

Verbeterd schattingsmodel voor korte afstandsmigratie in het regionaal prognosemodel PEARL

Rob Loke en Andries de Jong¹

Het Planbureau voor de Leefomgeving en het Centraal Bureau voor de Statistiek brengen om de twee jaar de regionaal demografische prognose uit. Deze prognose, met behulp van het model PEARL, geeft een beeld van regionale ontwikkelingen in de bevolking en huishoudens in komende drie decennia. In 2006 is de prognose voor het eerst uitgebracht en in 2011 voor de vierde keer. In deze prognose vormt de schatting van het aantal korte afstand migranten een belangrijk onderdeel. Hiertoe wordt gebruik gemaakt van het ruimtelijk interactiemodel, met een specificatie die door het PBL en het CBS toegesneden is op de regionale prognose. In de specificatie van dit model ten behoeve van de nieuwste prognose, is gebruik gemaakt van vier verklarende variabelen, te weten het inwonertal van de vestigingsgemeente, de reisafstand tussen vertrek- en vestigingsgemeente, de toegankelijkheid ten opzichte van alternatieve bestemmingen en de mutatie in de woningvoorraad in de laatste jaren. In de vorige prognose werden enkel de eerste twee variabelen gebruikt. Het ruimtelijk interactie model is in de nieuwste prognose voor elke gemeente afzonderlijk geschat, terwijl in de vorige prognose het model op provinciaal niveau is geschat. De nieuwe specificatie van het ruimtelijk interactiemodel leidde tot een aanzienlijke verbetering van de fit tussen geschatte en waargenomen migratiestromen

1. Inleiding

In 2006 hebben het Planbureau voor de Leefomgeving (in dat jaar nog het voormalig Ruimtelijk Planbureau) en het Centraal Bureau voor de Statistiek voor het eerst een regionale (bevolkings- en huishoudens)prognose gepubliceerd. De regionale prognose is opgesteld met behulp van het prognosemodel PEARL, het acroniem voor Projecting population Events At Regional Level. Voor een uitgebreide beschrijving van de methode en veronderstellingen van deze prognose wordt verwezen naar De Jong e.a. (2005). Een belangrijke component van de regionale prognose vormt de binnenlandse migratie. De hoofdmoot van de binnenlandse migratie wordt gevormd door verhuizingen over korte afstand, vanwege woonmotieven en levensloopgebeurtenissen. Dit type verhuizingen wordt gekenmerkt door een aanzienlijk afstandsverval; ofwel de kans op een verhuizing wordt beduidend kleiner met het toenemen van de afstand waarover wordt verhuisd. In het zogenaamde ruimtelijk interactiemodel vormt het afstandsverval een centrale parameter en daardoor is dit model zeer geschikt voor de schatting van dit type verhuizingen. In De Jong (2005) is ingegaan op de schatting van de verhuizingen over korte afstand met behulp van dit model, in het kader van het opstellen van de eerste regionale prognose. Er werd toen nog gebruik gemaakt van twee verklarende variabelen, te weten het inwonertal van de gemeente van vestiging en de afstand tussen twee gemeenten, gemeten in 'vogelvlucht'. Bij de update van de regionale prognose in 2008 is de variabele afstand anders geoperationaliseerd, namelijk aan de hand van afstand via de weg. Dit leidde tot een aanzienlijke verbetering van de fit tussen waargenomen en geschatte korte afstand verhuizingen (Wijngaarden en De Jong, 2008). In de prognose van 2011 is het model opnieuw aangepast. Er zijn twee extra variabelen opgenomen, namelijk de zogenaamde toegankelijkheid van alternatieve bestemmingen en de mutatie in de woningvoorraad in de

¹ Rob Loke was tot november 2011 werkzaam bij het Planbureau voor de Leefomgeving en Andries de Jong is hier ook werkzaam.

laatste jaren. In de vorige prognose werd het model per provincie geschat; in de nieuwe prognose werd het model voor elke gemeente afzonderlijk geschat. Deze inhoudelijke en methodologische aanpassingen leidde tot een verdere verbetering van de schatting van de korte afstand migratie. In dit artikel wordt ingegaan op de nieuwe specificatie van het ruimtelijk interactie model en de uitkomsten hiervan. De modellering van de korte afstand migratie is echter een afzonderlijke stap in de modellering van verhuizingen in de regionale prognose. Om de rol van deze stap toe te lichten, wordt in de volgende paragraaf eerst ingegaan op alle stappen van het modeleren van verhuizingen.

2. Het modeleren van verhuizingen

Verhuizingen zijn een belangrijke oorzaak van bevolkingsgroei c.q. bevolkingskrimp van gemeenten en vormen daarmee een cruciale component in de regionale prognose van het PBL en het CBS. De regionale prognose wordt gemaakt met het model PEARL en hierin worden de verhuizingen in een aantal stappen gemodelleerd.

In de eerste stap wordt een schatting gemaakt van het totaal aantal personen dat verhuist (binnen de betreffende gemeente of naar een andere gemeente). In 2009 zijn er in Nederland bijna 1.5 miljoen personen verhuisd. Dit aantal is duidelijk lager dan in de jaren daarvoor, met 1,63 miljoen in 2007 en 1,55 miljoen in 2008. De daling van het aantal verhuizingen in de laatste drie jaar hangt waarschijnlijk samen met de economische crisis. Zowel de nieuwbouw als de verkoop van bestaande koopwoningen is gedaald en dit leidde tot minder verhuizingen. In de regionale prognose wordt eerst voor de hele prognose periode het totaal aantal verhuizingen voorspeld en vervolgens wordt het aantal verhuizingen voor elke afzonderlijke gemeente bepaald.

In de tweede stap wordt per gemeente bepaald, welk gedeelte van alle verhuizende personen binnen de eigen gemeente verhuist. Van alle Nederlanders verhuisde in de afgelopen drie jaar rond 60% binnen de huidige woongemeente. Van deze groep van verhuizende personen is de bestemmingsgemeente bij voorbaat bekend. Dit geldt vanzelfsprekend niet voor de overige verhuizende personen, die naar een andere gemeente verhuizen. Het percentage 'binnenverhuizingen' verschilt overigens vrij sterk per gemeente. Dit loopt uiteen van 80 % in Kampen naar 5 % in Rozendaal.

In de derde stap wordt per gemeente de personen die naar een andere gemeente verhuizen, onderverdeeld in twee groepen: migranten die over lange afstand verhuizen en migranten die over korte afstand verhuizen. De verhuismotieven verschillen namelijk aanzienlijk tussen deze twee groepen en dat leidt tot andere verhuispatronen. Om deze reden wordt het verhuizen van elk van de twee groepen verschillend gemodelleerd. Kenmerkend voor de langeafstandsmigratie is dat het in hoge mate een 'structureel' karakter draagt: specifieke bestemmingsgemeenten trekken veel meer migranten aan dan andere gemeenten en deze aantrekkingskracht is over de jaren vrij stabiel. Deze langdurige aantrekkingskracht komt onder andere voort uit de aanwezigheid van bepaalde onderwijsvoorzieningen, zoals universiteiten, hogescholen en meer specifieke opleidingsinstituten zoals politiescholen, marine- en legeropleidingsinstituten. Ook grote steden zijn vaak erg in trek, enerzijds vanwege het sociaal-culturele klimaat (vooral jongeren waarderen de mogelijkheden om uit te gaan) en anderzijds vanwege de economische potentie (veel mensen vinden hier een baan). Dit type verhuizingen kent, in tegenstelling tot de korteaftandsmigratie, over het algemeen geen duidelijk afstandsverval: bijvoorbeeld veel jongeren uit Friesland trekken naar de stad

Groningen vanwege de aanwezigheid van een universiteit, terwijl de hoofdstad Leeuwarden in veel gevallen dichterbij ligt. Om deze reden is dit type migratie niet zo geschikt om te modelleren met een ruimtelijk interactiemodel, waarin de afstand centraal staat. Het criterium om langeafstandsmigratie te onderscheiden van korteafstandsmigratie is in de eerste prognose (zie De Jong e.a. 2006) bepaald op een verhuisafstand van 35 kilometer hemelsbreed. Voor elke vertrekgemeente wordt vervolgens nagegaan welk aandeel van de langeafstandsmigranten naar specifieke bestemmingsgemeenten verhuizen. Dit bestemmingspatroon wordt in de regionale prognose op de aantallen lange afstandsmigranten vanuit elke vertrekgemeente toegepast. Overigens worden deze bestemmingpatronen voor vier leeftijdsgroepen afgeleid (en toegepast); deze leeftijdsgroepen omvatten bij benadering verschillende fasen in de levensloop. Voorts worden alle migranten in de leeftijdsklasse 17 tot en met 24 jaar (dus ook degenen die verhuizen over een afstand korter dan 35 kilometer) gemodelleerd aan de hand van een dergelijk bestemmingspatroon. Het gaat hierbij om jongeren die gaan studeren of een eerste baan oppakken; ook voor hen geldt dat afstandsverval niet een grote rol speelt en dat daardoor het ruimtelijk interactiemodel niet zo geschikt is voor de modellering van deze groep migranten.

In de vierde stap worden de verhuizingen over korte afstand gemodelleerd. Bij het verhuizen over korte afstand spelen woonmotieven en veranderingen in de levensloop een doorslaggevende rol. Uit Feijten en Visser (2005) blijkt dat het voor een groot deel gaat om doorstromers, die in een andere woning dan wel andere woonomgeving willen wonen (en hiermee doorgaans een opwaartse stap zetten in hun wooncarrière). Naast doorstromers gaat het bij dit type migranten ook om starters op de woningmarkt (zoals kinderen die zelfstandig gaan wonen), dan wel semi-starters (mensen die vanwege een scheiding genoodzaakt zijn woonruimte te vinden). Bij het verhuizen over korte afstand speelt het afstandsverval een sterke rol. Dit hangt ten eerste samen met de factor informatie: mensen krijgen eerder informatie over nabijgelegen nieuwbouwwoningen of vrijgekomen (oudbouw)woningen, dan over woningen die vrij ver weg liggen. Ten tweede geldt dat bestaande contracten beter in stand kunnen worden gehouden als de nieuwe woonplek dicht bij de oude ligt; dit betekent dat voor het –frequent- bezoeken van familie, vrienden en kennissen het prettig is als de reistijd niet (veel) langer wordt. Aangezien in het ruimtelijk interactiemodel de focus ligt op de afstandsrelatie is dit type model zeer geschikt om verhuizingen over korte afstand te modelleren. In de regionale prognose wordt om deze reden van dit model gebruik gemaakt voor het modelleren van de korteafstandsmigratie. Dit type model kent in het algemeen drie categorieën van parameters: de aantrekkelijkheid van de vertrekgemeente, de aantrekkelijkheid van de vestigingsgemeente en de afstand tussen vertrek- en vestigingsgemeente. Op basis van de hierboven beschreven stappen is in de regionale prognose echter bekend hoeveel migranten uit elke vertrekgemeente verhuizen, en hierdoor vervalt de categorie parameters die betrekking hebben op de aantrekkelijkheid van de vertrekgemeente. Hierdoor kan er in de regionale prognose gebruik worden gemaakt van een eenvoudiger operationalisering, waarbij slechts twee categorieën van parameters hoeven te worden meegenomen, namelijk de categorie parameters voor de aantrekkelijkheid van de vestigingsgemeenten en de parameter voor de afstand. De geschatte parameters van het ruimtelijk interactiemodel worden in de regionale prognose gebruikt om voor elk prognose jaar een initieel patroon van korte afstandmigratie te berekenen. Dit patroon geeft dan de -gewenste- korteafstandsmigratiestromen, in een situatie dat iedereen kan gaan wonen waar hij wil en er dus voldoende woningen beschikbaar zijn. In dit artikel wordt op deze stap in detail ingegaan.

In de vorige stap wordt er voorbij gegaan aan een essentiële voorwaarde voor het laten plaatsvinden van verhuizingen, namelijk de beschikbaarheid van woningen. In een situatie van regionale krapte aan woningen, kan dit er toe leiden dat huishoudens niet kunnen verhuizen naar de gewenste vestigingsgemeente (zoals bepaald door het ruimtelijk interactiemodel). Naast deze regionale tekorten aan woningen, kan er sprake zijn van regionale overschotten. Dit is vooral het geval indien in een bepaalde gemeente een nieuwbouwwijk wordt opgeleverd, waardoor het aanbod van woningen plotseling sterk toeneemt. Bij krapte op de regionale woningmarkt heeft dit tot gevolg dat veel korteaafstandsmigranten naar deze gemeente verhuizen, mede doordat deze woningen zich vaak kenmerken door een gunstige prijs/kwaliteitsverhouding. Om rekening te houden met deze woningmarktaspecten, wordt in een separate vijfde stap gekeken naar vraag naar en aanbod van woningen; hierbij wordt overigens uitgegaan van huishoudens in plaats van personen. De vraag naar woningen wordt uitgeoefend door immigranten, kinderen die uit huis gaan, paren die uit elkaar gaan en langeafstandsmigranten en korte afstandsmigranten die in een bepaalde gemeente een woning zoeken. Het aanbod van woningen komt tot stand door emigratie, sterfte, alleenstaanden die gaan samenwonen, langeafstandsmigranten en korteaafstandsmigranten die uit een bepaalde gemeente vertrekken, en ten slotte de netto toevoegingen aan de woningvoorraad. Via een relatief simpel woningmarktmodel wordt dan per gemeente de vraag naar woningen toegewezen aan het aanbod van woningen. Hierdoor worden dan de (gewenste) korteaafstandsmigratiestromen uit de vierde stap aangepast. In het kader van dit artikel wordt niet nader op deze stap ingegaan.

