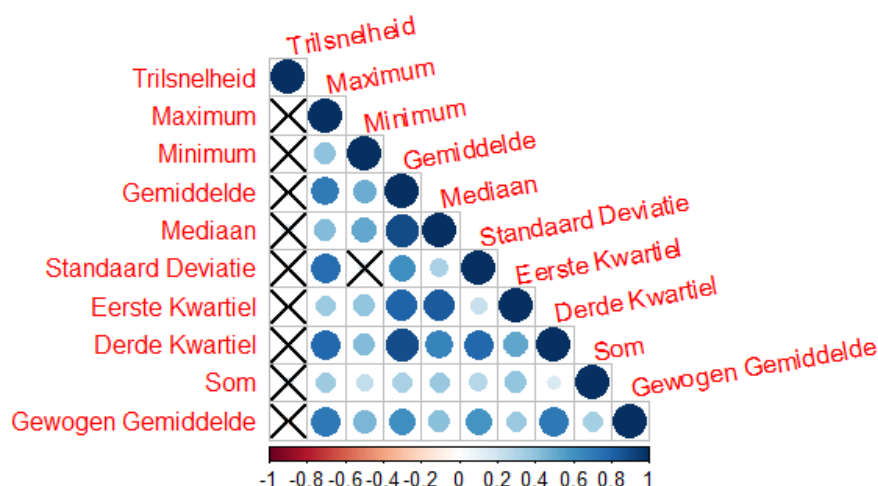
Figuur 24; Correlogram Conusweerstand vs. bodemparameters

2. Schachtwrijving



Figuur 25; Correlogram Schachtwrijving vs. bodemparameters

3. Wrijvingsgetal



Figuur 26; Correlogram Wrijvingsgetal vs. bodemparameters

Bij het analyseren van deze correlogrammen dient er gekeken te worden naar de eerste kolom. De grootte van de cirkel laat de correlatiegrootte tussen de parameter en de trilsnelheid zien. Hoe hoger dit getal, des te invloedrijker is de parameter. De kleur geeft de richting aan. Een rode kleur laat een negatief verband zien tussen de twee grootheden. Dus als de een omhoog gaat, gaat de ander omlaag en vice versa. Bij een positieve correlatiecoëfficiënt gaan de twee grootheden gelijk op. Een kruis door de rechthoek heen laat zien dat er geen statistisch significante correlatie is.

7.3 CONCLUSIE

Deze correlogrammen laten zien dat er geen statistisch significant verband te vinden is tussen de trilsnelheid en de bodemparameters uit de sondering.

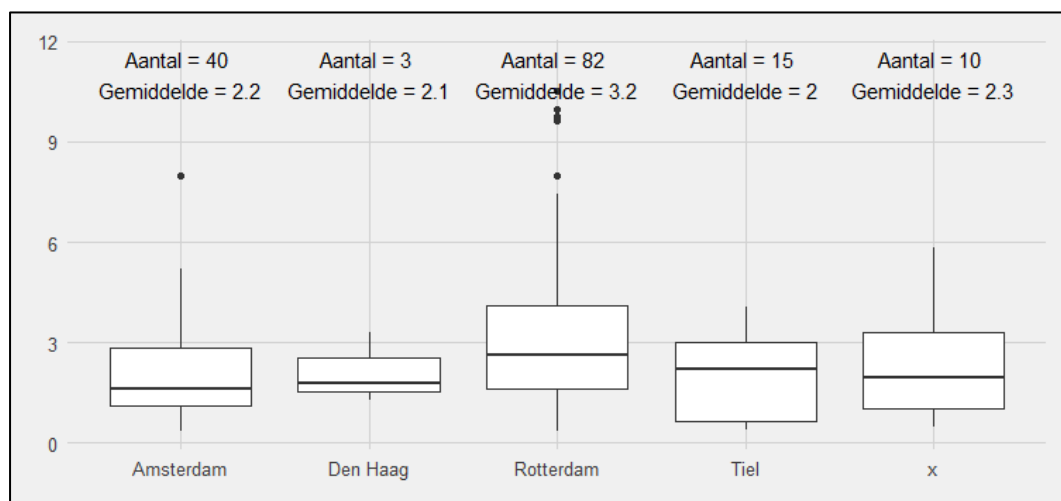
8. FEATURE IMPORTANCE

In het vorige hoofdstuk zagen we dat er geen invloedrijke bodemparementers zijn die invloedrijk zijn in het voorspellen van de trilsnelheid. In dit hoofdstuk gaan we verder met het zoeken naar invloedrijke variabelen. De variabelen die onderzocht zullen worden zijn als volgt;

- Frequentie trilblok [hz]
- Slagkracht [kN]
- Afstand Damwand Woning (r) [m]
- $1/r$ [m^{-1}]
- $1/\sqrt{r}$ [$m^{-0,5}$]
- Afstand Damwand Sondering [m]
- Oppervlakte Damwand [mm^2]
- Bodemprofiel [-]
- Lengte Damwand [m]

8.1 CORRELATIE

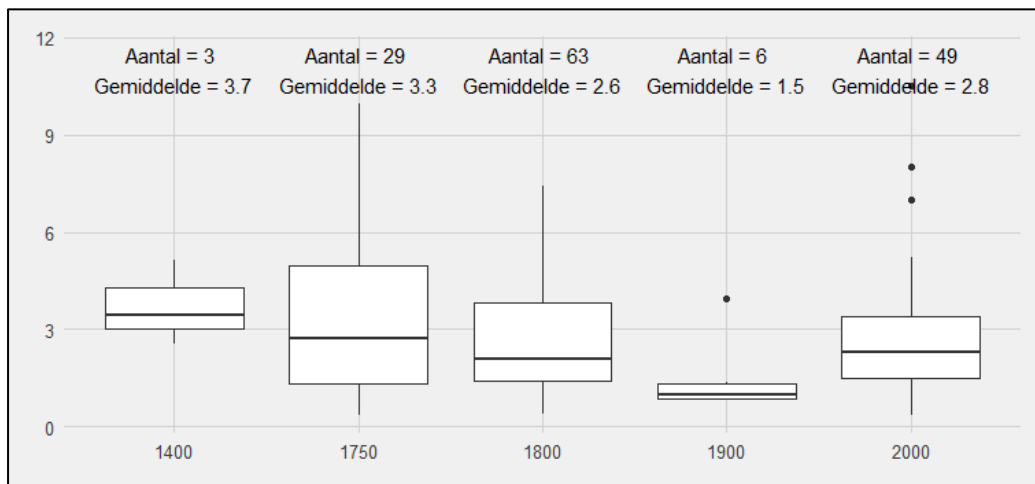
Invloed Bodemprofiel



Figuur 27; Boxplot Trilsnelheid vs. Bodemprofiel

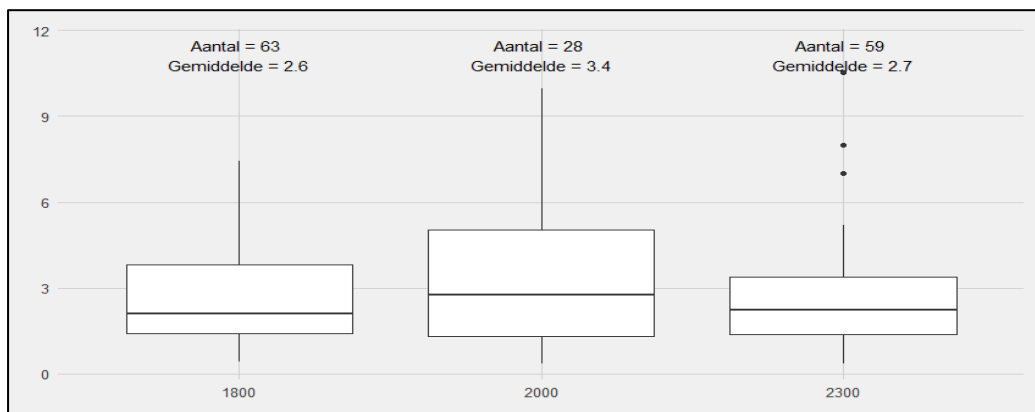
Op basis van de dataset kunnen we concluderen dat het maatgevende bodemprofiel weinig invloed heeft op de trillingsintensiteit. De enige noemenswaardige bodemprofiel is Rotterdam. Dit laat een hoger gemiddelde trilsnelheid zien. Bodemprofiel Rotterdam bevat een ongeveer 2 meters lange veenlaag kort onder de grond. Het effect hiervan dient verder onderzocht te worden in vervolgonderzoek. Ook dienen andere bodemprofielen onderzocht te worden die niet in de dataset aanwezig zijn.

Invloed Slagkracht



Figuur 28; Boxplot Trilsnelheid vs. Slagkracht

Invloed Frequentie Trilblok

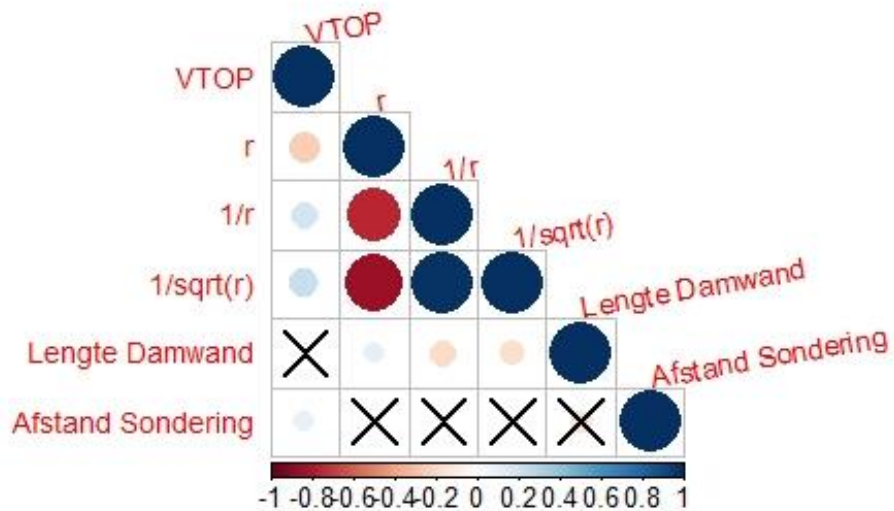


Figuur 29; Boxplot Trilsnelheid vs. Frequentie

Er is geen verband te zien tussen het type trilblok (slagkracht, frequentie) en de trilsnelheid.

Numerieke Waarden

Daarnaast kan er gekeken worden naar de relatie tussen de trilsnelheid en de numerieke systeemparameters. Uit de correlogram van hieribder blijkt dat alleen de afstand (r , $1/r$ en $1/\sqrt{r}$) en de afstand sondering enigzins een correlatie laten zien met de damwand.

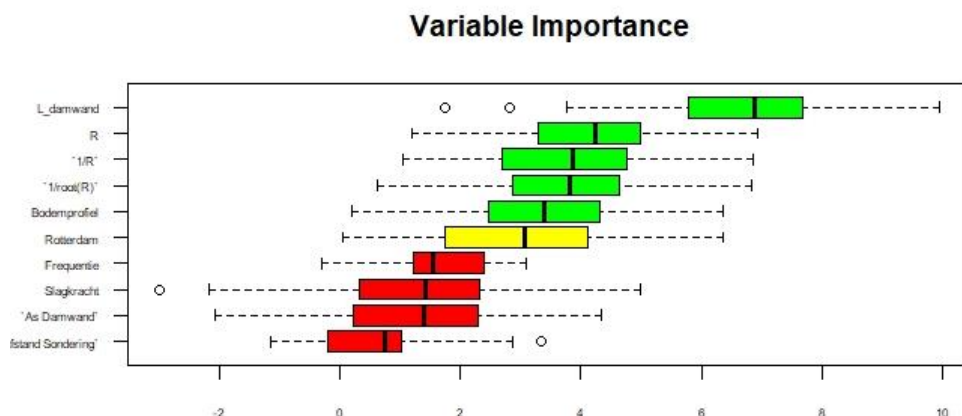


Figuur 30; Correlogram Numerieke Waarden

8.2 BORUTA

In de vorige paragraaf zagen we dat uit de correlatie analyse 4 parameters invloedrijk bleven. Een correlatie analyse is een methode om deze analyse uit te voeren. Echter bestaan er nog veel meer methodes om deze analyse uit te voeren. Een van deze methodes is Boruta feature importance. Boruta feature importance is een complexe techniek die buiten de scope van dit verslag ligt. Voor meer informatie hiervoor kunt u de volgende link raadplegen.

<https://www.datacamp.com/community/tutorials/feature-selection-R-boruta>



Figuur 31; Boruta Feature Importance Plot

De Boruta plot laat drie kleuren zien. De groene kleur betekent dat de parameter wordt gezien als invloedrijk. De rode kleur betekent dat een parameter niet invloedrijk is. Een gele waarde zit er tussenin. De Boruta feature laat ook de afstandsafhankelijkheid zien. Daarnaast wordt maatgevende bodemprofiel ook gezien als relevant. De maatgevende parameters zijn dus;

- r
- $1/r$
- $1/\sqrt{r}$
- Lengte Damwand
- IsRotterdam (Dit betekent of er spraken is van een Rotterdam bodemprofiel)
- Bodemprofiel

In de vorige paragraaf werd gesteld dat er een relatie bevindt tussen de afstand van de sondering met de trilsnelheid. In deze paragraaf wordt deze parameter gezien als de minst invloedrijke parameters. Gegeven het feit dat Boruta een robuustere techniek is en dat er geen verband lijkt te zijn tussen de trilsnelheid en de sondering, lijkt de kans dat de afstand van de sondering invloedrijk is nihil. Daarom is er gekozen om een voorspelling hier niet op te baseren

9. MACHINE LEARNING MODEL

In het vorige hoofdstuk zagen we welke parameters uiteindelijk invloedrijk waren. Bij het empirisch modelleren probeer je een model te maken met zo weinig mogelijk parameters. In deze hoofdstuk zullen we een aantal gemaakte modellen analyseren. Deze modellen zijn gemaakt m.b.v. machine learning. Voor een uitleg over wat machine learning en de uitleg over de gebruikte machine learning algoritmes verwijs ik u naar bijlage C.

Zoals in de bijlage is aangegeven is er spraken van twee type van supervised machine learning. Er is regressie en classificatie. Regressie is het voorspellen van een numerieke variabele. Classificatie gaat om het voorspellen van een categorische variabele. Zoals in de CUR 166 analyse kwam al naar voren dat je de trillingspredictie op twee manieren kunt beschouwen.

