

Altijd op tijd?

*Ontwikkeling van een voorspellingsmodel voor
de punctualiteit van de Nederlandse Spoorwegen*

Auteur: Ronald Vinke

 **Binnenlands Reizigers Vervoer**


Universiteit Twente
de ondernemende universiteit

Altijd op tijd?

Ontwikkeling van een voorspellingsmodel voor de punctualiteit van de Nederlandse Spoorwegen

Publicatiedatum:

September 2007

Auteur:

Ronald Vinke

Studentnummer 0024368

Een onderzoek in opdracht van:

Nederlandse Spoorwegen, Binnenlands Reizigers Vervoer, Productie-eenheid Hengelo

Bedrijfsbegeleiding:

Ton Timmerman

Ir. Bas ter Laak

Afstudeercommissie:

Dr. Ir. Leo van der Wegen

Dr. Peter Schuur

Voorwoord

“Statistieken hebben ongetwijfeld bewezen dat ze de helft van de tijd fout zijn.” (Dennis Miller)

Statistiek is nooit mijn favoriete vak geweest, maar toen ik de keuze kreeg om een opdracht te doen over het voorspellen (statistiek) van een onderwerp waar heel Nederland een mening over heeft (de punctualiteit van de NS), was mijn keuze snel gemaakt. Na vol enthousiasme aan het werk gegaan te zijn (de eerste modellen zagen na enkele weken al het licht), bleek dat mijn werkzaamheden met enige scepsis bekeken werden. Menig maal heb ik te horen gekregen dat het voorspellen van de punctualiteit al vele malen geprobeerd was en dat het ónmogelijk was om de punctualiteit te voorspellen. Mijn idee hierover was echter dat het zeker wél mogelijk moest zijn, maar misschien op een andere manier dan gebruikelijk binnen de NS.

De benadering die ik gekozen heb voor het voorspellen komt vanuit de kant van de statistiek: voorspellingen kunnen gemaakt worden door de data rondom de voorspellingen te bekijken. Binnen de NS werd over het algemeen gekeken naar “het systeem”: welke invloeden bestaan op het systeem en hoe rekenen deze door. De statistische benadering doet ongeveer hetzelfde, maar laat het systeem buiten beschouwing: welke relatie bestaat er tussen variabelen die invloed hebben op de punctualiteit.

Het resultaat van dit onderzoek is te vinden in dit verslag, en vertellen of het inderdaad mogelijk was om de punctualiteit te voorspellen zou teveel verklappen. Wel wil ik in dit voorwoord nog een aantal mensen in willekeurige volgorde bedanken: Steven van Dijk (voor het bedenken van de opdracht), Ton Timmerman (voor de goede ideeën en de begeleiding), Bas ter Laak (voor de ruimte om deze opdracht te doen), Erik Vos, Henk Dommerholt, Ingrid Klaver, Henk Mulder, Jacob Maatjes en Harald Vogd (voor het beantwoorden van al mijn vragen en de koffie), Peter Schuur en Leo van der Wegen (voor de begeleiding vanaf de universiteit), Rob van Kampen (voor de 700 mailtjes met data), Wiebo Drost en Gerrit Bronsema (voor de kritische noten en de data) en verder iedereen die mij van informatie heeft voorzien die onmisbaar is voor het maken van een voorspellingsmodel.

In het bijzonder wil ik mijn ouders bedanken voor het mogelijk maken van mijn studie, en mijn vriendin voor het onafgebroken vertrouwen in een goede afloop.

Ronald Vinke
Hengelo, 11 september 2007

Managementsamenvatting

Kader en doelstellingen

In het kader van de afstudeeropdracht voor de studie Technische Bedrijfskunde is in dit rapport onderzoek gedaan naar de (on)mogelijkheden van een voorspellingsmodel voor de landelijke punctualiteit van de Nederlandse Spoorwegen (NS). De punctualiteit – uitgedrukt in het percentage treinaankomsten dat binnen een marge van drie minuten ten opzichte van de geplande aankomsttijd valt – is voor de NS een belangrijke prestatie-indicator. Jaarlijks wordt vastgesteld, in samenspraak met de overheid, welke punctualiteit gehaald moet worden. Voor 2007 is de doelstelling 86,7 procent. Het palet van stakeholders maakt de punctualiteit een heikel punt. Door de opdeling van het spoor (ProRail, overheidsonderdeel), de vervoerder (NS, private onderneming met een meerderheidsbelang van de overheid) en de druk van andere partijen op het spoor (Railion en andere goederenvervoerders) die weinig belang hebben bij de punctualiteit van de NS is het maar beperkt mogelijk voor de NS om direct invloed uit te oefenen op de realisatie van de punctualiteit. Toch probeert de NS met de middelen waar zij over beschikt tot een maximale prestatie te komen. Het ontwikkelde voorspellingsmodel heeft als doel om de NS hierbij te helpen, en aan te geven waar eventuele problemen liggen in de uitvoer van het plan. De doelstellingen waarbij het model van waarde is, kunnen als volgt weergegeven worden:

1. Benchmarking van de eigen prestaties;
2. Verhogen van de dienstverlening op dagen met een (verwachte) lage punctualiteit;
3. Aanpassen bijsturingscapaciteit;
4. Waarschuwen van klanten voor ‘rampzalige’ dagen;
5. Middel voor onderhandelingen met de overheid.

De termijn voor de prognoses voor doelstellingen 1, 2, 3 en 4 is 24-36 uur. Voor doelstelling 1 en 5 zijn ook schattingen van de jaarpunctualiteit van belang.

Werkwijze

Voor het maken van voorspellingen zijn verschillende methodes beschikbaar. In dit onderzoek zijn vier verschillende methodes met elkaar vergeleken, met doel tot een goed voorspellend model te komen. De volgende methodes zijn vergeleken:

1. Meervoudige Lineaire Regressie (MLR);
2. Artificiële Neurale Netwerken (NN);
3. k-Nearest Neighbours Algoritme (k-NN);
4. ARIMA/Box-Jenkins.

De methodes vallen in principe in twee categorieën uiteen: de methodes die wel gebruik maken van de gegevens die bekend zijn over ‘morgen’: methode MLR en NN, en de methodes die geen gebruik maken van de gegevens die bekend zijn over ‘morgen’: k-NN en ARIMA/Box-Jenkins. Alle geselecteerde methodes maken gebruik van gegevens uit het verleden. Voor deze gegevens is een dataset gebruikt over de periode januari 2004 t/m maart 2007. De ingrijpende wijzigingen in de dienstregeling die ingegaan zijn op 10 december hebben niet tot dusdanige veranderingen in de punctualiteit geleid dat de data van voor deze wijziging niet gebruikt zou kunnen worden voor het vormen van de voorspellingsmodellen.

Om tot een voorspellingsmodel te komen is onderzocht in hoeverre verschillende omgevings- en interne variabelen invloed hebben op de punctualiteit. Hiervoor is data verzameld over deze variabelen en is bekeken in hoeverre de punctualiteit een correlatie heeft met deze variabelen. De volgende variabelen hebben (aantoonbaar) invloed op de punctualiteit:

1. Hoge temperaturen;
2. Bladval (in de relatie met de windsnelheid);
3. Sneeuwval;
4. Reizigersaantallen;
5. Historische punctualiteit.

Onderzoeksresultaten

Het MLR model heeft de volgende (gestandaardiseerde) vorm:

$$\hat{Y} = 63,9 - 0,080\tilde{x}_1 - 0,149\tilde{x}_2 - 0,228\tilde{x}_3 - 0,269\tilde{x}_4 \\ + 0,264\tilde{p}_1 + 0,207\tilde{p}_7 + 0,111\tilde{p}_{14}$$

Waarin:

$\hat{Y}$	De voorspelde punctualiteit in procent
x_1	De maximale temperatuur in 0,1K
x_2	De variabele ‘windbladval’ ¹ in m/s
x_3	Sneeuwval op de schaal van 0 (geen sneeuwval) tot 1 (hevige sneeuwval)
x_4	Het aantal reizigers gegeven door de aantallen reizigers in het jaarpatroon en weekpatroon
p_T	De punctualiteit van dag $t-T$ in procent, dus p_1 is de punctualiteit van gisteren

Van dit model is vastgesteld dat het model in het algemeen en alle variabelen op zich voldoen aan de voorwaarden die gesteld worden aan MLR-modellen. Uit de gestandaardiseerde vorm blijkt dat de factoren reizigersaantal, historische punctualiteit en sneeuwval de meeste invloed hebben op de punctualiteit.

Naast het MLR model zijn ook ARIMA, k-NN en NN modellen gebouwd. Het ARIMA model heeft de vorm ARIMA(1, 1, 1)(1, 1, 1)₇, wat betekent dat gekeken wordt naar de punctualiteit van gisteren, eergisteren, vorige week en twee weken geleden voor het maken van een voorspelling.

De k-NN en NN modellen zijn gebouwd met behulp van een in het kader van dit onderzoek ontwikkelde webapplicatie. Voor het neurale netwerk is een vorm gebruikt met zeven input neuronen voor de onafhankelijke variabelen, 105 neuronen in één verborgen laag (volgens Kolmogorov: $n[2n + 1]$) en één output neuron.

Het k-NN model gebruikt de voorgaande zeven dagen ($k = 7$) en neemt het gemiddelde van de beste drie overeenkomende dagen in het verleden.

Aan de hand van een drietal prestatie indicatoren (mean square error (MSE), mean absolute percentage error (MAPE) en de bias) is bepaald dat de NN en MLR modellen de beste voorspellingen geven.

Deze twee modellen worden gecombineerd door het gemiddelde te nemen in het model $G\{\text{NN}, \text{MLR}\}$. Het $G\{\text{NN}, \text{MLR}\}$ model scoort een MSE van 11,9.

Om het gebruik van het model eenvoudig te maken is een webapplicatie gebouwd. Met behulp van deze webapplicatie is het ook mogelijk om jaarvoorspellingen te maken op basis van het model, data uit het verleden en verwachtingen over, bijvoorbeeld, de groei van het aantal reizigers.

De relatief hoge MSE waarde kan goed verklaard worden. De invloed van de parameters wordt door het model goed berekend, maar de invloed van verstoringen die niet gerelateerd zijn aan de parameters zorgt voor afwijkingen van de punctualiteit van de voorspellingen. De belangrijkste oorzaak voor deze afwijkingen is het afwijken van het niveau van infrastructuurstoringen en materieeldefecten. Op dagen met relatief weinig storingen scoort de werkelijke punctualiteit *hoger* dan de door het model voorspelde punctualiteit, op dagen met veel storingen scoort de punctualiteit *lager* dan de voorspelde punctualiteit. De willekeurige aard van de storingen maakt het zeer moeilijk om deze mee te nemen in een voorspelling over de punctualiteit.

Na het vaststellen van het voorspellingsmodel is per doel bekeken in hoeverre het model voldoet aan de doelstellingen. Doordat het model een voorspelling geeft zonder in acht te nemen wat de invloed van storingen is, is goed te zien in hoeverre de NS goed of slecht presteert gegeven de factoren in het model. Het model is daarom goed bruikbaar voor doelstellingen één en vijf. Het model kan ook voor

¹ De windsnelheid op dagen dat er bladval is.

doelstellingen twee, drie en vier, zij het met enig voorbehoud, gebruikt worden. Aansluitend wordt ook aangegeven in hoeverre het model kan helpen om de gevoeligheid van de punctualiteit voor de versturende variabelen te verkleinen.

Conclusies en aanbevelingen

Concluderend kan vastgesteld worden dat het goed mogelijk is de punctualiteit aan de hand van invloedrijke variabelen te voorspellen. De invloed van willekeurige gebeurtenissen als materieeldefecten en infrastructuurstoringen is echter moeilijk te voorspellen, wat zorgt voor afwijkingen van de voorspelde punctualiteit en de werkelijke punctualiteit.

Afsluitend wordt de NS, naast enkele andere aanbevelingen, aanbevolen om de ‘zoektocht’ naar invloedrijke onafhankelijke variabelen voort te zetten. Wanneer een invloedrijke variabele gevonden wordt, kan deze relatief eenvoudig opgenomen worden in het MLR en NN model.

Management summary

Research area and objectives

This report contains the research about the (im)possibilities of the prediction of the punctuality of the Dutch Railways (NS) network. This research has been performed as a graduation assignment for the study Industrial Engineering and Management. The punctuality – defined as the percentage of train arrivals within three minutes of the planned arrival – is a very important performance indicator for the NS. The target for this indicator is defined yearly and is largely influenced by the opinion of the Dutch government. For 2007, the target punctuality is 86,7 percent.

The punctuality is a complex figure due to the different responsibilities of stakeholders involved with the realisation of this figure. The Dutch railway network is divided in the actual network (the ‘rails’) and the operators on this network. The actual network and the handling of the operators on the network are handled by ProRail, which is a governmental organization. Next to the NS, operators include other public transportation companies like Syntus, but also several cargo handlers like Railion. Although it is hard for the NS to realise the targets for punctuality due to all the different interests of the stakeholders, NS tries very hard to improve the punctuality to meet the targets. The model for forecasting punctuality can help the NS to improve their understanding and handling of punctuality.

The main targets for which the model will be used are:

1. Benchmarking of the performance;
2. Improving the service towards customers on days with an expected dip in the punctuality;
3. Adjustment of the capacity of calamity management;
4. Warning customers on bad days;
5. Tool for negotiations with the government.

The forecasts should be based on a 24 to 36 hour basis for targets one, two, three and four. Targets one and five also require estimations of the punctuality on a yearly basis.

Working method

There are several forecasting methods available. For this research, four different methods have been compared to find the most suitable model. The following methods have been considered:

1. Multiple linear regression (MLR);
2. Artificial Neural Networks (NN);
3. k-Nearest Neighbours Algorithm (k-NN);
4. ARIMA/Box-Jenkins.

The methods can be categorized in two categories: models that make use of information that is available about the future (MLR and NN), and models that do not use this information (k-NN and ARIMA). Both categories use historical information to ‘learn’ to make forecasts. The historical information used in this research includes the period from January 2004 until March 2007. The substantial modifications of the schedule in December 2006 did not lead to such ample changes of the punctuality that it would be impossible to use data from the period before December 2006 for the estimation of the prediction models.

The creation of the prediction models is started by the identification of the independent variables which are of influence on the dependent variable (punctuality). To identify the independent variables extensive data are gathered and correlations between this data and the punctuality are researched. The following variables are examined among others and are identified as useable for predictions:

1. High temperatures;
2. Leaf-fall;
3. Snow;
4. Number of travellers;
5. Historical punctuality.

Results

The MLR-model is determined as follows (standardised form):

$$\hat{Y} = 63,9 - 0,080\tilde{x}_1 - 0,149\tilde{x}_2 - 0,228\tilde{x}_3 - 0,269\tilde{x}_4 \\ + 0,264\tilde{p}_1 + 0,207\tilde{p}_7 + 0,111\tilde{p}_{14}$$

In which:

$\hat{Y}$	The predicted punctuality in percent
x_1	The maximal temperature in 0.1K
x_2	De variable ‘wind speed/leaf-fall’ ² in m/s
x_3	Snow, scaled from 0 (no snow) to 1 (heavy snow)
x_4	The number of travellers
p_T	The historical punctuality of $t-T$ in percent, so p_{t-1} is the punctuality of yesterday

This model and the separate variables fulfil the requirements for multiple linear regression models. The standardised form of the equation indicates that the variables ‘number of travellers’, ‘historical punctuality’ and ‘snow’ have the largest influences on punctuality.

The structure of the ARIMA is model is ARIMA(1, 1, 1)(1, 1, 1)₇. So the model includes two AR and two MA terms, with differences taken on seasonal (defined as a week in case of the punctuality) and daily level.

The k-NN and NN models have been created using a web application which has been built for this research. The NN model has seven input neurons for the seven variables influencing the punctuality, 105 neurons in one hidden layer (Kolmogorov: $n[2n + 1]$) and one output neuron.

The k-NN model uses the seven foregoing days ($k = 7$) and takes the mean of the best matching three days.

For the assessment of the different models, three indicators are used (mean square error (MSE), mean absolute percentage error (MAPE) and the bias). Using these indicators the models NN and MLR are selected. Using these models the model G{NN, MLR} is created, which takes the average of the predictions of the NN and MLR models. The G{NN, MLR} model scores a MSE of 11.9.

The usage of the model is made simple by the development of a web application. Using the web application, it is also possible to create predictions on a yearly basis using data from the past and predictions about, for instance, the growth of the number of travellers.

The relatively high value of the MSE can be explained by the lack of variables in the models concerning specific cases of disturbances on the planned rail services. These disturbances mainly consist of infrastructure failures and breakdowns of trains. Days with a low level of these disturbances result in higher punctualities than predicted, and days with a high level of these disturbances in lower punctualities than predicted. The main problem why these disturbances cannot be included in prediction models is the random nature of these disturbances.

After the determination of the forecasting model it is determined for each target as stated before whether the model is able to fulfil this target. As the model predicts the punctuality without considering the random disturbances as infrastructure failures, the model is very usable as a model to measure the performance given the parameters in the model. The model is therefore very usable for the targets one and five. The model can also be used for targets two, three and four, though with some restrictions. It is also indicated how the NS can use the information from the models to improve the punctuality by adjusting the parameters of the model through changing the sensitivity of the network for the independent variables.

² The wind speed during leaf-fall season.

Conclusions and future work

Concluding the research it can be determined that it is possible to predict the punctuality given a set of independent variables. The influence of random disturbances like train breakdowns and infrastructure failures is harder to predict, which results in errors in the predicted and realized punctuality.

The Dutch Railways are advised to invest time in future work to improve the prediction models. For instance, the inclusion of other influential independent variables in the prediction model can result in better predictions.

Inhoudsopgave

1	INLEIDING	1
1.1	NEDERLANDSE SPOORWEGEN	1
1.1.1	NS Groep	1
1.1.2	Organisatieonderdeel Knooppuntontwikkeling	3
1.1.3	Organisatieonderdeel Reizigersvervoer	3
1.2	PROBLEEMIDENTIFICATIE	5
1.2.1	Stakeholders	6
2	PROBLEEMSTELLING EN PLAN VAN AANPAK	9
2.1	PROBLEEMSTELLING	9
2.2	DOMEIN VAN HET ONDERZOEK	10
2.3	PLAN VAN AANPAK	10
2.3.1	Hoe wordt punctualiteit gemeten?	10
2.3.2	Hoe werkt voorspellen?	10
2.3.3	Is de oude data bruikbaar voor de nieuwe voorspellingen?	10
2.3.4	Welke externe en interne variabelen hebben invloed op de punctualiteit?	11
2.3.5	Hoe kunnen de verschillende variabelen worden gebruikt voor het maken van een voorspelling over de punctualiteit?	12
2.3.6	In hoeverre voldoet het model aan de doelstellingen en hoe kan de NS het model gebruiken?	12
2.3.7	Hoe kunnen de gesignaleerde versturende factoren aangepakt worden?	12
2.3.8	Leeswijzer	12
3	PUNCTUALITEIT	13
3.1	SOORTEN PUNCTUALITEIT	13
3.2	DEELNEMENDE PARTIJEN	13
3.2.1	NS	13
3.2.2	ProRail	14
3.2.3	Reizigers	15
3.2.4	Goederenvervoerders	15
3.3	HET METEN VAN PUNCTUALITEIT	16
3.3.1	Treinidentificatie en routing	16
3.3.2	Punctualiteitsmetingen	16
3.3.3	Uitgevallen treinen en (grote) calamiteiten	18
3.3.4	Reizigerspunctualiteit	18
3.4	REEDS UITGEVOERDE ONDERZOEKEN	18
3.4.1	Opbouw van vertraging	18
3.5	CONCLUSIE	20
4	VOORSPELLEN	21
4.1	HET DOEL VAN DE VOORSPELLING	21
4.1.1	Een voorspelling, en wat dan?	21
4.2	VOORSPELLINGSMETHODES	23
4.2.1	Kwalitatieve voorspellingsmethodes	23
4.2.2	Kwantitatieve voorspellingsmethodes	24
4.2.3	Combinatiemethodes	25
4.3	METHODE SELECTIE EN THEORIE	25
4.3.1	Methode selectie	25
4.3.2	Notatie	26
4.3.3	Meervoudige lineaire regressie	26
4.3.4	Box-Jenkins (ARIMA) methode	28
4.3.5	Artificiële Neurale Netwerken	30
4.3.6	k-Nearest Neighbor algoritme	34
4.3.7	Welke onafhankelijke variabelen worden gebruikt?	36
4.3.8	Welke (combinatie van) voorspellingen?	38

4.4	CONCLUSIE	39
5	OUDE DATA	41
5.1	KARAKTERISTIEKEN VAN DE DATA	41
5.2	VARIANTIEANALYSE PUNCTUALITEIT	42
5.3	VARIANTIEANALYSE TREINACTIVITEITEN	42
5.4	CONCLUSIE	44
6	INVLOEDEN OP PUNCTUALITEIT	45
6.1	EXTERNE VARIABELEN	45
6.1.1	<i>Weer</i>	45
6.1.2	<i>Overige externe invloeden</i>	51
6.1.3	<i>Reizigersaantallen</i>	51
6.1.4	<i>Vakanties</i>	52
6.1.5	<i>Files</i>	53
6.1.6	<i>Suïcides</i>	54
6.2	INTERNE VARIABELEN	55
6.2.1	<i>Ziektecijfers personeel</i>	55
6.2.2	<i>Vrijtijds aanspraak personeel</i>	56
6.2.3	<i>Treinactiviteiten</i>	57
6.2.4	<i>Materieel in onderhoud</i>	57
6.2.5	<i>Historische punctualiteit</i>	58
6.3	CONCLUSIE	59
7	VOORSPELLINGSMODELLEN	61
7.1	VALIDATIESET	61
7.2	MODELLEN	62
7.2.1	<i>Meervoudige lineaire regressie</i>	62
7.2.2	<i>ARIMA/Box-Jenkins</i>	66
7.2.3	<i>k-Nearest Neighbours</i>	68
7.2.4	<i>Neuraal netwerk</i>	70
7.3	MODELPRESTATIES	73
7.3.1	<i>ARIMA</i>	74
7.3.2	<i>k-NN, NN en MLR</i>	75
7.3.3	<i>Resultaten</i>	75
7.4	CONCLUSIE	80
8	TOEPASSINGEN	81
8.1	MATERIEELDEFECTEN EN INFRASTRUCTUURSTORINGEN	81
8.2	GEBRUIK VAN HET MODEL	81
8.2.1	<i>Workflow</i>	81
8.2.2	<i>Jaarprognose</i>	83
8.2.3	<i>Overige mogelijkheden van de applicatie</i>	84
8.3	TOEPASSING VAN HET MODEL	86
8.3.1	<i>Benchmarking (1) en onderhandelingsmiddel (5)</i>	86
8.3.2	<i>Waarschuwen van reizigers (4)</i>	86
8.3.3	<i>Verhogen dienstverlening (2)</i>	87
8.3.4	<i>Aanpassen bijsturingcapaciteit (3)</i>	87
8.4	CONCLUSIE	88
9	BEÏNVLOEDING FACTOREN	89
9.1	WINDBLADVAL	89
9.2	MAXIMALE TEMPERATUUR	90
9.3	REIZIGERSAANTALLEN	90
9.4	SNEEUWVAL	91
9.5	VERLEDEN PUNCTUALITEIT	91
9.6	MATERIEELDEFECTEN EN INFRASTRUCTUURSTORINGEN	92
9.6.1	<i>Materieeldefecten</i>	92
9.6.2	<i>Infrastructuurstoringen</i>	92
9.7	CONCLUSIE	92

10	CONCLUSIE	93
10.1	VOORSPELLINGSMODEL	93
10.2	AANBEVELINGEN	95
10.2.1	<i>Uitbreiding van de onafhankelijke variabelen met storingen</i>	95
10.2.2	<i>Inventarisatie van risico's en effect van storingen</i>	95
10.2.3	<i>Verbeter weerdata voor jaarprognoses</i>	96
10.2.4	<i>Laat punctualiteit niet leidend zijn: bepaal maatregelen vanuit klantperspectief</i>	96
10.2.5	<i>Blijf zoeken naar variabelen die invloed hebben</i>	96
	LIJST VAN AFKORTINGEN	97
	GEBRUIKTE AFKORTINGEN	97
	GEBRUIKTE AFKORTINGEN VAN STATIONS	97
	LIJST VAN FIGUREN EN TABELLEN	98
	FIGUREN	98
	TABELLEN	99
	LITERATUUR	100
	MET AUTEUR	100
	ZONDER AUTEUR	100
	BIJLAGE A: VOORSPELLINGEN EN WAARNEMINGEN	101
	BIJLAGE B: ARIMA MODEL	105

1 Inleiding

In het kader van de afstudeeropdracht voor de studie Technische Bedrijfskunde zal in dit rapport onderzoek gedaan worden naar de (on)mogelijkheden van een voorspellingsmodel voor de landelijke punctualiteit van de Nederlandse Spoorwegen (NS).

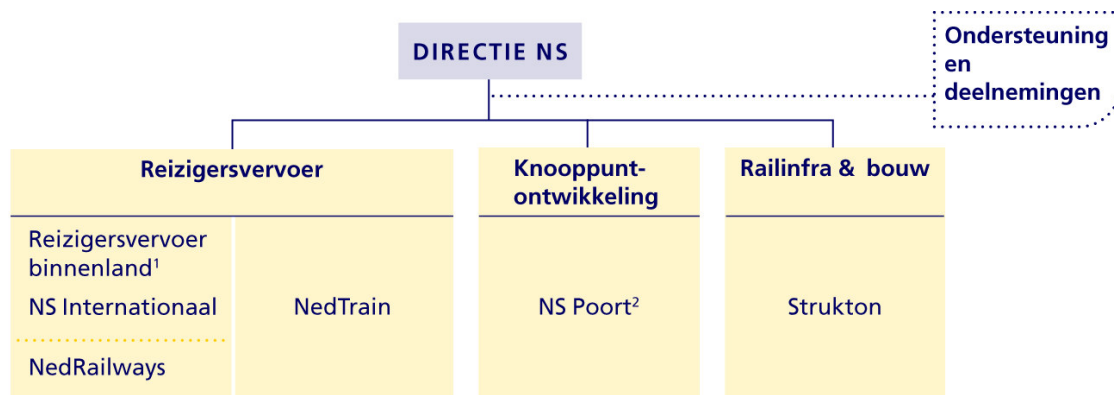
In deze inleiding zal een beschrijving geven van de NS, deze beschrijving begint met een historie van de NS, een kort financieel overzicht en een beschrijving van de bedrijfsonderdelen met bijzondere aandacht voor NS Reizigers. Vervolgens zal het probleem dat zich voordoet bij de NS geïdentificeerd worden, gevolgd door een identificatie van de belangrijkste stakeholders.

1.1 Nederlandse Spoorwegen

De NS is voornamelijk bekend om het vervoer van reizigers, echter de NS is een breder bedrijf dan alleen het vervoer van reizigers per trein. De NS is op te delen in een drietal takken: reizigersvervoer, knooppuntontwikkeling en bouw (Strukton). Onder deze takken vallen verschillende bedrijfsonderdelen die een gedeelte van de dienstverlening voor hun rekening nemen. De verschillende onderdelen zullen hieronder kort beschreven worden. Binnenlands Reizigers Vervoer (BRV), het onderdeel waar dit onderzoek plaats heeft gevonden, zal uitgebreider behandeld worden.

1.1.1 NS Groep

De NV NS is ontstaan in 1937 uit een fusie van de toenmalige Staatsspoorwegen (SS) en de Hollandse IJzeren Spoorwegmaatschappij (HIJSM). In de periode daarvoor is het spoornetwerk gegroeid van een enkele lijn Amsterdam – Haarlem tot het netwerk zoals dat nu ongeveer in Nederland ligt. De NS is in die periode nog zo goed als monopolist op het gebied van reizigersvervoer. Dit verandert in het interbellum wanneer de bus zijn intrede doet. In de jaren 60 loopt het reizigersaantal sterk terug doordat de auto bereikbaar werd voor iedereen, de tekorten die ontstaan worden gedekt door de staat. In de jaren 80 neemt de filedruk in Nederland sterk toe en ook milieubewustzijn zorgt voor een toename in reizigersaantallen. In 1992 stemmen de Eerste en Tweede Kamer in met de verzelfstandiging van de NS – mede gedwongen door een Europese verplichte scheiding netwerk en vervoerexploitatie. In 1995 splits de NS zich zoals aangegeven in een gedeelte netwerk (het huidige ProRail) en de NS Groep. In 2005 was de NS een bedrijf met meer dan 24.000 medewerkers (uitgedrukt in mensjaren), verdeeld over de takken reizigersvervoer (69%), knooppuntontwikkeling en –exploitatie (14%), bouw (14%) en overige (3%). (NS, 2007; NS Intranet, 2006)



¹ waarin NS Reizigers en NS Commercie

² waarin NS Stations, NS Vastgoed en NS Commercie Stationsontwikkeling

1.1.1.1 Financieel

De NS is financieel gezond. Het nettoresultaat over 2006 is uitgekomen op 197 miljoen euro, een kleine achteruitgang ten opzichte van 2005 (197 miljoen euro, zie ook Tabel 1.1), deze daling wordt veroorzaakt door het effect van de waardevermindering van de uitgestelde belastingvorderingen. De ROI bedraagt 3,7 procent, het behalen van een hoge ROI is voor een organisatie als NS van minder belang dan voor een for-profit organisatie. Voor NS zijn cijfers met betrekking tot punctualiteit, oordeel van de klant, zitplaatscapaciteit in de trein, etc., van veel groter belang. De NS heeft afspraken gemaakt met de overheid over de dienstverlening die ze moeten bieden. Deze afspraken zijn belangrijk voor de NS omdat in 2015 waarschijnlijk meer bidders zijn bij de openbare aanbesteding van het Nederlandse railnetwerk. Een goed presterende NS maakt meer kans om de biding te winnen dan een NS met onvoldoende resultaten. In Tabel 1.2 staat een overzicht van de klantbeoordelingen in 2004, 2005 en 2006.

	2006	2005	2004
Totaal bedrijfsopbrengsten	3.846	3.487	2.949
Totaal bedrijfslasten	3.536	3.186	2.805
Bedrijfsresultaat	310	288	144
Totaal financiële baten en -lasten	44	45	30
Resultaat voor belasting	354	333	174
Vennootschapsbelasting	-157	-112	-142
Nettoresultaat	197	221	32

Tabel 1.1 Geconsolideerde Winst-en-verliesrekening 2005 NV Nederlandse Spoorwegen, in miljoenen euro's (NS Jaarverslag 2006, 2007)

Een ander belangrijk kenmerk voor de NS is de punctualiteit. In 2005 werd een punctualiteit van 84,7 procent behaald (2004: 86%, 2006: 84,8%), het prestatiecontract met de overheid vereiste een punctualiteit van minimaal 86 procent in 2006 (NS, 2006). In 2007 wordt een punctualiteit van 86,7 procent geëist. In Tabel 1.2 staat een opsomming van de klantoordelen in 2005 en 2004. Het streefdoel van de NS is dat 70 procent van de klanten minimaal een zeven of hoger geeft.

Punctualiteit

Op het begrip punctualiteit zal later in dit verslag nog uitgebreid ingegaan worden. Voor nu volstaat te weten dat de punctualiteit gedefinieerd is als het percentage treinen dat op een aantal afgesproken stations voor, op of binnen een marge van drie minuten na de afgesproken aankomsttijd aankomt. Een meer nauwkeurige term voor deze punctualiteit is *aankomstpunctualiteit*. Naast aankomstpunctualiteit bestaat ook *vertrekpunctualiteit*, hier zal later nog op ingegaan worden.

Voor de intercity Deventer – Hengelo die om 9:27:00 zou moeten aankomen geldt dus:

Aankomst	Vertraging	Op tijd?
09:25:57	00:00:00	Ja
09:29:59	00:02:59	Ja
09:30:00	00:03:00	Nee
09:35:15	00:08:15	Nee

	2006	2005	2004
Op tijd rijden	39%	40%	39%
Informatie bij ontregeling in de trein	48%	47%	45%
Informatie bij ontregeling op het station	47%	48%	48%
Sociale veiligheid in de trein	74%	75%	72%
Zitplaatscapaciteit in de trein in de spits	71%	70%	
Reinheid in de trein	53%	55%	53%
Reinheid op het station	54%	52%	50%
Algemeen oordeel reizen per trein	68%	67%	65%

Tabel 1.2 Klantbeoordeling, percentage hoger dan een zeven (NS Jaarverslag 2006, 2007)

1.1.1.2 Missie

De missie van de NS luidt als volgt: *Meer reizigers veilig, op tijd, en comfortabel vervoeren via aantrekkelijke stations*. Voor het realiseren van deze missie zijn vijf hoofddoelstellingen opgesteld:

1. Op tijd rijden
2. Informatie verstrekken en service verlenen
3. Bijdragen aan sociale veiligheid
4. Voldoende vervoerscapaciteit creëren
5. Zorgen voor schone treinen en stations.

Aanvullend heeft de NS als doelstelling met betrekking tot de klantbeoordelingen dat minimaal 70% van de reizigers de NS een rapportcijfer van 7 of hoger geeft. Om te kunnen blijven investeren in de verbetering van de dienstverlening, heeft de ROI en streefdoel van 7%.

1.1.2 Organisatieonderdeel Knooppuntontwikkeling

Knooppunten zijn, kort gezegd, de grotere stations in Nederland. Knooppunten nemen een belangrijke plek in voor de strategie van de NS, welke gericht is op '[...] veilig, op tijd en comfortabel vervoeren [van passagiers] via aantrekkelijke stations' (NS Intranet, 2006). De component 'aantrekkelijke stations' in deze doelstelling houdt ondermeer in dat knooppunten ontwikkeld worden tot aantrekkelijke wacht- en verblijfplaatsen. Voorbeelden van ontwikkelingen die hiermee te maken hebben zijn onder andere de AH To Go winkels op NS stations en het aanbieden van vergaderfaciliteiten. Andere aspecten die hierbij horen is het schoonmaken van stations, de veiligheid op stations en de informatievoorziening aan reizigers. Al deze taken worden uitgevoerd door *NS Stations*. (NS Intranet, 2006).

De omgeving van stations is waardevol: de bereikbaarheid is groot en de locatie is veelal in het hart van de stad. *NS Vastgoed* is de vastgoedontwikkelaar en belegger op knooppunten van het openbaar vervoer. De NS heeft verspreid over Nederland ongeveer 4.700 hectare grond in bezit, veelal op binnenstedelijke locaties. Deze locaties worden door NS Vastgoed ontwikkeld tot woon-, werk- en winkelgebieden. De ontwikkeling van deze locaties heeft mede tot doel om het verblijf van klanten op stations te veraangemen. (NS Intranet, 2006).

1.1.3 Organisatieonderdeel Reizigersvervoer

1.1.3.1 NedRailWays

De NS is niet meer alleen in Nederland actief. Met het onderdeel *NedRailways* verwerft en exploiteert de NS raillicenties in het buitenland. NedRailways is in het buitenland hiermee te vergelijken met aanbieders als Syntus in Nederland. De liberalisering van de spoornetwerken in Europa zijn een belangrijke reden dat NedRailways bestaat. Zowel in Nederland als in de rest van Europa worden de netwerken (openbaar) aanbesteed. NedRailways exploiteert een groot netwerk in Groot-Brittannië (2800km spoor, 475 stations, 266 treindiensten/dag) en probeert met *NetBahnen* concessies in Duitsland te winnen.

1.1.3.2 NS Internationaal

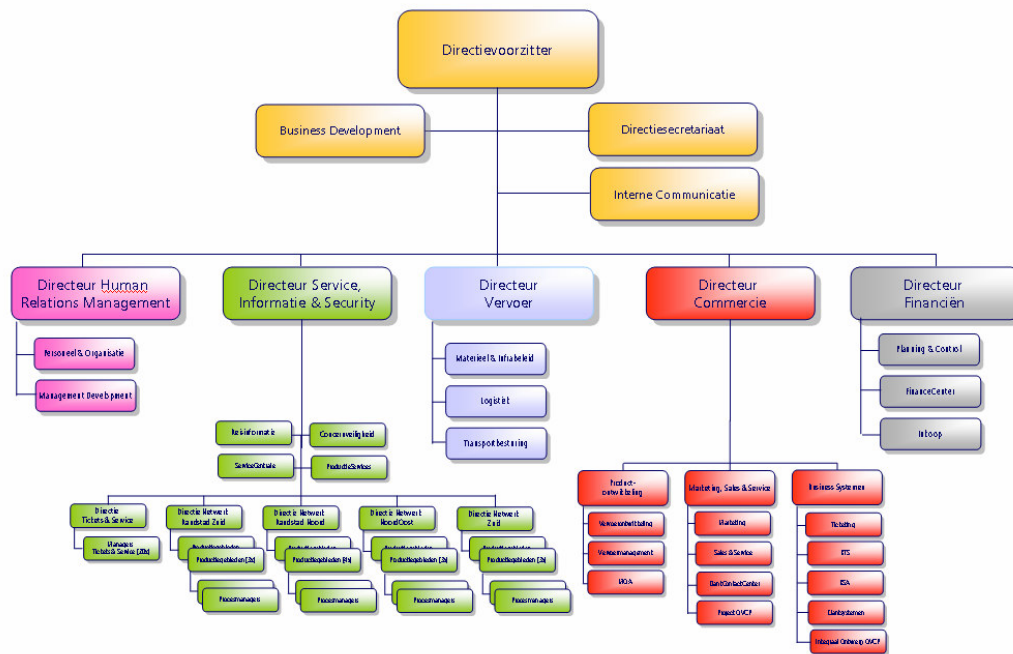
Andere internationale activiteiten worden uitgevoerd door *NS Internationaal*. NS Internationaal heeft eigen materieel, machinisten, trainmanagers en een eigen verkooporganisatie. NS Internationaal exploiteert de Thalys (Amsterdam – Parijs), ICE International (Amsterdam – Oberhausen, Duisburg, Düsseldorf, Keulen, Frankfurt), Intercity Brussel (Amsterdam – Brussel) en de Intercity Berlijn (Amsterdam – Berlijn).

1.1.3.3 NedTrain

Het onderhoud van treinstellen wordt verzorgd door *NedTrain*. NedTrain verzorgt, naast treinstellen van BRV, ook onderhoud van treinstellen van onder andere Syntus, NoordNed, Deutsche Bahn en de RET. (NedTrain, 2007).

1.1.3.4 Binnenlands Reizigers Vervoer

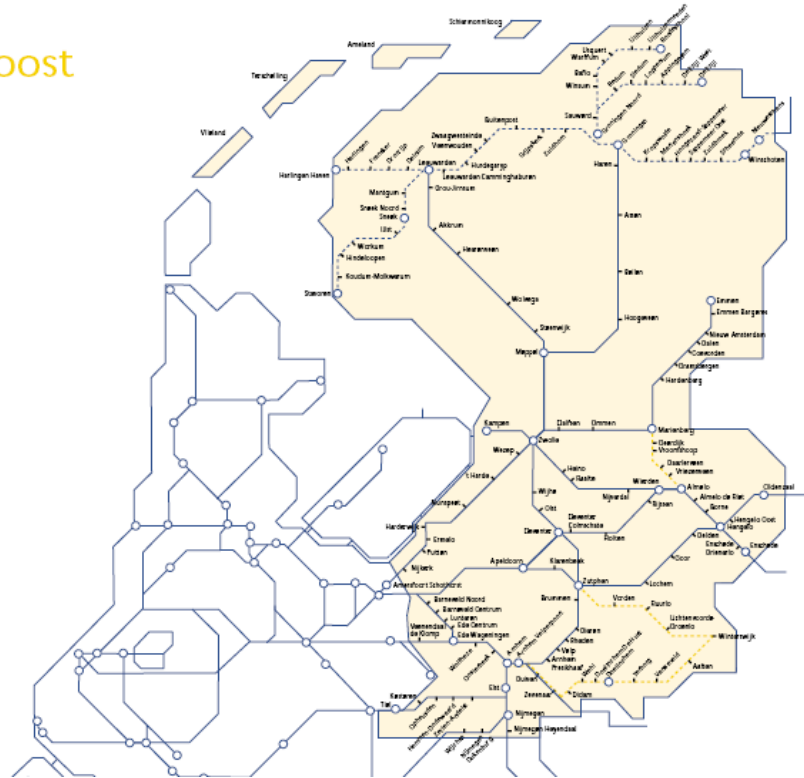
Het daadwerkelijke vervoer van reizigers binnen Nederland wordt uitgevoerd door *Binnenlands Reizigers Vervoer* (BRV). BRV zorgt dat treinen op tijd rijden, er voldoende comfortabel materieel is en er goede service, informatie en een adequate opvang bij verstoringen wordt gegeven. BRV is opgedeeld in vier netwerken: Randstad Zuid, Randstad Noord, Zuid en Noordoost. De netwerken zijn verdeeld in productie-eenheden (PE's). Netwerk Noordoost is verdeeld in drie PE's: Noord, Hengelo en Arnhem-Nijmegen. Dit onderzoek is uitgevoerd bij PE Hengelo. In Figuur 1.2 is netwerk Noordoost aangegeven, Figuur 1.1 laat de relatie van de verschillende netwerken tot de directie zien.



Figuur 1.1 Organigram Binnenlands Reizigers Vervoer

BRV vervoert dagelijks meer dan één miljoen reizigers met 5000 stop- en intercitydiensten. De totale lengte van het spoor is 2.800 km met 390 stations. In vergelijking met het NedRailways netwerk in Groot-Brittannië is het Nederlandse netwerk kleiner, maar drukker bezet (minder stations, maar veel meer treindiensten). Daar staat tegenover dat het gebied waar NedRailways opereert ook dunner bevolkt is dan Nederland.

Regio Noordoost



Figuur 1.2 Netwerk Noordoost

1.2 Probleemidentificatie

Zoals al eerder aangegeven heeft de NS afspraken met de Nederlandse overheid over verschillende aspecten van reizigersvervoer op het Nederlandse netwerk van sporen. In het kader van de concessieafspraken met het ministerie van Verkeer en Waterstaat publiceert NS jaarlijks een vervoerplan. In dit vervoerplan geeft de NS aan welke ambitie minimaal nagestreefd wordt met betrekking tot de volgende prestatie-indicatoren:

1. Het waarborgen van een verantwoorde mate van veiligheid ten behoeve van zowel de reizigers als het personeel.
2. Punctualiteit.
3. Procentuele beschikbaarheid van zitplaatsen.
4. Het zowel op stations als in de trein aan reizigers geboden serviceniveau, te weten reinheid van stations en treinen en de geboden informatie voorziening aan de reiziger (MinVW, 2005).

In het kader van dit onderzoek is voornamelijk punt twee, punctualiteit, van belang. De NS heeft in het vervoerplan 2006, in samenspraak met de staatsecretaris, een doel gesteld van 86 procent voor de punctualiteit (NS, 2006). Een andere belangrijke grens is een punctualiteit van 86,8 procent, aangezien de NS dan een extra tariefsverhoging mag doorvoeren in het daarop volgende jaar (MinVW, 2005). Zoals eerder vermeld is de punctualiteit die gehaald werd door de NS in 2006 te laag (84,8%). Het niet halen van de voorgenomen punctualiteit heeft voor de NS belangrijke nadelen. De publieke opinie over de NS loopt immers (weer) een deuk op, en daarnaast verzwakt dit de positie van de NS tegenover de overheid in toekomstige concessies. Sturen op punctualiteit is daarom erg belangrijk, echter, dit gebeurt nu allemaal achteraf. De NS is tot op heden niet in staat om de punctualiteit (dag, week, maand of jaar) gefundeerd te voorspellen, hoewel hier binnen de organisatie wel veel vraag naar is. Voorspellingen kunnen door verschillende mensen (operationeel managers) binnen de organisatie wel globaal gemaakt worden, maar dat is allemaal gebaseerd op impliciete kennis en vage relaties. Een aantal voorbeelden van deze relaties is:

- Laagstaande zon zorgt voor aanrijdingen.

- Bepaalde periodes van het jaar hebben een verhoogde kans op zelfdodingen (december feestdagen, mei/juni).
- Grote temperatuurverschillen (zowel warm-koud als koud-warm).
- Extreme weersituaties (wind, sneeuw, ijzel).
- Grote reizigersaantallen.

Om deze kennis expliciet te maken is NS op zoek naar een model waarmee deze relaties kwantitatief aangetoond en verklaard worden. De relaties zouden in een model gevat moeten worden, waarmee door input van een aantal variabelen die invloed hebben op de punctualiteit een voorspelling gedaan kan worden voor de punctualiteit. Wanneer blijkt dat de punctualiteit onvoldoende is kan er bijgestuurd worden zodat de punctualiteit waarschijnlijk toch nog boven de gewenste waarde uitkomt.

1.2.1 Stakeholders

Stakeholders ('belanghebbenden') zijn volgens Freeman (1984) individuen of groepen die belang hebben of betrokken zijn bij een specifieke organisatie of besluit. Stakeholders zijn dus geen aandeelhouders (*shareholders*): een stakeholder hoeft geen financieel belang te hebben bij een organisatie. Om stakeholders beter te kunnen categoriseren heeft Mitchell et al. (1997) aan deze definitie toegevoegd dat de stakeholders minimaal één van de volgende drie eigenschappen moet bezitten: macht, legitimiteit en urgentie (*power, legitimacy, urgency*). Met behulp van deze drie eigenschappen kan een verdere indeling gemaakt worden naar de soort eigenschappen dat een stakeholder bezit:

1. Definitive stakeholders
Deze stakeholders bezitten alle eigenschappen en zijn daarom zeer belangrijk.
2. Expectant stakeholders
Deze groep stakeholders heeft sterke belangen, maar omdat ze slechts twee van de drie mogelijke eigenschappen bezitten, zijn de minder belangrijk dan de definitive stakeholders.
 - a. Dependent ('afhankelijke') stakeholders
Legitimiteit en urgentie
 - b. Dangerous ('gevaarlijke') stakeholders
Macht en urgentie
 - c. Dominant ('dominante') stakeholders
Macht en legitimiteit
3. Latent stakeholders
De laatste groep stakeholders bezitten allemaal slechts één eigenschap en worden daarom als minder belangrijk beschouwd. Het belangrijkste is om deze groep geïnformeerd te houden.
 - a. Dormant ('slapende') stakeholders
Macht
 - b. Discretionary ('ingetogen') stakeholders
Legitimiteit
 - c. Demanding ('eisende') stakeholders
Urgentie

Hieronder zal een overzicht gegeven worden van de verschillende stakeholders met betrekking tot punctualiteit(voorspelling) gegeven worden, inclusief hun classificatie volgens Mitchell et al.

Stakeholder	NS Groep
Classificatie	Definitive
Beschrijving	De NS Groep heeft, door haar afspraken met de overheid, veel belang bij het behalen van het doel met betrekking tot de punctualiteit (urgentie), daarnaast heeft de directie van de NS Groep ook de macht om wijzigingen door te voeren met betrekking tot de factoren die invloed hebben op de punctualiteit. Als 'eigenaar van het proces' heeft de NS Groep ook legitimiteit om stakeholder te zijn voor dit probleem.

Stakeholder	Binnenlands Reizigers Vervoer (BRV)
Classificatie	Definitive
Beschrijving	Als uitvoerend bedrijfsonderdeel is BRV direct verantwoordelijk voor de behaalde punctualiteit (legitimiteit). BRV heeft ook, voor zover niet bepaald door externe factoren als weer, invloed op de punctualiteit die behaald wordt (macht). Daarnaast heeft BRV door haar verantwoordelijke rol richting de NS Groep ook de urgentie om de punctualiteit te verhogen.
Stakeholder	Nederlandse overheid
Classificatie	Definitive
Beschrijving	De Nederlandse overheid stelt het netwerk van sporen onder bepaalde afspraken beschikbaar aan de NS, maar is daarnaast (in de vorm van het ministerie van financiën) ook 100 procent aandeelhouder van de NS (macht). Als vertegenwoordiger van de burger heeft de overheid ook een legitiem belang bij de punctualiteit van de NS, om dezelfde reden heeft overheid ook urgentie bij een verhoging van de punctualiteit.
Stakeholder	Reizigers
Classificatie	Dependent
Beschrijving	De reizigers van de NS hebben urgentie en legitimiteit bij een hogere punctualiteit. Reizigers bezitten echter geen macht, aangezien de alternatieven ontoereikend of niet aanwezig zijn. NS is op individuele lijnen monopolist (net als Syntus en NoordNed monopolist zijn op hun eigen lijnen), dit limiteert de keuzevrijheid van de reiziger.
Stakeholder	ROVER (Vereniging Reizigers Openbaar Vervoer)
Classificatie	Dependent
Beschrijving	ROVER is een vereniging voor reizigers en heeft in die zin belang bij een hoge punctualiteit (urgentie). ROVER bezit echter geen formele macht binnen de NS, maar wordt wel gezien als een legitieme groep belanghebbenden bij het openbaar vervoer dat de NS aanbiedt. Deze legitimiteit uit zich onder andere in overleggen die ROVER heeft met de NS over verschillende onderwerpen. ROVER neemt ook deel in het LOCOV.
Stakeholder	LOCOV (Landelijk Overleg Consumentenbelangen Openbaar Vervoer)
Classification	Definitive
Beschrijving	De positie van LOCOV is ongeveer te vergelijken met die van ROVER, alleen heeft LOCOV wel formeel een adviespositie tegenover de NS, dit geeft LOCOV enige macht.
Stakeholder	Rijdend personeel
Classificatie	Dormant
Beschrijving	Het rijdende personeel heeft in principe weinig belang bij het op tijd rijden van de treinen, anders dan een verantwoordelijkheid tegenover de werkgever. Er bestaat geen financiële stimulans om treinen op tijd te laten rijden. Wel heeft het rijdende personeel de macht om treinen te laat te laten vertrekken. Enkele jaren geleden is dit ook gebeurd na een impopulaire beslissing van het management.
Stakeholder	(Bij)studerend management BRV
Classificatie	Definitive
Beschrijving	

Verschillende groepen mensen zijn binnen BRV bezig met het bijsturen van het proces op voorhand en in het geval van calamiteiten. Deze managers hebben in principe veel belang bij voorspellingen over de punctualiteit, omdat ze daarmee op voorhand al wijzigingen in de treinenloop kunnen doorvoeren om een hogere punctualiteit te realiseren. Naast dit belang hebben ze ook de macht om de treinenloop aan te passen waar dat nodig is.

Stakeholder ProRail

Classificatie Dominant

Beschrijving

ProRail regelt de dagelijkse treinenloop en kan de punctualiteit beïnvloeden door treinen voorrang te geven of te vertragen (macht). Als netwerkbeheerder is het ook een partij met een legitiem belang: een betere planning van de treinenloop door NS maakt het werk voor ProRail ‘makkelijker’. Urgentie ontbreekt, omdat ProRail in principe niet afgerekend wordt op lage punctualiteitscijfers.

De taken en verantwoordelijkheden van de stakeholders laat zien dat de totstandkoming van de punctualiteit een complex geheel is. Invloedrijke partijen (ProRail, goederenvervoerders) hebben veel minder belang bij een hoge punctualiteit en zijn dus ook minder geneigd om grote investeringen te doen die wellicht nodig zijn om de punctualiteit te verhogen. Een groep rijdend personeel die door correcte uitvoer een grote invloed heeft op de punctualiteit is benoemd als een dormant groep door gebrek aan belang bij een hoge punctualiteit. Een voorspellingsmodel kan helpen bij het toewijzen van punctualiteitsproblemen naar partijen die zonder ‘hard bewijs’ wellicht niet in actie zouden komen.

2 Probleemstelling en plan van aanpak

Met de probleemidentificatie uit hoofdstuk 1 als basis zal in dit hoofdstuk een probleemstelling en de deelvragen opgesteld worden. De tweede paragraaf zal per deelvraag een plan van aanpak geven.

2.1 Probleemstelling

Het doel van dit rapport is het opstellen van een model voor het voorspellen van de punctualiteit gegeven een aantal bekende of voorspelde parameters.

Om dit doel te bereiken, wordt de volgende onderzoeksvraag gebruikt:

Welke externe en interne factoren kunnen gebruikt worden voor het voorspellen van de punctualiteit van NS Reizigers en welke relatie bestaat tussen deze factoren en de punctualiteit?