3. Het ruimtelijk interactiemodel

Er bestaan diverse modellen om migratiestromen te simuleren. Stillwell en Congdon (1991) onderscheiden micro en macro migratiemodellen. Micro modellen gaan in op de processen die ten grondslag liggen aan het besluit van een potentiële migrant om in de huidige woning te blijven wonen dan wel te gaan verhuizen. Macro modellen gaan in op (geaggregeerde) migratiestromen. Een vrij bekend en veel gebruikt macro model is het zogenaamde 'migrationstock' model (zie o.a. van Imhoff e.a., 1997). In dit model wordt eerst het totaal aantal migranten in een bepaald gebied (zoals een land) bepaald, waarna deze migranten via een toedelingschema worden toegekend aan regio's hierbinnen (zoals gemeenten). Het nadeel van dit model is dat migrantenstromen tussen de regio's niet bekend zijn. Een ander bekend model is het zogenaamde ruimtelijke interactiemodel. In dit model worden stromen migranten tussen de verschillende regio's wel gesimuleerd. Het ruimtelijk interactiemodel is gebaseerd op het zwaartekrachtmodel van Newton en bevat de elementen schaal (plaatsen met veel inwoners zullen meer verhuizingen genereren en aantrekken dan plaatsen met weinig inwoners) en afstand (hoe groter de afstand tussen twee plaatsen, hoe minder interacties tussen die plaatsen). In formule:

$$M_{ij}=k(P_i P_j)/D_{ij}^\beta \quad (1)$$

waarbij

M_{ij} = het aantal verhuisde personen tussen vertrekgemeente i en bestemmingsgemeente j ;

P_i = het aantal inwoners van vertrekgemeente i ;

P_j = het aantal inwoners van bestemmingsgemeente j ;

D_{ij} = de afstand tussen gemeenten i en j ;
 β = exponent die het 'verval' van het afstandseffect bepaalt;
 k = constante.

Het ruimtelijk interactie model kent drie typen parameters: parameter(s) die de (on) aantrekkelijkheid van de vertreklocatie beschrijven (ofwel de 'push' factoren), parameter(s) die de aantrekkelijkheid van de bestemminglocatie beschrijven (ofwel de 'pull' factoren) en een parameter die de interactie tussen vertrek- en bestemmingslocatie beschrijft (dit gebeurt meestal in de vorm van afstand). In de toepassing van de regionale prognose wordt echter met een bepaalde variant van het ruimtelijk interactiemodel gewerkt. De reden hiervoor is dat door de gekozen modellering van de binnenlandse migratie in een eerdere stap al bekend hoeveel migranten in een bepaalde vertrekgemeente betrokken zijn bij de migratie over korte afstand. Dit betekent dat in dit geval slechts de parameter(s) voor de aantrekkelijkheid van de vestiginggemeente en de parameter voor de afstandsfunctie hoeven te worden geschat. In de technische literatuur (zie Fotheringham en O'Kelly, 1989) staat dit model bekend als het zogenaamde 'production-constrained' model (de 'constraint' wordt daarbij gevormd door de bekende aantallen vertrekkers). In de vorige regionale prognose van 2009 (en ook die daarvoor) is een simpele versie van dit production-constrained model gebruikt. Deze versie is geformuleerd als

$$M_{ij} = O_i P_j^\alpha D_{ij}^\beta / \sum_k (P_k^\alpha D_{ik}^\beta) \quad (2)$$

waarbij

M_{ij} = het aantal verhuisde personen tussen vertrekgemeente i en bestemmingsgemeente j ;
 O_i = het totaal aantal verhuisde personen vanuit vertrekgemeente i ;
 P_k = het aantal inwoners van bestemmingsgemeente j ;
 D_{ij} = de afstand tussen gemeenten i en j ;
 α = exponent die effect van inwonertal van bestemmingsgemeente j bepaalt;
 β = exponent die het 'verval' van het afstandseffect bepaalt;
 $\sum_k (P_k^\alpha D_{ik}^\beta)$ = normalisatiefactor die verwachte en waargenomen aantallen verhuisde personen aan elkaar gelijk maakt.

Zoals hierboven al vermeld, wordt het ruimtelijk interactiemodel enkel toegepast voor de korteaftandsmigratie. In de eerste editie van de regionale prognose, die in 2006 is gepubliceerd, is als criterium voor korte afstandsmigratie gekozen voor de grens van 35 km: mensen die binnen een straal van 35 km verhuizen worden gerekend tot de korteaftandsmigranten (en die over een grotere afstand verhuizen tot de langeaftandsmigranten). De Jong (2005) gebruikte bij de operationalisering van afstand de hemelsbrede afstanden tussen gemeenten, waarbij voor elke gemeente het 'zwaartepunt' is bepaald met de bijbehorende vierkantscode-coördinaten. Aan deze operationalisering kleefde echter een groot bezwaar. De afstand over de weg (hierna ook reisafstand genoemd) tussen twee gemeenten kan namelijk aanzienlijk verschillen van de hemelsbrede afstand. Deze situatie doet zich bijvoorbeeld voor als er water overgestoken moet worden: er moet dan worden omgereden voor een brug- of bootverbinding. Dit speelt niet alleen op de Zeeuwse eilanden en de Waddeneilanden, maar ook voor gemeenten aan weerszijden van rivieren en kanalen. Een groot verschil tussen hemelsbrede afstand en reisafstand kan de bereidheid om te

verhuizen doen afnemen, bijvoorbeeld omdat iemand contact wil blijven houden met personen of instellingen in de oude woonplaats. In de tweede regionale prognose van 2008 is daarom in het ruimtelijk interactiemodel de reisafstand gebruikt (Wijngaarden en de Jong, 2008). De reisafstanden zijn berekend als gewogen afstanden over de weg tussen gemeenten; de wegingsfactor is het inwonertal van viercijferige postcodegebieden (PC4-gebieden). Het zwaartepunt van een gemeente verschuift hierdoor naar PC4-gebieden met veel inwoners en dus veel potentiële verhuizers. Het is door de weging met inwonertallen mogelijk dat een reisafstand *korter* wordt dan een hemelsbrede afstand, bijvoorbeeld als twee gemeenten met dichtbevolkte PC4-gebieden naar elkaar toe liggen. De nieuwe operationalisering via reisafstanden, leverde een belangrijke verbetering van de fit van het ruimtelijk interactie model op.

In de vorige drie prognoses is het ruimtelijk interactie model voor elke provincie apart geschat. Hiertoe zijn voor een bepaalde provincie alle korte afstandsmigratiestromen vanuit alle (vertrek)gemeenten in die betreffende provincie meegenomen; deze stromen kunnen zowel betrekking hebben op bestemmingsgemeenten binnen de betreffende provincie als op bestemmingsgemeenten over de provinciegrens heen. Een belangrijke trend in het ruimtelijk modelleren is echter het exploreren van 'lokale' variaties in ruimtelijke relaties (Lloyd, 2007; Tate and Atkinson, 2001; Fotheringham e.a., 2002, Hu and Pooler, 2002). Door de overgang van 'globale' modellering naar 'lokale' modellering, is het mogelijk om specifieke regio's (zoals gemeenten) die sterk afwijken in hun ruimtelijke relaties, beter te modelleren. Gezien de mogelijke winst in voorspelkracht is onderzocht of het ruimtelijk interactie model in de update van de regionale prognose beter op het niveau van individuele gemeenten kan worden geschat. In paragraaf 5 wordt hier nader op ingegaan.

Ten behoeve van de update van de regionale prognose is voorts onderzocht of de voorspelkracht van het ruimtelijk interactie model verder kan worden verbeterd door het opnemen van meer variabelen in het model. Uit Fotheringham en Kelly (1989) blijkt namelijk dat de fit van het ruimtelijk interactiemodel belangrijk kan verbeteren door de opname van variabele die de zogenaamde 'toegankelijkheid' (in het Engels 'accessibility') van een bepaalde vestigingsgemeente ten opzichte van nabijgelegen gemeenten weergeeft. In geval van een groot aantal alternatieven is de mens namelijk niet in staat om alle mogelijke alternatieven te evalueren. Op basis van theoretische overwegingen, met betrekking tot de wijze waarop mensen informatie verwerken, veronderstellen Fotheringham en Kelly dat de toegankelijkheid ten opzichte van alternatieve bestemmingen een belangrijke rol speelt bij de keuze van een specifieke bestemming bij een verhuizing. Indien een bepaalde aantrekkelijke bestemming dicht ligt bij andere aantrekkelijke bestemmingen, dan ondervindt deze bestemming hiervan veel concurrentie en wordt de kans kleiner dat deze bestemming wordt gekozen. Fotheringham en Kelly adviseren om deze reden in het ruimtelijk interactiemodel een variabele op te nemen, die de concurrentie van alternatieve aantrekkelijke bestemmingen in de nabijheid van een bepaalde bestemming weergeeft. Het is daarbij gebruikelijk het inwonertal te gebruiken als proxy van de aantrekkelijkheid en de nabijheid te operationaliseren aan de hand van afstand. Fotheringham en Kelly benoemen deze variabele 'centrality', in navolging van hen wordt de variabele toegankelijkheid berekend aan de hand van de volgende formule:

$$C_{ij} = \sum_{k=1; k \neq i; k \neq j} P_k / D_{jk} \quad (3)$$

waarbij

C_{ij} = toegankelijkheid van bestemmingsgemeente j voor vertrekgemeente i ;

In deze formule is gewerkt met hemelsbrede afstand en loopt de sommatie ten behoeve van bestemmingsgemeente j over alle gemeenten (die betrokken zijn bij korteafstandsmigratie), behalve de vertrek gemeente i .

Stillwell en Congdon (1991) geven aan dat het meenemen van woningbouw ook tot een verbetering van de schatting van de korte afstandsmigratie kan leiden. Indien in een bepaalde gemeente relatief veel wordt gebouwd, dan kan dit de aantrekkingskracht van die gemeente vergroten en hierdoor leiden tot grotere migratiestromen naar deze gemeente.

Ten behoeve van de nieuwe regionale prognose is onderzocht of de opname van deze twee extra variabelen leidt tot een grotere voorspelkracht. Door de uitbreiding van het model wordt de formule voor de schatting van het ruimtelijk interactiemodel (volgens de 'production constrained' variant) als volgt:

$$M_{ij} = O_i P_j^{\alpha} D_{ij}^{\beta} C_{ij}^{\gamma} H_j^{\delta} / \sum_k (P_k^{\alpha} D_{ik}^{\beta} C_{ik}^{\gamma} H_k^{\delta}) \quad (4)$$

waarbij:

M_{ij} = het aantal verhuisde personen tussen vertrekgemeente i en bestemming j ;

O_i = het totaal aantal verhuisde personen vanuit vertrekgemeente i ;

P_j = het aantal inwoners van bestemmingsgemeente j ;

D_{ij} = de afstand tussen gemeenten i en j ;

C_{ij} = toegankelijkheid van bestemmingsgemeente j voor vertrekgemeente i ;

H_j = netto toevoegingen aan de woningvoorraad in bestemmingsgemeente j ;

α = exponent die effect van inwonertal van bestemmingsgemeente j bepaalt;

β = exponent die het 'verval' van het afstandseffect bepaalt;

γ = exponent die het effect van centrale ligging bepaalt;

δ = exponent die het effect van netto toevoegingen aan de woningvoorraad bepaalt;

$1 / \sum_k (P_k^{\alpha} D_{ik}^{\beta} C_{ik}^{\gamma} H_k^{\delta})$ = normalisatiefactor die verwachte en waargenomen aantallen verhuisde personen aan elkaar gelijk maakt.

In deze formule wordt gewerkt met 'machtsfuncties', aangezien de te schatten parameters als macht van inputvariabelen zijn opgenomen. Het voordeel van deze machtsfuncties is dat de schattingen van de parameters onafhankelijk zijn van de schaal waarin ze worden gemeten en dat hierdoor de schattingen afkomstig uit verschillende prognoses met elkaar kunnen worden vergeleken.

Merk op dat door de overgang van 'globale' op 'lokale' modellering deze specificatie van het ruimtelijk interactiemodel voor elke gemeente afzonderlijk dient te worden uitgevoerd. Dit betekent dat voor elke gemeente afzonderlijk de vier parameters uit formule 4 worden geschat. In de vorige versie van de regionale prognose werden de (twee) parameters (voor het effect van inwonertal van de bestemmingsgemeente en de afstand tussen vestiging en vertrekgemeente) nog voor elke provincie afzonderlijk geschat.

Om formule 4 via lineaire regressie (Ordinary Least Square) te kunnen schatten, dient deze nog lineair te worden gemaakt; dit is als volgt gedaan:

$$\ln M_{ij} - (1/n) \sum_j \ln M_{ij} = \alpha (\ln P_j - (1/n) \sum_j \ln P_j) + \beta (\ln D_{ij} - (1/n) \sum_j \ln D_{ij}) +$$

$$\gamma(\ln C_{ij} - (1/n) \sum_j \ln C_{ij}) + \delta(\ln H_j - (1/n) \sum_j \ln H_j), \quad (5)$$

Hierbij staat \ln voor het natuurlijk logaritme en n voor het aantal bestemmingsgemeenten. Merk op dat bij de schatting van deze regressievergelijking geen gebruik wordt gemaakt van een constante. Bij de schatting van de parameters aan de hand van formule 5 kan overigens het probleem opduiken dat bepaalde verklarende variabelen nul zijn en dan kan het natuurlijk logaritme niet worden bepaald. Tussen kleine gemeenten bestaat er vaak in bepaalde jaren geen migratieverkeer. In geval het aantal migranten tussen vertrekgemeente i en aankomstgemeente j (M_{ij}) nul bedraagt, is de waarde 1 gesubstitueerd. Ook de woningbouw in een bepaalde bestemmingsgemeente (H_j) kan in een bepaalde periode nul zijn of negatief (de onttrekkingen overtreffen dan de toevoegingen); ook in dat geval is de waarde 1 gesubstitueerd.