- Het voorspellen van de topsnelheid (Regressie)
- Het voorspellen van het overschrijden van de SBR-richtlijn (Classificatie)

De modellen zijn getraind op de volgende parameters:

- r
- $1/r$
- $1/\text{sqr}(r)$
- Lengte Damwand
- IsRotterdam (Dit betekent of er spraken is van een Rotterdam bodemprofiel)
- Bodemprofiel

9.1 REGRESSIE

LINEAIRE REGRESSIE

Lineaire regressie gebruikt alleen de parameters r en IsRotterdam. Dit doet het omdat de andere parameters het model niet significant verbeteren.

=== Formula ===

$$V_{\text{top}} = -0.0981 * R + 1.0976 * \text{IsRotterdam}=\text{TRUE} + 2.945$$

=== Summary ===

Correlation coefficient	0.3058
Mean absolute error	1.4271
Root mean squared error	1.947

RANDOM FOREST

=== Summary ===

Correlation coefficient	0.1997
Mean absolute error	1.4974

Root mean squared error 2.1538

NEURALE NETWERK

=== Summary ===

Correlation coefficient 0.0698
Mean absolute error 2.1207
Root mean squared error 2.7182

SAMENVATTING

Uit alle modellen voorspelt het lineaire regressie model het nauwkeurigst. Dit is tevens ook het meest simpele model om te gebruiken. Dit model is nauwkeuriger dan de CUR 166. Het model is echter niet nauwkeurig genoeg om tot een verandering van de CUR 166 te resulteren. De formule kan echter wel gebruikt worden om snel een indicatie te krijgen van de te verwachten trilsnelheid.

9.2 CLASSIFICATIE

DECISION TREE

=== Confusion Matrix ===

a	b		<-- classified as
3	24		a = TRUE
4	120		b = FALSE

De decision tree voorspelt van in 5% (7) van de gevallen een overschrijding van de SBR-richtlijn. Terwijl dit in minder dan de helft van de gevallen klopt. (3). Daarentegen is er in 27% van de gevallen een overschrijding van de SBR-richtlijn.

NEURALE NETWERK

=== Confusion Matrix ===

a	b		<-- classified as
3	24		a = TRUE
13	111		b = FALSE

De Neurale Netwerk voorspelt iets minder goed dan de decision tree. De decision tree geniet dan ook de voorkeur.

SUPPORT VECTOR MACHINE

=== Confusion Matrix ===

a	b		<-- classified as
3	24		a = TRUE
0	124		b = FALSE

De support vector machine voorspelt het nauwkeurigst. Echter is de nauwkeurigheid nog steeds ondermaats.

9.3 EMPIRISCH MODELLEREN

Door gebruik te maken van zelflerende algoritmes wordt er geen gebruik gemaakt van de aanwezige kennis omtrent trillingen. Vanuit het vakgebied van dynamiek bestaan er veel analytisch afgeleide formules. In deze paragraaf zullen we op een wijze modelleren die gebruikt maakt van deze kennis.

$$V_{top} = A * \frac{1^n}{r}$$

Met behulp van de Geobrain database is de formule geoptimaliseerd en is de gezocht naar de juiste waarden voor A en n. Hieruit kwam de volgende formule;

$$V_{top} = 4.79 * \frac{1^{0.3}}{r}$$

Uit dit model kwam een MAE gelijk aan 1.45 mm/s. Dit is vergelijkbaar met het CUR-model.

CONCLUSIE

De meest nauwkeurige algoritme voor het voorspellen van de trilsnelheid is een lineaire regressie. Dit is een opvallende uitkomst. Het neurale netwerk voorspelde het slechtst. Dit komt vooral omdat een neurale netwerk veel data nodig heeft om te trainen. Ook hebben de classificatie algoritmes niet geleid tot een goed waardevol model. De SBR-richtlijn voorspelt hierin nog slechter dan het CUR-model.

10. VERGELIJKING DEFINITIEVE MODEL MET CUR MODEL

Met behulp van machine learning zijn er verschillende modellen gemaakt. De fout van het model is nooit onder 1.42 mm/s gekomen. Dit is een verbetering, maar niet genoeg om tot een verandering van de CUR 166 te resulteren.

De CUR 166 berekening is echter veel complexer dan het machine learning model. Met twee simpele modellen waarin alleen de straal en het maatgevende bodemprofiel is gebruikt resulteerde al in een nauwkeuriger model dan de CUR.

Als we vervolgens gaan kijken naar de voorspelling van het machine learning omtrent de overschrijding van SBR-richtlijn dan zien we dat het CUR-model dit beter voorspelt.

11. CONCLUSIE

11.1 CONCLUSIE

Een van de meest waardevolle contributies aan het onderzoek naar grondtrillingen is een kwantificatie van de nauwkeurigheid van het CUR 166 model. Nog nooit is de CUR 166 model onderzocht met 150 trillingsmetingen. Dit onderzoek laat de onnauwkeurigheid van het CUR 166 model zien. Het laat zien dat het CUR 166 model niet gezien moet worden als een nauwkeurige voorspelling maar als een onnauwkeurige schatting. Echter in het onderzoeken van de 95%-waarde kwam iets interessants naar boven drijven. De 95%-waarde is veel te conservatief en is in praktijk een 99%-waarde. De 95%-waarde kon met een factor 2 gereduceerd worden volgens de dataset. Dit leidt ons tot een interessante conclusie. Het CUR 166 model is onnauwkeurig, maar niet zo onnauwkeurig als de CUR zelf suggereert.

Een ander interessant onderdeel is de literatuurstudie die zich in hoofdstuk 3 bevindt. In hoofdstuk 3 is onderzoek gedaan naar de complexiteit van trillingen. Hoe komt het dat grondtrillingen zo moeilijk voorspelbaar zijn? In dit hoofdstuk kwamen interessante concepten langs van non-lineaire dynamica tot aan chaos theorie. De conclusie die uiteindelijk getrokken kon worden was dat het exact voorspellen van de trilsnelheid onmogelijk is. Onzekerheid zal voor altijd verbonden blijven aan trillingsproblematiek.

Verder is er getracht om met behulp van machine learning een model te maken die beter presteert dan het CUR 166 model. Alhoewel dit lukte bij sommige modellen die ook nog eens veel makkelijker zijn om uit te voeren, leidde dit niet tot een significante verbetering. In het onderzoek ernaartoe kwamen er een aantal interessante dingen naar voren. We zagen dat er geen statistische relatie bestond tussen de trilsnelheid en de slagkracht, bodemeigenschappen, oppervlakte damwand en de frequentie van het trilblok. De dataset lijkt bijna op een hele grote wolk met ruis. Zowel het CUR 166 model, als de zelfgemaakte Machine Learning lijken een kleine beetje beter dan het NULL-model. Dit geeft een interessant inzicht in de werking van de grondtrillingen die ook weer een groot gedeelte van de literatuurstudie in hoofdstuk 3 valideert.

11.2 ADVIES VOOR VERDER ONDERZOEK

1. Het intensiveren van het vergaren van data

Hoe meer data er aanwezig is des te beter zijn we in staat zijn om statistisch valide uitspraken te doen over grondtrillingen.

2. Chaos Theorie

Zoals eerder in het verslag al aangegeven vertonen grondtrillingen chaotisch gedrag. Het zou van toegevoegde waarden kunnen zien om de analyse te doen vanuit het perspectief van chaos theorie.

3. Bayesiaanse Statistiek

Spreiding en onzekerheid zijn inherent verbonden aan grondtrillingen. In plaats van het doen van puntvoorspellingen zouden toekomstige modellen intervalvoorspellingen dienen te doen. Er wordt momenteel veel onderzoek gedaan naar Bayesiaanse Machine Learning. In de toekomst zou dit een interessante techniek kunnen zijn.

4. Eindige Elementen Model

Het trilsysteem wordt door onderzoekers gezien als nagenoeg deterministisch. Dat betekent dat de natuurkunde volledig begrepen is. Het probleem is dat de initiële condities (grondeigenschappen en interactiekrachten bijvoorbeeld) niet volledig te meten zijn. En omdat kleine veranderingen in deze condities kunnen leiden grote veranderingen in het trilgedrag is het nauwkeurig voorspellen momenteel niet mogelijk. Dit is ook de reden dat een eindige elementen model vaak niet gebruikt wordt. Je zou echter wel een eindige elementen model maken waarbij je kleine aanpassingen maakt en ruis toevoegt aan de initiële condities. Zo krijg je een reeks aan voorspellingen. Deze reeks voorspellingen kan dienen als kansverdeling voor de trilsnelheid.

5. Heroverwegen van de grootheid

Omdat de piek van de trilsnelheid zo lastig is om te voorspellen, zo je ervoor kunnen kiezen om over te gaan op een andere grootheid die makkelijker is om te voorspellen.

12. BIBLIOGRAPHY

- Bhande, A. (2018, 03 11). Opgehaald van Medium: <https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>
- Deckner. (2012). *Ground vibrations due to pile and sheet pile driving – prediction models of today*. Stockholm.
- Dinkgreve, S. P. (2018). *Interferentie (2018)*. Retrieved from Wetenschapsschool.
- Hölscher, P. (2016). *Soil Dynamics in Urban Areas*. Delft.
- Jain, P. (2019, 06 12). Retrieved from towardsdatascience: <https://towardsdatascience.com/complete-guide-of-activation-functions-34076e95d044>
- Kandan, H. (2017, 08 30). Retrieved from towardsdatascience: <https://towardsdatascience.com/understanding-the-kernel-trick-e0bc6112ef78>
- Lahrmam. (2018, 10 19). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Support-vector_machine#/media/File:SVM_margin.png
- net, S. (2017). *SBR Richtlijn A: Schade aan bouwwerken*.
- SBRCURnet. (2016). *CUR 166*. Delft: CROW.
- Sharma, S. (2017, 9 9). Retrieved from towardsdatascience: <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>
- Slimleren bv. (n.d.). *Slimleren*. Retrieved from <https://www.slimleren.nl/onderwerpen/rekenen/12.253/Boxplot>
- Tech, N. (2019, 5 15). Retrieved from dev.to: <https://dev.to/nexttech/classification-and-regression-analysis-with-decision-trees-jgp>
- Vlerick Busines School. (2019, 04 16). Retrieved from Vlerick: <https://www.vlerick.com/nl/research-and-faculty/knowledge-items/knowledge/artificiele-intelligentie-maakt-voorraadbeheer-efficiënter>

BIJLAGEN:

BIJLAGE A: STATISTIEK

In het rapport zal veel gebruikt worden van statistische concepten. In deze deel van de bijlage zullen deze veelgebruikte statistische concepten worden toegelicht.

MEAN ABSOLUTE ERROR

De meest simpele en robuuste methode voor het analyseren van de nauwkeurigheid van een model is de mean absolute error. Dit is simpelweg: De gemiddelde fout van het model. De fout wordt gedefinieerd als het verschil tussen de voorspellen en de meting.

R^2

Een andere grootte is de R^2 die ook wel de determinatiecoëfficiënt wordt genoemd. R^2 vertelt hoeveel van de variatie in je data wordt verklaard door je model. De R^2 varieert van 0 tot 1. Een R^2 van 0 is een waardeloos model en R^2 van 1 is een perfect model. Het voordeel van de R^2 is dat het onafhankelijk is van de schaalgrootte van de trilsnelheid. Of je de trilsnelheid nu meet in mm/s of in m/s er komt altijd dezelfde R^2 uit.

Waarschuwing:

De R^2 heeft ook zijn beperkingen. Bij grote uitschieters en weinig variatie kan de R^2 een vertekend beeld geven. Net zoals bij alle statistische toetsen is er sprake van beperkingen en is het visualiseren van de data nog altijd noodzakelijk. Bij het beoordelen van de nauwkeurigheid dienen meerdere technieken en visualisaties gebruikt te worden.

NULL MODEL

Een nul model is het meest simpele model die je kan bedenken. Bij een regressieanalyse is dat het gemiddelde. De gemiddelde van de trilsnelheid is 2,76 mm/s. Het nul model zegt dus dat alle trilsnelheden gelijk zullen zijn aan 2,76 mm/s. Het verschil tussen het CUR model en het nul model is een indicatie van hoeveel van de onderliggende processen het model weet te modelleren.

BOXPLOT

Binnen de beschrijvende statistiek is het vaak het doel om veel getallen samen te vatten in een getal die iets zegt of de distributie van een variabelen. Een van de wijzen waarop dit vaak wordt gedaan is dat de gemiddelde van een reeks getallen wordt getoond om iets te kunnen zeggen over het midden van een getallenreeks. Het samenvatten van veel getallen in een getal zorgt onomkeerbaar voor verlies aan informatie. Een andere benadering is door deze verdeling te visualiseren. De distributie van numerieke variabelen kan op de volgende wijzen gevisualiseerd worden.

- Histogram

Deze methode is al veel eerder gebruikt in dit verslag. Zie bijvoorbeeld figuur 12 en 13

- Density Plot

Een density plot vertelt iets over de kansdichtheid van variabelen. Dit is in feite een gladgetrokken versie van de histogram. Dit werkt echter niet goed bij kleine steekproeven en wordt door veel mensen ervaren als verwarrend.

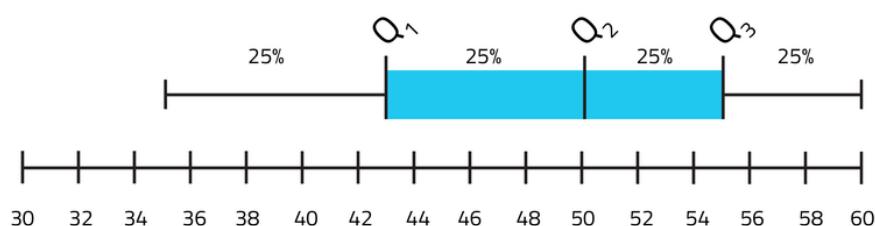
- Boxplot

Een boxplot visualiseert de volgende statistische waarden:

- Minimum
- Eerste Kwartiel Q1
- Mediaan Q2
- Derde Kwartiel Q3
- Maximum

Dit wordt ook wel de vijfgetallen samenvatting genoemd.

- Ook visualiseert het uitschieters



Figuur 32 ; Visuele Uitleg Boxplot (Slimleren bv., sd)

Hierboven is een voorbeeld van een boxplot te zien.

