# Scientific Software Engineers

Postbus 260 2600 AG DELFT

MEMO	MV/M15.045	tel. 015-285 012 fax. 015-285 012	
Datum	7 december 2015	vortech@vortech.n	
Auteur(s)	Misha Veldhoen en Jok Tang		
Onderwerp	Statistische Impactanalyse van Speciale TenneT	Projecten bij	

### Documentinformatie

Versie	Auteur	Datum	Opmerkingen	Review
0.1	MV	14-10-2015	Eerste preliminary versie	JT,FvE, RK
0.2	MV	28-10-2015	Tweede preliminary versie	JT,FvdM, RK
0.3	MV	02-11-2015	Eerste draft complete versie	JT,FvdM
0.4	MV	04-11-2015	Tweede draft complete versie	JT,MR, FvE
1.0	MV	10-11-2015	Uiteindelijke versie	FvE
1.1	MV	07-12-2015	Uiteindelijke versie	

# Disclaimer

The contents of this working document serve a purely informative purpose. Any application of the techniques, ideas, and suggestions in this document is at the reader's sole discretion and risk.

Versie 1.1, 7 december 2015

# Voorwoord

Voor u ligt de eindrapportage van de statistische impactanalyse die VORtech heeft uitgevoerd voor TenneT. Het omvat een gedegen statistische analyse die inzicht geeft in de vraag of er aantoonbare statistische invloeden te herkennen zijn van speciale projecten op de spanningskwaliteit. Als statisticus aan de TU Delft heb ik tijdens dit onderzoek VORtech ondersteund en geadviseerd over de verschillende relevante statistische methodes en modellen die gekozen kunnen worden. Ik sta volledig achter de statistische analyse zoals gepresenteerd in deze eindrapportage. Ik hoop dat TenneT de resultaten goed kan gebruiken bij verdere analyses en discussies.

dr.ir. F.H. van der Meulen Delft, 2 november 2015

# Management Summary

In deze eindrapportage is de statistische analyse gepresenteerd die VORtech heeft uitgevoerd voor TenneT. TenneT heeft van 8 locaties de tijdreeksen aangeleverd van verschillende spanningskwaliteitsaspecten alsmede van het schakelgedrag van het specifieke speciale project en overige relevante omgevingsfactoren.

In de uitgevoerde analyse is op een consistente en statistisch verantwoorde manier gekeken naar deze tijdreeksen. Op basis hiervan is voor iedere locatie de impact bepaald van het speciale project op de spanningskwaliteitsaspecten. Hieronder is een overzicht gepresenteerd van het antwoord op de vraag of er een statistisch significante impact is gevonden. Deze impacts zijn bepaald aan de hand van gevonden correlaties.

Impact: de aan- en afkoppeling van het speciale project valt statistisch samen								
met het spanningskwaliteitsaspect:								
Speciaal Project	Locatie	LSV1	LSV2	LSV3	Asym	THD1	THD2	THD3
HVDC NorNed	EEMS	ja	ja	ja	ja	ja	nee	nee
HVDC BritNed	MVL	nee	nee	nee	nee	ja	ja	ja
380kV-Condensatorbank	DMN	nee	nee	nee	ja	nee	nee	nee
220kV-Condensatorbank	WEW	nee						
150kV-Condensatorbank	MZ	ja						
110kV-Condensatorbank	HCL	nee	nee	nee	nee	ja	ja	ja
380kV-kabel	BWK	ja	ja	ja	nee	ja	ja	ja
380kV-kabel	WTR	ja	ja	ja	nee	ja	nee	nee

Tabel 1: In de statistische analyse is voor iedere locatie het covariëren van deze spanningskwaliteitsaspecten en het schakelen van het speciale project en de omgevingsfactoren onderzocht. De uiteindelijke analyse is gedaan op naar dag-waarden geaggregeerde tijdreeksen van de spanningskwaliteit. In dit overzicht wordt antwoord gegeven op de vraag of, gegeven de naar dag-waarde geaggregeerde data, de nulhypothese "geen correlatie met het speciale project" verworpen kon worden. Bij deze nulhypothese is een significantieniveau van 0.05 gehanteerd. Merk op dat het antwoord op de vraag geen uitspraak doet over de feitelijke grootte van het effect. Uitsluitend de kwalitatieve impact van het speciale project op de spanningskwaliteitsaspecten per locatie is hier gepresenteerd.

Inhoudsopgave

1	Inleiding		6
	1.1	Omschrijving van de Data	6
	1.2	Indeling van dit Rapport	7
<b>2</b>	Onderzoel	ksmethodiek	9
	2.1	Tijdsreeksen	11
	2.2	Het ARMA model	12
	2.3	Externe parameters	15
	2.4	Presentatie van de resultaten	18
	2.5	Analyse van Residuen	19
	2.6	Aggregatie van een tijdreeks	20
	2.7	Periodieke Effecten	20
	2.8	Analysestrategie	21
3	Resultater	a	<b>24</b>
	3.1	HVDC NorNed (EEMS)	25
	3.2	HVDC BritNed (MVL)	29
	3.3	380kV-Condensatorbank te Diemen (DMN) $\ldots \ldots \ldots \ldots \ldots$	35
	3.4	220kV-Condensatorbank te Weiwerd (WEW)	39
	3.5	150kV-Condensatorbank te Maarheeze (MZ)	45
	3.6	110kV-Condensatorbank te Harculo (HCL)	49
	3.7	380kV-kabel te Wateringen (WTR)	56
	3.8	380 kV-kabel te Bleiswijk (BWK)	61

# 4 Conclusies en Aanbevelingen

 $\mathsf{V}\mathcal{O}\mathsf{Rtech}$ 

Memo MV/M15.045

Versie 1.1, 7 december 2015

# Referenties

68

#### 6

# Hoofdstuk 1

# Inleiding

VORtech heeft voor TenneT een statistische analyse uitgevoerd op de gemeten spanningskwaliteit van zeven speciale projecten. Het doel van deze analyse is om te onderzoeken of de speciale projecten wel of niet een aantoonbare invloed hebben op de spanningskwaliteit, door gebruik te maken van verschillende statistische methoden.

# 1.1 Omschrijving van de Data

Voor ieder speciaal project zijn metingen van de spanningskwaliteit verricht, deze metingen vormen een tijdreeks (zie Sect. 2.1) met een interval van 10 minuten.

TenneT heeft aan VORtech tijdreeksen geleverd van vier verschillende parameters, die samen de spanningskwaliteit representeren:

- (a) Langzame Spanningsvariatie (LSV) De LSV is de gemiddelde spanning, gemeten over een periode van 10 minuten.
- (b) **Totale Harmonische Verstoring (THD)** De THD is een gewogen gemiddelde van verschillende harmonische, gemeten over een periode van 10 minuten.
- (c) **Asymmetrie** De asymmetrie is het quotiënt van de inverse en de normale component van de spanning.
- (d) **Flicker (PLT)** De PST is een maat voor in hoeverre een mens een bepaalde verstoring op het net als vervelend beschouwd als deze kijkt naar een aangesloten gloeilamp. De PLT is een lopend gemiddelde tot twee uur terug van de PST.

Van de LSV, THD en PLT zijn metingen verricht voor de drie verschillende fasen. In totaal zijn er dus 10 tijdreeksen beschikbaar per speciaal project. In dit rapport wordt uitsluitend gekeken naar de data die in 2013 is gemeten en door TenneT beschikbaar is gesteld aan VORtech.

Versie 1.1, 7 december 2015

De zeven speciale projecten die zijn onderzocht door VORtech zijn:

- 1. HVDC NorNed (EEMS) Een High Voltage Direct Current-Alternating Current installatie van 380 kV, die aangesloten is tussen Noorwegen, aangesloten te Feda, en Nederland, aangesloten te Eemshaven.
- 2. HVDC BritNed (MVL) Een High Voltage Direct Current-Alternating Current installatie van 380 kV, tussen Groot-Brittannië en Nederland, aangesloten op de Maasvlakte.
- 3. 380kV-Condensatorbank te Diemen (DMN).
- 4. 220kV-Condensatorbank te Weiwerd (WEW).
- 5. 150kV-Condensatorbank te Maarheeze (MZ).
- 6. 110kV-Condensatorbank te Harculo (HCL).
- 7. 380kV-kabel aangesloten tussen de 380kV-stations Wateringen (WTR) en Bleiswijk (BWK).

Voor de verschillende projecten zullen we de bovenstaande afkortingen hanteren in de rest van dit rapport. Een omschrijving van de speciale projecten is beschikbaar in Ref. [3].

Voor de statistische analyse is bekend op welke tijdstippen in het jaar 2013 deze speciale projecten aan het net geschakeld waren en wanneer niet. Daarnaast is ook bekend wanneer verschillende andere componenten aan hetzelfde net geschakeld waren.

In Fig. 1.1 staat een representatief voorbeeld van een tijdreeks, samen met de informatie van de aansluiting van een apparaat.

#### **1.2** Indeling van dit Rapport

Dit rapport is als volgt opgebouwd. In hoofdstuk 2 zal worden ingegaan op verschillende statistische methoden die zijn gebruikt om de tijdreeksen te analyseren, toegelicht met enkele voorbeelden. In hoofdstuk 3 zullen deze statistische methoden worden gebruikt om de tijdreeksen van de speciale projecten te analyseren. Tot slot zal in hoofdstuk 4 een overzicht worden gegeven van onze bevindingen.



Figuur 1.1: De assymetrie gemeten te Eemshaven, samen met het schakelen van de bekende externe factoren: EDC, EC6, EC7, SP401 en SP402. Het schakelen van het speciale project NorNed (EEM\_EDC) is aangegeven in oranje.

#### 9

# Hoofdstuk 2

# Onderzoeksmethodiek

De eerste stap in het beantwoorden van de onderzoeksvraag is het concreter maken van de vraag zelf. Hierbij laten we ons leiden door het rapport van Laborelec [4].



Meetwaarde Spanningskwaliteit

Figuur 2.1: Representatie van het mogelijke effect van een speciaal project op een spanningskwaliteitsaspect. Dit figuur is vergelijkbaar met figuur 13 uit [4].

