Voorspellen van verkiezingsuitslagen van de Tweede Kamer op basis van getelde stemmen

D. Doppenberg
Universiteit Twente.
d.doppenberg@student.utwente.nl,

26 juni 2016

Samenvatting

Dit onderzoek introduceert een nieuwe methode om voorspellingen te geven van de einduitslag van een Tweede Kamer verkiezing. Deze methode voorspelt de uitslag aan de hand van de oude verkiezingsuitslagen en de tot dan toe getelde stemmen van de huidige verkiezing. Deze voorspellingen zijn gebaseerd op het vinden van een transformatiematrix. Het doel van deze nieuwe methode is dat de schommelingen in opeenvolgende voorspellingen aanzienlijk kleiner zijn dan dat nu waargenomen wordt met de huidige voorspelmethode. De convergentie van de voorspellingen wordt onderzocht aan de hand van de fout van de voorspellingen ten opzichte van de daadwerkelijk einduitslag. De kwaliteit van de nieuwe manier van voorspellen is vergeleken met de kwaliteit van twee andere methodes, die ook worden geïntroduceerd in dit onderzoek. Uit deze vergelijkingen blijkt dat de voorspellingen van de nieuwe methode convergeren naar de einduitslag. De voorspellingen zijn ten slotte vergeleken met de bestaande voorspellingen van het ANP van 2012. Hieruit blijkt dat de voorspellingen van de nieuwe methode sneller convergeren naar de einduitslag dan de voorspellingen van de huidige methode.

kernwoorden: transformatiematrix, kleinste kwadraten methode, voorspellen verkiezingen

1. Inleiding

Verkiezingen voor de Tweede Kamer in Nederland vinden normaal gesproken eens in de vier jaar plaats wanneer de regeringstermijn is verlopen. Soms gebeurt dit ook wanneer de zittende regering is gevallen. De verkiezing is traditioneel op een woensdag en wanneer de stembussen om negen uur 's avonds sluiten, worden in elke gemeente de stemmen geteld. De uitslag van een gemeente bestaat uit een lijst met het aantal uitgebrachte stemmen per partij. Wanneer alle stemmen van alle gemeenten geteld zijn, wordt de kiesdeler berekend om de uitslag te bepalen. De kiesdeler is het getal dat aangeeft hoeveel stemmen een partij moet ontvangen om één zetel te krijgen. Een partij heeft dus recht op het aantal zetels dat gelijk is aan het aantal ontvangen stemmen gedeeld door de kiesdeler. Dit getal is (bijna) nooit een geheel getal en wordt daarom naar beneden afgerond. Het aantal zetels dat een partij op deze manier krijgt, wordt het aantal volle zetels genoemd. Op deze manier worden nog niet alle zetels verdeeld. De zetels die nog verdeeld moeten worden heten de restzetels. Voor een uitleg over de restzetelverdeling, zie het artikel van J.J. Paulus en J. Praagman over de Eerste Kamerverkiezingen [1].

Wanneer een gemeente haar stemmen heeft geteld, stuurt deze de lijst met uitslagen door naar de Publieke Omroep en het Algemeen Nederlands Persbureau (ANP). Voor dit onderzoek is vooral de data-uitwisseling met het ANP van belang. Het ANP probeert namelijk aan de hand van de *exit polls* en de toegestuurde data van gemeenten een voorspelling te geven van de uitkomst van de verkiezing. De *exit polls* zijn de uitslagen van statistisch onderzoek naar wat mensen bij de huidige verkiezing hebben gestemd. Het geven van een voorspelling van de einduitslag wordt gedaan door de 'Verkiezingsdienst' van het ANP. Het ANP kiest er zelf voor wanneer het een update van de voorspelling geeft. Het ANP zou er bijvoorbeeld voor kunnen kiezen om bij elke keer dat er van 80 gemeenten de uitslag is binnengekomen een nieuwe voorspelling te geven.

Het nadeel van de voorspellingen van het ANP is dat de voorspelde zetelverdeling van de nieuwe voorspelling vaak afwijkt van de voorspelde zetelverdeling van de vorige voorspelling. Deze afwijkingen in zetelverdeling worden schommelingen in de voorspellingen genoemd. Door deze schommelingen duurt het lang voordat de voorspellingen van het ANP een juiste weerspiegeling geeft van de echte einduitslag van de verkiezing.

Aangezien het algoritme van het ANP niet openbaar en daarom niet beschikbaar is, kan er geen verbetering voor het algoritme bedacht worden. Daarom ontwikkelt dit onderzoek een nieuw concept waarmee de einduitslag voorspeld kan worden aan de hand van de tot dan toe binnengekomen data. Het doel van dit nieuwe concept is dat de schommelingen in de voorspelde zetelverdeling van verschillende voorspelling flink worden gedempt. Bij de verkiezingen van 2012 waren alle voorspellingen van het ANP tot aan dat er 95 procent van de stemmen waren geteld niet correct [5]. In dit onderzoek zal de nieuwe methode worden getest in MATLAB. Deze methode probeert de schommelingen zo te dempen dat het mogelijk is om al eerder dan het moment dat er 95 procent van de stemmen is geteld een voorspelling te geven die niet afwijkt van de einduitslag. Deze methode zal in dit artikel de 'Methode A' genoemd worden. Het belang van eerder een betrouwbare voorspelling geven is dat partijen die bijvoorbeeld eerst een goede uitslag krijgen, vervolgens niet worden teleurgesteld als de voorspelling niet klopt. Daarnaast is het voor de media interessant om zo snel mogelijk een juiste uitspraak te kunnen doen over de verkiezingsuitslag.

De methode van dit onderzoek berust op het in kaart brengen van gemeentelijke verschuivingen in stemgedrag. Deze methode maakt gebruik van de aanname dat landelijke verschuivingen in stemgedrag terug te vinden zijn in de gemeentelijke verschuivingen in stemgedrag. Dit is gebaseerd op de veronderstelling dat landelijke trends in iedere gemeente terug te vinden zijn. De gemiddelde verschuiving op basis van alle gemeentelijke verschuivingen in stemgedrag wordt dan gezien als de landelijke verschuiving in stemgedrag.

In Hoofdstuk 2 zal de voorspelmethode die in dit onderzoek centraal staat besproken worden. Hier zullen alle losse onderdelen van deze methode uitvoerig worden toegelicht. Vervolgens zullen in Hoofdstuk 3 twee andere voorspelmethodes worden besproken. Deze methodes worden onderzocht zodat er voor de Methode A vergelijkingsmateriaal is. In Hoofdstuk 4 zullen de resultaten van deze methode worden onderzocht. De resultaten van de methode van Hoofdstuk 2 wordt in dit hoofdstuk vergeleken met methodes van Hoofdstuk 3. Hoofdstuk 4 presenteert de resultaten van de verschillende methodes. Ten slotte bevat Hoofdstuk 5 de conclusie en discussie.

2. Methode A

In dit hoofdstuk zal de nieuwe methode van voorspellen worden toegelicht. Paragraaf 2.1 zal zich richten op het construeren van technieken om een goede voorspelling te maken. Paragraaf 2.2 zal wat dieper in gaan op de wijze van voorspellen.

2.1. Transformatiematrix

In deze paragraaf zal aandacht worden besteed aan de benodigde technieken om een voorspelling te kunnen doen met Methode A. Eerst zal het begrip en het idee van een transformatiematrix worden toegelicht, gevolgd door de uitleg over het construeren van een transformatiematrix. Hoewel in Paragraaf 2.2 pas wordt uitgelegd hoe er voorspeld kan worden, zal deze paragraaf daar in het kort al een uitspraak over doen. Vervolgens zal er nog gekeken worden naar een efficiëntere en betere manier om een transformatiematrix te berekenen. Ten slotte zal de interpretatie van een dergelijke matrix worden uitgewerkt.

2.1.1 Constructie

Wanneer er getelde stemmen binnenkomen, kan er aan de hand van deze binnengekomen gemeentelijke uitslagen in combinatie met de data van diezelfde gemeenten van de vorige verkiezing een voorspelling gedaan worden. De manier van voorspellen in dit onderzoek berust op het vinden van een matrix T. Deze matrix is zo gekozen dat wanneer T vermenigvuldigd wordt met de vorige gemeentelijke verkiezingsuitslagen, dit de binnengekomen gemeentelijke verkiezingsuitslagen op een zo goed mogelijk manier oplevert. Een interpretatie van deze matrix T is dan als volgt: de matrix T is een matrix die de data van de vorige verkiezing afbeeldt op de huidige verkiezing. Deze matrix T wordt een transformatiematrix genoemd.

Elke keer wanneer er een nieuwe gemeentelijke uitslag binnenkomt, kan op basis van de tot dan toe binnengekomen uitslagen een nieuwe voorspelling van de einduitslag gegeven worden. Een belangrijke eis bij het afbeelden van de vorige verkiezingsuitslag van een gemeente op de nieuwe, is dat T daadwerkelijk dezelfde gemeente op dezelfde gemeente afbeeldt. Dit kan als volgt beschreven worden:
Definieer m als het aantal gemeenten die meedoen aan de verkiezingen en n als het aantal partijen die meedoen. Voor de waarde van n kan bijvoorbeeld gekozen worden voor 15. In dit geval zou dit onderverdeeld kunnen worden in elf 'grote' partijen, drie 'kansrijke' partijen en één partij voor de overige. In het vervolg van dit onderzoek wordt m gebruikt voor het aantal gemeenten en n gebruikt voor het aantal partijen.
Noteer $\mathbf{v^i} \in \mathbb{R}^n$ als de uitslag van gemeente i van de huidige verkiezing en noteer $\mathbf{u^i} \in \mathbb{R}^n$ als de uitslag van dezelfde gemeente i van de vorige verkiezing. Deze gemeentelijke uitslagen zijn in percentages. Laat $\mathbf{v^1}, \mathbf{v^2}, ..., \mathbf{v^k}, ..., \mathbf{v^m}$ de volgorde zijn waarop de getelde stemmen per gemeente van de huidige verkiezing binnenkomen. Laat $\mathbf{u^1}, \mathbf{u^2}, ..., \mathbf{u^k}, ..., \mathbf{u^m}$ de volgorde van de uitslag van de gemeenten van de vorige verkiezing zijn die correspondeert met de volgorde van binnenkomst van de huidige gemeentelijke verkiezingsuitslagen.
Wanneer de eerste gemeente binnenkomt, moet T voldoen aan het afbeelden van de uitslag van de vorige verkiezing van de eerst binnengekomen gemeente op die van de huidige verkiezing van dezelfde gemeente, met andere woorden: $T\mathbf{u^1} = \mathbf{v^1}$. Wanneer de tweede gemeente binnenkomt,

moet T ook voldoen aan $T\mathbf{u^2} = \mathbf{v^2}$, enzovoort. Definieer k als het aantal tot dan toe binnengekomen gemeentelijke uitslagen van de huidige verkiezingen, waarbij $1 \leq k \leq m$. Dan resulteert dat na k gemeenten in:

$$T(\mathbf{u^1}, \mathbf{u^2}, ..., \mathbf{u^k}) = (\mathbf{v^1}, \mathbf{v^2}, ..., \mathbf{v^k}).$$

Dit kan herschreven worden tot:

$$TU^k = V^k, \tag{1}$$

waarbij $U^k = (\mathbf{u^1}, \mathbf{u^2}, ..., \mathbf{u^k})$ en $V^k = (\mathbf{v^1}, \mathbf{v^2}, ..., \mathbf{v^k})$.

Zolang k kleiner is dan de orde van T, in dit geval n, dan is de transformatiematrix onderbepaald. Onderbepaald heeft als gevolg dat er oneindig veel verschillende T's gevonden kunnen worden die de vorige uitslagen van de binnengekomen gemeenten exact afbeelden op de nieuwe uitslagen van diezelfde gemeenten.

Voor k groter dan de orde van T, wordt T overbepaald. Wanneer een matrix overbepaald is, bestaat er geen T meer die een afbeelding exact kan realiseren. In het geval van een overbepaalde matrix wordt er daarom gezocht naar een T die U^k zo goed mogelijk afbeeldt op V^k. De standaardmethode om een dergelijk probleem op te lossen, is met behulp van de kleinste kwadraten methode. De kleinste kwadraten methode geeft een T die U^k op de best mogelijke manier afbeeldt op V^k. Dit komt omdat de oplossing van de kleinste kwadraten methode het kwadraat van de norm van $TU^k - V^k$ minimaliseert:

$$\min_T ||TU^k - V^k||^2 = \min_T \sum_{i=1}^{n} \sum_{j=1}^{k} (TU^k - V^k)_{ij}^2. \tag{2}$$

In het algemeen zijn kleinste kwadraten problemen niet gemakkelijk om op te lossen. Wanneer n en m bijvoorbeeld heel groot zijn, kost dit veel rekenkracht als gevolg van de vele variabelen die voor T bepaald moeten worden. In het geval van probleem (2) is er wel een eenvoudige manier om de beste T te vinden. Deze oplossing voor T kan op de volgende manier worden geconstrueerd: vermenigvuldig (1) rechts met de getransponeerde van U^k en definieer T^k als de transformatiematrix die wordt geconstrueerd wanneer er k gemeenten zijn binnengekomen, dan geldt:

$$T^k U^k U^{k^T} = V^k U^{k^T}. \tag{3}$$

Definieer nu $A^k = U^k U^{k^T}$ en $B^k = V^k U^{k^T}$, dan kan (3) als volgt worden herschreven:

$$T^k A^k = B^k. \tag{4}$$

Aangezien A^k een positief definiete matrix is, heeft A^k een inverse, zie Appendix A. Door (4) rechts te vermenigvuldigen met de inverse van A^k, ontstaat een uitdrukking voor T^k:

$$T^k = B^k A^{k^{-1}} = V^k U^{k^T} (U^k U^{k^T})^{-1}. \tag{5}$$

Deze T minimaliseert daadwerkelijk het kleinste kwadraten probleem (2), zie Appendix B. Hoe deze T gebruikt wordt om te voorspellen, zal verder worden uitgewerkt in Paragraaf 2.2.