GEWOGEN GEMIDDELDE

Om het begrip van gewogen gemiddelde uit te leggen beginnen we direct met een voorbeeld:

Meneer Mom begeleid drie afstudeergroepen genaamd groepen A, B en C. De groepen A en B bevatten beiden 5 studenten. Groep C bevat echter drie studenten. De gemiddelde leeftijden in groepen A, B en C is respectievelijk 20, 21 en 22 jaar oud. Een student zegt:

"De gemiddelde leeftijd van alle studenten is dus 21 jaar oud. " $(20 + 21 + 22) / 3$

Dit klopt niet omdat er sprake is van verschillende groepsgrootten. Hier dient een gewogen gemiddelde toegepast te worden. De juiste berekening wordt dan:

$$\frac{20 * 5 + 21 * 5 + 22 * 3}{5+5+3} = 20,85 \text{ Jaar}$$

In generieke termen wordt dit dan:

$$\frac{X_1 * w_1 + X_2 * w_2 + \dots + X_n * w_n}{w_1 + w_2 + \dots + w_n}$$

Een w die twee groter is wordt twee keer vaker geteld. Zo kunnen we dit ook gaan gebruiken in een sondering door getallen die belangrijker zijn vaker te tellen. Dit kunnen getallen zijn die dicht bij oppervlak liggen, omdat deze dan ook dicht bij de woning zitten.

BIJLAGE B FEATURE IMPORTANCE TECHNIKEN

In hoofdstuk 8 zijn een aantal technieken gebruikt voor het bepalen van het belang van de parameter. Dit werd gedaan om zo de dimensies te reduceren. Dit maakt het model simpeler en overzichtelijker. Ook kan dit de nauwkeurigheid van het model verbeteren. De feature importance technieken zijn als volgt

1. Correlatie
2. Boruta

CORRELATIE

De meest eenvoudige methode voor het bepalen van het belang van een parameter is met behulp van de correlatie coëfficiënt. Voor een uitleg naar correlatie verwijs ik naar bijlage C. Binnen de statistiek wordt er (in bijna alle gevallen) van een uit de drie correlatie coëfficiënten gekozen. Deze zijn;

1. Pearson Correlatie Coëfficiënt
2. Spearman Rank Correlatie Coëfficiënt
3. Kendall Rank Correlatie

Pearson Correlatie

De pearson correlatie meet de lineaire verband tussen 2 normaal verdeelde parameters. De pearson correlatie coëfficiënt kan alleen gebruikt worden als je variabelen voldoet aan de volgende voorwaarden;

- De twee parameters dienen normaal verdeeld te zijn.
- Tussen de twee parameters dient zich een lineair verband te bevinden.
- Homogeniteit van de variantie; Dit betekent dat de residuen om de lineaire lijn normaal verdeeld moeten zijn.

In de geobrain dataset kan aan geen enkele van deze voorwaarden worden voldaan. Vandaar dat de pearson correlatie coëfficiënt afvalt.

Spearman Rank en Kendall Rank Correlatie

De Spearman rank correlatie en de Kendall rank correlatie zijn beiden een non-parametrische test. Dit betekent dat het geen aanname doet over de onderliggende distributie. Ook dient er geen sprake te zijn van een lineair verband.

De geobrain dataset voldoet aan alle voorwaarden. De Kendall Rank Correlatie wordt voornamelijk gebruikt voor kleine datasets. De Spearman Rank Correlatie is de meest gebruikte methode. Daarom wordt in dit verslag de Spearman Rank Correlatie gebruikt.

Voor meer informatie over de verschillende correlatiecoëfficiënten raadpleeg dan de volgende link. [LINK](#)

Spearman Rank Correlatie

Zoals de naam al suggereert maakt de spearman rank correlatie gebruik van de rang van een waarneming in een getallenreeks. Alle getallen worden eerst gerangschikt van laag naar hoog. De Spearman rank correlatie kijkt naar de rangorde in plaats van de werkelijke

waarden. Zo wordt de bruikbaarheid van de correlatiecoëfficiënt onafhankelijk van de onderlinge distributie.

De Spearman rank Correlatie wordt berekend met de volgende formule:

$$r_s = 1 - \frac{6 \sum d^2}{n^3 - n}$$

Waarin	r_s	=	De Spearman Rank Correlatiecoëfficiënt
	d	=	Het verschil tussen de rangorde van de twee getallenreeksen
	n	=	Het aantal waarnemingen

Vergelijkbaar met de Pearson Correlatie coëfficiënt ligt de uitkomst tussen de -1 en +1. Een positief verband betekent een positieven relatie tussen de getallenreeksen en de een negatief correlatiecoëfficiënt duidt op een negatieve relatie. Om te kijken naar de sterkte van de relatie dienen we dus te kijken naar de absolute waarde van correlatiecoëfficiënt.

BORUTA

Boruta is een van de meest gebruikte feature importance technieken. Boruta pakt de waarden in een kolom en shuffelt dezen op een random wijze. Deze geshuffelde parameter wordt een shadow feature genoemd. Vervolgens wordt de belangrijkheid van de echte parameter vergeleken met die van de geshuffelde parameter.

Dit wordt gedaan door een model te maken gebaseerd op de echte parameter en een op de shadow feature. Het verschil in nauwkeurigheid representeert de belangrijkheid van de beschouwde parameter.

BIJLAGE C MACHINE LEARNING

Machine Learning is een discipline afkomstig uit de wereld van computerwetenschappen en statistiek waarbij computers taken leren zonder specifiek geprogrammeerd te zijn om deze taak uit te voeren. Een computer kan patronen herkennen in data, en een model creëren die deze taak zo goed mogelijk kan volbrengen. De aanwezigheid van een dataset en patronen in de dataset is cruciaal.

Als je geïnteresseerd bent in het maken van een model die in staat is om iets te voorspellen dien je data te hebben. De data

MACHINE LEARNING ALGORITMES

Er zijn globaal gezien 3 soorten machine learning algoritmes

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

SUPERVISED LEARNING

Bij supervised learning voorspel je gelabelde data. Dit betekent dat je een model maakt die een vooraf gespecificeerde grootte voorspelt. Een voorbeeld daarvan is het Geobrain project. Er dient een supervised machine learning model gemaakt te worden die in staat is om de trilsnelheid te voorspellen.

Supervised Learning algoritmes kunnen opgedeeld worden in twee groepen. De eerste groep verricht classificatie en de andere groep verricht regressie. Bij classificatie wordt geprobeerd een categorische grootte te voorspellen. Een voorbeeld daarvan is of de SBR trillingsrichtlijn zal worden overschreden. Bij regressie wordt geprobeerd een numerieke waarde te voorspellen. Een voorbeeld daarvan is wat de exacte maximale trillingsintensiteit zal zijn.

UNSUPERVISED LEARNING EN REINFORCEMENT LEARNING

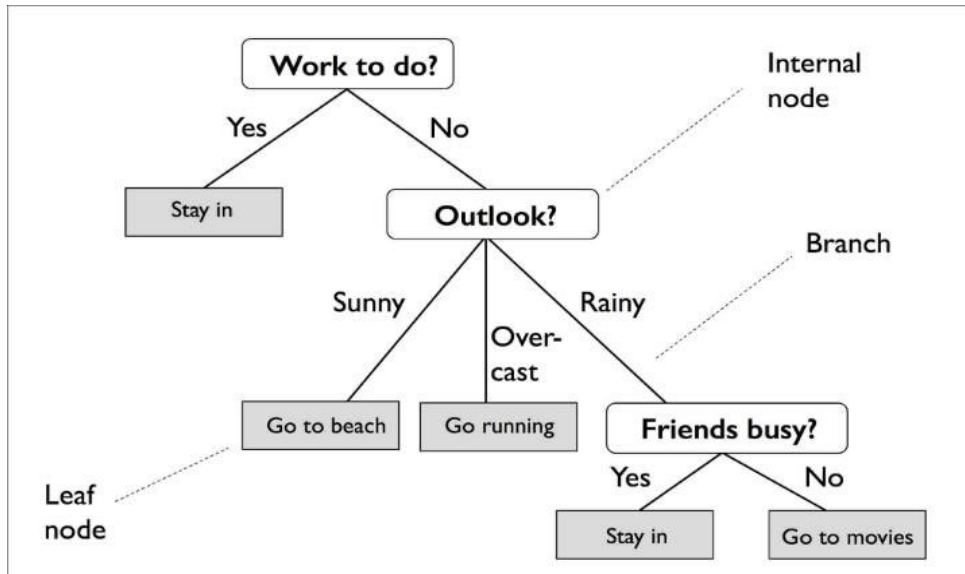
Bij unsupervised learning voorspel je niet gelabelde data. Dit betekent dat het algoritme een niet vooraf gespecificeerde grootte voorspelt. Een toepassing hiervan is clustering. Clustering is het groeperen van observaties die op elkaar lijken. Dit wordt vaak gebruikt bij webwinkels. Klanten worden geclusterd in een groep en de klantervaring wordt hierop aangepast. Reinforcement learning is de drijvende kracht achter zelfrijdende auto's. Reinforcement learning leert door in interactie te gaan met zijn omgeving, iets te doen, succes te definiëren en vervolgens de parameters te optimaliseren om succesvol de taak uit te kunnen voeren. Reinforcement leert door 'trial and error'.

Zowel unsupervised en Reinforcement learning algoritmes zijn niet gebruikt tijdens het onderzoek. Dit komt omdat het doel is om een vooraf gespecificeerde grootte te voorspellen.

CLASSIFICATIE ALGORITMES

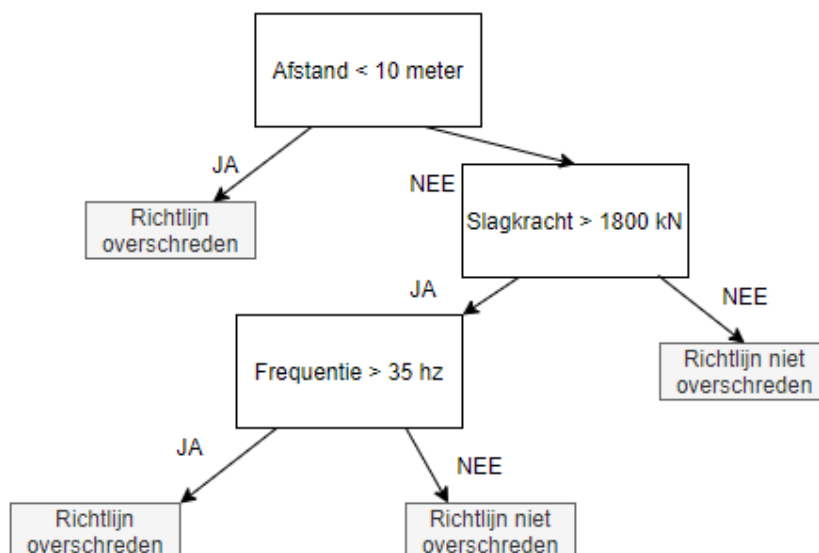
DECISION TREE

De beste manier om een decision tree uit te leggen is door te beginnen met een voorbeeld.



Figuur 33; Voorbeeld van een decision tree (Tech, 2019)

In figuur 24 is een simpel voorbeeld te zien. Deze voorbeeld laat zien hoe een beslissing gemaakt wordt. Een decision dan vragen die gesteld kunnen worden, waar consequenties aan verbonden kunnen worden. Als we deze techniek toe zouden passen dan zouden we volgende decision tree kunnen krijgen.



Figuur 34; Fictieve voorbeeld van een decision tree toegepast op trillingsproblematiek

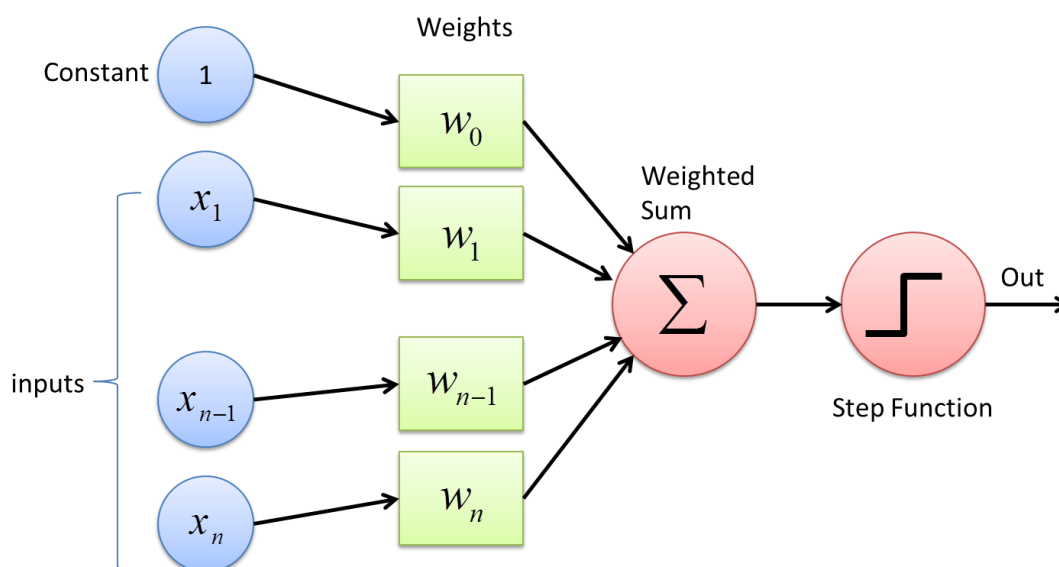
Een van de voordelen van een decision tree is dat het makkelijk interpreteerbaar is. Je kunt exact nagaan hoe een model een beslissing heeft gemaakt en welke variabelen het hiervoor heeft gebruikt. Een nadeel echter is dat een decision tree alleen makkelijke verbanden kan modelleren.

NEURALE NETWERKEN

Neurale Netwerken zijn een van de meest complexe machine learning algoritmes. Een van de redenen is omdat de werking van een neurale netwerk gebaseerd is op de werking van de hersenen. Het intuïtief begrijpen van de werken van een neurale netwerken duurt weken om te begrijpen. Het gebruiken van neurale netwerken wordt ook wel deep learning genoemd en er zijn volledige cursussen en universitaire vakken opgericht om uit te leggen hoe een neurale netwerk werkt. Deze paragraaf is bedoeld om korte inzicht te geven hoe een neurale netwerk eruit ziet. Voor een diepere uitleg in neurale netwerken verwijst ik naar de uitleg van de wiskundige Grant Sanderson.

<https://www.3blue1brown.com/neural-networks>

Een neurale netwerk kan gebruikt worden voor zowel regressie als classificatie. De meest simpele vorm van een neurale netwerk is een single layer perceptron. Voor een schematische weergave van een perceptron zie figuur 26.



Figuur 35; Perceptron (Sharma, 2017)

Al deze input nodes worden vermenigvuldigt met een gewicht.

Input Nodes

De blauwe cirkels heten de input nodes. Deze nodes zijn de variabelen waarmee je de voorspelling probeert te doen. In het voorbeeld van damwandtrillingen zouden dit de volgende variabelen kunnen zijn;

- Slagkracht Trilblok
- Afstand Damwand – Woning
- Frequentie Trilblok
- Etc....