In Fig. 2.1 geven we een voorbeeld van een mogelijk effect van een speciaal project op één van de spanningskwaliteitsaspecten (zie Sect. 1.1). Op de horizontale as van het figuur staat de meetwaarde van het relevante spanningskwaliteitsaspect uitgezet. De blauwe lijn

#### VORtech

Memo MV/M15.045

in dit figuur beschrijft de situatie vóór de aansluiting van het speciale project, of wanneer het speciale project niet aan het electriciteitsnetwerk is gekoppeld. Hoe hoger de blauwe lijn, hoe groter de kans dat de corresponderende waarde op de horizontale as voorkomt. De rode lijn in hetzelfde figuur beschrijft een mogelijke invloed van het speciale project. De groene lijn is de som van de rode en blauwe lijn, en beschrijft de situatie na de aansluiting van het speciale project. Merk op dat het hoogste punt van de groene lijn overeenkomt met een hogere meetwaarde dan het hoogste punt van de blauwe lijn. Dit wil zeggen dat in dit voorbeeld, door toedoen van het speciale project, de waarschijnlijkheid op een hoge waarde van het relevante spanningskwaliteitsaspect is vergroot<sup>1</sup>. Een manier om dit effect in de praktijk te kwantificeren is door een grote hoeveelheid metingen van de spanningskwaliteit te nemen en te onderzoeken of de gemiddelde waarde van het spanningskwaliteitsaspect beïnvloed wordt door de aan- of afwezigheid van het speciale project. Naast metingen van de verschillende spanningskwaliteitsaspecten is voor dit onderzoek ook informatie nodig over op welke tijdstippen het speciale project aan het electriciteitsnetwerk gekoppeld was en wanneer niet. Daarnaast is vergelijkbare informatie nodig over zo veel mogelijk andere bronnen die mogelijk de spanningskwaliteit zouden kunnen beïnvloeden.

In dit hoofdstuk beschrijven we de methodiek die we hebben gebruikt om te onderzoeken welke van de potentiële storingsbronnen de gemiddelde waarde van de spanningskwaliteit beïnvloeden. We zullen daartoe beginnen door een korte uitleg te geven over tijdreeksanalyse, met enkele voorbeelden om besproken concepten te verduidelijken. De voorbeelden in dit hoofdstuk hebben geen directe relatie met de data van TenneT, maar zijn random gegenereerd.

In Sectie 2.1 beginnen we met het bespreken van wat een tijdreeks is. We geven daar ook een kort overzicht op welke manier een aantal basale kenmerken van de tijdreeks bepaald kunnen worden. Vervolgens geven we een introductie tot het ARMA (Auto Regressief Moving Average) model in Sectie 2.2. Het ARMA model is een veelgebruikt model voor het modelleren van tijdreeksen en zal ook door ons worden gebruikt om de data van TenneT te analyseren. In Sectie 2.3 bespreken we het ARMAX model, een uitbereiding op het ARMA model wat ook externe variabelen (zoals het schakelen van componenten) kan meenemen. Vervolgens illustreren we in Sectie 2.4 welke informatie met behulp van een ARMAX model verkregen kan worden uit een tijdreeks en hoe deze informatie in hoofdstuk 3 gepresenteerd zal worden. Dit doen we aan de hand van het doorwerken van een simpel voorbeeld, uitgewerkt in het statistische software pakket R. In Sectie 2.5 bespreken we kort hoe de kwaliteit van de fit kunnen beoordelen. Daarna bespreken we in Sectie 2.6 hoe een tijdreeks door middel van aggregatie getransformeerd kan worden tot een tijdreeks die beter te modelleren is. Daarnaast bespreken we in Sectie 2.7 hoe kan worden omgegaan met periodieke effecten in een tijdreeks. Als laatste bespreken we de algemene onderzoeksstrategie die we gebruiken voor de analyse van de verschillende tijdreeksen (Sectie 2.8).

10

<sup>&</sup>lt;sup>1</sup>De verschillende lijnen in het figuur kunnen dusdanig worden geschaald dat hun integraal overeenkomt met 100%. In dit geval worden ze ook wel waarschijnlijkheidsverdelingen genoemd.

VORtech



Figuur 2.2: Twee voorbeelden van tijdreeksen, links: zonder autocorrelatie, rechts: met autocorrelatie. Het betreft hier geen echte data van TenneT, maar een voorbeeld van willekeurig gekozen getallen.

### 2.1 Tijdsreeksen

Met een **tijdreeks** bedoelen we een reeks van meetwaarden, die op opeenvolgende tijdstippen is gemeten. In principe hoeven de verschillende metingen in een tijdreeks niet op vaste tijdsintervallen te zijn gemeten. Een reden voor een niet vast tijdsinterval is bijvoorbeeld dat de meetapparatuur tijdelijk buiten gebruik was.

Tijdsreeksen komen veelvuldig voor in de natuurwetenschappen en in de economie. Enkele voorbeelden vanuit de natuurwetenschappen zijn: de luchtvochtigheid gedurende een dag, gemiddelde temperatuur per jaar, etc. Vanuit de economie kan men denken aan bijvoorbeeld de koersen van verschillende aandelen of beurzen. Ook de meetdata van TenneT zijn voorbeelden van tijdreeksen.

We noteren een tijdreeks als volgt:

$$y = \{y_1, y_2, y_3, \dots, y_{N-2}, y_{N-1}, y_N\},$$
(2.1)

waarbij met  $y_1, y_2, ...$  de opeenvolgende metingen worden aangeduid, en N de totale hoeveelheid metingen in de tijdreeks voorstelt. In Fig. 2.2 staan twee voorbeelden van hoe een tijdreeks er uit kan zien.

In veel tijdreeksen zijn de opeenvolgende metingen niet onafhankelijk. Dit wil zeggen dat in voorgaande metingen informatie zit over wat er in de toekomst waarschijnlijk gaat gebeuren. Een manier op erachter te komen of opeenvolgende metingen wel of niet van elkaar afhangen is door een zogenaamde lag-plot te bekijken. In deze plot worden de waarden van opeenvolgende metingen tegen elkaar uitgezet, dat wil zeggen: punten op coördinaten  $(y_1, y_2), (y_2, y_3), ..., (y_{N-1}, y_N)$  worden geplot in een figuur. In Fig. 2.3 staan

Versie 1.1, 7 december 2015

twee lag-plots, van de tijdreeksen uit Fig. 2.2. Het valt direct op dat op het linker panel een vrijwel "ronde" puntenwolk te zien is, terwijl er op het rechter paneel een duidelijke lijn zit. De lag-plots laten zien dat de opeenvolgende waarden van de eerste tijdreeks niet van elkaar afhangen, terwijl dit bij de tweede tijdreeks wel zo is. Met andere woorden: wanneer bij de tweede tijdreeks  $y_{t-1}$  laag/hoog is, dan zal de waarde van  $y_t$  waarschijnlijk ook laag/hoog zijn, maar bij de eerste tijdreeks is dit niet aan de hand. De afhankelijkheid die te zien is in het rechter paneel van Fig. 2.3 wordt ook wel aan **autocorrelatie** genoemd. Voor een tijdreeks y kan de ACF als volgt worden berekend<sup>2</sup>:

$$\rho(n) = \frac{\sum_{t=1}^{N-|n|} (y_t - \bar{y})(y_{t+n} - \bar{y})}{\sum_{t=1}^{N-|n|} (y_t - \bar{y})^2}, \quad \text{voor } -N < n < N,$$
(2.2)

waarbij  $\bar{y}$  het gemiddelde van de tijdreeks voorstelt.

Een (auto)correlatie is een getal tussen de -1 en 1, een autocorrelatie van 0 betekent geen afhankelijkheid, 1 betekent een sterke afhankelijkheid en -1 is ook een sterke afhankelijkheid, maar hoge waarden van de ene variable corresponderen dan met lage waarden van de andere en vice versa. Een autocorrelatie kan ook worden berekend met de waarde van 2, 3 of meer stappen terug. De autocorrelatie met n stappen terug wordt genoteerd met:  $\rho(n)$ en wordt ook wel de **autocorrelatie funcie (ACF)** genoemd. Voorbeelden van de ACF van de tijdreeksen in Fig. 2.2 zijn gegeven in Fig. 2.4. De blauwe stippellijnen geven het 95% betrouwbaarheidsinterval aan, dat wil zeggen dat voor een tijdreeks met Gaussische witte ruis valt te verwachten dat 95% van alle waarden voor  $n \neq 0$  tussen de twee lijnen valt. Hierbij wordt met **witte ruis** bedoeld dat alle waarden statistisch ongecorreleerd zijn. **Gaussische witte** ruis is witte ruis waarbij alle waarden uit een Gaussische verdeling zijn getrokken<sup>3</sup>. Merk op dat er door tijdreeks 1 inderdaad aan dit criterium wordt voldaan. Voor tijdreeks 2 wordt er duidelijk niet aan dit criterium voldaan, aangezien de eerste elf waarden van  $\rho(t)$  boven de blauwe lijn uitkomen.

#### 2.2 Het ARMA model

Twee van de meest voorkomende toepassingen van tijdreeksanalyse zijn: het extraheren van informatie over het onderliggende proces en het maken van voorspellingen over toekomstige metingen op basis van de beschikbare metingen. De analyse van de data van TenneT valt in de eerste categorie; we zijn geïnteresseerd in het extraheren van informatie over de invloed van speciale projecten. Zowel voor het extraheren van informatie als voor het maken van voorspellingen voor de toekomst moeten we een model vinden dat de gemeten data goed beschrijft. Eén van de meest gebruikte typen modellen om tijdreeksen te beschrijven is het zogenaamde ARMA model. Deze naam bestaat uit twee delen: AR staat voor "autoregressief" en MA staat voor "moving average". In deze sectie zullen we aan de hand van het simpelste ARMA model uitleggen hoe dit in grote lijnen in zijn werk gaat.

12

<sup>&</sup>lt;sup>2</sup>Dit heet eigenlijk de *sample* ACF, aangezien de berekening alleen is gebaseerd op het gemeten sample. <sup>3</sup>Een Gaussische verdeling is hetzelfde als een normaalverdeling.



Figuur 2.3: Lag plots van de tijdreeksen uit Fig. 2.2, links: zonder autocorrelatie,  $\rho(1) = 0$ , rechts: met autocorrelatie,  $\rho(1) = 0.8$ .



Figuur 2.4: De autocorrelatiefuncie (ACF) van de tijdreeksen uit Fig. 2.2. De blauwe stippellijnen geven het 95% betrouwbaarheidsinterval aan.