Over de robuustheid van Methode A kan een uitspraak gedaan worden op basis van de relatie tussen deze methode en de kleinste kwadraten methode. Bij een kleine verandering van een gemeentelijke uitslag zal de kleinste kwadraten methode een T vinden die niet veel afwijkt van de T die zonder deze verandering gevonden zou worden. Dit komt doordat het een kleine verandering is in een grote dataset, waardoor de oplossing die in gemiddelde zin voldoet aan de eisen niet veel verschilt. Dit geldt alleen als er wordt aangenomen dat er geen gemeenten zijn die zo groot zijn dat de uitslagen van andere gemeenten er niet meer toe doen voor de einduitslag. Dit zou namelijk betekenen dat de einduitslag in alle gevallen gelijk is aan de gemeentelijke uitslag van die gemeente. Hieruit kan geconcludeerd worden dat de methode robuust is.

2.1.2 Updatemethode

Elke keer wanneer er een nieuwe gemeente binnenkomt, kan er een nieuwe voorspelling gedaan worden. Dit betekent dat, wanneer de bovengenoemde methode van het construeren van een T wordt gebruikt om een voorspelling te geven, de matrices A^k en B^k elke keer opnieuw volledig berekend moeten worden. Het opnieuw berekenen van deze matrices kan op een efficiëntere manier gedaan worden aan de hand van een recursieve updatemethode:

$$\begin{aligned} A_{ij}^1 &= U_{i1}^1 U_{1j}^1 && \text{voor } k = 1, \\ A_{ij}^k &= A_{ij}^{k-1} + U_{ik}^k U_{kj}^{k^T} && \text{voor } 2 \leq k \leq m, \end{aligned}$$

zie Appendix C voor een afleiding van deze updatemethode.

Deze methode is efficiënter omdat er per element van A^k van k naar $k+1$ maar één product bepaald hoeft te worden, namelijk: $U_{ik}^k U_{kj}^{k^T}$.
Deze updatemethode geldt ook voor het bepalen van B^k en is op dezelfde manier af te leiden:

$$\begin{aligned} B_{ij}^1 &= V_{i1}^1 U_{1j}^{1^T} && \text{voor } k = 1, \\ B_{ij}^k &= B_{ij}^{k-1} + V_{ik}^k U_{kj}^{k^T} && \text{voor } 2 \leq k \leq m. \end{aligned}$$

Om te voorkomen dat de uitslagen van kleine gemeenten evenveel invloed hebben op de voorspelling als de uitslagen van grote gemeenten, worden er verschillende wegingen toegekend aan de uitslagen van verschillende gemeenten. De weging die wordt gebruikt per gemeente is gelijk aan het totaal aantal stemmen van deze gemeente gedeeld door het totaal aantal stemmen dat landelijk is uitgebracht. Een eenvoudige manier van wegen is het vaker toevoegen van een gemeentelijke uitslag aan de datasets, zodat T deze gemeentelijke uitslag van de vorige verkiezingen meerdere malen afbeeldt op de gemeentelijke uitslag van de huidige verkiezingen. Een voorbeeld van deze manier van weging is te vinden in Appendix D. Het op deze manier wegen van uitslagen komt overeen met de weging toekennen tijdens de updatemethode, zie Appendix D. Wanneer deze manier van wegen wordt geïmplementeerd in de updatemethode, levert dat nieuwe updatemethodes op. Definieer $w_i \in [0,1]$ als de weging voor gemeente i dan wordt de nieuwe updatemethode:

$$\begin{aligned} A_{ij}^1 &= w_1 \cdot U_{i1}^1 U_{1j}^1 && \text{en} \quad B_{ij}^1 = w_1 \cdot V_{i1}^1 U_{1j}^1 && \text{voor } k = 1, \\ A_{ij}^k &= A_{ij}^{k-1} + w_k \cdot U_{ik}^k U_{kj}^{k^T} && \text{en} \quad B_{ij}^k = B_{ij}^{k-1} + w_k \cdot V_{ik}^k U_{kj}^{k^T} && \text{voor } 2 \leq k \leq m. \end{aligned}$$

Voor de weegfactoren geldt dan ook dat $\sum_{i=1}^{m} w_i = 1$.

2.1.3 Interpretatie

Nu de constructie van een transformatiematrix bekend is kan er een betere interpretatie gegeven worden. De matrix T kan geïnterpreteerd worden als een matrix die de vorige verkiezingsdata afbeeldt op de huidige verkiezingsdata, zoals genoemd in Paragraaf 2.1.1. Om een betere interpretatie te krijgen van T, moet er gekeken worden naar de betekenis van alle elementen uit T en de invloed die deze elementen hebben op het transformeren van de data van de vorige verkiezing.

Stel $T \in \mathbb{R}^{n \times n}$, $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}^n$ en $t_{ij} \in T$ met $1 \leq i \leq n$ en $1 \leq j \leq n$. De vectoren $\mathbf{u}$ en $\mathbf{v}$ bevatten respectievelijk een gemeentelijke uitslag van de vorige verkiezing en de nieuwe verkiezing in percentages. Om de interpretatie van de elementen te af te leiden, wordt de volgende vergelijking gebruikt die volgt uit $T\mathbf{u} = \mathbf{v}$:

$$t_{11} \cdot u_1 + t_{12} \cdot u_2 + \ldots + t_{1n} \cdot u_n = v_1. \tag{6}$$

Deze vergelijking betekent in woorden dat voor het percentage stemmen dat partij 1 in de huidige verkiezing in deze gemeente ontvangt, geldt dat dit percentage een lineaire combinatie is van het percentage stemmen dat partij 1 in de vorige verkiezing in deze gemeente ontving, het percentage stemmen dat partij 2 in de vorige verkiezing in deze gemeente ontving tot en met het percentage stemmen dat partij n in de vorige verkiezing in deze gemeente ontving. Deze coëfficiënten t_{1i}, voor $1 \leq i \leq n$, kunnen worden geïnterpreteerd als het deel van de stemmers dat voor partij i stemde in oude verkiezing en nu voor partij 1 stemt in de huidige verkiezing.

Deze interpretatie kan in algemenere zin ook gebruikt worden. Een element van T in rij i en kolom j, waarbij $i \neq j$, kan geïnterpreteerd worden als het percentage stemmers dat van partij j naar partij i is overgestapt. Voor een element uit T in rij i in kolom i geldt een andere interpretatie. Deze elementen kunnen geïnterpreteerd worden als het percentage van de stemmers dat op partij i stemde bij de vorige verkiezing en die opnieuw op partij i stemde bij de huidige verkiezing. In andere woorden, de matrix T brengt de gemeentelijke verschuiving van het stemgedrag in kaart. In Appendix E staat een voorbeeld waarbij deze interpretatie wordt toegepast. Dit voorbeeld voedt deze interpretatie aangezien voor elke kolom van T geldt dat de elementen sommeren tot 1. Dit komt omdat kolom i het gedrag van alle stemmers van partij i weergeeft, ofwel honderd procent van de stemmers van partij i.
De verwachting is dat de diagonaalelementen van T in verhouding tot de andere elementen groot zijn, omdat over het algemeen veel stemmers hun partij trouw blijven. Uit Appendix F blijkt inderdaad dat de diagonaalelementen relatief groot zijn.

Hoewel deze interpretatie voor de hand lijkt te liggen, is het niet mogelijk om te controleren of de interpretatie daadwerkelijk klopt. Er zijn namelijk geen statistische onderzoeken die de gemeentelijke en landelijke verschuivingen in stemgedrag in kaart brengen.

2.2. Voorspellen

Wanneer voor een bepaalde k de transformatiematrix T^k op de eerder beschreven manier gevonden is, kan er een voorspelling gedaan worden van de einduitslag. De aanname waar de manier van voorspellen op gebaseerd is, is dat landelijke verschuivingen in stemgedrag terug te vinden zijn in de gemeentelijke verschuivingen in stemgedrag. Aangezien T^k voor de k als eerst binnengekomen gemeenten de best mogelijke gemeentelijke verschuiving in stemgedrag in kaart brengt, brengt T^k dus ook op de beste mogelijke manier de landelijke verschuiving in stemgedrag in kaart.

Door deze T te zien als een matrix die ook de landelijke verschuiving in kaart brengt, kan met deze T de einduitslag van de vorige verkiezing aangepast worden met de verschuiving in stemgedrag. Door deze verschuiving toe te passen, resulteert dat in een voorspelde einduitslag. Een formelere omschrijving van dit principe is als volgt:

Definieer $\mathbf{u} \in \mathbb{R}^n$ als de vector die de landelijke uitslag van de vorige verkiezing van n partijen bevat en $\mathbf{z^k} \in \mathbb{R}^n$ als de vector die de voorspelling van de landelijke uitslag van de huidige verkiezing na k gemeenten bevat, dan geeft T^k op de volgende manier na k gemeenten een voorspelling:

$$T^k \mathbf{u} = \mathbf{z^k}$$

Definieer nu $\mathbf{z} \in \mathbb{R}^n$ als de einduitslag van de huidige verkiezing. Om de kwaliteit van de voorspelling te controleren, wordt het kwadraat van de norm berekend van de verschilvector van $\mathbf{z^k}$ en $\mathbf{z}$:

$$Q = ||\mathbf{z^k} - \mathbf{z}||^2$$

Hoe groter Q, des te groter is het verschil tussen de de voorspelling en de daadwerkelijk uitslag. Belangrijk is dat de voorspelling elke keer dichterbij de uitslag komt, dus dat Q steeds kleiner wordt naarmate er meer gemeentelijke uitslagen binnenkomen. Wanneer Q erg klein is, convergeren de voorspellingen naar de einduitslag.

Definitie 2.1. *Definieer* $\mathbf{z^k} \in \mathbb{R}^\mathbf{n}$ *als de voorspelling na k gemeenten met* $1 \leq k \leq m$, *n het aantal partijen en m het aantal gemeenten. Definieer* $\mathbf{z} \in \mathbb{R}^\mathrm{n}$ *als de daadwerkelijke uitslag van de verkiezing. Dan convergeert een voorspelling naar de einduitslag wanneer er voor een gekozen* $0 < \epsilon \ll 1$ *een* $K \in [1, m]$ *bestaat, zodanig dat* $||\mathbf{z^k} - \mathbf{z}||^2 < \epsilon$ *voor alle* $K \leq k \leq m$.

De K waarvoor de voorspelling convergeert naar de einduitslag wordt het 'convergentiepunt van de voorspelling' genoemd. Hoewel een gekozen ϵ veel kleiner dan 1 moet zijn, moet ϵ niet te dicht bij nul liggen. Een ϵ te dicht bij nul zou er voor kunnen zorgen dat een voorspelling volgens Definitie 2.1 niet convergeert, terwijl de voorspelling dat redelijkerwijs wel doet. Een goede keuze voor een ϵ in dit onderzoek ligt in het interval $[1 \cdot 10^{-5}, 9 \cdot 10^{-5}]$.

3. Vergelijkmethodes

In dit hoofdstuk worden methodes geïntroduceerd die gebruikt worden om de resultaten van Methode A te kunnen beoordelen.

3.1. Methode B

De eerste andere methode berust op een simpel principe om een voorspelling te maken: gebruik de tot dan toe binnengekomen gemeentelijke uitslagen als einduitslag. Deze methode kan als volgt worden beschreven.

Definieer $\mathbf{a^k} \in \mathbb{R}^\mathrm{n}$ als het aantal stemmen van de huidige verkiezing die zijn binnengekomen na k gemeenten voor de n partijen. Deze vector $\mathbf{a^k}$ bevat dus het aantal binnengekomen stemmen per partij. Wanneer voor elke partij het aantal binnengekomen stemmen na k gemeenten gedeeld wordt door totaal aantal binnengekomen stemmen na k gemeenten, bevat de vector $\mathbf{a^k}$ de verdeling van het aantal stemmen per partij in percentages. Deze percentages worden dan gebruikt als voorspelling voor het percentage stemmen dat deze partij zal krijgen in de einduitslag.

Een voordeel van deze methode is dat wanneer alle gemeentelijke uitslagen zijn binnengekomen de voorspelling gelijk is aan de einduitslag. Een nadeel van deze methode is dat deze methode in het algemeen niet snel een weerspiegeling geeft van de einduitslag. Dit zal alleen zo zijn wanneer vooral gemeentelijke uitslagen binnenkomen die lijken op de einduitslag.

3.2. Methode C

Om de uitkomsten van de voorspellingen van Methode A beter te kunnen controleren, is er naast Methode B nog een voorspelmethode, te noemen 'Methode C'. Methode C berust ook op het vinden van een transformatiematrix Λ. Deze transformatiematrix Λ is een diagonaal matrix is.