Weights

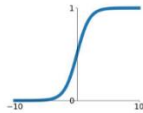
Al deze waarden in de input nodes worden vermenigvuldigt met een gewicht. De uitkomst van al deze vermenigvuldigingen worden opgesomd.

Activation Function

De uitkomst van deze opsomming gaan door een activation function. Een activation function is een non-lineaire verandering van een getal. Voorbeelden van veelgebruikte activation functions zijn te zien in de volgende figuur.

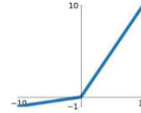
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



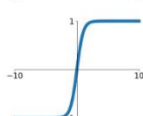
Leaky ReLU

$$\max(0.1x, x)$$



tanh

$$\tanh(x)$$

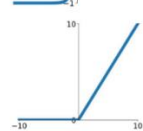


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

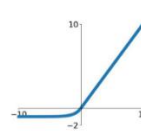
ReLU

$$\max(0, x)$$



ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

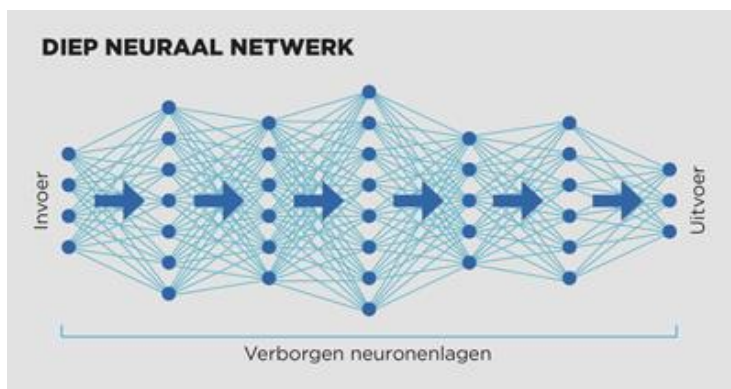


Figuur 36; Activation Functions (Jain, 2019)

Trainen Model

Aan het begin van het trainen zijn alle gewichten een random getal. Zo zijn de voorspellingen ook random. Een neurale netwerk traint via een proces genaamd back propagation en gradient descent. Via deze proces verandert het model de gewichten om een zo nauwkeurig mogelijk model te krijgen.

Dit was echter een voorbeeld met een node in het midden. Huidige modellen zijn veel complexer maar functioneren vanuit hetzelfde beginprincipes. Figuur 28 laat zo een model zien.



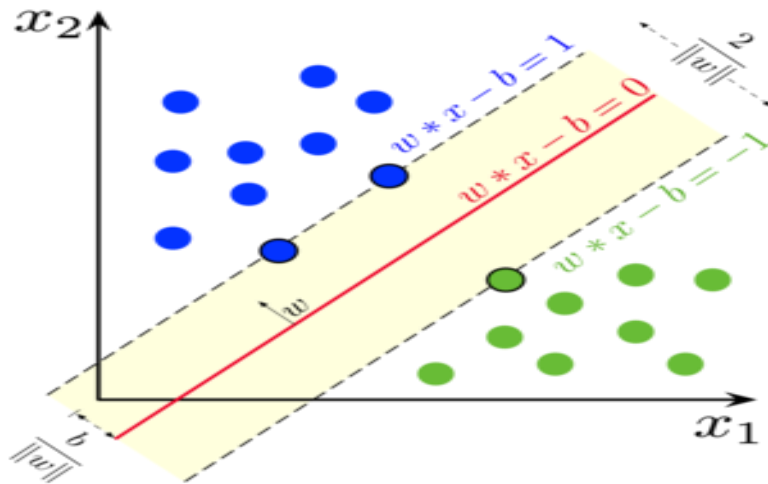
Figuur 37; Complexere neurale netwerk schematische weergave (Vlerick Business School, 2019)

- + Kan complexe verbanden modelleren, zoals het herkennen van gezichten in foto's.
- + Kan goed omgaan met veel dimensies
- Vergt veel data en kost veel rekenkracht

- Een neurale netwerkkennetwerk is moeilijk interpreteerbaar
- Een neurale netwerk kan in de optimalisatieproces hangen in iets dat een lokale minimum wordt genoemd. Dit is een model zonder de meest optimale parameters.

SUPPORT VECTOR MACHINE

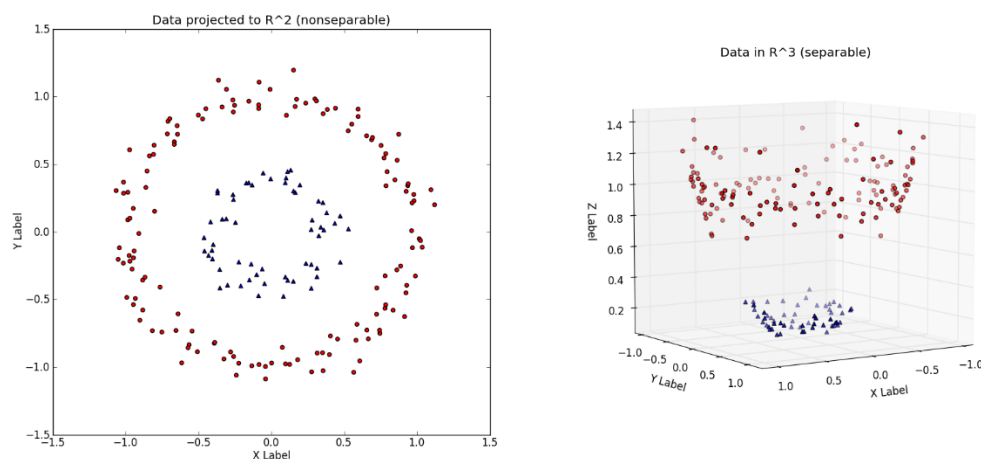
Een support vector machine is een classificatie algoritme dat gebruikt maakt van hyper planes voor het doen van voorspellingen. Een hyperplane is een scheiding tussen groepen die een dimensie lager is dan waar het zich in bevindt. Dus in een 2D-vlak is een hyperplane een lijn. In een 3D-ruimte is een hyperplane een vlak.



Figuur 38; Lineaire Hyperplane van een support vector machine (Lahrmam, 2018)

In de bovenstaande figuur is een voorbeeld van zo een scheidingslijn te zien. De rode lijn is de hyperplane. De gestippelde lijnen worden de support vectors genoemd. Vandaar ook de naam support vector machines. De afstand tussen de support vectors worden de support vector machines genoemd. Een SVM probeert de juiste hyperplane te kiezen om die afstand te maximaliseren.

Een hyperplane is altijd een lineaire functie. De groepen moeten dus lineair separeerbaar zijn. Om groepen lineair separeerbaar te maken maakt de SVM gebruik van een kernel trick. Een voorbeeld van een kernel trick is te zien in de volgende afbeelding.



Figuur 39; Example of a kernel trick (Kandan, 2017)

In eerste instantie zijn de twee groepen niet lineair te onderscheiden. Een SVM projecteert dan de punten op een andere dimensie om de groepen wel weer lineair separeerbaar te maken.

Een van de voordelen van een support vector machine is dat het niet blijft hangen in een lokale minima. (Convex optimisation) De correcte parameters worden dus gekozen voor het creëren van het correcte model.

REGRESSIE ALGORITMES

Regressie gaat dus om het voorspellen van een numeriek waarden. Er bestaan veel regressie algoritmes. In het onderzoek is er gebruik gemaakt van drie regressie algoritmes.

- Lineaire regressie
- Random forests
- Neurale Netwerken

Het gebruiken van neurale netwerken voor regressie is vrijwel identiek als classificatie. Voor een intuïtieve begrip van een neurale netwerk is de vorige paragraaf voldoende.

LINEAIRE REGRESSIE

Lineaire regressie is de meest simpele vorm van regressie en het werk volgens de formule.

$$\hat{y} = \alpha_1 x_1 + \alpha_2 x_2 + \dots \alpha_n x_n$$

In een lineaire regressie model is de voorspelling simpelweg een lineaire opsomming van variabelen die ieder vermenigvuldigd zijn met een factor. Alleen dienen de juiste factoren gekozen te worden.

RANDOM FORESTS

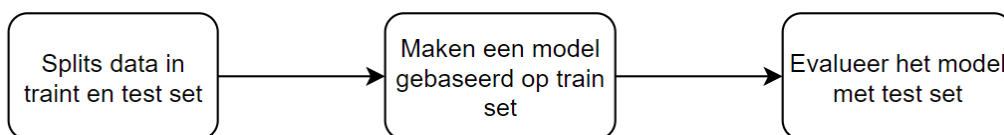
Een random forest kan zowel voor regressie als voor classificatie gebruikt worden. Een random forest bouwt voort op een decision tree. Een random forest maakt gebruik van soms wel honderden decision tree. De decision tree wordt getraind op subsets van de data. Een random forest doet dan een voorspelling gebaseerd op voorspellingen van honderden modellen. De voorspelling is dan het gemiddelde van alle voorspellingen. De reden dat dit gedaan wordt heeft te maken met overfitting, wat verder toegelicht zal worden in de volgende paragraaf.

*Opmerking

Het aantal algoritmes die er bestaan zijn talrijk. Tijdens het onderzoek is gekozen voor slechts een handvol aantal algoritmes. Dit wordt gedaan omdat dit de meest gebruikte algoritmes zijn.

MACHINE LEARNING WORKFLOW

De machine learning workflow is te zien in afbeelding 40.

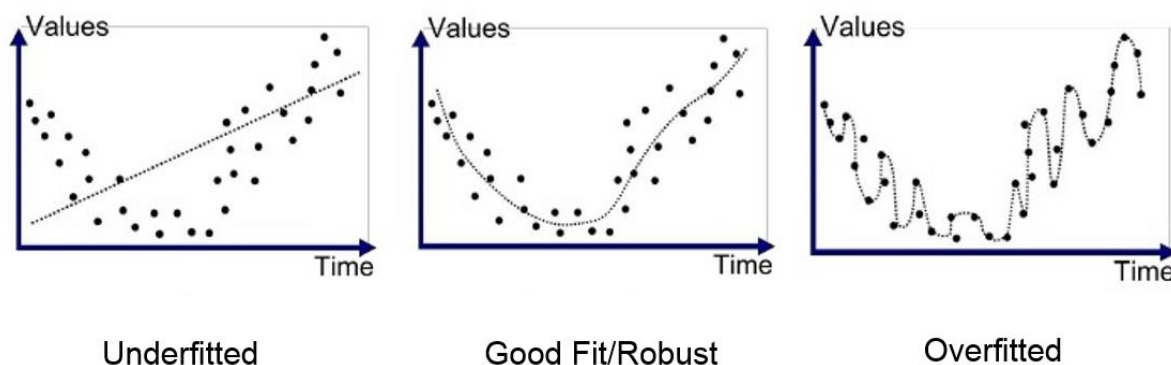


Figuur 40; Data Splitsing

De data deel je op in twee delen. Een train en test set. De volgende stap is om een model te maken gebaseerd op de train set. Vervolgens wordt met dit model een voorspelling gemaakt voor de waarnemingen die aanwezig zijn in de test set. Het verschil tussen de waargenomen waarde en de voorspelde waarde is een mate voor nauwkeurigheid.

De reden dat data opgedeeld wordt in twee delen is omdat het model generaliseerbaar moet zijn. Het moet nauwkeurig voorspellen op cases die het model nog niet gezien heeft, Anders vormt overfitting een te grote risico.

Overfitting is het proces waarbij een model gaat voorspellen gebaseerd op ruis. Het model werkt dan heel goed op de data waarop het getraind is, maar op het moment dat het model in productie gaat zal de nauwkeurigheid drastisch afnemen. Het proces van overfitting is te zien in figuur 41.



Figuur 41; Visualisatie overfitting en underfitting (Bhande, 2018)

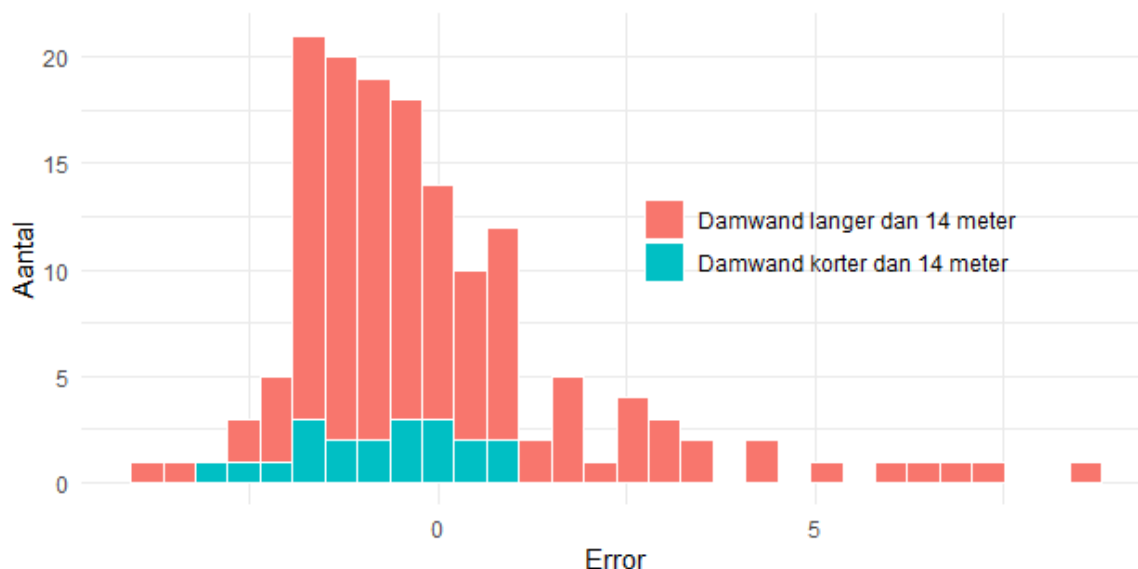
In figuur 41 wordt ook underfitting gevisualiseerd. Underfitting is het proces waarbij het model een te simplistische weergave geeft van de werkelijkheid. Het model mist zo belangrijke kenmerken van de werking van het systeem. Het verschil in nauwkeurigheid tussen de test en train set is een goede indicatie van overfitting. Neurale netwerken staan vaak bekend om het overfitten en bij lineaire regressie wordt vaak geunderfit.