Om deze onderzoeksvraag te kunnen beantwoorden is het allereerst van belang om het begrip van enkele onderdelen van de vraagstelling te vergroten. Om het begrip te vergroten, worden de volgende vragen beantwoord:

1. *Hoe wordt punctualiteit gemeten?*
 - a. *Welke partijen spelen een rol bij het meten van en het tot stand komen van de punctualiteit?*
 - b. *Wanneer is een trein 'op tijd'?*
 - c. *Welke wensen heeft de NS met betrekking tot punctualiteit?*
 - d. *Welke verschillen zijn te zien ten opzichte van het buitenland?*
2. *Hoe werkt voorspellen?*
 - a. *Wat is het doel van de punctualiteitvoorspelling?*
 - b. *Welke methodes bestaan voor het maken van voorspellingen?*
 - c. *Wat is de meest geschikte methode om de punctualiteit te voorspellen?*
 - d. *Hoe werkt deze methode?*

Het is bekend dat de NS sinds 10 december 2006 met een nieuwe dienstregeling werkt. Het wijzigen van de dienstregeling zou veel invloed kunnen hebben op de gegevens die bruikbaar zijn voor het maken van de voorspellingen. Het is daarom van belang om te onderzoeken of de data van voor 10 december 2006 gebruikt kan worden:

3. *Is de data van voor de invoering van de nieuwe dienstregeling bruikbaar voor voorspellingen over de nieuwe dienstregeling?*

Het volgende punt in de vraagstelling dat van belang is zijn de variabelen. Voor het voorspellen van de punctualiteit is het van belang om te weten welke variabelen invloed hebben op de punctualiteit. Met behulp van deelvraag vier zullen deze variabelen opgespoord worden.

4. *Welke externe en interne variabelen hebben invloed op de punctualiteit?*
 - a. *Welke externe variabelen hebben invloed op de punctualiteit en in welke mate?*
 - b. *Welke interne variabelen hebben invloed op de punctualiteit en in welke mate?*
 - c. *In hoeverre zijn de relaties van de bovenstaande variabelen en de punctualiteit statistisch aan te tonen?*

Om vervolgens een model te kunnen maken en de relatie tussen de variabelen *samen* en de punctualiteit aan te kunnen geven wordt vraagstuk vijf beantwoord:

5. *Hoe kunnen de verschillende variabelen worden gebruikt voor het maken van een voorspelling over de punctualiteit?*

Vervolgens zal vastgesteld worden aan welke doelstellingen het model voldoet en op welke manier, hiervoor zal de volgende deelvraag beantwoord worden:

6. *In hoeverre voldoet het geselecteerde voorspellingsmodel aan de gestelde doelen en hoe kan de NS de voorspellingen gebruiken?*
 - a. *Is het model accuraat genoeg om aan de doelstellingen (deelvraag 2a) te voldoen?*
 - b. *Hoe kan het model gebruikt worden door de NS om de punctualiteit te verhogen?*

Om de waarde van het vastgestelde model te benadrukken zal afsluitend de volgende vraag beantwoord worden:

7. *Hoe kunnen de gesignaleerde verstorende factoren aangepakt worden?*

2.2 Domein van het onderzoek

Het onderzoek zal zich in eerste instantie richten op het opstellen van een model voor een prognose van de punctualiteit voor de aankomende dagen op landelijk niveau voor NS Reizigers. Het model gaat dus niet expliciet in op de punctualiteit van andere partijen op het spoor (zoals Syntus, Veolia of goederentreinen). Het doel is om tot een eenvoudig te gebruiken applicatie te komen, waar belanghebbenden snel een prognose voor de punctualiteit kunnen maken aan de hand van enkele belangrijke factoren. Hierbij is belangrijk dat iemand die het model gebruikt in korte tijd een voorspelling kan maken voor een bepaalde tijdsperiode.

Aanvullend zou gekeken kunnen worden naar de verwachte punctualiteit van netwerkonderdelen (zoals Noordoost), productie-eenheden (Hengelo) en treinseries (1600, Enschede – Utrecht).

2.3 Plan van aanpak

In deze paragraaf zal per deelvraag aangegeven worden hoe deze vraag beantwoord zal gaan worden en welke informatie nodig is voor het beantwoorden van de vraag.

2.3.1 Hoe wordt punctualiteit gemeten?

De methodes waarop de punctualiteit gemeten wordt zijn vastgelegd in afspraken met de overheid en procedures die gevolgd worden door ProRail en NS Reizigers. Voor het verzamelen van informatie voor deze vraag kan de intranetsite van NS gebruikt worden, en daarnaast kan meer duidelijk worden door te praten met punctualiteit managers en operationeel managers.

De verschillen met het buitenland zullen door literatuurstudie en eventuele kennis van de eerder genoemde managers verzameld moeten worden. De beantwoording van deze vraag is voornamelijk van belang voor eventuele wijzigingen van het voorspellingsmodel om het geschikt te maken voor nieuwe/internationale richtlijnen in plaats van de Nederlandse richtlijnen.

2.3.2 Hoe werkt voorspellen?

Deze deelvraag is voornamelijk bedoeld om het begrip over voorspellingen te vergroten. Allereerst zal in dit hoofdstuk gekeken worden wat de waarde van een voorspelling is voor de NS. Daarna zal een overzicht gegeven worden van enkele voorspellingsmethodes waaruit vervolgens een keuze gemaakt wordt voor één of meerdere methodes die geschikt zouden kunnen zijn voor de beantwoording van de hoofdvraag. De theorie van de geselecteerde methodes zal vervolgens opgezocht worden in relevante literatuur.

2.3.3 Is de oude data bruikbaar voor de nieuwe voorspellingen?

Om te bekijken of oude data bruikbaar is, moet gekeken worden of de punctualiteit significant veranderd is ten opzichte van de oude dienstregeling. Om dit te vergelijken zal de beschikbare

punctualiteitdata vergeleken worden met data uit vergelijkbare periodes in de brondata. Om tot een (eensluidend) oordeel te komen zijn verschillende statistische toetsen beschikbaar welke gebruikt zullen worden om tot een oordeel te komen.

2.3.4 Welke externe en interne variabelen hebben invloed op de punctualiteit?

Deze deelvraag is opgedeeld in de externe en interne variabelen. Onder de interne variabelen worden de variabelen verstaan die beïnvloed kunnen worden door de NS. Hierbij kan dus gedacht worden aan hoeveelheid materieel, personeel, soort materieel, etc. De externe variabelen zijn de invloeden waar de NS geen of weinig invloed op heeft. Variabelen die hier dus onder vallen zijn het weer, ongelukken (door verschillende omstandigheden), gebreken in het netwerk, etc.

Niet alle variabelen die van invloed zijn zullen gemakkelijk gevonden kunnen worden. Doel van deze deelvraag is om tot een set variabelen te komen die significant invloed hebben op de punctualiteit.

Tweede voorwaarde voor deze variabelen is dat ze op het moment dat de voorspelling gedaan moet worden met voldoende zekerheid bekend zijn.

Het vinden van de variabelen die invloed hebben op de punctualiteit kan aan de hand van historische data van de NS. De verstoringen van de treinenloop worden goed bijgehouden. Dagelijks wordt een mailing (zie Voorbeeld 2.1) verzonden met daarin de belangrijkste verstoringen. Daarnaast wordt een uitgebreid systeem bijgehouden met daarin de aard van een verstoring en de afhandeling. Analyse van data uit deze mailings en eventuele aanvullende informatie over verstoringen kan een uitgebreide set variabelen opleveren. Om de variabelen onder gelijke noemers te kunnen behandelen zal een categorisering van de verstoringen gemaakt worden. Aan de hand van deze categorisering kan een logica gezocht worden.

Het vinden van data voor deze variabelen is, voor zover het gaat om variabelen die direct verbonden zijn aan de treinenloop, goed mogelijk. De NS en ProRail houden uitgebreide databanken bij over de treinenloop, en de NS doet aanvullend ook nog regelmatig onderzoeken naar bijvoorbeeld het reizigersaantal. Externe variabelen zijn wellicht lastiger te vinden. Deze data zal voornamelijk uit het publieke domein moeten komen. Een variabele waarvan verwacht wordt dat deze een relatief grote invloed zal hebben op de punctualiteit is het weer, voor deze variabele is in databanken van onder andere het KNMI voldoende informatie beschikbaar.

Toelichting dagpunctualiteit 6 februari 2007, behaalde punctualiteit: 86,3%

- Een stroomstoring te Kijfhoek (05.11 uur - 07.51 uur)
- Een sectie- en overwegstoring op het baanvak Breda - Roosendaal (02.06 uur - 07.30 uur)
- Een defecte trein op het baanvak Oss - Ravenstein (06.02 uur - 06.31 uur)
- Capaciteitsproblemen te Kijfhoek (06.54 uur - 09.22 uur)
- Een defecte reizigerstrein op het baanvak Den Bosch - Nijmegen (11.05 uur - 11.28 uur)
- Een defecte reizigerstrein op het baanvak Amersfoort - Barneveld (11.53 uur - 12.25 uur)
- Een defecte brug op het baanvak Alkmaar - Heerhugowaard (12.19 uur - 12.28 uur)
- Een wisselstoring op het baanvak Alphen aan den Rijn - Leiden (14.27 uur - 17.06 uur)
- Een rijrichtingstoring op het baanvak Rijssen - Snippeling aansluiting (17.31 uur - 18.17 uur)
- Een aanrijding met een dier op het baanvak Ede Wageningen - Wolfheze (19.41 uur - 21.05 uur)
- Een verdacht persoon op het baanvak Alkmaar - Uitgeest (21.33 uur - 22.05 uur)
- Een defecte reizigerstrein op het baanvak Baarn - Hilversum (21.49 uur - 22.23 uur)
- Een wisselstoring op het baanvak Breda - Lage Zwaluwe (22.25 uur - 22.58 uur)
- Een defecte reizigerstrein op het baanvak Hilversum - Naarden Bussum (22.30 uur - 22.55 uur)

Voorbeeld 2.1 Een toelichting dagpunctualiteit met de meest invloedrijke storingen

Wanneer de variabelen gevonden zijn, is het van belang om de relaties tussen de variabelen en de punctualiteit te bepalen met behulp van regressieanalyse of met één van de andere methodes die in hoofdstuk 4.3 beschreven worden. De gronden waarop een variabele geaccepteerd wordt als invloedhebbend op de punctualiteit moeten later vastgesteld worden.

2.3.5 Hoe kunnen de verschillende variabelen worden gebruikt voor het maken van een voorspelling over de punctualiteit?

De beantwoording van deze vraag moet leiden tot een model dat gebruikt kan worden voor het voorspellen van de functionaliteit. De in de tweede vraag geselecteerde methodes zullen geïmplementeerd worden voor de specifieke situatie, en vervolgens kan aan de hand van de resultaten een keuze gemaakt worden voor het best passende model, of er wordt een combinatie gemaakt tussen de voorspellingen van de aparte modellen.

2.3.6 In hoeverre voldoet het model aan de doelstellingen en hoe kan de NS het model gebruiken?

Deze deelvraag is gedeeltelijk gericht op een stuk beoordeling van de kwaliteit van het model: per doelstelling zal aangegeven worden in hoeverre het model accuraat genoeg is om de voorspellingen te gebruiken voor deze doelstelling. Daarnaast zal het gebruik van het model nader toegelicht worden. De gemiddelde gebruiker zal niet beschikken over de juiste kennis of software om iedere methode te kunnen gebruiken. Daarom zal aangegeven worden hoe de modellen geschikt gemaakt worden voor gebruik door gebruikers met een beperkte kennis over voorspellen/statistiek/excel/etcetera.

2.3.7 Hoe kunnen de gesignaleerde verstorende factoren aangepakt worden?

Voor de beantwoording van deze vraag zullen geen ‘out of the box’ oplossingen bedacht worden voor de verstorende factoren. Doel van deze vraag is om de personen die gaan werken met de voorspellingen handvatten te geven waarmee ze actie kunnen ondernemen wanneer de voorspelde punctualiteit niet aan de wensen voldoet.

2.3.8 Leeswijzer

De deelvragen kunnen per hoofdstuk teruggevonden worden in dit verslag. In het onderstaande schema kunnen de hoofdstuknummers gevonden worden waar de beantwoording van de deelvraag begint.

De belangrijkste hoofdstukken voor de vaststelling van het model en de mogelijkheden van het model zijn de hoofdstukken 7 en 8. Hoofdstuk 4 geeft een uitgebreide beschrijving van de gebruikte methodes voor de voorspellingsmodellen.

Deelvraag	Hoofdstuk
Hoe wordt punctualiteit gemeten?	3
Hoe werkt voorspellen?	4
Is de data van voor de invoering van de nieuwe dienstregeling bruikbaar voor voorspellingen over de nieuwe dienstregeling?	5
Welke externe en interne variabelen hebben invloed op de punctualiteit?	6
Hoe kunnen de verschillende variabelen worden gebruikt voor het maken van een voorspelling over de punctualiteit?	7
In hoeverre voldoet het geselecteerde voorspellingsmodel aan de gestelde doelen en hoe kan de NS de voorspellingen gebruiken?	8
Hoe kunnen de gesignaleerde verstorende factoren aangepakt worden?	9
Conclusies en aanbevelingen	10

3 Punctualiteit

In hoofdstuk 1 werd al een definitie gegeven van punctualiteit: *punctualiteit is gedefinieerd als het percentage treinen dat op een aantal afgesproken stations voor, op of binnen een marge van drie minuten na de afgesproken aankomsttijd aankomt*. In dit hoofdstuk zal de punctualiteit nader toegelicht worden en daarmee zal de eerste deelvraag beantwoord worden. Allereerst zal gekeken worden naar de soorten punctualiteit en vervolgens de partijen die (bewust of onbewust) invloed hebben op de punctualiteit. Vervolgens zal gekeken worden naar de verschillende soorten punctualiteit en de methode die wordt gebruikt voor het meten van de punctualiteit. Ook zal gekeken worden naar enkele onderzoeken die in het verleden gedaan zijn naar de oorzaken van het achterblijven van de punctualiteit.

3.1 Soorten punctualiteit

Punctualiteit valt uiteen in twee soorten: aankomstpunctualiteit en vertrekpunctualiteit. In de meeste gevallen wordt met punctualiteit de aankomst variant bedoeld, in dit rapport zal deze terminologie dan ook gebruikt worden. Wanneer vertrekpunctualiteit bedoeld wordt, zal dit ook expliciet vertrekpunctualiteit genoemd worden.

Het verschil tussen beide vormen is wat de naam doet vermoeden: vertrekpunctualiteit gaat over vertrekkende treinen, aankomstpunctualiteit over aankomende treinen. In het contract met de overheid zijn afspraken gemaakt over aankomstpunctualiteit. Vertrekpunctualiteit is echter ook van groot belang, een trein die te laat vertrekt zal immers ook hoogstwaarschijnlijk een vertraging hebben bij aankomst. ProRail en NS meten dus ook beide soorten punctualiteit.

3.2 Deelnemende partijen

Zoals al uit de stakeholderanalyse bleek, is niet alleen de NS verantwoordelijk voor het vervoer van reizigers over het Nederlandse spoorwegnetwerk. Los van partijen als Syntus en Veolia die op aparte lijnen werken, moet de NS haar lijnen delen met goederenvervoerders. Daarnaast wordt de verkeersregeling niet door de NS alleen gedaan, maar juist door de netwerkbeheerder ProRail. Naast goederenvervoerders en ProRail speelt ook het gedrag van reizigers een rol in de totstandkoming van de punctualiteit.

3.2.1 NS

Zoals duidelijk mag zijn speelt NS zelf een belangrijke rol in het behalen geplande rittijden. De NS maakt zelf een dienstregeling die volgens de NS haalbaar is. ProRail controleert deze dienstregeling vervolgens en past de dienstregeling aan waar dat nodig is. Met deze dienstregeling kan de NS dus gedeeltelijk invloed uitoefenen op de prestaties die zijzelf moet leveren. Wanneer bijvoorbeeld gepland wordt dat iedere trein 10 minuten wacht op ieder station, wordt het behalen van de punctualiteitsafspraken een stuk eenvoudiger dan wanneer de marge slechts één minuut is. Daarnaast kan de NS met minder treinen het netwerk minder belasten waardoor vertragingen minder doorrekenen (minder treinen hoeven op elkaar te wachten). Echter, de dienstregeling waarmee de NS (theoretisch) een punctualiteit van 100% zou kunnen behalen is waarschijnlijk zeer ongunstig voor reizigers, er moet dus een balans gevonden worden tussen een goede dienstregeling voor de klant, en een haalbare dienstregeling uit het oogpunt van punctualiteit.

Bij het realiseren van de gewenste punctualiteit speelt een aantal aspecten een rol:

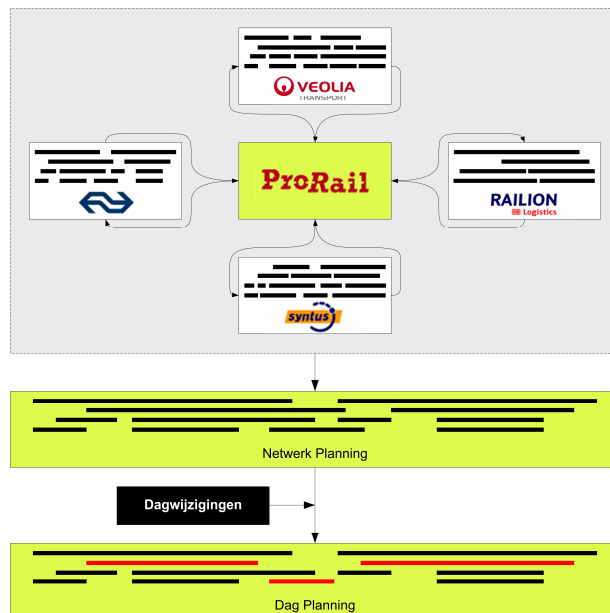
1. Eerste klap treinen

Eerste klap treinen zijn treinen die vertrekken met 'nieuw' personeel voor 8:00 uur. Wanneer het personeel te laat start met de trein, loopt de trein vanaf het begin van de dag vertraging op. De opgelopen vertraging moet vervolgens over de dag goed gereden worden, wat op druk bezette trajecten (bijvoorbeeld enkelspoor trajecten) zeer moeilijk is. Er wordt veel aandacht besteed aan de vertraging die eerste klap treinen oplopen, de vertrekpunctualiteit van deze treinen wordt met een marge van één minuut beoordeeld.

2. Op tijd vertrekken
Met de invoering van de nieuwe dienstregeling is de procedure voor het vertrekken van treinen ook aangescherpt. De conducteur bepaalt op welk moment de trein weggaat. Formeel moet de conducteur 20 seconden voor vertrek fluiten, waarna hij 10 seconden voor vertrek de deuren moet sluiten. De trein kan dan precies op het verspringen van de minuutwijzer weggrijpen. Wachten op passagiers die te laat zijn zorgen ervoor dat een trein te laat vertrekt, waardoor de kans op vertraging van de desbetreffende trein, of volgende treinen groter wordt. Op tijd vertrekken is van groot belang omdat de dienstregeling op sommige punten zo strak in elkaar zit dat één minuut vertraging al tot andere vertragingen kan leiden op het netwerk.
3. Volgen van procedures
Het besturen van treinen is aan strikte regels gebonden. Vanaf het moment dat een trein een baanvak inrijdt (een spoor tussen twee seinen) wordt dit baanvak voor deze trein gereserveerd. Verkeerd gebruik van het spoor kan voor vertragingen zorgen, bijvoorbeeld wanneer een trein niet ver genoeg doorrijdt bij een station en op die manier een extra spoor bezet houdt of wanneer een trein een stoptonend sein passeert (hier volgt bijna altijd een onderzoek op, resulterend in gestremd treinverkeer).
4. Veiligheid op het spoor
De veiligheid van reizigers is natuurlijk van groot belang, maar niet alleen voor het welzijn van de reizigers en het personeel. Ook voor de punctualiteit is het van belang om het aantal incidenten zo laag mogelijk te houden. Aanrijdingen en ongelukken rondom treinen (bijvoorbeeld bij in- & uitstappen) zorgen voor vertraging. Het voorkomen van deze incidenten is een gedeelde verantwoordelijkheid van ProRail en de NS. ProRail is in principe verantwoordelijk voor de beveiliging van het spoor, NS voor het veilig vervoeren van de reizigers.

3.2.2 ProRail

De planning van de treinenloop op het netwerk is een complex proces. In het onderstaande figuur is weergegeven hoe een dagplanning tot stand komt. De vervoerders op het netwerk zoals de NS, Veolia, Syntus, Rialion, etc. dienen een planning op die voor hun lijnen realiseerbaar is. ProRail kan vervolgens deze planning wijzigen zodat het past binnen de planning van andere partijen. Uiteindelijk komt hier een planning per dag uit die in principe geen conflicten heeft. Per dag kunnen nog wijzigingen opgenomen worden die rekening houden met eventueel onderhoud of andere bijzondere omstandigheden.



Figuur 3.1 Totstandkoming van de planning

ProRail regelt daarnaast ook de treinenloop van minuut tot minuut. Op het moment dat twee treinen op hetzelfde moment aanspraak willen maken op een baanvak, trekt de trein met vertraging altijd aan het kortste eind. De NS treinen (die strikt gebonden zijn aan vertrek- en aankomsttijden) krijgen dus niet altijd voorrang boven goederentreinen (die gebonden zijn aan de rijwegenplanning) waardoor vertraagde intercity's achter langzame goederentreinen komen en hierdoor nog meer vertraging oplopen.

Naast de planning van treinen is ProRail ook verantwoordelijk voor het onderhoud en controle van het spoor. Controlewerkzaamheden (schouwingen) kunnen niet altijd buiten de dienstregeling van de NS om gedaan worden, waardoor de treinen van de NS vertraging oplopen als een baanvak gecontroleerd wordt.

Het onderhoud van het Nederlandse spoorwegnetwerk moet ook op regelmatige basis gebeuren. In de afgelopen jaren is (verkeerd) bezuinigd op het onderhoud van het spoor, hierdoor is er op dit moment relatief veel onderhoud nodig aan het spoor. Ook zorgt deze achterstand in onderhoud voor veel infrastructuurstoringen zoals sein- en wisselstoringen. Aangezien onderhoud over het algemeen gepland is, heeft dit geen invloed op de punctualiteit, infrastructuurstoringen hebben echter wel invloed op de punctualiteit.

3.2.3 Reizigers

Ook reizigers spelen zelf een rol in het voorkomen van vertraging. Reizigers kunnen vertraging voorkomen door niet meer in te stappen na het fluitsignaal en deuren niet tegen te houden wanneer deze sluiten zodat de trein op tijd kan vertrekken. Ook spelen reizigersaantallen een rol: meer reizigers doen er immers langer over om in en uit te stappen. Een meer gespreide spits zou, hoewel moeilijk te realiseren, zorgen voor een betere punctualiteit.

Stoptrein Arnhem – Arnhem Presikhaaf (7500)

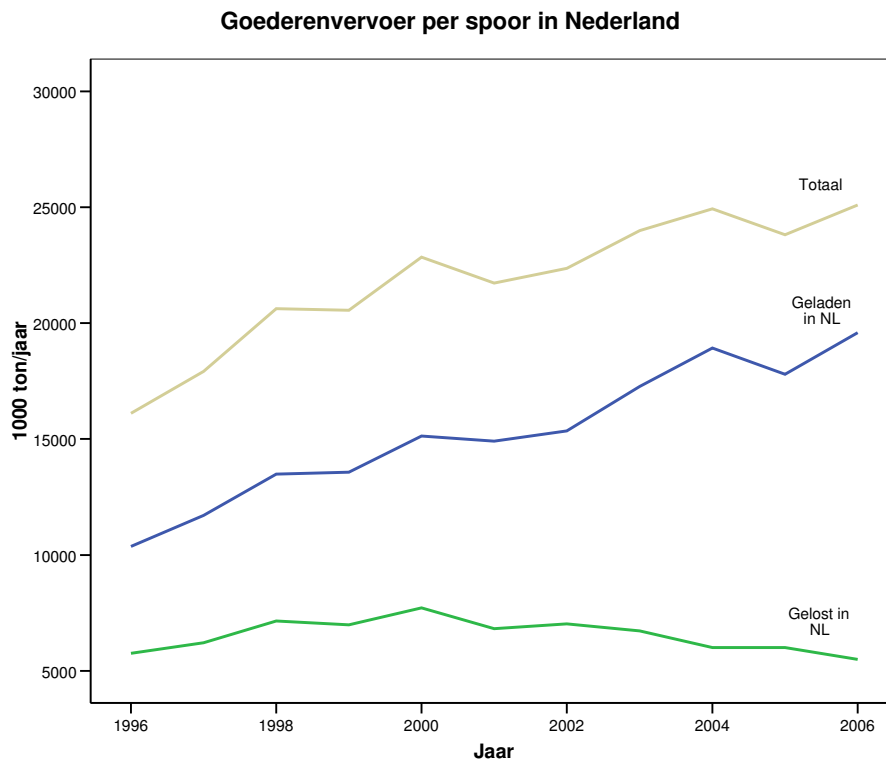
De stoptrein tussen Arnhem en Arnhem Presikhaaf had regelmatig vertraging. De oorzaak van de vertraging bleek een grote groep studenten te zijn die op Arnhem Presikhaaf achter in de trein stapten zodat ze op station Presikhaaf snel bij de trap waren (die aan het uiteinde van het station gelegen was). Het instappen van de studenten nam meer tijd dan nodig in beslag: alles moest door één deur, terwijl er ruimte genoeg was in de rest van de trein. Aanleg van een extra trap en pad bij Presikhaaf heeft deze problemen grotendeels verholpen: de 7500 rijdt nu beter op tijd.

Voorbeeld 3.1 Beter instappen leidt tot minder vertraging

3.2.4 Goederenvervoerders

Naast de NS maken ook goederenvervoerders gebruik van het netwerk en ook zij hebben te maken met punctualiteit ten opzichte van de afgesproken dienstregeling met ProRail. Een goederenvervoerder krijgt, net als de NS, een bepaalde periode een stuk spoor beschikbaar om daar met een goederentrein te rijden. Wanneer de trein uit zijn vooraf vastgelegde pad loopt, heeft deze trein vertraging (hoewel dat voor de ontvangende partij niet zo hoeft te zijn) en wordt deze trein ook als vertraagde trein behandeld door ProRail.

De afgelopen tien jaar is de hoeveelheid goederenvervoer per spoor sterk gegroeid, daarnaast worden sinds enkele jaren de goederentreinen gelijk behandeld als personentreinen. Hierdoor neemt de invloed van goederentreinen op de punctualiteit van de NS toe.



Figuur 3.2 Goederenvervoer per spoor in Nederland (CBS, 2007)

3.3 *Het meten van punctualiteit*

3.3.1 Treinidentificatie en routing

Iedere trein die rijdt heeft bij vertrek een uniek nummer gekregen, dit nummer wordt samengesteld uit de treinserie en moment waarop de trein vertrekt. Een cijfer eindigend op een even getal is over het algemeen een trein die van oost naar west rijdt, een oneven getal is dus van west naar oost. Verder duiden hoge cijfers op stoptreinen en lage cijfers op intercity's.

1751: Intercity Den Haag naar Enschede (17, 1 = oneven) en vertrekt midden op de dag (51)
1752: Intercity Enschede naar Den Haag (17, 2 = even) en vertrekt midden op de dag (52)
7911: Stoptrein Zwolle naar Enschede (79, 1 = oneven) en vertrek 's ochtends vroeg (11)
7984: Stoptrein Enschede naar Zwolle (79, 4 = even) en vertrek 's avonds (84)

Voorbeeld 3.2 Treinnummers

De planning van de treinen ligt vast over de dag, wanneer een trein wil vertrekken vanaf een station wordt automatisch de rijweg ingesteld en gereserveerd voor deze trein. Het toewijzen van de rijweg gebeurt door een computersysteem (Automatische Rijweg Instelling, ARI) onder verantwoordelijkheid van de treindienstleider. Wanneer een trein vertraging oploopt, valt de trein buiten zijn rijweg en zal deze trein een nieuwe rijweg moeten krijgen. Bij het toewijzen van deze nieuwe rijweg krijgen treinen zonder vertraging die aanspraak maken op dezelfde rijweg voorrang, de vertraagde trein krijgt hierdoor dus nog meer vertraging, maar de niet vertraagde trein loopt geen vertraging op.

3.3.2 Punctualiteitsmetingen

Een treinserie die vertrekt wordt op verschillende punten gemeten. Wanneer het een treinserie betreft die vertrekt voor 8:00 uur met 'nieuw' personeel wordt deze trein aangemerkt als 'eerste klap' en wordt de vertrekpunctualiteit gemeten door de procesleider perron (PLP). Over de rest van de dag

wordt de vertrekpunctualiteit gemeten bij seinen. Aangezien seinen vaak niet gelijk staan met het perroneinde of -begin, wordt een correctie van enkele seconden toegepast op de tijd dat een trein een sein passeert. Deze correctie is afhankelijk van de situatie. Het cijfer dat telt voor de punctualiteitsdoelstelling is de aankomstpunctualiteit. Deze wordt op een vergelijkbare manier gemeten als de vertrekpunctualiteit. Een trein is op tijd wanneer deze maximaal 2 minuten en 59 seconden later aankomt dan de tijd waarop de trein moest aankomen. Internationaal gezien ligt deze grens op 5 minuten. De NS meet de punctualiteit dan ook zowel op de Nederlandse 3 minuten norm als op de internationale 5 minuten norm. Gemeten op deze 5 minuten norm doet de NS het in Europa zeer goed, alleen Zwitserland presteert beter.

Land	Punctualiteit (5 minuten norm)
Zwitserland (SBB)	96,0%
Nederland	93,4%
België (NMBS)	91,9%
Frankrijk (SNCF)	82,6%*
Duitsland (DB)	76,6%*

* Op de hoofdlijnen

Tabel 3.1 Punctualiteit top vijf Europa, treinen binnen 5 minuten t.o.v. geplande aankomsttijd (NMBS, 2006)

Voor de punctualiteit tellen alleen treinen mee die gepland zijn. Een trein die uit de dienstregeling gehaald wordt telt niet mee. Wanneer, bijvoorbeeld door een calamiteit, de intercity Amsterdam Centraal – Enkhuizen (3200) niet meer kan rijden, tellen deze treinen niet meer mee voor de punctualiteitscijfers in Amsterdam. Rijdt de intercity Amsterdam – Enkhuizen door calamiteiten met een vertraging van 10 minuten, dan telt dit wel mee in de punctualiteit van Amsterdam.

Niet alle metingen tellen mee voor de punctualiteit. NS heeft met de overheid afspraken gemaakt over de stations waar aankomst punctualiteit gemeten wordt. In Tabel 3.2 is een lijst opgenomen met de stations waar de punctualiteit cijfers van meetellen voor de gerealiseerde punctualiteit.

Het systeem voor de punctualiteitsmeldingen is opgezet en wordt uitgevoerd door ProRail. NS heeft hier zelf dus beperkt invloed op.

Knoop	Pun.	Knoop	Pun.	Knoop	Pun.
1. Alkmaar	76%	13. Eindhoven	80%	25. Roermond	87%
2. Almelo	84%	14. Gouda	83%	26. Roosendaal	79%
3. Amersfoort	78%	15. Groningen	87%	27. Rotterdam Centraal	82%
4. Amsterdam Centraal	76%	16. Haarlem	82%	28. Schiphol	84%
5. Apeldoorn	84%	17. Heerlen	92%	29. Sittard	87%
6. Arnhem	74%	18. Hengelo	89%	30. Tilburg	85%
7. Breda	84%	19. 's-Hertogenbosch	76%	31. Utrecht Centraal	74%
8. Den Haag Centraal	80%	20. Hoorn	80%	32. Venlo	75%
9. Den Haag HS	82%	21. Leeuwarden	86%	33. Vlissingen	80%
10. Deventer	80%	22. Leiden	88%	34. Zutphen	80%
11. Dordrecht	80%	23. Maastricht	88%	35. Zwolle	82%
12. Duivendrecht	77%	24. Nijmegen	72%		

Tabel 3.2 Stations die meetellen voor de punctualiteit, in *cursief* zijn de stations aangegeven behorende tot netwerk Noordoost (MinVW, 2005)

De uiteindelijke punctualiteit gemeten over een aantal meetlocaties is het percentage van de treinen die op tijd zijn aangekomen op meetpunten. Reizigersaantallen spelen dus geen enkele rol in het meten van de punctualiteit. De invloed van het aantal treinen is wel groot. Een station met veel aankomsten (Amsterdam, 40/uur) telt zwaarder dan een station met weinig aankomsten (Groningen, 3/uur). Verder tellen intercity treinen vaak zwaarder dan stoptreinen. De 1700 (intercity Enschede – Den Haag) wordt

bijvoorbeeld onderweg 8 keer geteld (Hengelo, Almelo, Deventer, Apeldoorn, Amersfoort, Utrecht, Gouda en Den Haag), terwijl de 7000 (stoptrein Deventer – Almelo) maar één keer geteld wordt (Almelo).

3.3.3 Uitgevallen treinen en (grote) calamiteiten

Een aspect dat van relatief groot belang is voor de punctualiteit is het aantal uitgevallen treinen. Een uitgevallen trein wordt niet meer meegeteld in de punctualiteitscijfers. Wanneer de NS dus zou kiezen om op een dag één trein te laten rijden tussen Utrecht en Utrecht Overvecht en deze komt binnen de marge aan, dan heeft de NS een punctualiteit van 100% behaald. Om dit soort situaties te voorkomen zijn met de overheid ook afspraken gemaakt over de tevredenheid van klanten. De NS moet dus continu een afweging maken tussen (een hoge) punctualiteit en (een lagere) klanttevredenheid. Een ander aspect dat lijkt op vertraging maar niet als dusdanig meegenomen wordt is de situatie waarbij de dienstregeling op een traject gestremd is. Wanneer bijvoorbeeld tussen Hengelo en Almelo geen treinverkeer mogelijk is, worden er een pendeltrein op het traject Hengelo – Enschede ingezet en de intercity Schiphol – Enschede rijdt tot Almelo. De trein Hengelo – Enschede krijgt dan nieuwe nummers (en niet de ‘intercitynummers’ die deze normaal zou hebben) en rijdt volgens een eigen dienstregeling. Ook de intercity Schiphol – Almelo zal een eigen (nieuwe) dienstregeling krijgen. Deze treinen rijden dan dus niet met vertraging, ondanks dat de tijden wellicht anders zijn.

3.3.4 Reizigerspunctualiteit

Drie minuten vertraging is lang niet altijd even vervelend. Een vertraging van één minuut waardoor je een overstap niet haalt is veel vervelender dan een vertraging van vijf minuten op je eindbestemming. De punctualiteit die hiermee samenhangt heet reizigerspunctualiteit en is moeilijk continu te meten, maar geeft wel een veel beter beeld van de daadwerkelijke vertraging van reizigers. ProRail doet jaarlijks wel onderzoek naar de reizigerspunctualiteit, en over het algemeen ligt de reizigerspunctualiteit lager dan de (treinen)punctualiteit (ProRail, 2006).

3.4 Reeds uitgevoerde onderzoeken

In het verleden zijn verschillende onderzoeken gedaan naar verschillende componenten die met punctualiteit te maken hebben. Een drietal onderzoeken (Reijnen, 2005; Schriever, 2006; BCG, 2005) is nader bekeken, de belangrijkste vindingen uit deze onderzoeken zullen hieronder weergegeven worden.

3.4.1 Opbouw van vertraging

Zoals al eerder vermeld is onder andere de vertrekprocedure van een trein van belang voor de punctualiteit die een trein behaalt. Uit onderzoek (Reijnen, 2005) blijkt dat van de treinen die binnen een marge van één minuut vertrekken op het eerste vertrekstation 90,2 procent op tijd rijdt. Wanneer een trein eenmaal vertraging heeft, is de kans op extra vertraging groter dan wanneer een trein geen vertraging heeft. Wanneer bekeken wordt waarbij een (niet vertraagde) trein vertraging oploopt, levert dat het volgende op:

Handeling	Bron van vertraging
Vertrek van eerste vertrekstation	54%
Rijtijden	27%
Haltering ³	19%

Tabel 3.3 Handeling waarbij vertraging optreedt

Het “goedrijden⁴” van vertraging op een traject is lang niet altijd mogelijk: 69,8 procent van de treinen haalt tijd in op het schema. Ook bij de haltingen wordt geen tijd goedgeemaakt. Bij lange

³ Het proces van stoppen, wachten en wegrijden op een station.

vertragingen gaat er bij een haltering ongeveer 10 seconden van de vertraging af, bij weinig vertraging komt er ongeveer 10 seconden bij.

Percentage sneller gereden dan schematijd	
< -180 seconden	4,8%
< -120 seconden	16,3%
< -60 seconden	41,7%
< 0 seconden	69,8%
< 60 seconden	85,3%
< 120 seconden	91,9%
< 180 seconden	95,1%
< 240 seconden	96,8%

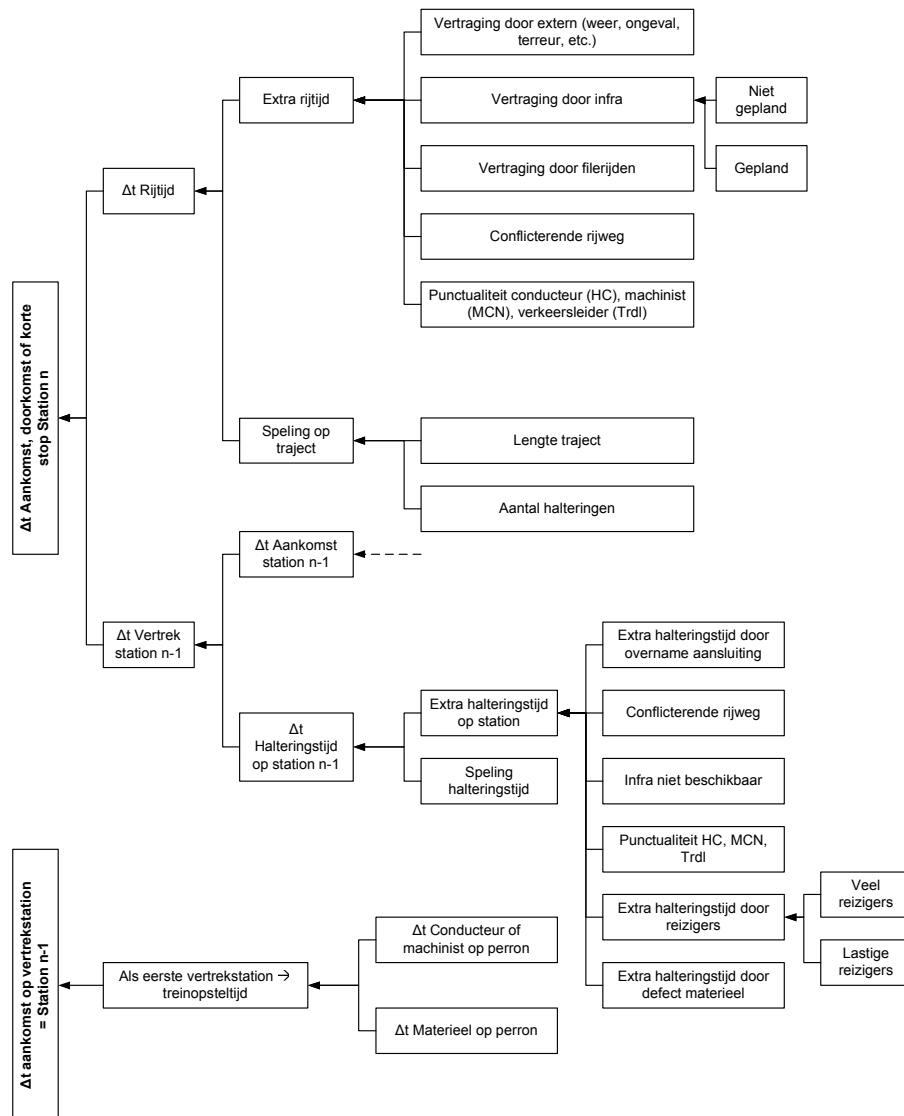
Tabel 3.4 Percentage sneller gereden dan schematijd

In Figuur 3.3 is een oorzaak en gevolg diagram opgenomen uit het onderzoek van Reijnen (2005). Hoewel deze figuur voor commentaar vatbaar is, geeft hij een aardig inzicht in de mogelijke oorzaken van vertraging. Het belangrijkste commentaar is dat de oorzaken van vertraging meer met elkaar samenhangen dan hier aangegeven is. Filerijden kan bijvoorbeeld goed veroorzaakt worden door een conflicterende rijweg of 'infrastructuur niet beschikbaar', daarnaast worden infrastructuurstoringen gedeeltelijk veroorzaakt door weersinvloeden. Het voorspellingsmodel dat onderzocht zal worden in dit rapport zal proberen enkele van de relaties in dit diagram te kwantificeren.

Het startproces van een trein valt volgens Schriever (2006) uiteen in de componenten gereedmaakproces, seinproces en vertrekproces. Onder vertragingen van het gereedmaakproces vallen materieel dat te laat aankomt, personeel dat te laat aankomt en defecten aan het materieel. Het seinproces omvat het ontstaan van een conflicterende rijweg en overige kleine infrastoringen. Het vertrekproces omvat spoorwijzigingen en reizigers (veel, te laat, lastig, etc.). Tijdens alle deelprocessen kan vertraging ontstaan die vervolgens doorrekent tijdens de dienst van de trein. Uit een analyse in 2006 (Schriever, 2006) blijkt onder andere dat:

- 12,5 procent van de onderzochte treinen is niet op tijd gereed
 - Defecten, personeel te laat
 - Problemen met een korte keer (trein die aankomt moet binnen korte tijd weer vertrekken in tegenovergestelde richting)
 - Aankomstvertraging
 - Treingereed maken duurt te lang
- 19 procent van de treinen heeft te laat veilig sein
 - Conflicterende rijwegen
 - Treinnummer te laat toegekend (een trein zonder treinnummer mag niet rijden)
- 14 procent van de treinen heeft reizigershinder

⁴ Machinisten kunnen voor beperkte tijd 10 procent harder rijden dan gepland (maar nooit harder dan de maximale snelheid op het traject) om een eventuele vertraging goed te rijden. Daarnaast is de geplande rijtijd op sommige trajecten langer dan de werkelijke rijtijd.



Figuur 3.3 Oorzaak en gevolg diagram (Reijnen, 2005).

3.5 Conclusie

Dit hoofdstuk heeft als doel om het begrip punctualiteit uit te leggen. De punctualiteit wordt niet alleen door de NS gerealiseerd, maar door een aantal verschillende partijen. Deze partijen (NS, ProRail, externe vervoersbedrijven en reizigers) hebben allen op verschillende aspecten van de punctualiteit invloed.

Punctualiteit valt uiteen in aankomstpunctualiteit en vertrekpunctualiteit. De Nederlandse overheid kijkt naar de aankomstpunctualiteit (naast andere kengetallen) voor het beoordelen van de prestaties van de NS. Een punctualiteit van 80% betekent grofweg dat 80% van de treinen zonder een vertraging of met een vertraging tot 3 minuten gereden heeft.

Eerder uitgevoerd onderzoek heeft aangetoond dat er vertragingen ontstaan door onder andere weer, dit onderzoek zal in de volgende onderzoeken proberen deze relaties te kwantificeren.

4 Voorspellen

Voor het voorspellen van de toekomst zijn vele methodes ontwikkeld variërend van persoonlijke intuïtie tot complexe wiskundige modellen. Exact de toekomst voorspellen blijft – ondanks alle moeite die in onderzoek naar effectief voorspellen gestoken is – zo goed als onmogelijk. Desondanks zijn prognoses van groot belang voor organisaties.

Hoe voorspellen werkt zal in dit hoofdstuk nader toegelicht worden, met doel deelvraag twee te beantwoorden. Allereerst zal gekeken worden wat het doel van de voorspellingen is, vervolgens zal gekeken worden naar de verschillende technieken voor het maken van voorspellingen. Aansluitend zullen de methodes die voor het punctualiteitsmodel geschikt lijken worden uitgelegd. Afsluitend zal de waarde van voorspellingen nader toegelicht worden.

Variabelen

Twee soorten variabelen spelen bij voorspellen een belangrijke rol:

1. **Onafhankelijke variabelen:** Dit zijn de variabelen die door de omgeving of de organisatie opgelegd worden. Voorbeelden van onafhankelijk variabelen (o.v.): Temperatuur, reizigersaantallen, ziekteverzuim, etc.
2. **Afhankelijke variabe(n):** Dit is de te voorspellen variabele. De afhankelijke variabele (a.v.) is in het geval van dit rapport dus de punctualiteit. De o.v. worden dus gebruikt om de a.v. te voorspellen.

4.1 Het doel van de voorspelling

Prognoses zijn onmisbaar voor het maken van een (productie)planning. Wanneer een planning drie weken van tevoren gemaakt moet worden, is het van groot belang dat *nu* al bekend is wat de vraag naar het te produceren product bij benadering zal zijn over drie weken. Is onbekend wat de vraag over drie weken zal zijn, dan is het maken van een planning haast onmogelijk.

Voor de NS geldt een iets andere situatie. De NS heeft een vaste dienstregeling, dus het is bekend hoeveel personeel op ieder moment nodig is. Echter, een aantal (on)bekende factoren hebben invloed op de dienstregeling. Om met deze (on)bekende factoren om te kunnen gaan, is het van belang dat de NS accurate voorspellingen kan maken. Zoals eerder aangegeven is het door afspraken met de overheid van belang dat de NS haar punctualiteit behaalt, en het behalen van de gewenste punctualiteit is (naast het behalen van indicatoren als zitplaatscapaciteit) voor de NS dus te vergelijken met het voorkomen van out-of-stock situaties bij productiebedrijven. De NS wordt – net als productiebedrijven – dus afgerekend op haar prestaties, gemeten in punctualiteit.

4.1.1 Een voorspelling, en wat dan?

Een voorspelling is niet alleen een ‘leuke indicatie van wat er gaat gebeuren’. Wanneer blijkt dat de punctualiteit te laag uitkomt in een periode (dag, week, maand), moet het proces op een dusdanige manier bijgestuurd worden dat de (verwachte) punctualiteit stijgt. De manier waarop de punctualiteit gemeten wordt beperkt de mogelijkheden voor ingrijpen. Op 18 januari (er raasde een storm over Nederland die voor veel overlast zorgde) reed de NS met een normaal dienstrooster en behaalde een punctualiteit van 67,7 procent. Op 8 februari (weeralarm, er werd veel sneeuw verwacht, en er viel ook sneeuw) reed de NS na de spits met een aangepaste dienstregeling (ongeveer 20 procent minder treinen ten opzicht van een normale donderdag), de punctualiteit die gerealiseerd werd was 69,4 procent. Op de stormdag viel al het treinverkeer aan het einde van de dag stil, wat resulteerde in 49 procent uitgevallen treinen. Op de sneeuwdag was het aantal treinen dat uitviel 21 procent, maar dat was gepland. Het is moeilijk om vast te stellen hoe beide dagen verlopen waren wanneer wel (in het geval van 18 januari) of geen (in het geval van 8 februari) maatregelen genomen zouden zijn, maar aan de hand van de cijfers kan wel gesteld worden dat de NS op 8 februari een betere dienstverlening heeft geleverd heeft dan op 18 januari. In zekere zin is het vergelijken van deze dagen niet terecht, aangezien het achterliggende productiemodel geheel verschillend is. Een extra factor die bijvoorbeeld voor vertraging zorgde was het wachten op aansluitingen van treindiensten die opgeknapt waren, dit zorgt direct voor een verlaging van de punctualiteit. Toch is het interessant om te zien dat van te voren

ingrijpen bij een verwachte lage punctualiteit niet direct de punctualiteit verhoogt, maar wel de dienstverlening aan de klant.

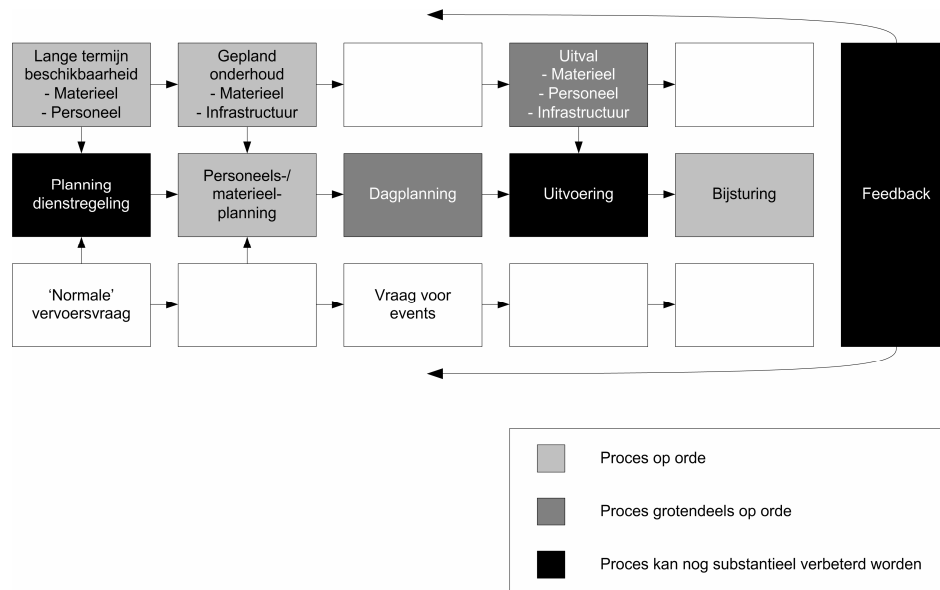
Voorspellingen kunnen dus onder andere gebruikt worden voor een aantal dingen:

1. Benchmarking van de eigen prestaties
Wanneer heb je het als PE of treinserie goed gedaan? Voorspellingen kunnen hier goed bij helpen. Wanneer het model voorspelt dat de punctualiteit op een bepaalde dag erg laag zal zijn, en een PE realiseert een (veel) hogere punctualiteit dan heeft deze PE voor de bewuste dag goed gepresteerd. Wordt lager gescoord, dan kan onderzocht worden waar dat precies aan lag.
2. Verhogen van de dienstverlening
Op dagen waar een zeer lage punctualiteit voorspeld wordt, kan de NS kiezen voor het aanpassen van de dienstregeling zodat een minimaal niveau van *zekere* dienstverlening geleverd kan worden. Een goed voorbeeld was de beperkte dienstregeling op 8 februari.
3. Aanpassen bijsturingscapaciteit
Op dagen dat een (extreem) lage punctualiteit verwacht wordt kan de bijsturingscapaciteit (operationeel managers, knooppunt coördinatoren, etc.) verhoogd worden. Door de capaciteit aan te passen kan sneller ingesprongen worden op calamiteiten die ontstaan.
4. Waarschuwen van klanten voor 'rampzalige' dagen
Een gewaarschuwde klant zal vertragingen eerder accepteren en kan rekening houden met een langere reistijd.
5. Middel voor onderhandelingen
De prognose kan ook gebruikt worden voor onderhandelingen met leidinggevend en overheden. Wanneer bijvoorbeeld bij (sterk) toenemende reizigersaantallen de overheid eist dat de punctualiteit verder verhoogd wordt, kan met behulp van de modellen gezien worden dat een stijgend reizigersaantal een lagere punctualiteit zal geven (ervan uitgaande dat deze relatie bestaat, wat natuurlijk onderzocht moet worden).

Er bestaat binnen de NS behoefte aan verschillende termijnen voor de prognoses. Op hoger managementniveau is vooral behoefte aan voorspellingen waar de gemiddelde punctualiteit op uit gaat komen aan het eind van het jaar in verband met de afspraken met de overheid. Voor deze prognose wordt op dit moment een doorlopend gemiddelde gebruikt van de afgelopen 365 dagen.