Om te kunnen voorspellen worden de binnengekomen stemmen per partij van elke gemeente bij elkaar opgeteld. Dit levert de vector $\mathbf{y^k} \in \mathbb{R}^n$ op, met n het aantal partijen en k het aantal binnengekomen gemeenten. Het element i van $\mathbf{y^k}$ bevat dus de som van het aantal stemmen van de eerste k gemeenten voor partij i. Hetzelfde wordt gedaan voor de oude uitslagen van dezelfde k gemeenten, dit levert de vector $\mathbf{x^k} \in \mathbb{R}^\mathrm{n}$ op. De transformatiematrix Λ beeldt dan de vector $\mathbf{x^k}$ af op $\mathbf{y^k}$:

$$\Lambda \mathbf{x^k} = \mathbf{y^k}. \tag{7}$$

Dit leidt tot de volgende set vergelijkingen:

$$\begin{aligned} \lambda_{11} x_1^k &= y_1^k \\ \lambda_{22} x_2^k &= y_2^k \\ &\vdots \\ \lambda_{nn} x_n^k &= y_n^k, \end{aligned}$$

waarbij $\lambda_{ii} \in \Lambda$ voor $1 \leq i \leq n$. Voor alle λ_{ii} kan eenvoudig een oplossing gevonden worden: $\lambda_{ii} = \frac{y_i^k}{x_i^k}$.

Wanneer Λ bepaald is, kan er op dezelfde manier als T een voorspelling geven worden:

Definieer $\mathbf{u} \in \mathbb{R}^n$ als de uitslag van de vorige verkiezing en $\mathbf{z^k} \in \mathbb{R}^n$ als de voorspelling na k binnengekomen gemeenten. Dan geldt:

$$\Lambda \mathbf{u} = \mathbf{z^k}$$

Uit de constructie van Λ blijkt dat deze transformatiematrix niet zoals T de verschuivingen van stemmers van partij naar partij in kaart brengt, maar enkel de verschuiving in het aantal ontvangen stemmen per partij. Het voorspelde aantal stemmen voor partij i in de huidige verkiezing is namelijk gelijk aan een λ_{ii} maal het aantal stemmen dat partij i in de vorige verkiezing ontving. Het element λ_{ii} geeft dan de verschuiving in het aantal stemmen voor partij i aan.

Het voordeel van Methode C is, net zoals bij Methode B, dat het zeker is dat deze voorspelling convergeert naar de einduitslag wanneer alle gemeenten geteld zijn. Dit is het resultaat van de keuze van $\mathbf{y^k}$: wanneer alle uitslagen van de gemeenten binnengekomen zijn dan is $\mathbf{y^k}$ gelijk aan de einduitslag. Een mogelijk nadeel van deze methode ten opzicht van Methode A is dat deze methode minder variabelen heeft om een verschuiving in kaart te brengen, wat zou kunnen leiden tot minder nauwkeurige voorspellingen.

4. Resultaten

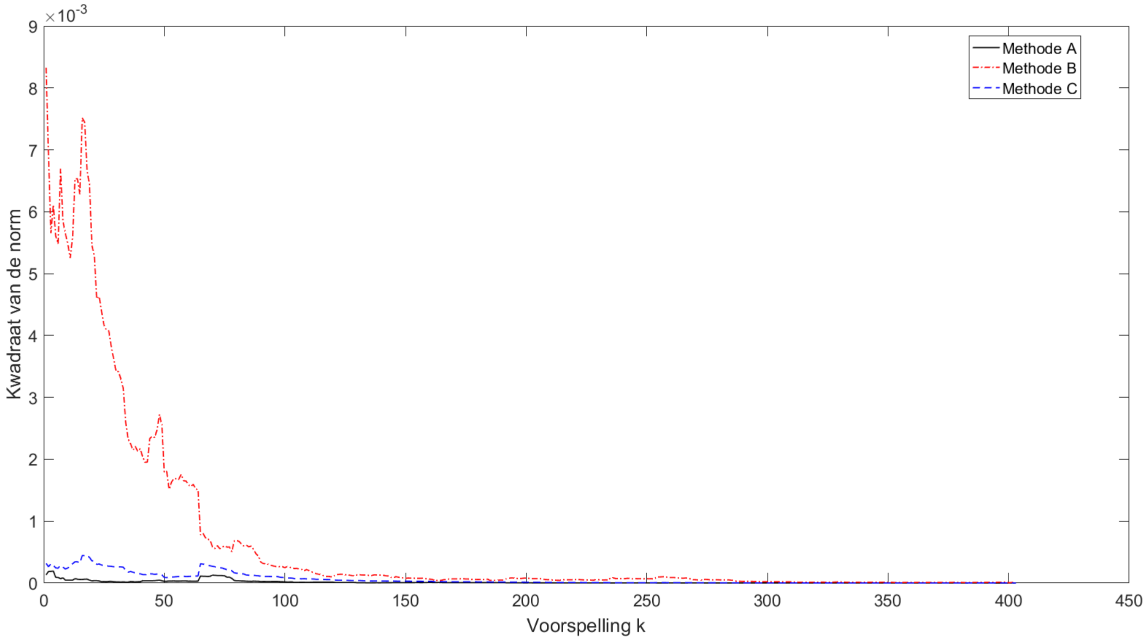
Om de werking van de Methode A aan te tonen, wordt de verkiezingsuitslag van 2012 voorspeld aan de hand van de verkiezinguitslag van 2010. Aangezien de uitslag van deze verkiezing al bekend is, kunnen de voorspellingen van de nieuwe methode gecontroleerd worden.
Bij de voorspellingen is er gebruikt gemaakt van 415 gemeenten, 13 partijen en een convergentiepunt van $\epsilon = 1 \cdot 10^{-5}$. De volgorde van partijen in de vectoren is van boven naar beneden: VVD, PvdA, PVV, CDA, SP, D66, GroenLinks, ChristenUnie, SGP, Partij voor de Dieren, Partij voor Mens en Spiritualiteit, Piraten Partij en overige partijen. Bij de voorspellingen wordt er gebruik gemaakt van de weegmethode, tenzij anders vermeld. In alle gevallen wordt er gebruik gemaakt van de echte volgorde van binnenkomst van gemeentelijke uitslagen, tenzij anders vermeld. Met de echte volgorde van binnenkomst wordt bedoeld de daadwerkelijke volgorde waarmee de gemeentelijke uitslagen bij de verkiezing van 2012 binnenkwamen. Deze volgorde wordt als representatief beschouwd, omdat de volgorde van binnenkomst voor een andere verkiezing niet veel zal afwijken. Dit komt doordat kleine gemeenten hun stemmen vaak eerder geteld hebben dan grote gemeenten.

4.1. Vergelijking

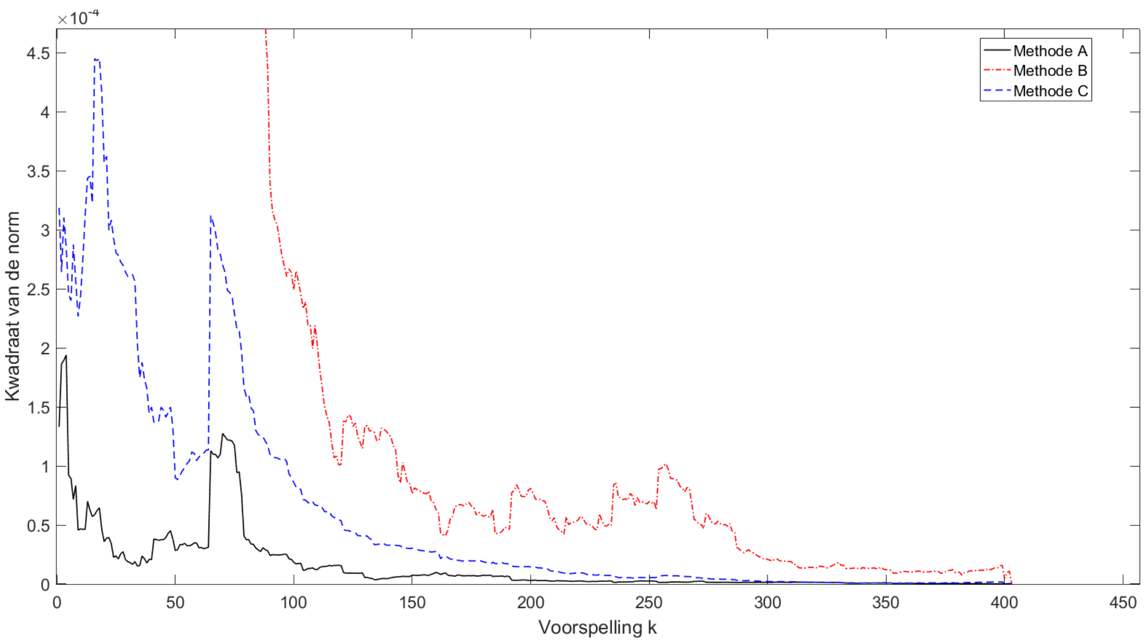
Allereerst worden de voorspellingen van Methode A vergeleken met de voorspellingen van de Methode B en C. Vervolgens worden de voorspellingen van Methode A vergeleken met de gevonden voorspellingen van het ANP.

4.1.1 Verschillende methodes

De voorspellingen van Methode A convergeren eerder naar de uitslag dan de andere twee methodes, zie Figuur 1, Figuur 2 en Tabel 1. Dit is ook geverifieerd voor meerdere volgordes, zie Appendix G. Bij de andere volgordes is te zien dat Methode A en Methode C altijd beter zijn dan Methode B. Daarnaast zijn de voorspellingen van Methode A niet op elk moment beter dan die van Methode C, maar de voorspellingen van Methode A convergeren in alle volgordes wel eerder dan die van Methode C.



Figuur 1: De convergentie van de voorspellingen van Methode A, Methode B en Methode C. Voor dit figuur is de echte volgorde van binnenkomst gebruikt.



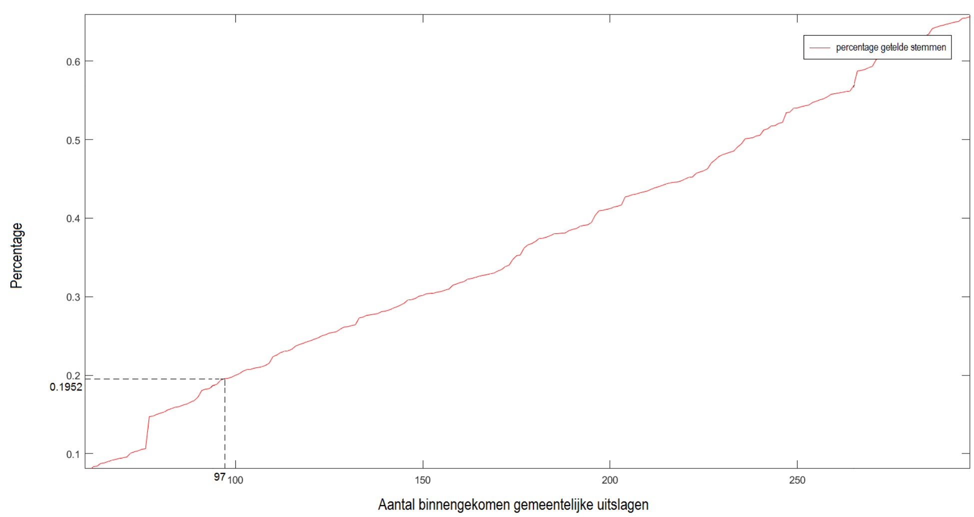
Figuur 2: De convergentie van de voorspellingen van Methode A, Methode B en Methode C. Voor dit figuur is de echte volgorde van binnenkomst gebruikt. (Dit figuur is een ingezoomde variant van Figuur 1)

Tabel 1: De convergentiepunten van de drie methodes bij de echte volgorde.

Methode	Convergentiepunt
Methode A	121
Methode B	403
Methode C	213

4.1.2 ANP

Het onderzoek van S. Dijkstra [6] heeft de voorspellingen van de uitslag in percentages met Methode A omgezet in een zetelverdeling. Uit zijn onderzoek blijkt dat alle voorspellingen met Methode A, die gemaakt zijn nadat er 97 gemeenten zijn binnengekomen, een evenbeeld zijn van de daadwerkelijke zetelverdeling van de einduitslag. In dit geval is de echte volgorde van binnenkomst van gemeentelijke uitslagen gebruikt.
De uitslagen van de eerste 97 gemeenten komt overeen met 19,52 procent van het totaal aantal stemmen, zie Figuur 3. Uit de liveblogs van het NRC Handelsblad van 2012 blijkt alle voorspellingen van het ANP niet correct zijn, zie de tabellen in Appendix H. Uit deze cijfers kan geconcludeerd worden dat Methode A daadwerkelijk beter is dan de onbekende voorspelmethode van het ANP.