BIJLAGE D: TOETSING STATISTISCHE SIGNIFICANTIE INVLOED LANGE DAMWANDLENGTE OP MODELPERFORMANCE

In de CUR 166 staat vermeld dat voorspellingen van trilsnelheden alleen gedaan kunnen worden op damwanden met een lengte korter dan 14 meter. De reden hiervoor is, is dat ten tijde van het opstellen van de empirische formules er geen of weinig trillingsproeven aanwezig waren van langer dan 14 meter. Dit was de juiste beslissing omdat bij empirisch onderzoek extrapoleren kan leiden tot grove fouten. Maar wel moet vermeld worden dat deze restrictie afkomstig is uit tekort aan kennis. Het zou dus kunnen zijn dat het model net zo nauwkeurig is bij damwanden korter dan 14 meter als damwanden langer dan 14 meter.

Ook zien we in de praktijk dat de CUR 166 trillingsprognose ook gebruikt wordt bij damwanden langer dan 14 meter. De vraag die we ons zelf moeten stellen is of dit wel terecht is. Gebruiken de praktijk een model op een situatie waar het niet voor bedoeld is? Of kunnen we vaststellen dat de CUR 166 ook gebruikt kan worden bij damwanden langer dan 14 meter?

VERKENNENDE DATA ANALYSE ERROR MODEL



In de bovenstaande afbeelding zien we dat alle grote fouten gemaakt worden bij damwanden langer dan 14 meter. Echter dienen we te beseffen dat we te maken hebben met 20 damwandproeven die korter zijn dan 14 meter. Dit heet in het jargon een 'Class Imbalance'. Het kan namelijk zo zijn dat het afwezig zijn van grote uitschieters een resultaat is van de kleine steekproefgrootte. Om hier een objectief oordeel over te vellen zijn statistische toetsen bedacht. Aan de hand van deze toetsen kun je hypothesen testen. Maar voor dat we dat doen gaan we eerst kijken naar statistische gegevens over de nauwkeurigheid van het model over deze twee steekproeven:

Korter dan 14 meter	Langer dan 14 meter
<ul style="list-style-type: none"> - Spearman Rank Correlatie: $R = 0.45$ - MAE = 1.1 mm/s 	<ul style="list-style-type: none"> - Spearman Rank Correlatie $R = 0.36$ - MAE = 1.5 mm/s

In gemiddelde zin zien we dan ook een verslechtering van het model. Is deze toename echter statistisch significant? De eerste stap is het formuleren van hypothesen:

1. Is de verdeling van de fout van het model van damwanden kleiner dan 14 meter en damwanden groter dan 14 meter afkomstig uit dezelfde distributie?
2. Is de gemiddelde absolute fout van het model gelijk over de twee subpopulaties?

Opmerking:

Veel statistische toetsen gaan uit van normaliteit (normale verdeling) van je data. Alhoewel we zagen dat de trilsnelheden lognormaal verdeeld waren, was de fout van het model dat niet. Veel van de gebruikelijke statistische toetsen vallen dan af. Aangezien de onderliggende distributie onbekend is, dient er gebruik gemaakt te worden van non-parametrische toetsen. Deze toetsen gaan niet uit van een onderliggende distributie. Voor meer info over het toetsen van de aard van de distributies zie dan bijlage B.

Hypothese 1:

Het toetsen of twee steekproeven afkomstig zijn uit dezelfde verdeling wordt getoetst met Kolmogorov-Smirnov test. Deze test geeft de volgende output:

Two-sample Kolmogorov-Smirnov test

data: korte_damwand\$Verschil and lange_damwand\$Verschil

D = 0.19231, p-value = 0.5431

alternative hypothesis: two-sided

Een P-waarden boven de 0.05 betekent dat de twee distributies niet statistisch significant van elkaar verschillen. We kunnen dus niet zeggen dat de twee afkomstig zijn uit verschillende distributies.

Hypothese 2:

Het statistisch toetsen wordt normaal gedaan met een T-toets. Maar aangezien de T-toets uitgaat van normaliteit kunnen we deze niet gebruiken. Een non-parametrische variant van de T-test is de Mann Whitney U-test. De output van deze test ziet er als volgt uit:

Wilcoxon rank sum test with continuity correction

data: Trilling_def2\$Abs_error by as.factor(Trilling_def2\$IsSmall)

W = 1530.5, p-value = 0.1018

alternative hypothesis: true location shift is greater than 0

Ook deze P-waarde is niet statistisch significant. Uit deze twee toetsen kunnen we afleiden dat er geen statistisch significant verschil waargenomen kan worden. Er is namelijk een 10% kans dat het verschil in MAE een gevolg is van toeval.

Conclusie:

In gemiddelde zin lijkt het model onnauwkeuriger te zijn bij damwanden langer dan 14 meter. In deze analyse is er echter geen statistisch significant verschil aangetroffen tussen de twee steekproeven. In praktijk wordt het model ook gebruikt bij situaties langer dan 14 meter. Daarom is de nauwkeurigheid van het model bij damwanden langer dan 14 meter ook relevant om te beoordelen.

BIJLAGE E: TOETSEN VAN NORMALITEIT VAN DE VERDELING

Het toetsen van de normaliteit van een verdeling kan op twee wijzen gebeuren.

1. Het gebruik maken van een statistische toets genaamd de 'Shapiro-Wilk test'.
2. Het visualiseren van de normaliteit middels een qqplot.

We willen twee parameters toetsen op normaliteit en lognormaliteit. Deze parameters zijn:

- Trilsnelheid
- Fout van de CUR 166

TRILSNELHEID NORMALITEIT

1. Shapiro-Wilk Test

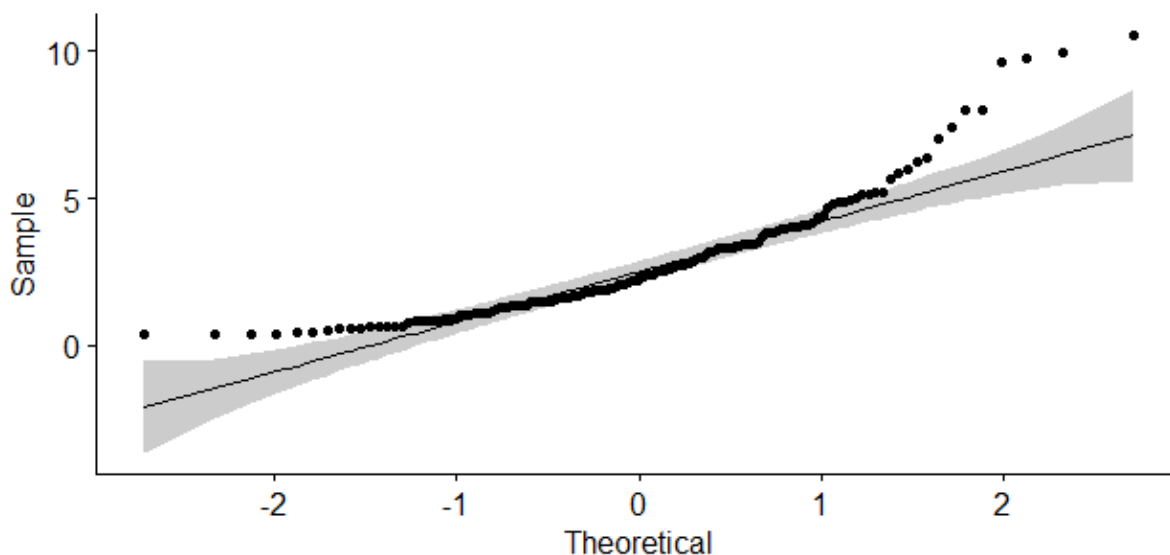
De Shapiro-Wilk Test genereert de volgende output:

Shapiro-Wilk normality test

data: Trilling_def2\$R0031_VTOP
W = 0.86657, p-value = 2.528e-10

De P-waarde is klein. Dit wil zeggen dat de distributie significant verschilt van een normaalverdeling. Hierdoor kunnen we concluderen dat de trilsnelheid niet normaalverdeeld is.

2. Qqplot



Bij een normaalverdeling dienen de punten een lineaire lijn te vormen om de lijn heen en binnen het grijze gebied te vallen. Dit is echter niet het geval.

Zowel de Shapiro-Wilk toets als de qqplot geeft aan dat de trilsnelheid niet normaal verdeeld is. De volgende stap is een lognormale verdeling.

TRILSNELHEID LOGNORMALITEIT

Een lognormale verdeling betekent de wanneer je de log van de parameter pakt je een normale verdeling krijgt. Dit kunnen we toetsen door weer dezelfde twee toetsen uit te voeren.

1. Shapiro-Wilk Test

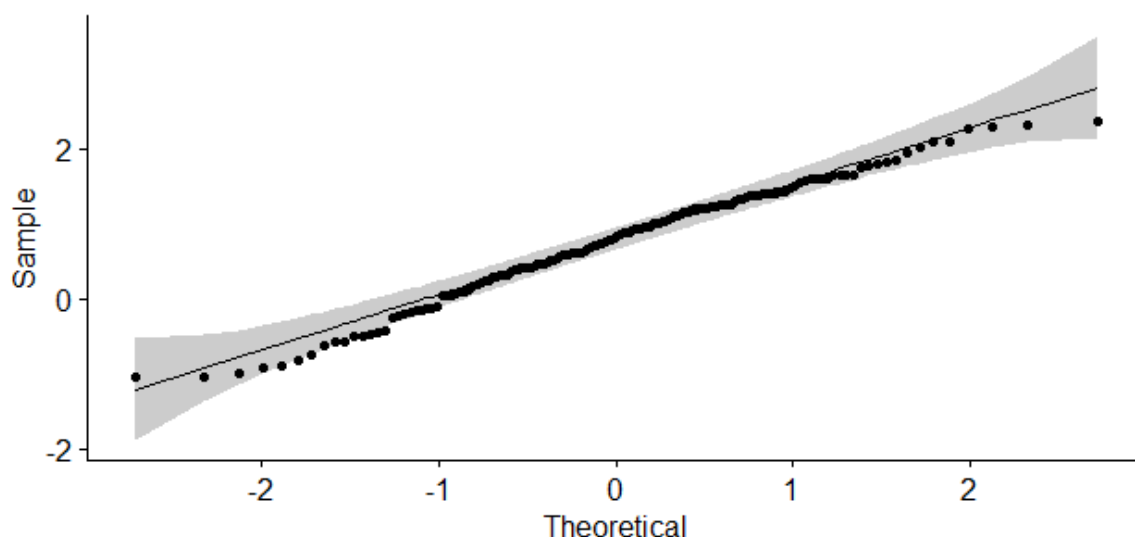
De Shapiro-Wilk Test genereert de volgende output:

Shapiro-Wilk normality test

data: `log(Trilling_def2$R0031_VTOP)`
W = 0.98423, p-value = 0.08431

We zien hier een P-waarden boven de 0.5. We kunnen dus niet met zekerheid zeggen dat de `log(Trilsnelheid)` niet normaal verdeeld is. Hierdoor dienen we dus te kijken naar de qqplot voor de normaliteitstoets.

2. Qqplot



We zien dat de `log(Trilsnelheid)` een normale verdeling volgt. Dit zien we doordat de punten een lineaire verloop volgen en binnen het grijze vlak blijven. We kunnen dus concluderen dat de trilsnelheid lognormaal verdeeld is.

ERROR NORMALITEIT

Als we kijken naar de Shapiro-Wilk test voor de error zien we de volgende output:

Shapiro-Wilk normality test

data: `Trilling_def2$Verschil`
W = 0.85832, p-value = 1.053e-10

Door de lage p-waarde kunnen we concluderen dat de error niet normaal verdeeld is. Ook laat qqplot zien dat de error niet normaal verdeeld is. De lognormale verdeling kan ook niet aangezien het niet mogelijk is om de log van een negatief getal te pakken. Hierdoor moeten we concluderen dat we de onderliggende verdeling van de fout van het cur-model niet weten.

BIJLAGE F: R-SCRIPTS

Script 1

Data Manipulatie

In deze script wordt de trillingsdata, de sonderingen, en data over de sondering ingeladen.

Deze data wordt gecombineerd, gemanipuleerd en schoon gemaakt.

prerequisites;

Script Geschreven volgens de Tidyverse Style Guide

Programmeur: Marwan el Morabet; Contact: marwanelmorabet@gmail.com

De juiste working directory keuze

```
setwd("C:/Users/morabet/OneDrive - Stichting Deltares/Desktop/Geobrain")
```

Programmeer libraries inladen

```
library(dplyr)
```

```
library(readr)
```

```
library(data.table)
```

```
library(xlsx)
```

Inladen van de data

Trillingsdata inladen

```
Constructie <- read.csv("C:/Users/morabet/OneDrive - Stichting
Deltares/Desktop/Data/Trillingdata3.csv",
                        sep=";", stringsAsFactors=FALSE)
```

Informatie over locatie van projectsondering inladen

```
Sondering_location <- read.csv("C:/Users/morabet/OneDrive - Stichting
Deltares/Desktop/Data/Fred/Admin_data.csv", stringsAsFactors=FALSE) %>%
  select(SONDNR, X_CRD, Y_CRD, MV_mm, MAXDEPTH_mm)
```

Projectsondering inladen

```
Sondering <- read.csv("C:/Users/morabet/OneDrive - Stichting
Deltares/Desktop/Data/Fred/Output_data.csv", stringsAsFactors=FALSE)
```

Extra DINO loket sonderingen inladen

```
Sondering <- read_delim("DINO.csv", ";", escape_double = FALSE, trim_ws = TRUE) %>%
  bind_rows(Sondering)
```

