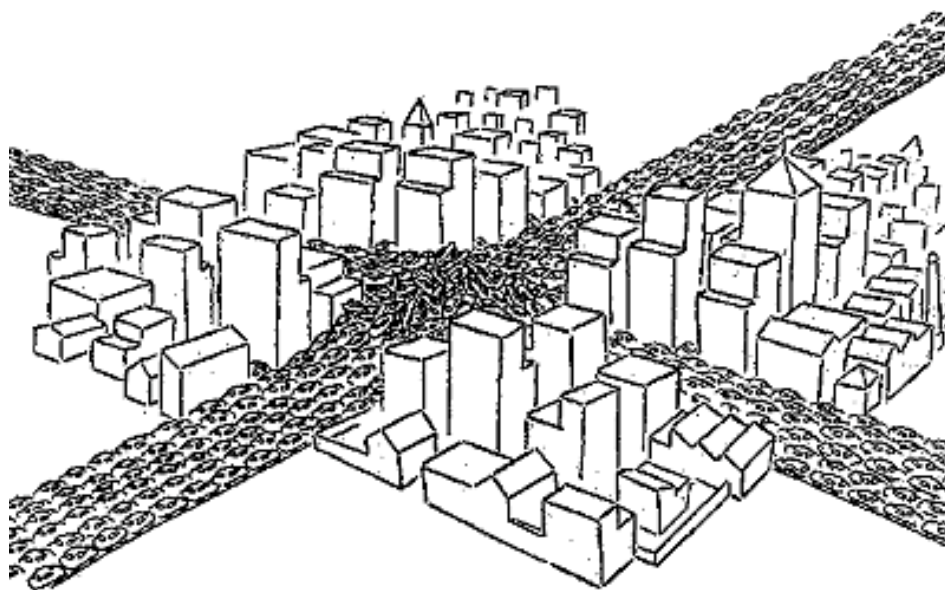




Universiteit Twente
de ondernemende universiteit

De wegen van de bedrijvigheid

*Uitbreiding van het vestigingsplaatsmodel van bedrijven binnen
een datagedreven landgebruik en transport interactie model*



Nico Veenstra

December 2005

Nico Veenstra

De wegen van de bedrijvigheid

Uitbreiding van het vestigingsplaatsmodel van bedrijven binnen een
datagedreven land en transport interactie model

Afstudeerscriptie

Begeleiding:

Prof.dr.ir. M.F.A.M. van Maarseveen

Dr.ir. F.Tillema

Universiteit Twente

Faculteit Construerende Technische Wetenschappen

Opleiding Civiele Techniek

Afdeling Verkeer, Vervoer & Ruimte

Enschede, december 2005

Voorwoord

Met trots presenteer ik u mijn afstudeerscriptie. Dit is het resultaat van ongeveer tien maanden onderzoek naar de interactie tussen bedrijfsactiviteiten en transport. In deze periode heb ik met veel plezier, maar ook met het nodige doorzettingsvermogen, gewerkt aan dit onderzoek. Dit is tevens de officiële afronding van mijn studie Civiele Techniek aan de Universiteit Twente. Hiermee komt een eind aan een periode van ruim zes jaar, waarin ik uiteraard veel kennis op heb gedaan. Daarnaast heb ik ook vele nieuwe ervaringen opgedaan en vele leuke mensen leren kennen. Kortom, het einde van een zeer leuke tijd hier in Enschede.

Dit onderzoek zou ik niet alleen hebben kunnen uitvoeren. Hiervoor heb ik de hulp gehad van een tweetal kundige begeleiders. Ten eerste is dit mijn dagelijkse begeleider, Frans Tillema. Het werk van mij is een voortzetting van eerder door hem verricht promotieonderzoek. Frans kan ik alleen maar omschrijven als een uitstekend begeleider. Hij heeft mij de ruimte gegeven om een goed onderzoek neer te zetten, maar hield wel steeds een vinger aan de pols. Altijd bereid om in korte tijd te reageren op ontstane problemen en hierbij kritische maar vaak zeer bruikbare suggesties te leveren. Daarnaast is er steeds een positieve benadering van het onderzoek geweest, wat mij zeer goed bevallen is.

Als tweede gaat mijn dank uit naar professor Van Maarseveen. Ondanks een tussentijds verblijf in Zuid Afrika, waarschijnlijk om zich te bezinnen op mijn onderzoek, heeft hij op een aantal momenten een belangrijke bijdrage geleverd. Door een onbevangen, scherpe blik en zijn grote kennis op het gebied van verkeer en vervoer kon hij elke keer, zoals hij dat zelf een keer genoemd heeft, "de advocaat van de duivel" spelen. Dit heeft vele waardevolle suggesties opgeleverd.

Nico Veenstra

Enschede, december 2005

Samenvatting

Er wordt in brede kringen aangenomen dat er een verband bestaat tussen landgebruik ontwikkelingen en transport. In het kort; waar mensen activiteiten ontplooiën, zijn transportfaciliteiten aanwezig en visa versa. In de traditionele vervoersplanning wordt echter geen rekening gehouden met landgebruik ontwikkelingen. Dit zorgt ervoor dat vooral strategische voorspellingen binnen de vervoersplanning als inconsistent kunnen worden betiteld, omdat bovengenoemde interactie voornamelijk op een strategische tijdschaal, 10 tot 30 jaren in de toekomst, plaatsvindt. Om deze reden worden in LUTI (Landgebruik en Transport Interactie) modellen landgebruik ontwikkelingen geïntegreerd in transportmodellen. Tot op heden is de toepasbaarheid van deze modellen en de empirische validatie van de wederzijdse beïnvloeding van landgebruikvariabelen en transport gebrekkig. Hierdoor zijn ook de LUTI modellen nooit tot volle bloei gekomen. Als reactie hierop is door Tillema (2004) een datagedreven LUTI model ontwikkeld.

Dit model bestaat uit drie belangrijke onderdelen: (potentiële) bereikbaarheid, locatiekeuzes van huishoudens en locatiekeuzes van bedrijven. Tillema toont in zijn onderzoek aan dat dit model geschikt is voor strategische voorspellingen. Dit onderzoek richt zich op de locatiekeuzes van bedrijven binnen het model. Hierbij is de volgende tweeledige doelstelling geformuleerd:

- (1) Het ontwerpen van een variabelenset waarmee de locatiekeuzes van bedrijven voorspeld kunnen worden en op basis hiervan het bedrijfslocatiemodel van Tillema te herdefiniëren.
- (2) Bepalen of het nieuwe bedrijfslocatiemodel toepasbaar is op elke willekeurige locatie in Nederland, door het model te verifiëren met data uit verschillende studiegebieden in Nederland.

Om deze doelstelling te realiseren is eerst een nieuw conceptueel model geformuleerd, waarmee voorspellingen worden gedaan over het aantal arbeidsplaatsen en het werkgelegenheidsaandeel van bedrijfssectoren op een locatie. Deze voorspellingen lijken een goede input te vormen voor de andere twee modelonderdelen. Om deze output goed te kunnen voorspellen is een variabelenset ontworpen waarvan aangenomen wordt dat deze de locatiekeuzes van bedrijven beïnvloedt. De selectie van variabelen is gebaseerd op basis van belangrijke locatiefactoren uit de literatuur en op basis van de beschikbaarheid van data. De beschikbaarheid van goede data is namelijk een belangrijke voorwaarde voor het werken met datagedreven modellen. Voor het voorspellen van het aantal arbeidsplaatsen zijn een viertal variabelen geselecteerd:

- Parkeerfaciliteiten
- Aanwezigheid adequaat opgeleid personeel
- Aantal bedrijfsvestigingen
- Stedelijkheid gebied

Deze variabelen zijn gebaseerd op factoren die in de literatuur naar voren gekomen als zeer bepalend voor locatiekeuzes van bedrijven. Dit is aangevuld met een viertal variabelen die specifiek de kenmerken van een vestigingsplaats beschrijven. Daarnaast worden hier variabelen aan toegevoegd die specifiek van belang zijn voor de voorspelling van het werkgelegenheidsaandeel van bedrijfssectoren. Met behulp van deze variabelen wordt verondersteld dat het bedrijfslocatiemodel in staat is het aantal arbeidsplaatsen in de sector industrie & bedrijvigheid, de sector handel & distributie en de sector diensten te voorspellen.

Voor het modelleren wordt gebruik gemaakt van een neuraal netwerk. Het neuraal netwerk wordt getraind door het invoeren van data uit de geselecteerde variabelenset en hierbij de corresponderende output te leveren. Als het model goed getraind wordt, is het in staat outputdata te reproduceren die overkomt met de werkelijkheid. De kalibratie van het model is gedaan met een dataset van heel Nederland. Daarnaast wordt de toepasbaarheid van het model op gebiedsniveau getoetst met behulp van een testcase uit de regio Twente. Het doen van goede voorspellingen op gebiedsniveau is een belangrijke voorwaarde voor een goed LUTI model. Met behulp van de testcase Twente wordt tevens inzicht gekregen of het model om kan gaan met bepaalde gebiedskarakteristieken die afwijken van het landelijke gemiddelde.

Uit het onderzoek blijkt dat het nieuwe bedrijfslocatiemodel goed in staat is het aantal arbeidsplaatsen in een gebied te voorspellen. Deze veronderstelling mag gedaan worden als meer dan 70 procent van de data die door het model wordt voorspeld overeen komt met de werkelijkheid. Het werkgelegenheidsaandeel van de sector handel & distributie en de sector diensten worden eveneens goed door het model voorspeld. Alleen is het model niet in staat goede voorspellingen te doen voor de sector industrie & productie. Dit betekent dat de locatiekeuzes van ondernemers uit deze sector afwijken van ondernemers uit andere sectoren. Tevens zijn de variabelen die volgens de literatuur specifiek op deze sector van toepassing zijn, niet toereikend om de locatiekeuzes van de sector industrie & productie te beschrijven. Dit zijn de variabelen "uitbreidingsmogelijkheden", "aanwezigheid toeleverende bedrijven" en "logistieke dienstverlening".

Het lijkt erop dat enkele locatiefactoren die door ondernemers genoemd worden in "stated preference" onderzoeken niet altijd overeenstemmen met de werkelijke locatiefactoren waarop zij hun keuzes baseren. Zo blijkt de invloed van twee hoofdvariabelen, "parkeerfaciliteiten" en "aanwezigheid van adequaat opgeleid personeel" geen significantie invloed hebben op locatiekeuzes van ondernemers. De invloed van de variabele "uitbreidingsmogelijkheden" is daarentegen in het bedrijfslocatiemodel groter dan op basis van de literatuur verwacht zou mogen worden. Het tweede dat opvalt, is dat de input van het bereikbaarheidsmodel de resultaten van het bedrijfslocatiemodel niet beïnvloedt. Het blijkt dat ondernemers hun locatievoorkeur laten beïnvloeden door andere vormen van bereikbaarheid. Dit wordt in het onderzoek aangetoond, doordat de variabelen "aanwezigheid infrastructuur" en "OV-bereikbaarheid" wel invloed blijken te hebben op locatiekeuzes van ondernemers. Dit betekent dat de bereikbaarheid wel degelijk invloed uitoefent op de locatiekeuzes van ondernemers. De theorie van het belang de vestigingsplaats voor locatiekeuzes van ondernemers wordt bevestigd in dit onderzoek. Twee variabelen die in de variabelenset opgenomen zijn, "agglomeratie-effect" en "aanwezigheid infrastructuur", blijken in het onderzoek aantoonbare invloed uit te oefenen op de locatiekeuzes van ondernemers. Alleen voor de sector diensten zijn de vestigingsplaats variabelen niet van invloed. Een verklaring hiervoor zijn de transportkosten, die sterk afhankelijk zijn van de vestigingsplaats, maar van minder belang zijn voor de sector diensten.