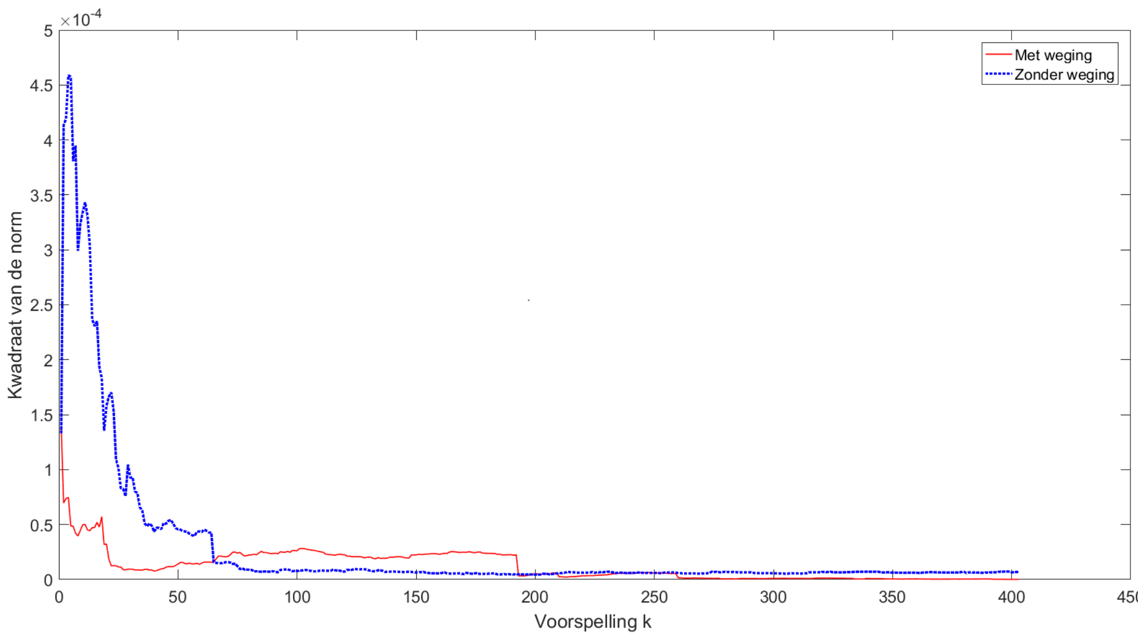
Figuur 3: Het aantal binnengekomen gemeenten uitgezet tegen het daarbij horende cumulatieve percentage van het totaal aantal stemmen.

4.2. Convergentie

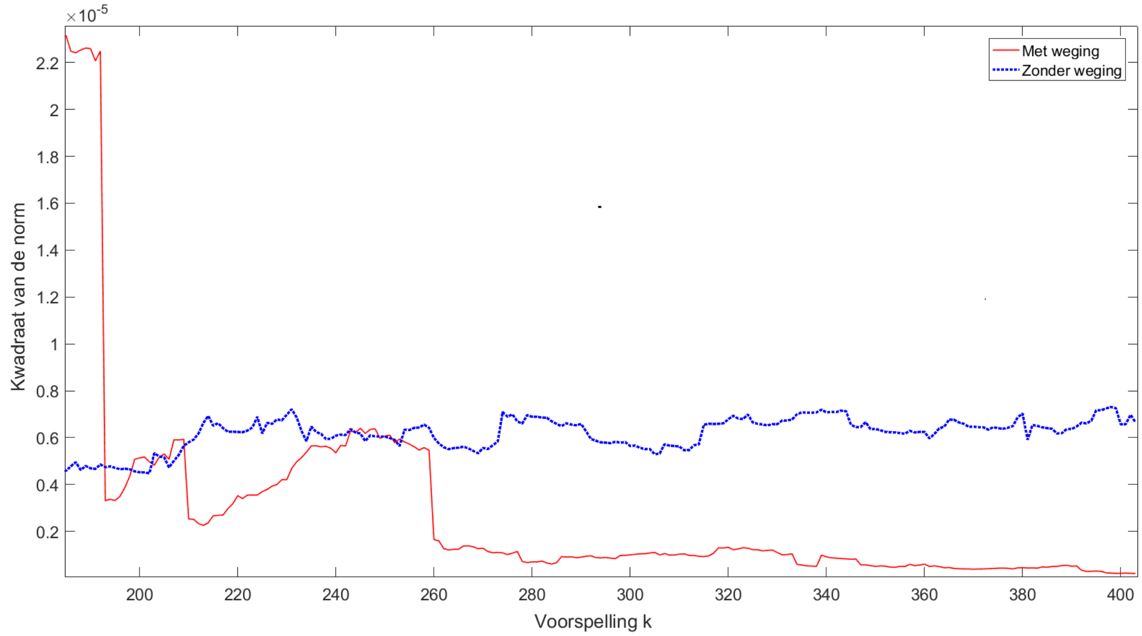
In deze paragraaf wordt gekeken naar de invloed van weging en verschillende volgorden van binnenkomst op de convergentie van de voorspelling. Ook wordt er onderzocht of het toekennen van verschillende wegingen aan gemeenten op basis van het aantal uitgebrachte stemmen een verbetering oplevert van de voorspellingen.

4.2.1 Invloed van weging

Uit Figuur 4 en Figuur 5 blijkt dat de fout van de voorspelling met weging soms tijdelijk groter is dan de fout van de voorspelling zonder weging. Belangrijk is dat de uiteindelijke voorspelling met weging wel dichterbij de einduitslag ligt. Aangezien de zetelverdeling in sommige gevallen bij een kleine verandering van de voorspelde uitslag een andere verdeling geeft, is het belangrijker dat de einduitslag dichterbij de werkelijke uitslag ligt dan dat deze eerder convergeert. Dus, hoewel de methode zonder weging sneller convergeert, is er een grotere kans dat deze voorspelling geen weerspiegeling zal zijn van de einduitslag. De nauwkeurigheid van de verschillende voorspellingen staat in Tabel 2. Uit deze tabel blijkt dat de voorspelling met weging vijf maal zo nauwkeurig is. Om deze reden is ervoor gekozen om het toekennen van wegingen mee te nemen in de ontwikkelde voorspelmethode.



Figuur 4: Het verloop van de fout de voorspellingen van Methode A met en zonder weging.



Figuur 5: Het verloop van de fout de voorspellingen van Methode A met en zonder weging. (Dit figuur is een ingezoomde variant van Figuur 4)

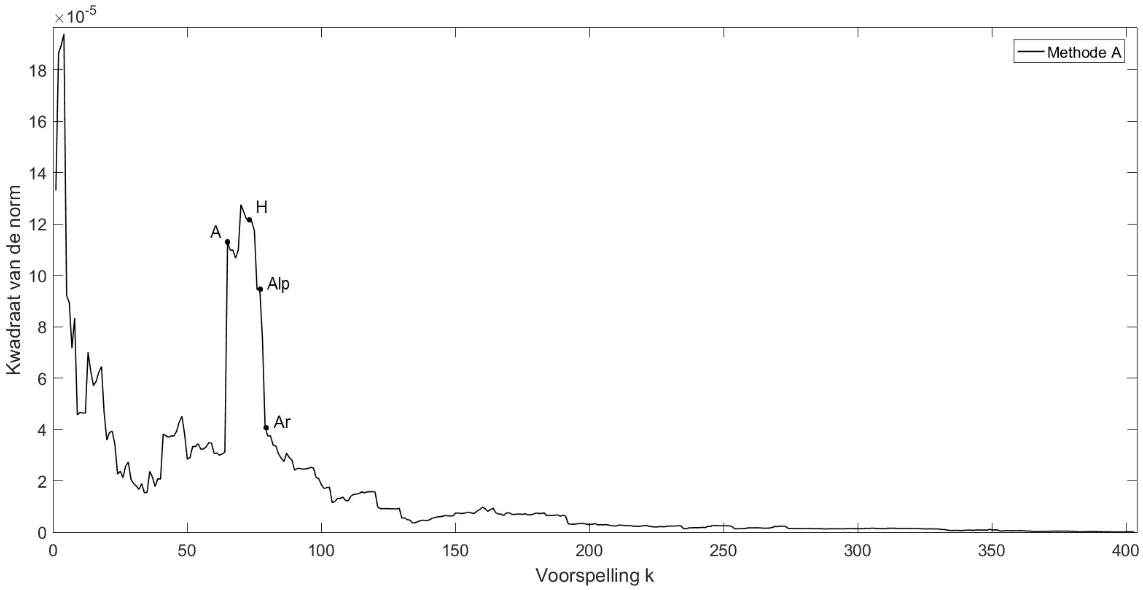
Tabel 2: De echte uitslag met daarnaast de uiteindelijke voorspelde uitslagen met en zonder weging in stemmen. De voorspelling zijn gedaan nadat alle gemeentelijke uitslagen zijn binnengekomen.

Partij	Echte uitslag	Eindvoorspelling met weging	Eindvoorspelling zonder weging
VVD	2.504.948	2.504.496	2.499.721
PvdA	2.340.750	2.339.820	2.362.035
PVV	950.263	953.634	946.405
CDA	801.620	801.394	805.697
SP	909.853	911.199	904.057
D66	757.091	755.784	750.685
GroenLinks	219.896	219.491	218.967
CU	294.586	293.452	293.104
SGP	196.780	196.179	196.872
Partij voor de Dieren	182.162	182.166	181.972
Partij voor Mens en Spiritualieteit	18.310	18.325	18.360
Piraten Partij	30.600	30.611	29.564
Overig	217.376	217.684	216.796
Totaal aantal stemmen	9.424.235	9.424.235	9.424.235
Verschil met echte uitslag	0	10.110	51.008

De waarden van de laatste rij in Tabel 2 zijn gevonden door het absolute verschil tussen te nemen tussen het voorspelde aantal stemmen en het echte aantal stemmen per partij. Vervolgens zijn al deze verschillen bij elkaar opgeteld. Dit resulteert dan in de waarden die in deze rij staan.

4.2.2 Norm

Zoals geïllustreerd in Figuur 6 convergeert de voorspelling met Methode B naar de daadwerkelijke uitslag van de verkiezing. Wat opvalt aan de convergentie is dat er bij voorspelling 65 een piek ontstaat. Vervolgens wordt deze piek weer gecorrigeerd bij voorspelling 79. In Tabel 3 staat de informatie over de punten die belangrijk zijn voor het verklaren van deze piek.



Figuur 6: Convergentie voorspellingen van Methode A. Punt 'A' is de voorspelling wanneer Amsterdam binnenkomt, punt 'H' die van Heemskerk, punt 'Alp' die van Alphen a/d Rijn en punt 'Ar' die van Arnhem.

Tabel 3: Informatie over punten in Figuur 6.

	Amsterdam	Heemskerk	Alphen a/d Rijn	Arnhem
CBS-code	363	396	484	202
Aantal stemmen	389267	22453	41613	80366
Positie lijst grootte gemeente	1	110	42	16
Positie lijst modelgemeenten	329	3	72	196
Voorspelling	65	76	78	79

Om Tabel 3 goed te kunnen begrijpen moeten eerst de begrippen 'lijst grootte gemeente' en 'lijst modelgemeenten' uitgelegd worden. De lijst met de grootte van de gemeente is een lijst gesorteerd op het aantal uitgebrachte stemmen per gemeente, zie Bijlage 2. De lijst met modelgemeenten is een lijst die gesorteerd is op hoe dicht een gemeentelijke uitslag bij de daadwerkelijk uitslag ligt, zie Bijlage 1.
De eerste piek bij voorspelling 65 ontstaat doordat de gemeente Amsterdam binnenkomt. Dat

er een piek ontstaat, komt doordat Amsterdam de grootste weegfactor heeft en tegelijkertijd een gemeentelijke uitslag heeft die niet dicht in de buurt ligt van de landelijke uitslag. Voor de andere drie punten geldt het tegenovergestelde: de gemeentelijke uitslagen van deze gemeenten liggen dichter of heel dicht in de buurt van de landelijke uitslag. Bovendien hebben Arnhem en Alphen a/d Rijn ook een hoge weegfactor.

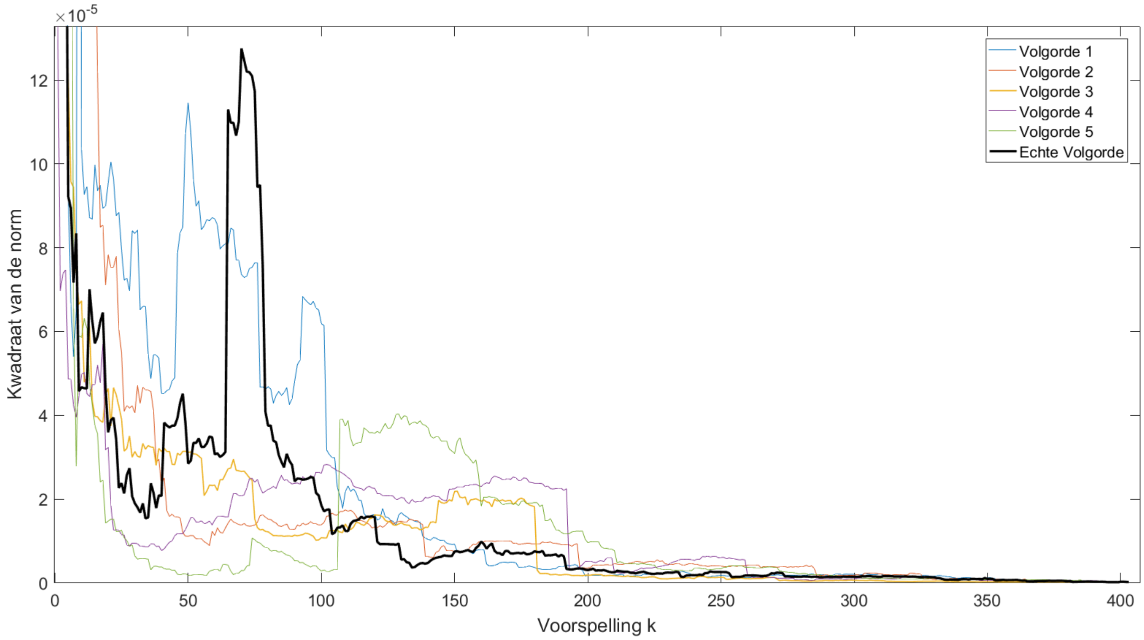
Dit roept de volgende vraag op: 'waarom ontstaat er niet een dergelijke piek wanneer andere grote gemeenten binnenkomen, zoals Rotterdam?'. De verklaring hiervoor ligt in het feit dat Rotterdam pas bij voorspelling 403 binnenkomt. Er zijn op dat moment al zoveel gemeenten binnengekomen dat Rotterdam geen grote invloed meer heeft op de voorspelling. Dit in tegenstelling tot Amsterdam waar er nog maar 64 uitslagen van andere gemeenten waren binnengekomen. Bovendien staat Rotterdam op positie 173 van de lijst met modelgemeenten in tegenstelling tot Amsterdam dat op positie 329 staat. Dit betekent dat de uitslag van Rotterdam relatief minder afwijkt van de uiteindelijke landelijke uitslag dan de uitslag van Amsterdam.