VORtech

Memo MV/M15.045

Versie 1.1, 7 december 2015

Zoals besproken in Sectie 2.1 kan men van een tijdreeks een autocorrelatie berekenen tussen opeenvolgende meetwaarden. Uit de linker panelen van Figuren 2.3 en 2.4 blijkt dat voor tijdreeks 1 de meetwaarden ongecorreleerd zijn, d.w.z. het gaat hier om witte ruis. Dit kan wiskundig worden geschreven als:

$$y_t = \epsilon_t, \tag{2.3}$$

waarbij  $\epsilon_t$  gedefinieerd is als een serie van random waarden, die doorgaans worden verondersteld normaal te zijn verdeeld. In het rechter paneel van Fig. 2.2 staat een voorbeeld van een tijdreeks met een autocorrelatie. Als een tijdreeks een autocorrelatie bevat, dan zit er in de voorgaande waarden informatie over de toekomstige waarden van de tijdreeks. In dat geval heeft het dus zin om een model op te stellen, waarbij de verwachte waarde van  $y_t$  uitgedrukt wordt in de gemeten waarden op eerdere tijdstippen plus een witte ruis term. In het simpelste geval kunnen we het volgende model opschrijven:

$$y_t = \phi y_{t-1} + \epsilon_t, \tag{2.4}$$

waarbij  $\phi$  een constante waarde heeft die voor dit model overeenkomt met de autocorrelatie op n = 1, d.w.z.:  $\rho(1)$ . Het model in vergelijking 2.4 wordt een autoregressief model genoemd, aangezien de "voorspeller" of "regressor" voor toekomstige waarden uit de tijdreeks zelf komt. Het model kan ook uitgebreid worden, namelijk door ook eerdere metingen mee te nemen als voorspellers. Het model wordt dan:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p} + \epsilon_t = \sum_{j=1}^p \phi_j y_{t-j} + \epsilon_t, \quad (2.5)$$

waarbij p staat voor de hoeveelheid termen die worden meegenomen en  $\phi_j$  verschillende modelparameters zijn, die min of meer de sterkte van de invloed weergeven. Dit model wordt ook wel genoteerd als "AR(p)", en het model in vergelijking 2.4 is dus het AR(1) model.

Een verdere veralgemenisering van het model kan gedaan worden door ook voorgaande witte ruis waarden ( $\epsilon_{t-1}, \epsilon_{t-2}, ...$ ) mee te nemen als voorspellers. Het model kan dan geschreven worden als:

$$y_t = \sum_{j=1}^p \phi_j y_{t-j} + \sum_{k=1}^q \theta_k \epsilon_{t-k} + \epsilon_t, \qquad (2.6)$$

waarbij q staat voor de hoeveelheid voorgaande witte ruis waarden die worden meegenomen als voorspellers en  $\theta_k$  modelparameters zijn. De termen die afhangen van voorgaande waarden van de reeks zelf worden wederom autoregressief genoemd, en de termen die afhangen van voorgaande waarden van de witte ruis worden "moving average" (MA) genoemd. Het volledige model in vergelijking 2.6 wordt daarom ook wel "ARMA(p, q)" genoemd.

Het modelleren van een gemeten tijdreeks gaat in twee stappen. Als eerste moet er bepaald worden welk model er het beste bij de data past. In het geval van het ARMA model betekent dit dat we moeten kiezen hoeveel termen worden meegenomen, d.w.z., de waarden p

VORtech

Versie 1.1, 7 december 2015

en q moeten worden gekozen. Deze keuze kan worden gemaakt op basis van inspectie van de ACF<sup>4</sup>, of op basis van het zogenaamde AICc criterium [2]. Het AICc criterium van een model is groter voor complexere modellen en kleiner als het waarschijnlijker is dat de data een realisatie is van het model<sup>5</sup>. Wanneer er besloten is welk model gebruikt gaat worden, moeten de optimale waarden van de relevante modelparameters ( $\phi_1, \phi_2, \dots, \theta_1, \theta_2, \dots$ ) bepaald worden. Dit gebeurd met de standaard fit algoritmen zoals geïmplementeerd in softwarepakketten als R [1] en Python. Het bepalen van de optimale waarden voor de modelparameters wordt ook wel het **fitten** van een model genoemd.

#### 2.3 Externe parameters

De vraag van TenneT betreft een eventuele invloed van externe invloeden op de spanningskwaliteit. Merk op dat in ons geval een speciaal project en andere componenten uit het electriciteitsnetwerk externe invloeden zijn. Van een gedeelte van de externe invloeden is bekend op welke tijdstippen ze aan- en afgekoppeld waren. Als we "aan" aanduiden met 1 en "uit" met 0, dan zou het schakelgedrag van een speciaal project er bijvoorbeeld uit kunnen zien zoals in het linker paneel van Fig. 2.5. In het rechter paneel van dit figuur staat een voorbeeld van een autoregressieve tijdreeks met  $\phi = 0.8$ , waarbij het gemiddelde van de tijdreeks (zie blauwe stippellijn) samenvalt met het "aan" en "uit" staan van het speciale project (zie ook Fig. 2.1).

In Fig. 2.6 laten we een histogram zien van alle waarden van deze tijdreeks, waarbij we onderscheid maken tussen de waarden tijdens het aan- en uitstaan van het speciale project. Uit dit figuur wordt duidelijk dat het speciale project inderdaad een invloed heeft zoals in Fig. 2.1.

In de modellen zoals beschreven door de uitdrukkingen 2.4–2.6 worden nog geen invloeden van buitenaf meegenomen. Om de speciale projecten en andere omgevingsinvloeden mee te nemen moeten we een uitbreiding maken op het ARMA model. We noteren voor de *l*-de externe invloed op tijdstip *t* als  $u_{t,l}$ . We definiëren een nieuwe variabele  $z_t$ , die afhangt van de meetwaarde  $y_t$ , een constante factor  $\mu$ , externe invloeden  $u_{t,l}$  en corresponderende weegfactoren  $\gamma_l$ :

$$z_t = y_t - \mu - \sum_{l=1}^{M} \gamma_l u_{t,l},$$
(2.7)

waarbij M het totale aantal bekende externe invloeden voorstelt. Vervolgens wordt het ARMA model met externe invloeden, ook wel het ARMAX model, gegeven door vergelijking 2.6 waarbij  $y_t$  wordt vervangen door  $z_t^6$ . Merk op dat de externe invloeden  $u_{t,l}$ 

<sup>&</sup>lt;sup>4</sup>Normaalgesproken worden zowel de ACF en cd PACF (partiële autocorrelatiefunctie) bestudeerd bij het kiezen van het juiste model. Voor meer informatie over de procedure refereren we naar [2].

<sup>&</sup>lt;sup>5</sup>De AICc waarde van een model wordt gegeven door: AICc =  $\frac{2kn}{n-k-1} - 2\ln L$ , waarbij k staat voor de hoeveelheid parameters van het model, n staat voor de hoeveelheid observaties in de reeks, en L staat voor de likelihood van het model, d.w.z., de waarschijnlijkheid op de huidige tijdreeks, gegeven het model.

<sup>&</sup>lt;sup>6</sup>Er bestaan ook andere representaties van het ARMAX model, deze zijn equivalent, maar de interpre-

VORtech



Figuur 2.5: Linker paneel: een voorbeeld van hoe het schakelgedrag van een speciaal project er uit kan zien als functie van de tijd. Rechter paneel: een autoregressieve tijdreeks met  $\phi = 0.8$  en een invloed van het speciale project. Met de blauwe stippellijn wordt het gemiddelde van het onderliggende model aangeduid. Het betreft hier geen echte data van TenneT, maar een voorbeeld van willekeurig gekozen getallen.



Figuur 2.6: Histogram van alle waarden uit de tijdreeks in het linker paneel van Fig. 2.5. De waarden tijdens het aan/uit staan van het speciale project zijn weergegeven in rood/blauw.

VORtech

Memo MV/M15.045

Versie 1.1, 7 december 2015

zelf ook tijdreeksen zijn. Een voorbeeld van een externe invloed is het schakelgedrag van een speciaal project, zie het linker paneel van Fig. 2.5. De ARMAX(1,0) tijdreeks op het rechter paneel van ditzelfde figuur is gemaakt met de parameters  $\phi = 0.8$ ,  $\theta = 0$ ,  $\mu = 100$ , M = 1 en  $\gamma_1 = 5$ , oftewel:

$$y_t = 100 + 5u_t + 0.8(y_{t-1} - 100 - 5u_{t-1}) + \epsilon_t.$$
(2.8)

Dit wil zeggen dat als we van deze tijdreeks een factor 100 aftrekken en ook vijf maal de tijdreeks van het schakelgedrag er van aftrekken, dat het resultaat dan beschreven kan worden beschreven door een ARMA(1,0) tijdreeks.

In de analyse van de tijdreeksen van TenneT zijn we, gegeven het schakelgedrag van een component, geïnteresseerd in het achterhalen van de waarde van de bijbehorende modelfactor  $\gamma$ . Dit kan gedaan worden met dezelfde algoritmen die eerder zijn genoemd in Sectie 2.2. Onder de voorwaarde dat de tijdreeks van de spanningskwaliteit en het schakelgedrag van een component covariëren kan de modelfactor  $\gamma$  worden geïnterpreteerd als de mate waarin het gemiddelde niveau van de spanningskwaliteit geassocieerd is met het schakelen van de component. Met covariëren bedoelen we dat variaties in de ene reeks overeenkomen met variaties in de andere reeks. Als bij een bepaalde component een modelparameter is gevonden die ongelijk is aan 0, dan zijn er twee mogelijkheden, het zou kunnen dat de component een invloed heeft op de spanningskwaliteit, maar het kan ook zijn dat er geschakeld is omdat de spanningskwaliteit een bepaald gedrag vertoond. Hierover kan de statistiek niets zeggen, en de interpretatie van de gevonden resultaten behoort tot het domein van de elektrotechniek.

Versie 1.1, 7 december 2015

In deze sectie laten we zien hoe binnen het softwarepakket R een ARMAX fit kan worden gedaan, en hoe de resultaten uiteindelijk in hoofdstuk 3 zullen worden gepresenteerd.

```
fit <- Arima(timeSeries3, order=c(1,0,0), xreg = externalInfluence)</pre>
summary(fit)
##
   Series: timeSeries3
   ARIMA(1,0,0) with non-zero mean
##
##
##
   Coefficients:
                        ar1 intercept externalInfluence
##
                     0.7754
                              99.9510
                                                   4.8523
                     0.0284
                               0.2521
                                                   0.2873
##
    s.e
```

Box 2.1: Een voorbeeld van de manier waarop de tijdreeks gefit kan worden met een ARMAX model. De eerste twee regels betreffen de R code en de daaropvolgende regels in het groen zijn de output van R. De gebruikte tijdreeks "timeSeries3" komt overeen met de tijdreeks uit het rechter paneel van Fig. 2.6, en de externe invloed "externalInfluence" komt overeen met het linker paneel van hetzelfde figuur. De getallen direct onder "ar1", "intercept" en "externalInfluence" komen respectievelijk overeen met de geschatte waarde van  $\phi$ ,  $\mu$  en  $\gamma_1$  (zie Vergelijking 2.8 en bijbehorende uitleg). De getallen onder de geschatte waarden komen overeen met de standaard error (s.e.) op de geschatte waarden.