Naast de jaargemiddelde bestaat ook behoefte aan een prognose op (zeer) korte termijn: vandaag weten wat morgen gaat gebeuren. Deze prognose bestaat nog niet binnen de NS, en in dit onderzoek zal geprobeerd worden een model te vinden dat deze prognose kan geven. Om te kunnen voldoen aan de gebruiksdoelen die hierboven genoemd zijn, zal het model vooral in staat moeten zijn om grote punctualiteitstijgingen en dalingen te voorspellen. Kleine verschillen zijn minder belangrijk: een voorspelde punctualiteit van 86,7% met een werkelijke punctualiteit van 86,8% is een zeer goed resultaat. Kleine verschillen zijn dus minder belangrijk dan grote verschillen: wanneer bijvoorbeeld de voorspelling van de punctualiteit een waarde heeft van 75%, zou dat voor de NS reden kunnen zijn om extra bijsturingscapaciteit te regelen. Wanneer er vervolgens niets aan de hand is (de bijsturing heeft een grote overcapaciteit door de extra bemensing) en de werkelijke punctualiteit komt op 86%, dan is er extra geld uitgegeven voor niets door een fout van het voorspellingsmodel. Een grote fout is dus belangrijker dan een kleine fout.

Het voorspellingsmodel past binnen een onderzoek dat uitgevoerd is door BCG. Het onderzoek van BCG (BCG, 2005) richtte zich voornamelijk op een vergelijking van de aankomstpunctualiteit 2004 en 2005. Belangrijkste conclusie van dit onderzoek is het hieronder weergegeven diagram. Een model voor het voorspellen van de punctualiteit, of een beter begrip van onderliggende oorzaken van vertraging kan bijdragen aan het verbeteren van de planningsprocessen.



Figuur 4.1 Planningproces met mogelijke verbeteringen (BCG, 2005)

4.2 Voorspellingsmethodes

Walonick (1992) noemt in zijn artikel een drietal veronderstellingen over voorspellingen, de volgende twee veronderstellingen (vertaald naar de relevante situatie) zijn van belang voor het punctualiteitmodel:

1. Er bestaat geen manier om de toekomst met volledige zekerheid te voorspellen. Onafhankelijk van de methode die gebruikt wordt zal er altijd een factor onzekerheid blijven bestaan.
2. [...]
3. Het verstrekken van voorspellingen aan procesbestuurders helpt hen om het proces anders (beter) te besturen. De veranderde besturing zal resulteren in een ander punctualiteitscijfer, waardoor de juistheid van de voorspelling afneemt.

Er zijn veel verschillende methodes voor het voorspellen van de toekomst. De (wetenschappelijke) onderzoeken zijn dan ook steeds gericht op het nauwkeuriger voorspellen van de toekomst. Er worden jaarlijks ook verschillende voorspellingswedstrijden georganiseerd, op formele (M-competition van het *International Institute of Forecasters*) en minder formele (waarvan de voetbal toto misschien wel de bekendste is) basis. De volgende methodes kunnen onderscheiden worden (naar Hanke, 2005; Walonick 1992), voor een uitgebreide beschrijving van de geselecteerde methodes zie paragraaf 4.3.

4.2.1 Kwalitatieve voorspellingsmethodes

De kwalitatieve voorspellingsmethodes zijn over het algemeen gebaseerd op intrinsieke kennis of vaardigheden van de voorspeller. Onder deze methodes vallen onder andere *genius forecasting*, *scenario's* en *consensus methodes*.

Genius forecasting is een methode gebaseerd op intuïtie. Deze methode berust volledig op de inzichten van de voorspeller waardoor het dus onmogelijk is om de waarde van een dergelijke voorspelling te bepalen totdat het voorspelde moment voorbij is. Wetenschappelijk gezien wordt deze methode niet erkend, omdat de wetenschap op dit moment niet ver genoeg is om deze verschijnselen te verklaren.

Scenario's is een voorspellingsmethode waarbij op basis van waargenomen variabelen een verhalende voorspelling gedaan wordt. Over het algemeen worden scenario's geschreven als een lange termijn verwachting en wordt het gebruikt om een organisatie voor te bereiden op een goede, normale en slechte voorspelling.

Consensus methodes gaan uit van de opinie en consensus van meerdere experts op een bepaald gebied. De meest betrouwbare techniek voor het maken van een consensus is de *Delphi* methode. De Delphi methode maakt gebruik van verschillende rondes waarin experts hun mening geven en vervolgens bijstellen aan de hand van wat andere experts voorspeld hebben. Op deze (geanonimiseerde) manier ontstaan redelijk snel een consensus over de afhankelijke variabele.

4.2.2 Kwantitatieve voorspellingsmethodes

Kwantitatieve voorspellingsmethodes zijn over het algemeen gebaseerd op wiskundige relaties die de werkelijkheid zo nauwkeurig mogelijk beschrijven. De relaties worden gevonden door het verleden te analyseren, kwantitatieve voorspellingsmethodes gaan er dan ook vanuit dat het verleden representatief is voor wat in de toekomst gaat gebeuren. Kwantitatieve methodes zijn dan ook alleen betrouwbaar als de omgeving niet teveel verandert, de wiskundige relaties zijn niet betrouwbaar (ver) buiten het interval waarop de relaties gebaseerd zijn. De methodes die vallen onder kwantitatieve methodes:

Een *beslisboom* (decision tree) kan gebruikt worden om op basis van de input van de gebruiker een voorspelling te maken. De beslisboom leidt de gebruiker door een set van vragen die uiteindelijk leiden tot een situatie die met de grootste kans voor zal komen. Beslisbomen zijn in staat om verschillende soorten variabelen (dus niet alleen kwantitatieve variabelen) in één model te gieten. Overigens is een beslisboom ook in staat om kwalitatieve prognoses te geven.

Het *k-nearest neighbor algoritme* (*k-dichtstbijzijnde burens*, *k-NN*) maakt – net als de meeste andere voorspellingsmethodes – gebruik van kennis uit het verleden. Deze kennis wordt echter niet in zijn geheel gebruikt in een vergelijking of model, maar wordt lokaal gebruikt rondom de te voorspellen variabele. Het algoritme berekent een prognose aan de hand van een *k* aantal nearest neighbours welke een minimale afstand hebben tot het punt waar een prognose gemaakt moet worden. Aan de hand van de geselecteerde neighbours kan vervolgens een (gewogen) gemiddelde gebruikt worden voor het bepalen van de prognose. In principe is het *k-NN* algoritme een extreem eenvoudige maar bewerkelijke manier voor het maken van een prognose. Het algoritme doet niets anders dan kijken of een zich voordoende situatie al eens eerder voorgekomen is, en wat er toen gerealiseerd is. De uitkomst van dat zoekproces is de voorspelling. Een groot voordeel van deze methode is dat er geen onafhankelijke variabelen bekend hoeven te zijn voor de dag waarvoor de prognose gemaakt moet worden aangezien de methode alleen kijkt naar de voorgaande dagen. Een belangrijk nadeel hiervan is dat de kennis die er wel is (het is bekend wat voor een weer het wordt) genegeerd wordt.

Trend extrapolatie probeert aan de hand van data uit het verleden trends en cycli te ontdekken. Aan de hand van deze trends kan met behulp van wiskundige methodes een voorspelling gemaakt worden. Het idee achter deze methode is dat de onafhankelijke variabelen die invloed uitoefenden op de afhankelijke variabele dit zullen blijven doen in de toekomst. Deze veronderstelling is over het algemeen geldig voor korte termijn voorspellingen, maar kan problematisch zijn voor langere termijn voorspellingen.

Een begrip dat helpt bij het bepalen of extrapolatie een geschikte methode is, is ‘*developmental inertia*’ (ontwikkelingstraagheid). Data waarbij de ontwikkelingstraagheid hoog is (dus kleine veranderingen) zijn uitermate geschikt voor voorspelling doormiddel van trend extrapolatie.

Wiskundige methodes die gebruikt kunnen worden voor trend extrapolatie zijn onder andere *weighted smoothing* en *enkelvoudige lineaire regressie*.

Simulatie methodes gebruiken analogieën om complexe systemen te modelleren. Een voorbeeld van een wiskundige analogie is een vergelijking voor het voorspellen van de groei van de economie. Onder de wiskundige analogie valt ook de *meervoudige lineaire regressie*. Deze methode probeert een relatie te vinden tussen een aantal verschillende onafhankelijke variabelen die invloed uitoefenen op de afhankelijke variabele, dit in tegenstelling tot simpele lineaire regressie waarbij de relatie tussen slechts één variabele en de afhankelijke variabele gezocht wordt.

De *cross-impact matrix methode* is gericht op het vinden van relaties tussen verschillende gebeurtenissen. De cross-impact matrix methode erkent dat het voorkomen van een gebeurtenis de kans dat een andere gebeurtenis voorkomt verhoogt (het regent en het is koud, dus de kans op ijsel neemt toe). Een belangrijk voordeel van deze methode is dat voorspellers moeten kijken wat de relatie tussen verschillende variabelen is, in plaats van alle variabelen apart te bekijken.

Een geavanceerde methode voor het maken van prognoses zijn *artificiële neurale netwerken* (ANN's, of kortweg neurale netwerken, NN's). Het bekendste voorbeeld van een (niet artificieel) neuraal netwerk is het menselijke brein: het brein bestaat uit een grote hoeveelheid neuronen (10^{11} , Kusters, 2006) en verbindingen. Neuronen krijgen verschillende signalen binnen en verwerken dit tot één signaal, de verbindingen geven de signalen door naar andere neuronen. Artificiële netwerken zijn een extreem vereenvoudigde (in grootte) benadering van het menselijke brein: met behulp van geprogrammeerde neuronen wordt een netwerk gemaakt dat in staat is om te voorspellen. Een NN 'leert' doormiddel van een zeer uitgebreide set voorbeelden wat de gewenste output is bij een gegeven set o.v., dit is vergelijkbaar met het leren van taken van het menselijke brein. Een groot voordeel van een NN is dat het model⁵ waarmee de voorspelling gedaan kan worden niet bekend hoeft te zijn. Het NN leert zelf welke 'relaties' bestaan tussen input variabelen en de output. Een regressiemethode vereist wel dat het model van te voren bekend is.

De *Box-Jenkins (ARIMA) methode* onderscheidt zich van de andere methodes (behalve k-NN) omdat deze methode geen onafhankelijke variabelen gebruikt. De methode gebruikt informatie uit de afhankelijke variabele om de afhankelijke variabele te voorspellen. De ARIMA (Autoregressive integrated moving average) methode gaat sterk uit van de autocorrelatie van data: dit betekent niet meer dan dat de waarde van de a.v. samenhangt met de a.v. van één dag, twee weken, drie maanden, etc. geleden. ARIMA is een methode die met behulp van enkele iteraties een model opstelt en verbetert totdat het de historische data accuraat voorspelt. De a.v. wordt voldoende goed voorspeld als de residuen⁶ klein zijn, normaal verdeeld zijn en geen informatie meer bevatten.

4.2.3 Combinatiemethodes

Zowel voor kwalitatieve als kwantitatieve methodes is het mogelijk om verschillende methodes te combineren en zo tot een meer betrouwbare voorspelling te komen. Een voorbeeld van het combineren van methodes is de Pollyvote, ontwikkeld door J. Scott Armstrong en Randall J. Jones (2006). De Pollyvote voorspelde de uitkomst van de presidentsverkiezingen in de VS op een overwinning voor Bush met 51,5 procent van de stemmen. De uiteindelijke uitslag was een winst van 51,3 procent voor de voorspelde winnaar.

4.3 Methode selectie en theorie

In deze paragraaf zal allereerst beredeneerd worden waarom voor een methode gekozen wordt. Vervolgens zal de theorie achter de methode(s) behandeld worden.

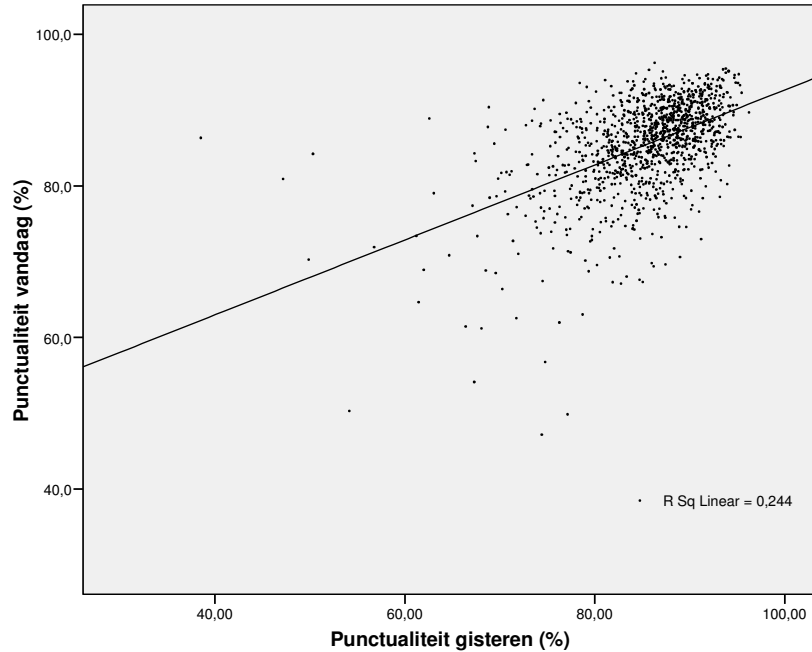
4.3.1 Methode selectie

Over het voorspellen van de punctualiteit kunnen de volgende veronderstellingen gemaakt worden:

1. De punctualiteit moet kwantitatief voorspeld worden.
 2. De punctualiteit wordt beïnvloed door (meerdere) onafhankelijke variabelen.
 3. Voorspellingen moeten snel en regelmatig gemaakt worden.
 4. Het model moet de punctualiteit van morgen kunnen voorspellen (dus 24 uur vooruit).
 5. De punctualiteit van vandaag is lijkt verbonden met de punctualiteit van gisteren.
- Zoals in het onderstaande figuur gezien kan worden, is er op het eerste oog een redelijk (lineair) verband tussen de punctualiteit van vandaag (x-as) en de punctualiteit van morgen (y-as). Het verband is echter niet heel sterk: er is een flink aantal uitzonderingen dat niet voldoet aan het (lineaire) verband ($R^2 \approx 0,26$). Er moeten dus meer invloeden zijn.

⁵ Bijvoorbeeld: de punctualiteit is lineair met de temperatuur, of: de punctualiteit is lineair met de *logaritme* van de temperatuur.

⁶ Het verschil tussen de voorspelling en de waargenomen waarde.



Figuur 4.2 Relatie punctualiteit gisteren - punctualiteit vandaag (data van 1 januari 2004 tot 11 maart 2007)

Deze veronderstellingen leiden dus tot een kwantitatieve analyse, met meerdere variabelen als input en een dagelijkse (snelle) behoefte aan voorspellingen. Op basis van deze voorwaarden lijken de volgende methodes het meest aan te sluiten:

1. Meervoudige regressie analyse
2. Box-Jenkins (ARIMA) methode
3. Artificiële neurale netwerken
4. k-Nearest Neighbor algoritme

Deze methodes zullen in de volgende paragrafen uitgebreider behandeld worden.

4.3.2 Notatie

Om de leesbaarheid van de volgende paragrafen te verhogen zal overal dezelfde notatie worden gebruikt voor de onafhankelijke variabelen en afhankelijke variabelen.

- De dataset bestaat uit K waarnemingen, k is een willekeurige waarneming.
- Een waarneming wordt benoemd met $W(k)$.
- Iedere $W(k)$ bestaat uit een vector met I onafhankelijke variabelen x waarvan (nog) niet bepaald is of deze invloed hebben op de afhankelijke variabele:

$$W(k) = \{x_{k,1}, x_{k,2}, \dots, x_{k,I}\} \quad (4.1)$$

- De voorspelde waarneming is $\hat{Y}$, de waargenomen waarde voor deze voorspelling is Y . De voorspelde waarneming voor $W(k)$ is dus $\hat{Y}_k$.

4.3.3 Meervoudige lineaire regressie

Meervoudige lineaire regressie (MLR) zoekt een vergelijking die gebruikt kan worden om een variabele Y ($Y_1, Y_2, \dots, Y_K$) te voorspellen aan de hand van variabelen $x_{k,i}$. Deze vergelijking heeft de volgende vorm:

$$Y_k = \beta_0 + \sum_{i=1}^N (\beta_i x_{k,i}) + \varepsilon_k \quad (4.2)$$

In deze vergelijking staat β_0 voor de beginwaarde van de vergelijking, $\beta_i x_{k,i}$ vertegenwoordigt de i -de variabele die van invloed is met waarde $x_{k,i}$ voor de k -de waarde uit de reeks $Y_1, Y_2, \dots, Y_K$. ε_k staat voor een onafhankelijke stochastische variabele met $\varepsilon_k \sim N(0, \sigma^2)$, dus normaal verdeeld. ε_k staat in de vergelijking om de afwijkingen van de individuele waargenomen variabelen (Y_k) te corrigeren. Om een goed voorspellingsmodel te maken, moeten de waarden van alle β_i voor $i = 1, \dots, I$ en van σ^2 geschat worden. Aangezien we nooit de precieze regressie-vergelijking te weten kunnen komen, wordt de aangepaste vergelijking geïntroduceerd. Deze vergelijking zal gebruikt worden als voorspellingsmodel, onderstaande vergelijking is een voorbeeld voor $I = 3$:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 \quad (4.3)$$

In deze vergelijking staat $\hat{Y}$ voor de schatting van een Y_k en $\hat{\beta}_i$ staat voor een geschatte waarde van β_i .

Een voorbeeld (getallen ontleend aan Kallenberg (2002), deels aangepast) is de relatie tussen de aanwezigheid van een proces bijzonder taken (PBT), het volwassen reizigersaantal, het onvolwassen reizigersaantal en de reisduur in uren. De te voorspellen variabele wordt dus verondersteld afhankelijk te zijn van drie variabelen (x_1, x_2, x_3 ; $I = 3$). Het aantal vrijheidsgraden wordt gedefinieerd als $K - (I + 1)$, dit hebben we later nodig voor het testen van de betrouwbaarheid van de $\hat{\beta}_i$'s. Het aantal vrijheidsgraden voor dit voorbeeld is dus $25 - (3 + 1) = 21$. De volgende gegevens zijn beschikbaar:

Waarneming	Reisduur (uren)	PBT aanwezig	Reizigersaantal (≥18 jaar)	Reizigersaantal (<18 jaar)
	Y_k	$X_{k,1}$	$X_{k,2}$	$X_{k,3}$
1	0,5	1	73	14
2	0,5	1	66	16
3	0,7	0	65	15
4	0,8	0	65	16
5	0,8	0	65	16
6	0,9	1	69	10
7	1,1	1	82	12
8	1,6	1	83	12
9	1,6	1	81	12
10	2,0	0	72	10
11	2,5	1	69	8
12	2,8	0	71	16
13	2,8	0	71	12
14	3,0	0	80	9
15	3,0	0	73	6
16	3,0	0	75	6
17	3,2	0	76	10
18	3,2	0	78	6
19	3,3	1	79	6
20	3,3	0	79	4
21	3,4	1	78	6
22	3,5	0	76	9
23	3,6	0	65	12
24	3,7	0	72	12
$n = 25$	3,7	0	80	6

Analyse van deze data met behulp van een statistisch programma levert de volgende output:

Variabelen in de vergelijking	
Variabele	$\hat{\beta}_i$
PBT aanwezig	-1,1757
Reizigersaantal (≥ 18)	0,03876
Reizigersaantal (< 18)	-0,15228
Constante	1,495

De aangepaste vergelijking (1.2) wordt dus:

$$\hat{Y} = 1,495 - 1,1757x_1 + 0,03876x_2 - 0,15228x_3 \quad (4.3)'$$

Volgens deze vergelijking zou een trein met een PBT, 70 volwassen reizigers en 8 onvolwassen reizigers een verwachte reisduur hebben van 1,8 uur:

$$\hat{Y} = 1,495 - 1,1757 \cdot 1 + 0,03876 \cdot 70 - 0,15228 \cdot 8 = 1,8 \text{ uur} \quad (4.3)'$$

4.3.3.1 Opnemen van een variabele

Of een variabele opgenomen moet worden in de vergelijking blijkt onder andere uit de correlatie van de onafhankelijke variabele met de afhankelijke variabele. Een hoge correlatie duidt op een sterke samenhang van de variabelen. Wanneer daarnaast ook nog een duidelijke oorzaak-gevolg relatie aanwezig is, dan kan de variabele een goede onafhankelijke variabele worden in de regressie vergelijking.

Om te bekijken of er een correlatie bestaat tussen twee variabelen kunnen deze variabelen tegen elkaar uitgezet worden in een grafiek. De relatie hoeft niet lineair te zijn om opgenomen te kunnen worden in de regressie vergelijking: een kwadratische ($Y \sim x^2$) of logaritmische ($Y \sim \log x$) relatie kan ook gebruikt worden.

4.3.4 Box-Jenkins (ARIMA) methode

De Box-Jenkins methode (vernoemd naar twee statistici die de methode uitgewerkt hebben, G.E.P. Box en G.M. Jenkins) bestaat uit twee componenten: het autoregressive model (AR) en het moving average model (MA). Beide modellen zullen hieronder uitgelegd worden (naar Hanke, 2005). De ARIMA (autoregressive integrated moving average) methode integreert de AR en MA modellen.

4.3.4.1 Autoregressive models

AR modellen lijken sterk op de normale regressievergelijkingen zoals beschreven in de vorige paragraaf, alleen zijn de onafhankelijke variabelen x vervangen door de waarden van de afhankelijke variabele in het verleden. Een AR model (waarin termen tot p perioden geleden meegenomen worden, een $AR(p)$ -model) ziet er als volgt uit:

$$Y_k = \phi_0 + \phi_1 Y_{k-1} + \phi_2 Y_{k-2} + \dots + \phi_p Y_{k-p} + \varepsilon_k \quad (4.4)$$

De waarden van de coëfficiënten ϕ_0 kunnen bepaald worden met een statistisch programma, bijvoorbeeld met de kleinste kwadraten methode. De Y_{k-1} , Y_{k-2} , etc. variabelen zijn de waarnemingen van respectievelijk 1, 2, etc. perioden geleden. Ook de AR modellen komt de ε_k waarde voor die de fout aangeeft tussen de waargenomen waarde en de voorspelde waarde Y_k .

4.3.4.2 Moving average models

Ook MA modellen lijken sterk op regressie modellen. Een $MA(p)$ model, waarin wederom termen tot p perioden geleden meegenomen worden luidt:

$$Y_k = \mu + \varepsilon_k - \omega_1 \varepsilon_{k-1} - \omega_2 \varepsilon_{k-2} + \dots + \omega_p \varepsilon_{k-p} \quad (4.5)$$

In deze vergelijking staat μ voor het constante gemiddelde van het proces, ε_k is de fout-term die de effecten van variabelen die niet in het model voorkomen representeert, ε_{k-1} , ε_{k-2} , ε_{k-p} zijn de fouten die waargenomen zijn in de voorgaande perioden in Y_k , ω_1 , ω_2 , ω_p zijn de coëfficiënten die geschat moeten worden.

Een belangrijk verschil tussen MA en AR modellen is dat MA modellen de waarde van Y_k voorspellen aan de hand van fouten uit de verleden, terwijl AR modellen Y_k voorspellen aan de hand van de waarde van Y_k in het verleden.

4.3.4.3 Autoregressive Moving Average Models

ARMA modellen combineren beide voorgaande methodes. Het model (ARMA(p, q)) ziet er dan als volgt uit:

$$\hat{Y}_k = \phi_0 + \sum_{n=1}^p \phi_n Y_{k-n} + \varepsilon_k - \sum_{m=1}^q \omega_m \varepsilon_{k-m} \quad (4.6)$$

Het opstellen van deze vergelijking verloopt in een aantal (iteratieve) stappen:

1. Identificatie van het model
Allereerst is het van belang om te kijken of de data stationair zijn: de waarden moeten bewegen rondom een vast gemiddelde. Wanneer dit niet het geval is, kan de data meestal alsnog stationair gemaakt worden door de waarden te vervangen door de verschillen tussen de waarden. Het ARMA(p, q) model verandert dan in een ARIMA(p, d, q) model, de d geeft aan hoe veel ‘verschillen’ genomen moet worden. Een ARIMA($p, 0, q$) model is dus een ARMA(p, q) model.
Vervolgens kan bepaald worden welk model gebruikt moet worden aan de hand van bestudering van de data.
2. Schatten van de coëfficiënten
Het schatten van de coëfficiënten kan gedaan worden met behulp van de kleinste kwadraten methode. Het toetsen van de significantie van de coëfficiënten gebeurt op ongeveer dezelfde manier als bij de normale regressie (zie paragraaf 0).
De fit van het model wordt bepaald met behulp van de residual mean square error:

$$s^2 = \frac{\sum_{k=1}^K e_k^2}{n-r} = \frac{\sum_{k=1}^K (Y_k - \hat{Y}_k)^2}{n-r} \quad (4.7)$$

Waarin n staat voor het aantal residuen en r voor het aantal coëfficiënten dat geschat is.

3. Checken van het model aan de hand van waargenomen waarden
Een ARIMA model is over het algemeen goed wanneer de residuen niet meer gebruikt kunnen worden om de voorspelling te verbeteren, wat zoveel betekent als dat de residuen willekeurig verdeeld zijn. Voor het bepalen van de kwaliteit van het model kan de chi-kwadraat toets gebruikt worden op basis van de Ljung-Box waarde:

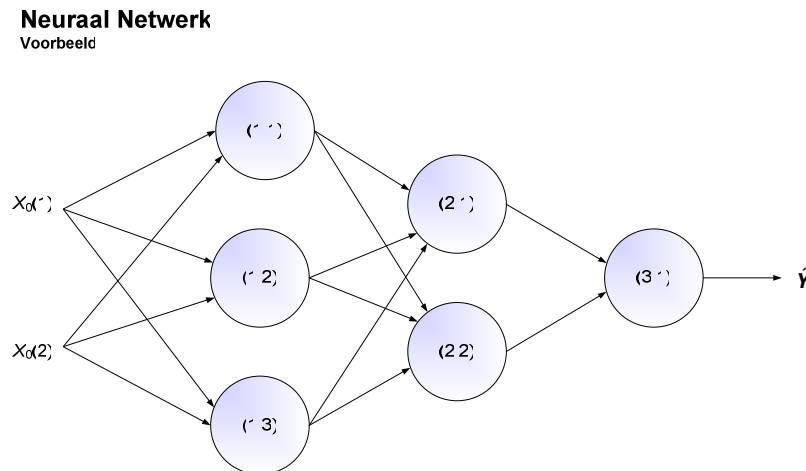
$$Q_Z = n(n+2) \sum_{z=1}^Z \frac{r_z^2(e)}{n-z} \quad (4.8)$$

Waarin n wederom het aantal residuen is, Z het aantal time lags (p voor AR, q voor MA) dat getest moet worden, $r_z(e)$ de autocorrelatie van de residuen bij time lag z . Wanneer de p -

- waarde horende bij de waarde van Q_z klein is (bijvoorbeeld $p < 0,05$), wordt het model niet als juist gezien.
4. Voorspellen met het model

4.3.5 Artificiële Neurale Netwerken

Een artificieel neuraal netwerk (NN) maakt gebruik van een netwerk van neuronen om aan de hand van een aantal gegeven variabelen (input) een gevraagde variabele (output) te berekenen. Deze berekening verloopt niet via een standaard formule, maar via de neuronen die in het netwerk opgenomen zijn. De neuronen geven verschillende outputs aan de hand van de input die zij van voorgaande neuronen ontvangen.



Figuur 4.3 Een voorbeeld van een (eenvoudig) NN, een bolletje is een neuron

4.3.5.1 Aanvullende notatie

In Figuur 4.3 staat een voorbeeld van een NN met drie lagen (laag 1: (1,1), (1,2), (1,3), etc.) en totaal zes neuronen. De x -waarden ($x_0(1)$ en $x_0(2)$) in de eerste laag bevinden zich in de *input* laag. Neuron zes bevindt zich in de *output* laag. De overige neuronen ((1, {1,2,3}) en (2, {1,2})) bevinden zich in de *verborgen* laag⁷. Allereerst is het van belang om te begrijpen hoe de nummering van de lagen en neuronen werkt. De laag wordt gedefinieerd door $m = 1, \dots, M$ en het neuron nummer (gedefinieerd per laag) door $n = 1, \dots, N$. Het neuron wordt aangegeven met (m, n) . Het eerste neuron in de derde laag heet dus (3,1), het vijfde neuron in de eerste laag heet (1,5). In de dataset worden de individuele waarnemingen, zoals eerder beschreven, genummerd met $k = 1, 2, \dots, K$. De waarnemingen blijven bestaan uit een vector $\{x_{k,1}, \dots, x_{k,l}\}$ welke dus vertaald worden naar $x_0(1), x_0(2)$, etc. Het verschil blijkt vooral uit de subscripts.

Variabelen in het netwerk worden aangegeven met $x_m(n)$, met m het laagnummer en n het neuronnummer zoals eerder. De input variabelen zijn $x_m(n)$ ($x_0(1)$ en $x_0(2)$), de output variabele is $\hat{Y}$. Omdat de output van een neuron gelijk is aan de input van een opvolgend neuron, is de notatie van de in- en output variabelen gelijk. De output van neuron drie in de eerste laag is dan dus $x_1(3)$ wat de input is voor de neuronen in laag twee. De waarde van een laag m wordt aangeduid met de net_m , wat een vector is met de waarden van de aparte neuronen: $net_m = \{x_m(1), \dots, x_m(N)\}$. De pijlen geven de verbindingen tussen de verschillende neuronen aan.

Het gewicht w tussen neuron i op laag m en neuron n op laag $m+1$ wordt $w_m(i, n)$ genoemd. Het gewicht tussen neuron (1,1) en neuron (2,3) is dus $w_1(1,3)$.

⁷ Het aantal verborgen lagen kan variëren, in dit voorbeeld zijn twee verborgen lagen gedefinieerd.

4.3.5.2 Netwerk

Het type van het netwerk in Figuur 4.3 is een *feed-forward* netwerk, of nauwkeuriger: een *multilayered perceptron*, aangezien de output van een neuron alleen doorgegeven wordt naar de volgende laag in het NN en niet terug via een cyclus. De waarde die een neuron aanneemt is afhankelijk van de input die het neuron ontvangt en het gewicht dat toegekend is aan de input van een voorgaand neuron, dus voor het eerste neuron in de eerste laag (met inputs $x_0(1)$ en $x_0(2)$):

$$x_1(1) = f_1(net_0) = \sum_{i=1}^2 [w_0(i,1) \cdot x_0(i)] = w_0(1,1) \cdot x_{0,1} + w_0(2,1) \cdot x_{0,2} \quad (4.9)$$

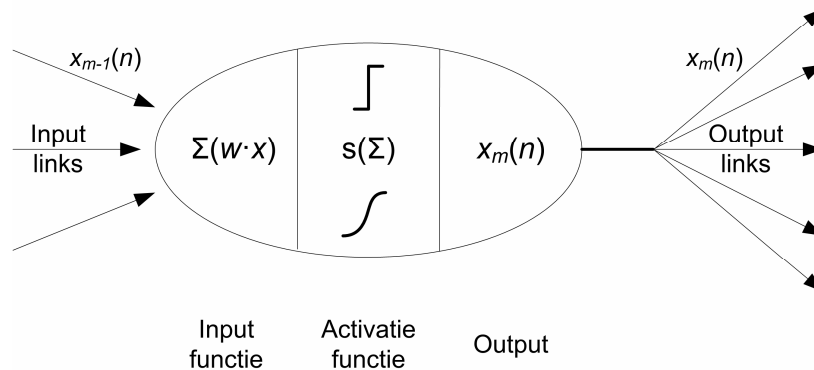
Met $w_0(i,1)$ het gewicht dat is toegekend aan de verbinding tussen neuron $(0, i)$ en neuron $(1, 1)$. Of algemener:

$$x_m(n) = f_n(net_{m-1}) = \sum_{i=1}^I [w_{m-1}(i,n) \cdot x_{m-1}(i)] \quad (4.10)$$

4.3.5.3 De werking van een neuron

Een neuron behandelt zijn input met vergelijking (4.10) (input functie), in Figuur 4.4 is een neuron weergegeven. Een neuron krijgt inputsignalen van zijn voorgaande neuronen. Deze signalen worden door het neuron opgeteld met vergelijking (4.10). Vervolgens wordt met de activatiefunctie de output van het neuron bepaald. De output van het neuron ligt altijd in het interval $(0, 1)$ of $(-1, 1)$, afhankelijk van het type activatiefunctie dat gebruikt wordt.

Neuron (m, n)



Figuur 4.4 Neuron met input en output (Kosters, 2006)

De activatiefunctie luidt in formulevorm:

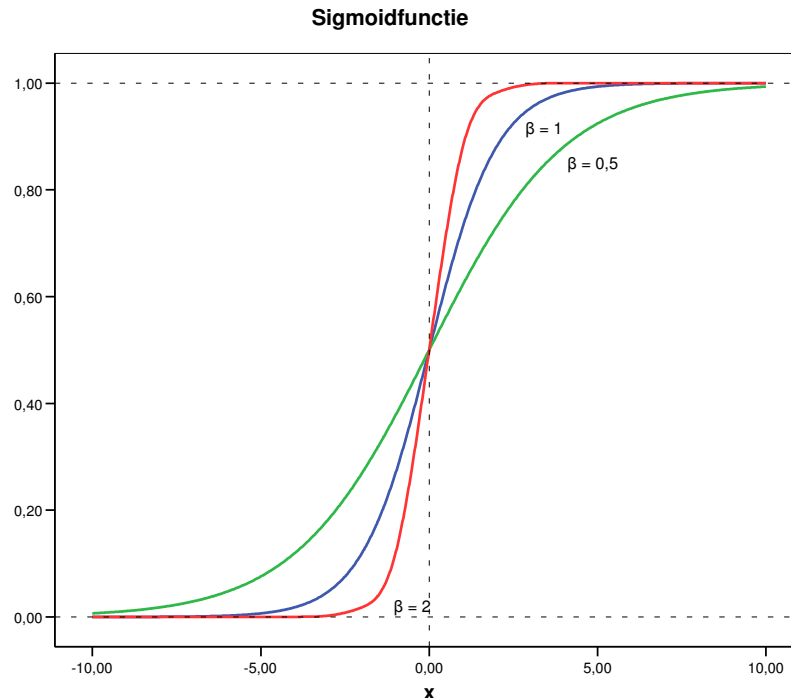
$$x_m(n) = s\left(\sum (w \cdot x_{m-1})\right) \quad (4.11)$$

Voor de functie $s(\dots)$ zijn verschillende opties beschikbaar, maar de sigmoidfunctie wordt vaak gebruikt. De belangrijkste reden hiervoor is dat de sigmoidfunctie continu is op $0 < y < 1$, zie ook Figuur 4.5. Tijdens het leerproces zullen de gewichten w tussen de neuronen langzaam aangepast worden en omdat de sigmoidfunctie continu is, zal de activatiewaarde van het neuron ook langzaam veranderen in plaats van wisselen tussen 0 en 1 bij de $\text{step}(x)$ functie. (Kosters, 2006)

De sigmoidfunctie luidt:

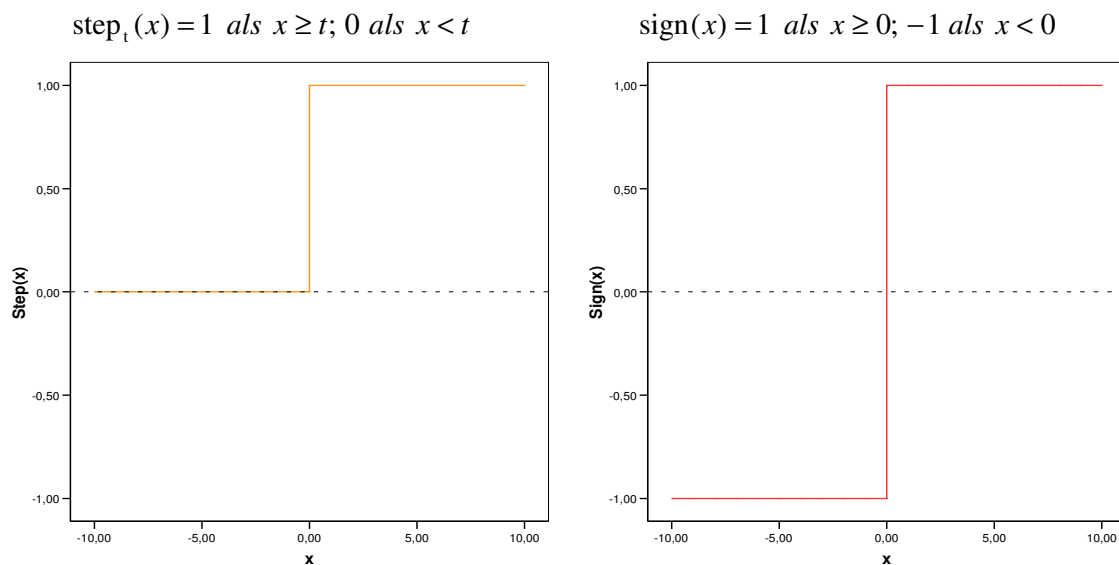
$$s(x) = \frac{1}{(1 + e^{-\beta x})} \quad (4.12)$$

De β kan gebruikt worden om de functie te verschuiven.



Figuur 4.5 De sigmoidfunctie voor verschillende β waarden.

Naast de sigmoidfunctie voor de output van een neuron bestaan nog andere functies. Twee voorbeelden hiervan zijn de step functie en de sign functie:



Figuur 4.6 Step en sign functie

4.3.5.4 Backpropagation

De gewichten (w) moeten bepaald worden met behulp van een leerproces. Een veelgebruikt proces hiervoor is *backpropagation*. Bij backpropagation wordt aan de hand van een dataset een groot aantal voorspelde waarden ($\hat{Y}$) berekend met behulp van het NN. De gewichten tussen de neuronen worden vervolgens aangepast aan de hand van de fout tussen de voorspelling en de waargenomen waarde. Om de gewichten (w) aan te passen in het geval van de sigmoid activatie functie, wordt de volgende vergelijking gebruikt:

$$\Delta w_m(i, n) = \frac{\eta}{K} \sum_{k=1}^K x_{m-1,k}(i) \cdot \delta_{m,k}(n) \quad (4.13)$$

In deze vergelijking staat de η voor de leersnelheid en $x_{m-1,k}$ voor de waarde van de k -de waarneming uit de dataset en in het n -de neuron op laag $m-1$. De complete dataset wordt in één keer bekeken. Aangezien het bekijken van een complete dataset voordat de gewichten aangepast worden erg tijdrovend is, worden over het algemeen per waarneming in de dataset de gewichten aangepast:

$$\Delta w_m(i, n) = \eta \cdot x_{m-1}(i) \cdot \delta_m(n) \quad (4.14)$$

De $\delta_m(n)$ wordt bepaald door de fout die een neuron n heeft ten opzichte van de gewenste waarde van het neuron in relatie tot het opvolgende neuron l :

$$\delta_m(n) = x_m(n) \cdot [1 - x_m(n)] \cdot w_{m+1}(n, l) \cdot \delta_{m+1}(l) \quad (4.15)$$

En in de output laag:

$$\delta_m(n) = \hat{Y} \cdot [1 - \hat{Y}] \cdot [Y - \hat{Y}] \quad (4.16)$$

Met $\delta_m(n)$ de fout in de output laag voor een meting k , Y de waargenomen waarde voor meting k en $\hat{Y}$ de met het NN berekende waarde voor meting k . (Picton, 2000)

Het leerproces verloopt in *epochs*. Een epoch omvat alle data (K) uit een trainingsset. Het algoritme verloopt als volgt:

```
Voor iedere k in trainingsset:
    Stel de invoerwaarden van het NN gelijk aan de waarden uit k
    Bereken de outputs  $x_m$ 
    Bereken de  $\delta_m(n)$ 's
    Update de  $w_m(i, n)$ 's
Tot het netwerk is 'geconvergeerd'
```

Code 4.1 Algoritme voor het bepalen van de gewichten (w) (Kosters, 2006)

4.3.5.5 Normalisatie

De beperking tot de intervallen (0, 1) en (-1, 1) betekent dat input data altijd vertaald moet worden als het interval van deze data anders is. Wanneer bijvoorbeeld van reizigersaantallen bekend is dat deze in het interval (5, 400) liggen, dan moet dit vertaald worden naar (0, 1) voor de sigmoidfunctie. Om een getal v uit het interval (c, d) te vertalen naar het interval (e, f) wordt de volgende functie gebruikt:

$$v' = (f - e) \cdot \left(\frac{v - c}{d - c} \right) + e \quad (4.17)$$

Dus een reizigersaantal van 252 vertaald naar het interval $(-1, 1)$ levert op:

$$v' = (1 - (-1)) \cdot \left(\frac{252 - 5}{400 - 5} \right) + (-1) \approx 0,25 \quad (4.17)'$$

Terugrekenen gaat met dezelfde formule, alleen zijn de intervallen dan omgedraaid. Of, in omgeschreven vorm:

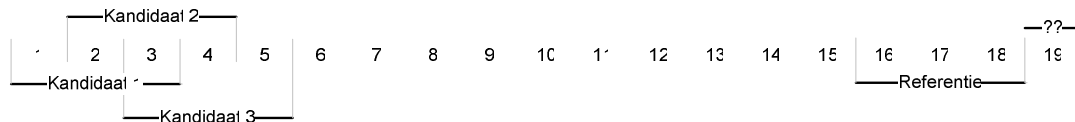
$$v = (d - c) \cdot \left(\frac{v' - e}{f - e} \right) + c \quad (4.17)'$$

4.3.5.6 De vorm van het neurale netwerk

De vorm van het neurale netwerk is afhankelijk van de functie van het netwerk. Er bestaan veel verschillende netwerken, inclusief complexe netwerken zonder een organisatie in lagen. Voor voorspellingsdoeleinden is een (eenvoudig) feed-forward netwerk met één of twee verborgen lagen voldoende (Swanson, 1997; Hippert, 2005; De Gooijer, 2006; Hornick, 1989; Picton, 2000). Het aantal neuronen in de inputlaag is gelijk aan het aantal input variabelen dat gebruikt wordt. Voor de verborgen lagen geldt dat het aantal benodigde neuronen volgens Kolmogorov (Picton, 2000) is ergens in de buurt van $n(2n+1)$. Kolmogorov stelde (kort gezegd) in 1957 dat iedere lineaire functie met n variabelen in een drie-laags perceptron benaderd kan worden met $n(2n+1)$ neuronen in de verborgen laag. Voorwaarde hierbij is wel dat de neuronen gebruik maken van een continue niet-lineaire outputfuncties, zoals de sigmoidfunctie. Een netwerk dat dus voorspellingen maakt aan de hand van 7 input variabelen, heeft ongeveer $7(14+1) = 105$ neuronen in de verborgen laag nodig.

4.3.6 k-Nearest Neighbor algoritme

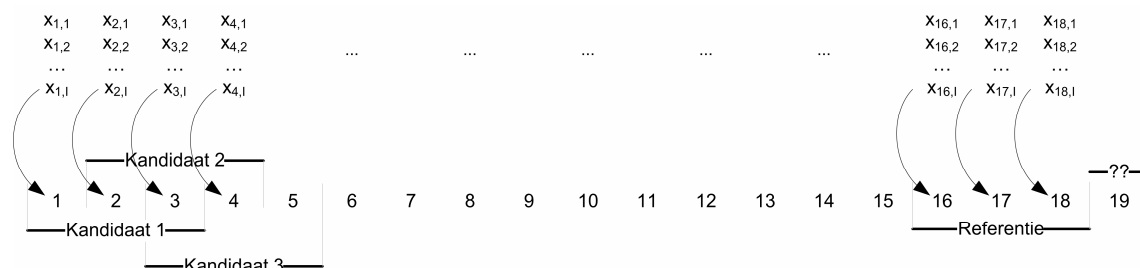
Het k-nearest neighbor algoritme (k-NN) zoekt in een dataset een vergelijkbare situatie als de situatie die voorafgaat aan het te voorspellen moment. Neem bijvoorbeeld de volgende (genummerde) set waarnemingen:



Figuur 4.7 De set waarnemingen

Deze waarnemingen bevatten een vector met waarden voor onafhankelijke variabelen die invloed hebben op de afhankelijke variabele, dus voor waarneming 16:

$$W(16) = \{x_{16,1}, x_{16,2}, \dots, x_{16,I}\} \quad (4.1)'$$

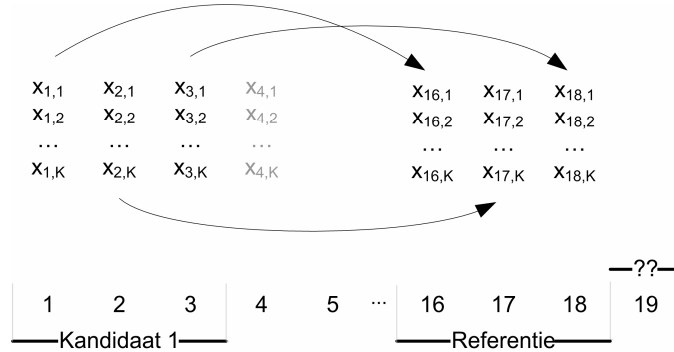


Figuur 4.8 Iedere waarneming k bestaat uit I onafhankelijke variabelen $x_{k,i}$

Het algoritme bekijkt een set van a ($a = 3$ in het voorbeeld, waarbij de a verwijst naar de k in k -NN) waarden vóór het moment waarvoor de voorspelling gemaakt moet worden (19), in het voorbeeld zijn dit dus waarnemingen 16, 17 en 18.

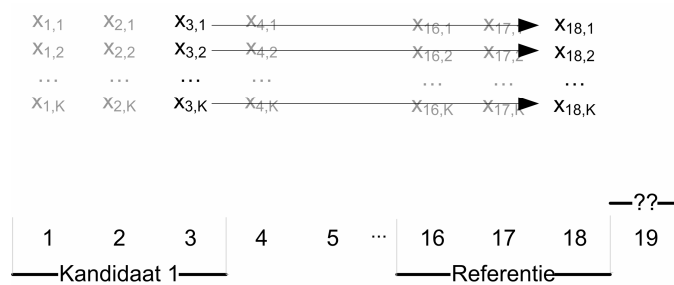
Vervolgens loopt het algoritme lineair door de dataset en berekent de fout tussen de referentie (16, 17, 18) en de kandidaat (bijvoorbeeld 2, 3, 4). De fout tussen de referentie en de kandidaat wordt bepaald door stuksgewijs de variabelen uit de vector $W(k)$ te vergelijken. De som van de procentuele fouten wordt gebruikt als maat voor de overeenkomst tussen de referentie en de kandidaat. Voor de procentuele fout tussen waarneming 16 en waarneming 2 wordt de relatieve afstand tussen de waarden van de individuele onafhankelijke variabelen ($x_{k,i}$) gebruikt:

$$d_{16,2} = \begin{cases} \sqrt{\sum_i \left(\frac{x_{2,i} - x_{16,i}}{x_{16,i}} \right)^2} & \text{voor } x_{16,i} \neq 0 \\ 1 & \text{voor } x_{16,i} = 0 \text{ en } x_{2,i} \neq x_{2,i} \\ 0 & \text{voor } x_{16,i} = x_{2,i} = 0 \end{cases} \quad (4.18)$$



Figuur 4.9 Kandidaat 1 {1, 2, 3} wordt als eerste vergeleken met de referentie {16, 17, 18}.

Met r de startwaarneming van de kandidaat, $x_{0,i}$ de i -de onafhankelijke variabele en $x_{1,i}$ de i -de onafhankelijke variabele uit de referentie. Het verschil tussen de twee waarnemingen $x_{2,i} - x_{16,i}$ wordt gedeeld door $x_{16,i}$ om te corrigeren voor schaalverschillen in de onafhankelijke variabelen. De uitzonderingssituatie dat $x_{16,i}$ gelijk is aan nul komt (voornamelijk) voor bij nominale variabelen (variabelen met de waarde 0 of 1), in dat geval betekent een verschil tussen de variabelen $x_{2,i}$ en $x_{16,i}$ altijd een verschil van 100% voor de berekening. De totale fout d is vervolgens een maat voor de overeenkomst tussen waarneming 16 en waarneming 2: wanneer $d = 0$, dan komen beide waarnemingen exact overeen, wanneer $d > 0$, dan komen de waarnemingen niet overeen, maar kan zonder verdere studie niets gezegd worden over wat er dan verschilt.



Figuur 4.10 Voor de bepaling van $d_{18,3}$ worden de o.v.'s van waarnemingen 3 en 18 met elkaar vergeleken met vergelijking (4.18).

De fout (F) tussen de referentie (16, 17, 18) en de kandidaat (2, 3, 4) is:

$$F = d_{16,2} + d_{17,3} + d_{18,4} \quad (4.19)$$

De kandidaat met de kleinste fout F wordt gebruikt voor de voorspelling: de waarneming *na* deze kandidaat is de prognose voor het te voorspellen moment. Als kandidaat 3, 4, 5 geïdentificeerd is met de kleinste fout, dan is de waarde van de prognose voor 19 dus de waarde van 6. Wanneer meerdere kandidaten gevonden worden die een (bijna) gelijke fout hebben, dan kan het gemiddelde van de opvolgende waarden genomen worden. Dus wanneer kandidaat twee en kandidaat drie een gelijke fout hebben, dan is de prognose voor 19 het gemiddelde van de waarden van 5 en 6.

4.3.7 Welke onafhankelijke variabelen worden gebruikt?

Niet alle onafhankelijke variabelen zullen statistisch gezien invloed hebben op de punctualiteit, daarom is het van belang om selectief te zijn in de variabelen die gebruikt worden. Meer onafhankelijke variabelen hebben voor MLR en ARIMA als risico dat het model overcorrigeert voor gebeurtenissen die weinig of geen invloed hebben, en meer variabelen zorgen voor langere rekentijden voor NN en k-NN.

Voor MLR zijn statistische toetsen beschikbaar die aangeven of een variabele in een vergelijking invloed heeft. Hieronder zal beschreven worden hoe deze statistische toetsen werken en wanneer een variabele gebruikt zal worden voor het voorspellen van de punctualiteit. De statistische toetsen die gebruikt worden voor het bepalen van de relevantie van componenten in de ARIMA vergelijkingen werken op een vergelijkbare manier, en zullen verder niet behandeld worden.

Voor NN geldt dat het ‘leren’ van het netwerk de o.v.’s die niet van invloed zijn zal uitschakelen, of de invloed zal sterk beperkt worden. Het is bij een NN dan ook niet nodig om op voorhand o.v.’s te selecteren. Desalniettemin zal dit toch gedaan worden, aangezien meer o.v.’s de rekentijd van een NN sterk vergroot. Als basis voor het selecteren van de o.v.’s zal de uitkomst van de selectie voor MLR gebruikt worden, o.v.’s die alsnog toegevoegd worden zullen verklaard worden.

Voor het k-NN geldt eveneens dat het vergroten van het aantal gebruikte o.v.’s de rekentijd sterk vergroot. Daarom zal ook voor het k-NN de analyse van het MLR gebruikt worden, met eventuele enkele (beargumenteerde) o.v.’s toegevoegd.