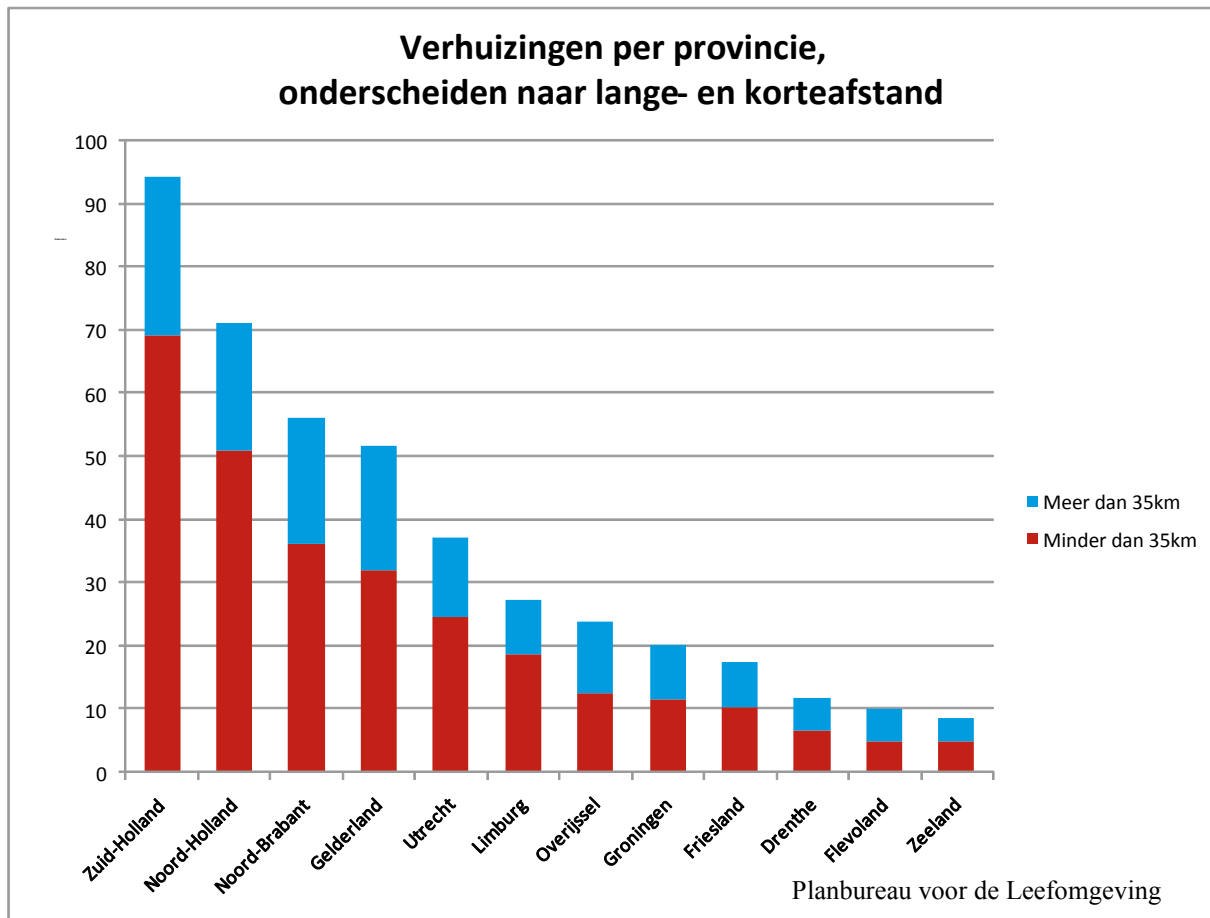
Merk op dat in het getransformeerde model voor de te verklaren variabele het natuurlijk logaritme van het aantal migranten wordt gebruikt in plaats van het werkelijke aantal migranten. Dit heeft tot gevolg dat de verklaarde variantie van het getransformeerde lineaire regressiemodel aangeeft in hoeverre de getransformeerde aantallen migranten worden verklaard, en dat is dus niet gelijk aan de werkelijke aantallen migranten. Hierdoor is de gebruikelijke maat om de voorspelkracht van het model te bepalen, te weten de verklaarde variantie, minder geschikt om na te gaan hoe goed het model de werkelijke migratiestromen weergeeft. Om deze reden wordt gebruikt gemaakt van andere maten die een indicatie geven van de voorspelkracht. Paragraaf 5 gaat hier nader op in.

4. Beschrijving van de data

Voor de periode 2006-2008 zijn gegevens verzameld over migratiestromen, inwoneraantallen en nieuwbouw van woningen; deze zijn ontleend aan de database StatLine van het CBS. Zowel voor de migratiestromen als de inwoneraantallen is het gemiddelde over de drie jaren berekend. Op deze wijze kon gecorrigeerd worden voor toevalsfluctuaties, die vooral bij migratiestromen tussen kleine gemeenten optreden. De gegevens over de reisafstanden zijn ontleend aan NAVTEQ (*verwijzing invoegen*).

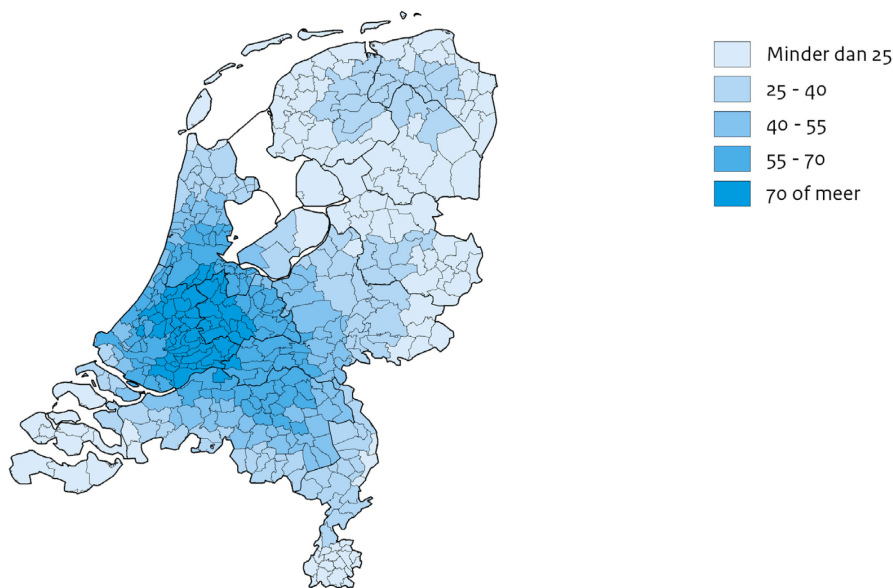
In het ruimtelijk interactie model worden korteaafstandsmigratiestromen tussen vertrek- en aankomstgemeenten verklaard. Hierbij is de 'gemeentelijst' van het kalenderjaar 2010 gebruikt. Om korte en lange afstandsmigratie te onderscheiden, wordt gebruikt gemaakt van hemelsbrede afstanden, waarbij de afstand van 35 kilometer als criterium om te splitsen is gebruikt. Op basis van deze afbakening blijkt dat het merendeel van de migratiestromen gecategoriseerd kan worden als korteaafstandsmigratie: ongeveer twee derde van het verhuisverkeer tussen de gemeenten in Nederland bestaat hieruit. In *figuur 1* is per provincie de verhouding tussen korte- en langeafstandsmigratiestromen weergegeven. In absolute aantallen zijn de korteaafstandsmigratiestromen veruit het grootst in Zuid-Holland, op enige afstand gevolgd door Noord-Holland. In deze twee provincies bestaat de migratie voor rond 70 % uit korteaafstandsmigranten. In Zeeland komt het aantal korteaafstandsmigranten het laagst uit en in Flevoland het op één na laagst. Ook het aandeel korteaafstandsmigranten ligt in deze twee provincies laag, met respectievelijk rond 55% en 50%.

Figuur 1: Migranten per provincie, onderscheiden naar lange- en korteaftandsmigranten.



Het aantal bestemmingsgemeenten dat betrokken is bij korteaftandsmigratie kan voor individuele vertrekgemeenten regionaal beduidend verschillen. In *kaart 1* is voor alle (vertrek)gemeenten in Nederland het aantal bestemmingsgemeenten weergegeven. Er valt hierin een duidelijk patroon waar te nemen: voor gemeenten gelegen in het gebied dat globaal overeenkomt met het 'hart' van de Randstad (gezien de ruimte die wordt opgespannen door de vier grote gemeenten) is het aantal bestemmingsgemeenten (met 70 tot 90) veel groter dan voor gemeenten aan de randen van Nederland (met 3 tot 20). Voor gemeenten die grenzen aan Duitsland en België komt de lage score deels door het feit dat ze ook migratieverkeer hebben met gemeenten over de grens (die niet worden meegenomen bij de berekening van het aantal bestemmingsgemeenten). Echter, de lage score geldt ook voor gemeenten aan randen van Friesland, Groningen en Zeeland die aan de zee gelegen zijn.

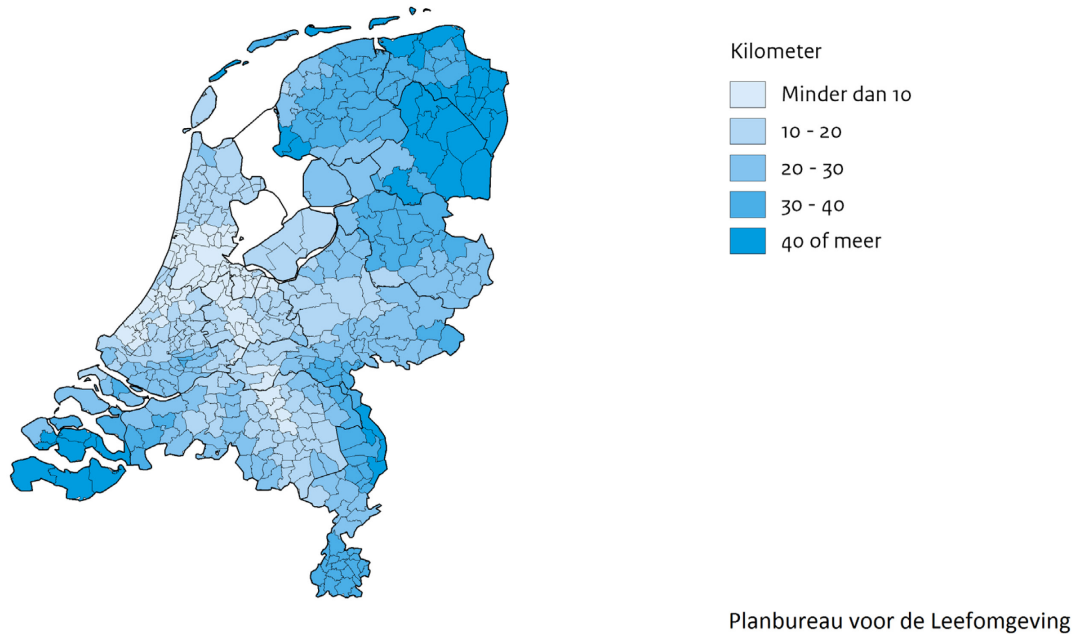
Kaart 1: Aantal migratiebestemmingen per (vertrek)gemeente



Planbureau voor de Leefomgeving

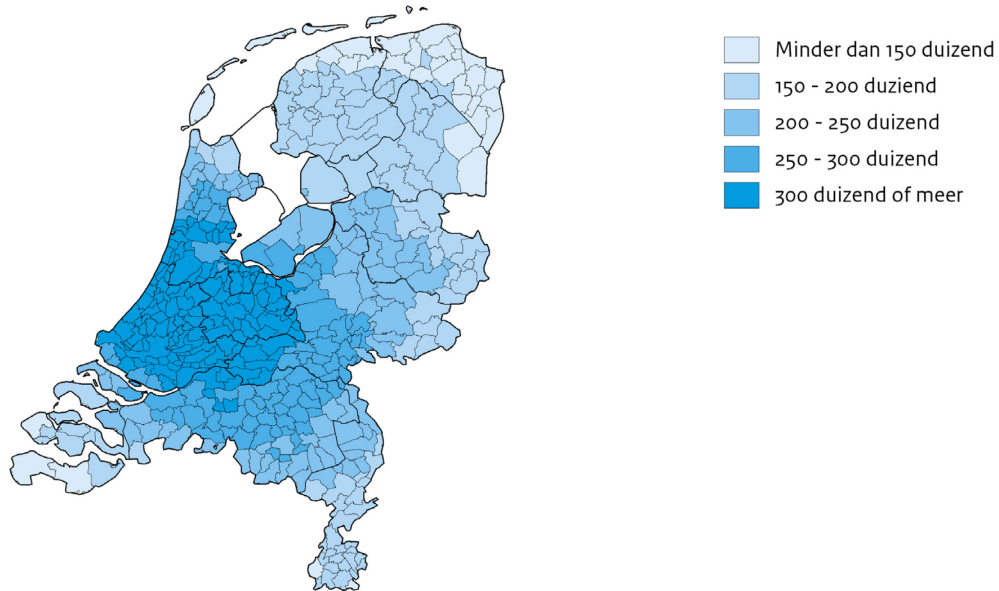
Hierboven is al aangegeven dat in het ruimtelijk interactiemodel ten behoeve van de operationalisering van de variabele afstand, reisafstanden in plaats van hemelsbrede afstanden worden gebruikt. *Kaart 2* geeft een indruk van het verschil tussen reisafstanden en hemelsbrede afstanden, uitgaande van de gemeente Amsterdam. Zoals hierboven al vermeld, is de afstand via de weg meestal langer als er sprake is van rivieren of meren op de (hemelsbrede) route, omdat er dan een omweg moet worden gemaakt om gebruik te kunnen maken van bruggen of tunnels. Ook kunnen wegen niet als een rechte lijn tussen twee postcodegebieden zijn aangelegd, vooral in geval van woongebieden, bedrijventerreinen of natuurgebieden. Uit kaart 2 blijkt dat in een cirkel van zo'n 20 kilometer rondom Amsterdam, de afwijking tussen reisafstand en hemelsbrede afstand het kleinst is. Dit komt ten dele door de vrij korte afstand tussen Amsterdam en de bestemmingen binnen deze cirkel, en ten dele door de fijnmazigheid van het wegennet. Voor steeds groter wordende cirkels van bestemmingen, valt echter op dat de afwijking tussen reisafstand en hemelsbrede afstand binnen deze cirkels niet meer gelijk is. In het algemeen wordt de afwijking groter bij langere reisafstanden vanaf Amsterdam, maar er zijn duidelijke stroken zichtbaar waar de afwijking aanzienlijk kleiner zijn. Zo is er een vrij smalle streek die diep naar het zuidoosten gaat, waar de afwijking veel kleiner is dan in de gemeenten gelegen naast deze streek. Deze lint lijkt vrij sterk overeen te komen met het traject van de autosnelweg A2. Min of meer hetzelfde geldt voor gemeenten gelegen aan de A1 en de A4, gezien de stroken die respectievelijk oostwaarts en zuidwaarts vanaf Amsterdam goed zichtbaar zijn.

Kaart 2: Verschil tussen afstand hemelsbreed en afstand via de weg voor bestemmingsgemeenten vanuit vertrekgemeente Amsterdam



In *kaart 3* is de score op de variabele toegankelijkheid van bestemmingsgemeenten weergegeven, berekend voor de vertrekgemeente Appingedam (een kleine gemeente in het noorden van de provincie Groningen). In de kaart komt duidelijk naar voren dat het aantal alternatieven voor een bestemmingsgemeente in het hart van de Randstad veel groter is dan aan de randen van Nederland. In het gebied dat opgespannen wordt door de vier grote gemeenten zijn er namelijk veel gemeenten met groot inwonertal, zodat de concurrentie tussen bestemmingsgemeenten hier veel groter is dan in de periferie van Nederland.