De resultaten van de voorspelling van het model voor de testcase Twente komen overeen de kalibratieresultaten van het model met de data van geheel Nederland. Er blijken echter wel een aantal afwijkingen in de data van Twente ten opzichte van het landelijke gemiddelde te zitten. Dat het model ondanks deze afwijkingen toch in staat is goede voorspellingen te doen, toont aan dat het model toepasbaar is op gebiedsniveau. Daarnaast wordt het model getoetst in een zestal regio's in Nederland, met elk een aantal specifieke karakteristieken. Hieruit blijkt dat het model op alle locaties in Nederland toepasbaar is, mits rekening wordt gehouden met twee factoren:

- De beperkte voorspellingskracht van het model voor de sector industrie & productie

- Het nadelige effect van de meetmethode in gebieden met lage aantallen arbeidsplaatsen

Als op het onderzoek wordt teruggezien kan worden geconcludeerd dat het doel van het onderzoek bereikt is: het bedrijfslocatiemodel van Tillema is geherdefinieerd op basis van een nieuwe variabelenset en getoetst op de toepasbaarheid op elke willekeurige locatie in Nederland. Het bedrijfslocatiemodel is op een aantal punten, zoals de definitie van basiselementen, het conceptuele model, het toetsen van de prestaties en toepasbaarheid verder uitgewerkt. De voorspellingen van het nieuwe model zijn voor drie van de vier onderdelen bevredigend. Op basis hiervan mag geconcludeerd worden dat er een bijdrage is geleverd aan de verbetering van het model van Tillema.

Als laatste dient het van belang van een tweetal zaken in dit onderzoek te worden noemen. Ten eerste is de interactie tussen landgebruik ontwikkelingen en transport vaak een dynamisch proces, wat in dit model buiten beschouwing is gelaten, maar wel degelijk van invloed is. Daarnaast is de kwaliteit van de uitkomsten van het model sterk afhankelijk van de kwaliteit van de inputdata en de configuratie van het neurale netwerk. Deze zaken zouden wellicht in toekomstig onderzoek beter uitgewerkt kunnen worden.

Inhoudsopgave

1	INLEIDING	1
1.1	ACHTERGROND	1
1.1.1	<i>Nieuw LUTI model</i>	1
1.1.2	<i>Projectkader</i>	2
1.2	DOELSTELLING EN ONDERZOEKSVRAGEN	3
1.2.1	<i>Doelstelling</i>	3
1.2.2	<i>Onderzoeksvragen</i>	3
1.3	LEESWIJZER	4
2	THEORIEVORMING	5
2.1	LANDGEBRUIK EN TRANSPORT INTERACTIE	5
2.1.1	<i>Achterliggende verbanden LUTI modellen</i>	5
2.1.2	<i>Verbindende factor activiteiten en transport: bereikbaarheid</i>	7
2.1.3	<i>Dynamisch karakter LUTI model</i>	7
2.1.4	<i>Toepassingsmogelijkheden LUTI modellen</i>	8
2.2	LOCATIEKEUZES VAN BEDRIJVEN	9
2.2.1	<i>Belangrijkste factor locatiekeuzes bedrijven</i>	10
2.2.2	<i>Andere factoren locatiekeuzes bedrijven</i>	10
2.2.3	<i>Locatievoorkeuren per bedrijfssector</i>	12
2.2.4	<i>Locatievoorkeuren per regio</i>	14
2.3	CONCLUSIES THEORIEVORMING	18
3	NEURALE NETWERKEN.....	19
3.1	WAT ZIJN NEURALE NETWERKEN?	19
3.2	HOE WERKEN NEURALE NETWERKEN?	20
3.3	TOEPASSINGSMOGELIJKHEDEN NEURALE NETWERKEN	21
3.4	VORMGEVING NEURALE NETWERK	22
3.4.1	<i>Kenmerken neuraal netwerk in Matlab</i>	22
3.4.2	<i>Beïnvloeding kwaliteit neuraal netwerk</i>	23
3.5	CONCLUSIES NEURALE NETWERKEN	25
4	DE BOUW VAN HET MODEL	27
4.1	ACHTERGROND TILLEMA'S BEDRIJFSLOCATIEMODEL	27
4.2	SPECIFICATIES NIEUWE BEDRIJFSLOCATIEMODEL	28
5	OPZET ONDERZOEK	31
5.1	ONDERZOEKSMODEL	31
5.2	ONDERZOEKSSTRATEGIE	32
5.3	ONDERZOEKSMATERIAAL	34
6	UITWERKING ONDERZOEK.....	37
6.1	OPERATIONALISEREN VARIABELEN.....	37

6.1.1	<i>Operationalisatie variabelenset aantal arbeidsplaatsen</i>	37
6.1.2	<i>Operationalisatie variabelenset werkgelegenheidsaandeel bedrijfssectoren</i>	39
6.2	TOELICHTING OUTPUT MODEL	40
6.2.1	<i>Categorisering uitkomstvariabelen</i>	41
6.3	STAPPENPLAN TESTEN VARIABELEN VOOR BEDRIJFSLOCATIEMODEL	43
6.4	UITWERKING TOETSING TOEPASBAARHEID MODEL OP ALLE LOCATIES IN NL	45
6.4.1	<i>Uitwerking selectie regio's</i>	45
7	RESULTATEN ONDERZOEK	47
7.1	INLEIDING	47
7.2	CONFIGURATIE NEURAAL NETWERK IN BEDRIJFSLOCATIEMODEL	47
7.2.1	<i>Configuratie voor kalibratie neuraal netwerk</i>	47
7.2.2	<i>Configuratie voor toetsing methodiek model op gebiedsniveau</i>	49
7.3	BEPALING DEFINITIEVE VARIABELENSET	50
7.3.1	<i>Overzicht definitieve variabelenset</i>	52
7.4	RESULTATEN KALIBRATIE MODEL	53
7.4.1	<i>Resultaten voorspelling aantal arbeidsplaatsen</i>	53
7.4.2	<i>Een deelresultaat uitgelicht</i>	56
7.4.3	<i>Resultaten voorspelling werkgelegenheidsaandeel per bedrijfssector</i>	57
7.5	RESULTATEN TOETSING METHODIEK MODEL OP GEBIEDSNIVEAU	60
7.5.1	<i>Resultaten voorspelling aantal arbeidsplaatsen</i>	60
7.5.2	<i>Resultaten voorspelling werkgelegenheidsaandeel per bedrijfssector</i>	63
7.6	RESULTATEN TOETSING TOEPASBAARHEID MODEL OP ALLE LOCATIES IN NL	65
8	CONCLUSIES EN AANBEVELINGEN	69
8.1	CONCLUSIE	69
8.1.1	<i>Ontwerp variabelenset voor locatiekeuzes bedrijven</i>	69
8.1.2	<i>Kalibratie bedrijfslocatiemodel</i>	69
8.1.3	<i>Toetsing methodiek bedrijfslocatiemodel op gebiedsniveau</i>	70
8.1.4	<i>Toetsing toepasbaarheid model op alle locaties in NL</i>	71
8.2	REFLECTIE	71
8.3	AANBEVELINGEN	72
	REFERENTIES	75
	BIJLAGEN	77

1 Inleiding

1.1 Achtergrond

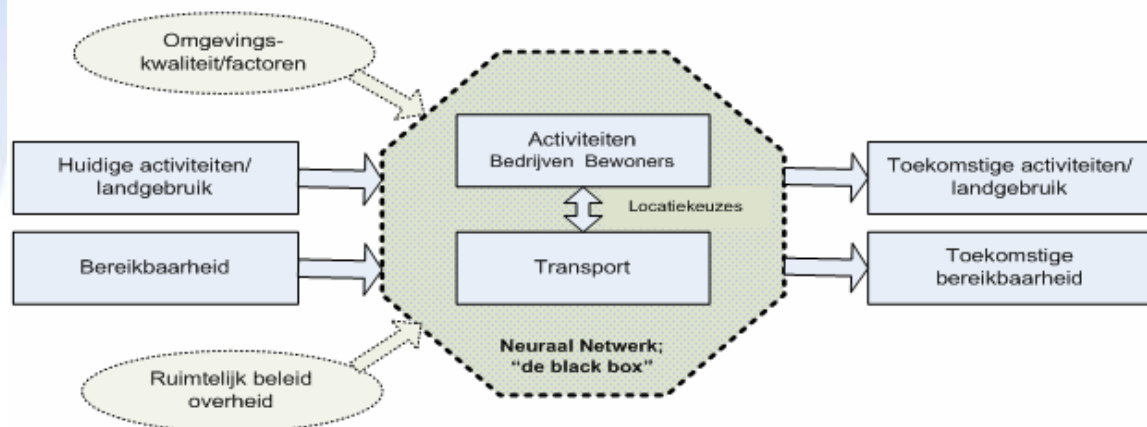
Het verplaatsen van personen, goederen en informatie is van cruciaal belang voor de samenleving. Mobiliteit is hierbij het sleutelbegrip. Strategische vervoersplanning heeft als doel het verkeerssysteem zo veilig, efficiënt en comfortabel mogelijk te laten functioneren, zodat in de huidige en toekomstige behoefte aan mobiliteit kan worden voorzien. De traditionele transportplanning maakt geen gebruik van landgebruik ontwikkelingen. Wel wordt algemeen aangenomen dat er een verband bestaat tussen landgebruik en transport. Hierdoor is in de loop der jaren uitgebreid onderzoek gedaan om inzicht te krijgen in de interactie tussen beide grootheden. Dit heeft geresulteerd in zogenaamde “Land use and transport interaction” (LUTI) modellen. Hiermee is het mogelijk om betere uitspraken te doen over strategische transportplanning. Dit betekent dat de vraagontwikkeling naar transport, beïnvloed door een (veranderende) omgeving, beter dan traditionele modellen voorspeld kan worden voor een periode tussen de 10 en 30 jaar in de toekomst.

Zoals gezegd bestaat er een verband tussen landgebruik en transport. Ontwikkelingen in bijvoorbeeld woongebieden hebben een duidelijke invloed op de mobiliteit. Immers, waar mensen wonen, worden activiteiten ontplooid, wat op haar beurt zorgt voor (vraag naar) verplaatsingen. Over de manier echter waarop beide grootheden elkaar beïnvloeden bestaat nog steeds veel onduidelijkheid. Het ontbreekt tot op heden aan een helder theoretisch kader waarin de relaties tussen landgebruikvariabelen en transportvariabelen beschreven worden. Er bestaat dan ook, ondanks alle inspanningen, nog geen LUTI model dat in brede kring geaccepteerd en gebruikt wordt.