4.3.7.1 Selectie van variabelen voor MLR

Vergelijking (4.3) heeft een drietal $\hat{\beta}_i$ ’s ($I = 3$) die *waarschijnlijk* een goede schatting geven voor de afhankelijkheid van de o.v. x_i . Het is dan ook van belang om de o.v. te testen of ze daadwerkelijk van (voldoende) invloed zijn op de $\hat{Y}$. Om dit te testen is er een drietal toetsen beschikbaar:

1. Ligt nul in het betrouwbaarheidsinterval van $\hat{\beta}_i$?

Het betrouwbaarheidsinterval geeft aan of nul (geen invloed) met een betrouwbaarheid van $1 - \alpha$ behoort tot het interval waarin $\hat{\beta}_i$ ligt. Wanneer nul in het interval ligt, is het onzeker of $\hat{\beta}_i$ invloed heeft op de te voorspellen variabele. Het betrouwbaarheidsinterval met betrouwbaarheid $1 - \alpha$ ziet er als volgt uit:

$$(\hat{\beta}_i - cS_{\hat{\beta}_i}, \hat{\beta}_i + cS_{\hat{\beta}_i}) \quad (4.20)$$

Met $S_{\hat{\beta}_i}$ de geschatte standaardafwijking (standard error) van $\hat{\beta}_i$ en c gegeven door de Student-verdeling:

$$P(T_{n-(k+1)} \leq c) = 1 - \frac{1}{2}\alpha \quad (4.21)$$

De uitdraai van de regressie analyse (bijvoorbeeld met Excel of SPSS), geeft de volgende informatie:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	1,495261	2,6371	0,5669
x_1	-1,1757	0,3156	-3,7258
x_2	0,03876	0,0319	1,2140
x_3	-0,1522	0,0501	-3,0391

De *coefficients* zijn de $\hat{\beta}_i$'s, *standard error* is de geschatte standaardafwijking van $\hat{\beta}_i$ ($S_{\hat{\beta}_i}$), *t Stat* wordt later gebruikt. Voor een betrouwbaarheid van 95% (standaardwaarde, $\alpha = 0,05$) wordt de waarde van c 2,08:

$$P(T_{21} \leq 2,08) = 0,975 \quad (4.22)$$

Voor het betrouwbaarheidsinterval van $\hat{\beta}_1$ wordt dus:

$$(-1,1757 - 2,08 \cdot 0,3156, -1,1757 + 2,08 \cdot 0,3156) = (-1,83; -0,52) \quad (4.20)'$$

Deze uitkomst betekent dat met een betrouwbaarheid van 95 procent dat treinen met een PBT 0,52 tot 1,83 uur sneller aankomen dan een trein zonder PBT. Aangezien nul niet voorkomt in het interval, kan redelijk verondersteld worden dat x_1 van invloed is op de reisduur (in aanwezigheid van de andere verklarende variabelen).

2. Toetsen of $\beta_i = 0$?

Een andere manier om vast te stellen dat x_i van invloed is op $\hat{Y}$ is het toetsen van deze hypothese. Voor het toetsen van statistische grootheden wordt een nulhypothese en een alternatieve hypothese vastgesteld. De nulhypothese wordt verworpen wanneer de overschrijdingskans van de gevonden waarde van de toetsingsgrootte kleiner of gelijk is aan een vooraf afgesproken grenswaarde α . De grenswaarde ligt meestal rond 5 procent (0,05), waardoor de nulhypothese niet snel verworpen wordt. Deze benadering is in principe te vergelijken met strafrechtelijk onderzoek: de verdachte is onschuldig (nulhypothese), tenzij met overtuigend bewijs aangetoond wordt dat de verdachte schuldig is. De nulhypothese wordt aangegeven met H_0 , de alternatieve hypothese met H_1 .

De nulhypothese voor de $\hat{\beta}_i$'s luidt: $H_0 : \beta_i = 0$ (dus: $\beta_i = 0$, tenzij zeer stellig bewezen wordt dat $\beta_i \neq 0$). De nulhypothese wordt verworpen wanneer $|T| \geq c$, met c gegeven door (4.21). De waarde van T volgt uit:

$$T = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}} \quad (4.23)$$

De waarde van T wordt in de uitdraai gegeven door *t Stat*. Voor het eerder gebruikte voorbeeld, kan de uitkomst dus als volgt geïnterpreteerd worden ($c = 2,08$, zoals eerder):

Nulhypothese	T	Conclusie
$H_0 : \beta_1 = 0$	-3,73	Verwerp H_0 ($ -3,73 \geq 2,08$): in aanwezigheid van x_2 en x_3 is x_1 statistisch van belang
$H_0 : \beta_2 = 0$	1,21	Accepteer H_0 ($ 1,21 < 2,08$): in aanwezigheid van x_1 en x_3 is x_2 niet statistisch van belang

$H_0 : \beta_3 = 0$	-3,04	Verwerp H_0 : in aanwezigheid van x_1 en x_2 is x_3 statistisch van belang
---------------------	-------	--

Hieruit blijkt dat (in aanwezigheid van x_1 en x_3) x_2 in principe overbodig is, echter, de variabele kan niet zomaar weggelaten worden. Deze uitkomst duidt er alleen op dat de voorgaande analyse opnieuw gedaan moet worden, met alleen β_0 , β_1 en β_3 als parameters.

3. Toetsen of $\beta_1 = \beta_2 = \dots = \beta_k = 0$?

Het uitvoeren van deze toets geeft aan of minimaal één van de β_i 's niet gelijk is aan nul, implicerend dat minimaal één factor (x_i) invloed heeft op de te voorspellen variabele. Voor het toetsen wordt de waarde F gebruikt:

$$F = \frac{MSR}{MSE} \quad (4.24)$$

De waarde van F , MSR en MSE worden gegeven in de computeroutput ($MSR^8 = MS$ Regression, $MSE^9 = MS$ Residual en $MS = SS/df$):

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	19,9327	6,6442	11,6982	0,0001
Residual	21	11,9273	0,5680		
Total	24	31,86			

De nulhypothese wordt verworpen wanneer $F \geq c$, met c bepaald door: $P(F_{n-(k+1)}^k > c) = \alpha$ (F -verdeling). De uitkomst van de toets is dan als volgt ($\alpha = 0,05$, $c = 3,06$):

Nulhypothese	<i>F</i>	Conclusie
$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$	11,70	Verwerp H_0 ($11,70 \geq 3,06$): Tenminste één variabele is statistisch van belang bij het voorspellen van de reisduur.

De conclusie kan ook direct bepaald worden aan de hand van de output, de waarde van *significance F* moet kleiner zijn dan α .

4.3.8 Welke (combinatie van) voorspellingen?

De voorspellingen van de genoemde methodes zullen zeer zeker niet gelijk zijn. Door het verschil in de verwerking van data zal bij dezelfde input bij iedere methode een (iets) andere uitkomst gegeven worden. De aansluiting van de verschillende methodes kan beoordeeld worden met verschillende kengetallen. Voor dit onderzoek wordt gekozen voor de mean square error (MSE) en de mean absolute percentage error (MAPE):

$$MSE(\hat{Y}) = \frac{1}{K} \sum_{j=k}^K (\hat{Y}_j - Y_j)^2 \quad (4.25)$$

De MSE is goed bruikbaar voor het vergelijken van voorspellingen van dezelfde schaal (punctualiteit ligt altijd tussen 0 en 100 procent) (Hyndman, 2006). Daarnaast bestraft de MSE grote afwijkingen extra door het kwadrateren van de fout, wat wenselijk is omdat grote afwijkingen erger zijn dan kleine

⁸ MSR: Mean Square Regression

⁹ MSE: Mean Square Error

afwijkingen. Naast de MSE zijn ook andere schaalonafhankelijke kengetallen beschikbaar zoals de veelgebruikte MAPE:

$$MAPE(\hat{Y}) = \frac{1}{K} \sum_{j=k}^K \left(100 \cdot \frac{|\hat{Y}_j - Y_j|}{Y_j} \right) \quad (4.26)$$

De MAPE wordt extra meegenomen omdat dit een eenvoudig te begrijpen maat is voor de afwijking, het geeft direct de gemiddelde procentuele afwijking, terwijl de MSE hierin erg onduidelijk is (wat betekent een MSE van 15?).

Naast de MAPE en de MSE is de bias ook een interessant kengetal voor de kwaliteit van een voorspelling. De bias is gedefinieerd als:

$$bias = \frac{\frac{1}{K} \sum_{j=k}^K \hat{Y}_j}{\frac{1}{K} \sum_{j=k}^K Y_j} \quad (4.27)$$

De bias heeft een ideale score van één, wat betekent dat het gemiddelde van de voorspellingen gelijk is aan het gemiddelde van de waarnemingen.

De methodes zullen allen gebruik maken van een grote dataset ($K > 1000$), welke in tweeën opgesplitst wordt. Het eerste (grote) gedeelte wordt gebruikt om de methodes in te stellen (de trainingsset): bepalen van de parameters voor MLR en ARIMA, leren voor het NN en kandidaten voor k-NN. Het tweede (kleine) gedeelte zal gebruikt worden om de methode te waarderen met behulp van de MSE (de validatieset). Alle methodes zullen hiervoor voorspellingen doen voor deze kandidaten en aan de hand van deze voorspellingen en de waargenomen waarden kan de MSE voor deze methodes uitgerekend worden.

Afhankelijk van de uitkomsten van deze analyse kan gekozen worden voor één van de methodes, of er kan een combinatie gemaakt worden van de verschillende methodes door bijvoorbeeld een gewogen gemiddelde te nemen van de voorspellingen van de verschillende methodes. De gewichten kunnen bepaald worden door de prestaties van de methodes (hoge MSE $\rightarrow$ laag gewicht, lage MSE $\rightarrow$ hoog gewicht).

4.4 Conclusie

In dit hoofdstuk is vastgesteld dat de NS op zoek is naar korte termijnvoorspellingen (vandaag voor morgen) waarbij vooral grote verschillen in de punctualiteit van belang zijn. Het is voor de NS belangrijk om een voorspelling te hebben voor de punctualiteit van morgen zodat:

1. De eigen prestaties beoordeeld kunnen worden;
2. De dienstverlening verhoogd kan worden;
3. De bijsturingcapaciteit aangepast kan worden aan de te verwachten situatie;
4. De klant gewaarschuwd kan worden voor ‘rampzalige’ dagen;
5. De NS in onderhandelingen over de punctualiteit op basis van de onafhankelijke variabelen een voorspelling kan doen over de te realiseren punctualiteit.

Om de punctualiteit te voorspellen zal gezocht worden naar een aantal onafhankelijke variabelen welke door gebruik van de methodes meervoudige lineaire regressie (MLR), k-nearest neighbours (k-dichtstbijzijnde burens, k-NN) en een neurale netwerk (NN) een voorspelling kunnen geven van de punctualiteit. Ook zal de methode ARIMA gebruikt worden voor het maken van een model voor het voorspellen van de punctualiteit. In tegenstelling tot de eerder genoemde methodes gebruikt ARIMA (in de basisvorm) geen onafhankelijke variabelen. Met behulp van correlatie gegevens en de beschikbare statistische testen voor MLR zullen de relevante onafhankelijke variabelen geselecteerd worden. De modellen worden gewaardeerd aan de hand van de MSE (en daarnaast de MAPE en *bias*)

waarna het best passende model gekozen wordt. Eventueel kan gekozen worden voor een combinatie van modellen.

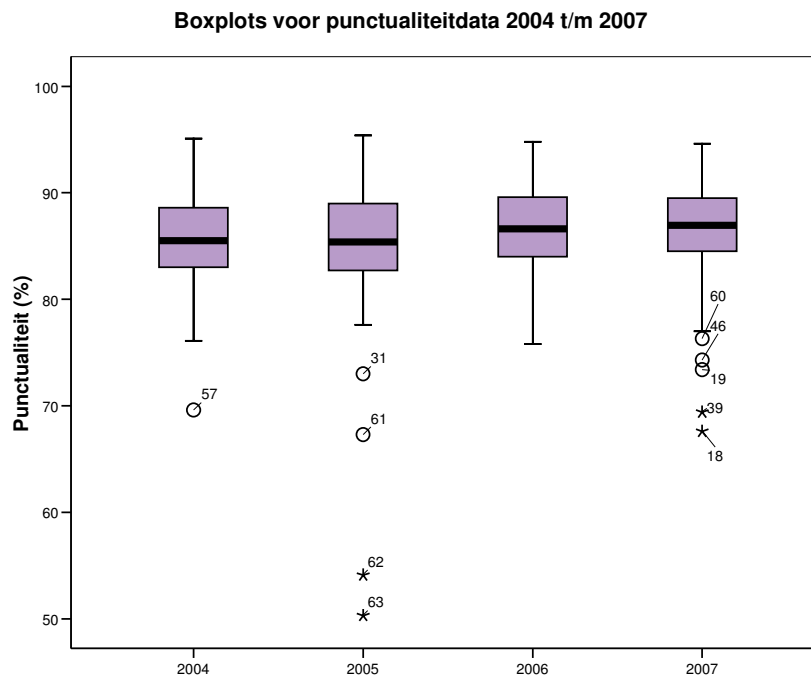
5 Oude data

De NS heeft op 10 december 2006 een nieuwe, ingrijpend gewijzigde dienstregeling ingevoerd. Door de nieuwe dienstregeling rijdt bijna iedere trein op een ander tijdstip en zijn sommige diensten opgeknipt om een verhoging van de punctualiteit te realiseren. De kern van de oude dienstregeling stamt uit 1970 en het aantal treinen dat eind 2006 reed was bijna het dubbele van het aantal treinen in 1970. Daarnaast zijn er met de HSL Zuid, de spoorverdubbeling tussen Utrecht en Amsterdam en de Betuweroute honderden kilometers spoor bijgekomen, reden genoeg dus om het spoorboekje geheel opnieuw op te bouwen.

Een onderliggende reden voor het wijzigen van de dienstregeling was het robuuster maken van de planning met als doel de punctualiteit op het netwerk te verhogen. Een voorbeeld hiervan is het opknippen van de lijn Zwolle – Arnhem – Nijmegen – Roosendaal in de lijnen Zwolle – Nijmegen en Arnhem – Roosendaal. Door de lijn korter te maken rekent een vertraging opgelopen in Zwolle niet door tot Roosendaal, maar tot Nijmegen, waar de trein vervolgens kan keren en zonder vertraging vertrekken. De punctualiteit met de nieuwe dienstregeling zou dus (significant) hoger moeten zijn, en de oude gegevens zouden dus niet bruikbaar meer moeten zijn. Om te onderzoeken of deze hypothese klopt, zal in dit hoofdstuk de data van 1 januari tot 11 maart¹⁰ van de jaren 2004, 2005, 2006 (oude dienstregeling) en 2007 (nieuwe dienstregeling) onderzocht worden.

5.1 Karakteristieken van de data

Voor het bestuderen van de data zijn de punctualiteitscijfers van 1 januari tot 11 maart voor ieder jaar genummerd van 1 tot en met 70 (2005, 2006, 2007) of 71 (2004, schrikkeljaar). Vervolgens is van de verschillende jaren een boxplot gemaakt. De data die vol buiten het ‘normale’ gebied vallen (aangegeven met een *), kunnen bestempeld worden als uitzonderingsgevallen (zie ook onder) en worden uit de dataset gehaald.



Figuur 5.1 Boxplots voor de data van 2004 t/m 2007

¹⁰ Op het moment van schrijven was niet meer data beschikbaar

Jaar	Case	Punctualiteit en oorzaak
2005	62	54,%, 3 maart, extreme kou, sneeuw
	63	50,4%, 4 maart, extreme kou, sneeuw
2007	18	67,6%, 18 januari, storm
	39	69,4%, 8 februari, sneeuw

Tabel 5.1 Gegevens van de verwijderde cases

5.2 Variantieanalyse punctualiteit

De boxplots lijken te wijzen op een sterke overeenkomst tussen de punctualiteitscijfers van 2007 en 2006, ook de cijfers van 2005 en 2004 lijken niet veel te verschillen. Om dit objectief te kunnen vaststellen wordt de gelijkheid van de data getoetst met de k -steekproevenstoets. Voorwaarde voor uit kunnen uitvoeren van deze toets is dat de waarnemingen onafhankelijk zijn van elkaar, de waarnemingen normaal zijn verdeeld en dat de varianties in de verschillende steekproeven gelijk zijn. Na de verwijdering van de genoemde cases in paragraaf 5.1 is dat het geval. De hypothesen voor deze toets luiden vereenvoudigd:

H_0 : De steekproeven zijn identiek; er is geen verschil tussen de waarnemingen uit de verschillende steekproeven.

H_1 : De steekproeven zijn significant anders; er is *wel* verschil tussen de waarnemingen uit de verschillende steekproeven.

	N	Gemiddelde	Std. Deviatie
2004	71	85,49	4,58
2005	68	85,64	4,90
2006	70	86,58	4,67
2007	68	86,79	4,42
Totaal	277	86,12	4,66

Tabel 5.2 Samenvatting van de data voor 2004 t/m 2007

De ANOVA (ANalysis Of VAriance) tabel voor de data luidt als volgt:

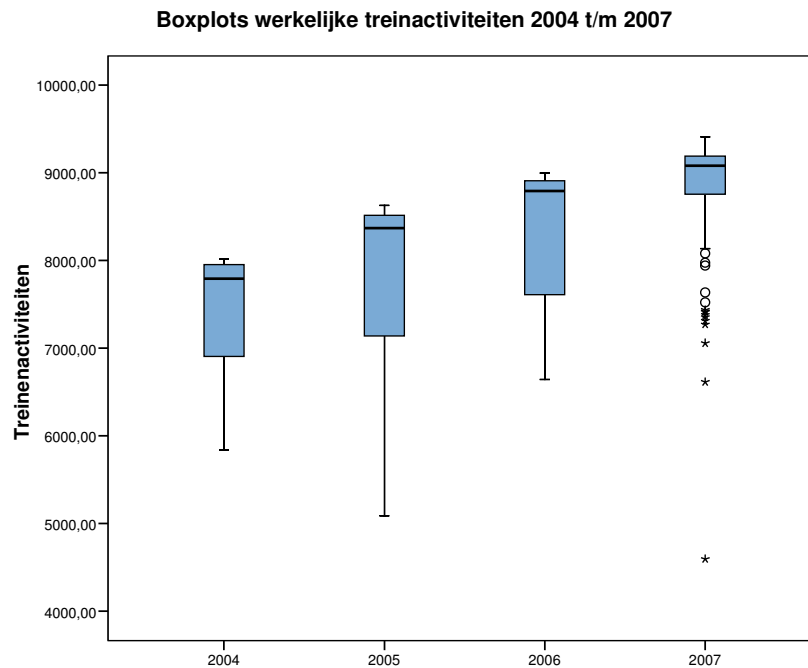
	Sum of Squares	df	Mean Square	F	Significantie
Jaar	89,43	3	29,81	1,38	0,25
Error	5891,96	273	21,58		
Totaal	5981,40	276			

Tabel 5.3 ANOVA tabel van de data voor 2005 t/m 2007

Aan de significantie is af te lezen ($0,25 > 0,05$) dat de nulhypothese (H_0) niet verworpen wordt. De conclusie is dus dat er geen verschil bestaat tussen de waarnemingen van 2004, 2005, 2006 en 2007.

5.3 Variantieanalyse treinactiviteiten

In Figuur 5.2 zijn boxplots opgenomen voor de treinactiviteiten tijdens de periode 1 januari – 20 maart voor de jaren 2004 t/m 2007. Zoals gezien kan worden is het aantal treinactiviteiten toegenomen in deze periode. Om een goede vergelijking te kunnen maken wordt een gecorrigeerde waarde van de treinactiviteiten gebruikt, deze gecorrigeerde waarde neemt mee dat het aantal treinactiviteiten afgestemd wordt op de reizigersaantallen. De reizigersgroei wordt vermeld in Tabel 5.4. De groei wordt gradueel meegenomen over de periode 1 januari tot 31 december, daarnaast wordt onderscheid gemaakt in de reizigersaantallen per dag. De resulterende waarde heeft de eenheid treinen/reiziger.



Figuur 5.2 Boxplots voor de werkelijke treinenactiviteiten

Jaar	Reizigers
2004	90,0
2005	91,6
2006	95,2
2007	100,0

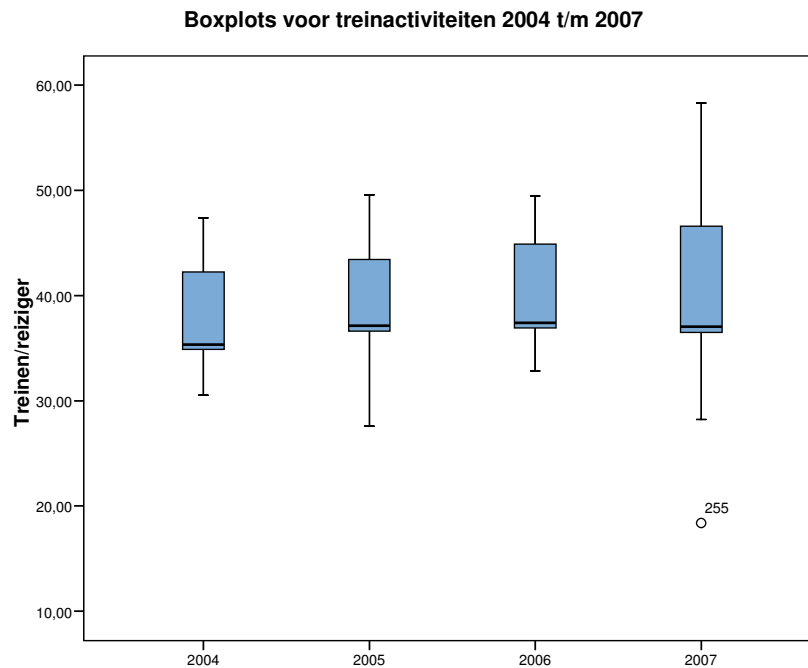
Tabel 5.4 Reizigersgroei

Basisvoorwaarde voor de variantieanalyse en de k -steekproevenstoets is dat de data normaal verdeeld zijn. Voor de gecorrigeerde treinen aantallen geldt dit niet: de Shapiro-Wilk toets geeft voor alle jaren een significantie van 0,000 ($< 0,05$), reden om de nul hypothese van de toets te accepteren: de verdeling van de treinactiviteiten is niet normaal verdeeld. Ook het apart bekijken van weekenden en weekdagen levert onvoldoende bewijs voor normaliteit.

	Jaar	Statistic	df	Sig.
Treinactiviteiten (gecorrigeerd)	2004	0,762	79	0,000
	2005	0,805	79	0,000
	2006	0,772	79	0,000
	2007	0,805	79	0,000

Tabel 5.5 Toets voor normaliteit (Shapiro-Wilk)

Wat rest is het vergelijken van de boxplots voor de treinactiviteiten. In het onderstaande figuur zijn deze boxplots gegeven. Op basis van dit figuur (en de kennis dat de punctualiteit wél vergelijkbaar is voor 2004 t/m 2007) kan gesteld worden dat alle data meegenomen kunnen worden, de boxplots laten een ongeveer gelijke verdeling zien voor de verschillende jaren. Jaar 2007 laat wel een hogere hoogste waarde zien dan de andere jaren, maar dit wordt niet al bezwaarlijk gezien.



Figuur 5.3 Boxplots voor treinactiviteiten 2004 t/m 2007

5.4 Conclusie

Aan de hand van de uitgevoerde statistische onderzoeken kan gesteld worden dat de gegevens van 2004 t/m 2006 gebruikt kunnen worden voor voorspellingen over de nieuwe dienstregeling. Aangezien de reizigersgroei gebruikt kan worden als input voor de modellen is het niet nodig om de treinactiviteiten ook aan te passen in de data.

6 Invloeden op punctualiteit

De punctualiteit wordt, zoals verwacht door personeel binnen BRV, beïnvloed door verschillende externe en interne variabelen. De verwachte relaties zijn echter allemaal gebaseerd op praktijkervaringen. Niet meer dan ‘anekdotisch’ bewijs ligt ten grondslag aan de verwachte relaties. Dit hoofdstuk heeft als doel om dit soort beweringen (statistisch) te bewijzen, of aan te tonen dat de relatie niet bestaat. Het daadwerkelijke modelleren van de relaties in een voorspellingsmodel, komt in hoofdstuk 7 aan de orde.

De opbouw van de beschouwing van de verschillende variabelen is steeds hetzelfde. Allereerst wordt gekeken wat de verwachte relatie is, vervolgens wordt gekeken welke gegevens beschikbaar zijn, en hoe deze relatie zich verhoudt tot de verwachte relatie. De relatie zal als aanwezig bestempeld worden wanneer een plot van de te testen variabele tegen de punctualiteit een verband laat zien en de Pearson-correlatie voldoende hoog is ($> 0,4$ of $< -0,4$).

Pearson-correlatie

Voor het testen van de relatie (*correlatie*) tussen een onafhankelijke variabele en de punctualiteit zal de Pearson-correlatie gebruikt worden. De Pearson-correlatie is een waarde tussen -1 en 1 die de mate van lineaire samenhang tussen twee variabelen aangeeft. Een waarde van 1 betekent een perfecte samenhang (wat wil zeggen dat als de onafhankelijke variabele met 1 toeneemt, de punctualiteit hetzelfde doet).

Een positieve uitkomst (de Pearson-correlatie heeft een voldoende positieve of negatieve waarde, dus ‘ver’ van 0) impliceert niet direct dat er een relatie is tussen de onafhankelijke en de afhankelijke variabele. Gedurende de periode 1960-1980 nam bijvoorbeeld het aantal geboortes in Duitsland sterk af, evenals het aantal ooeivaars. De correlatie tussen deze twee variabelen zal in die tijd hoog geweest zijn, maar in werkelijkheid is er natuurlijk geen enkel verband. Het is daarom van belang om naast de waarde van de correlatiecoëfficiënt te kijken naar een rationele verklaring van de relatie, ontbreekt deze dan is de relatie waarschijnlijk niet aanwezig.

6.1 Externe variabelen

Onder de externe variabelen vallen alle onafhankelijke variabelen waar de NS zelf geen invloed op kan uitoefenen. Een belangrijke externe o.v. is het weer, daarnaast zal onder andere gekeken worden naar de invloed van suïcides op het spoor en reizigersaantallen.

6.1.1 Weer

Van de weersinvloeden is de herfst wel het bekendst: ieder jaar daalt de punctualiteit in de herfst door bladval. In deze paragraaf zal deze en andere relaties bekeken worden vanuit statistisch perspectief.

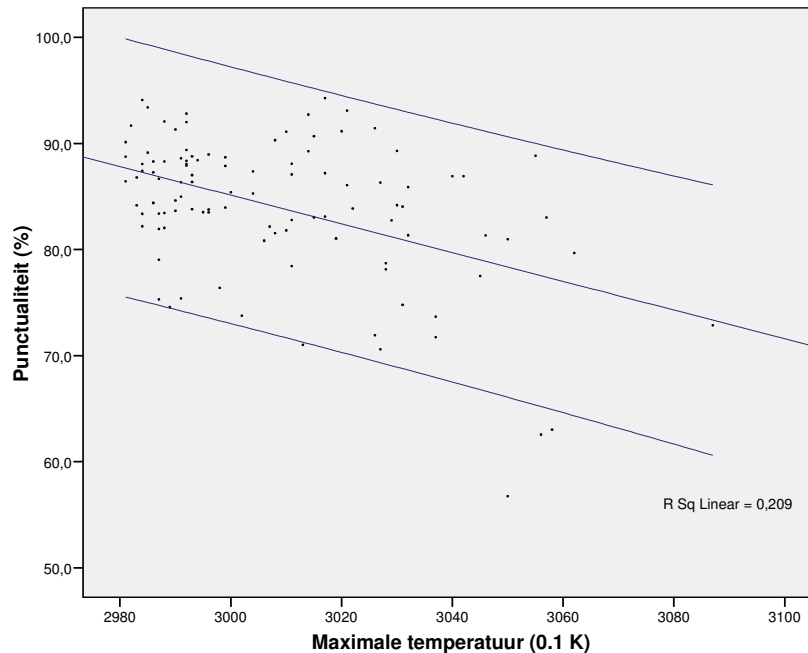
6.1.1.1 Hoge temperatuur

Verwachting: Een hogere temperatuur leidt tot spoorspatting¹¹ waardoor de punctualiteit daalt.

Data: Het KNMI stelt op dagelijkse basis weerdata beschikbaar. De relevante waarneming voor deze bewering is de maximum temperatuur gemeten op een dag, aangezien dit de temperatuur is die voor de verstoring zorgt.

Relatie: Omdat alleen hoge temperaturen een rol spelen – bij lage temperaturen kan heel goed een andere factor dan spoorspatting een rol gaan spelen – is uit de data een selectie gemaakt van alle dagen waar de maximum temperatuur boven de 25 graden Celsius lag. De precieze grens waar spoorspatting optreedt is niet duidelijk: langdurige zon bij 25 graden kan al zorgen voor extreem warme spoorstaven die vervolgens voldoende uitzetten om krom te trekken. In het onderstaande figuur is de relatie tussen maximale temperatuur en punctualiteit aangegeven. De middelste blauwe lijn geeft een fitvergelijking aan, de bovenste en onderste blauwe lijn markeren de grens waarbinnen 95% van de punten ligt. De (Pearson-)correlatie van maximale temperatuur-punctualiteit is -0,457, wat significant is met een betrouwbaarheid van 99%.

¹¹ Het kromtrekken van sporen waardoor treinen langzamer of helemaal niet kunnen rijden.



Figuur 6.1 Relatie maximale temperatuur – Punctualiteit.

Conclusie: Er bestaat een relatie tussen hoge temperaturen en lagere punctualiteit. Op een 99% betrouwbaarheidsniveau kan gesteld worden dat de punctualiteit negatief beïnvloed wordt door hoge temperaturen (mits de temperatuur hoger is dan 25 graden). Overigens is de tussenrelatie hoge temperatuur – spoorspatting – punctualiteit niet aangetoond: alleen de relatie hoge temperatuur – punctualiteit kan bewezen worden met deze data.

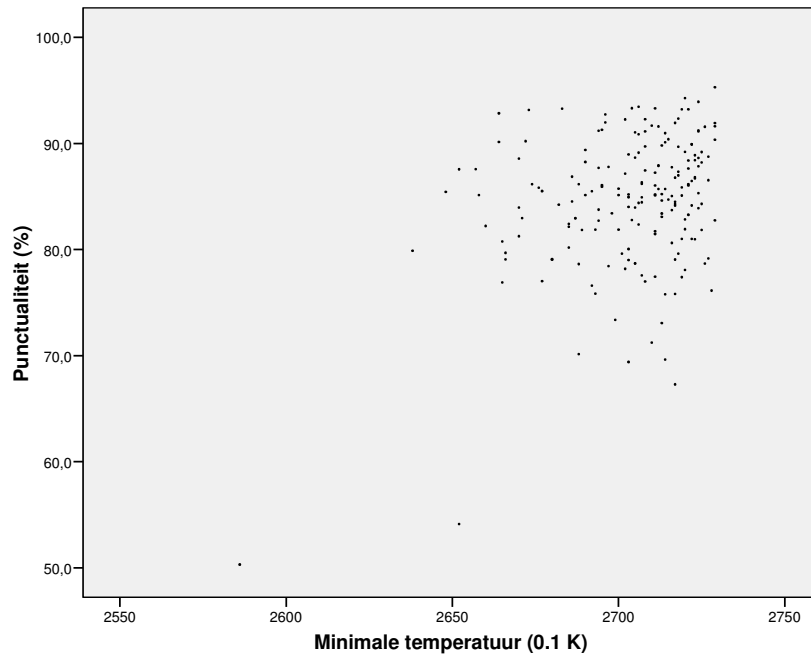
6.1.1.2 Lage temperatuur

Verwachting: Lage temperatuur zorgt voor knappende bovenleidingen en vastgevroren wissels waardoor de punctualiteit daalt.

Data: Voor deze relatie wordt de minimum temperatuur per dag gebruikt, eveneens van het KNMI.

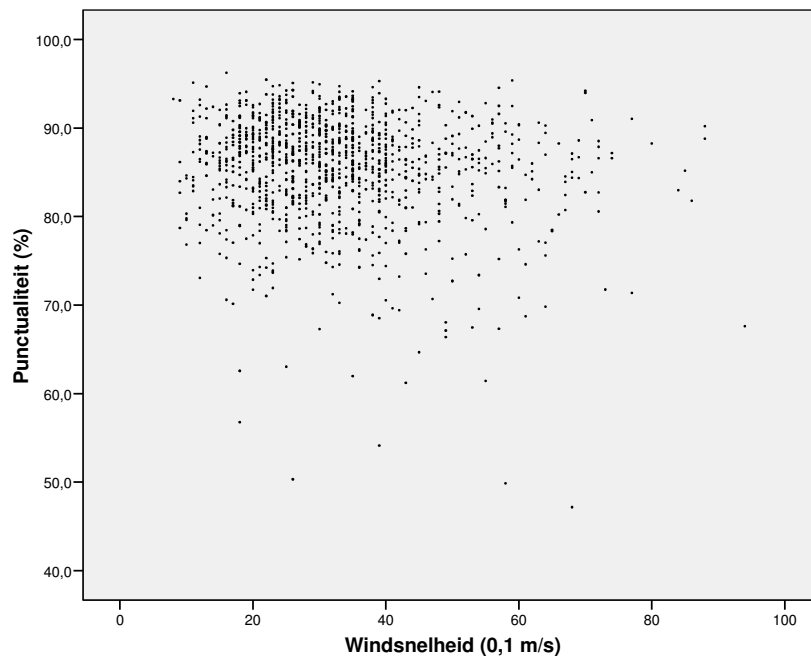
Relatie: Evenals bij de hoge temperaturen geldt dat deze relatie alleen van belang is op een bepaald gebied: temperaturen onder 0 graden Celsius worden geacht problemen te veroorzaken. In het onderstaande figuur is voor het relevante gebied de relatie minimale temperatuur-punctualiteit weergegeven. Zoals uit het onderstaande figuur blijkt is de relatie erg vaag: bij temperaturen rond 265 Kelvin (-8° Celsius) komen zowel hele lage als hele hoge waarden voor. Alleen een uitschieter rond de 258 Kelvin (-15° Celsius) en 265 Kelvin (-8 ° Celsius) geven wel een zeer lage punctualiteit, dit werd echter veroorzaakt door hevige sneeuwval.

Het gebrek aan een relatie is ook te verklaren: ProRail verwarmt wissels zodra de temperatuur onder de 0° Celsius komt, en het knappen van bovenleidingen door een lage temperatuur komt zelden voor.



Figuur 6.2 Relatie minimale temperatuur – Punctualiteit.

Conclusie: Er bestaat geen relatie tussen de minimale temperatuur en de punctualiteit.



Figuur 6.3 Er is geen relatie tussen windsnelheid en punctualiteit.

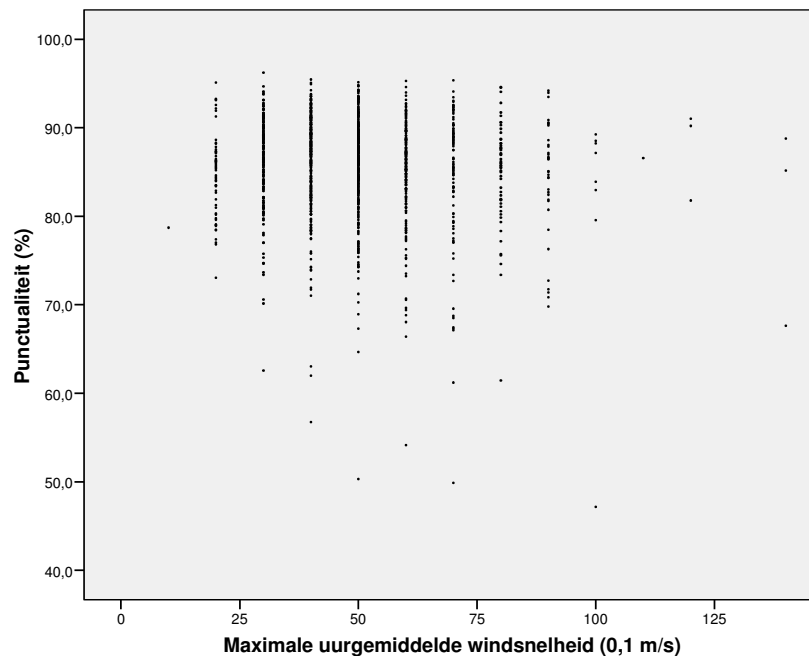
6.1.1.3 Windsnelheid

Verwachting: De punctualiteit daalt wanneer de windsnelheid toeneemt.

Data: Het KNMI publiceert de gemiddelde windsnelheid, maximale uurgemiddelde windsnelheid en maximale windstoot.

Relatie: De verwachting is gebaseerd op het ontstaan van calamiteiten bij hogere windsnelheden. Reden hiervoor is dat bij hogere windsnelheden meer op het spoor waait en meer bovenleidingen

knappen. Uit figuren 6.3 en 6.4 blijkt echter dat er nauwelijks een relatie bestaat tussen de (maximale uurgemiddelde) windsnelheid en de punctualiteit.



Figuur 6.4 Ook de maximale uurgemiddelde windsnelheid lijkt geen invloed te hebben.

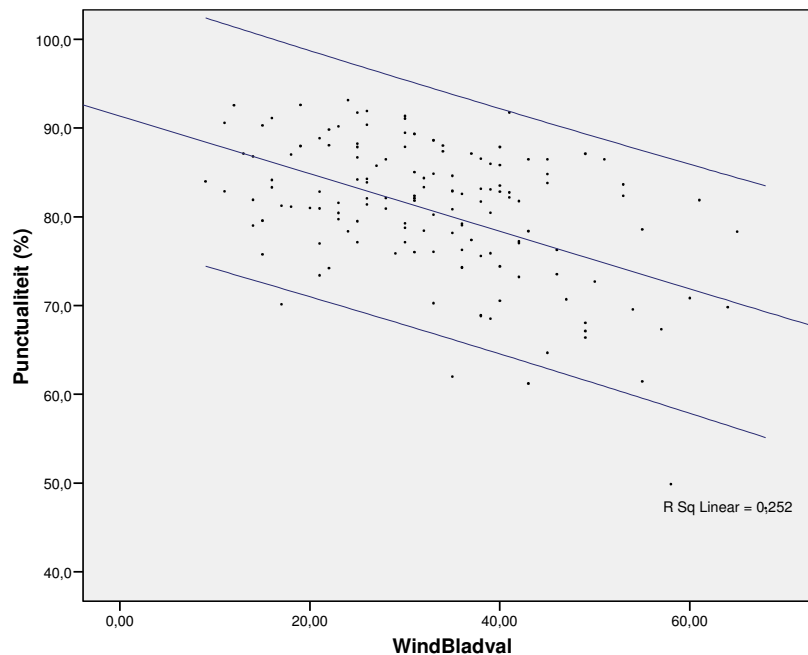
Conclusie: Er is geen relatie tussen de (maximale uurgemiddelde) windsnelheid en de punctualiteit.

6.1.1.4 Bladval

Verwachting: Bladval in combinatie met wind zorgt voor een lagere punctualiteit.

Data: Voor deze verwachting wordt een periode van bladval gebruikt van 15 oktober tot 31 november, en de gemiddelde windsnelheid per dag gepubliceerd door het KNMI. De te testen variabele is een samenstelling tussen de twee componenten: als er geen bladval is, dan is de waarde van de te testen variabele 0. Is er wel bladval, dan is de waarde van de te testen variabele de gemiddelde windsnelheid.

Relatie: Voor deze relatie worden alle gegevens gebruikt waar de 'windbladval' variabele groter dan 0 is (dus alle dagen waarop bladval verwacht werd). In het onderstaande figuur is duidelijk te zien dat er een relatie bestaat tussen de mate van bladval en de punctualiteit: naarmate de windsnelheid toeneemt bij bladval, daalt de punctualiteit. De Pearson correlatie voor deze relatie is -0,502, wat significant is op een betrouwbaarheidsniveau van 99%.



Figuur 6.5 Relatie WindBladval - Punctualiteit

Conclusie: Er bestaat een significante relatie tussen de variabele ‘windbladval’ en de punctualiteit.

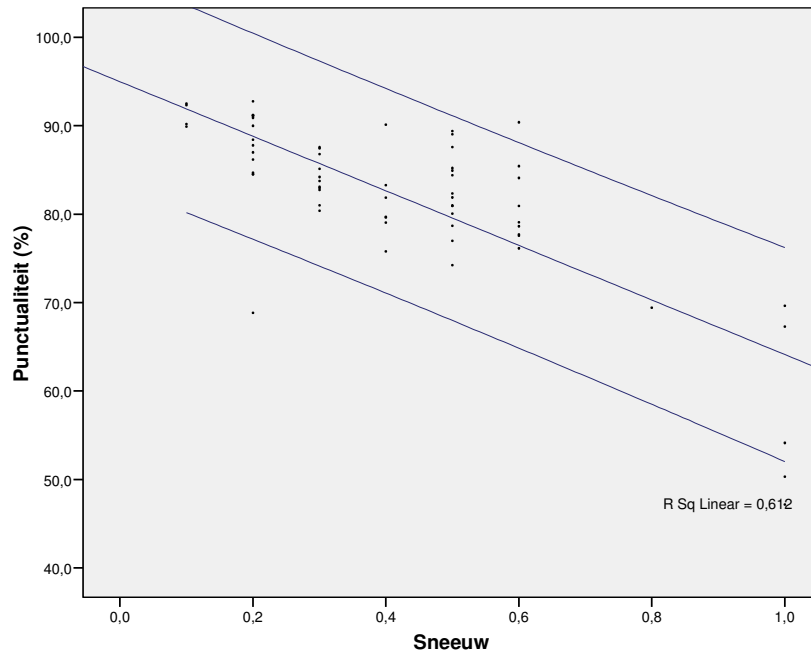
6.1.1.5 Sneeuwval

Verwachting: Bij sneeuwval daalt de punctualiteit.

Data: Sneeuwval wordt niet formeel bijgehouden door het KNMI, daarom is uit verschillende bronnen op internet informatie over sneeuwdagen opgezocht. Om de sneeuwval te kunnen beoordelen is een schaal gemaakt van 0 tot 1 en aan de hand van de beschrijvingen is een getal toegewezen aan de (sneeuw)dagen. De schaal is bewust niet gebaseerd op sneeuwhoogtes, aangezien op voorhand niet voorspeld is wat de sneeuwhoogte zal worden. De gebruikte terminologie komt overeen met weerberichten.

Sneeuwschaal	
0.0	Geen sneeuw
0.1	Natte sneeuw
0.2	Lichte sneeuw
0.3	
0.4	
0.5	“Normale” sneeuw
0.6	
0.7	
0.8	
0.9	
1.0	Heftige sneeuw(storm)

Relatie: Wederom is (nu voor de duidelijkheid) een selectie gemaakt uit de data. Alle data waar de sneeuwvariabele 0 is, zijn niet meegenomen. Er is een duidelijk verband tussen sneeuwval en punctualiteit te zien in het onderstaande figuur. De Pearson Correlatie van -0,782 is significant met een betrouwbaarheid van 99%.



Figuur 6.6 Relatie sneeuwval - Punctualiteit

Conclusie: Er bestaat een significante relatie tussen sneeuwval en de punctualiteit.

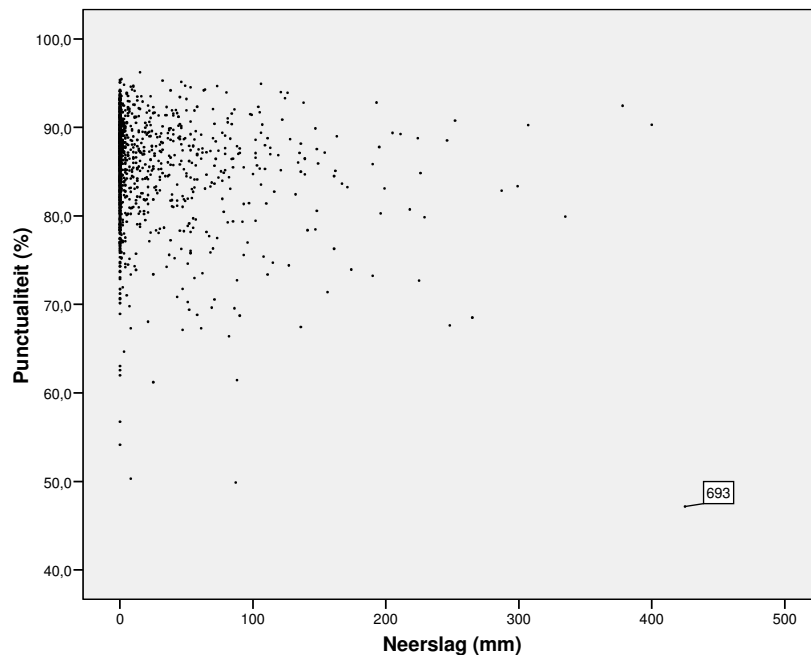
6.1.1.6 Neerslag(hoeveelheid)

Verwachting: Neerslag heeft geen invloed op de punctualiteit

Data: Neerslaggegevens van het KNMI.

Relatie: Treinen hebben in principe geen last van regen, tenzij de temperatuur erg laag is (sneeuw, zie eerder), er bladval is (zie eerder) of als er ijzel is (zie lage temperatuur). In het onderstaande figuur wordt het gebrek aan een relatie bevestigd. Het opvallende punt (aangegeven met 693) heeft een lage punctualiteit door sneeuwval.

Conclusie: Er is inderdaad geen relatie tussen neerslag(hoeveelheid) en punctualiteit.

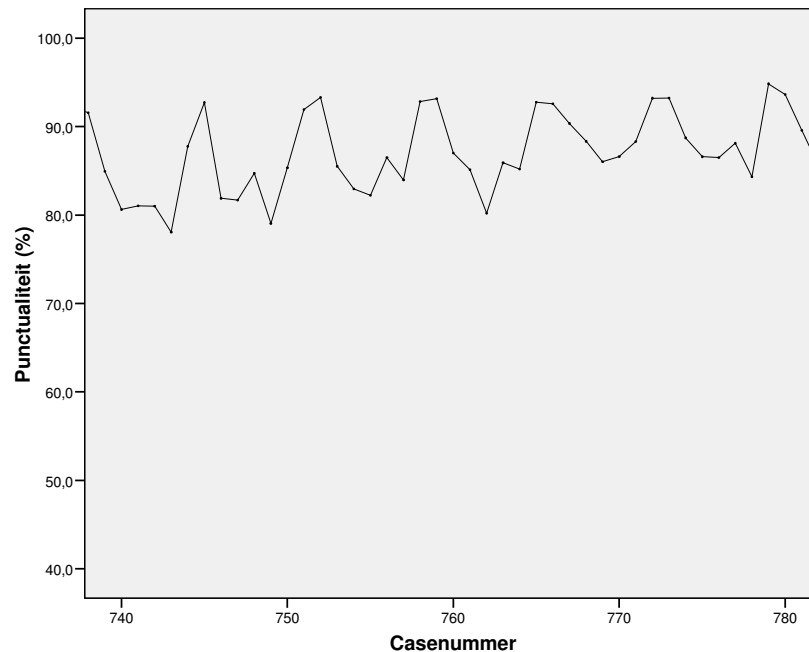


Figuur 6.7 Er is geen relatie tussen de neerslaghoeveelheid en de punctualiteit.

6.1.2 Overige externe invloeden

Verwachting: In het weekend ligt de punctualiteit hoger dan op werkdagen.

Relatie: In het onderstaande figuur is voor een willekeurige groep cases de punctualiteit geplot (op volgorde van het casenummer). Bij de weekenden is 'za' voor zaterdag en 'zo' voor zondag geplaatst. Zoals gezien kan worden is er een duidelijk verschil tussen de weekenden en de werkdagen. De punctualiteit in de weekenden is echter niet hoger *omdat het weekend is*. Relaties die daar invloed op kunnen hebben zijn onder andere het aantal reizigers, het aantal treinactiviteiten, of een andere variabele.



Figuur 6.8 Weekend, in het figuur is een duidelijk cyclisch patroon te vinden over 7 dagen.

Conclusie: De punctualiteit in de weekenden is over het algemeen hoger. De hogere punctualiteit wordt echter niet veroorzaakt doordat het weekend is, maar door een andere variabele (bijvoorbeeld reizigersaantallen).

6.1.3 Reizigersaantallen

Verwachting: Het aantal reizigers in het weekend is lager en daardoor stijgt de punctualiteit.

Relatie: Het wachten van treinen totdat alle reizigers in- en uitgestapt zijn, leidt tot vertragingen. Met een lager reizigersaantal is de kans veel kleiner dat een grote groep reizigers moet instappen bij één trein. Het lagere reizigersaantal in de weekenden wordt voornamelijk veroorzaakt door het ontbreken van de forenzen.

Data: Reizigers worden niet dagelijks nauwkeurig geteld. Conducteurs maken ruwe schattingen van het aantal reizigers dat aanwezig is in een trein tussen twee stations. Deze schattingen zijn over het algemeen gebaseerd op de bezetting in een bak, vermenigvuldigd met het aantal bakken van de trein. De conducteurstellingen worden per weekdag opgeslagen. Algemene vakantiedagen (nieuwjaarsdag, koninginndag, Pasen, Hemelvaart, Pinksteren, Kerst) worden behandeld als zondagen (155), semi-vakantiedagen (bevrijdingsdag, goede vrijdag, oudejaarsdag) worden behandeld als dagen met beperkte hoeveelheid reizigers (200).

Reizigersverdeling weekdays						
Maandag	Dinsdag	Woensdag	Donderdag	Vrijdag	Zaterdag	Zondag
242	250	250	251	252	166	155

Tabel 6.1 Reizigersverdeling per weekday

Naast deze tellingen zijn er ook gegevens bekend over reizigerskilometers (de totaal afgelegde afstand van reizigers over het spoor in een maand, over de periode januari 2004 tot maart 2007), welke per maand totaal opgeslagen worden. De reizigerskilometers liggen rond de 10^9 kilometer per maand. Door de ruwe tellingen van conducteurs en de reizigerskilometers met elkaar te combineren¹² wordt een relatief betrouwbare maat voor het aantal reizigers gecreëerd. De gecreëerde variabele heet kortweg ‘reizigers’.

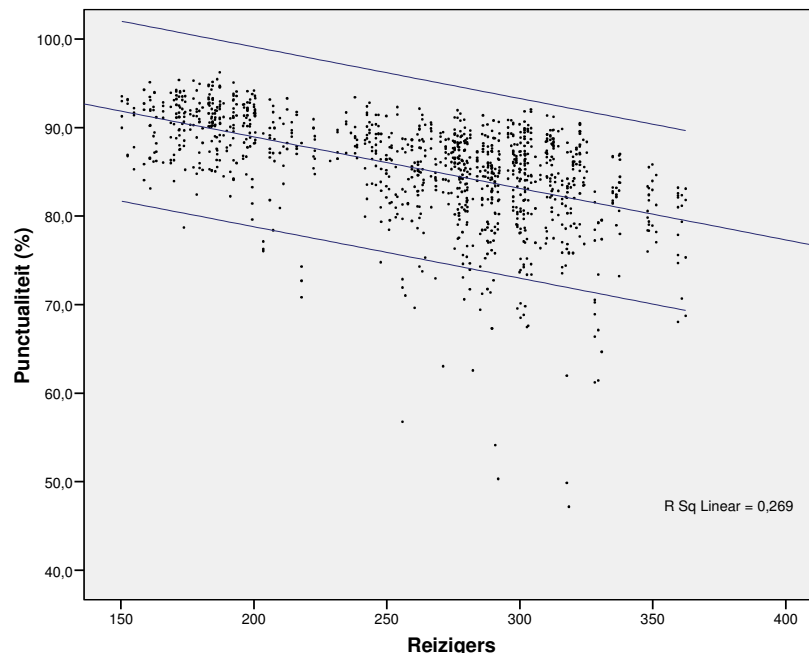
Wanneer de relatie bekeken wordt tussen reizigers en de punctualiteit ontstaat het beeld als weergegeven in figuur 6.9 op de volgende pagina. De lijnen geven de rechte lijnige interpolatie en de 95% betrouwbaarheidsintervallen voor deze lijn aan. De relatie tussen het aantal reizigers en de punctualiteit is duidelijk negatief: meer reizigers zorgen voor een dalende punctualiteit. De Pearson-correlatie is -0,519 en significant op met een betrouwbaarheid van 99%.

Conclusie: Er is een duidelijke relatie tussen het aantal reizigers en de punctualiteit.

6.1.4 Vakanties

Verwachting: De punctualiteit is in (school)vakanties en op feestdagen hoger.

Relatie: Dat de punctualiteit in (school)vakanties en op feestdagen hoger is kan teruggevonden worden in de data. Alleen parallel aan de weekenden geldt voor vakanties dat de punctualiteit niet hoger is *omdat* het vakantie is, de werkelijke achterliggende redenen zijn een lager reizigersaantal, aangepaste dienstregelingen, ander reizigerspatroon (beter verspreid over de dag), etc. Het meenemen van een variabele ‘vakantie’ is dus niet gewenst: andere onafhankelijke variabelen moeten het vakantie-effect vertegenwoordigen.