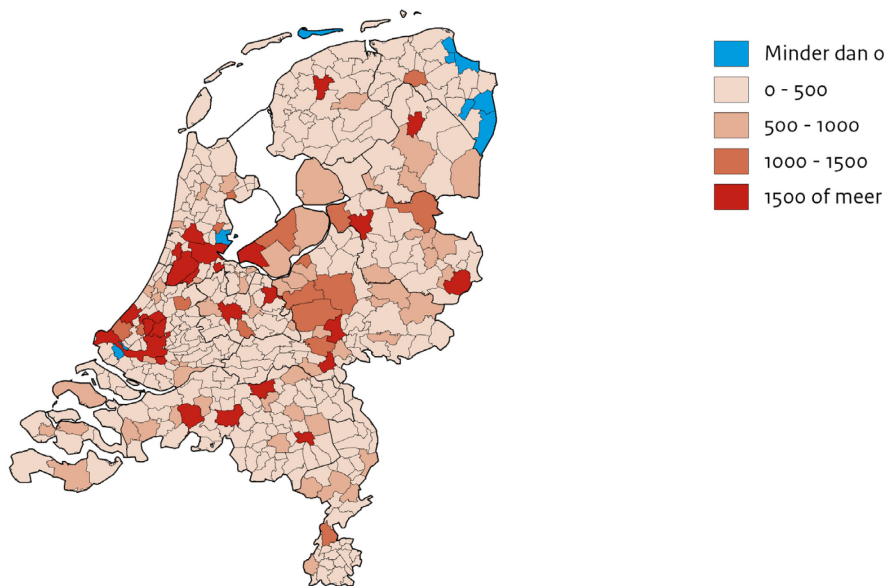
Kaart 3: Patroon van toegankelijkheid van bestemmingsgemeenten voor de vertrekgemeente Appingedam



Planbureau voor de Leefomgeving

In *kaart 4* is de (netto) woningbouw per gemeente afgebeeld; dit is berekend door de woningvoorraad op 1 januari 2006 af te trekken van de woningvoorraad op 1 januari 2009. De (netto) woningbouw wordt voornamelijk bepaald door nieuwbouw van woningen en ten dele ook door het saldo van de posten 'toevoegingen anderszins', onttrekkingen en administratieve correcties. Overigens valt de (netto) woningbouw in enkele gemeenten in het noordoosten van de provincie Groningen negatief uit: hier wordt meer gesloopt dan er wordt bijgebouwd. Diverse grotere gemeenten lichten duidelijk op in de kaart; hier is in de afgelopen jaren een aanzienlijk aantal huizen aan de woningvoorraad toegevoegd. Dit hangt onder meer samen met het beleid van de grotere gemeenten om (startende) gezinnen te behouden o.a. door het bouwen van VINEX wijken en hierdoor een halt toe te roepen aan de suburbanisatie naar randgemeenten, waar in de laatste decennia relatief veel eengezinswoningen werden gebouwd.

Kaart 4: Netto uitbreiding van de woningvoorraad tussen 1 januari 2009 en 1 januari 2006, per gemeente



Planbureau voor de Leefomgeving

5. Vergelijking van verschillende specificaties van het ruimtelijk interactiemodel

In paragraaf 3 is aangegeven dat ten behoeve van de nieuwe regionale prognose, onderzoek is gedaan naar hoe het ruimtelijk interactie model (verder) kan worden verbeterd. Hiertoe is enerzijds nagegaan of het schatten van het model voor afzonderlijke (vertrek)gemeenten in plaats van provincies kan leiden tot een verbetering van de voorspelkracht. Anderzijds is gekeken of het toevoegen van extra verklarende variabelen, te weten de zogenaamde toegankelijkheid ten opzichte van alternatieve bestemmingen en de netto woningbouw, ook tot een betere fit van de voorspelling aan de waarneming kan leiden.

Om te kunnen nagaan of deze wijzigingen inderdaad leiden tot een betere fit, is het nodig een indicator te hebben voor de voorspelkracht van de verschillende modellen. Hiertoe wordt gebruik gemaakt van twee indicatoren. In navolging van Fotheringham en Kelly (1989) wordt ten eerste gebruikt gemaakt van de zogenaamde SRMSE (de afkorting van de Engelse termen 'Standardized Root Mean Square Error'). Deze is als volgt gedefinieerd:

$$SRMSE = (1/M) [\sum_i \sum_j (M'_{ij} - M_{ij})^2 / n],$$

waarbij

M = totaal van de geobserveerde migratiestromen

M_{ij} = geobserveerde migratiestroom vanuit vertrekgemeente i naar bestemmingsgemeente j .

M'_{ij} = voorspelde migratiestroom vanuit vertrekgemeente i naar bestemmingsgemeente j .

n = aantal (geobserveerde) migratiestromen.

Hoe lager de waarde van de SRMSE uitvalt, hoe beter de voorspelling.

Als tweede indicator voor de voorspelkracht wordt gebruikt gemaakt van de zogenaamde PM (wat staat voor 'Percentage Misallocated'). Deze maat is gedefinieerd als in Hu en Pooler (2002):

$$PM = (50/M) \sum_i \sum_j \text{abs}(M'_{ij} - M_{ij}),$$

waarbij 'abs' staat voor de absolute waarde. Deze maat loopt van 0 (dat een perfecte fit aangeeft) tot 100.

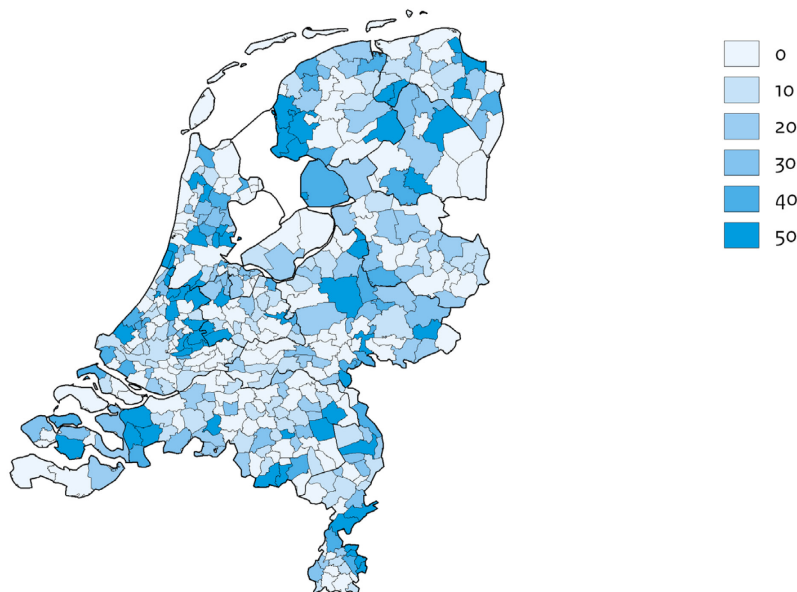
Het voordeel van beide maten is dat ze de voorspelde migratiestromen in de 'echte' ruimte evalueren en niet in de 'getransformeerde' ruimte, zoals de maat 'verklaarde variantie' die doorgaans wordt gebruikt om de voorspelkracht van een regressiemodel te beoordelen. De verklaarde variantie kan misleidend zijn, aangezien het model werkelijke migratiestromen dient te voorspellen en niet getransformeerde migratiestromen.

In de vorige drie regionale prognoses is het ruimtelijk interactie model op het niveau van de afzonderlijke provincies geschat. Ten behoeve van de nieuwe regionale prognose is onderzocht of het model beter op het niveau van individuele gemeenten kan worden geschat. Dit brengt echter wel het vraagstuk met zich mee, dat er moet worden nagegaan welke (regionale) migratiestromen precies bij de schatting moeten worden meegenomen. In de wetenschappelijke literatuur wordt dit aangeduid met de keuze van de zogenaamde 'korrelgrootte' (ofwel met de Engelse termen 'kernel size'). Deze korrelgrootte heeft dan betrekking op de straal rondom een bepaalde vertrekgemeente; bij de schatting van het ruimtelijk interactiemodel wordt dan van alle gemeenten die binnen de betreffende straal liggen alle korteaafstandsmigratiestromen meegenomen.

Bij de schatting van het ruimtelijk interactiemodel voor elke afzonderlijke vertrekgemeente ligt het voor de hand om alleen te kijken naar de migratiestromen tussen de betreffende vertrekgemeente en al haar bestemmingsgemeenten. In dat geval bedraagt de korrelgrootte nul. Uit kaart 1 blijkt echter dat in dat geval voor vrij veel gemeenten maar een relatief gering aantal bestemmingsgemeenten kan worden meegenomen. Dergelijke gemeenten zijn vooral gelegen aan de rand van Nederland; bovendien gaat het dan meestal om migratiestromen die uit een gering aantal personen bestaan. Dit kan resulteren in een lage verklaarde variantie en onbetrouwbare schattingen van de parameters van het model. In navolging van Fotheringham e.a. (2002) kan de schatting worden verbeterd door de korrelgrootte te verhogen. In de onderhavige analyse wordt hiertoe de korrelgrootte in vijf stappen van 10 kilometer (reisafstand) verhoogd. In de eerste stap van 10 kilometer worden dan bij de schatting van het ruimtelijk interactiemodel alle korteaafstandsmigratiestromen meegenomen die horen bij de (extra) vertrekgemeenten gelegen in een straal van 10 kilometer rondom de betreffende vertrekgemeente, waarvoor de parameters van het ruimtelijk interactiemodel dienen te worden bepaald. Bij een korrelgrootte van 20 kilometer worden dan alle gemeenten in een straal van 20 kilometer meegenomen, etc. Bij elke stap kan worden nagegaan hoe groot de voorspellingsfout is (aan de hand van de maat SRMSE). De korrelgrootte waarbij de voorspelfout het kleinst is, kan dan het beste worden gebruikt voor de schatting van de gemeentelijke parameters. Een nadeel van een variërende korrelgrootte per gemeente is overigens wel bij een andere korrelgrootte de parameters en statistische toetsen anders uitvallen; in de wetenschappelijke literatuur wordt dit aangeduid met het 'Modifiable Area Unit Problem' (zie Tate en Atkinson, 2001, en Fotheringham e.a., 2002). Het grote voordeel van een schatting op gemeentelijk niveau in plaats van op provinciaal niveau (zoals in de vorige regionale prognose is gedaan), is dat er nu meer recht wordt gedaan aan lokale patronen in het migratieverkeer.

Kaart 5 geeft voor alle (vertrek) gemeenten in Nederland een indruk van de korrelgrootte die leidt tot de laagste voorspelfout (waarbij het model met vier verklarende variabelen is gebruikt). Hieruit blijkt dat voor een meerderheid van de gemeenten een korrelgrootte van nul tot 10 kilometer voldoende is voor een adequate schatting van het ruimtelijk interactiemodel. Toch is er ook een aanzienlijk aantal gemeenten waarvoor het ruimtelijk interactiemodel met een korrelgrootte van 40 tot 50 kilometer dient te worden geschat. Hierin valt overigens geen duidelijk ruimtelijk patroon te ontdekken, hoewel het vrij vaak nodig is voor gemeenten aan de rand van Nederland.

Kaart 5: 'Korrelgrootte' dat de kleinste voorspelfout oplevert.



Planbureau voor de Leefomgeving

Op basis van de twee foutmaten SRMSE en PM, is nagegaan in hoeverre het schatten op gemeentelijk niveau in plaats van op provinciaal niveau leidt tot een betere voorspelling. Ook is ook nagegaan of het toevoegen van extra verklarende variabelen leidt tot een betere fit van het ruimtelijk interactie model. In *figuur 2* worden de twee maten SRMSE en PM per provincie getoond voor verschillende specificaties van het model. Wat betreft het aantal verklarende variabelen zijn er drie specificaties mogelijk: een model met twee verklarende variabelen (te weten het inwonertal van de vestigingsgemeenten en de afstand tussen vertrek- en bestemmingsgemeente), drie verklarende variabelen (met toevoeging van de variabele toegankelijkheid) en vier verklarende variabelen (met toevoeging van de variabele netto woningbouw). Bij de schatting van elk deze drie modellen kan er gekozen worden tussen een schatting op provinciaal niveau of op gemeentelijk niveau. Bij de schatting op het gemeentelijk niveau is er nog een extra model geschat: voor elke gemeente is gekeken welk aantal verklarende variabele de kleinste voorspelfout (op basis van de maat SRMSE) oplevert en vervolgens is dat aantal variabelen gebruikt.

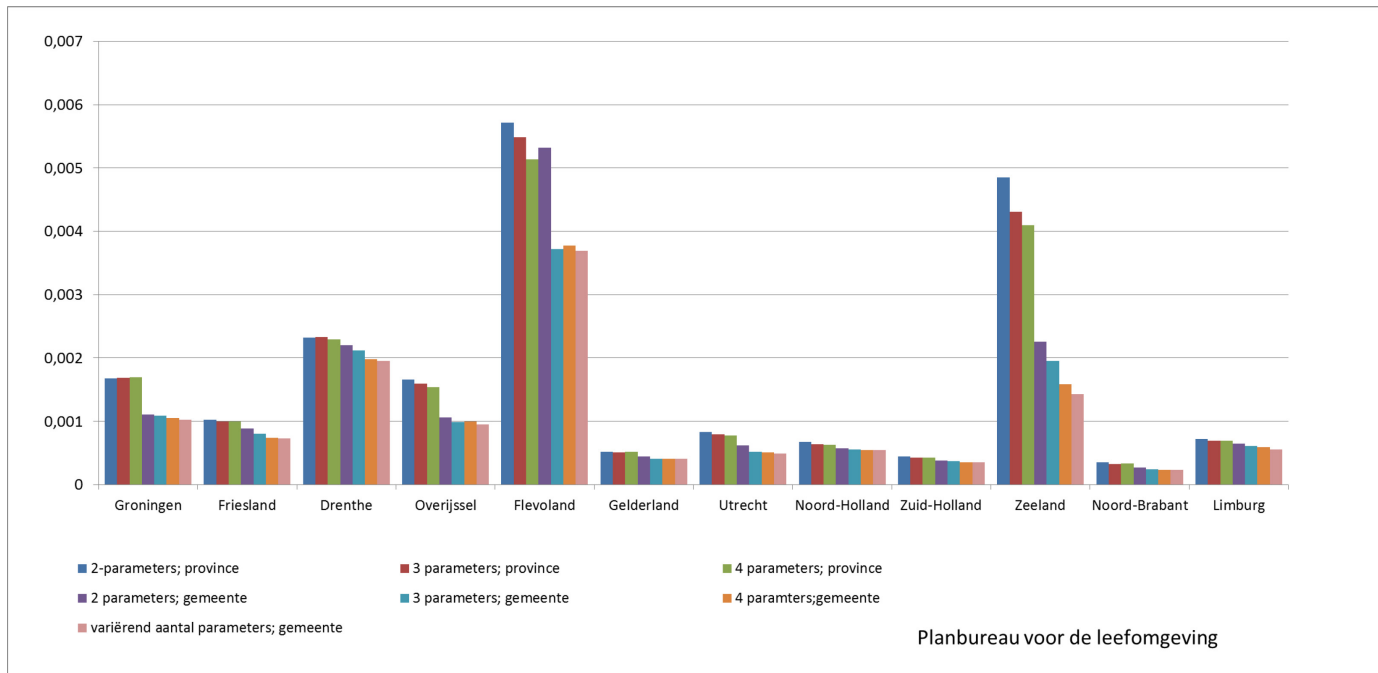
Figuur 2 laat zien dat de modellen die op gemeentelijk niveau zijn gespecificeerd, over het algemeen kleinere voorspelfouten hebben dan de modellen die provinciaal niveau zijn gespecificeerd. Op basis van de maat SRMSE, is de grootste reductie in voorspelfouten zichtbaar in de provincies Flevoland, Zeeland, Groningen en Overijssel. Indien gekeken wordt naar de maat PM, dan is de reductie van de voorspelfouten het grootst in Zeeland, Zuid-

Holland, Overijssel en Utrecht. Op basis van beide maten kan geconcludeerd worden dat het specificeren van het ruimtelijk interactie model op gemeentelijk niveau een grotere voorspelkracht oplevert.