In Box 2.1 staat een stuk R code wat een ARMAX fit uitvoert van de tijdreeks op het rechter paneel van Fig. 2.6, met als input de externe invloed die is weergegeven in het linker paneel van ditzelfde figuur. Zoals besproken in Sectie 2.3 is deze tijdreeks gegenereerd uitgaande van de modelparameters  $\phi^{w} = 0.8$ ,  $\mu^{w} = 100$  en  $\gamma_{1}^{w} = 5$ , waar het superscript "w" is toegevoegd om te verduidelijken dat het hier om de *werkelijke waarden* van de parameters gaat. De geschatte waarden van de modelparameters zijn:  $\hat{\phi} = 0.7754$  ("ar1"),  $\hat{\mu} = 99.9510$ ("intercept"), en  $\hat{\gamma}_{1} = 4.8523$  ("externalInfluence"), waar het dakje aangeeft dat het gaat om geschatte waarden. Naast de geschatte waarden van de modelparameters wordt ook de **standaard error (s.e.)** geschat voor iedere parameter. De standaard error is een maat voor de onzekerheid op geschatte waarden. Om precies te zijn: het is een schatter voor de standaard-afwijking van de geschatte waarde van de parameter.

In Fig. 2.7 staat een visuele representatie van het resultaat van Box 2.1. De rode balk geeft de grootte van de geschatte parameter  $\hat{\gamma}_1$ . De zwarte lijn geeft het 95% **betrouwbaar**heidsinterval (BI), wat gedefinieerd is als  $\hat{\gamma}_1 \pm 1.96$  s.e. In ons voorbeeld is de *werkelijke waarde* van  $\gamma_1$  bekend, namelijk  $\gamma_1^w = 5$ . Als we data zouden verzamelen van volgende jaren en voor ieder jaar dezelfde analyse zouden uitvoeren, dan zal in 95% van de gevallen het verkregen betrouwbaarheidsinterval de ware parameter  $\gamma_1^w$  bevatten.

Bij de analyse van de data van TenneT zijn de werkelijke waarden van de verschillende parameters  $\gamma_l$  niet bekend. In dit geval kunnen we wel een **hypothesetoets** doen, d.w.z.: we kunnen een aanname doen over de werkelijke waarde en de verdeling van  $\gamma_l$ , waarna de waarschijnlijkheid wordt bepaald dat onder de gekozen hypothese een waarde voor de fit parameter wordt gevonden gelijk aan of extremer dan de geobserveerde waarde. Als deze kans onder een bepaald niveau komt (het **significantieniveau**), dan noemen we



Figuur 2.7: Visuele representatie van het fitresultaat. De rode balk geeft de geschatte waarde van de modelparameter  $\gamma_1$  weer en de zwarte lijn geeft het 95% betrouwbaarheidsinterval aan, d.w.z.  $\hat{\gamma}_1 \pm 1.96$  s.e. De rode asterisk geeft aan dat het resultaat significant van 0 verschilt.

het resultaat **significant**, en verwerpen we de hypothese. De standaard keuze voor het significantieniveau is 0.05, ofwel 5%.

Dit illustreren we aan de hand van ons voorbeeld zoals hierboven behandeld. Als hypothese nemen we dat  $\gamma_1$  normaal is verdeeld met  $\gamma_1^w = 0$  en een standaard deviatie heeft gelijk aan de geschatte standaard error. We veronderstellen dus dat "tijdreeks 3" en het schakelen van de component niet covariëren. We weten dat onder de hypothese er een waarschijnlijkheid is van 95% dat de geobserveerde waarde  $\hat{\gamma}_1$  binnen het interval  $\hat{\gamma}_1 \pm (1.96 \times 0.2873) = \hat{\gamma}_1 \pm 0.56$ zit. Het is duidelijk dat de geobserveerde waarde ( $\hat{\gamma}_1 = 4.8523$ ) niet binnen dit interval valt. Derhalve verwerpen we daarom de hypothese dat tijdreeks 3 en het schakelen van de component niet covariëren.

Merk op dat de keuze voor het gewenste significantieniveau in principe willekeurig is; we hadden ook kunnen kiezen voor 0.01 of zelfs voor 0.001. Het meest gebruikte sgnificantieniveau is echter 0.05, wat we daarom ook in de rest van dit rapport zullen hanteren.

Een simpele manier om uit de visuele representatie van het fit resultaat (Fig. 2.7) te zien of het een significant resultaat betreft, is door te kijken of 0 binnen het betrouwbaarheidsinterval valt. In dit voorbeeld is dit niet zo, dus is het resultaat significant. Ter verduidelijking worden significante resultaten aangegeven met een asterisk (\*).

### 2.5 Analyse van Residuen

Nadat we een model hebben gevonden voor de beschrijving van de data, en de bijbehorende parameters hebben geschat met een fit algoritme, kunnen we gaan kijken naar het verschil tussen de modelwaarden en de gemeten waarden. Deze verschillen worden de "residuen" genoemd. Als eerste moet gekeken worden of er nog significante autocorrelaties zitten in de residuen, dit zou in principe niet zo moeten zijn. Daarnaast is het interessant om specifiek te kijken naar de grootste residuen. Deze grootste residuen kunnen toevalligerwijs ontstaan, maar kunnen ook wat te maken hebben met een externe invloed die niet bekend is.

Versie 1.1, 7 december 2015

# 2.6 Aggregatie van een tijdreeks

Het **aggregeren** van een tijdreeks wil zeggen dat de oorspronkelijke tijdreeks wordt opgedeeld in verschillende aaneengesloten subsets, waarna een nieuwe tijdreeks wordt geconstrueerd met als elementen een bepaalde eigenschap van deze subset. Dit illustreren wij aan het hand van het volgende voorbeeld. Stel dat we beschikken over een tijdreeks y met 10-minuten waarden:

$$y = \{1, 5, 3, 2, 6, 4, -5, 3, 5, \ldots\}.$$
(2.9)

Nu kunnen we besluiten om een nieuwe tijdreeks  $\tilde{y}$  te construeren met de gemiddelde waarde per half uur. De eerste waarde van deze reeks wordt gegeven door  $\tilde{y}_1 = (1+5+3)/3 = 3$ , de tweede door  $\tilde{y}_2 = (2+6+4)/3 = 4$ , etc. De geaggregeerde reeks ziet er dus als volgt uit:

$$\tilde{y} = \{3, 4, 1, \ldots\}.$$
(2.10)

Merk op dat de nieuwe reeks in dit geval drie maal minder datapunten bevat, aangezien elke drie datapunten geaggregeerd zijn tot één punt.

Er zijn verschillende redenen waarom het statistisch gezien verstandig kan zijn om een tijdreeks te aggregeren. Een belangrijke reden is als een reeks een zeer sterke autocorrelatie heeft, d.w.z. de autocorrelatiefunctie  $\rho(n)$  heeft waarden dicht bij 1 in een groot interval [0, n]. Dit gebeurt typisch wanneer opeenvolgende meetwaarden dusdanig kort na elkaar zijn gemeten dat ze nauwelijks verschillen. Voor een sterk gecorreleerde tijdreeks is het vaak lastig om een betrouwbaar model op te stellen. Een geaggregeerde tijdreeks heeft een minder sterke autocorrelatie dan de originele tijdreeks, en kan daarom gemakkelijker gemodelleerd worden.

Een andere reden kan zijn dat blijkt dat de ongecorreleerde ruis  $\epsilon_t$  niet normaal is verdeeld. Reguliere fit-algoritmen, zoals geïmplementeerd in de statistische analysebibliotheken van R en Python, gaan er van uit dat de waarden in  $\epsilon_t$  normaal verdeeld zijn. Wanneer dit niet het geval is, kunnen de geschatte waarden van de modelparameters en de geschatte onzekerheden op deze parameters minder betrouwbaar worden.

Een bijkomend voordeel van aggregeren kan zijn dat bepaalde periodieke structuur uit de data verdwijnt. Dit gebeurt als er geaggregeerd wordt over een tijdsinterval dat gelijk is aan (een veelvoud van) de periode van de periodieke structuur. Het weg-aggeregeren van een periodieke structuur is echter typisch niet de hoofdreden om te gaan aggregeren, aangezien er verschillende technieken bestaan om een periodieke structuur mee te nemen in het model. Twee van zulke technieken worden besproken in de volgende sectie.

### 2.7 Periodieke Effecten

Uit de analyses van de TenneT data blijkt dat vrijwel alle tijdreeksen van de spanningskwaliteit een duidelijke periodiciteit bevatten met de periode van een dag. Dit wil zeggen dat gemiddeld gezien de spanningskwaliteit iedere dag een zelfde soort verloop heeft. Na

20

VORtech

Memo MV/M15.045

Versie 1.1, 7 december 2015

aggregeren tot dagwaarden blijken veel van de geaggregeerde reeksen een periodiciteit te hebben met de periode van een week.

Er zijn verschillende manieren om deze periodiciteit mee te nemen in het ARMA model. In het geval dat het periodieke gedrag zeer voorspelbaar is, dat wil zeggen, het is elke dag vrijwel identiek, dan kan er voor worden gekozen om een tijdsreeks van het periodieke gedrag mee te nemen als een externe invloed  $u_t$ . Als het periodieke gedrag minder voorspelbaar is, dan kan het ook meegenomen worden als een stochastisch proces. Dit wil zeggen dat we niet alleen een klein aantal voorgaande waarden meenemen als voorspellers, maar ook waarden van een of enkel periodes geleden. Bijvoorbeeld, een AR(1) model met een periode van 24 tijdstappen kan worden gemodelleerd door:

$$y_t = \phi_1 y_{t-1} + \Phi_1 y_{t-24} + \epsilon_t, \tag{2.11}$$

waarbij  $\Phi_1$  een maat is voor de correlatie tussen de waarde van een periode geleden en de waarde op tijdstip t.