1.1.1 Nieuw LUTI model

Recent is door Tillema (2004) een nieuw LUTI model ontwikkeld dat gebaseerd is op de theorie achter data-gedreven neurale netwerken. Dit model onderscheidt zich doordat het probleem van een gebrek aan bewezen verbanden tussen landgebruik en transport omzeild wordt. Dit wordt opgelost door het gebruik van een neurale netwerk. Een neurale netwerk heeft een “black box” karakter, waardoor het inzicht in achterliggende verbanden ontbreekt, maar ondanks dat voorspellingen kunnen worden gedaan over de interactie tussen landgebruik en bereikbaarheid. Het model van Tillema is opgebouwd uit drie belangrijke onderdelen, te weten (potentiële) bereikbaarheid, locatiekeuzes van huishoudens en locatiekeuzes van bedrijven. In figuur 1.1 staat een schematisch overzicht van dit model.

Tillema concludeert in zijn onderzoek dat een datagedreven LUTI model in staat is de werkelijkheid voldoende nauwkeurig te voorspellen. Daarnaast blijkt uit het onderzoek dat het model in staat is logisch te anticiperen op veranderingen in omgevingsvariabelen.



Figuur 1.1: Schematisch overzicht model Tillemans

1.1.2 Projectkader

Tillemans concentreert zich in zijn onderzoek op de geschiktheid en de algemene toepasbaarheid van zijn datagedreven neurale transportmodel. Het model wordt geoperationaliseerd op basis van een standaard variabelenset voor landgebruik en transport. Tillemans gaat in zijn onderzoek voorbij aan de vraag welke factoren specifiek een sterke bijdrage leveren aan de kwaliteit van de voorspellingen van het model. Het onderzoek richt zich dan ook op deze vraag. Deze vraag wordt alleen getoetst op het onderdeel bedrijfslocatie van het model. Het onderzoek richt zich namelijk specifiek op het bedrijfslocatiemodel.

Dit betekent dat het onderzoek ingaat op de locatiekeuzes van bedrijven. Het is onduidelijk welke variabelenset nodig is voor een optimale kalibratie van het neurale netwerk, zodat het model in staat is het aantal bedrijven in een regio optimaal te voorspellen. In het verleden zijn uit onderzoek een aantal factoren naar voren gekomen waarvan men veronderstelt dat deze van invloed zijn de locatiekeuze van bedrijven. Doordat het hier echter om een nieuwe methode gaat zijn er geen resultaten bekend van factoren die binnen een datagedreven LUTI model van invloed zijn.

Naast het feit dat het bedrijfslocatiemodel voorspellingen kan doen over het aantal bedrijven in een regio, is het ook in staat een uitspraak te doen over de verdeling van desbetreffende bedrijven in een bepaalde bedrijfssector. Immers, een productiebedrijf heeft bijvoorbeeld mogelijk een andere invloed op het transport dan een distributiebedrijf. Tillemans is in zijn onderzoek niet ingegaan op de mogelijkheid dat er verschillen bestaan in voorkeuren tussen bedrijfstypen. Dit leidt wellicht ook tot verschillen in de optimale variabelenset tussen bedrijfstypen.

Een derde punt van aandacht is de vraag in hoeverre het bedrijfslocatiemodel toepasbaar is op elke willekeurige locatie in Nederland. Het model van Tillemans is getraind met een dataset van heel Nederland. De regio Twente wordt in het onderzoek van Tillemans gebruikt om te verifiëren of het model in staat is voorspellingen te doen voor een afgebakend gebied, wat het geval is. Of het model dit echter voor elke andere willekeurig gekozen regio in Nederland kan, is de vraag. De kwaliteit van voorspellingen van het model voor andere gebieden in Nederland, met eigen wellicht unieke karakteristieken, zijn niet onderzocht.

1.2 Doelstelling en onderzoeksvragen

In het voorgaande is ingegaan op de achtergrond van het LUTI vraagstuk en de resultaten die Tillema op dit gebied heeft geboekt met het nieuwe datagedreven LUTI model. Daarnaast is ingezoomd op het onderdeel bedrijfslocatie van het model en zijn een aantal zaken aangestipt die nader onderzoek vereisen. Dit heeft geleid tot de formulering van de doelstelling en de onderzoeksvragen zoals die in de volgende twee paragrafen vallen te lezen.

1.2.1 Doelstelling

De doelstelling van het onderzoek bestaat uit twee delen:

- (1) Het ontwerpen van een variabelenset waarmee de locatiekeuzes van bedrijven (& bedrijfssectoren) voorspeld kunnen worden en op basis hiervan het bedrijfslocatiemodel van Tillema te herdefiniëren.
- (2) Bepalen of het nieuwe bedrijfslocatiemodel op elke willekeurige locatie in Nederland toepasbaar is, door het model te verifiëren met data uit verschillende studiegebieden in Nederland.

1.2.2 Onderzoeksvragen

Op basis van de tweeledige doelstelling zijn een aantal onderzoeksvragen geformuleerd. Dit heeft als doel om aan te geven op welke terreinen onderzoek verricht dient te worden om zo de doelstelling te bereiken.

1. Welke variabelen die de locatiekeuzes van bedrijven beschrijven zorgen voor een betere kalibratie van het bedrijfslocatiemodel?
 - a. Welke factoren komen uit de literatuur naar voren die een substantiële invloed hebben op de locatiekeuzes van bedrijven?
 - b. Kunnen deze factoren omgezet worden in variabelen voor het bedrijfslocatiemodel en zo ja, hoe?
 - c. Hoe kan het neurale netwerk het beste worden vormgegeven om met deze nieuwe input om te gaan?
2. Zijn er verschillende optimale variabelensets om de aandelen van bedrijfssectoren in de totale werkgelegenheid in een gebied te voorspellen?
 - a. Wat zijn volgens de literatuur de verschillen tussen bedrijfssectoren in de mate waarin bepaalde factoren hun locatiekeuzes beïnvloeden?
 - b. Kunnen deze verschillen worden omgezet in aparte variabelensets voor de afzonderlijke sectoren en zo ja, hoe?
 - c. Hoe kan het neurale netwerk het beste worden vormgegeven om met deze nieuwe input om te gaan?

3. Is het bedrijfslocatiemodel op alle locaties in Nederland toepasbaar?
 - a. Zijn er factoren die locatiekeuzes van bedrijven beïnvloeden die in bepaalde regio's sterker of juist minder sterk aanwezig zijn dan in de rest van Nederland?
 - b. Mocht de toepasbaarheid van het model niet op elke willekeurige locatie in Nederland mogelijk zijn, wat zijn de ruimtelijke verschillen die dit veroorzaken?

1.3 Leeswijzer

De basis van dit onderzoek wordt gelegd in hoofdstuk 2. Hierin wordt ingegaan op de achterliggende theorieën en de toepassingsmogelijkheden van LUTI modellen. Daarnaast vindt een uitgebreide beschrijving plaats van de belangrijkste bevindingen op het gebied van locatiekeuzegedrag van ondernemers. Hierbij wordt ook ingegaan op de invloed van de vestigingsplaats op deze locatiekeuzes. In hoofdstuk 3 staat vervolgens een beschrijving van neurale netwerken en de vormgeving van dergelijk netwerken in de praktijk. Deze twee hoofdstukken vormen het eerste deel van het onderzoek, waarmee duidelijkheid wordt geschept in de achterliggende theorieën die van toepassing zijn op dit onderzoek. Daarna wordt in hoofdstuk 4 verder gegaan met de specificaties van het nieuwe bedrijfslocatiemodel. Hierna wordt ingegaan op de opzet van het onderzoek, zodat de onderzoeksvragen met behulp van dit nieuwe model beantwoord kunnen worden. Dit is te vinden in hoofdstuk 5. In hoofdstuk 6 wordt het onderzoek vervolgens uitgewerkt. Hier wordt de variabelenset gedefinieerd, die als input dient voor het bedrijfslocatiemodel. Daarnaast wordt de output van het model besproken en worden de regio's geselecteerd die nodig zijn voor het toetsen van de toepasbaarheid van het model op elke willekeurige locatie in Nederland. In hoofdstuk 7 staan de resultaten van dit onderzoek. Hier worden de prestaties van het model en de afzonderlijke variabelen onderzocht en beoordeeld. Daarnaast worden de prestaties van het model en de afzonderlijke variabelen in de regio Twente bekeken. Hiermee wordt de toepasbaarheid van het model op gebiedsniveau getoetst. Dit vormt het uitgangspunt van het laatste deel van het onderzoek, waarbij met behulp van de geselecteerde regio's, beoordeeld wordt of het model toepasbaar is op elke willekeurige locatie in Nederland. In hoofdstuk 8 staat vervolgens een opsomming van de belangrijkste conclusies die op basis van de resultaten getrokken kunnen worden. Daarnaast wordt ingegaan op de gevolgen die deze conclusies hebben voor het model van Tillemans en de algemene LUTI theorie. Het hoofdstuk wordt afgesloten met het geven van enkele aanbevelingen voor verder onderzoek.

2 Theorievorming

2.1 Landgebruik en Transport interactie

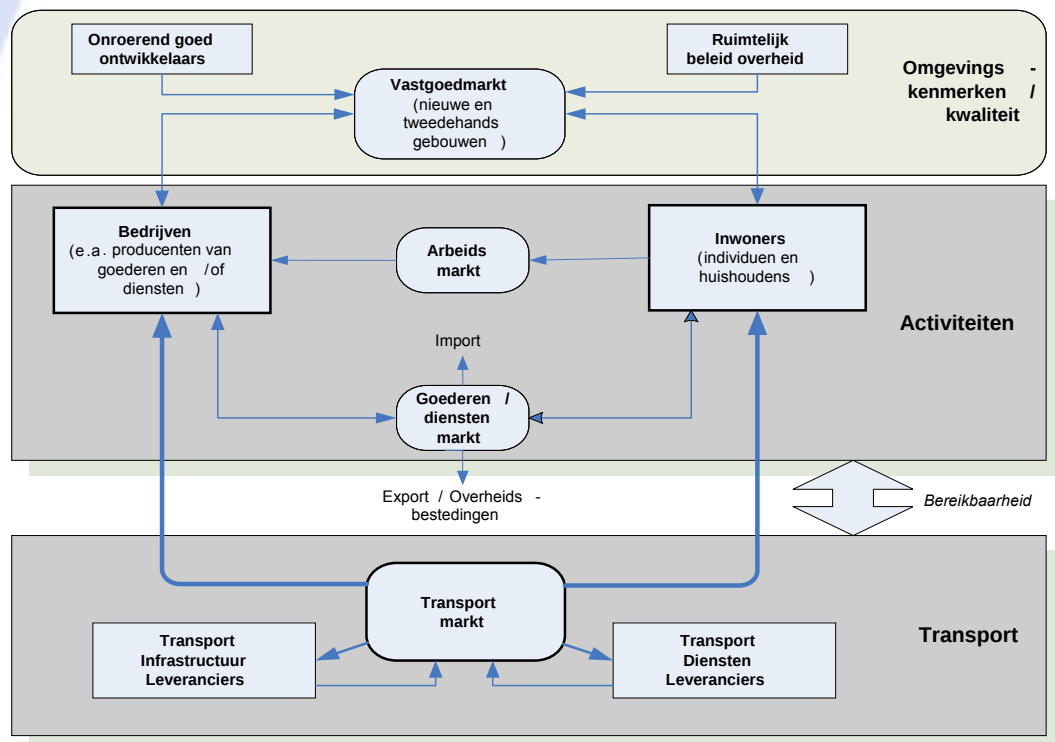
Er heeft sinds de zestiger jaren onderzoek plaatsgevonden naar de interactie tussen landgebruik en transport. Dit heeft geleid tot de ontwikkeling van een aantal Land Use and Transport Interaction (LUTI) modellen, elk met eigen karakteristieken. Toch heeft dit niet geleid tot één alom aanvaarde en eenduidige theorie over LUTI modellering. Timmermans (2003) concludeert dat op dit moment de manier waarop de interactie tussen landgebruik en transport wiskundig gemodelleerd wordt niet voldoende consistent is om in redelijke mate aan te sluiten op de huidige gangbare theorieën over de kosten/baten van transport. Ondanks dat de LUTI modellen in de loop van de jaren sterk verbeterd zijn, ondermeer door de beschikbaarheid van geografische informatie systemen (GIS), blijven de onderliggende theorieën nog steeds zwak en zijn de oude fundamentele problemen nog steeds niet opgelost.