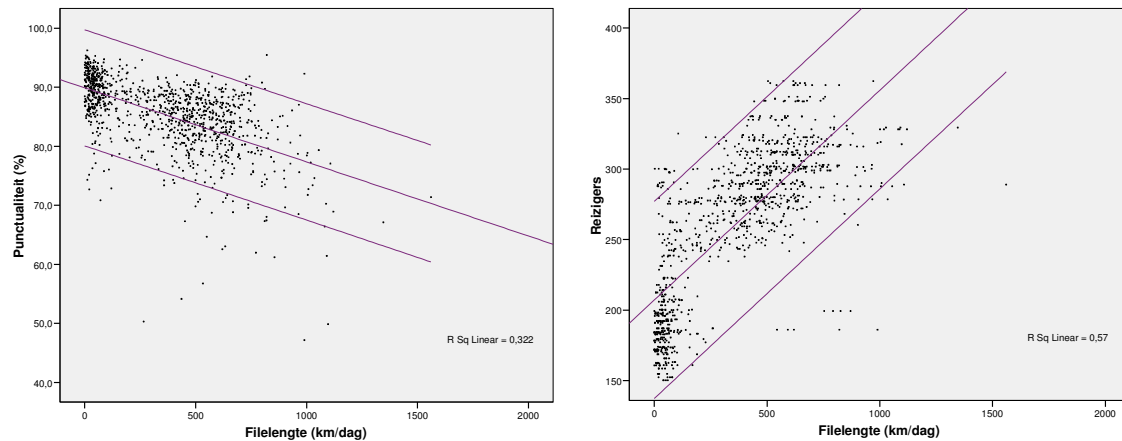


Figuur 6.9 Relatie aantal reizigers-punctualiteit

¹²
$$\text{Reizigers} = \frac{\text{reizigerskilometers} \cdot \text{conducteurstellingen}}{10^9}$$

6.1.5 Files

Het ligt niet in de lijn van verwachting dat files een directe invloed hebben op de punctualiteit: het mag duidelijk zijn dat files treinen niet belemmeren. Echter, uit het onderstaande figuur blijkt wel dat files en de punctualiteit een gelijk verloop vertonen. Met een Pearson-correlatie van $-0,567$ is deze relatie ook significant. Dit betekent echter nog niet dat de filelengte opgenomen moet worden als onafhankelijke variabele: er is sprake van een *statistisch* verband, maar niet direct van een *oorzakelijk* verband.

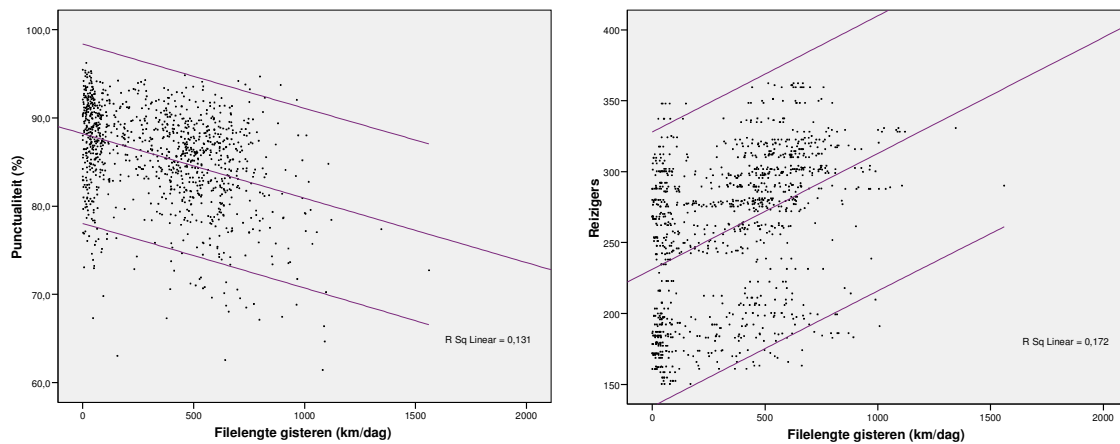


Figuur 6.10 Files, punctualiteit en reizigersaantallen

Nader onderzoek laat een zeer sterk (correlatie 0,755) verband zien tussen het reizigersaantal, wat zou kunnen betekenen dat zowel de NS als de snelwegen te leiden hebben onder grote reizigersaantallen: files worden langer en de punctualiteit daalt. (Filed data: RWS-AVV, 2007).

Een andere verklaring (eveneens via reizigersaantallen) voor een correlatie tussen filelengte en punctualiteit zou de 'keuzereiziger' kunnen zijn. Een keuzereiziger (snelle beslisser) kijkt op een willekeurig moment naar de filelengte en de situatie op het spoorwegnet en kiest vervolgens welk type vervoer er gekozen wordt: de trein of een alternatief (auto, fiets, bus, etc.). Op het moment dat de keuzereiziger waarneemt dat er file staat, zal de reiziger een treinkaartje kopen en de trein als vervoersmiddel kiezen. Neemt de keuzereiziger geen file waar, of juist een verstoring op het traject dat gereisd moet worden bij de NS, dan koopt de reiziger geen kaartje en kiest de auto als vervoersmiddel. Als deze relatie klopt, zou een relatie gezien moeten worden tussen de filelengte en de kaartverkoop op een dag. Daarnaast zou deze relatie wellicht nog sterker gelden in het verband vandaag file – morgen met de trein, of vandaag lange reistijd/lage punctualiteit – morgen met de auto. Er zijn echter onvoldoende gegevens beschikbaar om deze relatie(s) vast te kunnen stellen.

Voor de laatst genoemde mogelijke relatie zouden de correlaties toe moeten nemen voor vertraagde filed data, het ligt immers in de lijn van de verwachting dat meer forenzen een afweging in een periode van 12 tot 24 uur (de vandaag-morgen keuzereiziger) kunnen maken dan binnen 5 minuten (de snelle beslisser). Om deze relatie te testen zijn de onderstaande figuren interessant. De figuren laten wederom een relatie zien tussen de files gisteren – de punctualiteit vandaag (correlatie: $-0,361$) en files gisteren – reizigers vandaag (correlatie: $0,415$). Aangezien de correlaties afnemen, lijkt de eerder genoemde hypothese dat zowel files als punctualiteit te leiden hebben onder reizigersdruk een betere verklaring te geven dan de keuzereiziger. De relatief hoge correlatie wordt veroorzaakt door het weekpatroon: de files op werkdagen zijn relatief goed met elkaar te vergelijken, net als de punctualiteit en de reizigersaantallen.



Figuur 6.11 Files gisteren, punctualiteit en reizigersaantallen

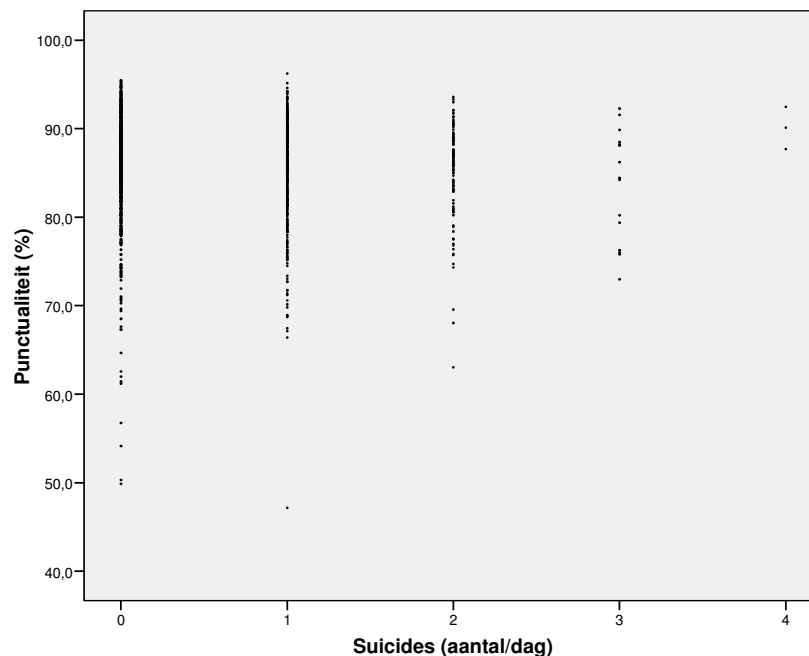
6.1.6 Suïcides

Verwachting: De punctualiteit daalt als het aantal suïcides toeneemt.

Data: Voor deze relatie wordt de registratie van suïcides op het spoor van de Inspectie Verkeer en Waterstaat gebruikt.

Relatie: Een (poging tot) suïcide op het spoor zorgt voor een ontregeling van de treinenloop op het traject waar de suïcide plaatsgevonden heeft. De ontregeling zorgt voor vertragingen en uitgevallen treinen aangezien de trein van de suïcide weggereden en gereinigd moet worden. Daarnaast nemen opruimwerkzaamheden tijd in beslag.

Uit het onderstaande figuur blijkt dat de impact van suïcides minimaal is: zelfs bij hoge suïcidecijfers blijft de punctualiteit redelijk hoog. Een verklaring hiervoor kan zijn dat suïcides vooral zorgen voor uitgevallen treinen, er kan immers geen enkele trein meer rijden op het traject waar de suïcide plaatsgevonden heeft. Wanneer de calamiteit opgelost is, kan de treinenloop grotendeels weer gewoon hervat worden.



Figuur 6.12 Relatie suïcides - punctualiteit

6.2 Interne variabelen

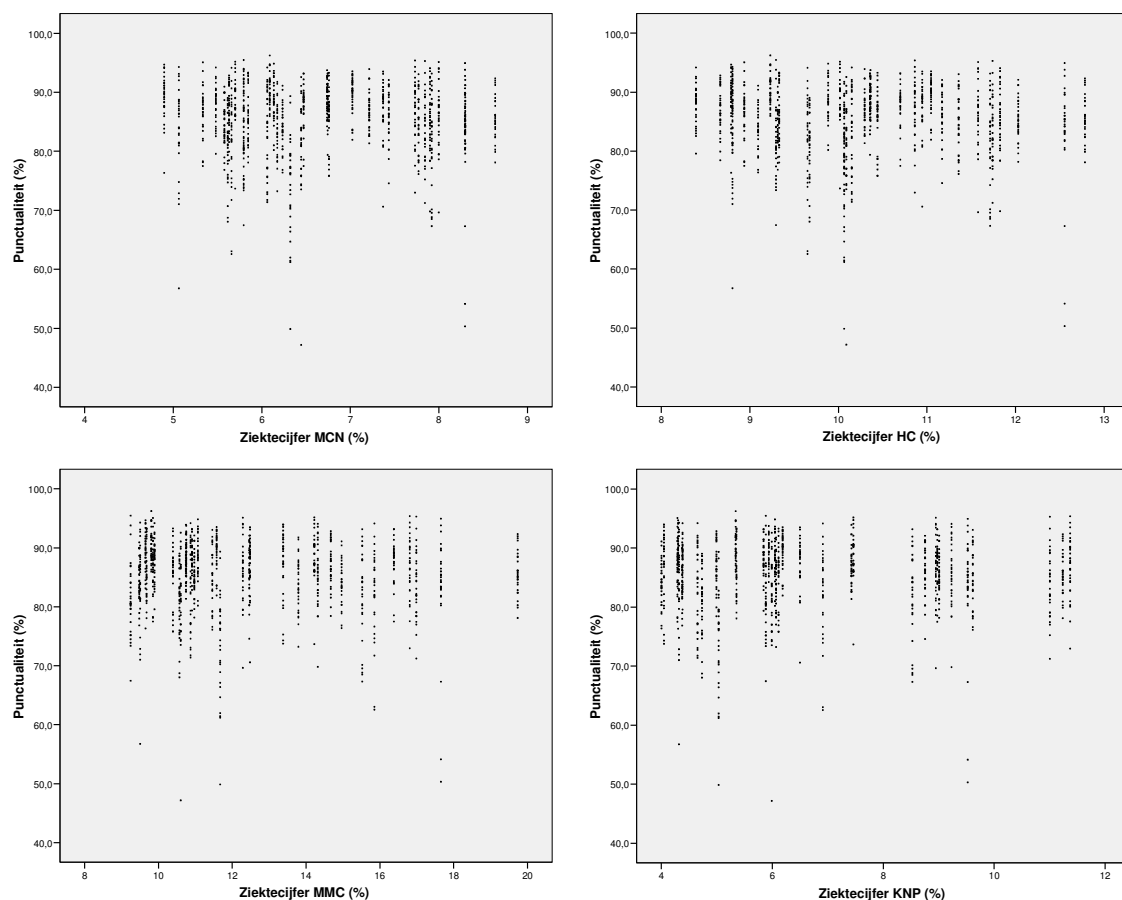
De interne variabelen zijn de o.v. waar de NS zelf wel invloed op heeft, of welke afhankelijk zijn van de prestaties van de NS. Variabelen die hiervoor bekeken zullen worden zijn onder andere ziektecijfers, vakantiedruk en de behaalde punctualiteit in het verleden.

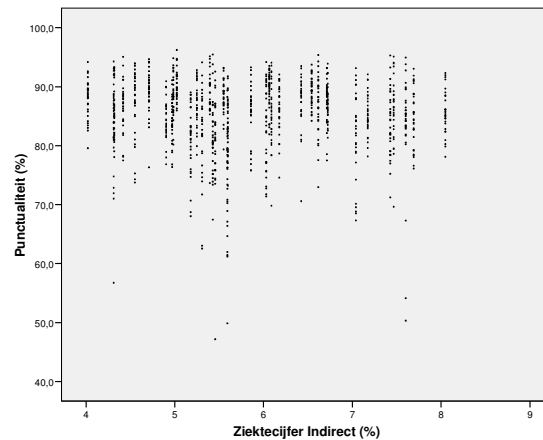
6.2.1 Ziektecijfers personeel

Verwachting: Hogere ziektecijfers leiden tot een lagere punctualiteit

Data: Het bedrijfseconomisch bureau heeft data geleverd over het ziektepercentage van machinisten (MCN), (hoofd)conducteurs (HC), medewerkers mobiele controle (MMC), knooppunt personeel (KNP) en indirect personeel.

Relatie: Een lagere beschikbaarheid van personeel zorgt voor een hogere druk op het personeel dat niet ziek is. Door deze druk daalt de motivatie ('ik kan geen vakantie nemen omdat er geen ruimte is in de planning'). Uit de onderstaande diagrammen blijkt dat deze relatie niet bestaat: uit geen enkel diagram volgt een relatie tussen het ziektepercentage en de punctualiteit.





Figuur 6.13 Relaties ziektecijfer – punctualiteit

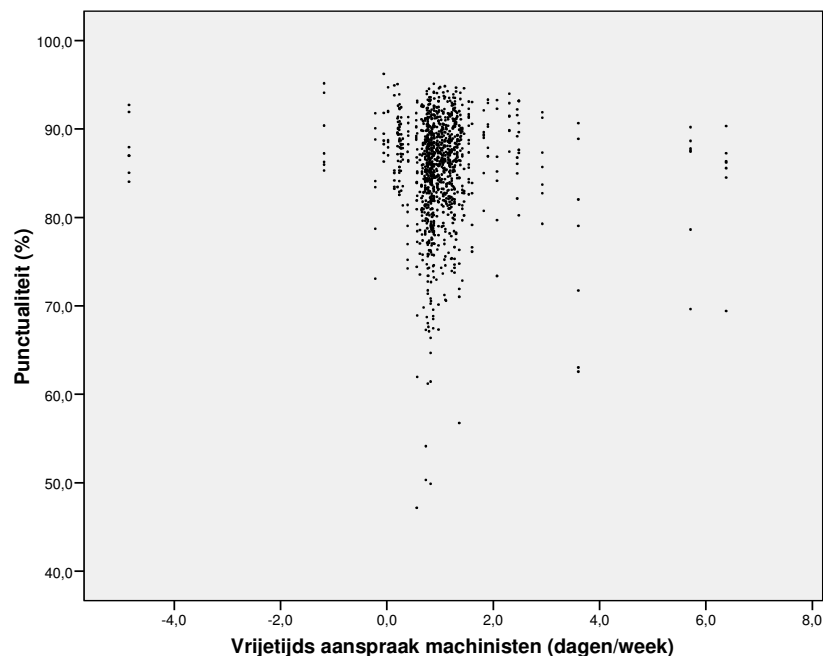
Conclusie: Er bestaat geen relatie tussen het ziektecijfer en de punctualiteit.

6.2.2 Vrijtijdsaanspraak personeel

Verwachting: Een hoge vrijtijdsaanspraak (VTA)¹³ van rijdend personeel zorgt voor een lagere punctualiteit.

Data: Voor deze relatie wordt de gemiddelde VTA-stand per personeelsgroep gebruikt. De opgenomen hoeveelheid VTA is VTA-stand (deze week) - VTA-stand (volgende week).

Relatie: De relatie kan op een gelijke manier verklaard worden als de relatie ziektecijfers-punctualiteit: meer VTA legt meer druk op het werkende personeel, wat leidt tot meer stress en een dalende punctualiteit. Deze relatie bestaat echter net als de relatie ziektecijfers-punctualiteit niet, zoals gezien kan worden in het onderstaande diagram. De figuren voor andere personeelsgroepen (HC, MMC, KNP, indirect) laten een vergelijkbaar beeld zien.



Figuur 6.14 Relatie VTA machinisten – punctualiteit. Negatieve VTA-waarden worden voornamelijk veroorzaakt door overwerken.

¹³ Het opnemen van vrije tijd.

Conclusie: Er is geen verband tussen VTA en punctualiteit.

6.2.3 Treinactiviteiten

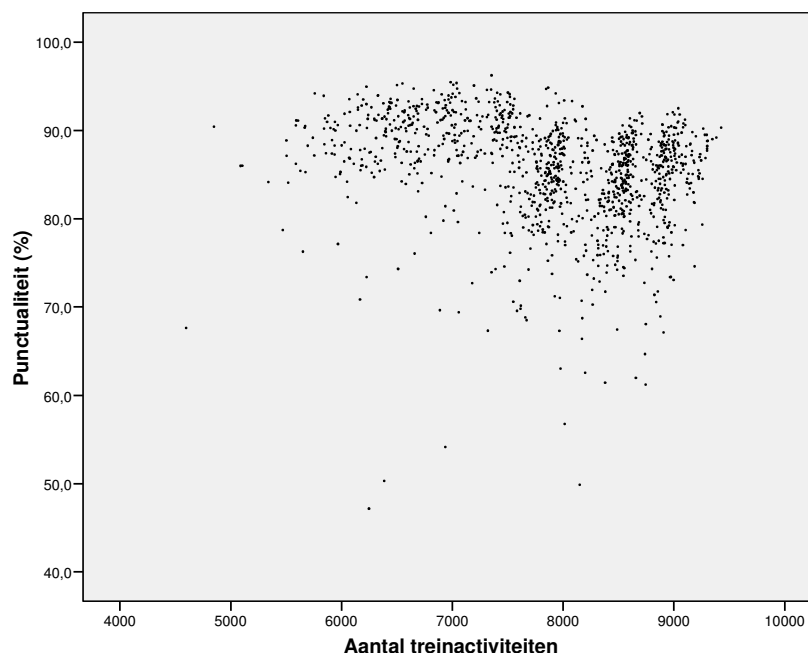
Verwachting: Als er minder treinen rijden daalt/stijgt de punctualiteit

Data: Als maat voor het aantal treinen dat gereden heeft op een dag wordt de gemeten treinactiviteit voor de punctualiteit genomen. Dit getal wordt op dagelijkse basis bekend gemaakt en staat goed in verhouding tot het aantal treinen dat daadwerkelijk gereden heeft. De treinactiviteit is het aantal treinen dat gemeten is op de knooppunten.

Relatie: De relatie tussen het aantal treinen en de punctualiteit is lastig. De eenvoudigste redenering is dat als minder treinen op het netwerk aanwezig zijn de vertragingen minder doorrekenen waardoor de punctualiteit stijgt. Een andere gehoorde redenering is dat wanneer minder treinen rijden de treinen met vertraging zwaarder tellen, maar deze redenering is niet direct waar¹⁴. De achterliggende reden van een dalende punctualiteit bij minder treinactiviteiten zit waarschijnlijk in andere factoren. Deze factoren zorgen voor een hoge uitval van treinen én een lage punctualiteit van treinen die wel rijden. Uit het onderstaande figuur blijkt ook dat de relatie dubieus is: zowel bij hoge als lage treinactiviteit worden zowel hoge als lage punctualiteiten behaald.

In de figuur zijn een drietal puntenwolken te onderscheiden die toe te wijzen zijn aan verschillende niveaus van treinactiviteiten in 2004, 2005 en 2006. Het apart plotten van de data voor deze jaren levert ook geen (duidelijk) verband op.

Conclusie: Er bestaat geen relatie tussen het aantal treinactiviteiten en de punctualiteit.



Figuur 6.15 Relatie aantal treinactiviteiten – punctualiteit. De puntenwolken worden veroorzaakt door de toename van het aantal ritten in de jaren 2004 – 2007.

6.2.4 Materieel in onderhoud

Verwachting: Een tekort aan materieel leidt tot een lagere punctualiteit.

Data: Het materieelregelcentrum houdt van dag tot dag (werkdagen) bij hoeveel materieel beschikbaar is ten opzichte van de planning, wat inhoudt dat er per dag een positieve of negatieve stand is ten

¹⁴ Vergelijk de volgende situaties:

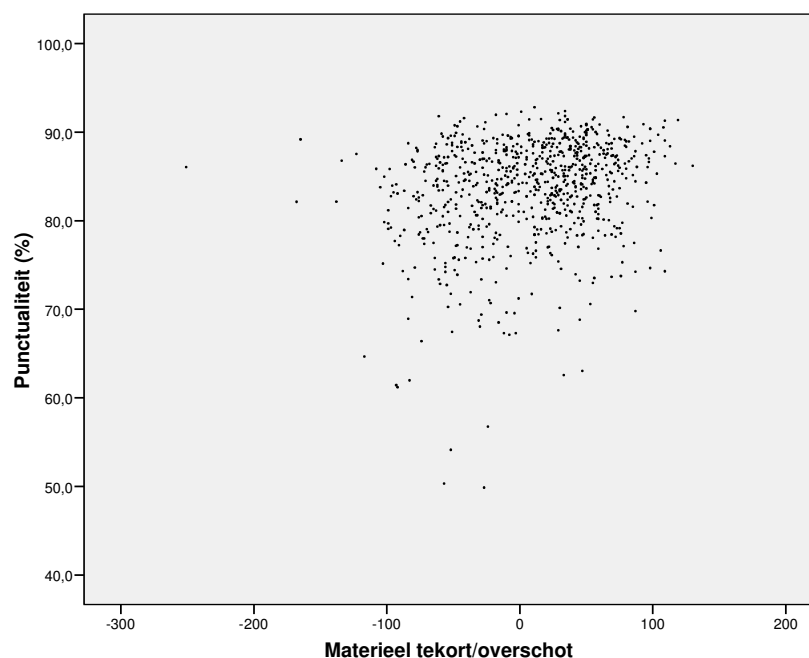
Op een dag met 10.000 treinactiviteiten hebben 2.000 treinen vertraging wat leidt tot een punctualiteit van 80%.
Op een dag met 8.000 treinactiviteiten (20% minder) hebben 1.600 treinen vertraging (20% minder) wat leidt tot een punctualiteit van wederom 80%.

opzichte van wat NedTrain binnen mag hebben. NedTrain verzorgt op periodieke basis onderhoud aan treinen, dit onderhoud duurt regelmatig langer of korter dan gepland. Daarnaast ontstaan er ook defecten aan treinstellen die niet gepland zijn voor onderhoud: deze treinstellen moeten ook onttrokken worden aan de treinenloop. Deze variaties leiden tot tekorten en overschotten op de planning.

Doordat de gegevens alleen voor werkdagen bijgehouden worden, kan de analyse alleen uitgevoerd worden voor deze dagen. De verwachting is dat de tekorten en overschotten in het weekend ook veel minder invloed hebben: door de meer beperkte dienstregeling in het weekend is er meer materieel beschikbaar en kunnen tekorten makkelijker opgevangen worden.

Relatie: De verwachte relatie is gebaseerd op een welbekende situatie: een trein is te kort om de alle reizigers te verwerken. Deze situatie leidt tot langere instap/uitstaptijden en dus tot vertraging. Uit het onderstaande figuur blijkt dat deze relatie niet bestaat: Er is een grote puntenwolk rond nul tekort - 86% punctualiteit met staarten naar beneden (lagere punctualiteit) en naar links (meer tekort). De correlatie is 0,220 en significant op een 99% zekerheidsniveau, wat duidt op een lineaire relatie. De variabele zou in een regressieanalyse meegenomen kunnen worden, maar met de kanttekening dat de gegevens niet compleet zijn (weekenden missen) en dat de relatie niet al te sterk is.

Conclusie: Er is een zwakke relatie tussen materieel tekorten/overschotten en punctualiteit, maar door gebrek aan complete data en de zwakke relatie zal de variabele niet meegenomen worden.



Figuur 6.16 Relatie materieel tekorten/overschotten – punctualiteit

6.2.5 Historische punctualiteit

Verwachting: De punctualiteit vandaag beïnvloedt de punctualiteit van morgen.

Relatie: De punctualiteit is per werkdag aardig te vergelijken. Het vergelijken van maandagen met maandagen, dinsdagen met dinsdagen, etc. levert een opvallend beeld op: de punctualiteit blijkt rond hetzelfde niveau te liggen. Deze waarneming duidt op autocorrelatie: de (vertraagde) afhankelijke variabele is zélf een onafhankelijke variabele. In het onderstaande figuur is een autocorrelatieplot opgenomen, duidelijk is dat er pieken zijn rond t-1, t-7 en t-14. De vertraging op de x-as duidt op het volgende:

za	zo	ma	di	wo	do	vr	za	zo	ma	di	wo	do	vr	za
14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

De sterke autocorrelatie met gisteren (t-1), een week geleden (t-7), twee weken geleden (t-14), etc. duidt op een goede toepasbaarheid van een Box-Jenkins (ARIMA) model. Een k-NN of NN model kan aangepast worden zodat deze ook de vertraagde punctualiteit meeneemt, en een MLR model kan als onafhankelijke variabele de punctualiteit van t-1, t-7 en t-14 meenemen. Het ligt overigens in de verwachting dat de autocorrelaties van t-7 en t-14 in het geval van het MLR model beter gemodelleerd kunnen worden met behulp van andere wekelijks terugkerende cijfers zoals het reizigersaantal. De autocorrelatie t-1 heeft wel een ‘interne’ relatie: uit de historische data blijkt dat een daling in de punctualiteit meestal voortgezet wordt. Redenen hiervoor zouden kunnen zijn:

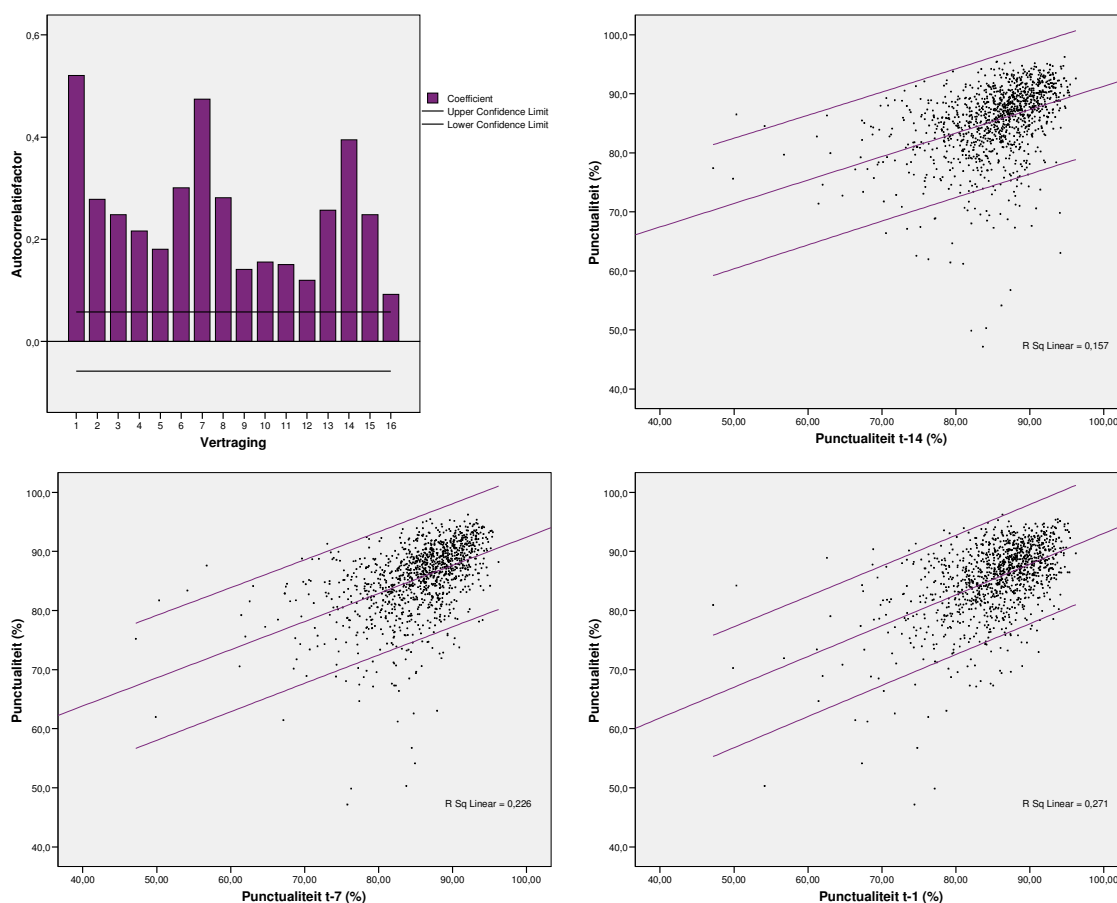
- Stressvolle dag voor medewerkers leidt tot een dip in motivatie de volgende dag;
- Materieel is na een slechte dag niet op de juiste plaats wat zorgt voor capaciteits- en opstartproblemen.

Een ‘externe’ relatie die zeker een rol speelt is de volgende:

- De omstandigheden die zorgden voor een daling (stijging) van de punctualiteit op t-1 gelden ook op moment t.

Deze relatie zou dus verklaard moeten worden door de genoemde omstandigheden.

De correlaties zijn 0,396, 0,476 en 0,521 voor respectievelijk t-1, t-7 en t-14.



Figuur 6.17 Autocorrelatie van de punctualiteit

Conclusie: De punctualiteit van t-1, t-7 en t-14 vertoont een duidelijk verband met de punctualiteit op dag t.

6.3 Conclusie

In de onderstaande tabel zijn de relaties van externe en interne onafhankelijke variabelen opgenomen. De geaccepteerde relaties zullen gebruikt worden in het volgende hoofdstuk om een

voorspellingsmodel voor de punctualiteit te maken. De hoge autocorrelatie pleit voor het gebruik van een Box-Jenkins (ARIMA) model.

Relatie	Geaccepteerd?	Pearson Correlatie
Maximale temperatuur – Punctualiteit	Ja	-0,457
Minimale temperatuur – Punctualiteit	Nee	0,063
Windbladval – Punctualiteit	Ja	-0,502
Windsnelheid – Punctualiteit	Nee	-0,126
Maximale uurgemiddelde windsnelheid – Punctualiteit	Nee	-0,089
Sneeuwval – Punctualiteit	Ja	-0,782
Neerslag(hoeveelheid) – Punctualiteit	Nee	-0,142
Aantal reizigers – Punctualiteit	Ja	-0,519
<i>Ziektecijfers</i>		
Ziektecijfer MCN – Punctualiteit	Nee	-0,014
Ziektecijfer HC – Punctualiteit	Nee	-0,047
Ziektecijfer MMC – Punctualiteit	Nee	-0,041
Ziektecijfer KNP – Punctualiteit	Nee	0,008
Ziektecijfer Indirect – Punctualiteit	Nee	-0,040
Materieel tekorten/overschotten – Punctualiteit	Nee	0,220
Suicides – Punctualiteit	Nee	-0,041
<i>Autocorrelatie punctualiteit</i>		
Punctualiteit (t-1) – Punctualiteit t	Ja	0,521
Punctualiteit (t-7) – Punctualiteit t	Ja	0,476
Punctualiteit (t-14) – Punctualiteit t	Ja	0,396

7 Voorspellingsmodellen

De gevonden relaties in het voorgaande hoofdstuk zullen in dit hoofdstuk gebruikt worden voor het opstellen van modellen voor de punctualiteit. Doel van de modellen is steeds om de dagpunctualiteit 24 uur vooruit te kunnen voorspellen. In de onderstaande tabel zijn de relaties wederom weergegeven, inclusief de symbolen die gebruikt gaan worden in de rest van dit hoofdstuk (x voor schaalbare variabelen, b voor nominale variabelen en p voor punctualiteitdata/autocorrelatie variabelen).

Relatie	Geaccepteerd?	Pearson Correlatie	Symbool	Eenheid
Maximale temperatuur – Punctualiteit	Ja	-0,457	x_1	0,1K
Windbladval – Punctualiteit	Ja	-0,502	x_2	m/s
Sneeuwval – Punctualiteit	Ja	-0,782	x_3	-
Aantal reizigers – Punctualiteit	Ja	-0,519	x_4	-
Vakantie – Punctualiteit	Indirect	-	b_1	-
Weekend – Punctualiteit	Indirect	-	b_2	-
Punctualiteit (t-1) – Punctualiteit t	Ja	0,521	p_1	%
Punctualiteit (t-7) – Punctualiteit t	Ja	0,476	p_7	%
Punctualiteit (t-14) – Punctualiteit t	Ja	0,396	p_{14}	%

Tabel 7.1 Geaccepteerde variabelen

7.1 Validatieset

Om de modellen die hieronder beschreven worden te kunnen beoordelen aan de hand van de MSE, de MAPE en de *bias*, moet een dataset gebruikt worden die verschillende situaties meeneemt zoals bladval en sneeuw. De data voor de validatieset zijn willekeurig geselecteerd uit de complete dataset en worden niet meer meegenomen voor het bepalen van de parameters van de modellen. De validatieset heeft een gemiddelde punctualiteit van 85,5% wat voldoende overeenkomt met de gemiddelde punctualiteit over de hele periode januari 2006 – maart 2007 van 85,4% om te veronderstellen dat de validatieset representatief is. Daarnaast heeft de validatieset ook een aantal hoge en lage waarnemingen.

#	Case#	Datum	Punctualiteit	Dag	Situatie ¹⁵
1	809	19-3-2006	90,9	Zondag	Aanrijdingen
2	812	22-3-2006	86,2	Woensdag	Weinig bijzonderheden
3	821	31-3-2006	87,1	Vrijdag	Weinig bijzonderheden
4	822	1-4-2006	91,7	Zaterdag	Weinig bijzonderheden
5	824	3-4-2006	90,6	Maandag	Ontsporing Nm, Stroomstoring Gvc-Gd
6	834	13-4-2006	87	Donderdag	Weinig bijzonderheden
7	850	29-4-2006	86,3	Zaterdag	Koninginnedag, Noodrentrekkingen Asd, Brandende bielzen Gvc
8	854	3-5-2006	86,1	Woensdag	Overwegstoringen Apd – Dv, Zp, O – Ht, Rvs – O
9	856	5-5-2006	84,2	Vrijdag	Ut: Gaslek, (beperkte) stroomstoring
10	862	11-5-2006	83,4	Donderdag	Stroomstoring Ut – Gd, defecte treinen
11	865	14-5-2006	92,6	Zondag	Weinig bijzonderheden
12	867	16-5-2006	90,5	Dinsdag	Weinig bijzonderheden
13	873	22-5-2006	83	Maandag	Uitloop werkzaamheden Amf- Zl, veel kleine storingen
14	881	30-5-2006	89,6	Dinsdag	Weinig bijzonderheden
15	883	1-6-2006	83,4	Donderdag	Weinig bijzonderheden
16	888	6-6-2006	86	Dinsdag	Ontruiming Mt, oplopende vertraging

¹⁵ Opvallende gebeurtenissen uit de RBV-meldkaarten (Bron: ProRail).

					Ut
17	890	8-6-2006	84,2	Donderdag	Spoorspatting, aanrijdingen met personen (3x)
18	891	9-6-2006	88,3	Vrijdag	Hoge temperatuur (spoorspatting, bruggen)
19	894	12-6-2006	78,1	Maandag	Hoge temperatuur, wisselstoringen Nm, Zp, Dv, Gvc
20	903	21-6-2006	87,2	Woensdag	Weinig bijzonderheden
21	905	23-6-2006	89,7	Vrijdag	Weinig bijzonderheden
22	917	5-7-2006	71,9	Woensdag	Zware weersomstandigheden: Warm & onweer/hagel/regen
23	920	8-7-2006	90,4	Zaterdag	Bermbrandjes
24	929	17-7-2006	81	Maandag	Hoge temperaturen, bielzen in brand, sectiestoringen
25	951	8-8-2006	91,1	Dinsdag	Weinig bijzonderheden
26	953	10-8-2006	92	Donderdag	Weinig bijzonderheden
27	958	15-8-2006	89,8	Dinsdag	Weinig bijzonderheden
28	966	23-8-2006	87,7	Woensdag	Weinig bijzonderheden
29	990	16-9-2006	85,4	Zaterdag	Weinig bijzonderheden
30	999	25-9-2006	76,9	Maandag	Stroomstoring Ut, verdacht pakketje Dt
31	1022	18-10-2006	81	Woensdag	Bladval
32	1025	21-10-2006	86,6	Zaterdag	Bladval
33	1027	23-10-2006	78,4	Maandag	Hevige bladval, defecte treinen
34	1032	28-10-2006	88	Zaterdag	Bladval, weinig bijzonderheden
35	1033	29-10-2006	88,6	Zondag	Bladval, defecte treinen
36	1047	12-11-2006	76,1	Zondag	Bladval, veel storingen, defecte treinen
37	1054	19-11-2006	77,1	Zondag	Bladval, defecte treinen
38	1059	24-11-2006	64,7	Vrijdag	Bladval, aanrijdingen, defecte treinen
39	1064	29-11-2006	78,4	Woensdag	Bladval, weinig bijzonderheden
40	1075	10-12-2006	88,9	Zondag	Uitloop werkzaamheden
41	1096	31-12-2006	94,2	Zondag	Weinig bijzonderheden
42	1100	4-1-2007	90,3	Donderdag	Weinig bijzonderheden
43	1104	8-1-2007	85,5	Maandag	Weinig bijzonderheden
44	1113	17-1-2007	84,7	Woensdag	Defecte treinen
45	1120	24-1-2007	81,9	Woensdag	Defecte treinen
46	1125	29-1-2007	87	Maandag	Weinig bijzonderheden
47	1131	4-2-2007	91,9	Zondag	Weinig bijzonderheden
48	1133	6-2-2007	86,3	Dinsdag	Defecte treinen
49	1134	7-2-2007	86,2	Woensdag	Weinig bijzonderheden
50	1148	21-2-2007	87,3	Woensdag	Weinig bijzonderheden

Tabel 7.2 Validatieset

7.2 Modellen

7.2.1 Meervoudige lineaire regressie

Meervoudige lineaire regressie (MLR) wordt als eerste gebruikt omdat deze methode een duidelijk inzicht geeft in het belang van verschillende onafhankelijke variabelen voor het voorspellen van de afhankelijke variabele. De onafhankelijke variabelen kunnen een duidelijk verband vertonen, maar hoeven nog niet significant van invloed te zijn in aanwezigheid van andere verklarende onafhankelijke variabelen. Door met behulp van MLR vast te stellen welke o.v. van grote invloed zijn, en welke niet, kan een inschatting gemaakt worden van het belang van deze o.v. in de andere methodes.

Voor het bepalen van de vergelijkingen wordt gebruik gemaakt van het statistische software pakket SPSS en alle overgebleven datapunten na uitsluiting van de validatieset en stakingsdagen. De vergelijking die gezocht wordt heeft de volgende vorm:

$$\hat{Y} = \beta_0 + \sum_{i=1}^4 \beta_i x_i + \beta_5 p_1 + \beta_6 p_7 + \beta_7 p_{14} \quad (7.1)$$

SPSS geeft voor de o.v. de volgende informatie:

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		β	Std. Error	β		
MLR ₁	Constante	63,889	5,986		10,673	,000
	Maximale temperatuur (x_1)	-,007	,002	-,080	-3,599	,000
	WindBladval (x_2)	-,079	,012	-,149	-6,704	,000
	Sneeuwval (x_3)	-12,479	1,225	-,228	-10,190	,000
	Reizigers (x_4)	-,031	,003	-,269	-11,008	,000
	Punctualiteit (t-1, p_1)	,264	,023	,264	11,486	,000
	Punctualiteit (t-7, p_7)	,207	,026	,207	8,044	,000
	Punctualiteit (t-14, p_{14})	,110	,025	,111	4,423	,000

Tabel 7.3 Data voor de lineaire regressie.

De toevoeging van de vertraagde punctualiteitwaarnemingen p_i voor $i = 1, 7$ en 14 levert een model autoregressief model op dat problemen kan geven bij de schattingen van de parameters met behulp van de kleinste kwadraten methode. Daarnaast is de waarde van de Durbin-Watson toets (zie onder) ook niet helemaal betrouwbaar meer. Eerdere modellen hebben laten zien dat het opnemen van autoregressieve termen noodzakelijk is: zonder deze toevoeging vertonen de fouten van de voorspellingen autocorrelatie (een Durbin-Watson waarde die ver van twee afligt). Er is veel onderzoek gedaan naar complexe methodes voor het correct opnemen van vertraagde afhankelijke variabelen als onafhankelijke variabelen, maar de beschreven methodes zijn te complex voor dit onderzoek. Volgens Tanizaki (2000) levert het gebruik van de kleinste kwadraten methode in combinatie met autoregressieve termen voornamelijk problemen op bij kleine datasets. De dataset gebruikt voor dit onderzoek bevat meer dan 1000 individuele waarnemingen, waarmee verondersteld kan worden dat de onnauwkeurigheid van de kleinste kwadratenmethode verwaarloosbaar is.

De onderlinge correlatie van de geselecteerde o.v. is ook van belang voor het selecteren van de o.v. De onderlinge correlatie is in de onderstaande tabel weergegeven. De gegevens in deze tabel geven geen reden voor het verwijderen van één van de variabelen uit de vergelijking.

	Maximale temperatuur (x_1)	WindBladval (x_2)	Sneeuwval (x_3)	Reizigers (x_4)	Punctualiteit (t-14, p_{14})	Punctualiteit (t-7, p_7)	Punctualiteit (t-1, p_1)
Maximale temperatuur (x_1)	1	-,058	-,306	-,028	,135	,123	,086
WindBladval (x_2)	-,058	1	,023	,207	-,165	-,237	-,301
Sneeuwval (x_3)	-,306	,023	1	,015	,005	-,032	-,174
Reizigers (x_4)	-,028	,207	,015	1	-,414	-,448	-,208
Punctualiteit (t-14, p_{14})	,135	-,165	,005	-,414	1	,487	,248
Punctualiteit (t-7, p_7)	,123	-,237	-,032	-,448	,487	1	,300
Punctualiteit (t-1, p_1)	,086	-,301	-,174	-,208	,248	,300	1

Tabel 7.4 Onderlinge (Pearson) correlatie van de o.v., een waarde van -1 of 1 geeft een perfecte (lineaire) correlatie weer.

Aan de hand van tabel 7.3 kan de volgende vergelijking (MLR₁) opgesteld worden:

$$\hat{Y} = 63,9 - 0,007x_1 - 0,079x_2 - 12,479x_3 - 0,031x_4 + 0,264p_1 + 0,207p_7 + 0,110p_{14} \quad (7.2)$$

Wanneer het belang van de afzonderlijke o.v. bekeken moet worden kan de gestandaardiseerde coëfficiënt gebruikt worden. Om tot de gestandaardiseerde coëfficiënt te komen, zijn de waarden van de o.v. vertaald in z-scores volgens de onderstaande formule, waarin μ het gemiddelde van de waarnemingen is, en σ de standaard deviatie:

$$\tilde{X} = \frac{X - \mu}{\sigma} \quad (7.3)$$

In vergelijking (7.3) wordt de normaal verdeelde o.v. X vertaald naar de *standaard* normaal verdeelde variabele $\tilde{X}$. De gestandaardiseerde vergelijking ziet er als volgt uit:

$$\hat{Y} = 63,9 - 0,080\tilde{x}_1 - 0,149\tilde{x}_2 - 0,228\tilde{x}_3 - 0,269\tilde{x}_4 + 0,264\tilde{p}_1 + 0,207\tilde{p}_7 + 0,111\tilde{p}_{14} \quad (7.4)$$

Uit deze vergelijking blijkt dat van de variabelen reizigers (x_4), punctualiteit (t-1, p_1) en sneeuwval (x_3) de meeste invloed hebben op de punctualiteit.

Met behulp van de eerder in paragraaf 4.3.7 besproken testen kan bekeken worden of de vergelijking een goede benadering van de realiteit geeft.

7.2.1.1 Ligt nul in het betrouwbaarheidsinterval van β_i ?

Het betrouwbaarheidsinterval wordt gegeven door:

$$(\hat{\beta}_i - cS_{\hat{\beta}_i}, \hat{\beta}_i + cS_{\hat{\beta}_i}) \quad (7.5)$$

Met c bepaald door de student-verdeling, of in het geval van een groot aantal vrijheidsgraden door de normale verdeling:

$$P(T_{n-(k+1)} \leq c) = 1 - \frac{1}{2}\alpha \quad (7.6)$$

In het geval van MLR₁ is de waarde van c 1,96. De betrouwbaarheidsintervallen worden dan:

Model		Unstandardized Coefficients		
		β	min	max
MLR ₁	Constante	63,889	52,143	75,634
	Maximale temperatuur (x_1)	-,007	-0,010	-0,003
	WindBladval (x_2)	-,079	-0,103	-0,056
	Sneeuwval (x_3)	-12,479	-14,882	-10,076
	Reizigers (x_4)	-,031	-0,036	-0,025
	Punctualiteit (t-1, p_1)	,264	0,219	0,310
	Punctualiteit (t-7, p_7)	,207	0,157	0,258
	Punctualiteit (t-14, p_{14})	,110	0,061	0,159

Tabel 7.5 Minimum en maximum waarden parameters

Voor geen van de coëfficiënten valt nul dus binnen het betrouwbaarheidsinterval, er kan dus verondersteld worden dat alle coëfficiënten van invloed zijn op $\hat{Y}$.

7.2.1.2 Geldt $\beta_1 = \beta_2 = \dots = \beta_6 = 0$ en $\beta_7 = 0$?

Toetsen of minimaal één van de coëfficiënten niet gelijk is aan nul wordt gedaan met de waarde van F :

$$F = \frac{MSR}{MSE} \quad (7.7)$$

De waarde van F voor MLR_1 kan uit de volgende (ANOVA) tabel gehaald worden:

ANOVA

Model		Sum of Squares	df	Mean Square (MS)	F	Sig.
MLR ₁	Regression	22178,970	7	3168,424	168,826	,000
	Error	20906,856	1114	18,767		
	Total	43085,826	1121			

Tabel 7.6 ANOVA tabel voor MLR_1

De F waarde in deze tabel is hoog genoeg ($Sig < 0,05$) om te kunnen veronderstellen dat minimaal één β_i niet gelijk is aan 0.

Toetsen of een coëfficiënt gelijk is aan nul geeft naast de voorgaande toets een goede indicatie of een variabele (in aanwezigheid van de andere variabelen) nuttig is. De waarde van t in Tabel 7.3 moet hiervoor kleiner zijn dan 0,05 bij een betrouwbaarheidsdrempel van 5%. Aangezien dit voor alle variabelen geldt, zijn alle variabelen van belang bij het voorspellen van de punctualiteit Y .

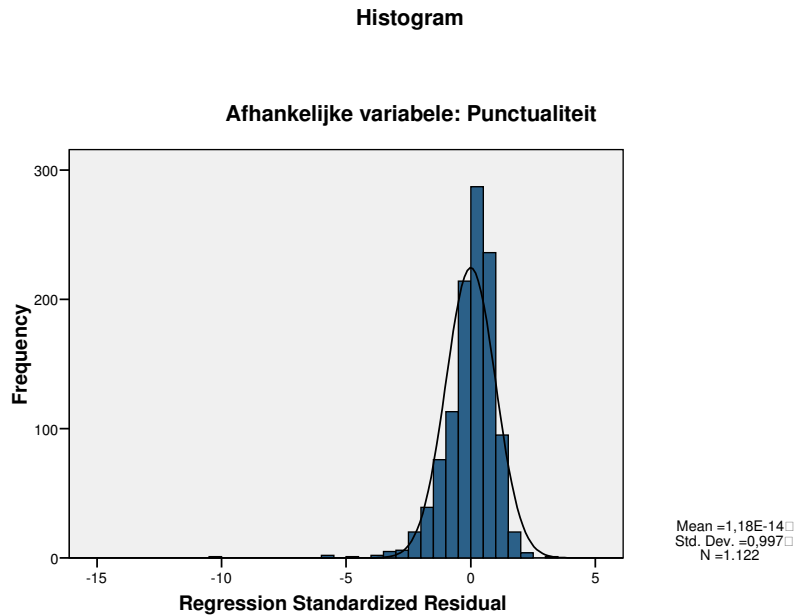
7.2.1.3 Overige statistieken

Andere interessante statistieken zijn de R^2 en Durbin-Watson getallen. Voor het MLR_1 model zijn in de onderstaande tabel deze waarden weergegeven. De waarde van 0,515 voor de R^2 geeft aan dat de vergelijking 51,5% van de schommelingen in de punctualiteit verklaart. De Durbin-Watson waarde van 1,819 geeft aan dat er hoogstwaarschijnlijk geen autocorrelatie meer in de data zit na de toevoeging van de p_1 , p_7 en p_{14} .

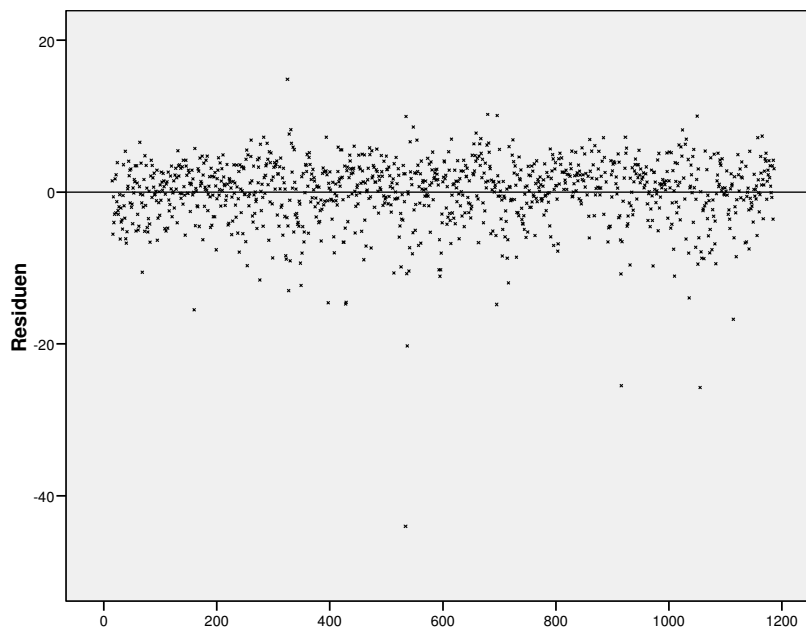
Model	R	R^2	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
MLR ₁	,717	,515	,512	4,33	1,82

Tabel 7.7 Statistieken voor MLR_1

De residuen laten ook een redelijk goed beeld zien: de verdeling is ongeveer standaard normaal ($N(0,1)$), zoals gezien kan worden in onderstaand histogram. Ook het hieronder opgenomen plot van de residuen laat geen trend of andere opvallende vorm zien.