4.2.3 Volgorde

Om te controleren of de volgorde van het binnenkomen van gemeentelijke uitslagen invloed heeft op de convergentiesnelheid, worden er naast de daadwerkelijk volgorde van binnenkomst nog vijf willekeurige volgordes getest. De reden hiervoor is dat verschillende volgordes verschillende datasets geven voor het voorspellen. Deze datasets zijn verschillend omdat op het moment dat er k gemeenten zijn binnengekomen dit voor de verschillende volgordes in het algemeen niet dezelfde k gemeenten zijn.
Uit Figuur 7 en Tabel 4 blijkt dat de volgorde invloed heeft op de convergentiesnelheid. Hoewel de willekeurige volgordes relatief laat convergeren ten opzichte van de echte volgorde, is het resultaat nog steeds positief: er zijn slechts 200 gemeenten nodig om een voorspelling te geven met een nauwkeurigheid van $1 \cdot 10^{-5}$. Dit betekent dat grote schommelingen van de voorspellingen na 200 gemeenten nauwelijks meer voor zullen komen.

Tabel 4: De convergentiepunten van de verschillende volgordes.

	Volgorde 1	Volgorde 2	Volgorde 3	Volgorde 4	Volgorde 5	Echte Volgorde
Convergentiepunt	147	172	181	193	200	121



Figuur 7: Convergentie van de voorspellingen van Methode A van de echte volgorde en vijf willekeurige volgordes.

Hoewel de convergentiesnelheden van de volgordes verschillend zijn, blijkt uit Figuur 7 dat alle volgordes wel naar dezelfde eindvoorspelling convergeren. Dit is logisch te verklaren door naar de updatemethode te kijken. Wanneer alle 415 gemeenten binnengekomen zijn, kan door middel van de commutativiteit van het optellen van matrices de volgorde van deze 415 gemeenten veranderd worden zonder dat de uitkomst verandert. Dit verklaart dat de eindvoorspelling van alle volgordes dezelfde is.

5. Conclusie en discussie

Het doel van dit onderzoek was het ontwikkelen van een nieuwe voorspelmethode die de fluctuaties in voorspelling snel kan dempen. De methode ontwikkeld in dit onderzoek bestaat uit het vinden van een transformatiematrix die de gemeentelijke uitslagen van de vorige verkiezing met behulp van een weging afbeeldt op de binnengekomen gemeentelijke uitslagen van de huidige verkiezing. Deze weging bestaat uit het aantal stemmen per gemeente gedeeld door het totaal aantal uitgebrachte stemmen.

Uit de resultaten blijkt dat de voorspellingen van Methode A grote fluctuaties kan dempen nadat er 200 gemeentelijke uitslagen zijn binnengekomen. Voordat deze methode daadwerkelijk gebruikt kan worden, moeten de voorspellingen in percentages omgezet worden in een zetelverdeling. Dit kan worden gedaan met het algoritme van S. Dijkstra [6].

Hoewel aan de hand van de voorspellingen van het ANP uit het liveblog van het NRC Handelsblad van 2012 geconcludeerd kan worden dat Methode A eerder met zekerheid uitspraak kan doen over de einduitslag van de verkiezing dan het ANP, kan niet met zekerheid gezegd worden dat dit waar is. Dit komt door de onbetrouwbaarheid van de bron. Ten eerste vermeldt het NRC niet de bron die de voorspellingen heeft gemaakt. Ten tweede vanwege het feit dat de data van het NRC niet volledig correct zouden kunnen zijn. Dit is al te zien aan de voorspelling wanneer er 95 procent van de stemmen is geteld: het voorspelde aantal zetels komt uit op 151 terwijl er maar 150 zetels te verdelen zijn. Om met zekerheid te kunnen concluderen dat Methode A beter werkt dan de huidige voorspelmethode van het ANP, zullen er meer gegevens beschikbaar moeten zijn over de voorspellingen van het ANP. Deze gegevens waren lastig te achterhalen omdat de meeste liveblogs waren verdwenen en de uitzendingen van de NOS niet meer beschikbaar waren. Een duidelijkere uitspraak kan hierover gedaan worden wanneer bij de volgende verkiezingen Methode A wordt getest en de voorspellingen van Methode A direct kunnen worden vergeleken met de voorspellingen van het ANP.

De limitaties van het onderzoek liggen vooral in de data die beperkt aanwezig of onvolledig zijn. Een onvolledigheid is te vinden in de data van de verkiezingen van 2010. Doordat er bij de verkiezingen van 2010 geen data beschikbaar zijn over bepaalde stemmers en gemeenten die wel meededen in 2012, zoals de briefstemmers en Bonaire, konden deze uitslagen niet meegenomen worden in het voorspellen van de einduitslag. Het wel meenemen van deze uitslagen zou kunnen resulteren in een uiteindelijke T die nauwkeuriger is.

Een andere limitatie is dat de conclusie enkel gebaseerd is op de verkiezingen van 2012. Dit komt mede doordat er geen volgorde van binnenkomst bekend is van andere jaren. Bovendien zijn er ook geen voorspellingen van het ANP van andere jaren beschikbaar om de resultaten van Methode A mee te kunnen vergelijken.

Mogelijkheden voor vervolgonderzoek liggen in de interpretatie van de transformatiematrix T. De interpretatie van T, gegeven in Paragraaf 2.1.3, verklaart nog niet waarom bepaalde elementen van T negatief of groter dan 1 zijn. Er is een vermoeden dat dit wordt veroorzaakt door het verdwijnen van oude stemmers en het ontstaan van nieuwe stemmers. Deze stroming van het verdwijnen en ontstaan van stemmers zou niet vastgelegd kunnen worden met een verschuiving, omdat verdwenen stemmers eerst op een partij stemden en daarna op niets en omdat nieuwe stemmers eerst op niets stemden en nu wel op een partij.

Tevens zou er een vervolgstudie gedaan kunnen worden naar zogenaamde 'modelgemeenten'. Modelgemeenten zijn gemeenten waarvan de gemeentelijke uitslag dichtbij de landelijke uitslag ligt. Tijdens dit onderzoek is al kort gekeken naar modelgemeenten en er blijkt dat de gemeente Apeldoorn structureel in de top drie van de beste modelgemeenten staat. Daarnaast is het interessant om bijvoorbeeld te onderzoeken hoe de voorspellingen zijn als de datasets enkel bestaan uit de 20 beste modelgemeenten. Een dergelijke voorspelling zou misschien een betere voorspelling kunnen geven dan Methode A, omdat de uitslagen van deze modelgemeenten dicht bij de einduitslag ligt.

Om dezelfde reden zou het interessant zijn om in deze 20 gemeenten structureel *exit polls* te houden. Wanneer er zulke *polls* van meerdere verkiezingen zijn verzameld, zou aan de hand van een transformatiematrix de verschuiving in de *exit polls* van de vorige verkiezing naar de huidige bepaald kunnen worden. Deze matrix kan daarna gebruikt worden om een voorspelling te geven van de einduitslag zelfs al voordat er gemeentelijke verkiezingsuitslagen binnenkomen.

6. Woord van dank

Ik bedank de volgende personen voor hun bijdrage aan mijn onderzoek. Frits van Beckum voor zijn ideeën over de opdracht, de eerste opzet en de interesse die hij toonde tijdens het uitvoeren van de opdracht. Mijn begeleider Harry Aarts voor zijn constructieve bijdrage en ondersteuning in het proces van dit onderzoek. Elise Smit, Frank Klein Schaarsberg en Aalt Doppenberg voor het herzien van de paper op duidelijkheid en leesbaarheid.

7. Lijst met afkortingen en veel gebruikte symbolen

Symbool	Betekenis
$\mathbb{N}$	De set van natuurlijke getallen
$\mathbb{R}$	De set van reële getallen
$\mathbb{R}_{>0}$	De set van alle reële getallen groter dan nul
$\mathbb{R}^n$	Euclidean n-ruimte
$\mathbb{R}^{n\times m}$	De set van reële $n \times m$ matrices
$\|\|.\|\|$	De norm
A^T	Transpose van matrix A
A^{-1}	Inverse van matrix A
$A^{\dagger}$	Pseudoinverse van A
T	Transformatiematrix Methode A
Λ	Transforamtiematrix Methode C
V^k	Matrix met de uitslagen van de huidige verkiezing van de k als eerst binnengekomen gemeenten
U^k	Matrix met de uitslagen van de vorige verkiezing van de k als eerst binnengekomen gemeenten
A^k	$U^k U^{k^T}$
B^k	$V^k U^{k^T}$
t_{ij}	Het element uit rij i en kolom j van T
λ_{ii}	Het diagonaal element uit rij i uit Λ
$\mathbf{u}$	De einduitslag van de vorige verkiezing
$\mathbf{u^i}$	Oude gemeentelijke uitslag in percentages van gemeente i
$\mathbf{v^i}$	Huidge gemeentelijke uitslag in percentages van gemeente i
$\mathbf{z^k}$	Voorspelling na k gemeenten
$\mathbf{z}$	Einduitslag van de huidige verkiezingen
$\mathbf{x^i}$	De som van alle oude gemeentelijke uitslagen in stemmen van de eerste i gemeenten
$\mathbf{y^i}$	De som van alle nieuwe gemeentelijke uitslagen in stemmen van de eerste i gemeenten
$\mathbf{a^i}$	De som van alle nieuwe gemeentelijke uitslagen in percentages van de eerste i gemeenten
w_i	De weegfactor van gemeente i
n	Aantal partijen
m	Aantal gemeenten
k	Tot dan toe binnengekomen aantal gemeentelijke uitslagen
K	Convergentiepunt
ϵ	De convergentiegrens
Methode A	De voorspelmethode uit Paragraaf 2.2
Methode B	De voorspelmethode uit Paragraaf 3.1
Methode C	De voorspelmethode uit Paragraaf 3.2

Referenties

[1] *J.J. Paulus, J. Praagman, Eerste Kamerverkiezingen, STAtOR 12 (2) (2011) ISSN 1567-3383.*

[2] *C.D. Meyer, Matrix analysis and applied linear algebra p.199 (2) (2000) SIAM ISBN 0-89871-454-0.*

[3] *U. Faigle, W. Kern, G. Still, Mathematical Optimization p.29 (2015).*

[4] *R. Penrose, On best approximate solutions of linear matrix equations, Mathematical Proceedings of the Cambridge Philosophical Society p. 17-19 1 (52) (1956) doi:10.1017/S0305004100030929.*

[5] *NRC Handelsblad (2012). Liveblog verkiezingen 2012 (deel I). Verkregen op 15 juni 2016 van:* `http://www.nrc.nl/nieuws/2012/09/12/live-de-uitslagen-van-de-tweede-kamerverkiezingen-2`

[6] *S. Dijkstra, De verdeling van Tweede Kamerzetels voorspellen op de verkiezingsavond (2016).*

A. Stellingen

In deze appendix zal een bewijs worden gegeven waarom de matrix A^k voor k groter dan of gelijk aan het aantal rijen in A een positief definiete matrix is en daarom een inverse heeft.