2.1.1 Achterliggende verbanden LUTI modellen

Verkeer en vervoer zijn voor het overgrote deel afgeleide activiteiten; ze zijn het gevolg van het feit dat we vele ruimtegebonden activiteiten (zoals wonen, werken, winkelen, bezoeken van familie en vrienden) op verschillende locaties uitvoeren. De afstanden tussen deze locaties dienen te worden overbrugd, hetgeen resulteert in verkeer en vervoer van personen, per auto, openbaar vervoer, (brom)fiets of anderszins. Daarmee is er een directe link tussen de ruimtelijke structuur van deze ruimtegebonden activiteiten en het vervoer van personen.

Goederenvervoer vloeit voort uit het feit dat productie en consumptie/gebruik ruimtelijk gescheiden zijn en uit het feit dat de productiefase uiteen valt in een groot aantal deelactiviteiten, variërend van grondstoffenwinning via de productie van halffabrikaten tot de productie van eindproducten, welke ruimtelijk gescheiden plaatsvinden (Van Wee, 1993).

Er is een aantal theorieën ontwikkeld die verbanden leggen tussen landgebruik factoren om zo de mobiliteitsbewegingen te verklaren. Onder andere Everaars en Brand-van Tuijn (2001) en Van de Vooren en Pauwels (2002) gaan in hun onderzoeken hierop in. Daarnaast heeft ook het Britse Ministerie van Transport (2003) door middel van een visueel concept dit inzichtelijk proberen te maken. In figuur 2.1 is een destillatie gemaakt uit de verschillende theorieën tot één totaaloverzicht.



Figuur 2.1: Actoren, markten en hun interacties in LUTI modellen

De twee belangrijkste subsystemen in het landgebruik en transport interactie model zijn, zoals ook uit figuur 2.1 valt af te lezen: activiteiten en transport. Binnen elk subsysteem is onderscheid gemaakt tussen de vraag en aanbod elementen die elkaar dusdanig beïnvloeden dat er een evenwicht bereikt wordt.

De locatie en interactie van activiteiten stellen de vraagzijde van het activiteiten subsysteem voor. Activiteiten veroorzaakt door bedrijven en inwoners gebeuren op specifieke plaatsen en beïnvloeden andere activiteiten om hun functies uit te voeren. De interactie tussen inwoners en bedrijven vindt plaats op de arbeidsmarkt en de goederenmarkt. Beide actoren gebruiken ook land en vloerruimte om hun functies uit te voeren. Deze ruimtes (denk daarbij ook aan de kwaliteit van de ruimtes) worden beschikbaar gesteld door ontwikkelaars op de vastgoedmarkt. Dit stelt de aanbodzijde voor in een bepaald gebied. Dit aanbod verschilt per gebied en is dus omgevingsafhankelijk. De interactie tussen deze twee elementen leidt tot een vastgoedmarkt die gereguleerd wordt door grondprijzen en onroerend goed prijzen. Daarnaast heeft de overheid invloed op deze markt door middel van subsidies of vergunningen.

Naast het feit dat activiteiten zorgen voor vraag naar land(ruimte), is er ook transport nodig voor de interactie tussen activiteiten. In het transport subsysteem wordt de vraag gezien als de behoefte aan reizen. Deze behoefte bestaat voor mensen uit het reizen naar hun werkplekken of naar diensten en voor goederen uit het bereiken van de plek van consumptie die ergens anders gesitueerd is dan de productieplaats. Een onderscheid wordt gemaakt tussen de fysieke beschikbaarheid van transport, zoals wegen, rails en vaarwegen en degenen die hiervan gebruiken maken om het transport daadwerkelijk te laten geschieden. Dit zijn de diensten leveranciers, zoals transportbedrijven, busbedrijven, luchtvaartmaatschappijen, of zelfs auto's en voetgangers.

Concreet kan figuur 2.1 vertaald worden in het volgende voorbeeld: stel dat in een gebied een nieuwe woonwijk wordt gebouwd (toename aanbod op vastgoedmarkt). Dit zorgt ervoor dat er meer mensen naar het gebied aangetrokken worden, wat leidt tot meer activiteiten in het gebied. Meer mensen heeft namelijk een toename in de vraag naar goederen en diensten tot gevolg, wat leidt tot meer bedrijvigheid in het gebied. Voor meer bedrijvigheid zijn meer arbeidsplaatsen nodig, wat opgevangen wordt door het grotere aanbod van mensen op de arbeidsmarkt. Er waren in het gebied immers meer mensen gaan wonen. Dit alles zorgt voor meer activiteiten, wat op haar beurt leidt tot een grotere vraag naar transport. Als reactie hierop kan de overheid besluiten het aanbod van transport te vergroten, door nieuwe infrastructuur aan te leggen of bestaande uit te breiden. Dit heeft een positieve invloed op het aantal activiteiten van bedrijven en inwoners. Meer bedrijvigheid zorgt daarnaast voor een grotere vraag naar bedrijfspanden op de vastgoedmarkt en voor het aantrekken van nieuwe mensen van buiten het gebied. Hiermee is de cirkel rond. Een dergelijk proces van opeenvolgende gebeurtenissen kan bij elk willekeurig hierboven genoemd onderdeel starten.

2.1.2 Verbindende factor activiteiten en transport: bereikbaarheid

Het activiteiten subsysteem en het transport subsysteem zijn onderling ook in evenwicht met elkaar. Het evenwicht in vraag en aanbod wordt bepaald door het begrip bereikbaarheid. Een goede bereikbaarheid zorgt voor meer activiteiten, terwijl meer activiteiten zorgen voor een grotere vraag naar transport, wat indirect weer resulteert in een afname van de bereikbaarheid.

Bereikbaarheid wordt algemeen gedefinieerd als de mate waarin de ruimtelijk-infrastructurele constellatie mensen (of goederen) in staat stelt om activiteiten (of bestemmingen) te bereiken met een bepaalde (combinatie van) vervoerswijze(n) (Geurs et al, 2001).

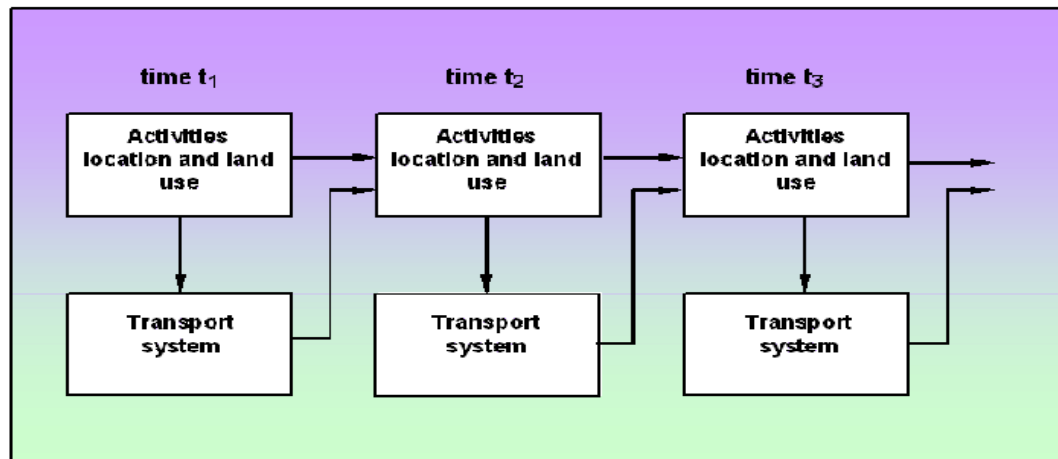
De manier waarop de bereikbaarheid het evenwicht reguleert tussen de activiteiten in een gebied en het benodigde transport hangt af van de manier waarop naar bereikbaarheid gekeken wordt. De verschillende benaderingen hebben in meer of mindere mate de verschillende (en onderling samenhangende) componenten van bereikbaarheid in zich: (a) een transport component, die de reistijd, -kosten en moeite beschrijft, en de waardering ervan door personen en bedrijven, (b) een ruimtelijke component, die de omvang en ruimtelijke spreiding van activiteiten weergeeft, (c) een tijdscomponent, die de tijdsrestricties van activiteiten (bijv. openingstijden) en personen weergeeft, en (d) een individuele component, die de individuele behoeften, capaciteiten en mogelijkheden weergeeft (de "nut" ervaring van mensen) (Geurs et al, 2001).

Uit bovenstaande blijkt dat activiteiten en transport elementen zijn die afhankelijk van elkaar zijn en elkaar wederzijds beïnvloeden. De interactie tussen activiteiten zorgt voor een vervoersvraag. Het transportevenwicht beïnvloedt op haar beurt de locatie en interactie van activiteiten en zodoende indirect ook de onroerend goed markt.

2.1.3 Dynamisch karakter LUTI model

Belangrijk kenmerk van een LUTI model is het feit dat de manier waarop de twee hoofdelementen elkaar beïnvloeden, landgebruik(activiteiten) en transport, dynamisch van aard is. Figuur 2.2 laat zien dat de vervoersvraag veroorzaakt door activiteiten in dezelfde tijdsperiode het aanbod van transport beïnvloedt. Het vraag-/aanbodevenwicht in het transportsysteem bepaald de bereikbaarheid tussen locaties in dezelfde tijdsperiode. Deze feedback heeft echter pas op een

later tijdstip, t_2 , invloed op het systeem. Hierdoor beïnvloedt de bereikbaarheid op tijdstip t_2 ook de verdeling van vervoersstromen. Op deze manier kan het enkele periodes duren, mede doordat verplaatsing van activiteiten vaak met vertraging geschiedt, voordat de verandering van het transportevenwicht volledig is verwerkt in het totale systeem (Modelistica, 2005).