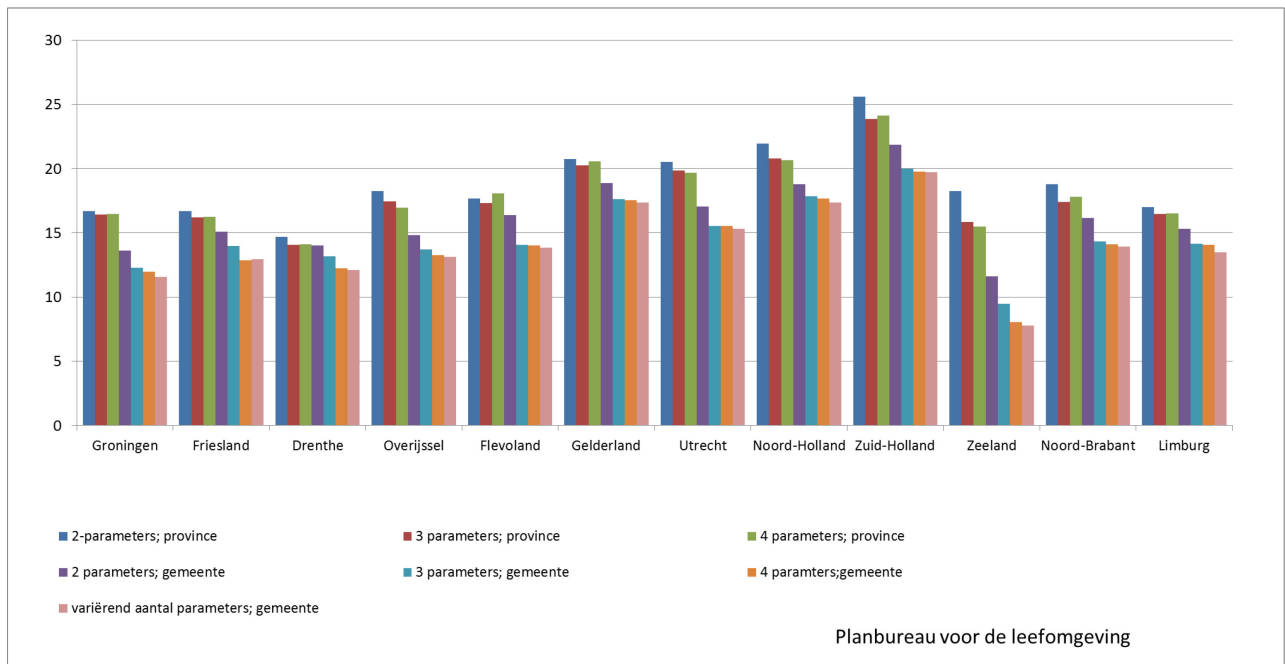
Wat betreft het aantal variabelen dat in het ruimtelijk interactiemodel is opgenomen, blijkt dat het model met drie of vier verklarende variabelen in een beperkt aantal provincies een beduidend grotere voorspelkracht heeft dan het model met twee verklarende variabelen. Dit geldt zowel voor de specificatie op provinciaal als op gemeentelijk niveau. In de gemeentelijke specificatie blijkt op basis van de maat SRMSE, dat vooral voor Flevoland het model met drie verklarende variabelen (dus met de toevoeging van de variabele toegankelijkheid) een beduidend kleinere voorspelfout oplevert. Op basis van de maat PM is in de meeste andere provincies ook een duidelijk winst in voorspelkracht zichtbaar, indien meer dan twee verklarende variabelen worden gebruikt. Het model met een wisselend aantal verklarende variabelen levert, zoals te verwachten was, de geringste voorspelfout op.

Figuur 2: Voorspelfout volgens de maten SRMSE en PM op basis van diverse specificaties van het Ruimtelijk interactiemodel

SRMSE per provincie



PM per provincie

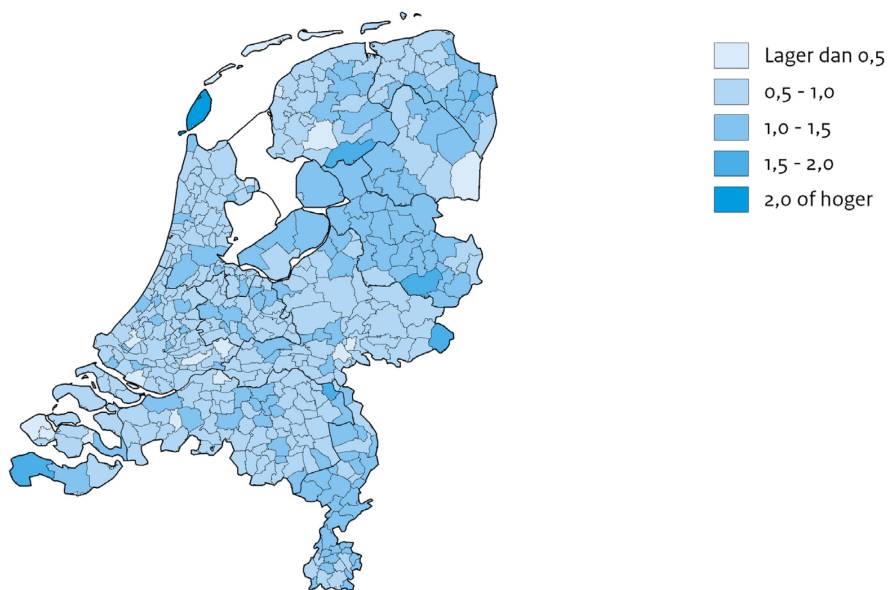


Op basis van de kaarten 6 tot en met 9 kan een beeld worden verkregen van de hoogte van elk van de geschatte parameters; hiertoe is het model gebruikt met vier verklarende variabelen dat per gemeente is geschat.

De geschatte gemeentelijke parameters van de variabele inwonertal van vestigingsgemeenten zijn, op een paar uitzonderingen na, allen positief (zie kaart 6). Dit betekent dat naar mate een vestigingsgemeenten meer inwoners heeft, hier meer migranten naar toe zullen verhuizen. Hoewel niet duidelijk zichtbaar, lijkt het dat voor gemeenten gelegen buiten de Randstad de 'pull factor' inwonertal van vestigingsgemeenten wat sterker is. Zo tonen gemeenten gelegen in het oosten van Groningen, grote stukken van Overijssel, midden Limburg en Zeeuws Vlaanderen verhoogde waarden voor deze parameter.

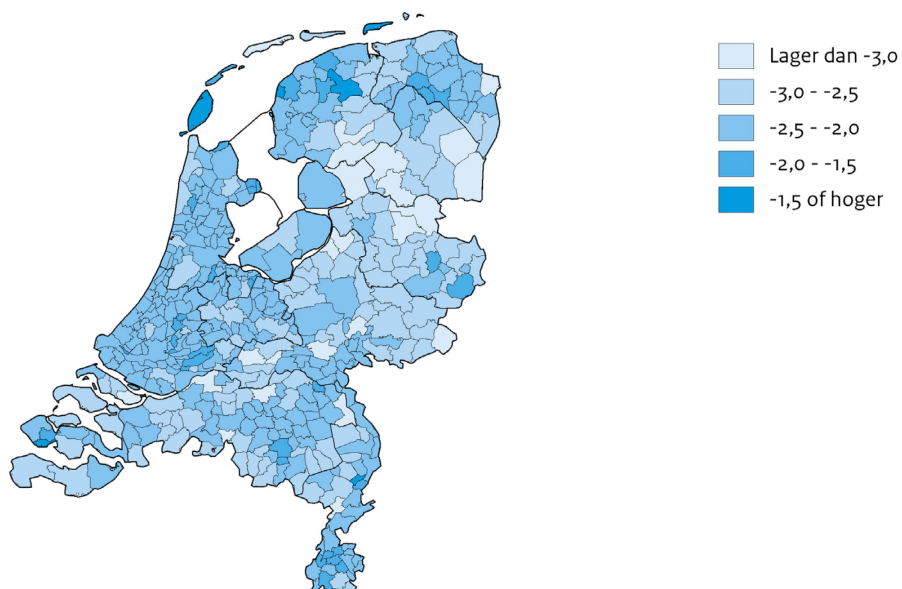
De parameter waarden van de variabele (reis)afstand zijn allen negatief (zie kaart 7). Deze negatieve waarden geven aan dat hoe groter de reisafstand wordt, hoe kleiner het aantal migranten wordt. In het kaartbeeld valt op dat buiten de Randstad de geschatte parameters over het algemeen sterker negatief zijn. Dit betekent dat de factor afstand hier een grotere rol lijkt te spelen bij het verhuizen. Overigens betogen Hu en Pooler (2002) dat bij een juiste specificatie van het ruimtelijk interactiemodel er geen ruimtelijke patronen in de parameter waarden mogen bestaan en dat dit dus kan duiden op voorspellingsfouten. Het valt moeilijk inhoudelijk te duiden waarom mensen die in de periferie wonen meer door afstand wordt 'afgeschrikt' bij het migreren, dan mensen die meer centraal wonen (de twee auteurs hadden bij hun analyse van migratieverkeer in de Verenigde Staten verhoogde waarden in dunbevolkte staten, zoals Noord Dakota, geconstateerd).

Kaart 6: 'Schatting van de parameter van inwonertal, per gemeente.



Planbureau voor de Leefomgeving

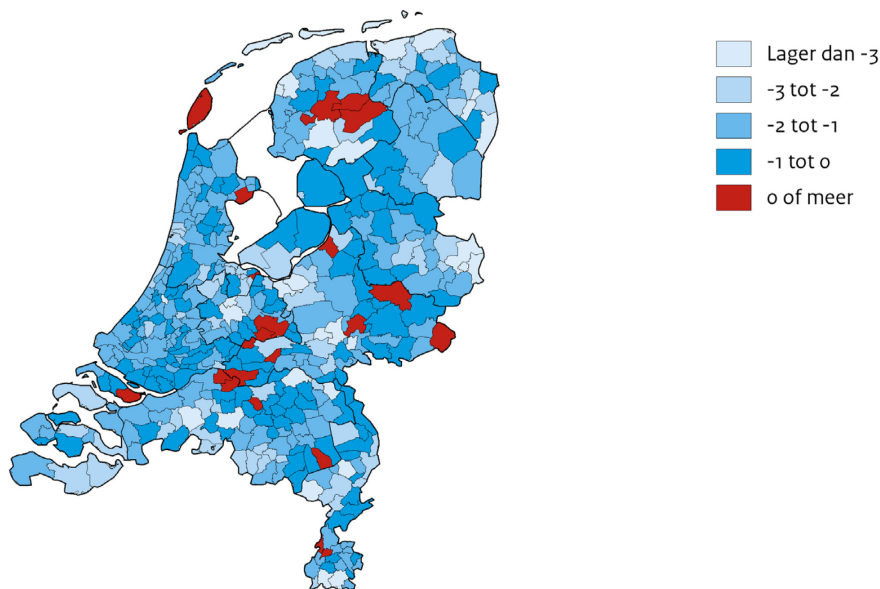
Kaart 7: 'Schatting van de parameter van afstand, per gemeente.



Planbureau voor de Leefomgeving

De schattingen van de parameter waarden van de variabele toegankelijkheid zijn voor veruit de meeste gemeenten negatief (zie kaart 8). De negatieve waarden geven aan dat er concurrentie bestaat tussen verschillende bestemmingsgemeenten, waardoor het voor migranten in een bepaalde vertrekgemeente moeilijker wordt een bepaalde bestemmingsgemeente te kiezen. Hoe negatiever de waarde van deze parameter is, hoe meer het migratieverkeer vanuit een bepaalde vertrekgemeenten beïnvloed wordt door de aanwezigheid van aantrekkelijke alternatieven bij de keuze van een bestemmingsgemeente. Met enige voorzichtigheid kan men stellen dat in gemeenten aan de randen van Nederland de waarde van de parameter voor centrale ligging in het algemeen wat sterker negatief zijn. Dit betekent dat het voor migranten uit deze gemeenten wat lastiger is om een specifieke bestemmingsgemeente te kiezen, doordat men uit meer geschikte alternatieven kan kiezen. In een beperkt aantal gemeenten is de geschatte parameter positief; volgens Thorsen en Gitlesen (1998) geeft dit aan dat 'agglomeratie-effecten' sterker zijn dan 'concurrentie-effecten'.

Kaart 8: 'Schatting van de parameter van toegankelijkheid per gemeente.