Locatiegegevens van DINO lokeet sonderingen inladen

```
DINO_locatie <- read.csv2("C:/Users/morabet/OneDrive - Stichting
Deltares/Desktop/Geobrain/DINO_location.csv", stringsAsFactors=FALSE) %>%
  select(SONDNR, X_CRD, Y_CRD, MV_mm, MAXDEPTH_mm) %>%
```

```
mutate(MV_mm = MV_mm * 1000, MAXDEPTH_mm = MAXDEPTH_mm * 1000)
```

```
# Datasets samenvoegen
```

```
Sondering_locatie <- rbind(Sondering_location, DINO_locatie)
```

```
# Dimensies trillingsdata reduceren
```

```
Constructie_short <- Constructie %>%
```

```
  select(Frequentie_Trilblok, Slagkracht, R0104_LOKATIE_X, R0105_LOKATIE_Y, R0201_AFSTAND_MEETPUNT,
    R0201_AFSTAND_MEETPUNT,
```

```
    R0717_IS_OVERSCHRIJDINGEN, R0711_LI_SBR_CATEGORIE, R0713_LI_X_AMPLITUDE,
    R0718_LI_Y_AMPLITUDE, R0719_LI_Z_AMPLITUDE,
```

```
    R0712_LI_WAARDE_VTOP, R0417_LI_TYPE_DAMWAND, R0515_DIEPTE_BEKEND, R0323_VALGEWICHT,
    R0108_PLAATSNAAM, R0412 LENGTE_PAAL_DAMWAND, R0413_PUNT_NIVEAU, R0031_VTOP, R0032_VD,
    R0031_VTOP,
```

```
    R0033_VKAR, R0034_VGRENS, R0035_VTOETSCALC)
```

```
#### Definitieve Code voor vinden dichtstbijzijnde waarneming ####
```

```
# Functie voor het vinden van de dichtstbijzijnde sondering
```

```
search_close <- function(xinput, yinput) {
```

```
  df <- Sondering_locatie %>% filter(MAXDEPTH_mm > 14000)
```

```
  df$distance <- sqrt((df$X_CRD - xinput)^2 + (df$Y_CRD - yinput)^2)
```

```
  return(df %>% filter(distance == min(df$distance)) %>% select(SONDNR) )
```

```
}
```

```
output <- data.frame()
```

```
for(i in 1:154) {
```

```
  X_coord <- Constructie[i, 17]
```

```
  Y_coord <- Constructie[i, 18]
```

```
  output[i, "SONDNR"] <- search_close(xinput = X_coord, yinput = Y_coord)
```

```
}
```

```
# Toevoegen van de juiste sondering en het berekene van de afstand tussen damwand en sondering
```

```
Trilling_Join <- cbind(Constructie_short, output) %>%
```

```
  left_join(Sondering_locatie, by="SONDNR") %>%
```

```
  mutate(distance = sqrt((R0104_LOKATIE_X - X_CRD)^2 + (R0105_LOKATIE_Y - Y_CRD)^2) ) %>%
```

```
  filter(R0108_PLAATSNAAM == "Lekkerkerk")
```

```
# Het creeren van een dataset met damwandinfo met sonderingwaarden
```

```
Sondering <- Trilling_Join %>% right_join(Sondering, by = "SONDNR") %>%
```

```
  filter(!is.na(R0104_LOKATIE_X)) %>%
```

```
  select(SONDNR, R0412 LENGTE_PAAL_DAMWAND, R0413_PUNT_NIVEAU, MAXDEPTH_mm, DEPTH,
    CONE_RESISTANCE, FRICTION, FRICTION_FACTOR) %>%
```

```
  mutate(R0413_PUNT_NIVEAU = abs(R0413_PUNT_NIVEAU))
```

```
# kolomnamen aanpassen
names(Sondering) <- c("SONDNR", "LENGTE_DAMWAND", "PUNT_NIVEAU", "MAXDEPTH_CPT", "DEPTH",
"CONE_RESISTANCE", "FRICTION", "FRICTION_FACTOR")

# Alleen lange sonderingen kiezen
Sondering <-Sondering %>% filter(DEPTH <= PUNT_NIVEAU) %>%
mutate(DEPTH_NEG = (1/DEPTH)^2)

#### Manipuleren van sondering ####

# Het berekenen van CPT averaging waarden (max, min, gemiddelde, mediaan, etc..... van
conusweerstand, wrijvingsgetal, schatwrijving)
Sondering <- Sondering %>% group_by(SONDNR) %>%
summarise(Eerste_meting = min(DEPTH),
MAX_CONE_RESISTANCE = max(CONE_RESISTANCE),
MIN_CONE_RESISTANCE = min(CONE_RESISTANCE),
MEAN_CONE_RESISTANCE = mean(CONE_RESISTANCE),
MED_CONE_RESISTANCE = median(CONE_RESISTANCE),
SD_CONE_RESISTANCE = sd(CONE_RESISTANCE),
FIRST_QUARTILE_CONE_RESISTANCE = summary(CONE_RESISTANCE)[2],
THIRD_QUARTILE_CONE_RESISTANCE = summary(CONE_RESISTANCE)[5],
SUM_CONE_RESISTANCE = sum(CONE_RESISTANCE),
WEIGHTED_MEAN_CONE_RESISTANCE = weighted.mean(CONE_RESISTANCE, DEPTH_NEG),
MAX_FRICTION = max(FRICTION),
MIN_FRICTION = min(FRICTION),
MEAN_FRICTION = mean(FRICTION),
MED_FRICTION = median(FRICTION),
SD_FRICTION = sd(FRICTION),
FIRST_QUARTILE_FRICTION = summary(FRICTION)[2],
THIRD_QUARTILE_FRICTION = summary(FRICTION)[5],
SUM_FRICTION = sum(FRICTION),
WEIGHTED_MEAN_FRICTION = weighted.mean(FRICTION, DEPTH_NEG),
MAX_FRICTION_FACTOR = max(FRICTION_FACTOR),
MIN_FRICTION_FACTOR = min(FRICTION_FACTOR),
MEAN_FRICTION_FACTOR = mean(FRICTION_FACTOR),
MED_FRICTION_FACTOR = median(FRICTION_FACTOR),
SD_FRICTION_FACTOR = sd(FRICTION_FACTOR),
FIRST_QUARTILE_FRICTION_FACTOR = summary(FRICTION_FACTOR)[2],
THIRD_QUARTILE_FRICTION_FACTOR = summary(FRICTION_FACTOR)[5],
SUM_FRICTION_FACTOR = sum(FRICTION_FACTOR),
WEIGHTED_MEAN_FRICTION_FACTOR = weighted.mean(FRICTION_FACTOR, DEPTH_NEG))
```



```
# Combineren van dataset over trillingen met de berekende CPT averages
```

```
Trilling_def2 <- Trilling_Join %>% left_join(Sondering, by="SONDNR") %>%
```

```
  select(Frequentie_Trilblok, Slagkracht, R0201_AFSTAND_MEETPUNT, R0713_LI_X_AMPLITUDE,
R0718_LI_Y_AMPLITUDE, R0104_LOKATIE_X, R0105_LOKATIE_Y, R0417_LI_TYPE_DAMWAND,
R0711_LI_SBR_CATEGORIE, MAXDEPTH_mm, distance, Eerste_meting, R0717_IS_OVERSCHRIJDINGEN,
SONDNR,
```

```
MAX_CONE_RESISTANCE, MIN_CONE_RESISTANCE, MEAN_CONE_RESISTANCE, MED_CONE_RESISTANCE, SD
_CONE_RESISTANCE, FIRST_QUARTILE_CONE_RESISTANCE, THIRD_QUARTILE_CONE_RESISTANCE,
```

```
SUM_CONE_RESISTANCE, WEIGHTED_MEAN_CONE_RESISTANCE, MAX_FRICTION, MIN_FRICTION, MEAN_FRI
CTION, MED_FRICTION, SD_FRICTION, FIRST_QUARTILE_FRICTION,
```

```
THIRD_QUARTILE_FRICTION, SUM_FRICTION, WEIGHTED_MEAN_FRICTION, MAX_FRICTION_FACTOR, MIN_FRI
CTION_FACTOR, MEAN_FRICTION_FACTOR, MED_FRICTION_FACTOR, SD_FRICTION_FACTOR,
```

```
  FIRST_QUARTILE_FRICTION_FACTOR, THIRD_QUARTILE_FRICTION_FACTOR,
SUM_FRICTION_FACTOR, WEIGHTED_MEAN_FRICTION_FACTOR, R0719_LI_Z_AMPLITUDE,
```

```
  R0031_VTOP, R0032_VD, R0033_VKAR, R0034_VGRENS,
R0035_VTOETSCALC, R0412 LENGTE_PAAL_DAMWAND)
```

```
# Overschrijdingen als een factor maken
```

```
Trilling_def2$R0717_IS_OVERSCHRIJDINGEN <- as.factor(Trilling_def2$R0717_IS_OVERSCHRIJDINGEN)
```

```
# Maatgevende bodemprofiel inladen
```

```
Bodemprofiel_Classificatie <- read_delim("C:/Users/morabet/OneDrive - Stichting
Deltares/Desktop/Data/Bodemprofiel Classificatie.csv",
```

```
  ";", escape_double = FALSE, trim_ws = TRUE)
```

```
# Combineren met de data
```

```
Trilling_def <- Trilling_def2 %>% left_join(Bodemprofiel_Classificatie, by = "SONDNR") %>%
```

```
  filter(R0711_LI_SBR_CATEGORIE == 3)
```

```
# damwandgegevens inladen
```

```
damwand_info <- data.frame(R0417_LI_TYPE_DAMWAND = c(0, 116, 119, 120, 124, 127, 130, 399),
```

```
  Naam_Damwand = c(NA, "AZ 13", "AZ 17", "AZ 18", "AZ 26", "AZ 36", "AZ 48", "AZ 36 - 700"),
```

```
  As_enkel = c(NA, 91.8, 86.9, 94.8, 124.6, 155.6, 177.8, 151.1),
```

```
  As_dubbel = c(NA, 183.6, 173.9, 189.6, 249.2, 311.2, 355.6, 302.2))
```

```
# Definitieve dataset voor het doen van analyses
```

```
Trilling_def <- Trilling_def %>% left_join(damwand_info, by="R0417_LI_TYPE_DAMWAND")
```

```
# Verwijderen van niet invloedrijke variabelen
```

```
remove(Bodemprofiel_Classificatie, Constructie, Constructie_short, DINO_locatie, output, Sondering,
Sondering_locatie, Sondering_location, Trilling_def2,
```

```
  i, X_coord, Y_coord, damwand_info, Trilling_Join, search_close)
```

Script 2

CUR 166 analyse

In deze script wordt de CUR 16 analyse die te zien is in hoofdstuk 4 uitgevoerd

Script Geschreven volgens de Tidyverse Style Guide

Programmeur: Marwan el Morabet; Contact: marwanelmorabet@gmail.com

Kiezen van de juiste working directory

```
setwd("C:/Users/morabet/OneDrive - Stichting Deltares/Desktop/Geobrain/Scripts/Def")
```

Laden van prerequisites

Runnen van de eerste script

```
source("1.Data_Schoonmaken.R")
```

Inladen van programming libraries

```
library(ggplot2)
```

```
library(ggthemes)
```

```
library(plotly)
```

```
library(RColorBrewer)
```

```
# library(ggpmisc)
```

```
# library(ggpubr)
```

```
# library(PerformanceAnalytics)
```

```
library(corrplot)
```

```
# library(BBmisc)
```

```
# library(randomForest)
```

CUR 166 damwand trillingsprognose in een functie

```
CUR166_Damwand_Trillingsprognose <- function(Bodemprofiel = "Amsterdam", straal = 8, slagkracht = 900, percentage = 95, statement = FALSE) {
```

Bodemparameters volgens de CUR

```
Bodemparameters <- data.frame(Bodemprofiel = c("Amsterdam", "Eindhoven", "Groningen", "Den Haag", "Rotterdam", "Tiel", "X"),
```

```
      V50 = c(1.6, 1.9, 1.7, 1.9, 1.6, 1.6, 1.6),
```

```
      C50 = c(0.002, 0.002, 0.002, 0.002, 0.002, 0.002, 0.002),
```

```
      V95 = c(9, 6.7, 13.4, 6.7, 9, 9, 9),
```

```
      C95 = c(0.011, 0.007, 0.016, 0.007, 0.011, 0.011, 0.011),
```

```
      V99 = c(18.3, 11.3, 31.6, 11.3, 18.3, 18.3, 18.3),
```

```
      C99 = c(0.023, 0.012, 0.037, 0.012, 0.023, 0.023, 0.023))
```

Het kiezen van de juiste overschreidingskans

```
Rij <- which(Bodemparameters$Bodemprofiel == Bodemprofiel)
```

```
if (percentage == 50) {
```

```

kolomv <- 2
kolomc <- 3
} else if (percentage == 95) {
  kolomv <- 4
  kolomc <- 5
} else if (percentage == 99) {
  kolomv <- 6
  kolomc <- 7
} else {
  return(print("Percentage should be either 50, 95 or 99 %"))
  # Verbreek de functie
}

# Selecteren van de juiste parameters voor de berekening
v0 <- Bodemparemeters[Rij, kolomv]
cvel <- Bodemparemeters[Rij, kolomc]

# vcor <- 2.5*v0 #+ cvel*(0.5 *slagkracht - 50)
vcor <- v0 + cvel*(slagkracht - 350)

# Het berekenen van de topsnelheid horende bij de ingevulde overschrijdingsgrens.
vmax <- round(vcor * 0.7 * sqrt(5/straal), digits = 2)

# Indien gewenst een print statement die de output verduidelijkt
if (statement == T) {
  print(paste("De", percentage, "% bovengrens van de trilsnelheid staat gelijk aan", vmax, "mm/s"))
}

# De Vmax aanwijzen als output van de functie
return(vmax)
}

# CUR 166 trillingsprognose over 150 cases
Trilling_def <- Trilling_def %>%
  rowwise() %>%
  mutate(Vcur = CUR166_Damwand_Trillingsprognose(Bodemprofiel = Bodemprofiel, straal =
R0201_AFSTAND_MEETPUNT, slagkracht = Slagkracht, percentage = 50)) %>%
  mutate(Verschil = Vcur - R0031_VTOP, Abs_error = abs(Verschil)) %>%
  mutate(Ispositive = Verschil >= 0, Is-small = R0412_LENGTE_PAAL_DAMWAND <= 14, fout_factor =
R0031_VTOP / Vcur) %>%
  mutate(Big_error = fout_factor >= 3) %>%
  as.data.frame()

# Frequentie en Slagkracht omzetten naar categorische variabelen

```

```
# Trilling_def$Frequentie_Trilblok <- as.factor(Trilling_def$Frequentie_Trilblok)
# Trilling_def$Slagkracht <- as.factor(Trilling_def$Slagkracht)

# Opsplitsen van dataset in twee datasets naar lengte damwand
korte_damwand <- filter(Trilling_def, lssmall == TRUE) %>% arrange(fout_factor)
lange_damwand <- filter(Trilling_def, lssmall == FALSE) %>% arrange(fout_factor)

#### Maken van statistische berekeningen ####
# Berekenen van spearman ranked correlatie over datasets
CUR_COR <- cor(Trilling_def$Vcur, Trilling_def$R0031_VTOP, method = "spearman")
CUR_COR14 <- cor(korte_damwand$R0031_VTOP, korte_damwand$Vcur, method = "spearman")
CUR_COR15 <- cor(lange_damwand$R0031_VTOP, lange_damwand$Vcur, method = "spearman")

# Mean average error berekenen
MAE <- mean(abs(Trilling_def$Verschil))
MAE_kort <- mean(abs(korte_damwand$Verschil))
MAE_lang <- mean(abs(lange_damwand$Verschil))

# Standaard Deviatie van error van model
sd_tril <- sd(Trilling_def$Verschil)
sd_kort <- sd(korte_damwand$Verschil)
sd_lang <- sd(lange_damwand$Verschil)

# Root mean squared error bereken van model
RMSE_tril <- sqrt(mean(Trilling_def$Verschil^2))
RMSE_kort <- sqrt(mean(korte_damwand$Verschil^2))
RMSE_lang <- sqrt(mean(lange_damwand$Verschil^2))

# Visualiseren van nauwkeurigheid CUR 166 zonder 14 meter onderscheiding (figuur 11)
ggplot() +
  geom_point(aes(x=Trilling_def$Vcur, y=Trilling_def$R0031_VTOP), size = 2, shape = 3) + xlab("Trilsnelheid  
Volgens het CUR 166 model") + ylab("Gemeten top snelheid") +
  geom_abline(slope = 1, color = "red", size = 1.3) + ylab("Gemeten top snelheid") +
  theme_light() + xlim(0, 11) + ylim(0, 12) + theme(
    legend.title = element_blank(),
    axis.title.x = element_text(size=14, face="bold"),
    axis.title.y = element_text(size=14, face="bold")) ; ggplotly()

# Verdeling van residuen (Figuur 12)
ggplot(Trilling_def, aes(x=Verschil, fill = lpositive)) +
  geom_histogram(bins = 40, color = "black") + theme_minimal() +
  xlab("CUR166 - Meting") + ylab("Aantal Waarnemingen") + labs(fill = "") +
  theme(legend.position = c(0.3, 0.58)) +
  scale_fill_discrete(labels=c("Meting hoger dan prognose", "Prognose hoger dan meting"))
```