Figuur 2.2: Dynamische relaties tussen activiteiten en transport (Modelistica, 2005)

Figuur 2.2 impliceert dat er te allen tijde een dynamische relatie tussen activiteiten en transport bestaat. Dit hoeft echter niet het geval te zijn. Zo kan bijvoorbeeld een sterke stijging in transportkosten (bijv. door hogere benzineprijzen), zorgen voor minder vraag naar transport. Dit heeft een gunstig effect op de bereikbaarheid. De betere bereikbaarheid zorgt alleen op tijdstip t_2 niet voor meer activiteiten. Het uitvoeren van deze activiteiten is namelijk minder aantrekkelijk geworden door de hogere transportkosten. Ook een toename in activiteiten hoeft niet in alle gevallen te leiden tot meer activiteiten op tijdstip t_2 . In een gebied met een grote restcapaciteit op de wegen zullen meer activiteiten wel leiden tot meer vraag naar transport via onder andere de weg, maar leidt dit niet tot een noemenswaardige daling in de bereikbaarheid. Dit betekent dat op tijdstip t_2 de activiteiten niet merkbaar beïnvloedt worden door het transportsysteem.

2.1.4 Toepassingsmogelijkheden LUTI modellen

Zoals eerder al vermeld worden LUTI modellen nog niet op grootschalige wijze toegepast in strategische transportplanning. Belangrijkste oorzaak hiervan is het feit dat er nog geen LUTI modellen ontwikkeld zijn die in staat zijn de werkelijkheid op een aanvaardbare wijze te benaderen. Dit wordt veroorzaakt door zwakke onderliggende theorieën. Het is op dit moment het grootste nadeel van LUTI modellen.

De belangrijkste voordelen van LUTI modellen:

- Het maakt het mogelijk om op een strategische tijdschaal voorspellingen te doen over de invloed van investeringen in nieuwe infrastructuur in een gebied.
- Legt een link tussen transport en landgebruik, waartussen, wat algemeen aangenomen wordt, een verband bestaat. Dit element ontbreekt bij veel andere traditionele modellen.

Naast het al eerder genoemde gebrek aan een goede theoretische onderbouwing zijn er nog enkele nadelen te noemen ten aanzien van LUTI modellen:

- Het black box probleem: zelfs degenen die het model hebben ontworpen begrijpen niet wat er binnen het model gebeurt. Er is geen duidelijke begrip van de verbanden/relaties die verantwoordelijk zijn voor de uiteindelijke resultaten.
- Niet generaliseerbaar: een ideaal model zou in voor elk doel generaliseerbaar moeten zijn, zodat het in elke landgebruik en transport situatie/omstandigheid toegepast kan worden. Deze aanname blijkt bij ontwikkelde modellen echter sterk betwifelbaar (Lee, 1973). Er bestaat een gebrek aan gestandaardiseerde theorieën en modellen die van toepassing kunnen zijn op elke stad of gebied. Het is gebleken dat het type stad en transport systeem van invloed is op de uitkomsten als geïntegreerde modellen worden gebruikt (Tillema, 2004).
- Beperkte kennis over dynamiek en tijdschalen, terwijl de modellen dynamische beslissingen proberen te voorspellen. De dynamiek van de interactie tussen activiteiten en transport is in paragraaf 2.1.3 besproken.

2.2 Locatiekeuzes van bedrijven

Het onderdeel van het LUTI model waar dit onderzoek zich op richt zijn de motieven van bedrijven waarop zij hun locatiekeuzes baseren. Er is ontzettend veel onderzoek gedaan naar dit onderwerp, vanuit verschillende wetenschappelijke vakgebieden. De ontwikkelde theorieën worden ook wel aangeduid als vestigingsplaatsleer of locatietheorie. In de volgende paragrafen worden de belangrijkste bevindingen uit deze onderzoeken behandeld.

Voordat bedrijven daadwerkelijk kiezen om zich in een andere regio te vestigen, dan wel een nieuwe vestiging in een andere regio te openen, hebben ze een aantal andere opties om de bedrijfsactiviteiten aan te passen op de huidige locatie. Dicken (1986) hanteert het onderstaande overzicht, waarin schematisch opgesomd wordt welke keuzes bedrijven hebben om zich aan te passen in veranderende omstandigheden, oftewel hoe ze kunnen omgaan met mogelijke kansen of bedreigingen.



Figuur 2.3: Locatiebeslissingen van bedrijven (Dicken, 1986)

Dicken hanteert twee typen van locatiebeslissingen, te weten “in situ” veranderingen en locatieverschuivingen. Locatiebeslissingen op de huidige locatie komen vaker voor dan locatiebeslissingen die betrekking hebben op een nieuwe locatie. Met betrekking tot de activiteiten op de huidige locatie zijn er drie mogelijkheden. Deze zijn: Uitbreiding van, Vervanging van en Inkrimping van activiteiten(het afbreken of buiten werking stellen van bestaande gebouwen of productielijnen).

Als een locatieverschuiving wordt overwogen, bestaan de opties uit: Investerings in nieuwe gebouwen en productielijnen op een andere locatie, overname van (een productie-unit van) een ander bedrijf en sluiting. De meest complexe locatiebeslissing betreft de relocatie, een verplaatsing van het gehele bedrijf. Van de locatiebeslissingen wordt volgens Dicken dit type beslissing het minst overwogen.

Van belang is te onderkennen dat bij de locatiebeslissingen van bedrijven niet alleen aan relocatie moet worden gedacht, maar dat bedrijven op hun huidige locatie ook vele mogelijkheden hebben om op veranderende omstandigheden te reageren. De vraag is nu welke omstandigheden van belang zijn voor bedrijven.

2.2.1 Belangrijkste factor locatiekeuzes bedrijven

De belangrijkste sturende variabele die de locatiekeuze van bedrijven beïnvloedt zijn de transportkosten, ook wel uitgedrukt in potentiële bereikbaarheid (Tillema, 2004) (Hayter, 2004) (Van Wee, 1993).

Potentiële bereikbaarheid wordt gedefinieerd als het aantal bestemmingen (bedrijven) die een persoon kan bereiken, gedeeld door de mate van inspanning voor de betreffende persoon om deze bestemming (bedrijven) te bereiken (Geurts en Ritsema van Eck, 2001).

Ondanks het feit dat de potentiële bereikbaarheid de belangrijkste beweegreden voor bedrijven is om zich te vestigen in een bepaalde regio constateert Van Wee (1993) dat de verwachte situatie in de praktijk niet hiermee overeenstemt. Om dit te bewijzen definieert Van Wee het mobiliteitsprofiel van een bedrijf of een publieke voorziening als het potentiële openbaarvervoer gebruik van werknemers en bezoekers van dat bedrijf of die voorziening. Daarnaast definieert Van Wee een bereikbaarheidsprofiel van een locatie als de bereikbaarheid van die locatie per openbaar vervoer en auto. Ook de parkeermogelijkheden vormen een deel van het bereikbaarheidsprofiel. De essentie van het onderzoek is, dat de bereikbaarheidsprofielen van locaties en de mobiliteitsprofielen van bedrijven op elkaar afgestemd dienen te zijn. Uit casestudies blijkt echter dat in de praktijk er grote verschillen bestaan. Slechts de helft van de bedrijven heeft zich op een theoretisch geschikte locatie gevestigd. Blijkbaar zijn er meer factoren die meespelen in de locatiekeuze van bedrijven.

2.2.2 Andere factoren locatiekeuzes bedrijven

Zoals eerder in dit hoofdstuk al vermeld is er veel onderzoek verricht naar de locatiekeuzefactoren van bedrijven. Er is om deze reden door TNO in samenwerking met de Universiteit Twente en de Rijksuniversiteit Groningen een overzicht gemaakt van de bestaande theorieën en onderzoeken over deze locatiekeuzefactoren en wellicht even belangrijk een onderscheid in keuzegedrag tussen

ondernemers met een verschillende achtergrond. Hierbij moet worden gedacht aan verschillen tussen ondernemers met betrekking tot sector, markt, strategie en schaal (Reijs et al, 2001).

De belangrijkste conclusies die Reijs trekt en die gelden voor het gehele Nederlandse bedrijfsleven, zijn als volgt: de meest essentiële eis die Nederlandse bedrijven aan hun omgeving stellen is bereikbaarheid. Deze eis wordt steeds vaker als essentiële eis genoemd. Dit is ook in overeenstemming met de bewering van in de vorige paragraaf vermelde bronnen. De vijf belangrijkste vestigingscriteria zijn respectievelijk bereikbaarheid over de weg, parkeergelegenheid voor personeel en bezoekers, beschikbaarheid van adequaat opgeleid personeel, representativiteit van het gebouw en de aanwezige telecommunicatievoorzieningen.

Een ander onderzoek dat veel duidelijk maakt over de voorkeuren van ondernemers is het werk van Hanemaayer (1998). In tabel 2.1 zijn de negen factoren opgenomen die volgens Hanemaayer's onderzoek het belangrijkste voor ondernemers zijn. In bijlage A staat een overzicht van alle factoren die Hanemaayer in zijn onderzoek heeft meegenomen. Belangrijke details uit tabel 2.1 zijn de verschillen tussen de geselecteerde bedrijfssectoren. Dit lijkt erop te wijzen dat er wel degelijk verschillen bestaan tussen locatievoorkeuren van ondernemers uit verschillende bedrijfssectoren.

Aspect/cluster	Industrie/Produc.	Handel	Verkeer/Vervoer	Diensten	Totaal
Bereikbaarheid over de weg	8,6	8,6	8,9	7,7	8,3
Parkeermogelijkheden	7,3	7,4	7,1	8,0	7,5
Beschikbaarheid adequaat opgeleid personeel	7,2	6,2	6,9	7,4	7,0
Representativiteit gebouw	6,1	6,4	6,5	7,3	6,6
Telecommunicatievoorzieningen	6,0	6,4	6,9	6,1	6,2
Laad- en losmogelijkheden op terrein	7,3	7,6	7,0	3,0	5,9
Huurprijs- of grondprijsniveau	5,2	5,4	6,0	5,3	5,4
Representativiteit omgeving	4,8	5,0	5,3	5,6	5,2
Bereikbaarheid via openbaar vervoer	4,3	4,3	4,1	5,5	4,7

Tabel 2.1: Waardering belangrijkste locatiefactoren (schaal 0 (=onbelangrijk) tot 10 (zeer belangrijk)) (Hanemaayer, 1998)

Het is van belang om op een aantal factoren dieper in te gaan. Opvallend is dat in veel onderzoeken voor hetzelfde begrip verschillende definities gehanteerd worden. Dat betekent dat voorzichtig moet worden omgegaan met vergelijkingen tussen onderzoeken.