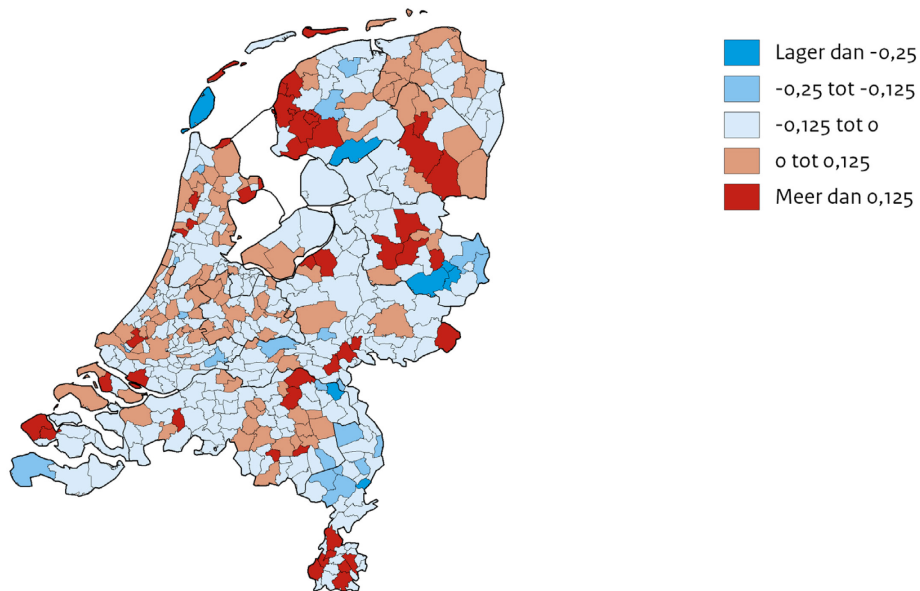
Planbureau voor de Leefomgeving

De schattingen van de parameter waarden van de variabele woningbouw vertonen verspreid over Nederland zowel hoge als lage waarden (zie kaart 9). Er zijn wel wat clusters zichtbaar van gemeenten met relatief hoge waarden. Zo ligt in het zuid oosten van Friesland een gebied met vrij hoge waarden. In deze regio wordt mogelijk relatief weinig gebouwd, zodat mensen die hier wonen vrij sterk worden aangetrokken door woningbouw elders (in het zuiden van Leeuwarden wordt bijvoorbeeld al meerdere jaren veel gebouwd). Van Wissen en Rima (1988) melden in dit kader dat bij de oplevering van grote nieuwbouwwijken (met een gunstige prijs/kwaliteit verhouding), dit tijdelijk zal leiden tot een sterke instroom van migranten uit omliggende regio's.

Natuurlijk is voor de opname van een bepaalde variabele in het ruimtelijk interactiemodel van groot belang de significantie van de geschatte waarde van de betreffende parameter. De parameters van de twee variabelen inwonertal en afstand blijken in alle gemeenten significant te zijn. De parameter van de variabele toegankelijkheid blijkt significant te zijn voor een flink

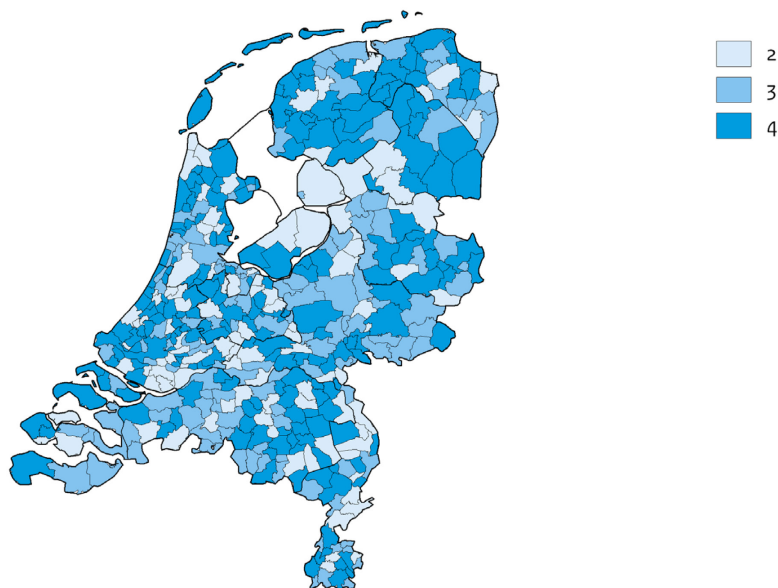
aantal gemeenten, terwijl de parameter van de variabele netto woningbouw slechts voor een beperkt aantal gemeenten significant is. Vooral op basis van deze significantie is de keuze gevallen op een verschillend aantal verklarende variabelen voor de afzonderlijke (vertrek)gemeenten. In bepaalde gevallen zijn er echter ook parameters opgenomen die niet significant waren, maar wel leiden tot een duidelijk betere voorspelling in de werkelijke ruimte. *Kaart 10* toont de gekozen specificatie naar aantal verklarende variabelen per gemeente. Hieruit blijkt dat voor de meeste gemeenten gekozen is voor een model met meer dan twee verklarende variabelen. Er valt overigens geen duidelijk ruimtelijk patroon te ontdekken in de specificatie van de modellen naar aantal verklarende variabelen. Wel is het zo dat, en met name bij het vier variabelen model, vrij vaak gemeenten die aan elkaar grenzen dezelfde specificatie hebben.

Kaart 9: 'Schatting van de parameter van netto woningbouw per gemeente.



Planbureau voor de Leefomgeving

Kaart 10: Aantal verklarende variabelen dat de kleinste voorspelfout oplevert¹, per gemeente.



Planbureau voor de Leefomgeving

1. Op basis van de indicator SRMSE.

6. Uitkomsten van de schatting met het ruimtelijk interactiemodel

In deze paragraaf wordt aan de hand van spreidingsdiagrammen een beeld gegeven van de uitkomsten van de twee modellen: het model dat in de vorige prognose is gebruikt (met twee verklarende variabelen dat op provinciaal niveau is gespecificeerd) en het model dat in de nieuwe regionale prognose van 2011-2040 wordt gebruikt (dat een wisselend aantal variabelen kent en op gemeentelijk niveau is gespecificeerd). Voor elke provincie worden de uitkomsten van beide modellen getoond, zodat een goed beeld kan worden verkregen van het effect van de twee verschillende specificaties. In de spreidingsdiagrammen zijn de verwachte aantallen verhuisde personen volgens het ruimtelijk interactiemodel uitgezet tegen de waargenomen aantallen. Elk punt op de diagram stelt het geschatte en het waargenomen aantal migranten tussen een bepaalde vertreksgemeente en een bepaalde vestigingsgemeente voor. Voor punten op de 45°-lijn geldt dat verwachte en waargenomen aantallen gelijk zijn; punten onder die lijn zijn het gevolg van onderschatting, terwijl punten boven de lijn het resultaat zijn van overschatting. Hoe dichter de puntenwolk dus tegen de 45°-lijn ligt, hoe beter dus de schatting van het ruimtelijk interactie model is. Bij een overschatting is een bepaalde vestigingsgemeente minder populair dan op basis van de verklarende variabelen mocht worden verwacht. Bij een onderschatting is een bepaalde vestigingsgemeente in werkelijkheid meer in trek dan het model had verwacht.

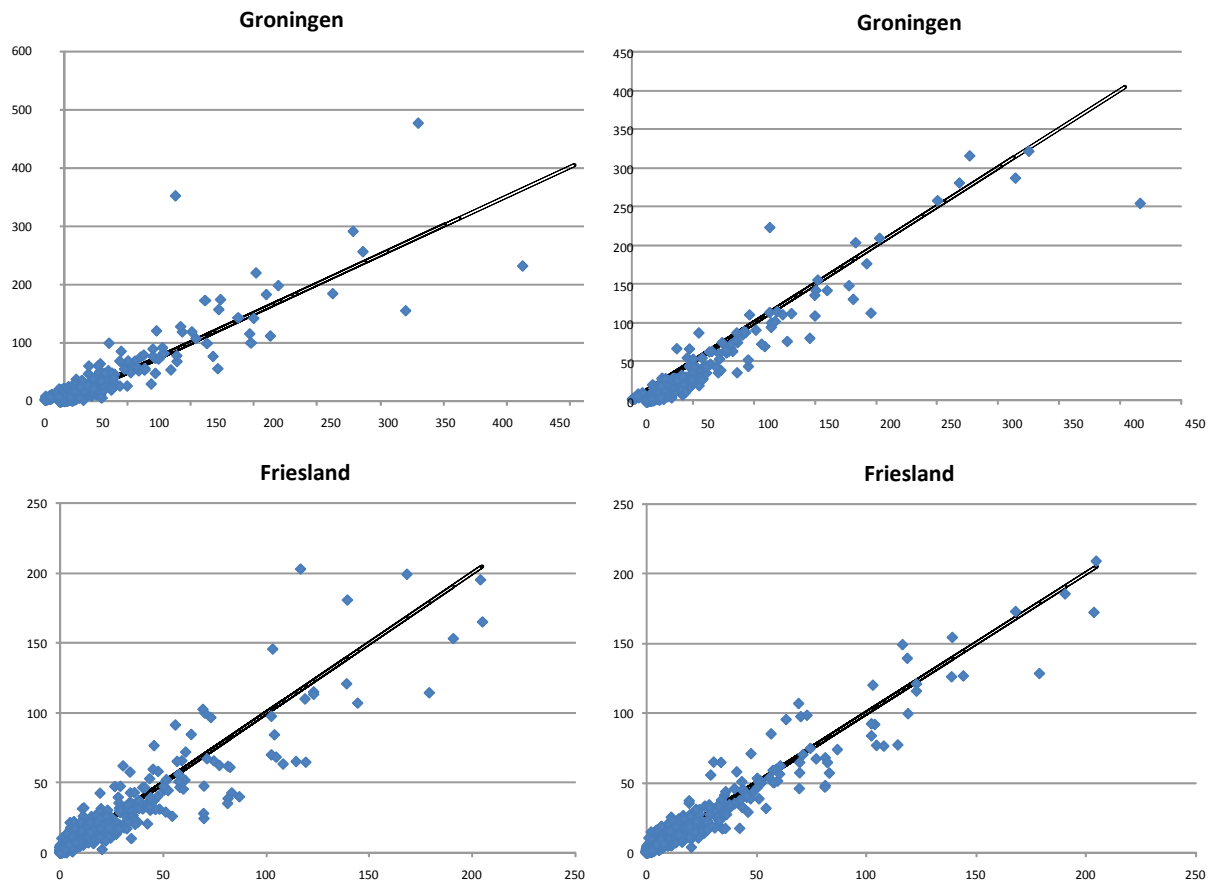
In geval dat de schatting sterk afwijkt van de observatie, kan dit worden veroorzaakt door structurele kenmerken van bepaalde gemeenten die niet in de modellering zijn meegenomen. Zo kunnen bepaalde gemeenten meer migranten trekken dan het model verwacht, doordat deze gemeenten zijn erg aantrekkelijk zijn door bijvoorbeeld de aanwezigheid van bepaalde voorzieningen (grote steden met hogere onderwijsinstellingen), het sociaal cultureel klimaat (grote steden met veel uitgaansmogelijkheden), het economisch potentieel (grote steden in de Randstad) en de leefomgeving (gemeenten gelegen in een fraai landschap). Ook het