Het meest algemene model dat gebruikt zal worden in onze analyse is het zogenaamde SARMAX model, waarbij de "X" wederom staat voor "externe factoren", en de "S" staat voor "seizoensafhankelijkheid". Het model wordt als volgt beschreven:

$$z_{t} = \sum_{j=1}^{p} \phi_{j} z_{t-j} + \sum_{k=1}^{q} \theta_{k} \epsilon_{t-k} + \sum_{j=1}^{P} \Phi_{j} z_{t-a \cdot j} + \sum_{k=1}^{Q} \Theta_{k} \epsilon_{t-a \cdot k} + \epsilon_{t}, \qquad (2.12)$$

waarbij *a* de periode is, en  $\Phi_i$  en  $\Theta_i$  modelparameters zijn en  $Y_i$  wordt gegeven in vergelijking 2.7. Dit model wordt ook wel geschreven als het SARMAX $(p,q) \times (P,Q)_a$  model. Merk op dat het model in vergelijking 2.11 een speciaal geval is van het algemene model in vergelijking 2.12, namelijk het SARMAX $(1,0) \times (1,0)_{24}$  model  $\mu = 0$ , en zonder externe parameters, dat wil zeggen:  $z_t = y_t$ .

In dit rapport concluderen we dat met de huidige data de meest betrouwbare strategie is om de naar dagwaarden geaggregeerde tijdsreeks te analyseren. Hierdoor is het niet nodig om de dagelijkse periodiciteit mee te modelleren. Bij de analyses van de ruwe data en de naar 60-minuten geaggregeerde data is echter wel rekening gehouden met de dagelijkse periodiciteit. In de meeste gevallen leek een fit met een SARMAX model hierbij een betere beschrijving te geven dan het ARMAX model waarbij een periodieke trend wordt meegenomen als externe invloed.

#### 2.8 Analysestrategie

In totaal zijn er voor elk speciaal project tien spanningskwaliteitsaspecten gemeten door TenneT. De analyse van elk van deze tijdreeksen is in principe handwerk, aangezien elke tijdreeks uniek gedrag vertoont. In grote lijnen kunnen we echter één algemene analysestrategie definiëren. In deze sectie bespreken we deze algemene analysestrategie aan de hand van de modellen en concepten die eerder in dit hoofdstuk zijn geïntroduceerd.

21

Versie 1.1, 7 december 2015

- 1. Overzicht van de tijdreeks We beginnen met het maken van een overzicht van de tijdreeksen van de spanningskwaliteitsaspecten en de bekende (mogelijke) externe invloeden. We verwijderen foutieve meetwaarden die bijvoorbeeld ontstaan zijn door het incorrect werken van de meetapparatuur. TenneT heeft gegevens van het schakelgedrag van mogelijke externe invloeden geleverd en we bekijken welke van deze invloeden meegenomen kunnen worden in de analyse. Een reden om bepaalde mogelijke externe invloeden niet mee te nemen in de analyse is als ze (vrijwel) niet schakelen; in dat geval is het namelijk onmogelijk om de situatie te vergelijken waarbij deze aangekoppeld of afgekoppeld is. Een andere mogelijkheid is dat twee mogelijke externe invloeden op exact dezelfde manier schakelen. In dit geval nemen we maar één van de twee invloeden mee en kan de analyse geen onderscheid maken tussen tussen de twee invloeden. De invloed die we dan bepalen moet worden geïnterpreteerd als een gecombineerde invloed van beide factoren. Als laatste bestuderen we ook de correlatiestructuur van de tijdreeks (zie Sectie 2.1), om een indruk te krijgen van welk model goed bij de data zou passen en of er een periodieke structuur in de data zit.
- 2. Bekijken van verdelingen Voor het speciale project bekijken we of er al nuttige informatie gevonden kan worden door een verdeling te maken van de spannings-kwaliteit wanneer het speciale project aangesloten/afgeschakeld is (zie bijvoorbeeld Fig. 2.6). Deze methode is voornamelijk zinvol als er voldoende metingen zijn waarbij het speciale project is aangesloten en afgeschakeld, en wanneer er voldoende wille-keurig wordt geschakeld gedurende de gehele meetperiode.
- 3. Fit met een ARMAX model Vervolgens bepalen we een geschikt ARMAX model voor de data. Zoals besproken in Sect. 2.2 begint dit met het bepalen van de waarden voor p en q (zie de vergelijking 2.6). Vervolgens bepalen we waarden van de modelparameters  $\phi_i$ ,  $\theta_i$  en  $\gamma_l$  (zie de vergelijking 2.7) door het uitvoeren van een fit.
- 4. Bekijken van de residuen Nadat we een model hebben gevonden voor de beschrijving van de data bestuderen we de residuen, zoals beschreven in Sect. 2.5.
- 5. Aggregatie Het valt op dat in alle gemeten tijdreeksen die door TenneT zijn geleverd de autocorrelatie zeer groot is. Zoals besproken in Sect. 2.6 is een reden hiervoor dat het interval tussen de metingen redelijk kort is. Daarnaast bleek het ook dat de residuen<sup>7</sup> vrijwel altijd niet normaal verdeeld waren. Dit zijn twee redenen om de data te aggregeren alvorens een model te fitten.

In deze analyse hebben voor alle tijdreeksen ook een 60-minuten geaggregeerde reeks en de 1-dag geaggregeerde reeks geanalyseerd. Vrijwel altijd kon hierdoor een simpeler model worden gekozen om de data te beschrijven. Daarnaast blijkt het dat voor de geaggregeerde reeksen de de residuen van de fit veel beter beschreven kunnen worden met een normale verdeling, wat de conclusies die voortkomen uit het model

<sup>&</sup>lt;sup>7</sup>Ofwel de waarden van de  $\epsilon_t$  term.

#### VORtech

Memo~MV/M15.045

betrouwbaarder maken. De verbetering is het grootst wanneer er wordt geaggregeerd tot dag-waarden.

- 6. **Periodiciteit** Wanneer uit de (P)ACF blijkt dat de geaggregeerde reeks een periode bevat, dan wordt dit op een stochastische manier meegenomen in het model (zie Sect. 2.7).
- 7. **Overzicht** Tot slot geven we voor ieder spanningskwaliteitsaspect van ieder project een overzicht van de grootte van de invloed van de bekende externe factoren op het gemiddelde van de tijdreeks, met bijbehorende onzekerheidsmarges. Deze bespreking richt zicht op de resultaten van de tijdeeksen van de naar dag-waarden geaggregeerde data. Er is tijdens de analyse naar verschillende modellen gekeken, maar in het overzicht worden alleen de uitkomsten gepresenteerd van de modellen met de optimale AICc waarde. De zoekruimte van modellen is hierbij beperkt tot modellen met  $p \leq 3$ en  $q \leq 3$ .

# Hoofdstuk 3

# Resultaten

In dit hoofdstuk bespreken we de resultaten van de analyse van de spanningskwaliteit voor de verschillende speciale projecten, waarbij elk speciaal project wordt besproken in een aparte sectie. Voor de langzame spanningsvariatie, asymmetrie, en totale harmonische verstoring worden de resultaten van de analyses van de tot dag-waarden geaggregeerde tijdreeksen gepresenteerd in de vorm van een staafdiagram, met daarin de gefitte waarde van de parameters  $\gamma$  van iedere component die is meegenomen in de analyse (zie Vergelijking 2.7). Deze parameters zijn een maat voor de covariatie van het schakelgedrag van de component en de spanningskwaliteit. Zoals besproken in Sectie 2.4 is de onzekerheid op de gefitte parameters gegeven in termen van het 95% betrouwbaarheidsinterval, en zal per parameter met een asterisk worden aangegeven of het resultaat significant is met een significantie niveau van 0.05.

De modellen die gebruikt zijn voor het genereren van de resultaten zijn gekozen op basis van het AICc criterium (zie Sect. 2.2), wat wil zeggen dat het model een optimale balans heeft tussen een lage complexiteit en een zo goed mogelijke fit aan de data. De tijdreeksen van de PLT bleken niet gemodelleerd te kunnen worden binnen de klasse van (S)ARMAX modellen, doordat de PLT een bewerkt signaal is (zie Sect. 1.1). Voor enkele projecten bleek dat het schakelen van bepaalde componenten samenging met een duidelijke sprong in de PLT. Dit hebben we geïllustreerd door enkele representatieve voorbeelden te laten zien.

Bij de analyse van alle tijdreeksen is expliciet gekozen om geen trend uit de tijdreeks te halen, aangezien we juist geïnteresseerd zijn in veranderingen in de trend door toedoen van de verschillende componenten. Het zou kunnen dat een eventuele trend in de data niet helemaal goed beschreven kan worden door (een combinatie van) de aan ons bekende componenten. Een symptoom hiervan kan zijn dat de residuen niet precies normaal zijn verdeeld, en we plaatsen een opmerking bij de speciale projecten waarbij dit aan de orde is. Bij twee speciale projecten (HCL en WEW) bleek dat er een duidelijke afname was in het gemiddelde van de LSV tegen het einde van de tijdreeks. In deze twee gevallen hebben we ervoor gekozen om het einde van de tijdreeks niet mee te nemen in de analyse. VORtech

Memo MV/M15.045

Versie 1.1, 7 december 2015

Per project zal worden aangegeven welke componenten zijn meegenomen in de analyse, en wanneer een component niet is meegenomen in de analyse, dan zal deze keuze worden gemotiveerd.

Een voorbeeld van een tijdreeks van de spanningskwaliteit (asymmetrie bij Eemshaven) is gegeven in Fig. 1.1. Een overzicht van alle tijdreeksen die zijn geanalyseerd is te vinden in Referentie [3].

# 3.1 HVDC NorNed (EEMS)

De eerste dataset die we geanalyseerd hebben betreft de hoogspanningskabel NorNed. In Fig. 3.1 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden te Eemshaven. NorNed (EEMS\_EDC) staat in dit figuur aangegeven in het oranje. Het eerste wat opvalt is dat veel mogelijke invloeden gedurende het hele jaar aangesloten zijn (EEM\_EOSZ, EEM\_MEEW, etc.). Deze mogelijke invloeden kunnen daarom niet worden meegenomen in de analyse, aangezien er geen informatie beschikbaar is over de situatie wanneer deze mogelijke invloeden afwezig zijn.



Figuur 3.1: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit te Eemshaven gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. NorNed (EEMS\_EDC) is aangegeven in het oranje.