Definitie A.1. *Laat $S \in \mathbb{R}^{nxn}$ een symmetrische matrix zijn, dan is S een positief semi-definiete matrix als voor alle $\mathbf{x} \in \mathbb{R}^n$ geldt $\mathbf{x}^T S\mathbf{x} \geq 0$. De matrix S is positief definiet als voor alle $\mathbf{x} \in \mathbb{R}^n / \{\mathbf{0}\}$ geldt $\mathbf{x}^T S\mathbf{x} > 0$.*

Lemma A.2. *Laat $A \in \mathbb{R}^{n\times m}$ dan geldt dat AA^T positief semi-definiet is.*

Bewijs. Laat $\mathbf{x} \in \mathbb{R}^n$. Dan geldt $\mathbf{x}^T AA^T\mathbf{x} = (\mathbf{x}^T A)(\mathbf{x}^T A)^T = ||\mathbf{x}^T A||^2 \geq 0$. Hieruit volgt dat AA^T positief semi-definiet is. □

Lemma A.3. [Rank-nullity theorem] *Laat $A \in \mathbb{R}^{nxm}$, dan geldt dat de rank en de dimensie van de kernel van A samen optellen tot het aantal kolommen van A, met andere woorden: $rank(A) + dim(kern(A)) = m$. [2]*

Stelling A.4. *Laat $A \in \mathbb{R}^{n\times m}$ met $rank(A) = n$, dan geldt dat AA^T positief definiet is.*

Bewijs. Gegeven is dat $A \in \mathbb{R}^{n\times m}$ en $rank(A) = n$. Uit Lemma A.2 volgt dan dat AA^T positief semi-definiet is. Laat $\mathbf{x} \in \mathbb{R}^n$. Dan geldt $\mathbf{x}^T AA^T\mathbf{x} = ||A^T\mathbf{x}||^2$. Te bewijzen is dan dat $||A^T\mathbf{x}||^2 = 0$ desda $\mathbf{x} = \mathbf{0}$.
Uit Lemma A.3 volgt dat: $rank(A^T) + dim(kern(A^T)) = n$. Aangezien $rank(A^T) = rank(A) = n$, volgt dat $dim(kern(A^T)) = 0$. Dit implicieert $A^T\mathbf{x} = \mathbf{0}$ desda $\mathbf{x} = \mathbf{0}$. Dan geldt ook $||A^T\mathbf{x}||^2 = 0$ desda $\mathbf{x} = \mathbf{0}$. Hieruit volgt $\mathbf{x}^T AA^T\mathbf{x} = 0$ desda $\mathbf{x} = \mathbf{0}$, dus AA^T is positief definiet. □

Lemma A.5. *Laat $A \in \mathbb{R}^{n\times n}$ een symmetrische matrix zijn. Dan bestaat er een inverteerbare matrix $Q \in \mathbb{R}^{n\times n}$ zodanig dat*

$$D = QAQ^T$$

een diagonaal matrix is. [3]

Lemma A.6. *Laat $A \in \mathbb{R}^{n\times n}$ een symmetrische matrix zijn en $Q \in \mathbb{R}^{n\times n}$ een inverteerbare matrix zodanig dat $D = QAQ^T$ diagonaal is. Dan is A positief definiet dan en slechts dan als alle diagonaal elementen van D strict positief zijn.*

Bewijs. Uit Lemma A.5 volgt dat Q bestaat.
Stel dat A postief definiet is. Laat $\mathbf{x} \neq \mathbf{0}$. Dan geldt $\mathbf{x}^T A\mathbf{x} > 0$. $D = QAQ^T$ implicieert $A = Q^{-1}D(Q^T)^{-1}$. Dan geldt $\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T Q^{-1}D(Q^T)^{-1}\mathbf{x} > 0$. Kies $\mathbf{y} = (Q^T)^{-1}\mathbf{x}$, dan ook $\mathbf{y} \neq \mathbf{0}$. Dan geldt

$$\mathbf{y}^T D\mathbf{y} = \sum_{i=1}^{n} d_{ii}y_i y_i = \sum_{i=1}^{n} d_{ii}y_i^2 > 0.$$

Aangezien $\mathbf{y} \neq \mathbf{0}$ geldt $d_{ii} > 0$ voor alle $1 \leq i \leq n$.
Stel nu dat alle elementen van D strict positief zijn. Laat opnieuw $\mathbf{y} \neq \mathbf{0}$. Dan $d_{ii} > 0$ voor alle $1 \leq i \leq n$. Dus $\sum_{i=1}^{n} d_{ii}y_i^2 = \mathbf{y}^\mathbf{T} D\mathbf{y} > 0$ voor alle $\mathbf{y} \neq \mathbf{0}$. Kies nu $\mathbf{y} = (Q^T)^{-1}\mathbf{x}$, dan ook $\mathbf{x} \neq \mathbf{0}$. Dan geldt:

$$\begin{aligned}
((Q^T)^{-1}\mathbf{x})^T D (Q^T)^{-1}\mathbf{x} &> 0 \\
\mathbf{x}^T ((Q^T)^{-1})^T D (Q^T)^{-1}\mathbf{x} &> 0 \\
\mathbf{x}^T Q^{-1} QAQ^T (Q^T)^{-1}\mathbf{x} &> 0 \\
\mathbf{x}^T A\mathbf{x} &> 0.
\end{aligned}$$

Dus A is postief definiet. □

Stelling A.7. *Elke positief definiete matrix heeft een inverse.*

Bewijs. Laat $S \in \mathbb{R}^{nxn}$ positief definiet zijn. Dan bestaat er volgens Lemma A.6 en Lemma A.5 een inverteerbare matrix Q en een diagonaal matrix D met strict positieve diagonaal elementen zodanig dat $D = QSQ^T$ ofwel $S = Q^{-1}DQ^{T^{-1}}$. Aangezien D volgens Lemma A.6 alleen maar strict positieve elementen op de diagonaal heeft, kan D als volgt worden geschreven: $D = \sqrt{D}\sqrt{D}$. De matrix $\sqrt{D}$ is de matrix waarvan de elementen gelijk zijn aan elementsgewijs de wortel nemen van de elementen van D. Bovendien zijn D en daarom ook $\sqrt{D}$ inverteerbaar, omdat deze matrices in elke rij een pivot element hebben.
Laat $B = Q^{-1}\sqrt{D}$, dan geldt:

$$\begin{aligned}
S &= Q^{-1}\sqrt{D}\sqrt{D}(Q^T)^{-1} \\
&= (Q^{-1}\sqrt{D})(((Q^T)^{-1})^T\sqrt{D^T})^T \\
&= (Q^{-1}\sqrt{D})(Q^{-1}\sqrt{D^T})^T \\
&= (Q^{-1}\sqrt{D})(Q^{-1}\sqrt{D})^T \\
&= BB^T.
\end{aligned}$$

Aangezien Q en $\sqrt{D}$ inverteerbaar zijn, is het product ook inverteerbaar. $B = Q^{-1}\sqrt{D}$ is dus inverteerbaar. Aangezien B inverteerbaar is en daarom ook B^T, is het product van deze matrices ook inverteerbaar. $S = BB^T$ is dus ook inverteerbaar. Hieruit volgt dat S een inverse heeft. □

Corollary A.8. *Laat $A \in \mathbb{R}^{n\times m}$ en $rank(A) = n$, dan is AA^T positief definiet en heeft dus een inverse.*

Bewijs. Dit is een direct gevolg uit Stelling A.4 en Stelling A.7.

□

Deze stellingen zijn nodig om de volgende bewerking te kunnen gebruiken:

$$\begin{aligned}
T^kU^k &= V^k && (8)\\
T^kU^kU^{k^T} &= V^kU^{k^T} && (9)\\
T^kA^k &= B^k && (10)\\
T &= B^kA^{k^{-1}}. && (11)
\end{aligned}$$

Waarbij de stap van 10 naar 11 volgt uit Corollary A.8. Deze stelling kan enkel worden gebruikt in de veronderstelling dat de matrix U^k volledige rijrang heeft. Aangezien de elementen in U^k getallen met zestien decimalen zijn, is de kans dat de rijen lineair onafhankelijk van elkaar zijn vrijwel één.

B. Kleinste kwadraten methode

In deze appendix zal bewezen worden dat de oplossing $T = BA^{-1} = VU^T(UU^T)^{-1}$ het kleinste kwadraten probleem $||TU - V||^2$ minimaliseert. Het bewijs zal bestaan uit het vinden van ondergrens voor de norm en er zal worden aangetoond dat deze ondergrens de kleinste ondergrens is. In het bewijs zal gebruik worden gemaakt van de volgende identiteit:

$$||AP + (I - AA^{\dagger})Q|| = ||AP|| + ||(I - AA^{\dagger})Q|| \text{ [4]}, \tag{12}$$

waarbij $A^{\dagger} = (A^T A)^{-1} A^T$ de pseudoinverse is van A.

In het bewijs zullen de volgende variabelen regelmatig voor elkaar worden gesubstitueerd:

$$\begin{aligned} A^{\dagger} &= (A^T A)^{-1} A^T \\ U^{\dagger} &= U^T (UU^T)^{-1} \\ U &= A^T \\ Q^T &= -V \\ P^T &= T - VU^{\dagger} \end{aligned}$$

waarbij $U^{\dagger} = U(UU^T)^{-1}$ de pseudoinverse is van U.

.

Stelling B.1. *Het kleinste kwadraten probleem $||TU - V||^2$ heeft een minimale oplossing voor $T = VU^T(UU^T)^{-1}$.*

Bewijs. Om identiteit 12 te kunnen gebruiken, zal $||TU - V||$ omgeschreven moeten worden in de vorm van 12. In de volgende bewerkingen zullen de eerder genoemde substituties gebruikt worden.

De eerste bewerking is als volgt

$$\begin{aligned} ||TU - V|| &= ||TU - VU^{\dagger}U + VU^{\dagger}U - V|| && (13) \\ &= ||(T - VU^{\dagger})U + (-V)(I - U^{\dagger}U)||. && (14) \end{aligned}$$

Vervolgens zal vergelijking 14 zo omgeschreven worden totdat de linkerkant van het gelijkteken van identiteit 12 zichtbaar wordt, waarna deze identiteit toegepast kan worden.

$$\begin{aligned}
||(T-VU^{\dagger})U+(-V)(I-U^{\dagger}U)|| &= ||P^TU+Q^T(I-U^{\dagger}U)|| \\
&= ||P^TU+Q^T(I-(U^T(UU^T)^{-1})U)|| \\
&= ||P^TA^T+Q^T(I-A(A^TA)^{-1}A^T)|| \\
&= ||P^TA^T+Q^T(I-((A^TA)^{-1}A^T)^TA^T))|| \\
&= ||P^TA^T+Q^T(I-(A^{\dagger})^TA^T)|| \\
&= ||(P^TA^T+Q^T(I-(A^{\dagger})^TA^T))^T|| \\
&= ||AP+(I-AA^{\dagger})Q|| \\
&= ||AP||+||(I-AA^{\dagger})Q|| \\
&= ||(AP)^T||+||((I-AA^{\dagger})Q)^T|| \\
&= ||P^TA^T||+||Q^T(I-(A^{\dagger})^TA^T)|| \\
&= ||P^TA^T||+||Q^T(I-((A^TA)^{-1}A^T)^TA^T)|| \\
&= ||P^TA^T||+||Q^T(I-A(A^TA)^{-1}A^T)|| \\
&= ||P^TU||+||Q^T(I-U^T(UU^T)^{-1}U)|| \\
&= ||P^TU||+||Q^T(I-U^{\dagger}U)|| \\
&= ||(T-VU^{\dagger})U||+||(-V)(I-U^{\dagger}U)||.
\end{aligned}$$

Hieruit kan geconcludeerd worden dat

$$||(T-VU^{\dagger})U+(-V)(I-U^{\dagger}U)|| = ||(T-VU^{\dagger})U||+||(-V)(I-U^{\dagger}U)||.$$

Vervolgens kan de ondergrens worden bepaald voor $||TU-V||^2$:

$$\begin{aligned}
||TU-V||^2 &= ||(T-VU^{\dagger})U+(-V)(I-U^{\dagger}U)||^2 \\
&= (||(T-VU^{\dagger})U||+||(-V)(I-U^{\dagger}U)||)^2 \\
&= ||(T-VU^{\dagger})U||^2+2||(T-VU^{\dagger})U||\cdot||(-V)(I-U^{\dagger}U)||+||(-V)(I-U^{\dagger}U)||^2 \\
&\geq ||(-V)(I-U^{\dagger}U)||^2.
\end{aligned}$$

Het gelijkteken geldt alleen als $(T-VU^{\dagger})U=0$, dit is alleen als $T=VU^{\dagger}=VU^T(UU^T)^{-1}$. Dus het kleinste kwadraten probleem $||TU-V||^2$ heeft een minimale oplossing voor $T=VU^T(UU^T)^{-1}$.

□

C. Updatemethode

De updatemethode voor A^k is op de volgende wijzen af te leiden:

$$\begin{aligned}
A^k_{ij} &= \sum_{l=1}^{k} U^k_{il} U^{k^T}_{lj} && (15)\\
&= \sum_{l=1}^{k-1} U^k_{il} U^{k^T}_{lj} + U^k_{ik} U^{k^T}_{kj} && (16)\\
&= \sum_{l=1}^{k-1} U^{k-1}_{il} U^{k-1^T}_{lj} + U^k_{ik} U^{k^T}_{kj} && (17)\\
&= A^{k-1}_{ij} + U^k_{ik} U^{k^T}_{kj} && (18)
\end{aligned}$$

voor $2 \leq k \leq m$ met $A^1_{ij} = U^1_{i1} U^1_{1j}$

De stap van (16) naar (17) is als volgt:

$$\sum_{l=1}^{k-1} U^k_{il} U^{k^T}_{lj} = \sum_{l=1}^{k-1} \left[\mathbf{u^1}, \mathbf{u^2}, \ldots, \mathbf{u^k}\right]_{il} \begin{bmatrix} \mathbf{u}^{1^\mathrm{T}} \\ \mathbf{u^2}^{T} \\ \vdots \\ \mathbf{u}^{\mathrm{k}^\mathrm{T}} \end{bmatrix}_{lj} = u^1_i u^{1^T}_j + u^2_i u^{2^T}_j + \ldots + u^{k-1}_i u^{k-1^T}_j =$$

$$= \sum_{l=1}^{k-1} \left[\mathbf{u^1}, \mathbf{u^2}, \ldots, \mathbf{u^{k-1}}\right]_{il} \begin{bmatrix} \mathbf{u^1}^{\mathrm{T}} \\ \mathbf{u^2}^{\mathrm{T}} \\ \vdots \\ \mathbf{u^{k-1}}^{\mathrm{T}} \end{bmatrix}_{lj} = \sum_{l=1}^{k-1} U^{k-1}_{il} U^{k-1^T}_{lj}$$

D. Weging

In deze appendix zal worden laten zien de twee methodes van wegen hetzelfde zijn.