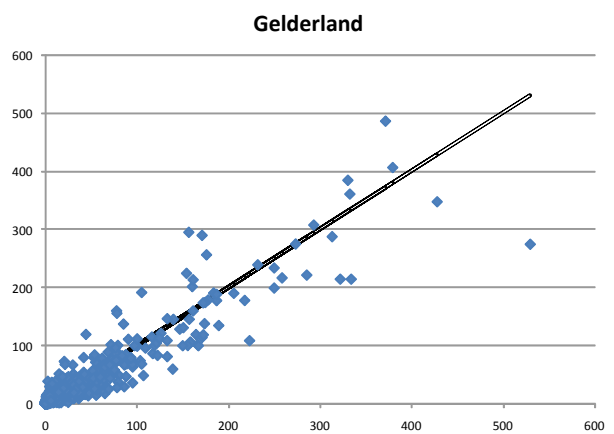
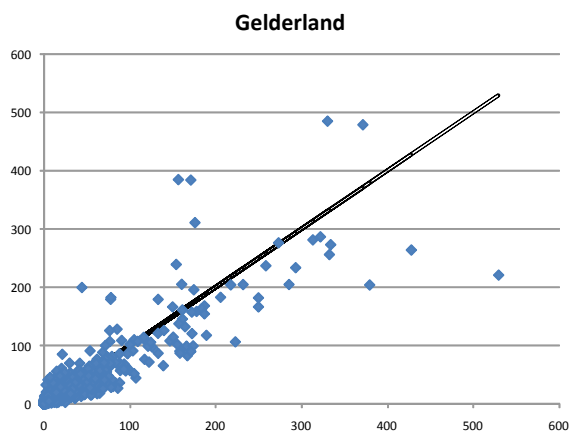
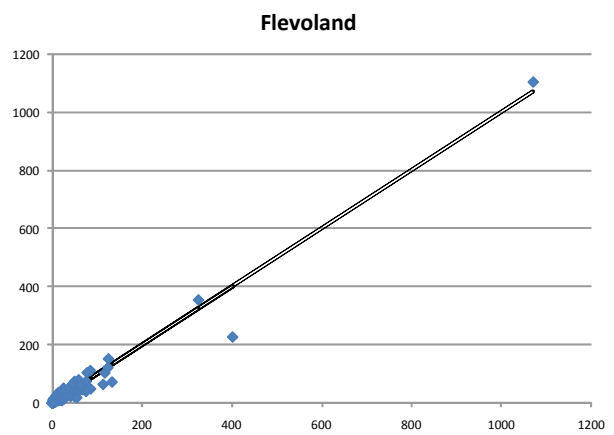
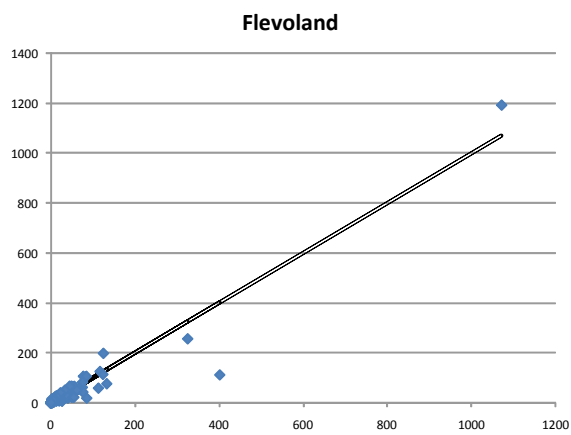
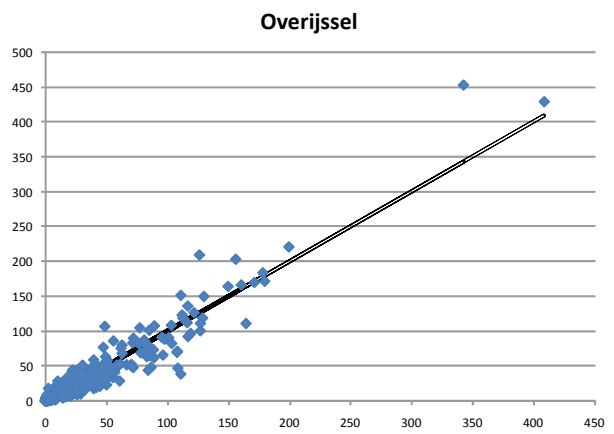
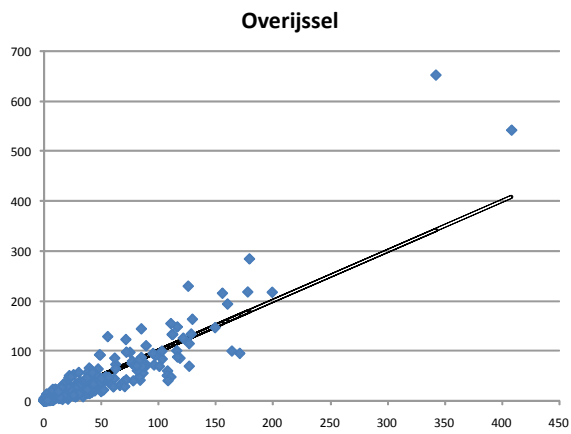
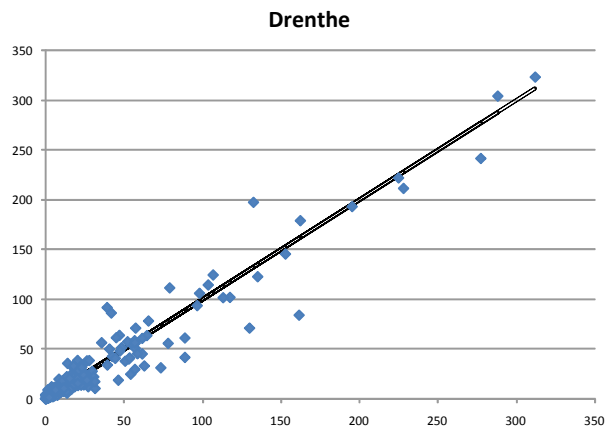
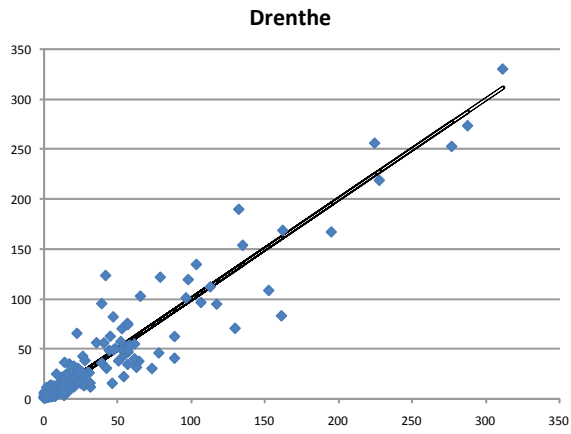
omgekeerde kan gelden: door de afwezigheid van voorzieningen en banen kunnen bepaalde vestigingsgemeenten weinig in trek zijn. Deze vorm van (on)aantrekkelijkheid kan in een lange periode effect blijven uitoefenen. In het prognosemodel PEARL wordt rekening met dit type fout gehouden door de schatting volgens het ruimtelijk interactiemodel aan te passen aan de hand van een correctiefactor. Aan de hand van de uitbijters op de spreidingsdiagrammen kunnen dergelijk gevallen worden getraceerd en kan door vergelijking van de geschatte met de waargenomen aantallen migranten de hoogte van de correctiefactor worden bepaald.

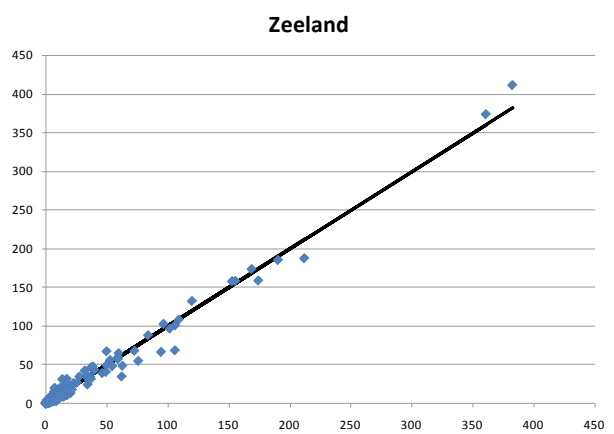
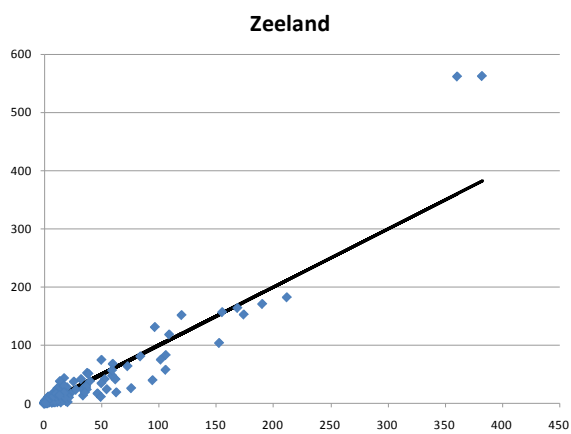
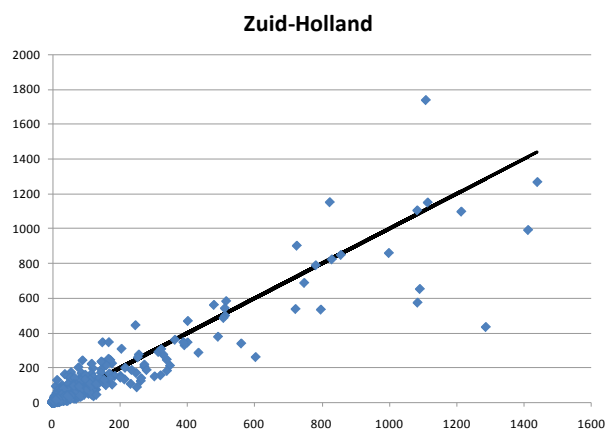
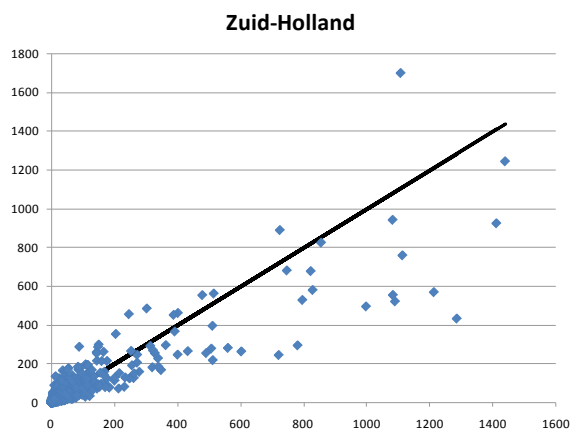
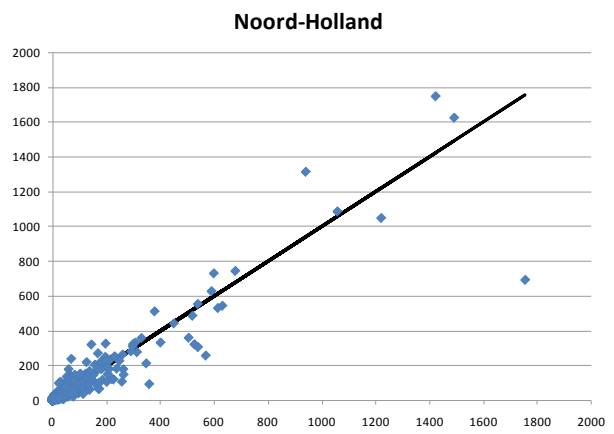
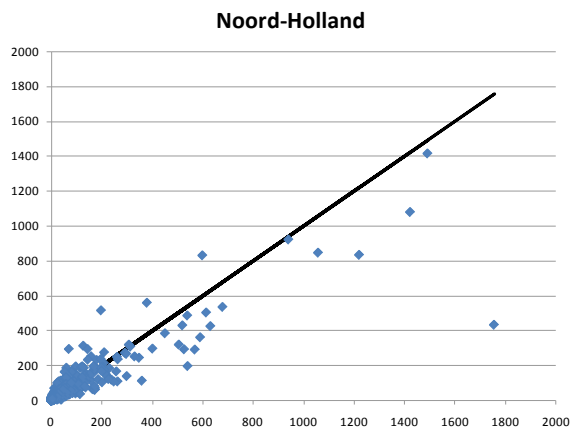
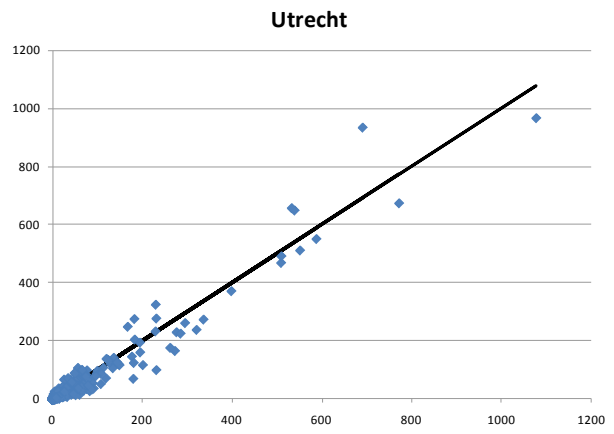
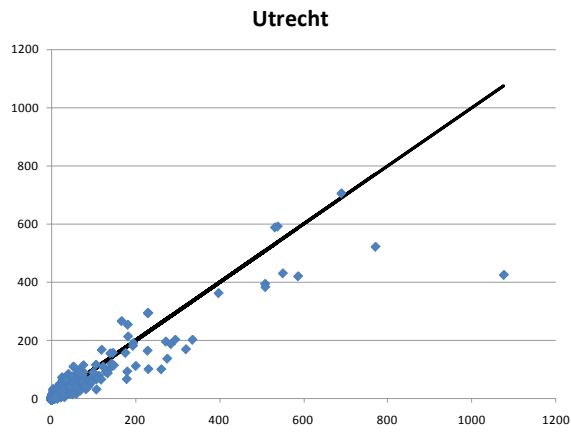
In het model dat bij de vorige prognose werd gebruikt, behoorde woningbouw niet tot de verklarende variabelen. Voor de verklaring van binnenlandse migratiestromen lijkt het effect van de woningmarkt echter behoorlijk groot. Indien in een bepaalde gemeente recent een nieuwe woonwijk recent is opgeleverd, dan kan dit leiden tot een (meestal tijdelijk) krachtige migratiestroom naar deze gemeente. Krapte op de woningmarkt, hetgeen meestal samengaat met hoge huizenprijzen, heeft juist een omgekeerd effect: gemeenten met een tekort aan woningen trekken weinig migranten aan. In het ruimtelijk interactie model dat voor de update van de regionale prognose is gebruikt, is de variabele woningbouw nu wel opgenomen (geoperationaliseerd als de toename van woningvoorraad in tussen 1 januari 2006 en 1 januari 2009). Echter, in de toepassing van (de geschatte parameters van) het ruimtelijk interactie model in het prognosemodel PEARL dat bij de nieuwe regionale prognose is gebruikt, wordt deze variabele *niet* meegenomen als verklarende variabele. De reden hiervoor is dat het effect van de woningmarkt op de (korte afstand) migratie in een aparte stap wordt gemodelleerd (zie paragraaf 2). Hiertoe werkt het model PEARL met veronderstellingen over de jaarlijkse netto toevoegingen aan de bestaande woningvoorraad en wordt niet alleen op basis van korteaftandsmigratie maar ook op basis van andere componenten (buitenlandse migratie, sterfte, huishoudensovergangen en ten slotte langeafstandsmigratie) gekeken naar vraag en aanbod van woningen. Doordat de woningbouw nu wél is meegenomen bij de schatting van het ruimtelijk interactie model, zijn de parameters van de overige verklarende variabelen gecorrigeerd voor eventuele effecten van de woningmarkt (hetgeen niet het geval was bij de parameters die bij vorige regionale prognose zijn gebruikt).

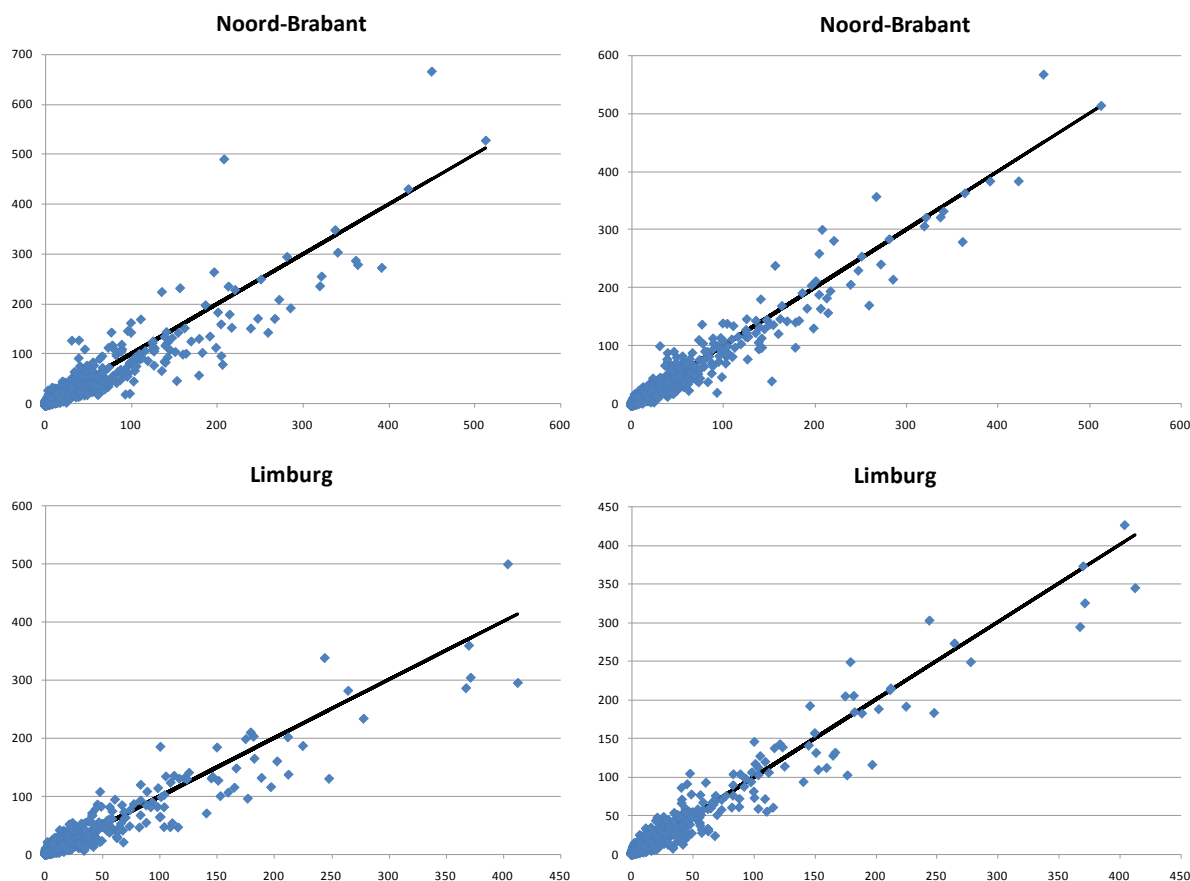
Hieronder wordt per provincie een bespreking van de uitkomsten van de twee specificaties van het ruimtelijk interactiemodel gegeven.