VORtech

Memo MV/M15.045

Versie 1.1, 7 december 2015

#### $\mathbf{LSV}$

Voor de naar dag-waarde geaggregeerde data van de LSV te Eemshaven blijkt het SARMAX $(1,0) \times (0,2)_7$  model de beste fit te geven. In Fig. 3.2 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van NorNed lijkt samen te gaan met een afname van de LSV met 1 - 1.5 kV.



Figuur 3.2: Parameters van het  $SARMAX(1,0) \times (0,2)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de LSV te Eemshaven. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### Asymmetrie

Voor de naar dag-waarde geaggregeerde data van de asymmetrie te Eemshaven blijkt het ARMAX(2,1) model de beste fit te geven. In Fig. 3.3 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van NorNed lijkt samen te gaan met een toename van de asymmetrie van ongeveer 0.07.

#### THD

Voor de naar dag-waarde geaggregeerde data van de THD te Eemshaven blijkt het AR-MAX(1,2) model de beste fit te geven. In Fig. 3.4 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van Nor-Ned lijkt samen te gaan met een toename van de THD1 van ongeveer 0.1. Voor THD2 en THD3 is geen significant resultaat gevonden.



Figuur 3.3: Parameters van het ARMAX(2,1) model, gefit aan de naar dag-waarden geaggregeerde data van de Asymmetrie te Eemshaven. De parameters zijn een maat voor de covariatie van de component en de Asymmetrie. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

asym

#### PLT

In Fig. 3.5 zijn de eerste tweehonderd 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.5 wekt echter niet de suggestie dat de sprongen in de PLT gerelateerd zijn aan het schakelen van de aan ons bekende invloeden.



Figuur 3.4: Parameters van het ARMAX(1,2) model, gefit aan de naar dag-waarden geaggregeerde data van de THD te Eemshaven. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.



Figuur 3.5: Tweehonderd 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Eemshaven, vanaf 31 december 2012, 23:00 tot 2 januari 2013, 8:10. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

Versie 1.1, 7 december 2015

### 3.2 HVDC BritNed (MVL)

In Fig. 3.6 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden op de spanningskwaliteit op station Maasvlakte. BritNed (MVL\_MDC) staat in dit figuur aangegeven in het oranje. Alleen componenten waarvan een redelijke hoeveelheid aan- en uit metingen zijn kunnen goed geanalyseerd worden. Een redelijk aantal componenten zijn vijwel volledig anti-gecorreleerd en kunnen daarom niet geanalyseerd worden. De volgende componenten zijn om die reden niet meegenomen in de analyse: CST\_SMH\_MVL\_W, MVL\_SMH\_CST\_W, CST\_SMH\_MVL\_Z, MVL\_SMH\_CST\_Z.



Figuur 3.6: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit op station Maasvlakte gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. BritNed (MVL\_MDC) is aangegeven in het oranje.

#### LSV

Voor de naar dag-waarde geaggregeerde data van de LSV op station Maasvlakte blijkt het  $SARMAX(1,2) \times (1,0)_7$  model de beste fit te geven. De residuen van deze fit bevatten een klein aantal waarden die lager zijn dan verwacht zou worden bij normaal verdeelde residuen. In Fig. 3.7 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van BritNed en de waarde van de LSV.

29

Versie 1.1, 7 december 2015

Voor de naar dag-waarde geaggregeerde data van de asymmetrie op station Maasvlakte blijkt het SARMAX $(1,2) \times (2,0)_7$  model de beste fit te geven. In Fig. 3.8 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van BritNed en de waarde van de asymmetrie.

#### THD

Voor de naar dag-waarde geaggregeerde data van de THD op station Maasvlakte blijkt het SARMAX $(1,0) \times (3,0)_7$  model de beste fit te geven. In Fig. 3.9 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van BritNed lijkt samen te gaan met een afname van de THD van ongeveer 0.1 - 0.15.

#### PLT

In Fig. 3.2 zijn de eerste achthonderd 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.2 wekt echter niet de suggestie dat de sprongen in de PLT gerelateerd zijn aan het schakelen van BritNed.

VORtech



Figuur 3.7: Parameters van het  $SARMAX(1,2) \times (1,0)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de LSV op station Maasvlakte. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.



Figuur 3.8: Parameters van het  $SARMAX(1,2) \times (2,0)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de Asymmetrie op station Maasvlakte. De parameters zijn een maat voor de covariatie van de component en de Asymmetrie. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

VORtech



Figuur 3.9: Parameters van het  $SARMAX(1,0) \times (3,0)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de THD op station Maasvlakte. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

VORtech



Figuur 3.10: Achthonderd 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 op station Maasvlakte, vanaf 31 december 2012, 23:00 tot 6 januari 2013, 12:10. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

### **3.3** 380kV-Condensatorbank te Diemen (DMN)

In Fig. 3.11 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden op de spanningskwaliteit te Diemen. De speciale projecten DIM\_C01 DIM\_C02 staan in dit figuur aangegeven in het oranje. Alleen componenten waarvan een redelijke hoeveelheid aan- en uit metingen zijn kunnen goed geanalyseerd worden. De volgende componenten zijn om die reden niet meegenomen in de analyse: DIM\_LLS\_Z, LLS\_DIM\_Z DIM\_OZN\_G, OZN\_DIM\_G, DMN\_C01, DMN\_KV.

Daarnaast kunnen ook componenten die vrijwel geheel gecorreleerd zijn niet allebei worden meegenomen, tussen deze componenten kan namelijk geen onderscheid gemaakt worden. Het blijkt dat de paren DIM\_LLS\_W, LLS\_DIM\_W en KIJ\_DIM\_W, DIM\_KIJ\_W zeer sterk gecorreleerd zijn. Om deze reden zijn LLS\_DIM\_W en KIJ\_DIM\_W niet meegenomen in de analyse.



Figuur 3.11: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit te Diemen gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. De condensatorbanken (DIM\_C01 en DIM\_C02) zijn aangegeven in het oranje.

#### $\mathbf{LSV}$

Voor de naar dag-waarde geaggregeerde data van de LSV te Diemen blijkt het SARMAX $(1,0) \times (2,2)_7$  model de beste fit te geven. In Fig. 3.12 wordt een overzicht gegeven van geschatte

VORtech
---------

modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de LSV.



Figuur 3.12: Parameters van het  $SARMAX(1,0) \times (2,2)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de LSV van te Diemen. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### Asymmetrie

Voor de naar dag-waarde geaggregeerde data van de asymmetrie te Diemen blijkt het SARMAX $(1,2) \times (1,1)_7$  model de beste fit te geven. In Fig. 3.13 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van condensatorbank 2 lijkt samen te gaan met een afname van de asymmetrie met ongeveer 0.05.

#### THD

Voor de naar dag-waarde geaggregeerde data van de THD te Diemen blijkt het SARMAX $(1,0) \times (1,1)_7$  model de beste fit te geven. De residuen van deze fit lijken niet helemaal normaal verdeeld, maar hebben wat zwaardere staarten. In Fig. 3.14 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis

#### VORtech

Memo MV/M15.045



Figuur 3.13: Parameters van het  $SARMAX(1,2) \times (1,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de Asymmetrie te Diemen. De parameters zijn een maat voor de covariatie van de component en de Asymmetrie. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de THD.

#### $\mathbf{PLT}$

In Fig. 3.15 zijn vijfhonderd 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.15 wekt echter niet de suggestie dat de sprongen in de PLT gerelateerd zijn aan het schakelen van de aan ons bekende invloeden.



Figuur 3.14: Parameters van het SARMAX $(1,0) \times (1,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de THD te Diemen. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.



Figuur 3.15: Vijfhonderd 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Diemen, vanaf 21 juli, 8:10 tot 24 juli, 19:30. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

## 3.4 220kV-Condensatorbank te Weiwerd (WEW)

In Fig. 3.16 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden op de spanningskwaliteit te Weiwerd. Het speciale project WEW\_C01 staat in dit figuur aangegeven in het oranje. Alleen componenten waarvan een redelijke hoeveelheid aan- en uit metingen zijn kunnen goed geanalyseerd worden. Om deze reden kunnen de volgende componenten niet worden meegenomen in de analyse: WEW\_TR201, WEW\_DES\_Z, en WEW\_DZO\_W.

Daarnaast zijn er verschillende componenten die precies hetzelfde schakelen, en in dat geval kan slechts een van de componenten worden meegenomen. De gepresenteerde resultaten van de analyse moeten dan geïnterpreteerd worden als het effect door de gecombineerde invloed van de componenten die hetzelfde schakelen. De componenten MEE\_WEW\_O, RBB\_WEW\_G, RBB\_WEW\_G, MEE\_WEW\_B, en RBB\_WEW\_P worden niet meegenomen, aangezien ze respectievelijk hetzelfde schakelen als: WEW\_MEE\_O, WEW\_RBB\_G, WEW\_RBB\_G, WEW\_KV, en WEW\_RBB\_P.



Figuur 3.16: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit te Weiwerd gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. De condensatorbanken (WEW\_C01) zijn aangegeven in het oranje.

Versie 1.1, 7 december 2015

#### $\mathbf{LSV}$

In Fig. 3.17 staat een overzicht van de LSV te Weiwerd. Het valt op dat eind november de gemiddelde waarde van de LSV plotseling sterk daalt. Deze daling kan duidelijk niet verklaard worden vanuit de aan ons bekende externe factoren. Om deze reden is alleen de data tot en met halverwege november meegenomen in de analyse.

Voor de naar dag-waarde geaggregeerde data van de LSV te Weiwerd blijkt het ARMAX(2,3) model de beste fit te geven. In Fig. 3.18 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de LSV.



Figuur 3.17: Dag-waarden van de LSV te Weiwerd. De sterke daling aan het eind van de tijdreeks is aangegeven met rode ellipsen.

#### Asymmetrie

Voor de naar dag-waarde geaggregeerde data van de asymmetrie te Weiwerd blijkt het ARMAX(1,0) model de beste fit te geven. In Fig. 3.19 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de asymmetrie.

### THD

Voor de naar dag-waarde geaggregeerde data van de THD te Weiwerd blijkt het SARMAX $(1,0) \times (3,0)_7$  model de beste fit te geven. In Fig. 3.20 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de asymmetrie.

40





Figuur 3.18: Parameters van het SARMAX(2,3) model, gefit aan de naar dag-waarden geaggregeerde data van de LSV te Weiwerd. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### PLT

In Fig. 3.21 zijn duizend 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.21 suggereert echter een duidelijk verband tussen het aanschakelen van de condensatorbank en het tijdelijk toenemen van de PLT.