Parkeerfaciliteiten; het belang hiervan is ondermeer aangetoond in een studie van Teunissen (1992) naar bedrijfsverplaatsingen conform het locatiebeleid in Amsterdam. In andere onderzoeken wordt echter aangetoond dat de rol van parkeergelegenheid minder groot is dan gedacht (Van Wee, 1993). Ook Everaars en Brand-van Tuijn (2001) tonen aan dat een goede OV verbinding leidt tot minder verplaatsingen met de auto wat het belang van parkeerfaciliteiten doet afnemen.

Telecommunicatievoorzieningen; is een belangrijke locatiefactor. Bedrijven hechten de laatste jaren steeds meer belang aan dit type voorziening, waarschijnlijk voornamelijk onder invloed van de vele nieuwe ontwikkelingen op dit gebied.

Beschikbaarheid adequaat opgeleid personeel (kennispotentiaal); er bestaat de laatste jaren een toenemende behoefte aan hoog opgeleide arbeidskrachten en is daarom een belangrijke locatiefactor. Zij worden aangetrokken door een brede range van socio-culturele en vrijetijdsvoorzieningen, wat leidt tot een toenemend belang van quality-of-life (QOL) factoren (Reijs et al, 2001). De factor beschikbaarheid van adequaat opgeleid personeel hangt dus nauw samen met de quality-of-life factoren in een gebied. Het belang van de omgevingskwaliteit; de leefkwaliteit, de gesteldheid van de gezondheidszorg, sociaal-culturele, vrije tijd- en sportvoorzieningen, scholen, kwaliteit lucht, water, omgeving etc. is dus indirect ook groot voor bedrijven.

Huurprijs/ grondprijsniveau; een analyse van de grondkosten is bijzonder complex. Er bestaan grote verschillen tussen locaties in grondsoorten, de bijbehorende eigenschappen en kwaliteit en het lokale beleid van de overheid. Dit komt tot uiting in een gesegmenteerde grondmarkt waar verschillende prijzen gehanteerd worden voor “groene” (landbouw en natuur) en “rode” (wonen, werken en infrastructuur) functies en waar per sector weer grote ruimtelijke prijsverschillen voorkomen (Rietveld en Koonen, 2004).

Daarnaast blijkt dat, ondanks dat deze factor niet in het onderzoek van Hanemaayer naar voor komt, in voor sommige bedrijven de factor milieuaspecten ook een rol speelt. Dit geldt vooral bij bedrijven met hoge milieulasten. Deze factor wordt echter vaak in combinatie met andere factoren genoemd en niet vaak als hoofdfactor aangewezen. Zo zal een grote investering in nieuwe innovatieve milieuvriendelijke oplossingen niet alleen gebaseerd zijn op het beperken van de milieukosten, maar ook ter verbetering van de concurrentiepositie (Van der Veen, 2004). De factor milieuaspecten wordt dus niet vaak als hoofdzaak gezien en dit is waarschijnlijk ook de reden dat het minder nadrukkelijk uit andere onderzoeken naar voren komt.

In deze paragraaf zijn een aantal belangrijke factoren behandeld waarvan is bewezen dat deze van invloed zijn op de locatiekeuzes van bedrijven. De bevindingen van de onderzoekers lijken in grote lijnen met elkaar overeen te komen. Als er specifiek naar bepaalde factoren wordt gekeken, blijken er toch onduidelijkheden te bestaan. De oorzaken hiervan zijn te vinden in het feit dat bepaalde factoren elkaar deels versterken dan wel verzwakken, de interventie van de overheid op sommige gebieden en indirecte factoren, zoals de omgevingskwaliteit. Daarnaast lijken er duidelijke verschillen te bestaan tussen locatievoorkeuren van ondernemers uit verschillende bedrijfssectoren. Om deze reden zal in paragraaf 2.2.3 nader worden ingegaan op bepaalde verschillen tussen ondernemers met een verschillende achtergrond.

2.2.3 Locatievoorkeuren per bedrijfssector

Everaars en Brand-van Tuijn (2001) tonen aan dat het type bedrijf van de ondernemer van invloed is op waarde die de betreffende ondernemer aan de algemene locatiefactoren geeft. Als voorbeeld stellen zij dat over het algemeen een transportbedrijf in zijn locatiekeuze bereikbaarheid over de weg zwaarder zal laten meewegen dan de uitstraling van het bedrijfspand. Zo zullen ook de milieukosten en de factor laad/los capaciteiten voor de sector industrie & productie van veel groter belang zijn dan voor bijvoorbeeld de sector diensten. Er zijn echter vele indelingen mogelijk om bedrijven in een bepaalde groep neer te zetten. Tillema (2004) maakt in zijn onderzoek onderscheid tussen drie bedrijfstypen, te weten industrie, commercie en andere sectoren. Reijs (2001) en min of meer op dezelfde manier ook Everaars en Brand-van Tuijn (2001) gaan op hun beurt in op de specifieke locatiekeuzefactoren aan de hand van een vijftal bedrijfstypen op regionaal niveau, te weten industrie- en productieondernemingen, verkeer en

vervoersondernemingen, handelsbedrijven, dienstverlenende bedrijven en hoofdkantoren. Deze specifieke locatiekeuzefactoren zijn hieronder per bedrijfstype in tabel 2.2 neergezet.

	Uitbreidings- mogelijkheden	Laad- en los- mogelijkheden	Aanwezigheid toeleveranciers	Openbaar vervoer	Logistieke dienstverlening	Huur- en grond prijsniveau	Bereikbaarheid	Aanwezigheid zee/luchthaven	Internationale bedrijvigheid	QOL factoren	Parkeer- mogelijkheden	Kwaliteit/imago gebouw	Uitstraling omgeving	Telecomm. voorzieningen
Industrie/ Productie	x	x	x		x									
Verkeer en Vervoer		x			x	x	x	x	x					
Handel		x			x									
Diensten				x						x	x	x	x	
Hoofdkantoren								x		x				x

Tabel 2.2: Factoren lokatiekeuze per bedrijfssector (Reijs, 2001 & Everaar en Brand-van Tuijn, 2001)

Nederlandse industriële en productieondernemingen hechten een meer dan gemiddeld belang aan de factoren uitbreidingsmogelijkheden, laad- en losmogelijkheden, toeleverende bedrijven en logistieke dienstverlening. De locatiekeuze van een industrie of productieonderneming hangt af van de gekozen concurrentiestrategie. Ook de fase van een product in de levenscyclus kan een bepalende factor in de locatiekeuze zijn. Zo zal een bedrijf met een relatief nieuw product met veel groeipotentieel en een agressieve concurrentiestrategie, de factor uitbreidingsmogelijkheden extra belangrijk vinden.

Voor bedrijven uit de verkeer- en vervoersector blijkt dat de factoren huurprijs- en grondprijsniveau, bereikbaarheid over de weg, de internationale bedrijvigheid en de nabijheid van een zee- of luchthaven en vaarwater van relatief meer belang zijn dan in andere sectoren. Everaars en Brand-van Tuijn voegen daar nog de factoren laad- en losmogelijkheden op het terrein en de logistieke dienstverlening aan toe.

Voor de Nederlandse handelsondernemingen zijn de factoren laad- en losmogelijkheden op het terrein en de logistieke dienstverlening relatief van meer belang.

Professionele dienstverleners vestigen en zullen zich in de stedelijke gebieden blijven vestigen. De nabijheid van de klanten is van belang, maar ook de aanwezigheid van hoog opgeleid personeel. Hoog opgeleid personeel wordt aangetrokken door QOL factoren. Nederlandse dienstverlenende bedrijven hechten relatief veel belang aan parkeermogelijkheden, de kwaliteit van het gebouw, de uitstraling van de omgeving en de bereikbaarheid via het openbaar vervoer.

De hoofdkantoren blijven zich in de toekomst vestigen in grootstedelijke gebieden (Reijs, 2001). Het percentage executives en professionals dat op een hoofdkantoor werkt, is relatief hoog ten opzichte van het gemiddelde. Dit betekent dat er relatief veel hoog opgeleide mensen werkzaam zijn in deze sector, waardoor de QOL factor van belang is. Het verband tussen deze twee factoren is in paragraaf 2.2.2 uitgelegd. De nabijheid van luchthaven en de kwaliteit van de telecommunicatievoorzieningen zijn belangrijke factoren.

Het is nu duidelijk wat de verschillen zijn bij een indeling van bedrijven in sectoren. Er zijn echter meerdere mogelijkheden om bedrijven in te delen naar hun aard. Zo zijn er waarschijnlijk verschillen in locatiekeuzes tussen binnenlandse en van oorsprong buitenlandse bedrijven aan te wijzen, een indeling gehanteerd door het ministerie van Economische Zaken (2003). Een andere optie die hierop aansluit is de indeling die door Mooij (2004) wordt gehanteerd. Hier worden bedrijven ingedeeld in de categorie multinationals en midden- en kleinbedrijf (MKB). Als laatste dient in dit rijtje de indeling van Maanen, Verroen en Heerema (1992) genoemd worden, die een indeling maken in “hoogwaardige bedrijven”, bedrijven met relatief veel werknemers met een hoog inkomen en “laagwaardige bedrijven”, bedrijven met relatief veel werkers met een laag inkomen en een hoog aantal vrouwen. Het lijkt erop dat al deze indelingen verschillen in locatiekeuzegedrag tussen de gecreëerde categorieën opleveren. Dit betekent dat bij de herdefinitie van het bedrijfslocatiemodel, als output verschillende indelingen van bedrijfstypen gehanteerd kunnen worden. De voordelen van een indeling in bedrijfssectoren zijn dat voor deze indeling uitgebreid onderzoek gedaan is en dat er duidelijke verschillen aangetoond zijn. Het lijkt daarom voor de hand te liggen om in dit onderzoek hierop voort te borduren en te kiezen voor een indeling in sectoren in het bedrijfslocatiemodel.

2.2.4 Locatievoorkeuren per regio

Deze paragraaf gaat over de waardering van plaatskenmerken in Nederland, los van de andere factoren waarin een dergelijke locatie goed of slecht scoort. Schotten (1997) zegt over de attractiviteitsbepaling van vestigingsplaatsen het volgende: fysische geschiktheid, potentiaal (=optelsom van positieve en negatieve zaken die zich in de buurt bevinden, waarbij kleine afstanden grotere invloed hebben) en beleid bepalen ieder een deel van de aantrekkelijkheid van een gebied voor een bepaalde functie.