Figuur 7.1 Foutenverdeling van MLR_1



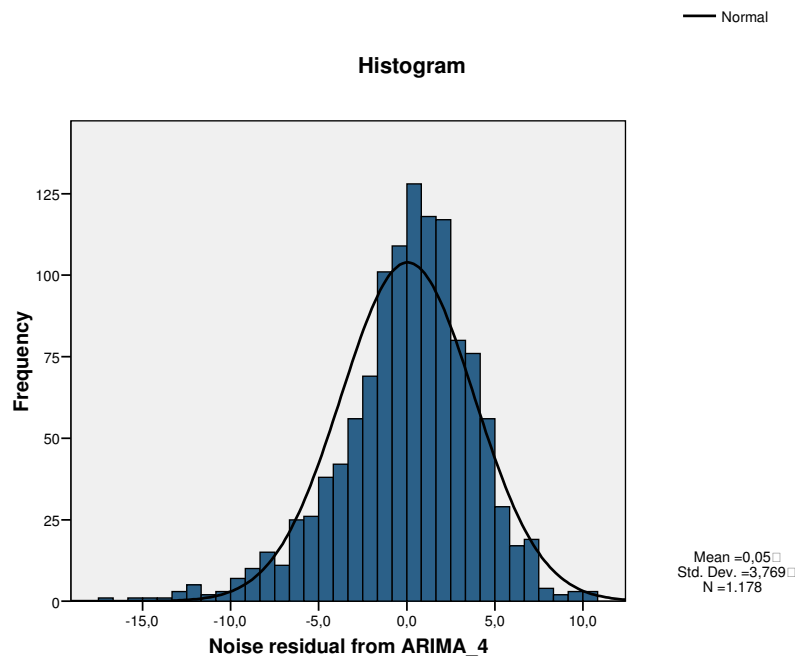
Figuur 7.2 De residuen zijn willekeurig verdeeld

7.2.2 ARIMA/Box-Jenkins

In aanvulling op het MLR model is ook gekeken naar modellen die alleen kijken naar de punctualiteit in het verleden om de toekomst te voorspellen. Het beschreven MLR model in de vorige paragraaf zit dicht bij een ARIMA model door te toevoeging van de punctualiteit $t-1$, $t-7$ en $t-14$ waarden. Het volledige proces waarmee het model vastgesteld is kan gevonden worden in bijlage B. Het ARIMA model is in eerste instantie gemaakt met alle datapunten, dus inclusief de validatieset die eerder gegeven is. De validatieset zal later, met behulp van de gevonden modelvorm, gebruikt worden als werkelijke validatieset met als basis voor het model de voorgaande data. Het gevonden ARIMA model heeft de vorm $ARIMA(1, 1, 1)(1, 1, 1)_7$ (kortweg $ARIMA_4$) wat neerkomt op één korte

differentie, één seizoensdifferentie, en zowel voor de seizoen als korte termen een AR en een MA term.

In het onderstaande figuur kan gezien worden dat de verdeling van de residuen redelijkerwijs als normaal beschouwd kan worden.



Figuur 7.3 Foutenverdeling van ARIMA₄.

De verdere statistieken van dit model zien er als volgt uit:

Model	Stationary R-squared	Ljung-Box Q(18)			Number of Outliers
		Statistics	DF	Sig.	
ARIMA ₄	,695	17,871	14	,213	12

Tabel 7.8 Statistieken voor ARIMA₄

De Ljung-Box statistiek laat een goede waarde die groter is dan 0,05 en dat betekent dat de residuen willekeurig verdeeld zijn en geen informatie meer bevatten die gebruikt kan worden voor het verbeteren van de voorspelling.

De parameters van het model voor de complete dataset zijn als volgt:

		Parameter	SE	t	Sig.
AR	Lag 1	,347	,036	9,669	,000
Differenties		1			
MA	Lag 1	,895	,018	50,674	,000
AR, Seizoens-	Lag 1	,095	,032	2,916	,004
Seizoensdifferenties		1			
MA, Seizoens-	Lag 1	,970	,011	92,291	,000

Tabel 7.9 Parameters voor ARIMA₄

Met een significantie kleiner dan 0,05 kan verondersteld worden dat alle parameters terecht in het model staan.

De differenties zorgen voor het volgende beeld van de variabelen:

$$\begin{aligned}\Delta Y_t &= Y_t - Y_{t-1} \\ \Delta_7 Y_t &= Y_t - Y_{t-7} \\ \Delta \Delta_7 Y_t &= \Delta(Y_t - Y_{t-7}) = Y_t - Y_{t-1} - Y_{t-7} + Y_{t-8} = W_t\end{aligned}\quad (7.8)$$

Het model heeft dan de volgende vorm:

$$W_t = 0,347W_{t-1} - 0,895\varepsilon_{t-1} + 0,095W_{t-7} - 0,970\varepsilon_{t-7}\quad (7.9)$$

Of in de uitgeschreven vorm:

$$\begin{aligned}Y_t &= Y_{t-1} + Y_{t-7} - Y_{t-8} + 0,347(Y_{t-1} - Y_{t-2} - Y_{t-8} + Y_{t-9}) \\ &\quad - 0,895\varepsilon_{t-1} + 0,095(Y_{t-7} - Y_{t-8} - Y_{t-14} + Y_{t-15}) - 0,970\varepsilon_{t-7}\end{aligned}\quad (7.10)$$

Om te controleren of de modelvorm opgaat voor de gehele tijdsperiode van het valideringsinterval zijn modellen gemaakt met data van 1 maart, 1 april, 1 juni, 1 september, 1 december 2006 en 1 maart 2007. Bij al deze datasets bleef de Ljung-Box significantie boven de 5% en bleef de autocorrelatie tussen de residuen binnen acceptabele waarden. De modelvorm ARIMA(1, 1, 1)(1, 1, 1)₇ kan dus op voor alle validatiepunten gebruikt worden.

7.2.3 k-Nearest Neighbours

Zoals beschreven doorzoekt het k-Nearest Neighbours (k-dichtstbijzijnde burens) een tijdserie voor overeenkomende situaties. Om de beschikbare dataset te kunnen doorzoeken is een programma geschreven dat deze zoekopdracht uit kan voeren. Het programma is geschreven in PHP¹⁶ en is benaderbaar via een webpagina wat het zeer eenvoudig maakt om van dag tot dag voorspellingen te maken.

7.2.3.1 Werking algoritme/programma

Het is onwenselijk om in dit rapport uitgebreid in te gaan op de code die geschreven is, daarom zal een korte omschrijving van de werking van de code gegeven worden. Het k-NN zoekprogramma werkt met de klasse `class knn` waarvan een object gemaakt kan worden met de variabelen `$k` en `$x`, waarbij:

- `$k` het aantal burens waarnaar gekeken moet worden (*a* in paragraaf 4.3.6) en
- `$x` het aantal best overeenkomende dagen waarvan het gemiddelde genomen moet worden voor de voorspelling.

In het geval dat er gekeken wordt naar de voorgaande drie dagen, dan heeft de variabele `$k` de waarde '3'. Wanneer de best overeenkomende vier dagen geselecteerd worden uit de database met gegevens, heeft de variabele `$x` de waarde '4'.

De opbouw van de klasse is verder als volgt (niet alle functies zijn weergegeven):

```
<?php

class knn{

    //constructor
    function knn($k, $x, $field = "punctualiteit"){...}

    //selecteer de tabel die gebruikt moet worden voor het voorspellen
```

¹⁶ Recursief acroniem: PHP Hypertext Preprocessor. PHP is een populaire (script-)taal die voornamelijk gebruikt wordt voor webapplicaties.

```
function setTable($table){...}  
  
//stel het moment in waarvoor een voorspelling gemaakt moet worden  
function setPrediction($ref = ''){...}  
  
//stel de referentieset in  
function setReference(){...}  
  
//stel in naar welke (onafhankelijke) variabelen gekeken moet worden  
function setActiveFields($array){...}  
  
//maak de voorspelling, zoek de waarden die het best overeenkomen  
function getKey(){...}  
  
//print de voorspelling  
function printForecast(){...}  
  
}  
?>
```

Code 7.1

Voor het maken van een voorspelling voor 5 juni 2007 aan de hand van de maximale temperatuur en het aantal reizigers, met 3 burens als referentie, en het gemiddelde van de beste 4, is dus vereenvoudigd de volgende code nodig:

```
<?php  
  
//maak een object aan van de klasse knn  
$predictor = new knn(3, 4);  
  
//stel het voorspelmoment in  
$predictor->setPrediction('2007-06-05');  
  
//selecteer de velden  
$predictor->setActiveFields(array('maxtemp', 'reizigers'));  
  
//maak de voorspelling  
$predictor->getKey();  
  
//print de voorspelling  
$predictor->printForecast();  
  
?>
```

Code 7.2

De functie **getKey()** voert het daadwerkelijke zoeken uit. De functie gebruikt hiervoor procentuele verschillen tussen de referentie en de kandidaten. De output (**printForecast()**) bevat, naast enkele statistieken, de voorspelling voor het gewenste moment. Naast een ongewogen gemiddelde van de beste s_x overeenkomende kandidaten, wordt ook een gewogen gemiddelde gegeven aan de hand van de procentuele overeenkomst van de data.

7.2.3.2 Welke parameters?

De waarden van het aantal burens s_k (zie ook paragraaf 4.3.6, de variabele a), en het aantal best overeenkomende kandidaten s_x (zie ook paragraaf 4.3.6, het aantal best overeenkomende dagen dat gebruikt wordt voor de voorspelling) kunnen niet numeriek bepaald worden en moeten daarom proefondervindelijk vastgesteld worden. Aan de hand van de testset moet een keuze gemaakt worden voor de beste waarden.

7.2.3.3 Gebruikte model

Voor het k-NN model zijn ongeveer dezelfde datasoorten gebruikt als voor het MLR_1 -model, te weten maximale temperatuur, reizigers, sneeuwval en windbladval. In tegenstelling tot het MLR_1 -model

wordt de punctualiteit op t-1, t-7 en t-14 niet meegenomen, maar wordt de werkelijk gemeten punctualiteit op het moment van de referentie en de kandidaat gebruikt. De punctualiteit van t-14, t-7 en t-1 wordt immers al gebruikt voor de waardering van de kandidaten op die momenten. Omdat per situatie de optimale parameters $\$k$ en $\$x$ verschillen, is voor verschillende parameters gekeken wat de optimale voorspellingen oplevert op een verder niet behandelde aparte testset. Allereerst is gekeken $\$x = \{3\}$ en $\$k = \{2, 4, 6, 7, 8, 10\}$.

$\$k$	2		4		6		7		8		10	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
	4,70	26,85	4,90	24,11	4,17	21,91	4,01	20,53	4,56	26,10	4,74	26,05

Tabel 7.10 Selectie van het aantal burens

Vervolgens is gekeken naar $\$k = \{7\}$ en $\$x = \{1, 2, \dots, 6\}$, als optimum geldt hier $\$x = 3$.

$\$x$	1		2		3		4		5		6	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
	4,80	25,29	4,30	23,46	4,01	20,54	4,06	21,07	4,10	21,28	4,13	20,80

Tabel 7.11 Selectie van het aantal best overeenkomende kandidaten

Voor het voorspellen van de punctualiteit lijkt een combinatie $\$k = 7$, $\$x = 3$ dus optimaal.

7.2.4 Neuraal netwerk

De werking van een neuraal netwerk is redelijk complex: het overzicht in gewichten, neuronen en waarden van de neuronen is snel verloren. De rekenkracht die van een computer gevraagd wordt om de gewichten van het netwerk uit te rekenen (backpropagation) is ook groot: het aantal gewichten voor een netwerk met 10 input neuronen, 210 ($n[2n + 1]$) verborgen neuronen in één laag en 1 output neuron is al $10 \cdot 210 + 210 = 2310$ gewichten. Het stuk voor stuk aanpassen van deze gewichten met een doorrekenende fout en vervolgens deze gewichten opslaan vraagt veel tijd. De programmacode (klasse) die geschreven is voor het berekenen van de gewichten is dan ook sterk geoptimaliseerd om in korte tijd toch veel iteraties te kunnen uitvoeren. Het script voert ongeveer 13.000 iteraties uit in 415 seconden, afhankelijk van de beginwaarden van de gewichten (welke willekeurig gekozen worden) zijn tussen de 2 en 10 maal 12.000 iteraties nodig.

7.2.4.1 Werking algoritme

Ook het neurale netwerk is geprogrammeerd in PHP, de gewichten worden opgeslagen in een MySQL¹⁷ database. Er worden 3 klassen gebruikt voor het neurale netwerk: `class network`, `class learn` en `class dataNormalization`. De eerste klasse wordt gebruikt om een neuraal netwerk object te maken, dit object bevat alle neuronen en gewichten die het netwerk laten functioneren. De klasse `learn` wordt gebruikt om de gewichten van het netwerk aan te passen, hiervoor heeft het object van de klasse `learn` een object `network` van de klasse `network`, van dit object worden in verschillende iteraties de gewichten aangepast zodat het netwerk bij bepaalde input waarden de juiste output waarde (voorspelling) gegeven wordt. De klasse `dataNormalization` wordt gebruikt om de inputwaarden te normaliseren naar waarden tussen 0 en 1.

Voordeel van deze wijze van werken is dat een leerproces voor een netwerk snel geschreven is:

```
<?php
//code voor het ophalen van data is niet weergegeven
//de array $input bevat de input waarden, de array $desired bevat
//de gewenste output waarden
```

¹⁷ MySQL is een open source relationele database management systeem (RDBMS), dat gebruik maakt van Structured Query Language (SQL).

```
//definieer het model: laag 1 met 10 neuronen, laag 2 met 210 neuronen
$networkmodel = array(1 => 10, 2 => 210, 3=> 1);

//maak een object van de klasse learn die het netwerk gaat leren, de
//gewichten worden opgeslagen in $stable
$teacher = new learn($networkmodel, $input, $desired, $stable);

//doe een 10 keer een epoch met de data in $desired
for($i = 0; $i < 10; $i++){
    $teacher->doEpoch();
}
?>
```

Code 7.3

Ook het berekenen van een voorspelling kan met relatief weinig code gebeuren:

```
<?php

//de array $input bevat de (genormaliseerde) input waarden.

//definieer het model: laag 1 met 10 neuronen, laag 2 met 210 neuronen
$networkmodel = array(1 => 10, 2 => 210, 3=> 1);

//maak een object van de klasse network die het netwerk representeert,
//de gewichten worden geladen uit $stable
$nnp = new network($networkmodel, $input, $stable);

$nvalue = $nnp->calculateNetwork();

//print de voorspelling, afgerond op 1 decimaal.
echo 'voorspelling: ' . round($nvalue, 1) . '%';
?>
```

Code 7.4

Door de korte schrijfwijze voor het leren en uitrekenen van neurale netwerken is het goed mogelijk om veel verschillende modellen te maken. Zoals vermeld in paragraaf 4.3.5 is het nodig om het aantal verborgen neuronen te variëren en te bepalen welke aantal optimaal werkt. Daarnaast is het startpunt van het leerproces van invloed op het succes van het leerproces. Het leerproces kan gezien worden als een balletje dat rolt over een landschap met heuvels (slechte resultaten) en dalen (goede resultaten), en het is goed mogelijk dat het balletje blijft steken in een dal dat niet het laagste (beste) dal is. Een dergelijke oplossing wordt een lokaal optimum genoemd, maar is niet het optimum van de hele ruimte. Er zijn geavanceerde methodes beschikbaar (bijvoorbeeld Plagianakos et al, 1998) om te voorkomen dat het leerproces blijft steken in een lokaal optimum. Inbouwen van deze methodes gaat te ver voor dit rapport: herhaaldelijk het leerproces uitvoeren kan ook gebruikt worden voor het voorkomen van suboptimale oplossingen.

7.2.4.2 Gebruikte model

Voor het voorspellen van de punctualiteit zijn verschillende modellen geprobeerd. Het best lerende en voorspellende model (NN₄) is een model met 7 input neuronen voor de onafhankelijke variabelen, 105 verborgen neuronen in één laag en één output neuron. De input bestaat uit sneeuwval, maximale temperatuur, windbladval, reizigers, punctualiteit t-1, punctualiteit t-7 en punctualiteit t-14. De input is dus hetzelfde als bij het MLR model. Bij het testen van andere NN-modellen is gebleken dat het gebruiken van irrelevante input zorgt voor ongewenste effecten, daarnaast convergeert een klein model ook sneller naar een goed voorspellingsmodel. Voor het instellen/leren van de parameters is voor de cases in de dataset vele malen in willekeurige volgorde een voorspelling berekend. Het aanbieden van een de dataset in willekeurige volgorde bleek betere resultaten te geven dan het aanbieden van de dataset in chronologische volgorde. Reden hiervoor kan zijn dat het model steeds aan het eind van een leercyclus het model aanpaste aan de seizoenssituatie van de laatste waarnemingen. De leersnelheid η , zoals gedefinieerd in paragraaf 4.3.5., is ingesteld op 0,2. Het NN₄ model gaf na ongeveer 48000 iteraties (1-1,5 uur) een goede voorspelling.

7.2.4.3 Modeluitbreidingen

Het model dat nu gebruikt is voor het voorspellen van de punctualiteit is een relatief eenvoudig model met één tussenlaag. Het aantal tussenlagen kan echter uitgebreid worden, en ook het aantal neuronen in de tussenlagen kan gevarieerd worden. De in PHP/MySQL geschreven klasse loopt echter tegen het probleem aan dat het opslaan van de gewichten tussen de neuronen veel tijd in beslag neemt: door restricties binnen de programmeertaal is voor ieder gewicht een aparte opdracht nodig om deze op te slaan, wat bij veel gewichten zorgt voor lange reketijden voordat de gewichten van het model bepaald zijn.

De sterk vereenvoudigde code voor het berekenen van de gewichten komt neer op het volgende:

```
<?php

for(ieder leerpunt){

    1. Bereken netwerkwaarde
    2. Bereken de fout van de netwerkwaarde ten opzicht van de gerealiseerde
       waarde

    for(ieder gewicht){

        1. Bereken de error van de neuron/het gewicht
        2. Pas het gewicht aan

        for(ieder gewicht){
            1. MySQL: sla het gewicht op
        }
    }
}
?>
```

Code 7.5

Ieder leerpunt in de bovenstaande code is één iteratie voor het aanpassen van de gewichten. De complexiteit van een algoritme of een stuk programmacode kan gegeven worden door de grote-O notatie. De grote-O notatie geeft aan in hoeverre de benodigde middelen (CPU¹⁸, geheugen, tijd, etc.) toeneemt naarmate het probleem dat opgelost moet worden groter wordt. Voor het neurale netwerkmodel is het interessant om te kijken in hoeverre het neurale netwerk uitgebreid kan worden en wat de ‘kosten’ hiervan zijn in termen van tijd en bijvoorbeeld CPU kracht. De grote-O notatie geeft aan hoe de aanspraak op middelen verandert met de veranderingen in het model. Bij een algoritme met een $O(N)$ -orde (of –tijd) neemt de benodigde tijd lineair toe met de toename van de probleemgrootte N . Wanneer de tijd benodigd voor het uitvoeren van een iteratie voor het neurale netwerk bekeken wordt, komt dat op het volgende neer (met H het aantal leerpunten en G het aantal neuronen) (Burch, 2006; Hasti, 2003):

```
for (H) {

    O(1)-tijd

    for (G) {

        O(1)-tijd

        for (G) {
            O(1)-tijd
        }
    }
}
```

Code 7.6

¹⁸ CPU: Central Processing Unit, de processor van de computer.

Alle stukken code vervangen door $O(1)$ -tijd kunnen in een tijd onafhankelijk van de variabelen H en G uitgevoerd worden. Een stijging in het aantal gewichten heeft dus geen invloed op de tijd die nodig is om deze individuele functies uit te voeren. De *for*-loops om de $O(1)$ -tijd opdrachten hebben echter wel invloed op de tijd die nodig is voor het uitvoeren van de iteratie. Aangezien H het aantal iteraties bepaalt, kan deze nu buiten beschouwing gelaten worden. Alleen het aantal gewichten G speelt nu nog een rol om de tijd nodig voor één iteratie te bepalen.

De tijd benodigd voor één iteratie, gegeven door de functie $I(G)$, is dan:

$$I(G) = O(G \cdot (1 + (G \cdot 1))) \quad (7.11)$$

Wat neerkomt op:

$$I(G) = O(G^2) \quad (7.12)$$

De tijd nodig voor het uitvoeren van één iteratie neemt dus kwadratisch toe met het aantal gewichten.

Wanneer het aantal neuronen in laag k gedefinieerd wordt als n_k , dan geldt voor een drielaagsnetwerk (input ($k = 0$), tussenlaag en output ($k = 2$)) dat het aantal gewichten G uitkomt op:

$$G = n_0 \cdot n_1 + n_1 \cdot n_2 \quad (7.13)$$

Voor iedere laag die toegevoegd wordt, komt er een term $n_k \cdot n_{k+1}$ bij, dus in het algemeen voor een netwerk met K lagen:

$$G = \sum_{k=0}^{K-1} n_k \cdot n_{k+1} \quad (7.14)$$

Met behulp van vergelijkingen (7.12) en (7.14) kan gezien worden dat een vergroting van het aantal lagen grote gevolgen heeft voor de tijd waarbinnen een iteratie uitgevoerd kan worden. Wanneer bijvoorbeeld gekozen wordt voor een vierlaagsnetwerk met 7 input neuronen, 50 en 55 neuronen in de twee verborgen lagen en één output neuron, dan neemt het aantal gewichten toe van 840 tot $7 \cdot 50 + 50 \cdot 55 + 55 \cdot 1 = 3155$ terwijl het aantal neuronen gelijk is gebleven (105 in de verborgen lagen). Gegeven vergelijking (7.12), zal de tijd om een iteratie uit te voeren kwadratisch toenemen met het aantal gewichten, wat dus neerkomt op ruim 14 keer¹⁹ zoveel tijd. Een netwerk met 105 neuronen min of meer gelijk verdeeld over drie verborgen lagen vraagt ruim 10 keer zoveel tijd, het aantal gewichten neemt immers af ten opzichte van het vierlaagsnetwerk.

Toevoegen van een tiental neuronen in de verborgen laag in het drielaagsnetwerk levert een minimale toename op naar 1,2 keer de tijd van het oorspronkelijke netwerk. Toevoegen van een input neuron inclusief de daarvoor benodigde toename volgens Kolmogorov van het aantal neuronen in de verborgen laag naar 136 zorgt voor een dubbele tijd die nodig is voor het doen van één iteratie.

Gezien de gevoeligheid van de code voor het toevoegen van neuronen (en gewichten) is het raadzaam om het model zo klein mogelijk te houden. Het toevoegen van input neuronen is mogelijk, maar zal vragen om veel tijd voor het instellen van de gewichten.

7.3 Modelprestaties

Om de modellen met elkaar te vergelijken en te beoordelen op hun waarde voor de NS zijn voor met alle modellen voorspellingen gemaakt voor de eerder genoemde geselecteerde dagen. De modellen worden vergeleken op basis van de MAPE en de MSE zoals beschreven in paragraaf 4.3. In de

¹⁹ $\left(\frac{3155}{840}\right)^2$

tabellen met voorspellingen en waarnemingen zal verwezen worden naar de data door het nummer van de datum dat weergegeven is in de eerste paragraaf van dit hoofdstuk.

7.3.1 ARIMA

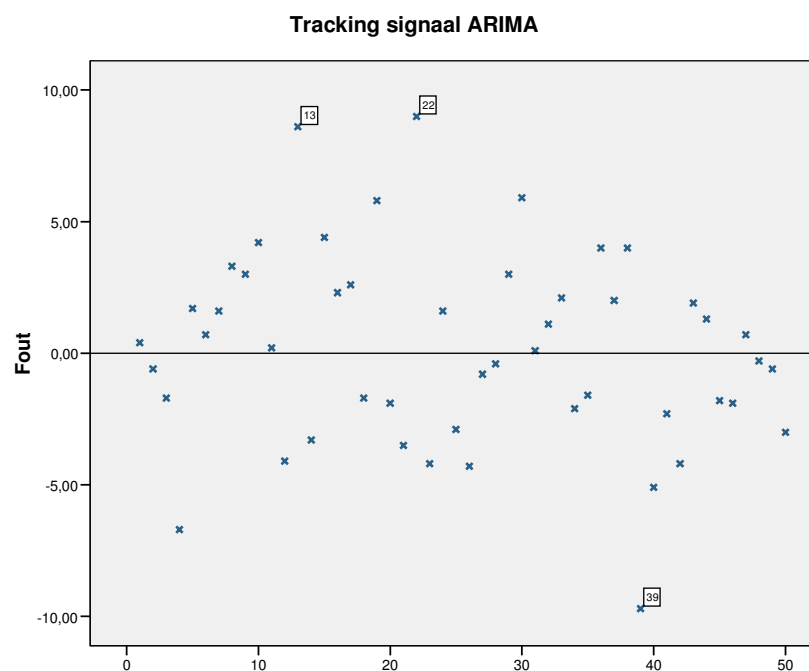
Het maken van voorspellingen met ARIMA verloopt punt voor punt. Voor ieder punt moet opnieuw bepaald worden wat de waarde van de parameters in het model moet zijn en of de vorm van het model nog voldoet. Voor het voorspellen van de validatieset is steeds een $ARIMA(1, 1, 1)(1, 1, 1)_7$ -model ($ARIMA_4$) gebruikt, dat gecontroleerd is op de Ljung-Box test ($p > 0,05$) en op de significantie van de autocorrelatie van de residuen. Daarnaast is als tracking signaal de fout van de voorspelling gebruikt. Wanneer één van deze indicatoren een waarde aanneemt die wijst op een slecht passende modelvorm, dan moet de modelvorm aangepast worden. Voor de indicatoren geldt dat een goede waarde is:

- Ljung-Box test: $p > 0,05$;
- Aantal significante autocorrelaties residuen ≈ 0 , wat neerkomt op minder dan twee significante autocorrelaties op relevante momenten (lag 1, 7, 14, etc.), en minder dan drie significante autocorrelaties op niet-relevante momenten;
- Tracking signaal: geen trend, geen (onverklaarbare) extreme fouten.

Het tracking signaal is hier vooral ter illustratie opgenomen. Een tracking signaal is voornamelijk van belang wanneer van dag tot dag voorspellingen gedaan worden met een model, het is dan zeer belangrijk om in de gaten te houden of de gekozen modelvorm nog voldoet. De waarde van het tracking signaal is in het geval van de validatieset beperkt omdat er geen voorspellingen gemaakt worden voor opeenvolgende dagen, het is daardoor onmogelijk om te zien of er een trend zit in de fouten. Het tracking signaal zal hier dan ook alleen gebruikt worden om extreme fouten te identificeren.

Voor alle cases in de validatieset is voldaan aan deze voorwaarden. Het tracking signaal is weergegeven in onderstaand figuur, de cases die genummerd zijn wijzen op verklaarbare uitschieters in de fouten. Overigens vallen deze fouten binnen de verdeling van de fouten van het basismodel $ARIMA_4$ (zie figuur 7.3).

- 13: Veel kleine storingen, uitloop werkzaamheden;
- 22: Zware weersomstandigheden, warmte;
- 39: (Extreme) bladval.



Figuur 7.4 Tracking signaal voor het $ARIMA_4$ model.

7.3.2 k-NN, NN en MLR

Het NN- en het MLR-model hebben de parameters geschat aan de hand van de dataset, met uitzondering van de cases in de validatieset. De methode k-NN heeft geen leerproces en kijkt *altijd* de complete dataset voor het moment waarvoor de voorspelling gemaakt moet worden, uitsluiten van de validatieset heeft hier dus geen zin.

Wanneer is een MSE/MAPE-waarde goed?

Wat een goede MSE is, is afhankelijk van de data die voorspeld moet worden. Een getallen voorbeeld geeft wel enig idee bij wat goed is:

Stel dat een reeks voorspeld wordt met consequent een afwijking van 0,1. De MSE is dan $(0,1)^2 = 0,01$. Wanneer de reeks voorspeld wordt met een consequente afwijking van 10, dan is de MSE $(10)^2 = 100$, een verschil van een factor 10.000!

De grootte van de te voorspellen variabele is daarnaast ook van belang: een MSE van 0,01 voor een variabele met waarden tussen de 0,001 en 0,01 is heel groot, maar een MSE van 1 voor een variabele met waarden tussen de 100.000 en 1.000.000 is heel klein. De te voorspellen punctualiteit heeft waarden tussen de 0 en 100.

Wanneer voor de punctualiteitsvoorspellingen een afwijking van 5 procentpunt als goed gezien kan worden, dan is een MSE kleiner dan $(5)^2 = 25$ in principe een goede schatting. Dat betekent echter niet dat *alle* voorspellingen binnen deze grens liggen, de MSE is immers een gemiddelde van een aantal waarnemingen, waarbij grote afwijkingen door het kwadrateren ook nog eens zwaarder wegen. De MAPE is wat dat betreft eenvoudiger te begrijpen: deze waarde geeft de directe gemiddelde procentuele afwijking van de voorspellingen. Echter, de MAPE is ook een gemiddelde.

Het is dus moeilijk te zeggen wanneer een model goed presteert door alleen naar de MSE en de MAPE te kijken, de waarden geven een goede onderlinge vergelijking van modellen, maar alleen het bekijken van de voorspellingen op detailniveau kan een oordeel geven over het model.

7.3.3 Resultaten

De modellen worden gewaardeerd aan de hand van de MSE en de MAPE waarde. Voor de validatieset scoren de modellen de MSE en MAPE waarde als aangegeven in tabel 7.13. De MSE-waarde van het k-NN model is veruit het hoogste: een waarde van 22,6 betekent dat het model een aantal grove fouten maakt die slecht gecompenseerd worden door goede voorspellingen. Momenten waarop het k-NN model slecht voorspelt zijn dagen waar externe omstandigheden als veel bladval of een zeer hoge temperatuur een grote rol spelen (case 22, 36, 37, 38; alleen een $SE > 100$). Het k-NN model geeft voor deze dagen een voorspelling die, zonder de speciale omstandigheden, niet slecht zou zijn (rond de 86-87% punctueel), echter, de omstandigheden zorgen voor een lagere punctualiteit. Een oorzaak hiervoor kan zijn dat bepaalde situaties in het verleden niet voorgekomen zijn, of dat de onafhankelijke variabelen een trendbreuk laten zien (bijvoorbeeld: een week lang 20 °C, waarna opeens 30 °C).

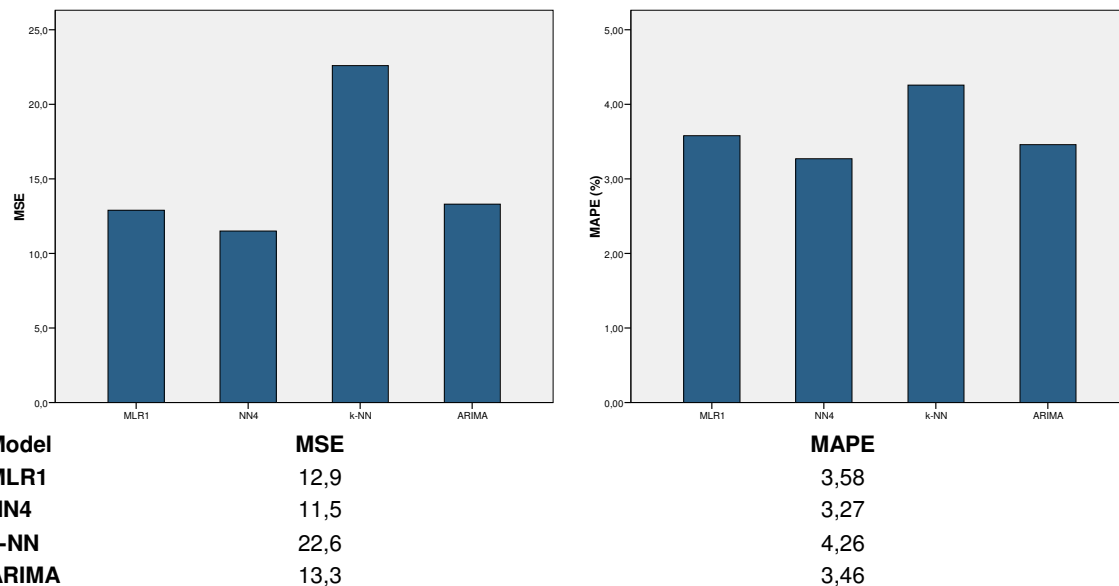
In tegenstelling tot k-NN houden de MLR en NN modellen wel rekening met de omstandigheden die gelden op de dag van de voorspelling en dit vertaalt zich in betere prestaties: MLR en NN vertonen een vergelijkbare MSE van 12,9 en 11,5 respectievelijk. In tabel 7.12 is een overzicht gegeven van de cases waarvoor de modellen een slechte voorspelling ($MSE > 25$) gaven. Opvallend is dat voor bijna alle cases een slechte voorspelling door beide modellen tegelijkertijd gedaan werd. Voor dagen met weinig bijzonderheden, voorspellen beide modellen regelmatig te laag. Dagen met veel verstoringen, en met name verstoringen die gerelateerd zijn aan de infrastructuur of het materieel, worden te hoog voorspeld. Beide modellen hebben geen parameters die rekening houden met infrastructuurverstoringen of materieeldefecten, de modellen voorspellen punctualiteitsdalingen door deze oorzaken dan ook niet. De ARIMA methode, die wat informatie over de te voorspellen dag evenveel meeneemt als k-NN, realiseert een MSE van 13,3. Deze score plaatst de methode dicht bij MLR en NN. Het is moeilijk om te onderscheiden waarom ARIMA een slechte voorspelling geeft. Een mogelijke oorzaak voor een slechte voorspelling is een toevallige combinatie van een slechte punctualiteit op $t-1$ en $t-7$ (wat een lage voorspelling zal opleveren) en een goede punctualiteit op t .

Case	Datum	Situatie	MLR	NN
12	16-5-2006	Weinig bijzonderheden	3	3
19	12-6-2006	Veel wisselstoringen	1	1
22	5-7-2006	Zware weersomstandigheden	1	
25	8-8-2006	Weinig bijzonderheden	3	4
26	10-8-2006	Weinig bijzonderheden	3	3
30	25-6-2006	Stroomstoring Ut	1	1
34	28-10-2006	Bladval, weinig bijzonderheden	3	4
36	12-11-2006	Veel storingen	1	1
37	19-11-2006	Bladval, defecte treinen	2	1
38	24-11-2006	Bladval, aanrijdingen, defecte treinen	1	1
39	29-11-2006	Bladval, defecte treinen	1	1
42	4-1-2007	Weinig bijzonderheden	3	4

Legenda

1	MSE > 25, voorspelling te hoog
2	MSE > 10 (als van de andere methode de MSE > 25), voorspelling te hoog
3	MSE > 25, voorspelling te laag
4	MSE > 10 (als van de andere methode de MSE > 25), voorspelling te laag

Tabel 7.12 Slechte voorspellingen MLR en NN



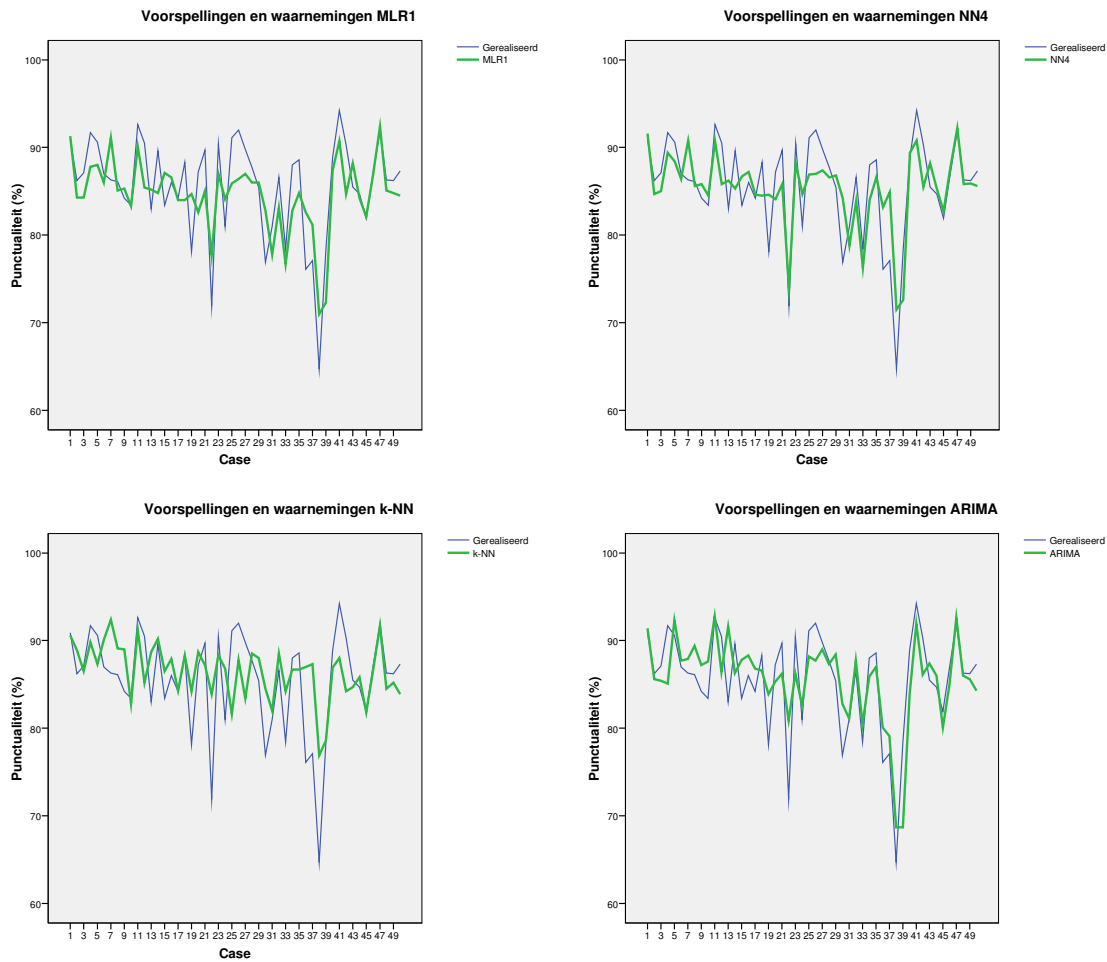
Tabel 7.13 MSE en MAPE waarde voor de modellen.

In figuur 7.5 is voor iedere methode het verloop van de voorspellingen en de waarnemingen weergegeven. Het is van belang om bij deze te figuren te zien dat de cases elkaar wel opvolgen, maar dat er in werkelijkheid geen directe opeenvolging is, er zitten immers (extra) dagen tussen cases. Wanneer een punctualiteit tussen de 80 en 90 als 'normaal' beschouwd wordt, dan kan alles daaronder als 'dal', en alles daarboven als 'piek' benoemd worden. Deze grenzen zijn natuurlijk arbitrair, maar maken het mogelijk om de uitschieters in de grafieken te bekijken.

Allereerst vallen twee dalen op: case 22 en case 36. MLR, NN en ARIMA voorspellen voor deze cases een lagere punctualiteit. Bij de pieken vallen cases 11, 26, 41 en 47 op. Voor deze pieken geldt dat case 47 door alle modellen (ook k-NN) exact voorspeld wordt, cases 11 en 41 worden redelijk voorspeld door MLR, NN en ARIMA. Case 26 wordt slecht voorspeld door alle methodes.

Validatieset voorspellingen						
Case	Gerealiseerd	MLR ₁	NN ₄	k-NN	ARIMA ₄	G{NN, MLR}
1	90,9	91,3	91,6	90,6	91,4	91,5
2	86,2	84,3	84,7	88,9	85,6	84,5
3	87,1	84,3	85,0	86,5	85,4	84,7
4	91,7	87,8	89,4	89,8	85,1	88,6
5	90,6	88,0	88,4	87,3	92,3	88,2
6	87,0	86,0	86,4	90,1	87,7	86,2
7	86,3	91,1	90,8	92,4	87,9	87,3
8	86,1	85,1	85,6	89,1	89,4	85,4
9	84,2	85,3	85,8	89,0	87,2	85,6
10	83,4	83,3	84,5	82,8	87,6	83,9
11	92,6	90,1	90,7	91,2	92,8	90,4
12	90,5	85,4	85,8	85,2	86,4	85,6
13	83,0	85,2	86,2	88,7	91,6	85,7
14	89,6	84,8	85,3	90,2	86,3	85,1
15	83,4	87,1	86,7	86,5	87,8	86,9
16	86,0	86,6	87,2	87,9	88,3	86,9
17	84,2	84,0	84,6	84,3	86,8	84,3
18	88,3	84,0	84,5	88,3	86,6	84,3
19	78,1	84,7	84,6	84,2	83,9	84,7
20	87,2	82,6	84,1	88,7	85,3	83,4
21	89,7	85,0	85,8	87,2	86,2	85,4
22	71,9	77,7	74,0	83,9	80,9	75,9
23	90,4	87,0	88,2	88,3	86,2	87,6
24	81,0	84,1	84,7	86,8	82,6	84,4
25	91,1	85,9	86,9	81,6	88,2	86,4
26	92,0	86,4	87,0	87,8	87,7	86,7
27	89,8	87,0	87,4	83,4	89,0	87,2
28	87,7	86,0	86,6	88,5	87,3	86,3
29	85,4	86,0	86,8	88,0	88,4	86,4
30	76,9	82,8	84,2	84,5	82,8	83,5
31	81,0	77,8	78,8	82,0	81,1	78,3
32	86,6	83,1	83,9	88,5	87,7	83,5
33	78,4	76,7	76,3	84,2	80,5	76,5
34	88,0	82,8	84,0	86,7	85,9	83,4
35	88,6	84,8	86,6	86,7	87,0	85,7
36	76,1	82,6	83,2	87,0	80,1	82,9
37	77,1	81,2	84,9	87,3	79,1	83,1
38	64,7	71,0	71,5	76,9	68,7	71,3
39	78,4	72,3	72,6	78,6	68,7	72,5
40	88,9	87,4	89,4	86,9	83,8	88,4
41	94,2	90,7	90,8	88,0	91,9	90,8
42	90,3	84,7	85,5	84,2	86,1	85,1
43	85,5	88,2	88,2	84,7	87,4	88,2
44	84,7	84,2	85,3	85,8	86,0	84,8
45	81,9	82,1	82,7	81,8	80,1	82,4
46	87,0	86,8	87,6	86,5	85,1	87,2
47	91,9	92,4	92,2	91,7	92,6	92,3
48	86,3	85,1	85,8	84,5	86,0	85,5
49	86,2	84,8	85,9	85,2	85,6	85,4
50	87,3	84,5	85,6	83,9	84,3	85,1
MSE		12,9	11,5	22,6	13,3	11,9
MAPE		3,58	3,27	4,26	3,46	3,39

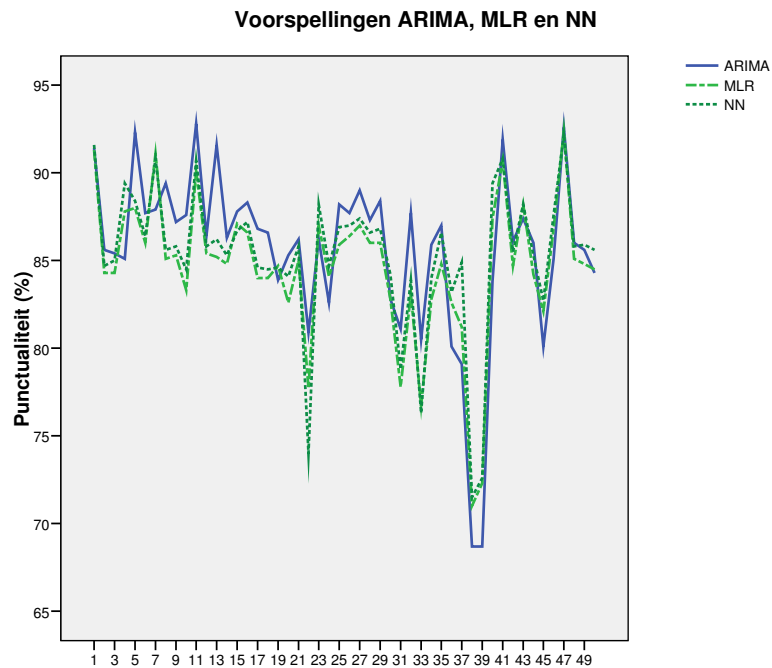
Tabel 7.14 Waarnemingen en voorspellingen van de verschillende modellen voor de validatieset



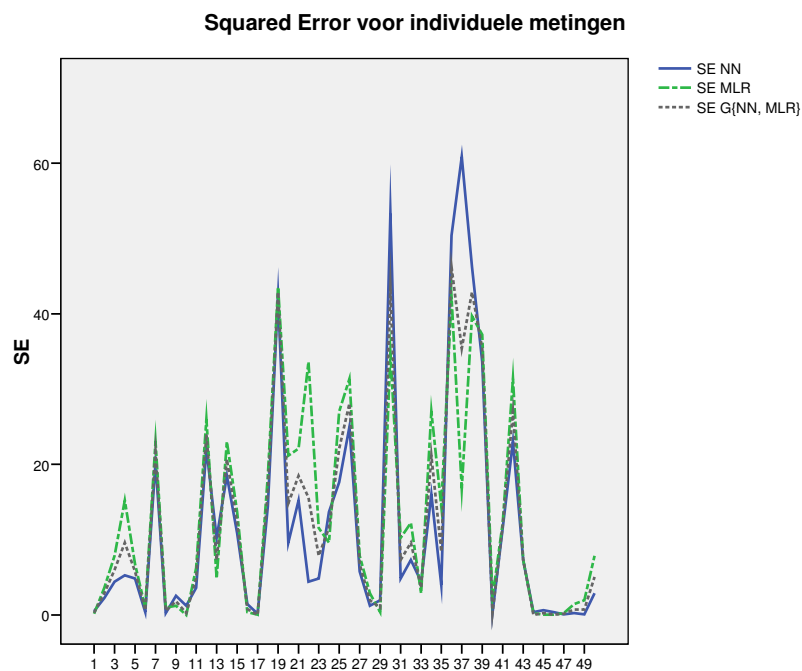
Figuur 7.5 Voorspellingen en waarnemingen voor de verschillende methodes. Grotere versies van deze figuren staan in bijlage A.

Wanneer de methodes MLR, NN en ARIMA nog direct met elkaar vergelijken (zie figuur 7.6), valt op dat de methodes NN en MLR zeer gelijkende voorspellingen geven. De benadering van de MLR-vergelijking door NN is dus zeer goed: de prestaties ten opzichte van de werkelijkheid valt voor NN zelfs iets positiever uit.

Een mogelijkheid om de voorspellingen beter te maken is het combineren van verschillende methodes doormiddel van een gemiddelde of een gewogen gemiddelde. Wanneer de beste twee methodes (NN en MLR) gecombineerd worden met een direct gemiddelde ($G\{NN, MLR\}$), heeft deze methode een MSE van 11,9 en een MAPE van 3,39. Hoewel dit een stijging is van de MSE en MAPE ten opzichte van de beste methode, is het nemen van een gemiddelde waardevol. In figuur 7.7 kan gezien worden dat de SE van de individuele voorspellingen van NN niet continu onder die van MLR ligt. Door gebruik van het gemiddelde worden de goede voorspelmomenten van NN gecombineerd met de slechte van MLR, en andersom, waardoor een betere voorspelling ontstaat.

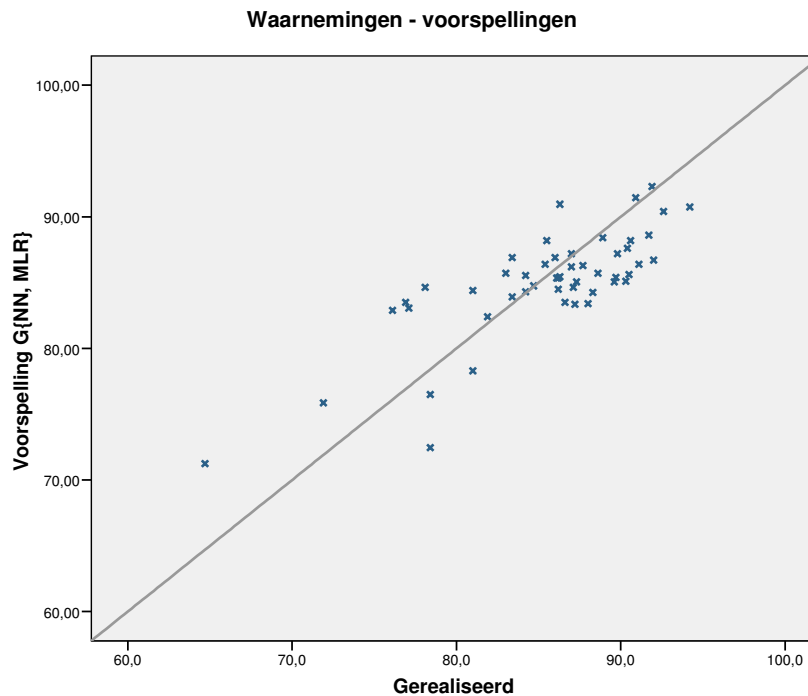


Figuur 7.6 Vergelijking van de beste drie voorspellingsmethoden.



Figuur 7.7 Squared Error (SE) voor NN, MLR en $G\{NN, MLR\}$.

Een plot met de waarde van de daadwerkelijke punctualiteit uitgezet tegen de voorspelling geeft een goed beeld van de kwaliteit van de voorspelling. In het onderstaande figuur is te zien dat de meeste voorspellingen dicht bij de (grijze) lijn voorspelling = waarneming liggen.



Figuur 7.8 Vergelijking voorspellingen – waarnemingen. De grijze lijn is voorspelling = waarneming.

Verder is het interessant om te zien dat de voorspellingen een gemiddelde van 85,0% hebben, en de waarnemingen een gemiddelde van 85,5%. Dit resulteert in een bias van:

$$bias = \frac{\bar{\hat{Y}}}{\bar{Y}} = \frac{85,0}{85,5} = 0,99 \quad (7.15)$$

Een bias van één is een perfecte score, de bias van het model is dus minimaal.

7.4 Conclusie

De gebouwde modellen vallen in twee categorieën uiteen: modellen die wel rekening houden met de factoren die invloed hebben op de punctualiteit en modellen die dat niet doen. De modellen die wel rekening houden met de onafhankelijke factoren op de dag van de voorspelling zijn NN en MLR. Uit de MSE/MAPE waarden van de modellen komt naar voren dat de NN (MSE 11,5) en MLR (MSE 12,9) modellen beter presteren dan de ARIMA (MSE 13,3) en k-NN (MSE 22,6) modellen. Een belangrijke reden waarom de modellen een relatief hoge fout hebben is de onmogelijkheid van de modellen om infrastructuur- en materieelstoringen te voorspellen. De modellen zijn wél in staat om de invloed van de geselecteerde onafhankelijke variabelen te bepalen.

Voor verder gebruik worden de MLR en NN modellen geselecteerd. Van de vier modellen scoren deze het beste op de MSE en MAPE, en daarnaast verdienen de methodes extra waardering voor hun eenvoud. Zowel het MLR als het NN model zijn erg gebruiksvriendelijk: de input van de verwachte waarden van de onafhankelijke variabelen volstaat voor een voorspelling. Ook heeft het MLR model een relatief hoge ‘werkelijkheids’-waarde: het model is eenvoudig te begrijpen zonder kennis te hebben van ingewikkelde wiskundige concepten. Het NN model scoort op dat vlak slechter, maar heeft grote voordeel dat het model zelflerend is.

De combinatie van beide modellen: G{NN, MLR} geeft een MSE van 11,9 en een bias van 0,99. In het opvolgende hoofdstuk zal verder gekeken worden wat de mogelijkheden zijn van dit model.

8 Toepassingen

Het beste voorspellingsmodel, $G\{NN, MLR\}$, heeft op de validatieset een MSE van 11,9, een gemiddelde afwijking (MAPE) van 3,39 en een bias van 0,99. Deze statistieken duiden, in combinatie met de analyse uit het vorige hoofdstuk, op een model dat in staat is om de invloeden van de geïdentificeerde invloedrijke onafhankelijke factoren op de punctualiteit te voorspellen. Echter, het model is niet in staat om de invloed van materieeldefecten en infrastructuurstoringen op de punctualiteit te voorspellen, wat een verklaring kan zijn voor de relatief hoge MSE.

Dit hoofdstuk zal verder ingaan op de materieeldefecten en infrastructuurstoringen en de relatie tot (de voorspelling van) de punctualiteit. Daarnaast zal dit hoofdstuk beschrijven hoe de modellen gebruiksvriendelijk gemaakt zijn doormiddel van een webapplicatie. Vervolgens zal de waarde van dit model voor de NS uiteengezet worden.