Figuur 3.19: Parameters van het ARMAX(1,0) model, gefit aan de naar dag-waarden geaggregeerde data van de Asymmetrie te Weiwerd. De parameters zijn een maat voor de covariatie van de component en de Asymmetrie. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.



Figuur 3.20: Parameters van het  $SARMAX(1,0) \times (3,0)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de THD te Weiwerd. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

VORtech



Figuur 3.21: Duizend 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Weiwerd, vanaf 16 november 2012, 09:30 tot 23 november 2013, 08:10. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

### 3.5 150kV-Condensatorbank te Maarheeze (MZ)

In Fig. 3.22 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden op de spanningskwaliteit te Maarheeze. Het speciale project MZ\_COBA staat in dit figuur aangegeven in het oranje. Alleen componenten waarvan een redelijke hoeveelheid aan- en uit metingen zijn kunnen goed geanalyseerd worden. De volgende componenten zijn om die reden niet meegenomen in de analyse: BDL\_MZ\_W, BDL\_MZ\_Z, EHVO\_EZMZ\_W, EHVO\_EZMZ\_Z, EHVZ\_EOMZ\_W, EHVZ\_EOMZ\_Z, MZ\_BDL\_W, MZ\_BDL\_Z

Daarnaast kunnen ook componenten die vrijwel geheel gecorreleerd zijn niet allebei worden meegenomen, tussen deze componenten kan namelijk geen onderscheid gemaakt worden. Het blijkt dat de paren MZ\_EHZO\_Z, EHVZ\_EOMZ\_Z en MZ\_EHZO\_W, EHVO\_EZMZ\_W zeer sterk gecorreleerd zijn. Daarnaast is het schakelgedrag van MZ\_TR4 vrijwel antigecorreleerd met het schakelgedrag van MZ\_TR1. De som van deze twee reeksen is dus niet te onderscheiden van een constante invloed. Om deze reden zijn MZ\_EHZO\_Z, MZ\_EHZO\_W, en MZ\_TR4 niet meegenomen in de analyse.



Figuur 3.22: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit te Maarheeze gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. De condensatorbank ( $MZ_COBA$ ) is aangegeven in het oranje.

Versie 1.1, 7 december 2015

#### LSV

Voor de naar dag-waarde geaggregeerde data van de LSV te Maarheeze blijkt het SARMAX $(2,0) \times (0,1)_7$  model de beste fit te geven. In Fig. 3.23 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van de condensatorbank lijkt samen te gaan met een toename van de LSV met 1 - 1.5 kV.



Figuur 3.23: Parameters van het  $SARMAX(2,0) \times (0,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de LSV te Maarheeze. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### Asymmetrie

Voor de naar dag-waarde geaggregeerde data van de asymmetrie te Maarheeze blijkt het SARMAX $(2, 1) \times (1, 1)_7$  model de beste fit te geven. In Fig. 3.24 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van de condensatorbank lijkt samen te gaan met een afname van de asymmetrie met ongeveer 0.04.

Versie 1.1, 7 december 2015



Figuur 3.24: Parameters van het  $SARMAX(2,1) \times (1,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de Asymmetrie te Maarheeze. De parameters zijn een maat voor de covariatie van de component en de Asymmetrie. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### THD

Voor de naar dag-waarde geaggregeerde data van de THD te Maarheeze blijkt het SARMAX $(1,2) \times (1,1)_7$  model de beste fit te geven. In Fig. 3.25 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van de condensatorbank lijkt samen te gaan met een toename van de THD met ongeveer 0.1-0.3.

#### $\mathbf{PLT}$

In Fig. 3.26 zijn achthonderd 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.26 suggereert echter een duidelijk verband tussen het aanschakelen van de condensatorbank en het tijdelijk toenemen van de PLT.

#### VORtech



Figuur 3.25: Parameters van het  $SARMAX(1,2) \times (1,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de THD te Maarheeze. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.



Figuur 3.26: Achthonderd 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Maarheeze, vanaf 31 december 2012, 23:00 tot 6 januari 2013, 12:10. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

# 3.6 110kV-Condensatorbank te Harculo (HCL)

In Fig. 3.27 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden op de spanningskwaliteit te Harculo. De speciale projecten HCL\_CBANK1 en HCL\_CBANK2 staan in dit figuur aangegeven in het oranje. Alleen componenten waarvan een redelijke hoeveelheid aan- en uit metingen zijn kunnen goed geanalyseerd worden. Een van de speciale projecten (HCL\_CBANK2) schakelt nooit, en kan dus niet geanalyseerd worden. Daarnaast zijn er verschillende componenten die precies hetzelfde schakelen, en in dat geval kan slechts een van de componenten worden meegenomen. De gepresenteerde resultaten van de analyse moeten dan geïnterpreteerd worden als het effect door de gecombineerde invloed van de componenten die hetzelfde schakelen.

In deze analyse worden de componenten RT\_HCL\_W, ZLH\_HCL\_W en ZLW\_HCL\_W niet meegenomen, omdat ze op dezelfde wijze schakelen als DVTP\_HCL\_W. Daarnaast worden ook de componenten OL\_HCL\_R, ZLH\_HCL\_Z en ZLW\_HCL\_Z niet meegenomen, aangezien deze hetzelfde schakelen als ZLH\_HCL\_Z.



Figuur 3.27: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit te Harculo gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. De condensatorbanken (HCL\_CBANK1 en HCL\_CBANK2) zijn aangegeven in het oranje.

Versie 1.1, 7 december 2015

#### $\mathbf{LSV}$

In Fig. 3.28 staat een overzicht van de LSV te Harculo. Het valt op dat eind november de gemiddelde waarde van de LSV plotseling sterk daalt. Deze daling kan duidelijk niet verklaard worden vanuit de aan ons bekende externe factoren. Om deze reden is alleen de data tot en met halverwege november meegenomen in de analyse.

Voor de naar dag-waarde geaggregeerde data van de LSV te Harculo blijkt het ARMAX(3, 2) model de beste fit te geven. In Fig. 3.29 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de LSV.



Figuur 3.28: Dag-waarden van de LSV te Harculo. De sterke daling aan het eind van de tijdreeks is aangegeven met rode ellipsen.

#### Asymmetrie

In de tijdreeks van de asymmetrie zaten een klein aantal zeer hoge waarden, die niet leken te corresponderen met foutieve metingen van de LSV. Deze waarden leken ook niet te corresponderen met het schakelen van de aan ons bekende componenten. De aanwezigheid van deze outliers maakt het modelleren lastig, dus alle waarden asym > 0.3 zijn uit de tijdreeks verwijderd.

Voor de naar dag-waarde geaggregeerde data van de asymmetrie te Harculo blijkt het  $SARMAX(1,2) \times (0,1)_7$  model de beste fit te geven. De residuen van deze fit bevatten een klein aantal waarden die hoger zijn dan verwacht zou worden bij normaal verdeelde residuen. In Fig. 3.30 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de asymmetrie.

Versie 1.1, 7 december 2015

### THD

Voor de naar dag-waarde geaggregeerde data van de THD te Harculo blijkt het SARMAX $(3,0) \times (2,0)_7$  model de beste fit te geven. De residuen van deze fit bevatten een klein aantal waarden die hoger zijn dan verwacht zou worden bij normaal verdeelde residuen. In Fig. 3.31 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aanstaan van het speciale project lijkt samen te gaan met een afname van de THD van ongeveer 0.5.

### PLT

In Fig. 3.32 zijn vijfhonderd 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.32 wekt echter niet de suggestie dat de sprongen in de PLT gerelateerd zijn aan het schakelen van de aan ons bekende invloeden.

VORtech



Figuur 3.29: Parameters van het ARMAX(3,2) model, gefit aan de naar dag-waarden geaggregeerde data van de LSV van te Harculo. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

VORtech



Figuur 3.30: Parameters van het  $SARMAX(1,2) \times (0,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de Asymmetrie te Harculo. De parameters zijn een maat voor de covariatie van de component en de Asymmetrie. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.



Figuur 3.31: Parameters van het  $SARMAX(3,0) \times (2,0)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de THD te Harculo. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

VORtech



Figuur 3.32: Duizend 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Harculo, vanaf 21 januari, 18:50 tot 28 januari, 17:30. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

### 3.7 380kV-kabel te Wateringen (WTR)

In Fig. 3.33 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden op de spanningskwaliteit te Wateringen. Het speciale project staat in dit figuur aangegeven in het oranje. Alleen componenten waarvan een redelijke hoeveelheid aan- en uit metingen zijn kunnen goed geanalyseerd worden. Voor dit project konden alle componenten worden meegenomen in de analyse.



Figuur 3.33: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit te Wateringen gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. De componenten van de ondergrondse kabel tussen Bleiswijk en Wateringen (BWK\_WTRW\_O en BWK\_WTRZ\_O) zijn aangegeven in het oranje.

#### $\mathbf{LSV}$

Voor de naar dag-waarde geaggregeerde data van de LSV te Wateringen blijkt het SARMAX $(1,0) \times (0,2)_7$  model de beste fit te geven. In Fig. 3.34 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aangesloten zijn van het speciale project lijkt samen te gaan met een toename van de LSV met 1.0 kV.

VORtech

WTR\_TR413 WTR\_TR412





Figuur 3.34: Parameters van het  $SARMAX(1,0) \times (0,2)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de LSV te Wateringen. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### Asymmetrie

Voor de naar dag-waarde geaggregeerde data van de asymmetrie te Wateringen blijkt het SARMAX $(1,0) \times (1,0)_7$  model de beste fit te geven. In Fig. 3.35 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de asymmetrie.

#### THD

Voor de naar dag-waarde geaggregeerde data van de THD te Wateringen blijkt het SARMAX $(1,0) \times (1,1)_7$  model de beste fit te geven. In Fig. 3.36 wordt een overzicht gegeven van geschatte

VORtech



Figuur 3.35: Parameters van het  $SARMAX(1,0) \times (1,0)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de LSV te Wateringen. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

modelparameters van de verschillende bekende externe invloeden. Het aangesloten zijn van BWK\_WTRW\_O lijkt samen te gaan met een toename van de THD1 van 0.1. Er is echter geen significant verband is gevonden voor THD2 en THD3.

#### PLT

In Fig. 3.37 zijn dertienhonderd 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

#### VORtech

Memo MV/M15.045



Figuur 3.36: Parameters van het  $SARMAX(1,0) \times (1,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de THD te Wateringen. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.37 laat zien dat het inschakelen van het speciale project gepaard lijkt te gaan met een sprong in de PLT. Dit is nog iets duidelijker te zien als we wat inzoomen (zie Fig. 3.44).