Beschouw eerst de volgende methode: het vaker afbeelden van een gemeente op de einduitslag. Stel er zijn k gemeenten binnengekomen en alle gemeenten hebben een weging van 1 behalve gemeente i, $1 \leq i \leq k$, die heeft weging $w \in \mathbb{N}$. Dat betekent dat gemeente i w keer wordt afgebeeld:

$$T(\mathbf{u^1}, \mathbf{u^2}, ..., \mathbf{u^{i-1}}, \mathbf{u^i}, \mathbf{u^i}, ..., \mathbf{u^i}, \mathbf{u^{i+1}}, ..., \mathbf{u^k}) = (\mathbf{v^1}, \mathbf{v^2}, ..., \mathbf{v^{i-1}}, \mathbf{v^i}, \mathbf{v^i}, ..., \mathbf{v^i}, \mathbf{v^{i+1}}, ..., \mathbf{v^k}) \tag{19}$$

Wanneer A^k recursief bepaald wordt via de updatemethode geeft dat:

$$\begin{aligned}
A^k_{ij} &= U^1_{i1}U^{1^T}_{1j} + U^2_{i2}U^{2^T}_{2j} + ... + U^{i-1}_{i(i-1)}U^{i-1^T}_{(i-1)j} + U^i_{ii}U^{i^T}_{ij} + U^i_{ii}U^{i^T}_{ij} + ... + U^i_{ii}U^{i^T}_{ij} + U^{i+1}_{i(i+1)}U^{i+1^T}_{(i+1)j} + ... + U^k_{ik}U^{k^T}_{kj} \\
&= U^1_{i1}U^{1^T}_{1j} + U^2_{i2}U^{2^T}_{2j} + ... + U^{i-1}_{i(i-1)}U^{i-1^T}_{(i-1)j} + w \cdot U^i_{ii}U^{i^T}_{ij} + U^{i+1}_{i(i+1)}U^{i+1^T}_{(i+1)j} + ... + U^k_{ik}U^{k^T}_{kj}
\end{aligned}$$

Nu is duidelijk te zien dat bij het bepalen van A^k de weging w pas meegeven in de updatemethode hetzelfde effect heeft als gemeente i w keer afbeelden. In de algemenere zin, wanneer de overige wegingen niet nul zijn, geeft dit de volgende expressie voor A^k:

$$A^k_{ij} = w_1 \cdot U^1_{i1}U^{1^T}_{1j} + w_2 \cdot U^2_{i2}U^{2^T}_{2j} + ... + w_k \cdot U^k_{ik}U^{k^T}_{kj} \tag{20}$$

waarbij $w_i \in \mathbb{N}$ de weging is voor gemeente i.

Dat deze weegmethode daadwerkelijk hetzelfde is als het vaker afbeelden van een gemeente, wordt verduidelijkt aan de hand van een voorbeeld.

Voorbeeld 1. *Laat $u_1 \in \mathbb{R}^{2\times 1}$ en $u_2 \in \mathbb{R}^{2\times 1}$ de oude uitslagen van gemeente 1 en 2 zijn. Laat $v_1 \in \mathbb{R}^{2\times 1}$ en $v_2 \in \mathbb{R}^{2\times 1}$ de nieuwe uitslagen van deze gemeenten zijn. In dit geval is er dus sprake van 2 partijen waarop gestemd kan worden. Wanneer gemeente 1 zwaarder moet meetellen dan gemeente 2 zou dit op de boven beschreven manier kunnen. Aan gemeente 1 wordt een weging van 3 gegeven en aan gemeente 2 een weging van 1. Dit leidt dan tot de volgende expressie:*
$T^2U^2 = T(u_1, u_1, u_1, u_2) = (v_1, v_1, v_1, v_2) = V^2$
In dit voorbeeld heeft gemeente 1 een weging van drie en gemeente 2 een weging van één. Kies nu $u_1 = (a, b)^T$, $u_2 = (c, d)^T$, $v_1 = (e, f)^T$ en $v_2 = (g, h)^T$ met $a, b, c, d, e, f, g, h \in [0, 1]$. Voor de updatemethode zijn de volgende matrices nodig:

$$U^2 = \begin{bmatrix} a & a & a & c \\ b & b & b & d \end{bmatrix}$$

$$D{=}U^2U^{2^T} = \begin{bmatrix} a & a & a & c \\ b & b & b & d \end{bmatrix} \begin{bmatrix} a & b \\ a & b \\ a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 + a^2 + a^2 + c^2 & ab + ab + ab + cd \\ ab + ab + ab + cd & b^2 + b^2 + b^2 + d^2 \end{bmatrix} = \begin{bmatrix} 3a^2 + c^2 & 3ab + cd \\ 3ab + cd & 3b^2 + d^2 \end{bmatrix}$$

$V^2 = \begin{bmatrix} e & e & e & g \\ f & f & f & h \end{bmatrix}$

E=$V^2 U^{2^T} = \begin{bmatrix} e & e & e & g \\ f & f & f & h \end{bmatrix} \begin{bmatrix} a & b \\ a & b \\ a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ae + ae + ae + cg & be + be + be + dg \\ af + af + af + hc & bf + bf + bf + dh \end{bmatrix} = \begin{bmatrix} 3ae + cg & 3be + dg \\ 3af + hc & 3bf + dh \end{bmatrix}$

Nu kan worden geverifieerd dat wanneer de weging in de updatemethode wordt toegepast dezelfde matrices als hierboven oplevert. De updatemethode van stap $k = 1$ naar stap $k = 2$ is:

$$U^2 U^{2^T} = U^1 U^{1^T} + U^2 U^{2^T} \text{ en } V^2 U^{2^T} = V^1 U^{1^T} + V^2 U^{2^T}$$

Vervolgens wordt de weging toegekend:

$$P = U^2 U^{2^T} = 3 \cdot U^1 U^{1^T} + 1 \cdot U^2 U^{2^T} = 3 \cdot \begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} a & b \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \begin{bmatrix} c & d \end{bmatrix}$$

$$= 3 \cdot \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix} + \begin{bmatrix} c^2 & cd \\ cd & d^2 \end{bmatrix} = \begin{bmatrix} 3a^2 + c^2 & 3ab + cd \\ 3ab + cd & 3b^2 + d^2 \end{bmatrix}$$

$$R = V^2 U^{2^T} = 3 \cdot V^1 U^{1^T} + 1 \cdot V^2 U^{2^T} = 3 \cdot \begin{bmatrix} e \\ f \end{bmatrix} \cdot \begin{bmatrix} a & b \end{bmatrix} + \begin{bmatrix} g \\ h \end{bmatrix} \begin{bmatrix} c & d \end{bmatrix}$$

$$= 3 \cdot \begin{bmatrix} ae & be \\ af & bf \end{bmatrix} + \begin{bmatrix} cg & dg \\ hc & hd \end{bmatrix} = \begin{bmatrix} 3ae + cg & 3be + dg \\ 3af + dg & 3bf + hc \end{bmatrix}$$

Hieruit volgt dat $P = D$ en $R = E$, dus de twee methoden hebben dezelfde resultaten.

E. Interpretatie T

Voorbeeld 2. *Kies* $T \in \mathbb{R}^{3\times 3}$ *en* $U = \begin{bmatrix} 0,25 & 0,35 & 0,28 \\ 0,30 & 0,25 & 0,30 \\ 0,45 & 0,40 & 0,42 \end{bmatrix}$ *en* $V = \begin{bmatrix} 0,30 & 0,30 & 0,35 \\ 0,25 & 0,35 & 0,25 \\ 0,45 & 0,35 & 0,40 \end{bmatrix}$ *zijn matrices die de uitslagen van drie gemeenten van de vorige verkiezing en respectievelijk van dezelfde drie gemeenten van de huidige verkiezing bevatten. In de kolommen van deze matrices staan op de volgende volgorde de partijen: in rij 1 staan de uitslagen van partij 1, in rij 2 die van partij 2 en rij 3 die van partij 3. Dan beeldt* T U *af op* V*:* $TU = V$. *Op de eerder beschreven manier in paragraaf 2.1.1 kan* T *worden berekend:*

$$T = \begin{bmatrix} 0,55 & 2,2167 & -1,1167 \\ 0,85 & -1,15 & 0,85 \\ -0,40 & -0,0667 & 1,2667 \end{bmatrix}$$

De interpretatie dat T *de verschuiving van stemgedrag in kaart brengt wordt versterkt door het feit dat voor elke kolom van* T *geldt dat de elementen in die kolom sommeren tot 1. Aan de hand van de interpretatie van* T *in paragraaf 2.1.3 kunnen de elementen in kolom 1 op de volgende manier gezien worden:* t_{11} *is het deel van de stemmers dat partij 1 is trouw gebleven,* t_{21} *is het deel van de stemmers dat partij 1 in de vorige verkiezing stemde maar in de huidige verkiezing partij 2 stemt en* t_{31} *is het deel van de stemmers dat partij 1 in de vorige verkiezing stemde maar in de huidige verkiezing partij 3 stemt. Aangezien dit het gedrag van alle oude stemmers van partij 1 is, moet gelden dat deze elementen van* T *sommeren tot* 1. *Voor een vervolg studie is het interessant om te onderzoeken wat de betekenis is van de elementen van* T *die groter zijn dan* 1 *en kleiner dan 0.*

F. Uiteindelijke T

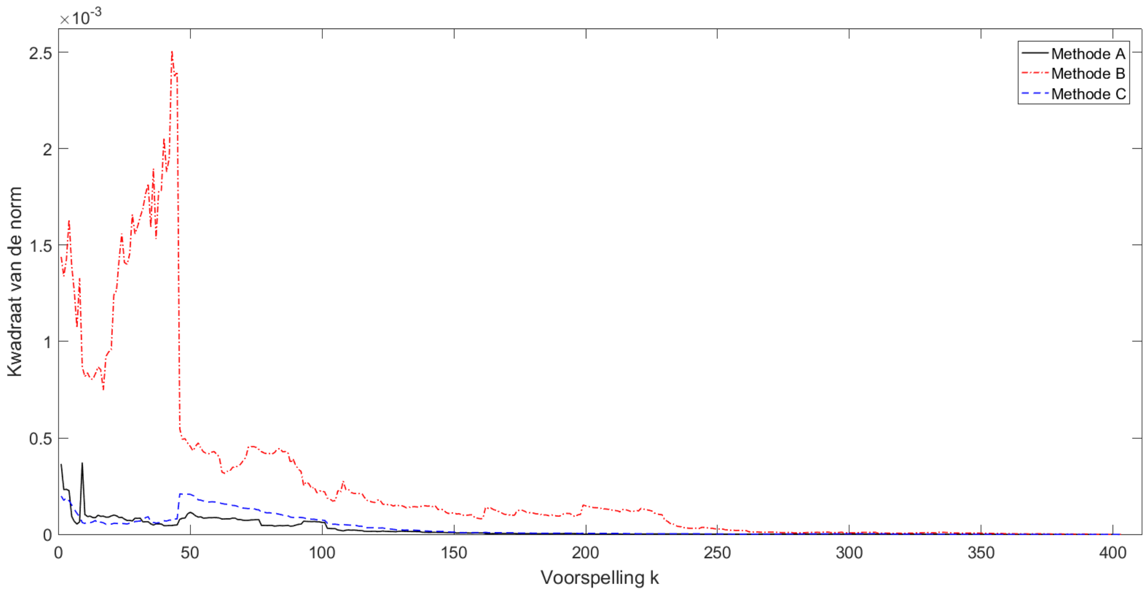
In deze appendix staat de matrix T die wordt gevonden nadat alle gemeentelijke uitslagen zijn binnengekomen.

$T =$

$$
\begin{vmatrix}
1{,}196 & -0{,}112 & 0{,}083 & 0{,}284 & -0{,}064 & -0{,}127 & 0{,}252 & -0{,}133 & 0{,}041 & -0{,}187 & -1{,}888 & -3{,}069 & 0{,}593 \\
-0{,}009 & 1{,}059 & 0{,}078 & 0{,}023 & 0{,}278 & -0{,}173 & 0{,}128 & 0{,}127 & -0{,}085 & -0{,}266 & 1{,}658 & 1{,}573 & -0{,}340 \\
-0{,}015 & -0{,}001 & 0{,}689 & -0{,}042 & -0{,}026 & -0{,}045 & -0{,}044 & 0{,}044 & -0{,}009 & 0{,}656 & -0{,}583 & 1{,}916 & 0{,}282 \\
-0{,}015 & 0{,}087 & -0{,}021 & 0{,}669 & -0{,}071 & -0{,}106 & -0{,}041 & 0{,}033 & -0{,}076 & 0{,}174 & 1{,}512 & 0{,}670 & -0{,}924 \\
-0{,}133 & 0{,}008 & 0{,}083 & 0{,}022 & 0{,}939 & 0{,}264 & -0{,}015 & -0{,}066 & 0{,}011 & -0{,}386 & 0{,}569 & -1{,}276 & 0{,}514 \\
-0{,}025 & -0{,}017 & 0{,}026 & 0{,}045 & -0{,}029 & 1{,}066 & 0{,}254 & -0{,}059 & 0{,}007 & -0{,}298 & -0{,}250 & -1{,}782 & -0{,}177 \\
-0{,}022 & -0{,}016 & 0{,}015 & 0{,}008 & -0{,}032 & 0{,}082 & 0{,}452 & -0{,}008 & 0{,}010 & -0{,}172 & -0{,}225 & -1{,}118 & -0{,}136 \\
-0{,}005 & -0{,}005 & 0{,}004 & 0{,}001 & -0{,}005 & 0{,}016 & 0{,}001 & 1{,}012 & -0{,}010 & 0{,}010 & -0{,}264 & -0{,}496 & 0{,}059 \\
-0{,}006 & -0{,}003 & 0{,}005 & 0{,}003 & -0{,}008 & 0{,}019 & -0{,}006 & 0{,}063 & 1{,}114 & -0{,}048 & -0{,}006 & -0{,}010 & 0{,}076 \\
0{,}007 & 0{,}004 & 0{,}005 & -0{,}008 & -0{,}013 & 0{,}002 & 0{,}043 & -0{,}001 & 0{,}002 & 1{,}168 & -0{,}042 & 0{,}405 & 0{,}001 \\
0{,}000 & -0{,}001 & 0{,}002 & 0{,}000 & 0{,}000 & -0{,}003 & 0{,}008 & 0{,}000 & 0{,}000 & 0{,}019 & 0{,}532 & 0{,}001 & -0{,}010 \\
-0{,}002 & 0{,}002 & 0{,}003 & -0{,}001 & 0{,}000 & 0{,}012 & 0{,}006 & 0{,}001 & 0{,}000 & 0{,}010 & -0{,}080 & 1{,}405 & 0{,}044 \\
0{,}030 & -0{,}005 & 0{,}029 & -0{,}004 & 0{,}031 & -0{,}007 & -0{,}038 & -0{,}014 & -0{,}005 & 0{,}321 & 0{,}067 & 2{,}782 & 1{,}019
\end{vmatrix}
$$