8.1 Materieeldefecten en infrastructuurstoringen

Zoals eerder beschreven hebben materieeldefecten en infrastructuurstoringen (kortweg ‘storingen’) effect op de kwaliteit van de voorspelling: het model is onvoldoende in staat om het ontstaan en het effect van storingen te voorspellen. Kwantificeren van deze invloeden is zeer complex: storingen zijn moeilijk met elkaar te vergelijken. Een trein met remproblemen op de lijn Zwolle-Kampen heeft landelijk geen enkele invloed omdat deze lijn geïsoleerd is. Een wisselstoring op Utrecht Centraal heeft enorm veel invloed: 29 treinseries komen door Utrecht. Maar welke treinseries invloed hebben op welke andere treinseries is weer onduidelijk. Een onderzoek naar de invloed op de punctualiteit van de storingen op bepaalde trajecten gaat te ver voor dit onderzoek, maar zou zeker een interessant vervolgonderzoek zijn.

Voor nu is het in ieder geval belangrijk om te constateren dat veel (of invloedrijke) storingen zorgen voor een verlaging van de punctualiteit ten opzichte van wat verwacht mag worden op basis van de weersomstandigheden en het reizigersaantal. Weinig (of invloedarme) storingen resulteert in een verhoging van de punctualiteit. Op basis van deze waarneming kan dus gesteld worden dat er een ‘standaardniveau’ is van storingen en punctualiteit. Wat dit niveau is kan gezien worden op dagen dat het model goed voorspelt, maar een harde uitspraak doen over deze storingen is moeilijk door het eerder beschreven probleem van invloedrijke en –arme storingen. Voor dag-tot-dag analyses van de impact van storingen op de punctualiteit kan dit model wel gebruikt worden. Het model wordt dan niet gebruikt als voorspellingsmodel, maar als referentiemodel voor de verwacht punctualiteit en de impact van storingen die dag.

8.2 Gebruik van het model

Een voorspellingsmodel moet continu gecontroleerd worden op de kwaliteit van de voorspellingen die het doet, maar niet iedereen is in staat om dit uit te voeren. Om het voorspellingsmodel eenvoudig toegankelijk te maken voor iedereen is een (web)applicatie gebouwd. Het voordeel van de applicatie is dat door middel van het invullen van enkele velden en een druk op de knop een voorspelling gedaan kan worden. Verder worden de voorspellingen opgeslagen, zodat in één oogopslag door een expert ingeschat kan worden of het model nog voldoet aan de eisen. In de hierop volgende paragrafen zal beschreven worden hoe de webapplicatie gebruikt kan worden voor het maken van voorspellingen.

8.2.1 Workflow

Om het voorspellingproces zo eenvoudig mogelijk te maken is de applicatie gebaseerd op een vaste procedure (workflow). De applicatie ‘praat’ de gebruiker min of meer door het proces heen. Wanneer de gebruiker de applicatie benadert, wordt een overzicht gegeven van de globale stappen die de voorspeller moet nemen om tot een voorspelling te komen.



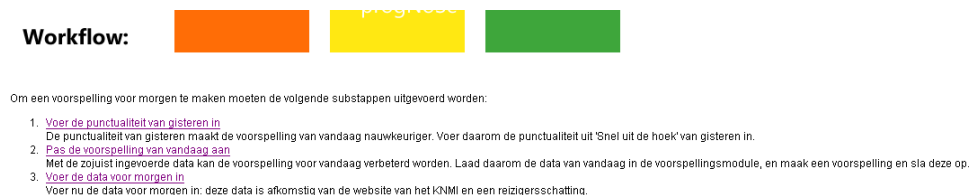
Figuur 8.1 Openingsscherm ProgNoSe

Wanneer de gebruiker op ‘verzamel data’ klikt, krijgt de gebruiker de locaties te zien waar de informatie opgezocht moet worden. Deze locaties bestaan uit ‘Snel uit de hoek’, een dagelijks infomail met daarin de punctualiteit van gisteren, en de website van het KNMI waar weerdata voor de toekomst, en van het verleden gevonden kan worden.



Figuur 8.2 Eerste stap voor het maken van een voorspelling

Wanneer de gebruiker de data bij de hand heeft, kan geklikt worden op ‘zet data in progNoSe’. De gebruiker krijgt dan de informatie te zien in het onderstaande figuur. Aangezien vandaag de punctualiteit van gisteren bekend geworden is, kan deze toegevoegd worden aan de database (1). Vervolgens kan de voorspelling voor vandaag nauwkeuriger gemaakt worden met behulp van de ingevoerde punctualiteit van gisteren (immers, de punctualiteit van gisteren is een ‘onafhankelijke variabele’ voor het $G\{NN, MLR\}$ -model). De voorspelling wordt nauwkeuriger omdat er al een voorspelling gemaakt is voor vandaag op $t-1$, de nieuwe voorspelling neemt de gerealiseerde punctualiteit van gisteren mee in plaats van de voorspelde punctualiteit voor gisteren. Na de voorspelling voor vandaag verbeterd te hebben, kunnen de gegevens voor morgen ingevoerd worden.



Figuur 8.3 Tweede stap voor het maken van een voorspelling

Vervolgens kan de voorspelling voor morgen gemaakt worden (zie onderstaande figuur), de gebruiker gaat hiervoor naar de laatste stap. Als laatste substap moeten de gegevens van gisteren nog gecorrigeerd worden met de gerealiseerde (weer)data. Hiervoor kunnen de historische weergegevens van het KNMI gebruikt worden. Het bijwerken van de geprognosticeerde weerdata naar de

gerealiseerde weerdata is van belang om in het in de toekomst mogelijk te maken een nieuw model te definiëren.

Workflow:



1. [Maak de voorspelling voor morgen](#)

Nu de data ingevoerd is, kan de voorspelling voor morgen gemaakt worden. Laad de data van vandaag, en maak de voorspelling. Sla de voorspelling ook op.

2. [Corrigeer data van gisteren](#)

Om de data in de database consistent te houden, moet de data van gisteren gecorrigeerd worden met de data die nu (zeker) bekend is. Gebruik hiervoor de pagina van het KNMI (» Klimatologie/Verleden weer » Gegevens in tabelvorm per station » De Bilt & gewenste dag selecteren)

Figuur 8.4 Laatste stappen voor het maken van een voorspelling in ProgNose

8.2.2 Jaarprognose

Zoals eerder vermeld maakt de NS afspraken met de overheid over onder andere de gemiddelde punctualiteit die de NS wil realiseren. Voor 2007 is de doelstelling vastgelegd op 86,8%. De applicatie kan naast de 24-uurvoorspellingen ook gebruikt worden voor een prognose op jaarbasis. De werkwijze van de applicatie is in principe eenvoudig. Voor de jaarprognose gebruikt de applicatie alle punctualiteitgegevens die al beschikbaar zijn in een jaar, en maakt daarbij prognoses voor de dagen die nog niet bekend zijn. Deze methode neemt meer informatie mee dan de prognose die nu gebruikt wordt op basis van de gegevens die tot nu toe beschikbaar zijn, in combinatie met de gegevens van de resterende periode vorig jaar. Dus op 15 augustus 2007 worden de gegevens t/m 14 augustus van 2007 plus de gegevens van 15 augustus t/m 31 december gebruikt voor het maken van een prognose voor de jaarpunctualiteit.

Een probleem bij het prognosticeren op lange termijn met het $G\{NN, MLR\}$ -model is het ontbreken van voorspellingen voor het weer. Prognoses voor reizigersaantallen zijn wel bekend. Om toch een voorspelling te kunnen maken voor jaar j , gebruikt de applicatie de gemiddelde weerdata van de voorgaande twee jaar, waarvan verondersteld wordt dat dit een goede schatting is voor het huidige jaar. Mochten er extremen voorkomen in de weerdata, dan worden deze gesignaleerd en uitgesloten. Hiervoor wordt de weerdata van jaar $j-1$ vergeleken met de weerdata van jaar $j-2$, en wanneer deze meer dan 50% van elkaar verschillen wordt de meest op de vorige waarneming gelijkende waarde gebruikt, en niet het gemiddelde.

Voor het maken van een jaarprognose selecteert de gebruiker de module 'jaarprognose'. De gebruiker ziet dan het onderstaande beeld. De gebruiker kan vervolgens een verwachte groei (in absolute aantallen reizigers of in reizigerkilometers) aangeven, en eventueel week- of jaarpatronen wijzigen. Vervolgens klikt de gebruiker op 'maak voorspelling' om de voorspelling te laten berekenen.

Jaarprognose

Voor het maken van een jaarprognose wordt de reeds bekende punctualiteitsdata tot vandaag gebruikt van dit jaar, plus geprognosticeerde punctualiteitsdata voor de rest van het jaar. Om de prognoses te kunnen maken, wordt u gevraagd de week en jaarpatronen van reizigers in te vullen. De overige data wordt uit de database gehaald. Wanneer u de data ingevuld heeft, kunt u een voorspelling maken.

Verwachte reizigersgroei

In percentage groei (reizigers en/of reizigerkilometers)

Groei:

Weekpatroon reizigers

Maandag:	242
Dinsdag:	250
Woensdag:	250
Donderdag:	251
Vrijdag:	252
Zaterdag:	166
Zondag:	155

Jaarpatroon reizigers

Januari:	1.198
Februari:	1.119
Maart:	1.280
April:	1.207
Mei:	1.290
Juni:	1.240
Juli:	1.023
Augustus:	1.101
September:	1.339
Oktober:	1.438
November:	1.312
December:	1.151

Voorspelling

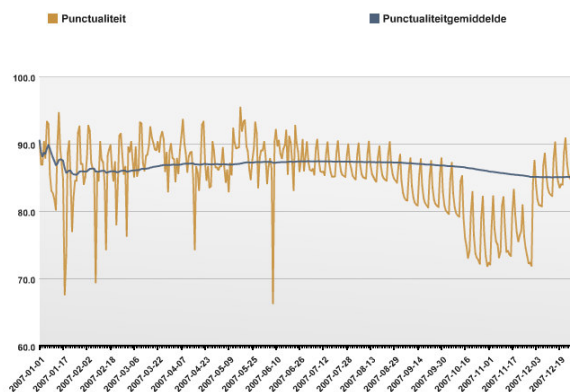
[Maak voorspelling](#)

Figuur 8.5 Instellen van de parameters voor de jaarprognose

De output van de jaarprognose levert een getal (de daadwerkelijke prognose) en een grafiek met daarin de waarnemingen tot nu toe, de prognoses per dag en de ontwikkeling van de gemiddelde punctualiteit. Voor 2007 komt de prognose op 85,1%, bij een reizigersgroei van 1%. In de grafiek is duidelijk de invloed van de herfst en de veranderende reizigersaantallen door het jaar te zien.

Voorspelling jaarpunctualiteit: 85.1%

! data755.xml verwijderd
! data410.xml verwijderd



Figuur 8.6 Resultaat van de jaarprognose

8.2.3 Overige mogelijkheden van de applicatie

Het is van belang dat een voorspellingsmodel gecontroleerd wordt op de kwaliteit van de voorspellingen. Hiervoor zijn twee middelen beschikbaar:

1. Grafiek met voorspelde en waargenomen punctualiteit

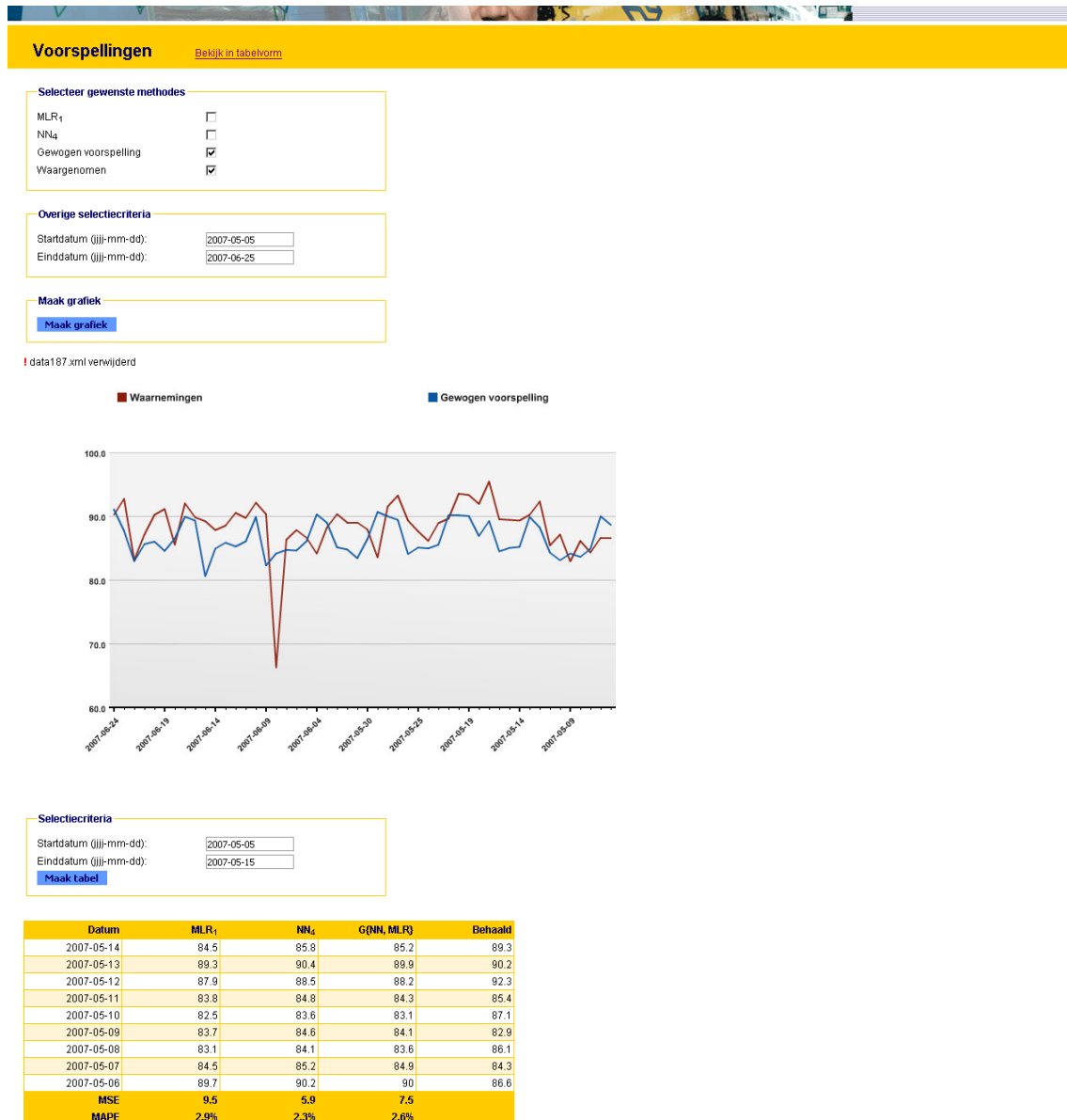
De grafiek laat in één oog opslag zien hoe de voorspellingen zich verhouden tot de waarnemingen. In het onderstaande figuur is bijvoorbeeld te zien dat op 8 juni een grote

misser geweest is. Deze misser werd veroorzaakt door twee stroomstoringen op twee grote knooppunten.

2. Tabel met voorspelde en waargenomen punctualiteit

Naast de grafiek kan ook in een tabelvorm gekeken worden naar de data. De tabelvorm geeft ook de MSE en MAPE waarde voor de geselecteerde periode.

Een statistisch onderlegd persoon kan aan de hand van de MSE en MAPE waarde en de grafiek inschatten of het model toe is aan een update. Belangrijk hierbij is dat goed gekeken wordt naar de oorzaak van een afwijkende voorspelling. Het is 'normaal' dat het model een te hoge punctualiteit voorspelt als er veel infrastructuurstoringen en materieeldefecten zijn. Ook kan het model een te lage punctualiteit voorspellen als er juist weinig storingen zijn.



Figuur 8.7 Voorspellingen en waarnemingen (grafiekvorm en tabelvorm)

8.3 Toepassing van het model

In hoofdstuk 4 is beschreven wat de doelen waren waarbij het model behulpzaam kan zijn. De doelen worden hier kort herhaald:

1. Benchmarking van de eigen prestaties;
2. Verhogen van de dienstverlening;
3. Aanpassen bijsturingscapaciteit;
4. Waarschuwen van klanten voor ‘rampzalige’ dagen;
5. Middel voor onderhandelingen.

Gezien de prestaties van het model kan de geschiktheid van het model voor ieder van deze doelen bepaald worden.

Het model is vooral interessant voor dagen dat het slecht²⁰ gaat aangezien dat de dagen zijn waarop NS klappen oploopt voor de doelstellingen.

8.3.1 Benchmarking (1) en onderhandelingsmiddel (5)

Het is voor de NS belangrijk om een vergelijkingswaarde te hebben voor de punctualiteit die zij behaalt. De NS wordt over het algemeen vergeleken met bedrijven in het buitenland zoals de Zwitserse spoorwegen (SBB). Echter, de meetmethoden, invloeden op de punctualiteit en situaties op het netwerk zijn niet gelijk voor alle spoorbedrijven. De vergelijking NS – SBB kan wel gemaakt worden, maar geeft niet per definitie een eerlijk beeld.

Een alternatief is om de NS te vergelijken met de NS in het verleden. Het model doet dit min of meer: de voorspellingen zijn gebaseerd op prestaties uit het verleden. De voorspelling van het model geeft dus aan welke punctualiteit de NS zou moeten behalen bij de gegeven weersomstandigheden, gegeven de prestaties in het verleden en de huidige situatie. Scoort de NS structureel beter dan de voorspelling, dan kan gesteld worden dat de NS op de punctualiteit goed presteert.

Naast het beoordelen van de eigen prestaties kan het model op een vergelijkbare manier gebruikt worden voor onderhandelingen met bijvoorbeeld de overheid. Door jaarschattingen voor de onafhankelijke variabelen te gebruiken kan een schatting gemaakt worden voor de gemiddelde punctualiteit voor het hele jaar. Hiervoor moeten 365 voorspellingen gemaakt worden voor een heel jaar, waarmee de gemiddelde punctualiteit berekend kan worden. De NS kan hiermee de invloed van stijgende reizigersaantallen laten zien: de eis die op dit moment aan de NS gesteld wordt om én 5% reizigersgroei te realiseren, én de punctualiteit te verhogen kan op basis van het model als onmogelijk bestempeld worden: hogere reizigersaantallen zorgen (bij een verder gelijkblijvende situatie) voor een lagere punctualiteit²¹. Wil de NS naast reizigersgroei een verhoging van de punctualiteit realiseren, dan zal de relatie tussen de punctualiteit en het aantal reizigers verzwakt moeten worden, bijvoorbeeld door efficiëntere in- en uitstapprocessen of door een uitbreiding van de capaciteit van en op het netwerk.

8.3.2 Waarschuwen van reizigers (4)

De NS heeft met de overheid, naast afspraken over de punctualiteit, afspraken gemaakt over de ‘informatie bij ontregelingen’ (IBO). Graadmeters hiervoor zijn de cijfers die reizigers de NS geven over de aspecten van IBO. Een manier om de reiziger beter te informeren over ontregelingen zou het op voorhand waarschuwen over verwachte punctualiteitsdalingen kunnen zijn. Het gebruiken van dit model voor dit soort waarschuwingen zou moeten gebeuren met grote voorzichtigheid. Het model voorspelt zelden dalen als deze niet voorkomen, maar het model mist wel regelmatig een dal (bijvoorbeeld case 19 uit de validatieset). De opinie van de reiziger kan door dit soort gemiste dalen een extra deuk oplopen. Reizigers hebben de neiging om negatieve ervaringen beter te onthouden van positieve ervaringen, een gemist dal zal dan ook beter onthouden worden dan alle keren dat een dal

²⁰ ‘Slecht’ is een ruim begrip: het is geheel afhankelijk van de afspraken die de NS gemaakt heeft met de overheid. Over het algemeen kan gesteld worden dat een punctualiteit onder de 80% als echt slecht beoordeeld wordt.

²¹ Een stijging van het aantal reizigers met 5 procent zorgt voor een daling van de punctualiteit (ceteris paribus, op dagniveau) met ongeveer 0,2 procentpunt.

wel voorspeld was. Het risico op meer negatieve effecten van de waarschuwingen dan positieve is dus relatief groot.

Om de accuraatheid van het model voor dit doeleinde te bepalen zou het model een periode achter de schermen gebruikt kunnen worden voor het informeren van rijdend personeel. Op dagen dat een lage punctualiteit voorspeld wordt kan het rijdend personeel dan extra gericht zijn op het informeren van de reiziger. Wanneer na deze periode blijkt dat de juiste dagen een dal voorspeld werd, kan overwogen worden de punctualiteitswaarschuwing publiek te maken.

8.3.3 Verhogen dienstverlening (2)

Samenhangend met het vorige doel, kunnen ook drastischer maatregelen genomen worden wanneer de voorspelling van de punctualiteit een dal vertoont. Op 8 februari 2007 verminderde de NS de treinenloop en vergrootte de NS het inzet van personeel om de reiziger zo vlot mogelijk door de beperkte dienstregeling te loodsen. Reden hiervoor was het afgegeven weeralarm van het KNMI. Hoewel de punctualiteit op deze dag alsnog een (grote) dip liet zien, kan het goed zijn dat de dienstverlening naar reizigers wel als goed beoordeeld is. De treinen die reden, konden door de beperkte dienstregeling ook echt rijden, en werden niet gehinderd door vollopende stations en defecte treinstellen.

Een dergelijke beperking van de dienstregeling is een ingrijpende maatregel die alleen bij extreme omstandigheden gebruikt moet worden. Extreme weersomstandigheden zullen door het KNMI voorspeld worden met een weeralarm, het is voor de NS logischer om daarop te reageren in plaats van een model dat de voorspellingen van het KNMI weer interpreteert. Het is dus niet direct wenselijk het voorspellingsmodel te gebruiken voor ingrijpende wijzigingen in de dienstregeling naar aanleiding van voorspellingen van een lage punctualiteit.

8.3.4 Aanpassen bijsturingcapaciteit (3)

Bijsturingcapaciteit kan op dagen met veel vertragingen de beperkende factor bij het afhandelen van de dienstregeling zijn. Op het moment dat een (reizigers)trein veel²² vertraging oploopt moet het pad van deze trein opnieuw ingelegd worden, wat betekent dat de trein op andere tijdstippen een baanvak toegewezen krijgt dan in de dienstregeling staat (herspecificeren). Het herspecificeren wordt gedaan door netwerkbesturing, een afdeling die valt onder ProRail, en wordt niet uitgevoerd door ARI. Het vraagt dus extra personele capaciteit op het moment dat er meer vertragingen zijn.

Een lagere voorspelde punctualiteit kan ook duiden op een groter aantal calamiteiten. Echter, deze relatie is niet erg hard. Calamiteiten leiden over het algemeen tot het uitvallen van treinen, en uitgevallen treinen tellen niet meer mee in de punctualiteit. Er bestaan ook calamiteiten waarbij de treindienst zoveel mogelijk in stand gehouden wordt, een goed voorbeeld hiervan is een wisselstoring waarbij een gedeelte van een station onbereikbaar wordt. Op dat moment blijft de treindienst in stand, maar moeten treinen soms wachten tot er een leeg perron ontstaat op het station.

Bij uitgebreide calamiteiten waarbij de treindienst ontregeld raakt, is er veel capaciteit nodig om de afhandelingsscenario's te kunnen uitvoeren. Een probleem waar regelmatig tegenaan gelopen wordt is dat het afhandelingsscenario niet haalbaar is, bijvoorbeeld omdat er onvoldoende bussen beschikbaar zijn bij de buspartner. Bij een voorspelde lagere punctualiteit kan hierop ingespeeld worden: de buspartner kan vooraf gewaarschuwd worden om een voldoende aantal oproepbare bussen klaar te hebben staan. Daarnaast zou de hoeveelheid beschikbaar personeel binnen de NS dat de calamiteiten afhandelt vergroot kunnen worden wanneer een dip in de punctualiteit verwacht wordt.

De betrouwbaarheid van het model zou ook hiervoor eerst getest moeten worden. Bijvoorbeeld het oproepen van extra personeel en het 'reserveren' van bussen zal extra geld kosten. Het is wenselijker om de afhandelingsscenario's en afspraken met (bus)partners nauwkeurig te bekijken en deze waar mogelijk te stroomlijnen. Het model kan vervolgens gebruikt worden om op de dagen met een voorspelde lagere punctualiteit (extra) actie te ondernemen.

²² De grens voor reizigerstreinen verschilt en is niet vastgelegd, voor goederentreinen ligt de grens landelijk op 4 minuten vertraging.

8.4 Conclusie

Gezien de prestaties van het voorspellingsmodel ($G\{NN, MLR\}$) kan het model in beperkte mate gebruikt worden voor het plannen van bedrijfsprocessen. De in hoofdstuk 4 genoemde doelen voor het voorspellingsmodel zijn in dit hoofdstuk als volgt beoordeeld:

1. Benchmarking van de eigen prestaties
Het model geeft inzicht hoe de NS zou moeten presteren (gezien de prestaties uit het verleden) bij een gegeven situatie. Wanneer de impact van storingen een gemiddeld niveau heeft, en de NS presteert op of boven de voorspelde punctualiteit, dan kan gesteld worden dat de NS een goede prestatie heeft geleverd. Wanneer de NS structureel niet in staat is om de voorspelde punctualiteit te behalen, kan gesteld worden dat de NS slecht presteert op de indicator punctualiteit.
2. Verhogen van de dienstverlening
Weersomstandigheden waardoor de punctualiteit sterk daalt worden door het KNMI aangekondigd, het zijn immers eenvoudige factoren als bladval (in combinatie met de windsnelheid) en sneeuwval. De NS kan met behulp van het model eventueel wel eventuele combinaties van factoren die leiden tot een verlaging van de punctualiteit signaleren (bijvoorbeeld veel reizigers en een hoge temperatuur). Direct gebruik van het model om de dienstregeling drastisch aan te passen is niet aan te raden: hoewel het model niet vaak te laag voorspelt, is het model niet betrouwbaar genoeg om dit soort beslissingen alleen hierop te baseren.
3. Aanpassen bijsturingscapaciteit
Een belangrijk aspect bij het afhandelen van calamiteiten is de bijsturingscapaciteit. De bijsturingscapaciteit bestaat bijvoorbeeld uit operationeel managers die de leiding hebben bij de afhandeling van een calamiteit, maar ook uit beschikbaar vervangend vervoer. Het model kan goed gebruikt worden om voorzorgsmaatregelen te nemen voor slechte dagen.
4. Waarschuwen van klanten voor ‘rampzalige’ dagen
Om het model hiervoor te gebruiken zou het model (en de methode van voorspellen) enige tijd ‘stil’ gebruikt moeten worden. Wanneer men het model voldoende accuraat vindt in het voorspellen van ‘rampzalige’ dagen, kan het model gebruikt worden om reizigers te waarschuwen. Hierbij moet wel rekening gehouden worden met de neiging van reizigers om vooral de slechte dingen van de NS te onthouden. Niet waarschuwen wanneer dat wel nodig geweest was zou tot nieuwe imago-deuken van de NS kunnen leiden.
5. Middel voor onderhandelingen
Het model toont aan dat er een relatie bestaat tussen een groei in reizigersaantallen en de punctualiteit. Eisen van de overheid voor een groei in het reizigersaantal en een stijging van de punctualiteit zijn volgens het model (bij de gegeven situatie) onmogelijk. Stijging van reizigersaantallen zorgt in het model per definitie voor een daling van de punctualiteit.

Het model kan dus zeker gebruikt worden, maar met de kanttekening dat niet iedere verstoring juist voorspeld wordt. Een verder onderzoek naar de oorzaak en de invloed van storingen kan een grote meerwaarde hebben voor de voorspellingen. Echter, het lijkt of storingen zich voordoen op een willekeurige basis, de gebruikte factoren in het model zijn in ieder geval geen oorzaak.

9 Beïnvloeding factoren

De geselecteerde onafhankelijke variabelen windbladval (gemiddelde windsnelheid wanneer er bladval is), maximale temperatuur, reizigersaantallen, sneeuwval, punctualiteit $t-1$, punctualiteit $t-7$ en punctualiteit $t-14$ zouden beïnvloed kunnen worden teneinde de punctualiteit te verhogen. De factoren hebben via een relatie $\beta \cdot factor$ invloed op de punctualiteit. Wanneer de afhankelijkheid van de punctualiteit van de verschillende onafhankelijke variabelen moet veranderen, kan gesleuteld worden aan het proces of de omgeving waardoor de β of de *factor* veranderd. Wanneer de invloed van de onafhankelijke factoren geëlimineerd moet worden, komt dat neer op:

1. $\beta \rightarrow 0$
2. $factor \rightarrow \gamma \quad \forall \beta$

Waarbij γ staat voor een gewenst niveau dat verschilt per *factor*.

Optie 1 betekent dat de gevoeligheid van de treinenloop voor de *factor* aangepakt wordt. Optie 2 betekent dat de *factor* zelf aangepakt wordt, in geval van een positieve relatie ($\beta > 0$) met de punctualiteit betekent dat een verhoging van de *factor*, bij een negatieve relatie ($\beta < 0$) een verlaging van de *factor*.

Per factor zal in dit hoofdstuk bekeken worden wat mogelijke acties zijn voor de factoren. Steeds zal gekeken worden welke relatie bestaat tussen de punctualiteit en de *factor*, en vervolgens zal bepaald worden welke acties haalbaar zijn. Dit hoofdstuk heeft niet als doel om met volledig nieuwe oplossingen te komen: binnen de NS wordt al geruime tijd gewerkt aan oplossingen voor bijvoorbeeld de bladval.

9.1 Windbladval

De invloed van meer reizigers op de punctualiteit is negatief. De relatie in de MLR₁-vergelijking ($\dots - 0,079x_2 \dots$) zegt op zichzelf niets: de β in deze relatie zegt alleen iets over de invloed van windbladval op de punctualiteit *in aanwezigheid* van de overige factoren. Er kan echter wel verondersteld worden dat de invloed van windbladval op de punctualiteit negatief is. Het beïnvloeden van de *factor* windbladval is zo goed als onmogelijk. Bladval en wind zijn natuurlijke verschijnselen die niet beïnvloed kunnen worden. Daarom is het bij deze relatie van belang om de gevoeligheid van de treinenloop te verkleinen door de β te vergroten tot nul.

Rijtijden zijn gebaseerd op een adhesie (quotiënt van de normaalkracht van het wiel en de tractiekracht/remkracht in de rijrichting van het wiel op de spoorstaaf) van 0,07. Door bladval (maar ook door chemische vervuiling, water/condens en roest) kan de adhesie dalen tot onder deze kritische waarde. Wanneer de waarde onder de 0,05 komt, wordt ook remmen binnen de aangegeven ruimte een probleem. Wanneer de adhesie onder de 0,07 komt, ontstaat op het bewuste traject vertraging omdat treinen de berekende rijtijd niet kan halen. Treinen die last hebben van de lage adhesie vallen over het algemeen niet uit, maar komen met een vertraging aan op stations, hier ondervinden andere treinen weer last van de vertraagde trein, waardoor de vertraging zich snel verspreid over Nederland.

De invloed van bladval kan verminderd worden door de adhesie te verhogen gedurende de herfst. De NS en ProRail hebben samen in de afgelopen jaren ingezet op het verminderen van de overlast door het gebruik van Sandite en Anti Blokkeer Systemen (ABS). ProRail brengt op gevoelige baanvakken Sandite aan, Sandite is een gel die ervoor zorgt dat het spoor aan ruwheid wint. De NS heeft treinen uitgerust met ABS wat ervoor zorgt dat de wielen van treinen niet blokkeren bij remmen bij gladheid. Echter, deze maatregelen zijn nog niet voldoende om de overlast door bladval te elimineren. Hiervoor is het ADREM (Adhesie Remedie) project gestart, wat een samenwerkingsverband is tussen de Universiteit Twente, TU Delft en Universiteit Wageningen.

De combinatie van Sandite, ABS en de eventuele uitkomsten van het ADREM project kunnen de β verlagen.

9.2 Maximale temperatuur

Hoge temperaturen zorgen (vaak in combinatie met veel zon) voor het kromtrekken van spoorstaven: 'spoorspatting'. De relatie tussen de maximale temperatuur en de punctualiteit is negatief ($\dots - 0,007x_1\dots$). In de warme julimaand van 2006 waren er 143 valide meldingen van spoorspatting. Spoorspatting zorgt voor het verlagen van de snelheid op de baanvakken waar het voorkomt, of zelfs voor het volledig stilleggen van treinverkeer over het baanvak.

Aangezien hoge temperaturen – net als bladval en wind – natuurlijke verschijnselen zijn, is het corrigeren van de β naar nul de enige haalbare optie. Aanpakken van spoorspatting is echter een dure kwestie aangezien sporen die recentelijk aangelegd zijn een minimale ruimte tussen de aparte spoorstaven hebben om geluidsoverlast te verminderen. Deze sporen zijn extra gevoelig voor spoorspatting aangezien de sporen weinig ruimte hebben om uit te zetten. Het ballastbed moet in principe de krachten die ontstaan door de uitzetting van de spoorstaven kunnen opvangen, maar dit is regelmatig niet het geval. Volgens experts kan beter onderhoud en toezicht het aantal gevallen van spoorspatting sterk verminderen. Ook het gebruik van scheve (glijdende) verbindingen tussen de spoorstaven kan het ontstaan van spoorspatting verminderen.



9.3 Reizigersaantallen

Reizigersaantallen is in het model een aanzienlijke factor die de punctualiteit beïnvloedt. De relatie tussen de factor reizigersaantallen en de punctualiteit is negatief ($\dots - 0,031x_4\dots$). Beperking van de reizigersaantallen past niet binnen het streven van de NS en de overheid om meer mensen te laten reizen met het openbaar vervoer. Doel moet dus zijn om de gevoeligheid van de treindienst te verlagen ($\beta \rightarrow 0$).

De belangrijkste veroorzaker van vertraging in relatie tot reizigersaantallen is het instapproces. Wanneer veel reizigers moeten in- en uitstappen bij een haltering loopt de tijd die nodig is voor de haltering op. Redenen die hiervoor genoemd worden zijn:

1. De 'deurcapaciteit' is onvoldoende
Vooral bij internationale treinen is de instapcapaciteit van de deur te klein. Bijvoorbeeld de Duitse IC rijkstugten zijn voorzien van een enkele deur, in tegenstelling tot de (meeste) Nederlandse rijkstugten. De reizigersstroom doet er bij een enkele deur langer over om in- en uit te stappen dan bij een dubbele deur. Naast de deur naar buiten kunnen in de trein ook limiteringen optreden: bijvoorbeeld de enkele deuren in de ICM-treinstellen (Koplopers) tussen het balkon en de coupé.
2. Reizigers verzamelen zich op 'strategische punten'
Reizigers verzamelen zich op punten die de kortste looproute opleveren, wat meestal neerkomt op een verzameling reizigers rond het midden van het perron. Rond het midden van het perron duurt het instappen vervolgens ook het langst, terwijl bij het begin en het einde van trein bijna niemand instapt.
3. Het in- en uitstapproces verloopt chaotisch
Vooral op drukke tijdstippen verloopt het in- en uitstappen met aardig wat duw- en trekwerk. Daarnaast krijgen uitstappers niet altijd de ruimte die ze nodig hebben.

De NS doet onderzoek naar het stroomlijnen van het instapproces, maar de huidige treinen kunnen niet snel aangepast worden op bijvoorbeeld de breedte van deuren. Echter, enkele eenvoudige maatregelen zouden het instapproces al kunnen beïnvloeden:

1. Licht reizigers in over de stopplaats en lengte van treinen
Wanneer reizigers weten waar de trein stopt en waar de eerste/tweede klas is, kan de reiziger op voorhand een rustige plaats op het perron uitzoeken zodat hij/zij snel kan instappen. Met behulp van de nieuwe digitale CTA-borden en de lettering op de perrons kan redelijk nauwkeurig aangegeven worden hoe een trein in elkaar zit.

2. Bestudeer problematische stations
Sommige stations hebben een opbouw die zorgt voor een beperkte in- en uitstap op bepaalde locaties. Een goed voorbeeld is het eerder genoemde probleem (en oplossing) voor Arnhem Presikhaaf.
3. Creëer in- en uitstapdeuren
Een verregaande maatregel is het scheiden van in- en uitstapplaatsen. Door dit te doen wordt een stroom reizigers door het rijtuig gecreëerd die op elkaar aansluit, in plaats van eens stroom die tegen elkaar ingaat.

Het gaat te ver om te denken dat dit de oplossingen zijn voor het instapprobleem, maar het geeft in ieder geval aan dat er oplossingen zijn die geprobeerd kunnen worden. Het oplossen van het instapprobleem heeft een redelijk hoge prioriteit gezien de sterke afhankelijkheid van de punctualiteit met reizigers aantallen.

9.4 Sneeuwval

Ook sneeuwval is een factor die zelf niet beïnvloedbaar is, wel kan de gevoeligheid van de treinenloop voor sneeuwval verlaagd worden ($\beta \rightarrow 0$). Enkele problemen die kunnen ontstaan door sneeuwval zijn:

1. Seinen worden onleesbaar doordat ze dichtsneeuwen;
2. Koppelingen van treinen vriezen vast waardoor niet meer gecombineerd/gesplitst kan worden;
3. Meer defect materieel;
4. Gedeeltelijk toegankelijke perrons.

Afhankelijk van de hevigheid en verspreiding van de sneeuw heeft de NS meer of minder last. Maatregelen die op 8 februari 2007 genomen zijn, waren gericht op het voorkomen van de hierboven genoemde problemen. De invloed van sneeuwval op de punctualiteit kan verminderd worden door op sneewdagen de gevoelige situaties zo veel mogelijk te voorkomen. Ook het beperken van de dienstregeling, zoals op 8 februari, kan een goede oplossing zijn om materieel vrij te maken zodat materieeldefecten opgevangen kunnen worden.



9.5 Verleden punctualiteit

De invloed van verleden punctualiteit op de punctualiteit vandaag is relatief groot (en positief), wat in principe betekent dat wanneer de punctualiteit verhoogd wordt voor een langere tijd, dit waarschijnlijk door kan rekenen naar de punctualiteit in de toekomst. De redenen hierachter zijn vaag, maar kunnen waarschijnlijk gevonden worden in medewerkermoraal, structurele conflicten in de treindienst door capaciteitsgebrek (de treindienst van iedere zondag is in principe gelijk) en andere (interne) factoren die niet in cijfers uit zijn te drukken. Voor deze factor geldt dat voornamelijk de factor zelf aangepakt moet worden, aangezien dit mogelijk is. Het verminderen van de afhankelijkheid (β) is hier minder belangrijk.

De afhankelijkheid van de punctualiteit in het verleden blijft de NS stimuleren om de interne processen te blijven verbeteren, met als doel een permanente verhoging van de punctualiteit. De acties die hiervoor ondernomen kunnen worden liggen op het stapsgewijs verbeteren van planningen, het verbeteren van de houding van medewerkers ten opzichte van punctueel werken en het durven loslaten van vaste werkpatronen.

De relatie vandaag-gisteren wordt verondersteld veroorzaakt te worden door het vaste patroon waarin treinstellen gepland zijn. Het loslaten van een strikte planning (treinstel 4218 is op dinsdag 10 juli om 10:00 in Enschede) zou het oplossen van calamiteiten eenvoudiger maken. In plaats van plannen op treinstellen zou gepland moeten worden op capaciteit: op dinsdag 10 juli om 10:00 staat er een trein

met 400 zitplaatsen gereed voor vertrek naar Den Haag/Rotterdam. Of deze trein dan een ICM of een VIRM is, maakt niet uit. Zolang de juiste combinaties maar gemaakt kunnen worden.

9.6 Materieeldefecten en infrastructuurstoringen

De factoren die niet meegenomen zijn in het voorspellingsmodel maar wél invloed uitoefenen zijn materieeldefecten en infrastructuurstoringen.

9.6.1 Materieeldefecten

Materieeldefecten hebben in zekere zin te maken met tactische/strategische beslissingen van de NS over de termijn waarop materieel gecontroleerd moet worden op defecten. Op dit moment ligt deze grens op 48 uur, iedere materieelsoort wordt dus eens in de twee dagen gecontroleerd op defecten. Wanneer de NS wil dat het aantal materieeldefecten verlaagd worden, kan gekozen worden om deze termijn te verlagen en het materieel vaker op defecten te controleren. Ook kan gevoelig materieel uit de omloop genomen worden zodra de capaciteit dit toestaat.

9.6.2 Infrastructuurstoringen

Infrastructuurstoringen (storingen) zijn voornamelijk de verantwoordelijkheid van ProRail. ProRail investeert op het moment redelijk veel in vernieuwing van het spoor, wat de kans op storingen sterk moet verkleinen. Continu investeren in het verbeteren van het spoornetwerk is van groot belang voor het realiseren van een hoge punctualiteit. Niet alleen de NS is verantwoordelijk voor de realisatie van de punctualiteit: de rol van ProRail is minstens even groot.

9.7 Conclusie

De beïnvloeding van de factoren die de punctualiteit beïnvloeden is voor de meeste onafhankelijke variabelen erg lastig. De NS moet dan ook streven naar het robuust maken van de treinenloop voor deze factoren, en daarmee de gevoeligheid van de punctualiteit voor versturende factoren te verkleinen. In dit hoofdstuk is aangegeven waar de NS al mee bezig is en waar dit past in de beschreven modellen. De activiteiten komen neer op efficiënter instappen, het minder gevoelig maken van het materieel en het netwerk voor weersomstandigheden en het veranderen van de werkhouding van rijdend personeel.

10 Conclusie

Dit onderzoek is begonnen met de volgende onderzoeksvraag:

Welke externe en interne factoren kunnen gebruikt worden voor het voorspellen van de punctualiteit van NS Reizigers en welke relatie bestaat tussen deze factoren en de punctualiteit?

In de hierop volgende paragrafen zal uiteengezet worden wat het resultaat van dit onderzoek is, teneinde de onderzoeksvraag te beantwoorden. Vervolgens zullen aanbevelingen gedaan worden voor verder onderzoek en gebruik van de uitkomsten van dit onderzoek.

10.1 Voorspellingsmodel

Voor het voorspellingsmodel zijn een vijftal doelen van belang die voornamelijk te maken hebben met het planningsproces. Het voorspellingsmodel is een belangrijk hulpmiddel voor het realiseren van deze doelen:

1. Benchmarking van de eigen prestaties;
2. Verhogen van de dienstverlening;
3. Aanpassen bijsturingscapaciteit;
4. Beter informeren van klanten voor/over/tijdens ‘rampzalige’ dagen;
5. Onderbouwde onderhandelingen over punctualiteitsdoelstellingen;

De voorspellingshorizon ligt voor doelen twee, drie en vier op 24-36 uur, vandaag moet een voorspelling voor morgen gemaakt kunnen worden. Doelen één en vijf zijn meer gericht op reflectieve processen: “hoe hebben we gepresteerd, en hoe hadden we kunnen presteren gegeven de omstandigheden?”. Voor doelen één en vijf zijn schattingen van de jaarpunctualiteit op basis van reizigersgroei ook belangrijk.

Aansluitend is bij de doelen is voor de selectie van een geschikte voorspellingsmethode vastgesteld dat:

1. De voorspellingen een kwantitatieve verwachting van de punctualiteit moeten geven;
2. De punctualiteit beïnvloed wordt door (meerdere) onafhankelijke variabelen;
3. Voorspellingen snel en regelmatig gemaakt moeten worden;
4. Het model de punctualiteit van morgen moet kunnen voorspellen (dus 24-36 uur vooruit);
5. De punctualiteit van vandaag verbonden lijkt met de punctualiteit van gisteren.

Aan de hand van deze constatering is een keuze gemaakt voor een vier-/vijftal methodes voor het voorspellen de punctualiteit, welke aan de hand van de mean square error (MSE), mean absolute percentage error (MAPE) en de bias met elkaar vergeleken zijn:

1. Meervoudige lineaire regressie (MLR);
2. Box-Jenkins (of ARIMA);
3. k-Nearest Neighbours (k-dichtstbijzijnde burens, k-NN);
4. Artificiële Neurale Netwerken (NN);
5. (Combinatie van de bovenstaande methodes).

Om tot een selectie te komen van de modellen zijn voorspellingen gemaakt voor een validatieset afkomstig uit een dataset over de jaren 2004, 2005, 2006 en begin 2007. Aangetoond is dat, ondanks de ingrijpende wijzigingen in de dienstregeling, de gegevens uit de voorgaande jaren probleemloos gebruikt kunnen worden voor het opstellen van voorspellingsmodellen. Aan de hand van de prestaties van de modellen zijn de best presterende modellen geselecteerd, te weten MLR₁ en NN₄ welke hieronder toegelicht worden. De prestaties van het ARIMA en k-NN model zijn als onvoldoende beoordeeld. De geselecteerde modellen baseren hun voorspelling voor punctualiteit op dag t op de onafhankelijke variabelen windbladval (gemiddelde windsnelheid wanneer er bladval is), maximale temperatuur, reizigersaantallen, sneeuwval, punctualiteit $t-1$, punctualiteit $t-7$ en punctualiteit $t-14$. Van deze variabelen is vastgesteld dat ze invloed hebben op de punctualiteit.

Het voorspellingsmodel MLR_1 heeft de vorm:

$$\hat{Y} = 63,9 - 0,007x_1 - 0,079x_2 - 12,479x_3 - 0,031x_4 + 0,264p_1 + 0,207p_7 + 0,110p_{14} \quad (10.1)$$

Met:

x_1	Maximale temperatuur in 0,1K.
x_2	Windbladval in m/s gedurende bladvalperiodes.
x_3	Sneeuwval, een variabele tussen 0 en 1 die de intensiteit van de sneeuwval aangeeft.
x_4	Reizigersaantallen, een samengestelde variabele van reizigerskilometers en conducteurstellingen, variërend tussen 200 en 350.
p_r	Punctualiteit $t-r$ in procent.

Wanneer het model genormaliseerd wordt, blijkt dat de invloed van de parameters windbladval, de historische punctualiteit en sneeuwval de grootste invloed hebben op de punctualiteit.

Het NN_4 -model heeft een inputlaag met zeven neuronen voor de onafhankelijke variabelen, één verborgen laag met 105 neuronen (gebaseerd op Kolmogorov: $n[2n+1]$ en testen met andere hoeveelheden neuronen), en één outputlaag met één neuron voor de voorspelde punctualiteit.

De modellen MLR_1 en NN_4 zijn gecombineerd in het model $G\{NN, MLR\}$. Het $G\{NN, MLR\}$ neemt een direct gemiddelde van de voorspelling van NN_4 en MLR_1 . Voor de validatieset behaalde het $G\{NN, MLR\}$ -model een MSE van 11,9 en een MAPE van 3,39%.

De MSE-score van 11,9 is niet enorm goed, maar betekent niet dat het model niet gebruikt kan worden. De relatief hoge waarde van 11,9 kan verklaard worden door dagen met een afwijkend niveau van infrastructuurstoringen en materieeldefecten (kortweg storingen). Het model $G\{NN, MLR\}$ voorspelt goed wanneer het aantal (of de impact van) storingen op een gemiddeld niveau blijft. Dagen waarop véél storingen plaatsvinden hebben een (veel) lagere punctualiteit dan voorspeld, dagen met weinig storingen hebben een (veel) hogere punctualiteit dan voorspeld.

Hoewel vanuit de NS de wens bestaat om dit soort factoren mee te nemen in een voorspelling, lijkt het onmogelijk om dit te realiseren. Storingen ontstaan door legio factoren, waarvan vele op een willekeurige basis voorkomen. Een trein raakt bijvoorbeeld defect doordat een component versleten is, of door vandalisme. Voorspellen van dit soort factoren is glazen-bol-kijken en is dus moeilijk (of zelfs niet) mogelijk. Verder onderzoek zou op een gedetailleerd niveau kunnen kijken wat onderliggende oorzaken zijn voor storingen, en bepalen of hier een voorspelling van gemaakt kan worden.

Het $G\{NN, MLR\}$ model – en in meer detail de afzonderlijke NN en MLR modellen – laat zien dat er factoren zijn die invloed hebben op de punctualiteit. Daarmee is in principe de hoofdvraag beantwoord.

Om het model bruikbaar te maken binnen de NS, is een (web)applicatie gebouwd. Met behulp van deze applicatie kan eenvoudig een voorspelling per dag gemaakt worden. De applicatie biedt ook de mogelijkheid om opeenvolgende voorspellingen op te slaan, en te controleren in hoeverre het model nog voldoet. Naast het voorspellen van dag-tot-dag kan de applicatie, op basis van weerdata uit het verleden, ook een voorspelling maken voor de gemiddelde punctualiteit over een jaar.

Het is belangrijk te benadrukken dat een voorspellingsmodel onder continue controle moet staan op de geldigheid van het model. Hierbij zijn de volgende aspecten van belang:

1. Het model is uitsluitend geldig op het bereik van de variabelen die gebruikt zijn voor het bepalen van de parameters;
2. Het model is ook uitsluitend geldig bij de omstandigheden (netwerk, dienstregeling, etc.) waaronder de punctualiteit tot stand is gekomen die gebruikt is voor het bepalen van de parameters.

De indicatoren voor afwijkingen van deze twee voorwaarden zijn de MSE en MAPE waarden van de voorspellingen over een recente periode (bijvoorbeeld de voorspellingen van de afgelopen maand). Wanneer deze waarden hoger worden, zonder dat daar een goede verklaring voor is (bijvoorbeeld veel dagen met veel storingen), dan moet het model opnieuw onder de loep genomen worden. Bijstellen van de parameters kan dan al voor een voldoende resultaat zorgen.



Figuur 10.1 Screenshot van de ontwikkelde applicatie 'ProgNoSe'

Concluderend kan dus gesteld worden dat de NS in beperkte mate met de factoren maximale windbladval (gemiddelde windsnelheid wanneer er bladval is), maximale temperatuur, reizigersaantallen, sneeuwval, punctualiteit $t-1$, punctualiteit $t-7$ en punctualiteit $t-14$ een prognose kan maken voor de punctualiteit van morgen. De relatie tussen deze factoren en de punctualiteit wordt gegeven door de MLR_1 , NN_4 en $G\{NN, MLR\}$ modellen.

10.2 Aanbevelingen

10.2.1 Uitbreiding van de onafhankelijke variabelen met storingen

Het grootste gebrek van het huidige model is het ontbreken van de invloedrijke materieeldefecten en infrastructuurstoringen. Hoewel het onmogelijk lijkt om hiervoor goede onafhankelijke variabelen te vinden die op tijd beschikbaar zijn, is het waardevol om hier een onderzoek naar te doen. Belangrijke aspecten hiervoor zijn storingen die een terugkerend karakter hebben, of die op de een of andere manier verbonden kunnen worden met activiteiten op het spoor zoals werkzaamheden.

10.2.2 Inventarisatie van risico's en effect van storingen

Om beter te kunnen omgaan met storingen zou een onderzoek naar de effecten van storingen op bepaalde baanvakken waardevolle informatie kunnen geven. Dit onderzoek kan resulteren in baanvakken die voor de treindienst van cruciaal belang zijn – uitval van deze baanvakken leidt onherroepelijk tot grote punctualiteitsdalingen. De aandacht die de NS en ProRail nu hebben voor de staat van het spoornetwerk zou zich dan sterk kunnen concentreren op de baanvakken die als cruciaal bestempeld zijn. In principe geldt hier een analogie met een productiebedrijf waar de machines in serie een product fabriceren en waarbij één machine de bottleneck is. De bottleneck moet ten allen tijden draaiend gehouden worden om de maximale productie te behalen. Hetzelfde geldt voor de treinenloop: de cruciale baanvakken moeten een beschikbaarheid hebben die dicht bij de 100% ligt om een maximale punctualiteit te behalen.

10.2.3 Verbeter weerdata voor jaarprognoses

De jaarprognoses die door de applicatie gemaakt worden kunnen (sterk) verbeterd worden door nauwkeuriger weerdata te gebruiken. Op dit moment gebruikt de applicatie de weerdata van de afgelopen twee jaar (met correcties voor uitschieters) om een voorspelling te maken voor dit jaar.