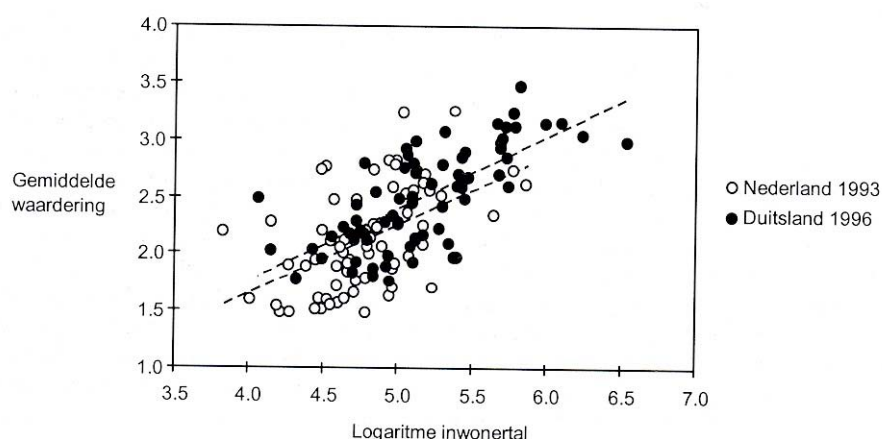
De factor vestigingsplaats

Meester (1999) heeft uitgebreid onderzoek verricht naar de waardering van plaatskenmerken in Nederland, los van de andere factoren waarin een dergelijke locatie goed of slecht scoort. Er worden een viertal te onderzoeken plaatskenmerken c.q. vestigingsplaatsfactoren gedefinieerd; ligging, agglomeratie-effecten, infrastructuur, ligging ten opzichte van de externe markt. Om dit te meten introduceert Meester een viertal variabelen om deze waardering te meten.

- *Ligging*; geoperationaliseerd als de afstand van de plaats in kwestie tot de plaats die uit het onderzoek als hoogst gewaardeerde centrale plaats naar voren gekomen is (voor Nederland is dit Utrecht).

Een voorbeeld van het belang van de ligging volgt uit de volgende discussie: er wordt nogal eens gepleit voor het beter verbinden van perifere regio's met kernregio's om de eerste in de vaart der volkeren op te stuwen. Zo is er de discussie over de magneetweefbaan van de Randstad naar Groningen die beoogt Groningen aantrekkelijker te maken als locatie voor bedrijvigheid. Soortgelijke motieven spelen bij de HSL-Zuid van Amsterdam naar Parijs, die Nederland met het hart van Europa moet verbinden. Inzichten uit de nieuwe economische geografie suggereren dat een dergelijk beleid contraproductief kan zijn voor de perifere regio (Martin en Rogers, 1995).

- *Agglomeratie-effecten*; geoperationaliseerd als het logaritme van het inwonertal van de betreffende agglomeratie. Deze veronderstelling wordt ondersteund door figuur 2.4.



Figuur 2.4: Inwonertal en gemiddelde waardering plaatsen (Meester, 1999)

Mooij (2004) legt dit agglomeratie-effect als volgt uit: de belangrijkste factor waarmee bedrijven rekening houden zijn de transportkosten. Als grotere afstanden zijn gemoeid met het vervoer van goederen, betekent dit meer kosten. Transportkosten impliceren dat clustering van bedrijvigheid voordelig kan zijn voor alle bedrijven. Omdat individuele bedrijven geen rekening houden met het voordeel van hun aanwezigheid op andere bedrijven, is hier sprake van externaliteiten, ook wel agglomeratie-externaliteiten genoemd. Daarbij zijn niet alleen transportkosten relevant. Men kan ook denken aan bijvoorbeeld leereffecten: in het productieproces dragen bedrijven bij aan een lokale voorraad kennis waar andere bedrijven van kunnen profiteren. Andere voorbeelden zijn gereduceerde zoekkosten op gezamenlijke arbeidsmarkten (toevallige positieve en negatieve fluctuaties in vraag of aanbod naar arbeid heffen elkaar op) en het verschijnen van een markt voor gespecialiseerde halffabrikaten. Het gevolg is dat de economieën door toenemende meeropbrengsten worden gekarakteriseerd: meer bedrijvigheid verhoogt de productiviteit van andere bedrijven.

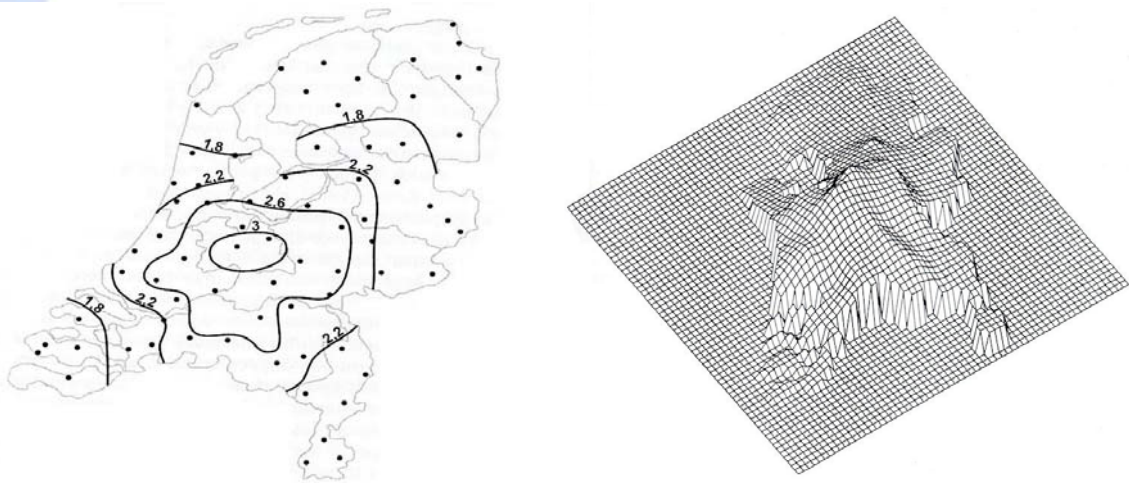
De andere twee variabelen die Meester (1999) voor de waardering van vestigingsplaatsen introduceert zijn:

- *Infrastructuur*; geoperationaliseerd als een binaire variabele die aangeeft of de betreffende plaats wel of niet is aangesloten op het nationale net van autosnelwegen.
- *Ligging t.o.v. de externe markt*; geoperationaliseerd door de kaartvariabelen van een locatie als onafhankelijke variabelen in het model op te nemen.

Meester concludeert met behulp van een regressieanalyse dat de waardering van plaatsen door ondernemers een belangrijke rol speelt in de waardering van locaties. Bij een goede analyse van locatiekeuzes van bedrijven betekent dit ook dat deze factoren mee dienen te worden genomen.

Subjectieve waardering regio

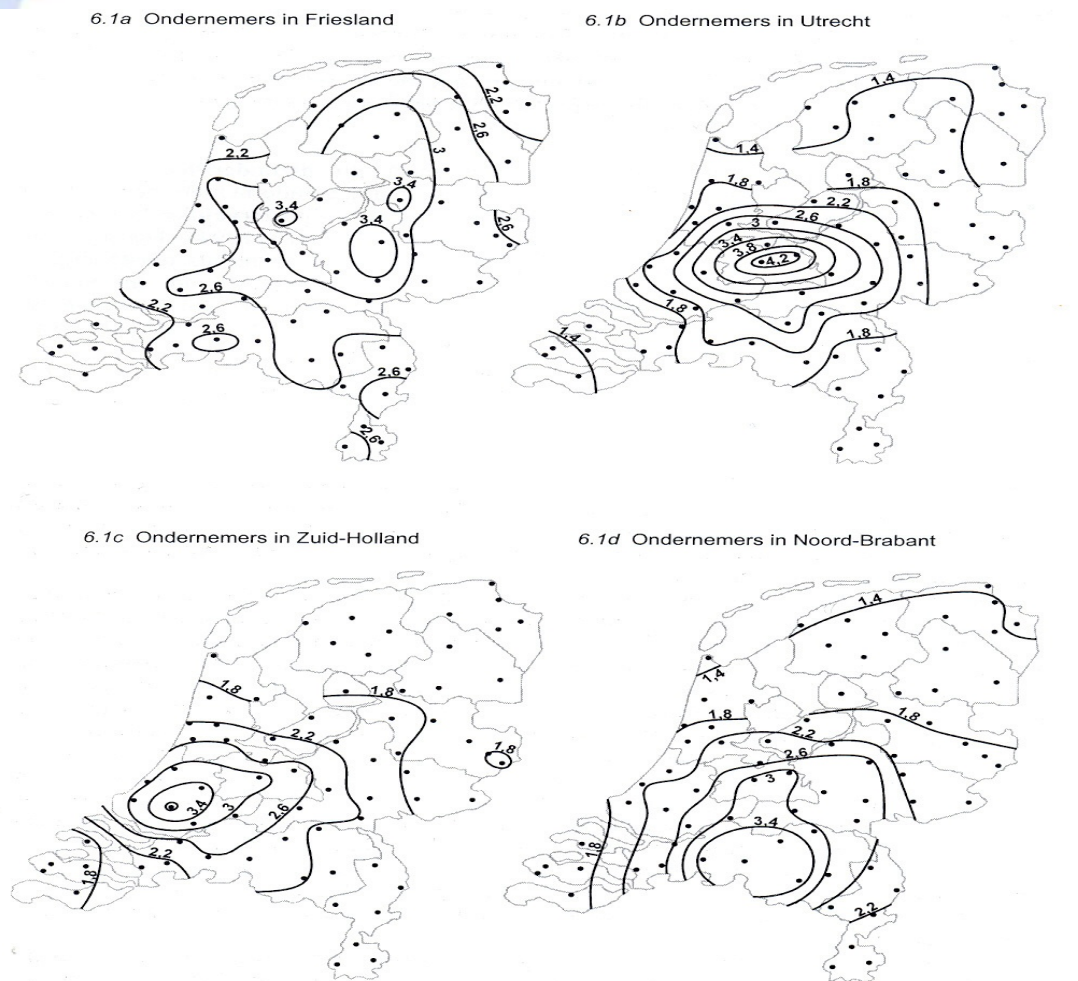
Meester (1999) toont in zijn proefschrift een overzicht van de gemiddelde waardering van vestigingsplaatsen in Nederland uit een enquête uit 1993. In figuur 2.5 staat dit overzicht.



Figuur 2.5: Gemiddelde waardering vestigingsgebieden Nederland in 1993 (links uitgedrukt m.b.v. isoplethen, rechts driedimensionaal uitgedrukt) (Meester, 1999)

Figuur 2.5 is een vrij globale weergave van de waardering van vestigingsgebieden in Nederland. Het blijkt dat de regio Utrecht de meest gewaardeerde vestigingsplaats is. De vraag is of er een verband bestaat tussen ruimtelijk schaalniveau en de waardering van vestigingsmilieus. Zo gaat Meester nader in op Noord Nederland. De waarderingspatronen blijken als gevolg van de detaillering complexer te worden. Het patroon zoals dat voor Nederland als geheel is gevonden, is nog steeds zichtbaar in de vorm van het aflopen van de waarderungen van de zuidrand van Noord-Nederland als deelgebied naar de uiterste noordelijke periferie (buitenkant, buitenste gebied), maar bovenop deze algemene structuur is, als nieuw element, een patroon te zien van hoge waarderungen voor de grotere agglomeraties binnen Noord-Nederland.