G. Verschillende Volgordes

G.1. Volgorde 1

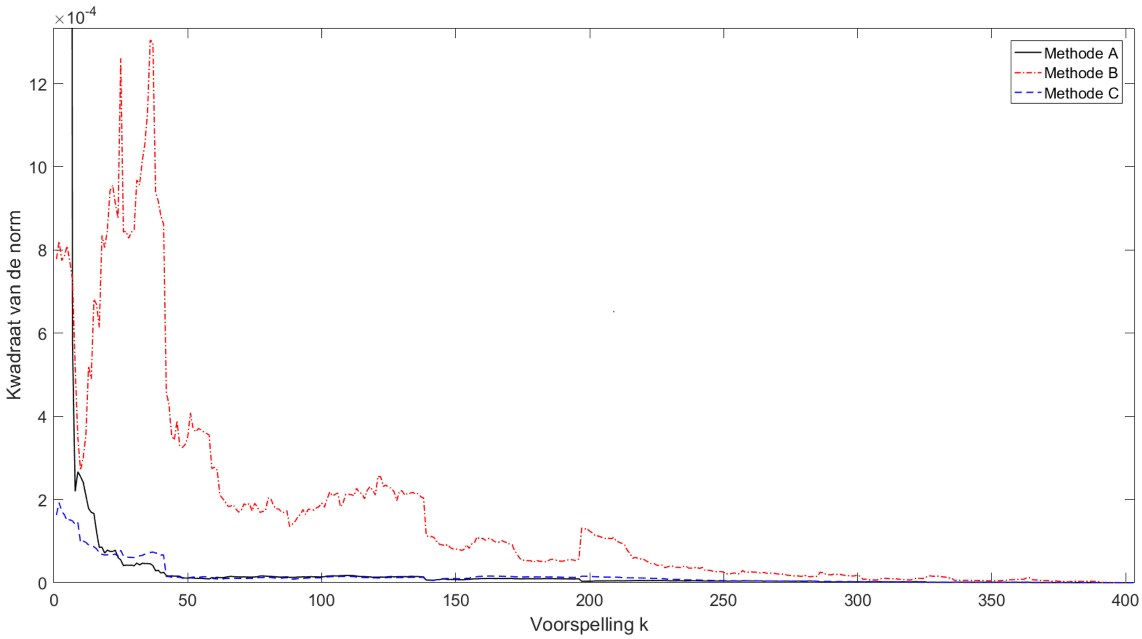


Figuur 8: De convergentie van de drie methoden bij Volgorde 1.

Tabel 5: De convergentiepunten van de drie methoden bij Volgorde 1.

Methode	Convergentiepunt
Methode A	147
Methode B	337
Methdoe C	175

G.2. Volgorde 2

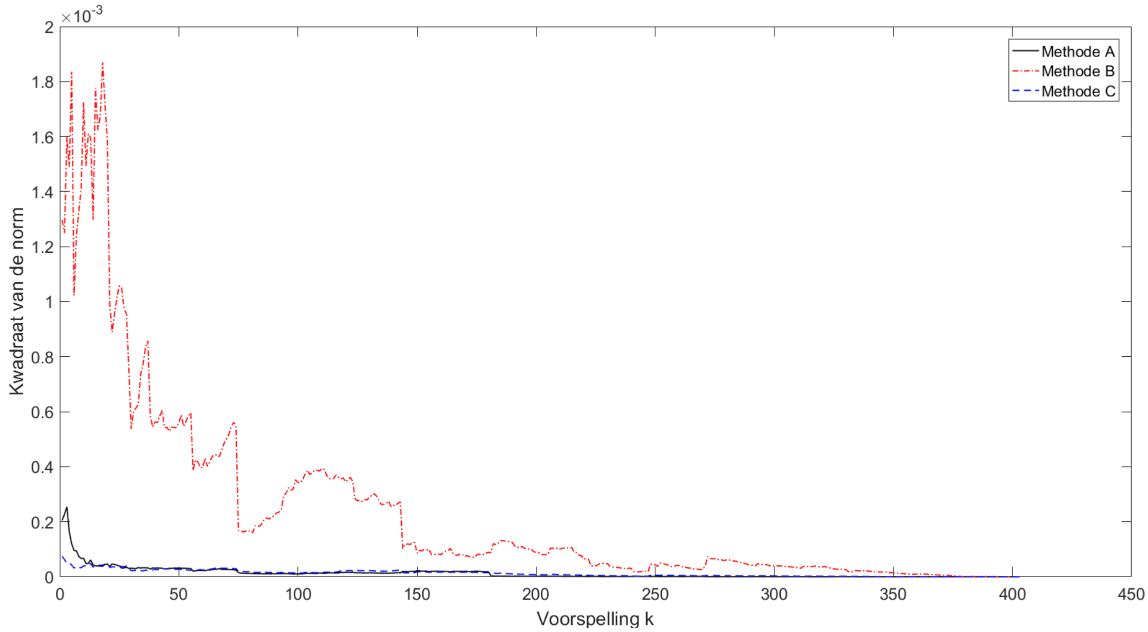


Figuur 9: De convergentie van de drie methoden bij Volgorde 2.

Tabel 6: De convergentiepunten van de drie methoden bij Volgorde 2.

Methode	**Convergentiepunt**
Methode A	172
Methode B	365
Methode C	239

G.3. Volgorde 3

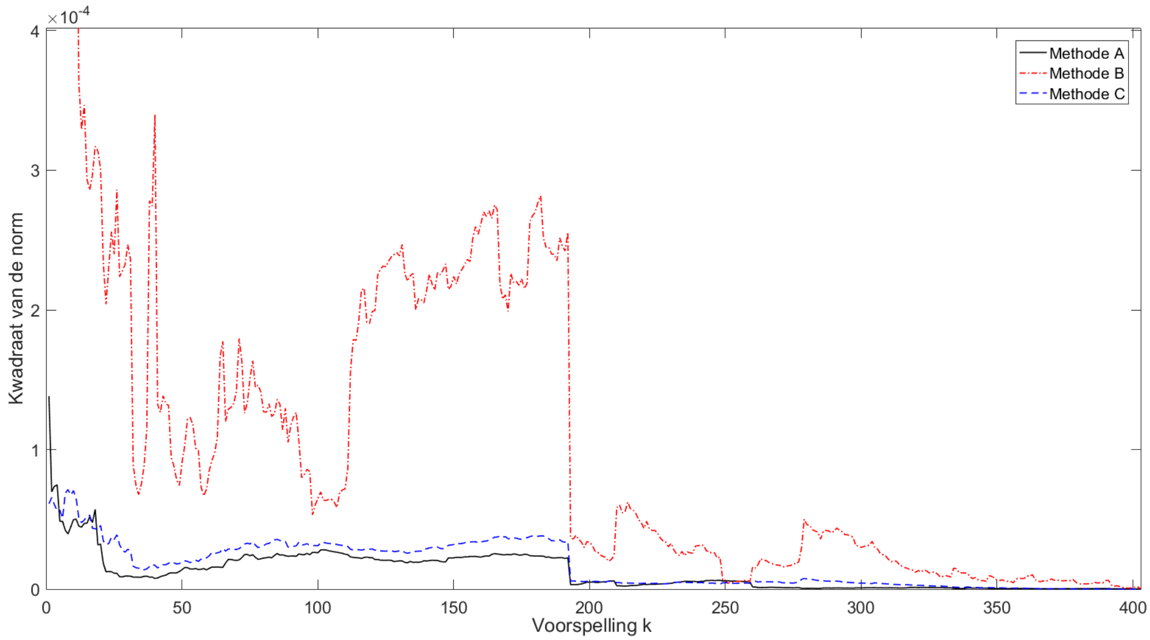


Figuur 10: De convergentie van de drie methoden bij Volgorde 3.

Tabel 7: De convergentiepunten van de drie methoden bij Volgorde 3.

Methode	**Convergentiepunt**
Methode A	181
Methode B	356
Methode C	204

G.4. Volgorde 4

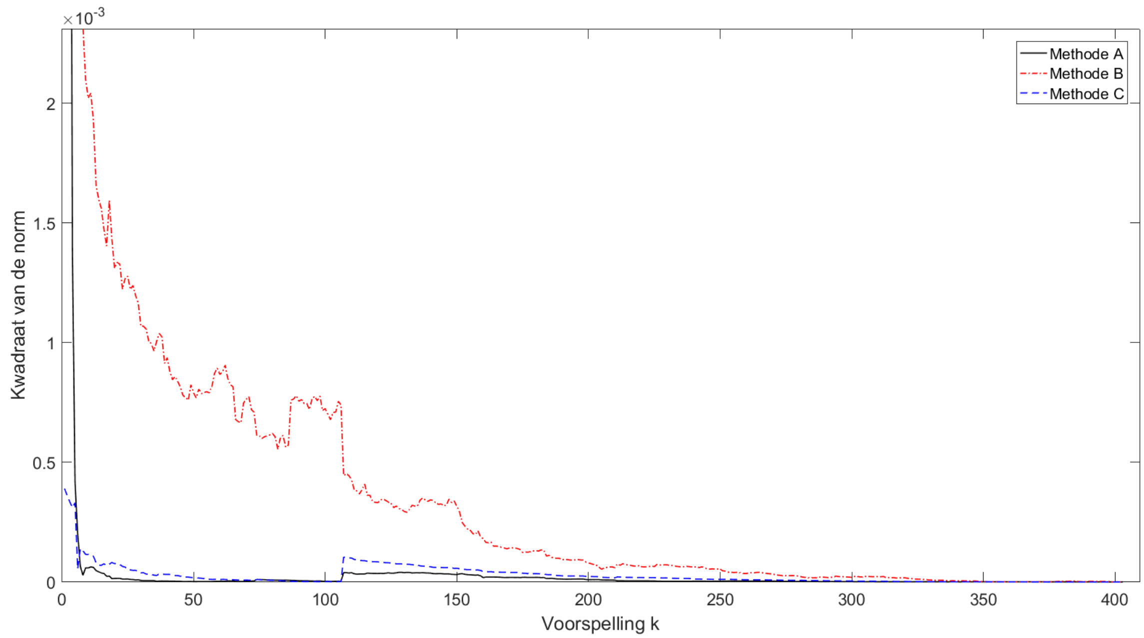


Figuur 11: De convergentie van de drie methoden bij Volgorde 4.

Tabel 8: De convergentiepunten van de drie methoden bij Volgorde 4.

Methode	**Convergentiepunt**
Methode A	193
Methode B	364
Methode C	205

G.5. Volgorde 5



Figuur 12: De convergentie van de drie methoden bij Volgorde 5.

Tabel 9: De convergentiepunten van de drie methoden bij Volgorde 5.

Methode	**Convergentiepunt**
Methode A	200
Methode B	329
Methode C	270

H. Voorspellingen ANP 2012

Tabel 10: Deel 1 van de voorspellingen van het ANP op basis van het liveblog van het NRC 5. De percentages geven aan hoeveel procent van het totaal aantal stemmen geteld is. De waarden in het rood zijn de verkeerd voorspelde waarden.

	Echte Uitslag	Voorspelling bij 95 %	Voorspelling bij 74,5%	Voorspelling bij 67,2%
VVD	41	41	40	40
PVDA	38	39	39	39
PVV	15	15	15	15
CDA	13	13	13	13
SP	15	15	15	15
D66	12	12	12	12
GroenLinks	4	3	3	3
ChristenUnie	5	5	5	5
SGP	3	3	3	3
PvdD	2	3	3	3
50+	2	2	2	2
aantal zetels	150	151	150	150

Tabel 11: Deel 2 van de voorspellingen van het ANP op basis van het liveblog van het NRC 5. De percentages geven aan hoeveel procent van het totaal aantal stemmen geteld is. De waarden in het rood zijn de verkeerd voorspelde waarden.

	Echte Uitslag	Voorspelling bij 58,1%	Voorspelling bij 25%	Voorspelling bij 13,6%
VVD	41	40	41	41
PVDA	38	39	39	37
PVV	15	15	15	15
CDA	13	13	13	13
SP	15	15	15	16
D66	12	12	12	12
GroenLinks	4	3	3	3
ChristenUnie	5	5	5	4
SGP	3	3	3	4
PvdD	2	3	2	3
50+	2	2	2	2
aantal zetels	150	150	150	150