Wanneer er een trend in het weer zit, wordt deze niet door de applicatie opgepikt.

De weergegevens zouden in samenwerking met meteorologische bureaus verbeterd kunnen worden, zodat rekening gehouden kan worden in de jaarprognose met bijvoorbeeld een warme zomer of een koude winter.

10.2.4 Laat punctualiteit niet leidend zijn: bepaal maatregelen vanuit klantperspectief

Tijdens dit onderzoek is naar voren gekomen dat de punctualiteit binnen de NS erg belangrijk is. Het belang van de punctualiteit wordt bevestigd door de media-aandacht die er is voor dit aspect. Echter: punctualiteit zegt niets over de beleving van klanten. Het is in principe mogelijk dat klanten een zeer goed oordeel geven over de NS terwijl de punctualiteit voor NS-begrippen laag was. Wanneer gestuurd gaat worden op voorspelde punctualiteit, moet altijd het klantperspectief in ogenschouw gehouden worden.

Een mogelijkheid om de punctualiteit te verhogen is het schrappen van gevoelige treinseries, echter, de klant is hier helemaal niet mee geholpen. De capaciteit van de series die op ongeveer hetzelfde traject rijden is waarschijnlijk onvoldoende, en de aansluitingen die klanten gewend zijn kunnen opeens niet meer gehaald worden. Het klantoordeel zal dan waarschijnlijk ook dalen. Soms is het beter een lagere punctualiteit te incasseren.

Een voorspellingsmodel maakt het alleen maar aantrekkelijker om te sturen op punctualiteit, echter, het klantperspectief zou altijd leidend moeten blijven.

10.2.5 Blijf zoeken naar variabelen die invloed hebben

De variabelen die gebruikt zijn in de modellen zijn in staat om rond de 50% van de schommelingen in de punctualiteit te verklaren. De overige 50% wordt dus veroorzaakt door onafhankelijke variabelen die niet in het model opgenomen zijn. Extra onafhankelijke variabelen, naast de ‘moeilijke’ variabele storingen, kunnen het model een hogere nauwkeurigheid geven.

Lijst van afkortingen

Gebruikte afkortingen

ARI	Automatische Rijweginstelling
BRV	Binnenlands Reizigers Vervoer
HC	(Hoofd)conductor
k-NN	k-Nearest Neighbours algoritme
KNP	Knooppuntcoördinator
MAPE	Mean Absolute Percentage Error
MCN	Machinist
MLR	Meervoudige lineaire regressie
MMC	Medewerker mobiele controle
MSE	Mean square error
MSR	Mean square regression
NN	Artificieel Neuraal Netwerk
NS	Nederlandse Spoorwegen
PBT	Proces bijzondere taken
PE	Productie Eenheid
PLP	Procesleider perron
VTA	Vrijtijdsaanspraak

Gebruikte afkortingen van stations

Amf	Amersfoort
Apd	Apeldoorn
Asd	Amsterdam
Dt	Delft
Dv	Deventer
Es	Enschede
Esd	Enschede Drienerlo
Gd	Gouda
Gvc	's Gravenhage
Hgl	Hengelo
Ht	's Hertogenbosch
Mt	Maastricht
Nm	Nijmegen
O	Oss
Rvs	Ravenstein
Zl	Zwolle
Zp	Zutphen

Lijst van figuren en tabellen

Figuren

Figuur 1.1 Organigram Binnenlands Reizigers Vervoer.....	4
Figuur 1.2 Netwerk Noordoost.....	5
Figuur 3.1 Totstandkoming van de planning.....	14
Figuur 3.2 Goederenvervoer per spoor in Nederland (CBS, 2007).....	16
Figuur 3.3 Oorzaak en gevolg diagram (Reijnen, 2005).....	20
Figuur 4.1 Planningproces met mogelijke verbeteringen (BCG, 2005).....	23
Figuur 4.2 Relatie punctualiteit gisteren - punctualiteit vandaag (data van 1 januari 2004 tot 11 maart 2007).....	26
Figuur 4.3 Een voorbeeld van een (eenvoudig) NN, een bolletje is een neuron	30
Figuur 4.4 Neuron met input en output (Kosters, 2006)	31
Figuur 4.5 De sigmoidfunctie voor verschillende β waarden.....	32
Figuur 4.6 Step en sign functie.....	32
Figuur 4.7 De set waarnemingen.....	34
Figuur 4.8 Iedere waarneming k bestaat uit I onafhankelijke variabelen $x_{k,i}$	34
Figuur 4.9 Kandidaat 1 $\{1, 2, 3\}$ wordt als eerste vergeleken met de referentie $\{16, 17, 18\}$	35
Figuur 4.10 Voor de bepaling van $d_{18,3}$ worden de o.v.'s van waarnemingen 3 en 18 met elkaar vergeleken met vergelijking (4.18).....	35
Figuur 5.1 Boxplots voor de data van 2004 t/m 2007	41
Figuur 5.2 Boxplots voor de werkelijke treinenactiviteiten	43
Figuur 5.3 Boxplots voor treinactiviteiten 2004 t/m 2007	44
Figuur 6.1 Relatie maximale temperatuur – Punctualiteit.....	46
Figuur 6.2 Relatie minimale temperatuur – Punctualiteit.....	47
Figuur 6.3 Er is geen relatie tussen windsnelheid en punctualiteit.....	47
Figuur 6.4 Ook de maximale uurgemiddelde windsnelheid lijkt geen invloed te hebben.....	48
Figuur 6.5 Relatie WindBladval - Punctualiteit	49
Figuur 6.6 Relatie sneeuwval - Punctualiteit.....	50
Figuur 6.7 Er is geen relatie tussen de neerslaghoeveelheid en de punctualiteit.....	51
Figuur 6.8 Weekenden, in het figuur is een duidelijk cyclisch patroon te vinden over 7 dagen.....	51
Figuur 6.9 Relatie aantal reizigers-punctualiteit	52
Figuur 6.10 Files, punctualiteit en reizigersaantallen.....	53
Figuur 6.11 Files gisteren, punctualiteit en reizigersaantallen.....	54
Figuur 6.12 Relatie suicides - punctualiteit.....	54
Figuur 6.13 Relaties ziektecijfer – punctualiteit.....	56
Figuur 6.14 Relatie VTA machinisten – punctualiteit. Negatieve VTA-waarden worden voornamelijk veroorzaakt door overwerken.....	56
Figuur 6.15 Relatie aantal treinactiviteiten – punctualiteit. De puntenwolken worden veroorzaakt door de toename van het aantal ritten in de jaren 2004 – 2007.....	57
Figuur 6.16 Relatie materieel tekorten/overschotten – punctualiteit.....	58
Figuur 6.17 Autocorrelatie van de punctualiteit.....	59
Figuur 7.1 Foutenverdeling van MLR_1	66
Figuur 7.2 De residuen zijn willekeurig verdeeld	66
Figuur 7.3 Foutenverdeling van $ARIMA_4$	67
Figuur 7.4 Tracking signaal voor het $ARIMA_4$ model.....	74
Figuur 7.5 Voorspellingen en waarnemingen voor de verschillende methodes. Grotere versies van deze figuren staan in bijlage A.....	78
Figuur 7.6 Vergelijking van de beste drie voorspellingsmethoden.....	79
Figuur 7.7 Squared Error (SE) voor NN, MLR en $G\{NN, MLR\}$	79
Figuur 7.8 Vergelijking voorspellingen – waarnemingen. De grijze lijn is voorspelling = waarneming.....	80

Figuur 8.1 Openingsscherm ProgNoSe	82
Figuur 8.2 Eerste stap voor het maken van een voorspelling.....	82
Figuur 8.3 Tweede stap voor het maken van een voorspelling	82
Figuur 8.4 Laatste stappen voor het maken van een voorspelling in ProgNose.....	83
Figuur 8.5 Instellen van de parameters voor de jaarprognose	84
Figuur 8.6 Resultaat van de jaarprognose	84
Figuur 8.7 Voorspellingen en waarnemingen (grafiekvorm en tabelvorm)	85
Figuur 10.1 Screenshot van de ontwikkelde applicatie 'ProgNoSe'	95

Tabellen

Tabel 1.1 Geconsolideerde Winst-en-verliesrekening 2005 NV Nederlandse Spoorwegen, in miljoenen euro's (NS Jaarverslag 2006, 2007).....	2
Tabel 1.2 Klantbeoordeling, percentage hoger dan een zeven (NS Jaarverslag 2006, 2007)	2
Tabel 3.1 Punctualiteit top vijf Europa, treinen binnen 5 minuten t.o.v. geplande aankomsttijd (NMBS, 2006).....	17
Tabel 3.2 Stations die meetellen voor de punctualiteit, in <i>cursief</i> zijn de stations aangegeven behorende tot netwerk Noordoost (MinVW, 2005)	17
Tabel 3.3 Handeling waarbij vertraging optreedt	18
Tabel 3.4 Percentage sneller gereden dan schematijd	19
Tabel 5.1 Gegevens van de verwijderde cases	42
Tabel 5.2 Samenvatting van de data voor 2004 t/m 2007	42
Tabel 5.3 ANOVA tabel van de data voor 2005 t/m 2007	42
Tabel 5.4 Reizigersgroei.....	43
Tabel 5.5 Toets voor normaliteit (Shapiro-Wilk).....	43
Tabel 6.1 Reizigersverdeling per weekdag	52
Tabel 7.1 Geaccepteerde variabelen.....	61
Tabel 7.2 Validatieset.....	62
Tabel 7.3 Data voor de lineaire regressie.	63
Tabel 7.4 Onderlinge (Pearson) correlatie van de o.v., een waarde van -1 of 1 geeft een perfecte (lineaire) correlatie weer.	63
Tabel 7.5 Minimum en maximum waarden parameters	64
Tabel 7.6 ANOVA tabel voor MLR ₁	65
Tabel 7.7 Statistieken voor MLR ₁	65
Tabel 7.8 Statistieken voor ARIMA ₄	67
Tabel 7.9 Parameters voor ARIMA ₄	67
Tabel 7.10 Selectie van het aantal burens	70
Tabel 7.11 Selectie van het aantal best overeenkomende kandidaten	70
Tabel 7.12 Slechte voorspellingen MLR en NN	76
Tabel 7.13 MSE en MAPE waarde voor de modellen.	76
Tabel 7.14 Waarnemingen en voorspellingen van de verschillende modellen voor de validatieset	77

Literatuur

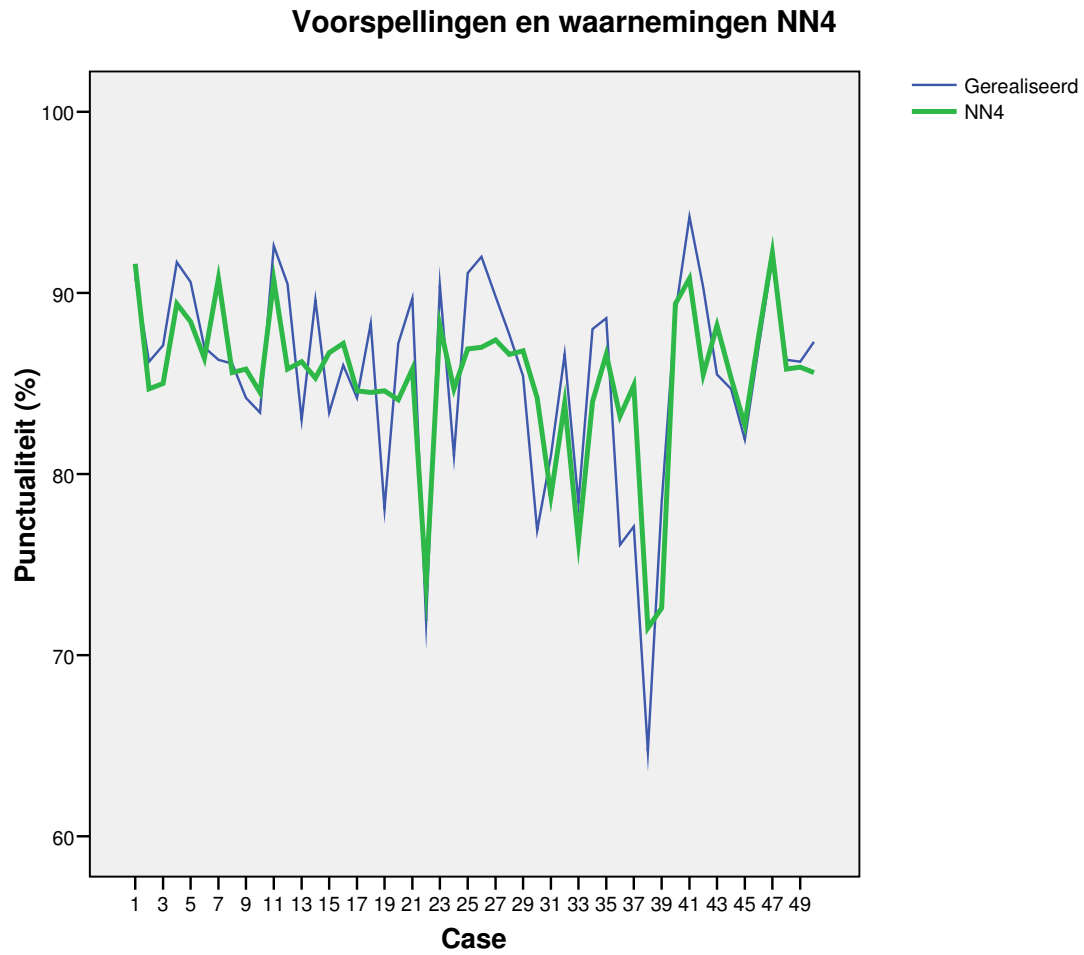
Met auteur

- Burch, C., 2003: *Computational Complexity*, session notes for *Java Programming*, Hendrix College.
- Freeman, R.E., McVea, J. A *Stakeholder Approach to Strategic Management*. The Darden School, University of Virginia.
- Gooijer, J.G. de, Hyndman, R.J., 2006: *25 years of time series forecasting*. International Journal of Forecasting, 22, pp 443-473.
- Hanke, John E., Wichern, Dean W., 2005: *Business Forecasting*. Eight (international) edition, Pearson Prentice Hall, New Jersey. ISBN 0-13-122856-0.
- Hasti, R. 2006: *Complexity*, session notes for *Introduction to Data Structures*, University of Wisconsin-Madison.
- Hippert, H.S., Bunn, D.W., Souza, R.C., 2005: *Large neural networks for electricity forecasting: Are they over fitted?* International Journal of Forecasting, 21, pp 425-434.
- Hornick, K., Stinchcombe, M., White, H., 1989: *Multilayer feedforward networks are universal approximators*. Neural Networks, 2(5), pp. 359-366.
- Hyndman, R.J., Koehler, A.B., 2006: *Another look at measures of forecast accuracy*. International Journal of Forecasting, 22, pp 679-688.
- Kallenberg, W.C.M., 2002: *Dictaat Statistiek II voor TBK en CiT*. Universiteit Twente, faculteit der Toegepaste Wiskunde, juni 2002.
- Kosters W.A., 2007, *Kunstmatige Intelligentie: Neurale Netwerken (collegesheets naar Russel, Norvig, Artificial Intelligence, A Modern Approach)*. Leiden, voorjaar 2006
- Mitchell, R.K., Agle, B.R., Wood, D.J., 1997, *Toward a theory of stakeholder identification and salience: defining the principle who and what really counts*. Academy of management review, vol. 22, no. 4, pp. 853-886, 1997.
- Picton, P., 2000: *Neural Networks*. Second edition, Palgrave, New York. ISBN 0-333-80287-X.
- Plagianakos, V.P., Sotiropoulos, D.G., Vrahatis, M.N., 1998: *An improved backpropagation method with adaptive learning rate*. Technical Report No. TR98-02.
- Reijnen, R., Schriever, B., de Vries, B., 2005: *Kwantificeren punctualiteit (1^e fase)*. CQM bv, Eindhoven, december 2005.
- Schriever, B., van Doremalen, J., 2006: *Analyse startpunctualiteit en impact op de KPI aankomstpunctualiteit (Fase 2a)*. CQM bv, Eindhoven, april 2006.
- Swanson, N.R., White, H., 1997: *Forecasting economic time series using flexible versus fixed specifications and linear versus nonlinear econometric models*. International Journal of Forecasting, 13, pp 439-461.
- Tanizaki, H., 2000: *Bias correction of ordinary least squares estimation in the regression model with lagged dependent variables*. Computational Statistics & Data Analysis 34, pp. 495-511

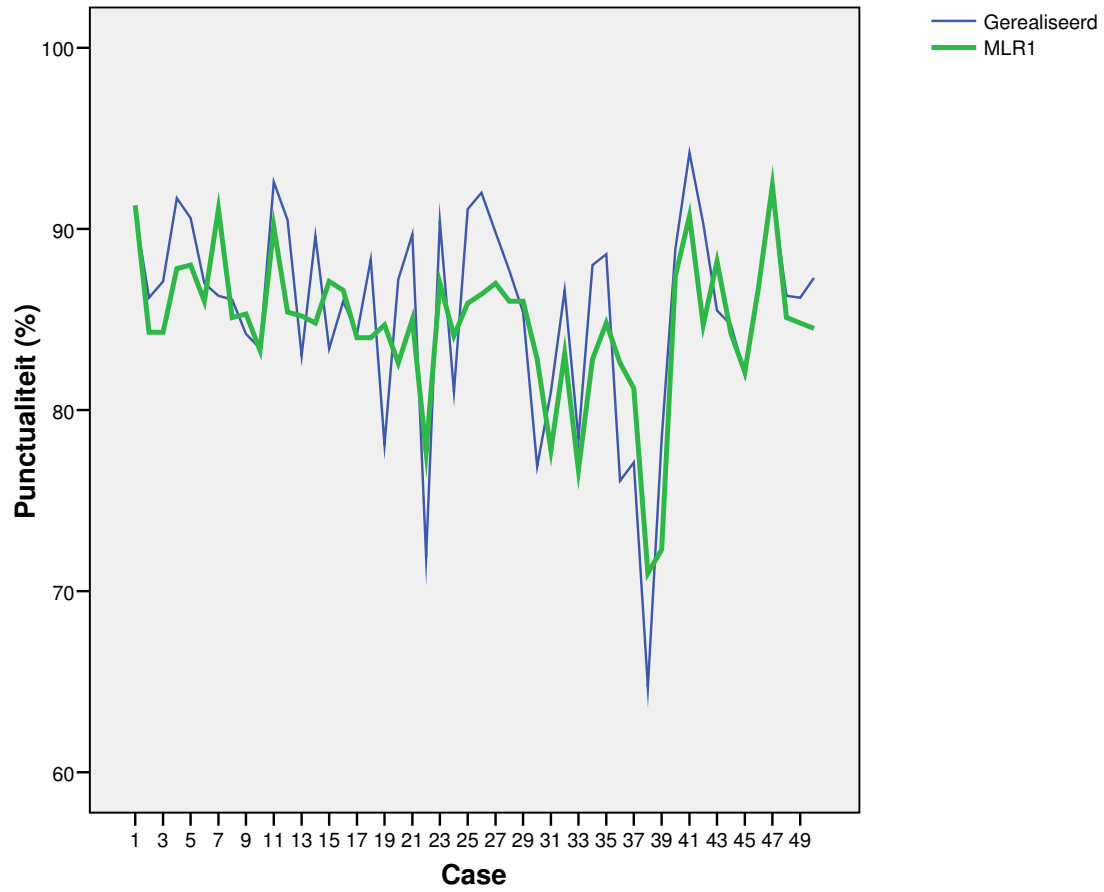
Zonder auteur

- BCG, 2005: *Samenvatting bevindingen aankomstpunctualiteit*. The Boston Consulting Group B.V., december 2005.
- CBS, 19 maart 2007, *Maandcijfers internationaal goederenvervoer per spoor*. CBS Statline, www.cbs.nl, geraadpleegd 19 maart 2007.
- MinVW, 2005: *Vervoerconcessie voor het hoofdtrainnet*. 2005, Den Haag.
- Nedtrain, 2007: *Website Nedtrain* (www.nedtrain.nl). Geraadpleegd 05 februari 2007.
- NMBS, 2006: *Intrabel: De stiptheid verbetert lichtjes tijdens 1^e semester van 2006*. Persbericht NMBS, 06 september 2006.
- NS Intranet, 2006: *Intranet NS/Beleid/Huidige Koers/Koers*. Geraadpleegd 05 februari 2007.
- NS Jaarverslag 2006, 2007: *Ga mee...: NS Jaarverslag 2006*. 15 februari 2007, Utrecht.
- NS, 2006: *Vervoerplan NS 2006*. Directie Nederlandse Spoorwegen, 30 januari 2006, Utrecht.
- ProRail, 2006: *Monitoring Spoorgebruik 2005: Ontwikkelingen 2005 ten opzichte van 2004*. Publicatie ProRail, 28 november 2006.

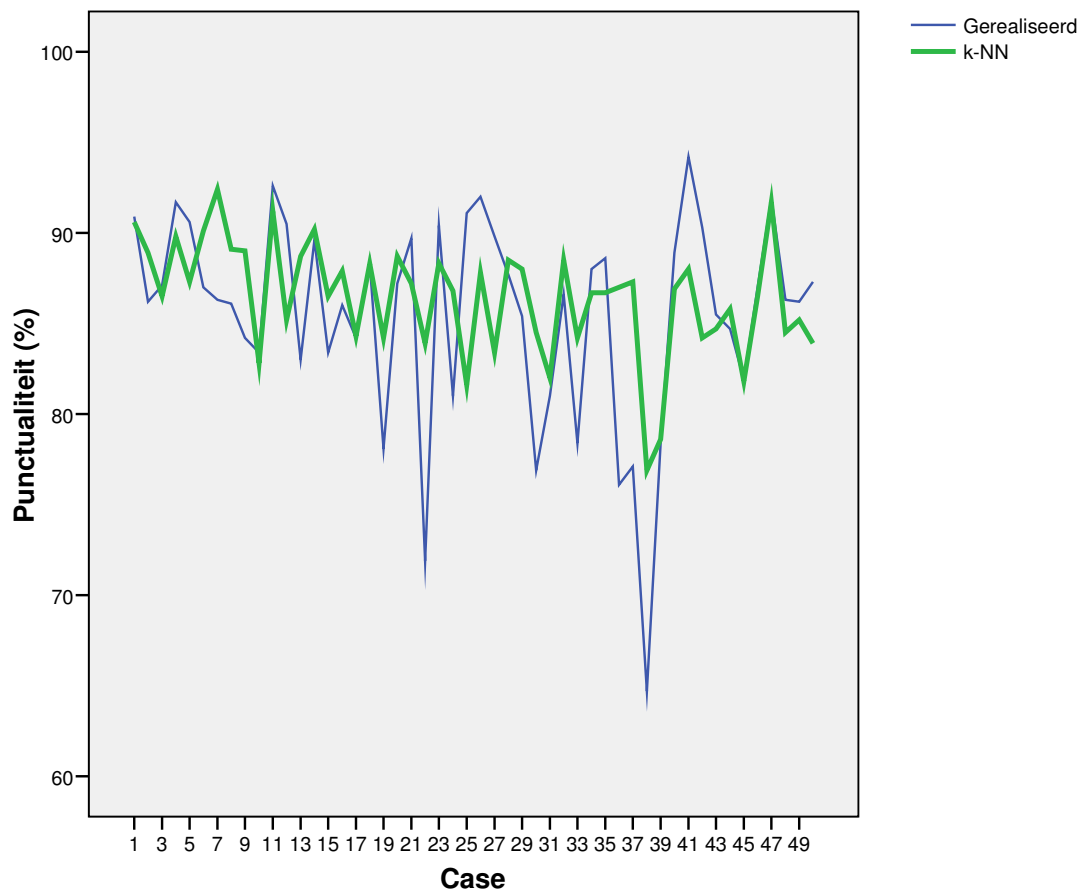
Bijlage A: Voorspellingen en waarnemingen



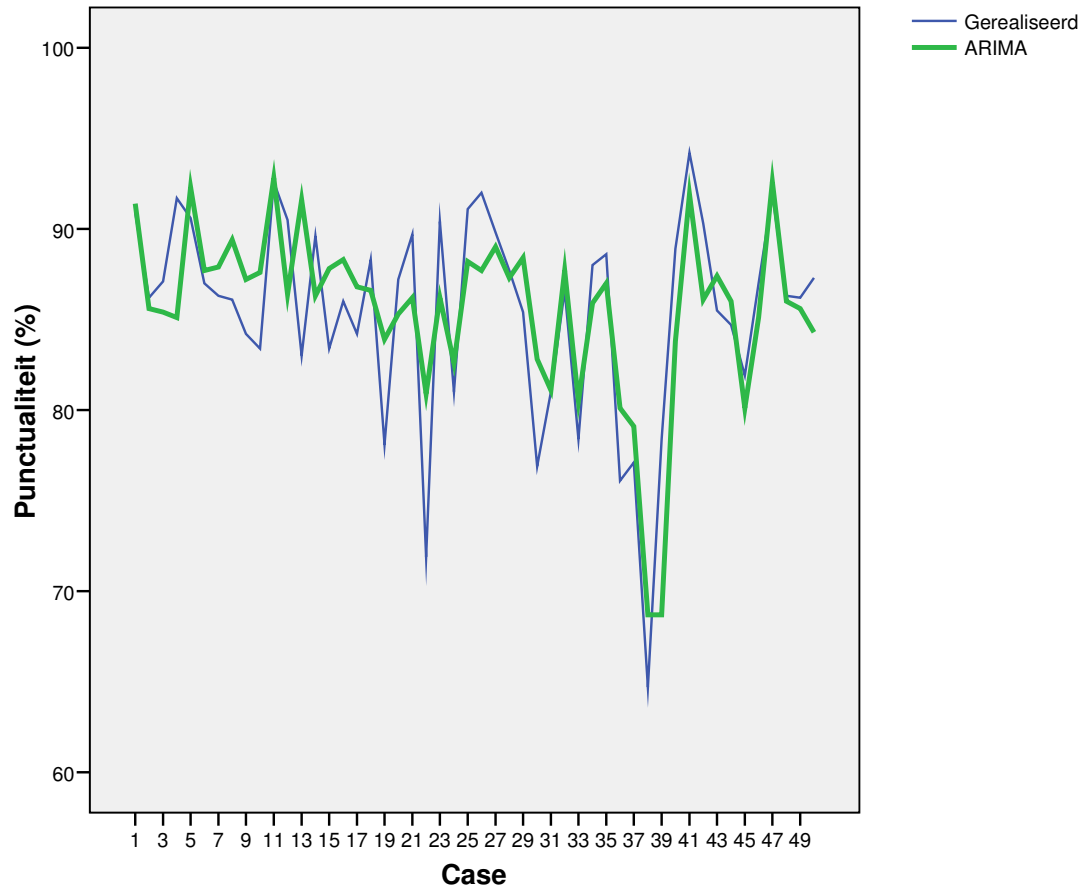
Voorspellingen en waarnemingen MLR1



Voorspellingen en waarnemingen k-NN

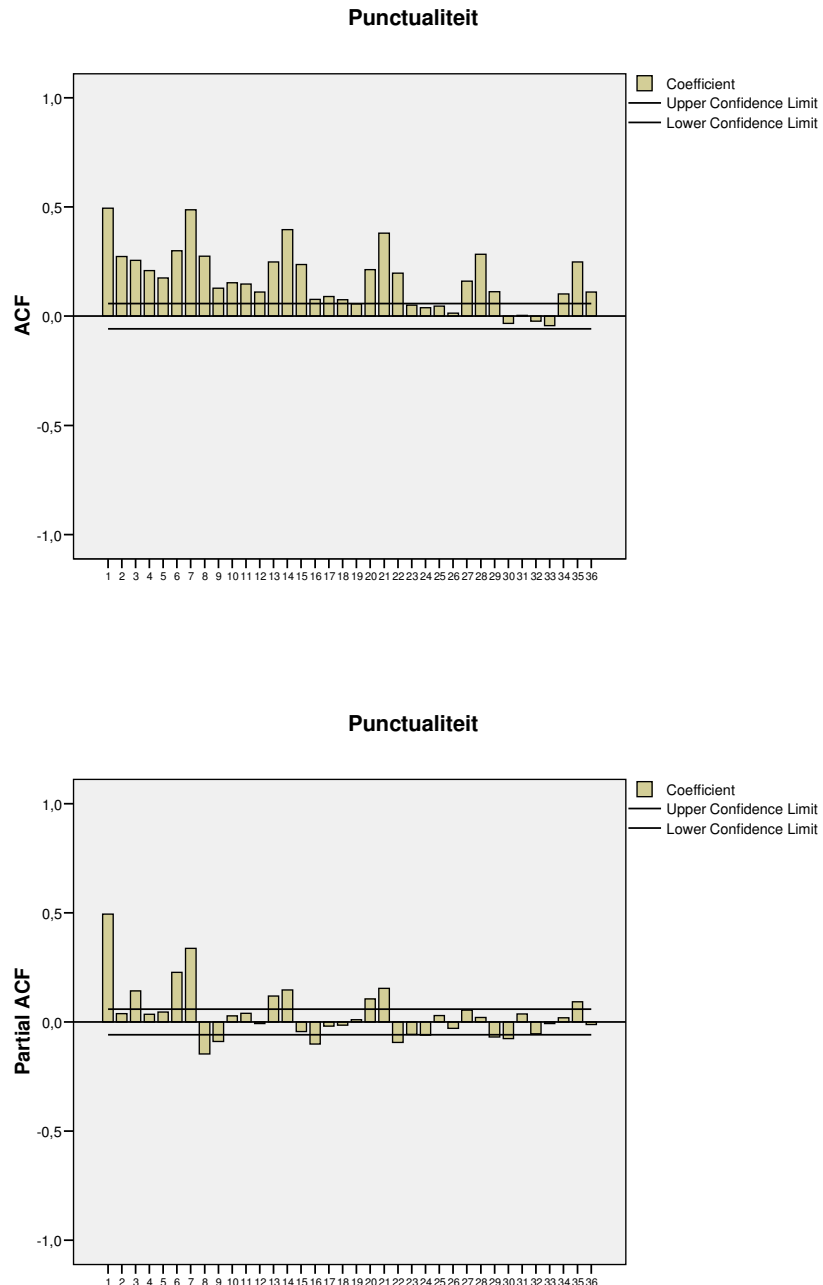


Voorspellingen en waarnemingen ARIMA



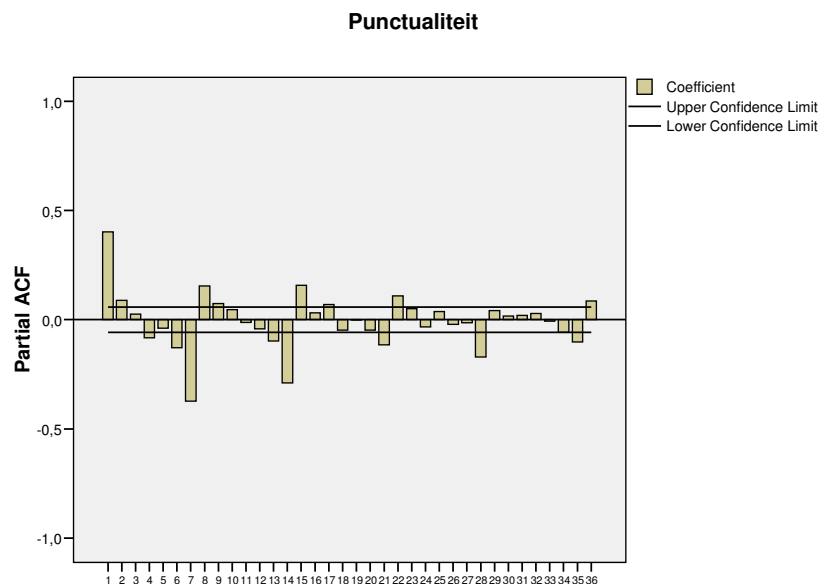
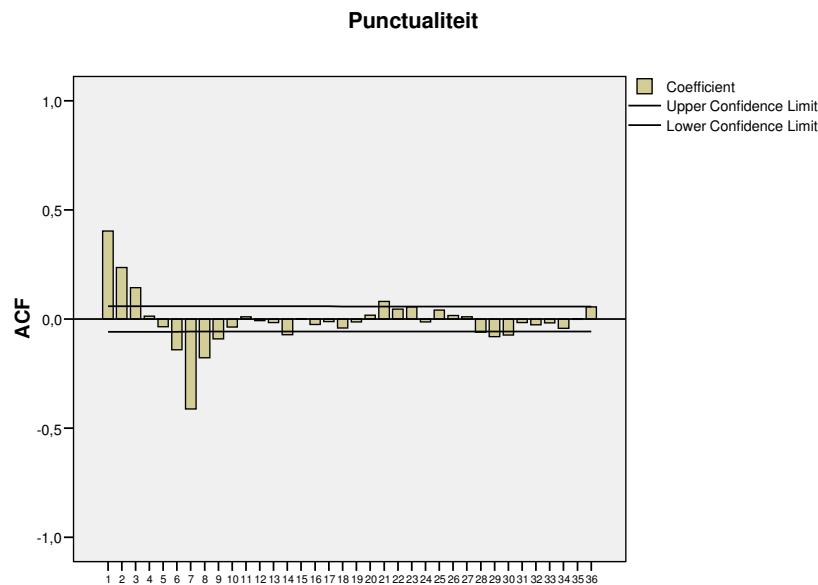
Bijlage B: ARIMA model

Allereerst wordt voor het maken van een ARIMA model gekeken naar de (partiële-) autocorrelatie (autocorrelation factor (ACF) en partial autocorrelation factor (PACF)) van de afhankelijke variabele. De ACF en PACF plaatjes voor de standaard data zien er als volgt uit:



Uit deze figuren kan geen patroon gehaald worden dat gebruikt kan worden voor een ARIMA model. Om de dubbele autocorrelatie (omdat vandaag met gisteren autocorrelatie vertoont, vertoont gisteren ook autocorrelatie met eergisteren) uit de data te halen wordt eerst een seizoensdifferentie ($Y_t - Y_{t-7}$) genomen. Er wordt gekozen voor een seizoensdifferentie omdat de seizoensautocorrelatie het beste in

stand blijft (hoogste pieken die langzaam uitdoven). Deze differentie levert de volgende ACF en PACF figuren op:



In deze figuren is het volgende te zien:

Korte termijn:

Het ACF figuur laat een exponentieel naar nul verlopend patroon zien voor de niet-seizoensfactoren.

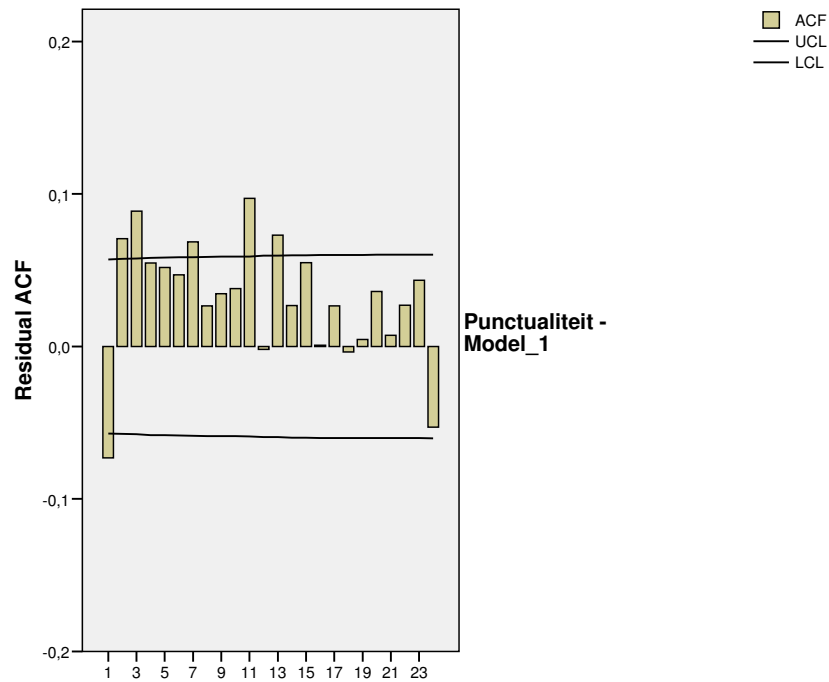
Het PACF figuur laat (verhoudingsgewijs t.o.v. het ACF figuur) 1 piek zien. Dit duidt op een (1, 0, 0) model voor de niet-seizoensfactoren.

Seizoenstermijn:

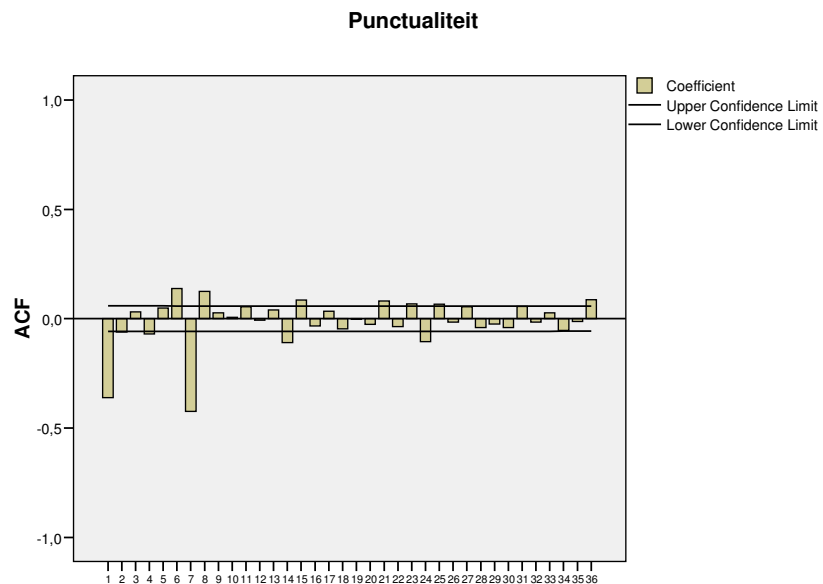
Voor de seizoenstermijn laten de figuren precies het tegenovergestelde beeld zien van de korte termijn: 1 piek bij de ACF, en een exponentieel afnemend PACF. Dit duidt op een (0, 1, 1) model voor de seizoensfactoren.

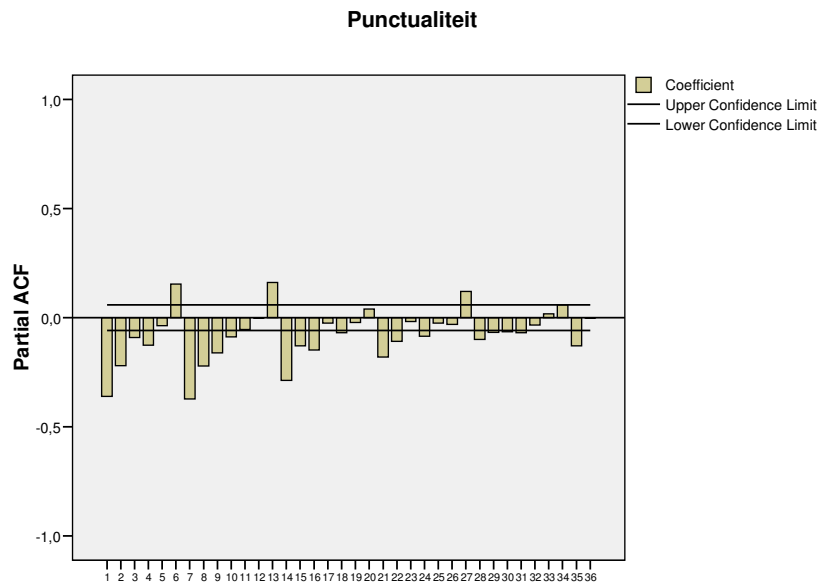
Totaal wordt dit dus een $ARIMA(1, 0, 0)(0, 1, 1)_7$ model ($ARIMA_1$).

Voor het ARIMA₁-model wordt weer een ACF-figuur getekend met daarin de ACF van de residuen (de fouten in de voorspellingen). Uit dit figuur blijkt dat er nog steeds informatie in de residuen zit aangezien er nog steeds significante AC-factoren zijn tussen de residuen (bijvoorbeeld op lag 1 en lag 3).



Om dit te voorkomen, wordt er nogmaals een differentie genomen op de korte termijn. De ACF en PACF figuren zien er dan als volgt uit:





Het beeld van de factoren is nu sterk veranderd ten opzicht van het begin: het resultaat is ook beter dan in het begin.

Korte termijn:

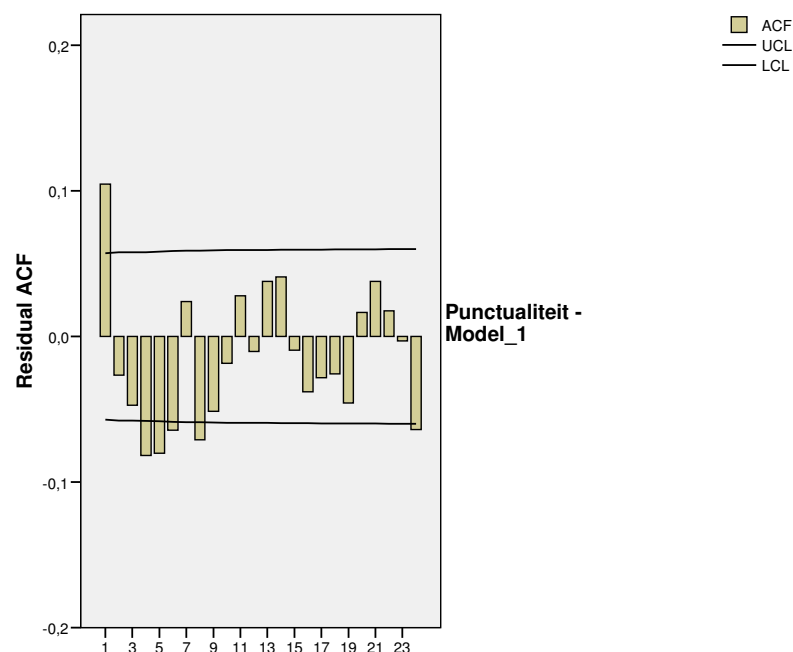
Het ACF figuur laat duidelijk één piek zien voor de korte termijn. De PAC-factoren lijken min of meer exponentieel naar nul te gaan. Dit duidt op een (0, 1, 1) model.

Seizoenstermijn:

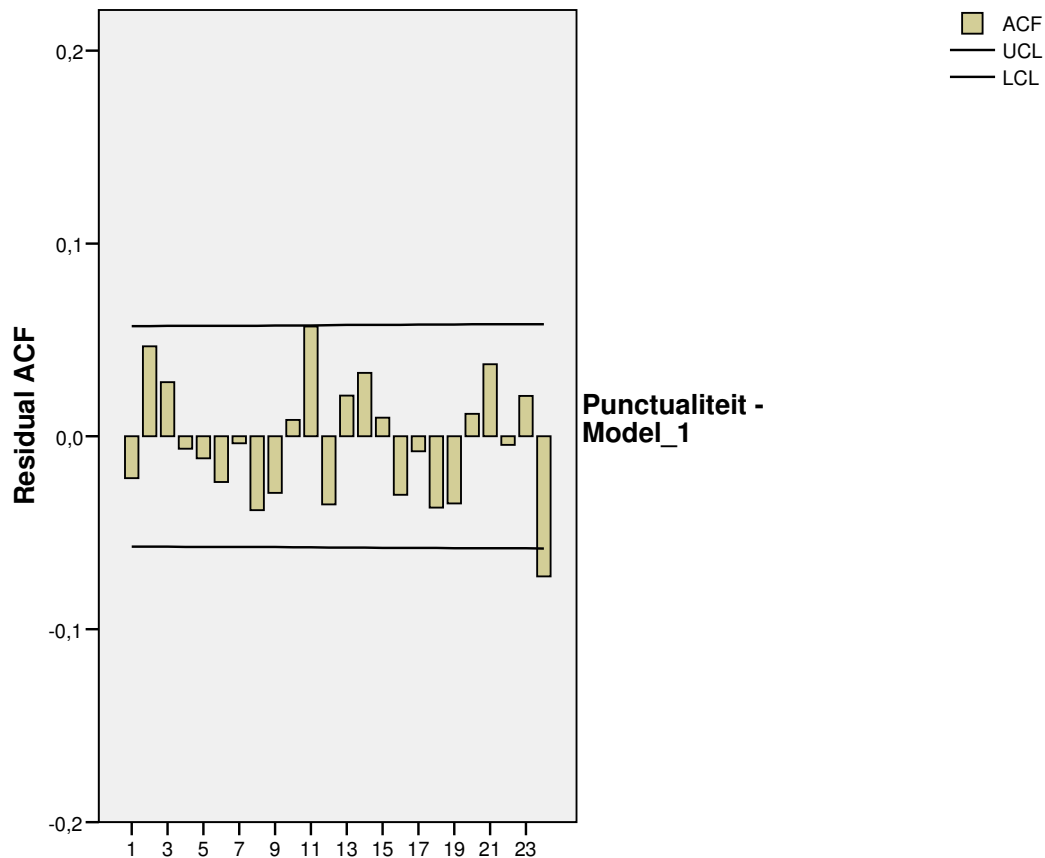
Op seizoenstermijn laat het ACF-figuur één piek zien, en het PACF-figuur laat een min of meer exponentieel verloop naar nul zien. Het model is dan ook (0, 1, 1).

Het totale model komt dan op $ARIMA(0, 1, 1)(0, 1, 1)_7$ ($ARIMA_2$).

Voor het $ARIMA_2$ -model zien de AC-factoren er als volgt uit:

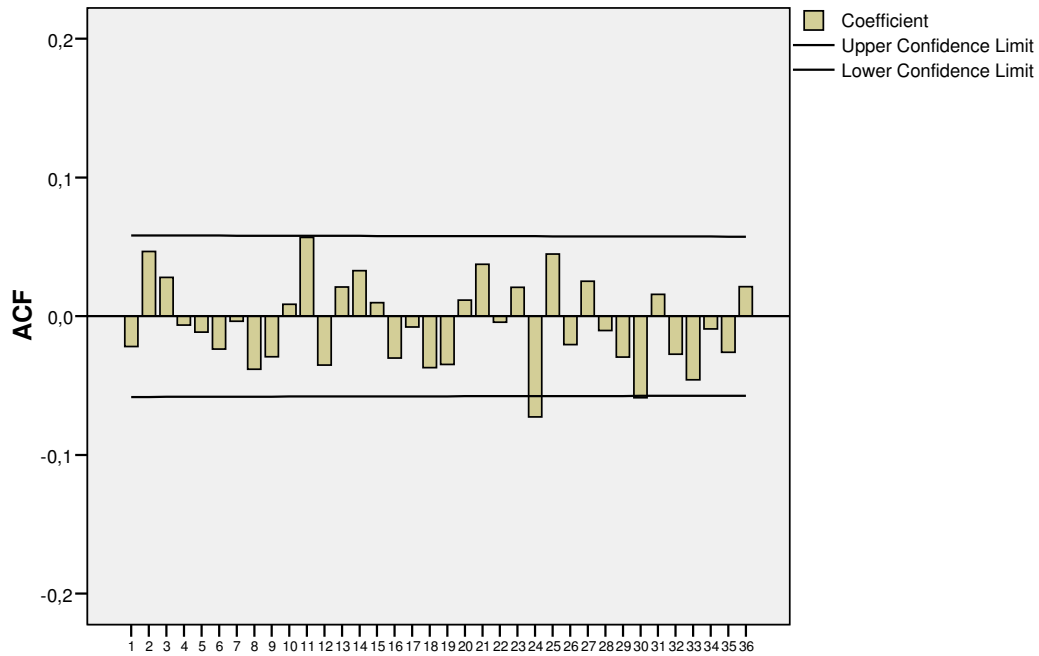


Ook dit resultaat is nog niet goed genoeg om het ARIMA₂-model te accepteren. Om de sterke autocorrelatie voor lag 1 te compenseren, voegen we een AR-term toe aan het korte model: ARIMA(1, 1, 1)(0, 1, 1)₇ (ARIMA₃).

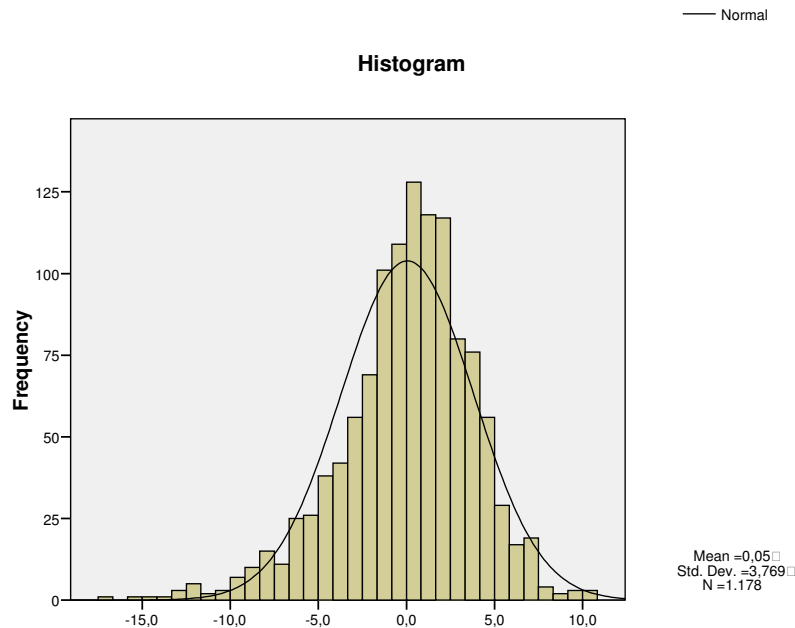


Na toevoeging van deze term blijft een significante autocorrelatie in de residuen op lag 7 over, wat ondervangen kan worden door het toevoegen van een AR-term in het seizoensmodel: ARIMA(1, 1, 1)(1, 1, 1)₇ (ARIMA₄). Dit levert het volgende beeld op:

Noise residual from Punctualiteit-Model_1



Op een significante term bij lag 24 en 30 na ligt alles binnen de zekerheidslimieten: er kan dus redelijkerwijs verondersteld worden dat de residuen bestaan uit 'witte ruis'. Daarnaast scoort het model op de normaliteitstoets van Shapiro-Wilk een waarde van 0,973, wat duidt op een normale verdeling.



De verdere statistieken van dit model zien er als volgt uit:

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
	Stationary R-squared		Statistics	DF	Sig.	
Punctualiteit-Model_1	0	,695	17,871	14	,213	12

De Ljung-Box statistiek laat een goede waarde die groter is dan 0,05.

De parameters van het model zijn als volgt:

ARIMA Model Parameters

		Estimate	SE	t	Sig.
AR	Lag 1	,347	,036	9,669	,000
Difference		1			
MA	Lag 1	,895	,018	50,674	,000
AR, Seasonal	Lag 1	,095	,032	2,916	,004
Seasonal Difference		1			
MA, Seasonal	Lag 1	,970	,011	92,291	,000

De differenties zorgen voor het volgende beeld van de variabelen:

$$\begin{aligned}
 \Delta Y_t &= Y_t - Y_{t-1} \\
 \Delta_7 Y_t &= Y_t - Y_{t-7} \\
 \Delta \Delta_7 Y_t &= \Delta(Y_t - Y_{t-7}) = Y_t - Y_{t-1} - Y_{t-7} + Y_{t-8} = W_t
 \end{aligned}
 \tag{0.1}$$

Het model heeft dan de volgende vorm:

$$W_t = 0,347W_{t-1} - 0,895\varepsilon_{t-1} + 0,095W_{t-7} - 0,970\varepsilon_{t-7} \tag{0.2}$$

Of in de uitgeschreven vorm:

$$\begin{aligned}
 Y_t &= Y_{t-1} + Y_{t-7} - Y_{t-8} + 0,347(Y_{t-1} - Y_{t-2} - Y_{t-8} + Y_{t-9}) \\
 &\quad - 0,895\varepsilon_{t-1} + 0,095(Y_{t-7} - Y_{t-8} - Y_{t-14} + Y_{t-15}) - 0,970\varepsilon_{t-7}
 \end{aligned}
 \tag{0.3}$$