VORtech

#### Versie 1.1, 7 december 2015



Figuur 3.37: Dertienhonderd 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Wateringen, vanaf 24 mei 2013, 14:40 tot 2 juni 2013, 15:20. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.



Figuur 3.38: Driehonderd 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Wateringen, vanaf 31 mei 2013, 13:20 tot 2 juni 2013, 15:20. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

# 3.8 380kV-kabel te Bleiswijk (BWK)

In Fig. 3.39 staat een overzicht van het schakelgedrag van de mogelijke externe invloeden op de spanningskwaliteit te Bleiswijk. Het speciale project staat in dit figuur aangegeven in het oranje. Alleen componenten waarvan een redelijke hoeveelheid aan- en uit metingen zijn kunnen goed geanalyseerd worden. De volgende componenten zijn om die reden niet meegenomen in de analyse: BWK\_TR412, BWK\_TR414, KIJ\_BWKW\_O, BWK\_KIJW\_ALL, BWK\_KIJZ\_ALL, KIJ\_BWKZ\_O.



Figuur 3.39: Overzicht van het schakelgedrag van alle mogelijke externe invloeden op de spanningskwaliteit te Bleiswijk gedurende het jaar 2013. Wanneer de lijn behorende bij een component "hoog" is dan is deze aan het net gekoppeld; wanneer de lijn laag is, dan is deze ontkoppeld van het net. De exacte hoogte van de lijn heeft niets te maken met de bijdrage van de externe invloed op de spanningskwaliteit. De componenten van de ondergrondse kabel tussen Bleiswijk en Wateringen (BWK\_WTRW\_O en BWK\_WTRZ\_O) zijn aangegeven in het oranje.

#### $\mathbf{LSV}$

Voor de naar dag-waarde geaggregeerde data van de LSV te Bleiswijk blijkt het ARMAX(1,0) model de beste fit te geven. In Fig. 3.40 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aangesloten zijn van het speciale project lijkt samen te gaan met een toename van de LSV met 1.0 - 1.2 kV. VORtech

Memo MV/M15.045



Figuur 3.40: Parameters van het ARMAX(1,0) model, gefit aan de naar dag-waarden geaggregeerde data van de LSV het station te Bleiswijk. De parameters zijn een maat voor de covariatie van de component en de LSV. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### Asymmetrie

Voor de naar dag-waarde geaggregeerde data van de asymmetrie te Bleiswijk blijkt het  $SARMAX(1,2) \times (1,1)_7$  model de beste fit te geven. In Fig. 3.41 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Op basis van de huidige analyse kan geen duidelijke uitspraak worden gedaan over de relatie tussen het aangesloten zijn van het speciale project en de waarde van de asymmetrie.

#### THD

Voor de naar dag-waarde geaggregeerde data van de THD te Bleiswijk blijkt het SARMAX $(1,0) \times (1,1)_7$  model de beste fit te geven. In Fig. 3.42 wordt een overzicht gegeven van geschatte modelparameters van de verschillende bekende externe invloeden. Het aangesloten zijn van het speciale project lijkt samen te gaan met een toename van de THD van 0.5 - 1.0

VORtech

Memo MV/M15.045



Figuur 3.41: Parameters van het  $SARMAX(1,2) \times (1,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de Asymmetrie te Bleiswijk. De parameters zijn een maat voor de covariatie van de component en de Asymmetrie. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.

#### PLT

In Fig. 3.43 zijn driehonderd 10-minuten waarden van de PLT tijdreeksen uitgezet tegen de tijd. Het valt direct op dat de PLT redelijk constant is, maar zo nu en dan ineens verspringt naar een grotere waarde. Deze sprongen worden veroorzaakt door de manier waarop de PLT variabele geconstrueerd is. De PLT is een lopend gemiddelde van twee uur van een andere variabele, waardoor een "plateau-achtige" structuur ontstaat.

Een tijdreeks als de PLT is niet te modelleren binnen de standaardklasse van ARMA(X) modellen. Fig. 3.43 laat zien dat het schakelen van enkele componenten gepaard lijkt te gaan met een sprong in de PLT. Als we echter kijken naar een later tijdstip in het jaar (zie Fig. 3.44), dan is dit verband niet te bestaan.



Figuur 3.42: Parameters van het SARMAX $(1,0) \times (1,1)_7$  model, gefit aan de naar dag-waarden geaggregeerde data van de THD te Bleiswijk. De parameters zijn een maat voor de covariatie van de component en de THD. De zwarte lijnen binnen de balken zijn een maat voor de geschatte onzekerheid.



Figuur 3.43: Driehonderd 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Bleiswijk, vanaf 25 maart 2013, 12:50 tot 27 maart 2013, 14:50. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

VORtech



Figuur 3.44: Tweeduizend 10-minuten waarden van de tijdreeksen PLT1, PLT2 en PLT3 te Bleiswijk, vanaf 2 november, 12:10 tot 16 november, 9:30. Met de rode stippellijnen is aangegeven waar er "sprongen" in de PLT tijdreeksen voorkomen.

# Hoofdstuk 4

# Conclusies en Aanbevelingen

We hebben voor TenneT een analyse uitgevoerd van tijdreeksen van tien verschillende spanningskwaliteitsaspecten, gemeten op acht verschillende locaties. De vraag is of bepaalde externe factoren, waaronder "speciale projecten" onder beheer van TenneT invloed hebben op de spanningskwaliteit. Van deze externe factoren is bekend op welke tijdstippen ze aan het netwerk gekoppeld waren en op welke tijdstippen niet. Een standaardtechniek om deze vraag te beantwoorden is door de gegeven data te fitten met een ARMAX model en daarmee te bekijken of het schakelen van de externe factoren co-varieert met de tijdreeksen van de spanningskwaliteit.

De originele data bestaat uit 10-minuten waarden, en bevatte over het algemeen een zeer sterke autocorrelatie en een duidelijke periodieke structuur met een periode van een dag. Daarnaast leverde ARMAX fits van de originele data residuen op die duidelijk niet normaal zijn verdeeld. Deze twee problemen zijn opgelost door de data te aggregeren naar dagwaarden. In dit rapport zijn de uitkomsten van de ARMAX fits aan de geaggregeerde data gerapporteerd. De tijdreeksen van de PLT kunnen niet gemodelleerd worden met een ARMAX model, doordat de PLT een sterk bewerkt signaal is (e.g., een lopend gemiddelde).

Waar naar gekeken is, is de mate waarin de spanningskwaliteit en het schakelen van het speciale project covariëren, en of het het verband statistisch significant is met een significantieniveau van 0.05. Wanneer dit zo is, dan spreken we van een "impact". Een overzicht van de impact van de speciale projecten op de spanningskwaliteit is te vinden in Tabel 4.1.

Met de huidige analyse wordt op een consistente en statistisch verantwoorde manier gekeken naar alle tijdreeksen. In een eventuele vervolgstudie kan worden gekeken naar de volgende zaken.

- In plaats van naar de PLT kan worden gekeken naar de onderliggende variabele. Mogelijk valt deze beter te modelleren binnen de klasse van ARMA modellen.
- Zoals besproken zijn de residuen van de niet-geaggregeerde data niet normaal verdeeld, mede om deze reden is besloten om de data te aggregeren. Er zou kunnen

#### VORtech

Memo MV/M15.045

#### Versie 1.1, 7 december 2015

Impact: de aan- en afkoppeling van het speciale project valt statistisch samen								
met het spanningskwaliteitsaspect:								
Speciaal Project	Locatie	LSV1	LSV2	LSV3	Asym	THD1	THD2	THD3
HVDC NorNed	EEMS	ja	ja	ja	ja	ja	nee	nee
HVDC BritNed	MVL	nee	nee	nee	nee	ja	ja	ja
380kV-Condensatorbank	DMN	nee	nee	nee	ja	nee	nee	nee
220kV-Condensatorbank	WEW	nee						
150kV-Condensatorbank	MZ	ja						
110kV-Condensatorbank	HCL	nee	nee	nee	nee	ja	ja	ja
380kV-kabel	BWK	ja	ja	ja	nee	ja	ja	ja
380kV-kabel	WTR	ja	ja	ja	nee	ja	nee	nee

Tabel 4.1: In de statistische analyse is voor iedere locatie het covariëren van deze spanningskwaliteitsaspecten en het schakelen van het speciale project en de omgevingsfactoren onderzocht. De uiteindelijke analyse is gedaan op naar dag-waarden geaggregeerde tijdreeksen van de spanningskwaliteit. In dit overzicht wordt antwoord gegeven op de vraag of, gegeven de naar dag-waarde geaggregeerde data, de nulhypothese "geen correlatie met het speciale project" verworpen kon worden. Bij deze nulhypothese is een significantieniveau van 0.05 gehanteerd. Merk op dat het antwoord op de vraag geen uitspraak doet over de feitelijke grootte van het effect. Uitsluitend de kwalitatieve impact van het speciale project op de spanningskwaliteitsaspecten per locatie is hier gepresenteerd.

worden gekeken naar de mogelijkheden om de niet-normale verdeling op te nemen in het model, waardoor er over een minder lange periode hoeft te worden geaggregeerd. Onze analyses wijzen uit dat in het geval van de LSV, de asymmetrie en de THD zou kunnen worden gewerkt met een t-verdeling.

• De huidige dataset bevat alleen schakeldata, d.w.z., het is alleen bekend wanneer een invloed aan het netwerk gekoppeld was en wanneer niet. Des te informatiever externe voorspellers zijn, des te nauwkeuriger kan er informatie uit de tijdreeks verkregen worden op grond van deze voorspellers. Geschiktere voorspellers zouden mogelijkerwijs aangeleverd kunnen worden door een elektrotechnicus, aan de hand van gedetailleerde kennis van het netwerk en elektrotechnische parameters van de componenten.

# Referenties

- [1] The R foundation. R. https://www.r-project.org/.
- [2] Robert H Shumway and David S Stoffer. *Time series analysis and its applications:* with R examples. Springer Science & Business Media, 2010.
- [3] Frans van Erp. Nulmeting ten behoeven van de impactanalyse. Technical Report PUAM14777, TenneT, December 2014.
- [4] Michiel van Lumig. Energiekamer NMa Eindrapport advies over spanningskwaliteit in elektriciteitsnetten. Technical Report LBE02710239, Laborelec, November 2012.