Null Model Residuen (Figuur 13)

```
verschil_nulmodel <- data.frame(a = mean(Trilling_def$R0031_VTOP) - Trilling_def$R0031_VTOP) %>%
  mutate(Ispositive = a >= 0)
MAE_nul <- mean(abs(verschil_nulmodel$a))
```

```
ggplot(verschil_nulmodel, aes(x=a, fill = Ispositive)) +
  geom_histogram(bins = 35, color = "black")+ theme_minimal() +
  xlab("Error Null Model") + ylab("Aantal Waarnemingen") + labs(fill = "") +
  theme(legend.position = c(0.25, 0.58)) +
  scale_fill_discrete(labels=c("Meting hoger dan prognose", "Prognose hoger dan meting"))
```

Verschil groter of kleiner dan 14 meter (Figuur 15)

```
ggplot() +
  geom_point(aes(x=Trilling_def$Vcur, y=Trilling_def$R0031_VTOP, color = Trilling_def$Issmall), size = 2,
  shape = 3) + xlab("Trilsnelheid Volgens de CUR 166 model") + ylab("Gemeten top snelheid") +
  geom_abline(slope = 1, color = "deeppink3", size = 1.3) + ylab("Gemeten top snelheid") +
  theme_light() + xlim(0, 11) + ylim(0, 11) +
  scale_color_discrete(labels=c("Damwand langer dan 14 m.", "Damwand korter dan 14 m.))+
  theme(
    legend.background = element_rect(fill = "transparent"),
    legend.box.background = element_rect(fill = "transparent"),
    legend.position = c(0.72, 0.15),
    legend.direction = "vertical", legend.title = element_blank(),
    legend.text = element_text(size = 10),
    axis.title.x = element_text(size=14, face="bold"),
    axis.title.y = element_text(size=14, face="bold")) + annotate("text", x=6.5, y=9.5, label = paste0("R^2 = ",
  round(CUR_COR^2, digits = 3), "0"), size = 6, color = "brown1") +
  annotate("text", x=6.5, y=8.5, label = paste0("R^2 = ", round(CUR_COR14^2, digits = 3)), size = 6, color =
  "cyan")
```

Lengte Damwand vs. Error (Figuur 16)

```
ggplot(Trilling_def, aes(x=R0412 LENGTE_PAAL_DAMWAND, y=Abs_error)) + geom_point() +
  xlab("Lengte Damwand [m]") + ylab("Absolute Error Model [mm/s]") + theme_light() +
  geom_smooth(method = "lm", se=F) +
  theme(
    legend.background = element_rect(fill = "transparent"),
    legend.box.background = element_rect(fill = "transparent"),
    legend.text = element_text(size = 10),
    axis.title.x = element_text(size=14, face="bold"),
    axis.title.y = element_text(size=14, face="bold"))
```

Lengte Damwand vs. Vtop (figuur 17)

```
ggplot(Trilling_def, aes(x=R0412 LENGTE_PAAL_DAMWAND, y=R0031_VTOP)) + geom_point() +
```

```
xlab("Lengte Damwand [m]") + ylab("V top [mm/s]") + theme_light() + geom_smooth(method = "lm",
se=F) +
```

```
theme(
  legend.background = element_rect(fill = "transparent"), # get rid of legend bg
  legend.box.background = element_rect(fill = "transparent"), # get rid of legend panel bg
  legend.text = element_text(size = 10),
  axis.title.x = element_text(size=14, face="bold"),
  axis.title.y = element_text(size=14, face="bold"))
```

```
# Simpele Boxplot grafieken
```

```
ggplot(Trilling_def, aes(y=Verschil, x= Frequentie_Trilblok)) +
  geom_boxplot() + coord_flip()
```

```
ggplot(Trilling_def, aes(y=Verschil, x= Slagkracht)) +
  geom_boxplot() + coord_flip()
```

```
# Error Correlatie Analyse
```

```
# Creeeren van een functie voor toevoegen voor statistieken aan grafiek
```

```
stat_box_data <- function(y, upper_limit = max(Trilling_def$Abs_error) * 1.15) {
  return(
    data.frame(
      y = 0.95 * upper_limit,
      label = paste('Aantal =', length(y), '\n',
        'Gemiddelde =', round(mean(y), 1), '\n')
    )
  )
}
```

```
# Invloed op bodemprofiel op error model
```

```
Bodem_freq <- Trilling_def %>%
```

```
  group_by(Bodemprofiel) %>% summarise(Aantal = n(), mean_error = mean(Abs_error), mean_vibr =
  mean(R0031_VTOP)) %>%
```

```
  filter(Aantal > 8)
```

```
# (Afbeelding 18)
```

```
Trilling_def %>%
```

```
  filter(Bodemprofiel %in% Bodem_freq$Bodemprofiel) %>%
```

```
  ggplot(aes(x=Bodemprofiel, y= Abs_error)) +
```

```
  geom_boxplot() + xlab("Bodemprofiel ") + ylab("Absolute fout model") +
```

```
  stat_summary(
```

```
    fun.data = stat_box_data,
```

```
    geom = "text",
```

```
    hjust = 0.5,
```

```
    vjust = 0.9
```

) +

theme_classic()

Invloed op type damwandtype op error model

Damwand_freq <- Trilling_def %>%

group_by(Naam_Damwand) %>% summarise(Aantal = n(), mean_error = mean(Abs_error), mean_vibr = mean(R0031_VTOP)) %>%

filter(Aantal > 10)

(Afbeelding 19)

Trilling_def %>%

filter(Naam_Damwand %in% Damwand_freq\$Naam_Damwand) %>%

ggplot(aes(x=Naam_Damwand, y= Abs_error)) +

geom_boxplot() + xlab("Damwand ") + ylab("Absolute fout model") +

stat_summary(

fun.data = stat_box_data,

geom = "text",

hjust = 0.5,

vjust = 0.9

) +

theme_classic()

Correlatieplotten

Selecteren van de juiste variabelen

Trilling_num <- Trilling_def %>% select(R0031_VTOP, Abs_error, R0412 LENGTE_PAAL_DAMWAND, R0201_AFSTAND_MEETPUNT, distance, Slagkracht, Frequentie_Trilblok, fout_factor)

Veranderen van kolomnamen

names(Trilling_num) <- c("Trilsnelheid", "Absolute Error", "Lengte Damwand", "Afstand Woning", "Afstand Sondering", "Slagkracht", "Frequentie", "Fout Factor")

Dataset over sondering

Trilling_numson <- Trilling_def[,c(55, 15:41,43)]

Correlatie test functie maken

cor.mtest <- function(mat, ...) {

mat <- as.matrix(mat)

n <- ncol(mat)

p.mat<- matrix(NA, n, n)

diag(p.mat) <- 0

for (i in 1:(n - 1)) {

for (j in (i + 1):n) {

tmp <- cor.test(mat[, i], mat[, j],method = "spearman", ...)

```

    p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
  }
}
colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
p.mat
}

# Correlatiematrix plotten

# Error CUR met Systeemparameters (Figuur 20)
p.mat <- cor.mtest(Trilling_num[2:7])
m <- cor(Trilling_num[2:7], method = "kendall")
corrplot(m, type = "lower", p.mat = p.mat, sig.level = 0.05, tl.srt = 10, tl.cex = 0.6)

# Error CUR met Sondering (Figuur 21)
p.mat <- cor.mtest(Trilling_numson)
m <- cor(Trilling_numson, method = "spearman")
corrplot(m, type = "lower", p.mat = p.mat, sig.level = 0.1, tl.srt = 10, tl.cex = 0.6)

# 95% overschrijdingskans
Trilling_95 <- Trilling_def %>%
  rowwise() %>%
  mutate(Vcur = CUR166_Damwand_Trillingsprognose(Bodemprofiel = Bodemprofiel, straal =
R0201_AFSTAND_MEETPUNT, slagkracht = Slagkracht, percentage = 95)) %>%
  mutate(Verschil = Vcur - R0031_VTOP, Abs_error = abs(Verschil)) %>%
  mutate(Ispositive = Verschil >= 0, Is-small = R0412 LENGTE_PAAL_DAMWAND <= 14, fout_factor =
R0031_VTOP / Vcur) %>% filter(fout_factor > 0)

# Visualiseren van nauwkeurigheid CUR 166 bij 5% overschrijdingskans (Figuur 22)
ggplot() +
  geom_point(aes(x=Trilling_95$Vcur, y=Trilling_95$R0031_VTOP), size = 2, shape = 3) + xlab("95%-
waarde") +
  geom_abline(slope = 1, color = "red", size = 1.3) + ylab("Gemeten topsnelheid") +
  geom_abline(slope = 1, intercept = -6.2, color = "blue", size = 1.3, linetype = "dashed") +
  theme_light() + xlim(0, 29) + ylim(0, 11) + theme(
  legend.title = element_blank(),
  axis.title.x = element_text(size=14, face="bold"),
  axis.title.y = element_text(size=14, face="bold"))

# 0.5 * 95%-Vcur plaatje (Figuur 23)
ggplot() +
  geom_point(aes(x=0.50*Trilling_95$Vcur, y=Trilling_95$R0031_VTOP), size = 2, shape = 3) + xlab("0.5 *
Vcur") + ylab("Gemeten top snelheid") +

```



```
geom_abline(slope = 1, color = "red", size = 1.3) + ylab("Gemeten top snelheid") +
theme_light() + xlim(0, 29) + ylim(0, 11) + theme(
  legend.title = element_blank(),
  axis.title.x = element_text(size=14, face="bold"),
  axis.title.y = element_text(size=14, face="bold"))
```

CUR 50 categorische evaluatie

```
Trilling_def$Grens <- Trilling_def$Frequentie_Trilblok * 0.1 / 60 + 1
Trilling_def$Overschrijding <- Trilling_def$Grens < Trilling_def$R0031_VTOP
Trilling_def$CUR_Overschrijding <- Trilling_def$Grens < Trilling_def$Vcur
Trilling_def$Overschrijding2 <- ifelse(Trilling_def$CUR_Overschrijding, "Positive", "Negative")
Trilling_def$Waarheid <- ifelse(Trilling_def$Overschrijding == Trilling_def$CUR_Overschrijding, "True",
"False")
Trilling_def$Waarheid <- paste(Trilling_def$Waarheid, Trilling_def$Overschrijding2)
Trilling_def %>% group_by(Waarheid) %>% summarise(aantal = n())
```