Figuur 3: Spreidingsdiagram per provincie (x-as: waargenomen migratiestromen, y-as voorspelde migratiestromen; linker grafiek: specificatie die in de vorige prognose is gebruikt, rechter grafiek: specificatie die in de nieuwe prognose is gebruikt).









Voor de provincie Groningen geldt dat de geschatte migratiestromen in het nieuwe model over het algemeen veel dichterbij de 45-graden lijn liggen dan in het oude model. Zowel voor het oude als nieuwe model geldt dat in werkelijkheid meer migranten naar Tynaarlo verhuizen dan wordt voorspeld door het model; de landschappelijke schoonheid van deze gemeente kan hierbij een rol spelen. Omgekeerd zijn in het nieuwe model veel minder migranten naar Bedum verhuisd, dan door het model wordt verwacht. In het oude model wordt met name de migratiestroom van Groningen naar Assen verkeerd geschat: in werkelijkheid ligt dit veel hoger dan door het model wordt voorspeld. In de nieuwe modellering wordt deze migratiestroom vrijwel correct voorspeld.

De spreidingsdiagram voor Friesland laat voor het vorige model relatief veel uitschieters zien bij gemeenten die in de oostelijke helft van de provincie liggen. In het nieuwe model zijn deze uitschieters sterk gereduceerd.

Voor het Drenthe geldt dat het oude model geen echte uitbijters laat zien. De voorspelfout is echter in het nieuwe model over het algemeen geringer.

Wat betreft Overijssel laat het vorige model diverse uitbijters ten opzichte van de 45-gradenlijn zien. Zo worden de verhuizingen tussen Enschede en Hengelo duidelijk overschat. In het nieuwe gemeentelijke model wordt enkel de migratiestroom van Enschede naar Hengelo duidelijk overschat.

In Flevoland komen in het oude model de schattingen vrijwel overeen met de waarnemingen, op twee belangrijke uitzonderingen na. De eerste uitzondering betreft de verhuisstroom van Almere naar Lelystad; er werd aanzienlijk vaker verhuisd dan wordt verwacht door het model. Ook in de nieuwe modellering duikt deze voorspelfout weer op. De huizenprijzen liggen in Lelystad in vergelijking met Almere relatief laag en dit heeft mogelijk een aanzuigend effect op de bewoners van Almere. De tweede belangrijke voorspelfout van het vorige model betreft de verhuisstroom van Almere naar Amsterdam: er verhuizen veel minder vaak mensen van

Almere naar Amsterdam dan door het model wordt verwacht. In de nieuwe modellering is deze voorspelfout vrijwel verdwenen.

Voor Gelderland geldt dat het vorige model flink veel uitbijters laat zien. In het nieuwe model is het voorspelfout in het algemeen kleiner, hoewel er nog wel vrij veel uitbijters blijven bestaan. De grootste voorspelfout, zowel in het oude als het nieuwe model, betreft de migratie van Arnhem naar Rheden: het voorspelde aantal verhuizingen ligt ver onder het waargenomen aantal verhuizingen. De voorspelfout kan mogelijk worden veroorzaakt door het feit dat Rheden zeven kernen kent, waarvan Velp en Dieren de grootste zijn. Velp ligt tegen Arnhem aan en fungeert deels als overlooptgemeente; dit leidt mogelijk tot extra verhuizingen. Dit feit komt mogelijk te weinig tot uitdrukking in zowel het oude als nieuwe model.

Wat betreft Utrecht geldt dat in het vorige model sprake is van een sterke onderschatting van de migratiestroom van Utrecht naar Amsterdam. Ook in het nieuwe model is er sprake van een onderschatting, maar deze is veel kleiner geworden. Amsterdam is voor migranten een zeer aantrekkelijke stad, gezien de vele opleidingsinstituten en ruimere kansen op een baan. De migratiestroom van Utrecht naar de Utrechtse heuvelrug wordt ook (licht) onderschat. De landschappelijke schoonheid van deze gemeente trekt waarschijnlijk extra veel migranten aan. Opvallend is dat het nieuwe model de migratiestroom vanuit Utrecht naar Nieuwegein en in mindere mate naar Maarsen en De Bilt overschat. Mogelijk zijn er beperkte mogelijkheden om in de laatste twee gemeenten te vestigen, mede door hoge huizenprijzen aldaar.

De schattingen voor Noord-Holland laten met het oude model diverse uitbijters zien, waarvan verschillende in de schattingen met het nieuwe model zijn verdwenen. Op basis van het nieuwe model wordt verwacht dat er meer migranten vanuit Amsterdam naar zowel Haarlem en Zaanstad verhuizen, dan er in werkelijkheid heeft plaatsgevonden. Vanuit Amsterdam verhuizen naar Almere in werkelijkheid ongeveer drie keer zoveel mensen als door het model wordt voorspeld. In het vorige model is de onderschatting van deze verhuisstroom nog veel groter. De grote aantrekkingskracht van Almere wordt waarschijnlijk veroorzaakt door een ruim aanbod van (relatief goedkope) nieuwbouwwoningen. Dit wordt kennelijk niet goed 'opgepikt' in het nieuwe model, waarin de woningbouw als variabele is opgenomen. Dit kan mogelijk het gevolg zijn van het feit dat er de laatste jaren relatief weinig is gebouwd in Almere. Desondanks zit er kennelijk nog veel ruimte in de woningvoorraad (zoals leegstaande woningen en vrijgekomen woningen door mensen die uit Almere wegtrekken), zodat mensen toch wel een woning in Almere kunnen vinden.

Zuid-Holland laat in beide modellen een vrij groot aantal onderschatte verhuisbewegingen zien. Het gaat vooral om verhuizingen tussen naburige gemeenten in het westelijk deel van de provincie, zoals verhuizingen van Den Haag naar Zoetermeer en verhuizingen van Rotterdam naar Lansingerland, Barendrecht en Capelle aan de IJssel. De laatste drie gemeenten fungeren als overlooptgemeenten voor Rotterdam en Zoetermeer vervult die rol voor Den Haag; dit verklaart waarschijnlijk de onderschatting. Wederom wordt dit effect niet of te weinig opgepikt door de variabele woningbouw in het nieuwe model. De verhuizingen van Den Haag naar Rijswijk worden in beide modellen juist overschat. Mogelijk is dit het gevolg van weinig nieuwbouw dan wel relatief hoge huizenprijzen in Rijswijk.

Voor Zeeland laat het nieuwe model zeer geringe afwijkingen zien tussen waarneming en voorspelling. Dit geldt veel minder voor het oude model, waar sprake is van twee uitbijters, namelijk de migratiestromen tussen Vlissingen en Middelburg.

Terwijl de spreidingsdiagram van het oude model voor Noord-Brabant tamelijk veel beduidende over- en onderschattingen laat zien, geldt dit voor de grafiek van het nieuwe model veel minder. Het nieuwe model laat toch nog enkele redelijk grote overschattingen zien. Het model voorspelt te veel verhuizingen vanuit Eindhoven naar de 'dure' buurgemeenten Veldhoven en Nuenen, en Gerwen en Nederwetten. Ook het aantal verhuizingen van Tilburg naar Breda wordt duidelijk overschat.

Voor Limburg is er in het oude model over het algemeen sprake van relatief geringe voorspelfouten. In het nieuwe model is de voorspelfout over de gehele linie nog verder teruggelopen.

4. Conclusie en evaluatie

In dit artikel is nagegaan in hoeverre het ruimtelijk interactiemodel dat in de vorige regionale prognose is gebruikt voor de schatting van korte afstandsmigratie verder kan worden verbeterd. Op basis van dit onderzoek is het ruimtelijk interactie model dat ten behoeve van de nieuwe regionale prognose is gebruikt op twee manieren aangepast. Het eerste type aanpassing betreft het toevoegen van extra verklarende variabelen in de regressievergelijking, ten behoeve van de schatting van het ruimtelijk interactie model. In dit kader is ten eerste de variabele 'toegankelijkheid' ten opzichte van alternatieve bestemmingen toegevoegd en ten tweede de mutatie in de woningvoorraad in de laatste jaren. Met name de toevoeging van de toegankelijkheid variabele leidde tot een belangrijke verbetering van de schatting van het migratieverkeer in diverse provincies en in het bijzonder voor Flevoland. Het tweede type aanpassing betrof de overgang van een schatting op 'globaal niveau' naar 'lokaal niveau': in de vorige prognose is het model geschat voor provincies, terwijl het in de nieuwe prognose is geschat voor individuele (vertrek)gemeenten. De modellen die op gemeentelijk niveau zijn gespecificeerd blijken over het algemeen kleinere voorspelfouten hebben dan de modellen die provincie niveau zijn gespecificeerd. Een grote reductie in voorspelfouten is met name van toepassing voor de provincies Flevoland, Zeeland, Groningen en Overijssel. Op basis van deze resultaten is in de nieuwe prognose gekozen voor de specificatie van het ruimtelijk interactie mode op gemeentelijk niveau met afwisselend twee, drie of vier verklarende variabelen (afhankelijk van de kleinste voorspelfout van de verschillende specificaties).

Ondanks de duidelijke verbetering van de fit met de nieuwe specificatie van het ruimtelijk interactiemodel, blijven er aantal migratiestromen met beduidende voorspelfouten over. In een aantal gevallen gaat het om de verhuizing naar landschappelijk aantrekkelijke gemeenten, waarbij het model de migratiestroom onderschat. Ook het migratie verkeer naar de grote gemeenten is in diverse gevallen onderschat. Zo is de grote aantrekkelijkheid van Amsterdam voor migranten uit Utrecht door het model onderschat. De operationalisering van aantrekkelijkheid door de variabele inwonertal in de regressievergelijking op te nemen, schiet dan kennelijk tekort. De grootste onderschatting treedt op bij de migratiestroom van Amsterdam naar Almere. De laatste gemeente fungeert al lang als overloopgemeente en de ruime woningmarkt lokt extra verhuizingen uit, hetgeen door het model niet goed wordt opgepikt. Ook in andere gevallen wordt de verhuisstroom naar overloopgemeenten onderschat.

Het feit dat in een aantal gevallen de migratiestroom niet goed door het ruimtelijk interactie model wordt voorspeld, heeft overigens geen ernstige gevolgen voor de prognose van korte afstand migratie in het regionaal prognosemodel PEARL. De schattingen door het ruimtelijk interactie model wordt namelijk gebruikt als een 'initiële' schatting van het migratieverkeer. In een vervolgstap van het model PEARL worden pas de definitieve migratiestromen bepaald. Dit gebeurt aan de hand van een beknopt woningmarktmodel, waarin de vraag naar woningen wordt aangepast aan het aanbod van woningen, hetgeen leidt tot verschuivingen in door het ruimtelijk interactie model bepaalde korteaftandsmigratiestromen.

Literatuur

Feijten, P. en P. Visser (2005), Binnenlandse migratie: verhuismotieven en verhuisafstand. Bevolkingstrends 53(2), blz. 75–81. CBS, Voorburg/Heerlen.

- Fotheringham A.S., C. Brunsdon and M. Charlton (2002), *Geographically weighted regression: the analysis of spatially varying relationships*, John Wiley and Sons.
- Fotheringham, AS and ME O'Kelly (1989), *Spatial interaction models. Formulations and applications*. Dordrecht/Boston/London: Kluwer Academic Publishers.
- Hu, P. and J. Pooler (2002), An empirical test of the competing destinations model, *Journal of Geographical Systems* 4, pp. 301-323, Springer-Verlag.
- Imhoff, E. van, N. van der Gaag, L. van Wissen & P. Rees (1997), The selection of internal migration models for European regions. *International Journal of Population Geography*, 3: 137-159.
- Jong, A. de. (2005), Regionale verschillen in migratie over korte afstand: een ruimtelijk interactiemodel. *Bevolkingstrends* 53(2), blz. 82–92. CBS, Voorburg/Heerlen.
- Jong, A. de, M. Alders, P. Feijten, P. Visser, I. Deerenberg, M. van Huis, D. Leering (2005), *Achtergronden en veronderstellingen bij het model PEARL. Naar een nieuwe regionale bevolkings- en allochtonenprognose*. RPB/CBS, NAI uitgevers.
- Lloyd, C.D. (2007), *Local models for spatial analysis*. CRC Press.
- Stillwell J. and P. Congdon (1991), *Migration models: macro and micro approaches*. Belhaven press, London.
- Tate, N. J. Tate and P.M. Atkinson (2001), *Modelling scale in geographical information science*. John Wiley and Sons.
- Wijngaarden, P. and A. de Jong (2008), Verbeterde schattingswijze van migratie over korte afstand in het model PEARL. *Bevolkingstrends* 56(3), pp. 55–64. CBS, Voorburg/Heerlen.
- Wissen, L. van en A. Rima (1988), *Modelling Urban Housing Market Dynamics*, Studies in Regional Science and urban Economics. Amsterdam: North Holland.