CUR 95 categorische evaluatie

```
Trilling_95$Grens <- Trilling_95$Frequentie_Trilblok * 0.1 / 60 + 1
#Trilling_95$Grens <- Trilling_95$Frequentie_Trilblok * 0.25 / 60 + 2.5
Trilling_95$Overschrijding <- Trilling_95$Grens < Trilling_95$R0031_VTOP
Trilling_95$CUR_Overschrijding <- Trilling_95$Grens < Trilling_95$Vcur
Trilling_95$Overschrijding2 <- ifelse(Trilling_95$CUR_Overschrijding, "Positive", "Negative")
Trilling_95$Waarheid <- ifelse(Trilling_95$Overschrijding == Trilling_95$CUR_Overschrijding, "True", "False")
Trilling_95$Waarheid <- paste(Trilling_95$Waarheid, Trilling_95$Overschrijding2)
Trilling_95 %>% group_by(Waarheid) %>% summarise(aantal = n())
```

```
# Script 3
## CPT averaging Techniques and Feature Importance
# In deze hoofdstuk worden de analyse van hoofdstuk 7 en 8 van het verslag uitgevoerd.

# Script Geschreven volgens de Tidyverse Style Guide
# Programmeur: Marwan el Morabet; Contact: marwanelmorabet@gmail.com

# Kiezen van de juiste working directory
setwd("C:/Users/morabet/OneDrive - Stichting Deltares/Desktop/Geobrain/Scripts/Def")

#Inladen libraries
library(Boruta)

# Laden van prerequisites
# Runnen van de tweede script
source("2.CUR166.R")

# Correlatie Vtop vs. Conusweerstand (Figuur 24)
p.mat <- cor.mtest(Trilling_numson[, c(29,2:10)])
m <- cor(Trilling_numson[, c(29,2:10)], method = "spearman")
colnames(m) <- c("Trilsnelheid", "Maximum", "Minimum", "Gemiddelde", "Mediaan", "Standaard Deviatie",
"Eerste Kwartiel", "Derde Kwartiel", "Som", "Gewogen Gemiddelde")
rownames(m) <- c("Trilsnelheid", "Maximum", "Minimum", "Gemiddelde", "Mediaan", "Standaard
Deviatie", "Eerste Kwartiel", "Derde Kwartiel", "Som", "Gewogen Gemiddelde")
corrplot(m, type = "lower", p.mat = p.mat, sig.level = 0.05, tl.srt = 10)

# Correlatie Vtop vs. Schachtwrijving (Figuur 25)
p.mat <- cor.mtest(Trilling_numson[, c(29, 11:19)])
m <- cor(Trilling_numson[, c(29, 11:19)], method = "spearman")
colnames(m) <- c("Trilsnelheid", "Maximum", "Minimum", "Gemiddelde", "Mediaan", "Standaard Deviatie",
"Eerste Kwartiel", "Derde Kwartiel", "Som", "Gewogen Gemiddelde")
rownames(m) <- c("Trilsnelheid", "Maximum", "Minimum", "Gemiddelde", "Mediaan", "Standaard
Deviatie", "Eerste Kwartiel", "Derde Kwartiel", "Som", "Gewogen Gemiddelde")
corrplot(m, type = "lower", p.mat = p.mat, sig.level = 0.05, tl.srt = 10)

# Correlatie Vtop vs. Wrijvingsgetal (Figuur 26)
p.mat <- cor.mtest(Trilling_numson[, c(29, 20:28)])
m <- cor(Trilling_numson[, c(29, 20:28)], method = "spearman")
colnames(m) <- c("Trilsnelheid", "Maximum", "Minimum", "Gemiddelde", "Mediaan", "Standaard Deviatie",
"Eerste Kwartiel", "Derde Kwartiel", "Som", "Gewogen Gemiddelde")
rownames(m) <- c("Trilsnelheid", "Maximum", "Minimum", "Gemiddelde", "Mediaan", "Standaard
Deviatie", "Eerste Kwartiel", "Derde Kwartiel", "Som", "Gewogen Gemiddelde")
corrplot(m, type = "lower", p.mat = p.mat, sig.level = 0.05, tl.srt = 10)
```

```
# Dataset met belangrijke parameters
```

```
Trilling_systeem <- Trilling_def %>% mutate(rinverse = 1/R0201_AFSTAND_MEETPUNT, rinverse_root =  
sqrt(1/R0201_AFSTAND_MEETPUNT)) %>%
```

```
  select(R0031_VTOP, R0201_AFSTAND_MEETPUNT, rinverse, rinverse_root, Bodemprofiel, distance,  
As_dubbel, Frequentie_Trilblok, Slagkracht, R0412 LENGTE_PAAL_DAMWAND) %>%
```

```
  mutate(IsRotterdam = Bodemprofiel == "Rotterdam")
```

```
# Laatste systeemparemeters (Figuur 30)
```

```
p.mat <- cor.mtest(Trilling_systeem[, c(1,2,3,4,9,10)])
```

```
m <- cor(Trilling_systeem[, c(1,2,3,4,9,10)], method = "pearson", use = "pairwise.complete.obs")
```

```
colnames(m) <- c("VTop", "r", "1/r", "1/sqrt(r)", "Lengte Damwand", "Afstand Sondering")
```

```
rownames(m) <- c("VTop", "r", "1/r", "1/sqrt(r)", "Lengte Damwand", "Afstand Sondering")
```

```
corrplot(m, type = "lower", p.mat = p.mat, sig.level = 0.1, tl.srt = 10, tl.cex = 0.9)
```

```
# Numerieke variabelen transformeren naar categorische variabelen
```

```
Trilling_systeem$Bodemprofiel <- as.factor(Trilling_systeem$Bodemprofiel)
```

```
# Kolomnamen veranderen
```

```
names(Trilling_systeem) <- c("VTop", "R", "1/R", "1/root(R)", "Bodemprofiel", "Afstand Sondering", "As  
Damwand", "Frequentie", "Slagkracht", "L_damwand", "Rotterdam")
```

```
# Boruta feature importance
```

```
boruta_output <- Boruta(VTop ~ ., data=na.omit(Trilling_systeem))
```

```
# Plotten van boruta plot (Figuur 31)
```

```
plot(boruta_output, cex.axis=0.5,pos = 1, las=1, xlab="", ylab = "", main="Variable Importance", horizontal  
=T,
```

```
  whichShadow = c(F,F,F))
```

```
# Boxplots (Figuur 27, 28 en 29)
```

```
stat_box_data <- function(y, upper_limit = max(Trilling_def$R0031_VTOP) * 1.15) {
```

```
  return(
```

```
    data.frame(
```

```
      y = 0.95 * upper_limit,
```

```
      label = paste('Aantal =', length(y), '\n',
```

```
        'Mean =', round(mean(y), 1), '\n')
```

```
    )
```

```
  )
```

```
}
```

```
Trilling_def %>% ggplot(aes(y=R0031_VTOP, x=as.factor(Bodemprofiel))) + geom_boxplot() +
```

```
  stat_summary(
```

```
    fun.data = stat_box_data,
```

```
geom = "text",
hjust = 0.5,
vjust = 0.9
) +
theme_fivethirtyeight()
```

```
Trilling_def %>% ggplot(aes(y=R0031_VTOP, x=as.factor(Slagkracht))) + geom_boxplot() +
stat_summary(
  fun.data = stat_box_data,
  geom = "text",
  hjust = 0.5,
  vjust = 0.9
) +
theme_fivethirtyeight()
```

```
Trilling_def %>% ggplot(aes(y=R0031_VTOP, x=as.factor(Frequentie_Trilblok))) + geom_boxplot() +
stat_summary(
  fun.data = stat_box_data,
  geom = "text",
  hjust = 0.5,
  vjust = 0.9
) +
theme_fivethirtyeight()
```

```
Trilling_def %>% ggplot(aes(y=R0031_VTOP, x=as.factor(As_dubbel))) + geom_boxplot() +
xlab("Oppervlakte") + ylab("Trilsnelheid") +
stat_summary(
  fun.data = stat_box_data,
  geom = "text",
  hjust = 0.5,
  vjust = 0.9, size = 3
) +
theme_fivethirtyeight()
```

```
Trilling_save <- Trilling_systeem[, c(1:5, 10, 11)] %>%
write.csv("Regressie.csv")
```

```
Trilling_systeem <- Trilling_def %>% mutate(rinverse = 1/R0201_AFSTAND_MEETPUNT, rinverse_root =
sqrt(1/R0201_AFSTAND_MEETPUNT)) %>%
select(R0031_VTOP, R0201_AFSTAND_MEETPUNT, rinverse, rinverse_root, Bodemprofiel, distance,
As_dubbel, Frequentie_Trilblok, Slagkracht, R0412 LENGTE_PAAL_DAMWAND, Overschrijding) %>%
mutate(IsRotterdam = Bodemprofiel == "Rotterdam") %>%
write.csv("Classificatie.csv")
```

Script 4

Empirisch Modelleren

In deze hoofdstuk worden de analyse van hoofdstuk 9.2 van het verslag uitgevoerd.

Script Geschreven volgens de Tidyverse Style Guide

Programmeur: Marwan el Morabet; Contact: marwanelmorabet@gmail.com

Kiezen van de juiste working directory

```
setwd("C:/Users/morabet/OneDrive - Stichting Deltares/Desktop/Geobrain/Scripts/Def")
```

Runnen van de tweede script

```
source("3.CPT_Avg_Feat_Imp.R")
```

Selecteren variabelen

```
Trilling_nls <- Trilling_def %>% select(R0031_VTOP, R0201_AFSTAND_MEETPUNT)
```

```
names(Trilling_nls) <- c("Vtop", "r")
```

Stap 1: $A = 5$, $n = 0.37$

```
nls_mod <- nls(Vtop ~ A * (1/r)^n ,
```

```
  data = Trilling_nls,
```

```
  start = list(A = 10, n = 1), trace = T)
```

BIJLAGE G: 2 BLADZIJDEN CUR 166 EVALUATIE

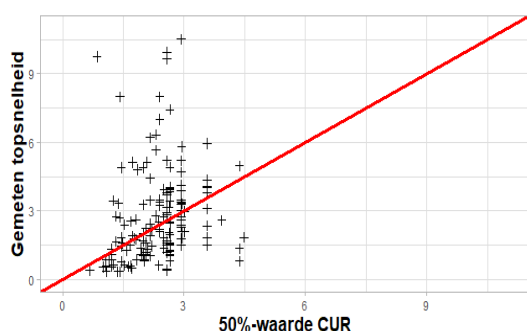
Een trillingspredictie is een rekenkundige analyse naar de trillingsintensiteit veroorzaakt door bouwwerkzaamheden. Een veelgebruikte methode voor het berekenen van de maatgevende trillingsintensiteit voor het trillend installeren van damwanden staat vermeld in de CUR 166. De berekende trillingsintensiteit wordt vervolgens getoetst aan de hand van de grenswaarden vermeld in de SBR richtlijn A: schade aan gebouwen. In dit document zal de betrouwbaarheid

Numerieke Evaluatie

Figuur 1 toont de relatie tussen de voorspelde trilsnelheid en de gemeten trilsnelheid.

Op de horizontale as staat de 50%-waarde volgens het CUR-model. Dit is de waarde die in 50% van de gevallen overschreden wordt. De 50%-waarde is gebruikt omdat dit het dichtst bij de gemeten waarde zou moeten zitten. Op de verticale as staat de maximaal gemeten trilsnelheid. (V_{top}) Een goed model volgt zo dicht mogelijk de rode lijn waarbij

$$V_{measured} = V_{cur}$$



Figuur 42; 50%-waarde CUR vs. Gemeten Topsnelheid

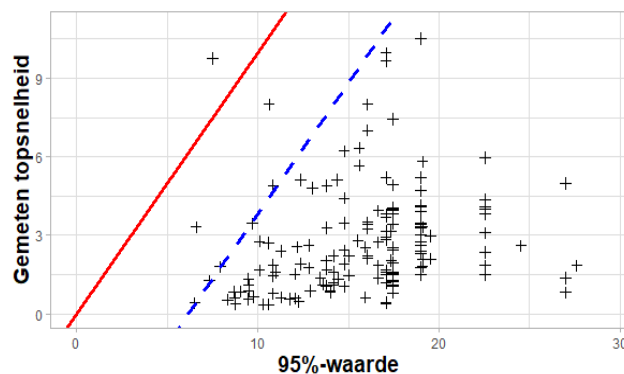
Bij het CUR-model zien we geen verband tussen de voorspelde en de gemeten waarde. Ook is het model niet in staat om grote uitschieters correct te voorspellen. Het model zit er gemiddeld 1.45 mm/s per seconde naast.

Om de nauwkeurigheid inzichtelijk te maken is er gebruik gemaakt van een NULL-model. Het null-model is in veel gevallen het meest simpele model die je kan bedenken. Bij een regressiemodel (een model die numerieke puntpredictie doet) is dit het gemiddelde. Als het CUR-model vervangen wordt door het gemiddelde (2.76 mm/s) dan zitten we er gemiddeld 1.54 mm/s naast. Dit laat zien dat het CUR-model slechts een heel klein beetje beter is dan het NULL-model. Dus vervang het

CUR-model door het getal 2,76 en je verliest maar weinig nauwkeurigheid.

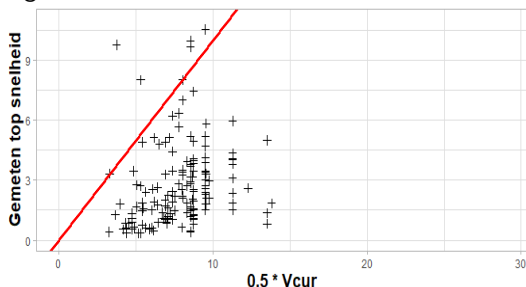
Bij het beoordelen of trillen mogelijk is, wordt er gebruik gemaakt van de 95%-waarde. Dit is de waarde die slechts in 5% van de gevallen overschreden wordt. Figuur 2 toont de relatie tussen de gemeten en de 95%-waarde. We zien het volgende.

De 95%-waarde wordt in slechts 1 van de 150 gevallen overschreden. Dit komt overeen met 0,67% van de data. Dit laat zien dat 95%-waarde uiterst conservatief is. De blauwe gestippelde lijn laat de empirisch bepaalde 95%-waarde zien. Het verlagen van de 95%-waarde van de CUR met 6 mm/s zou zorgen voor een betere benadering van de 95%-waarde.



Figuur 43; 95%-waarde CUR vs. Gemeten Topsnelheid

Op basis van de data kunnen we ook concluderen dat de 95%-waarde met een factor 2 kan worden gereduceerd. Dit zorgt voor een betere benadering van de 95%-waarde. Dit laat echter wel zien dat de CUR een uiterst veilige benadering is. Dit is te zien in figuur 3.



Figuur 44: $\frac{1}{2} \times 95\%$ -waarde versus de gemeten

Figuur 3 laat zien dat in 5 gevallen het resultaat van de CUR wordt overschreden.

Categorische Evaluatie

Vanuit praktijk oogpunt is het waardevoller om te weten of correct wordt voorspeld of de SBR-richtlijn wordt overschreden of niet. Een exacte voorspelling van de trilsnelheid is in dit geval minder belangrijk. In dit geval zijn er vier verschillende mogelijke uitkomsten.

- Het CUR-model verwacht een overschrijding en de overschrijding wordt gemeten. (TP)
- Het CUR-model verwacht geen overschrijding en de richtlijn wordt niet overschreden. (TN)
- Het CUR-model verwacht geen overschrijding maar de norm wordt wel overschreden. (FN)
- Het CUR-model verwacht een overschrijding maar de norm wordt niet overschreden. (FP)

In de volgende tabel zijn de aantallen van de vier mogelijkheden aangegeven.

Tabel 3; Confusion Matrix 50%-waarde CUR

50%-waarde		Werkelijk	
		Overschrijding	Geen Overschrijding
Voorspeld	Overschrijding	5	8
	Geen Overschrijding	22	115

In de bovenstaande tabel is laten zien dat in de 27 gevallen waarbij de SBR overschreden werd dit in slechts 5 gevallen (19%) als zodanig door de CUR is voorspeld. Ook kunnen we zien dat wanneer de CUR voorspeld dat de SBR-richtlijn niet overschreden wordt dit in 16% van

Tabel 4; Confusion Matrix 95%-waarde CUR

95%-waarde		Werkelijk	
		Overschrijding	Geen Overschrijding
Voorspeld	Overschrijding	27	123
	Geen Overschrijding	0	0

de gevallen niet klopt. Figuur 5 toont dezelfde analyse, maar dan met de 95%-waarde.

De 95%-waarde voorspelt in alle gevallen een overschrijding van de SBR-richtlijn, terwijl de overschrijding in slechts 18% van de gevallen optreedt.

Conclusie

De analyse van de nauwkeurigheid leert ons de volgende drie dingen.

- De voorspelling van de CUR laat geen sterk verband zien met de gemeten topsnelheid.
- De 95%-waarde wordt in minder dan 1% van de gevallen overschreden en is een factor 2 te conservatief.
- Het CUR-model is niet in staat om een overschrijding van de SBR-norm te voorspellen.

Als ingenieur die bezig is met trillingspredicties dien je te beseffen dat de CUR-waarden niet beschouwd moeten worden als de ultieme werkelijkheid, maar als een indicatie die iets nauwkeuriger is dan het gemiddelde.

De onnauwkeurigheid van het CUR-model laat het volgende zien.

- De noodzaak van monitoring tijdens de bouw.
- De noodzaak naar meer onderzoek over trillingen.
- De noodzaak van het empirisch valideren van modellen en indirect daarmee ook het verzamelen van data.