Een belangrijke reden hiervoor is dat de actuele vestigingsplaats als bedrijfskenmerk van grote invloed is op de ruimtelijke voorkeuren van ondernemers. Dit geldt voor alle regio's in Nederland. Ondernemers waarderen de huidige eigen regio altijd relatief hoog. Dit lijkt ook logisch te verklaren uit het feit dat vooral voor kleinere ondernemers de lokale markt van belang is, waarmee zij ook het meest bekend en vertrouwd zijn. Figuur 2.6 verduidelijkt dit beeld.



Figuur 2.6: Waardering vestigingsgebieden Nederland in 1993, naar herkomstgebied (Meester, 1999)

Daarnaast heeft Meester nog drie bedrijfskenmerken onderzocht en hun invloed op ruimtelijke voorkeuren van Nederlandse ondernemers. Het kenmerk grootteklasse heeft geen aantoonbare invloed, type activiteit en omvang van het afzetgebied daarentegen wel. De invloed van de sector uit zich vooral in een relatief hoge waardering van grotere agglomeraties door bedrijven in de zakelijke dienstverlening, en een relatief sterke voorkeur van de groothandel voor het centrale deel van het land. In invloed van de omvang van het afzetgebied uit zich in het feit dat de ruimtelijke voorkeuren minder geprononceerd lijken te zijn naarmate het afzetgebied groter is. Bedrijven met een internationale markt hebben een minder sterke voorkeur voor centrale locaties dan bedrijven met een nationale markt.

Uit bovenstaande wordt het belang bevestigd dat het nodig is om het bedrijfslocatiemodel van Tillemans te onderzoeken op de toepasbaarheid op elke willekeurige locatie in Nederland. Niet automatisch mag worden aangenomen dat als het model, waarvan in Twente de methodiek op gebiedsniveau getoetst zal worden, ook in staat is goede voorspellingen te geven voor andere regio's in Nederland. Daarnaast blijkt ook nogmaals de verschillen in voorkeuren tussen verschillende bedrijfssectoren.

2.3 Conclusies theorievorming

In dit hoofdstuk wordt ingegaan op de kenmerken van LUTI modellen. Daarnaast wordt een onderdeel van LUTI modellen nader bekeken. Dit zijn de motieven van bedrijven waarop zij hun locatiekeuzes baseren.

In theorie zijn LUTI modellen een goed uitgangspunt om de interactie tussen landgebruik en transport te voorspellen. Algemeen wordt aangenomen dat deze relatie aanwezig is. De theorie over LUTI modellen is gebaseerd op locatiekeuzes. Het gaat hier specifiek om locatiekeuzes van huishoudens en bedrijven die invloed hebben op het landgebruik in een gebied. De locatie en interactie van activiteiten van zowel huishoudens als bedrijven, wordt beïnvloed door de bereikbaarheid. De activiteiten en de transportmogelijkheden in een gebied zijn onder invloed van de bereikbaarheid in evenwicht met elkaar. Daarnaast hebben LUTI modellen een dynamisch karakter hebben, wat betekent dat veranderingen in transportevenwicht enkele periodes duren voordat deze volledig zijn verwerkt in het systeem.

Tot op heden zijn de onderliggende relaties tussen de variabelen in LUTI modellen niet goed, waardoor LUTI modellen niet op brede schaal worden toegepast in strategische transportplanning. Dit komt doordat de locatietheorieën die arbeidsmigratie en landgebruik beschrijven te eenvoudig zijn. Naast dit nadeel hebben LUTI modellen ook een belangrijk voordeel. Deze modellen maken het mogelijk om op een strategische tijdschaal voorspellingen te doen over de invloed van investeringen in nieuwe infrastructuur in een gebied.

Empirische studies worden gebruikt om de factoren die locatiekeuzes van bedrijven beïnvloeden te begrijpen. De bereikbaarheid is de belangrijkste factor voor bedrijven om zich in een gebied te vestigen. Andere belangrijke algemene factoren zijn:

- aantal parkeermogelijkheden
- aanwezigheid van adequaat opgeleid personeel
- representativiteit van het gebouw
- telecommunicatievoorzieningen

Daarnaast toont Meester (1999) aan dat de specifieke kenmerken van een vestigingsplaats ook een belangrijke invloed hebben op locatiekeuzes van ondernemers. De vier belangrijkste kenmerken van een vestigingsplaats die door Meester worden gedefinieerd zijn:

- ligging
- agglomeratie-effecten
- kwaliteit infrastructuur
- ligging t.o.v. de externe markt

In onderzoek is aangetoond dat er verschillen bestaan in locatievoorkeuren tussen ondernemers uit verschillende bedrijfssectoren. Een veel gehanteerde indeling van bedrijfssectoren is een indeling in vijf sectoren, te weten industrie- en productieondernemingen, verkeer- en vervoers-ondernemingen, handelsbedrijven, dienstverlenende bedrijven en hoofdkantoren. Elke sector heeft een aantal factoren die de locatiekeuze beïnvloeden, die specifiek van relatief meer belang zijn dan voor andere sectoren. Een overzicht hiervan staat in tabel 2.2. Met behulp van deze factoren kunnen voorspellingen over locatiekeuzes van bedrijven beter worden gedaan in gebieden waar een specifieke sector sterk is vertegenwoordigd.

3 Neurale netwerken

Het model van Tillema (2004), waarvan in dit onderzoek het onderdeel bedrijfslocatie wordt behandeld, maakt gebruik van een neuraal netwerk. Tillema bewijst in zijn dissertatie dat neurale netwerken gebruikt kunnen worden voor LUTI modellen. De vormgeving van het neurale netwerk speelt een belangrijke rol bij de kwaliteit van het model. Om daadwerkelijk in staat te zijn verbeteringen aan te brengen in het bedrijfslocatiemodel, dienen dan ook de eigenschappen van een neuraal netwerk helder te zijn.

3.1 Wat zijn neurale netwerken?

Een artificieel neuraal netwerk kan worden gekarakteriseerd als een informatie verwerkend model dat gebaseerd is op structuur en de werking van de menselijke hersenen, een voorbeeld van een biologisch neuraal netwerk. Het basisconcept van dit model bestaat uit een groot aantal dataverwerkende elementen, genaamd neuronen, wat in onze hersenen aangeduid wordt als cellen. Deze neuronen vormen onderling een groot aantal verbindingen, zodat ze in staat zijn gezamenlijk een probleem op te lossen. Artificiële neurale netwerken zijn in staat, net als mensen, te leren door voorbeelden.

Het onderzoek naar neurale netwerken lijkt een recente ontwikkeling te zijn. Toch werd het eerste artificiële neuron al in 1943 door McCulloch en Pitts ontwikkeld. Geschikte technologie om daadwerkelijk gebruik te maken van deze bevindingen was toen nog niet beschikbaar. Het verlangen om een computer te maken die zich gedraagt als de hersenen van een mens komt direct voort uit het verlangen om een computer te maken die meer menselijk is. In de decennia hierna raakte het onderzoek naar neurale netwerken in verval. Dit werd goed weergegeven door een boek gepubliceerd door Minsky en Papert (1969), waarin de gebreken van neurale netwerken en de frustratie hierover uiteengezet werd. In de jaren tachtig vinden er toch weer belangrijke doorbraken plaats, waardoor de belangstelling voor neurale netwerken en hun toepassingsmogelijkheden nieuw leven ingeblazen wordt.

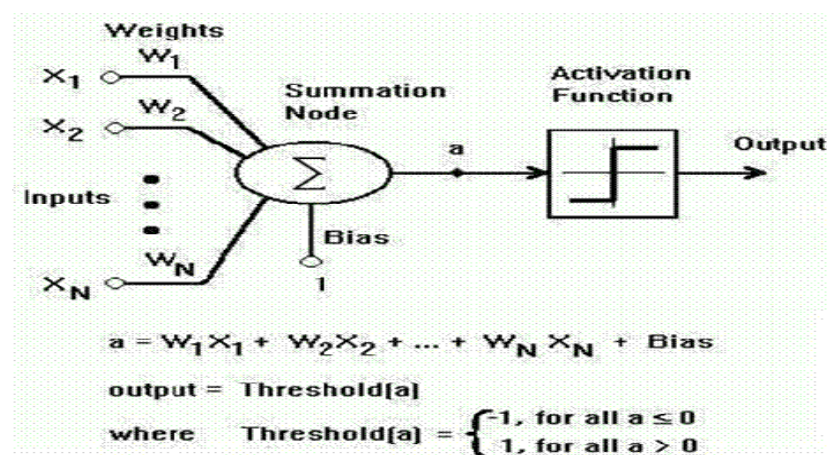
Mensen zijn goed in beeld- en spraakherkenning, redenering, begrijpen en omgaan met onzekerheden, zaken waarmee een computer grote problemen heeft. Zo kunnen mensen eenvoudig twee familieleden herkennen op basis van een aantal gemeenschappelijke gezichtskenmerken. Voor een computer is het herkennen van zulke patronen echter een grote opgave. Mensen hebben de eigenschap om kennis te verkrijgen door te leren en het geleerde toe te passen in nieuwe situaties. Computers gebruiken een rekenkundige, stapsgewijze aanpak om tot de oplossing van een probleem te komen. Neurale netwerken proberen de hierboven beschreven menselijke eigenschappen over te brengen in computer software door de simulatie van een groot aantal eenvoudige verwerkingseenheden die met elkaar verbonden zijn in een netwerk. Een dergelijke neurale computer kan getraind worden om een probleem op te lossen. Doordat neurale netwerken zelf verantwoordelijk zijn hoe ze leren om een oplossing te vinden, zijn ze in staat zichzelf efficiënt te programmeren.

Bij het vormgeven van neurale netwerk modellen draait alles om “leren”. Het neurale netwerk maakt hierbij gebruik van zogenaamde trainingsdata. Dit betekent dat er data in het neurale netwerk wordt ingevoerd dat bestaat uit een serie van voorbeelden van het probleem en de gewenste oplossing(en) voor die opgave. Op deze manier is het neurale netwerk in staat, mits er voldoende

trainingsmateriaal aanwezig is, de onderliggende principes van het probleem boven water te halen, oftewel te leren het probleem op te lossen en toe te passen in vergelijkbare situaties. Deze manier van leren wordt “Supervised Learning” genoemd. Daarnaast zijn er nog twee andere manieren om een neurale netwerk te trainen, namelijk “Unsupervised Learning” en “Reinforcement Learning”. Dit laatste lijkt op Supervised leren, echter, hierbij wordt het netwerk niet de gewenste uitvoer gegeven. Het netwerk ontvangt een positief of negatief signaal naar aanleiding van de geproduceerde uitvoer (Stoianov 2001). Als laatste kan “Unsupervised Learning” worden omschreven als wat bij mensen geldt als leren door ervaringen. Hierbij is de gewenste uitkomst van de ingevoerde data onbekend. Het leren gebeurt op basis van associaties met vergelijkbare situaties, oftewel het neurale netwerk is in staat patronen te herkennen en deze in te delen in bepaalde klassen. Deze toepassing van een neurale gebied is ontdekt door Kohonen in zijn onderzoek op het gebied van spraakherkenning (Computer Associates International, 1999) (Stergiou en Siganos, 1996) (Beale and Jackson, 1990).

3.2 Hoe werken Neurale Netwerken?

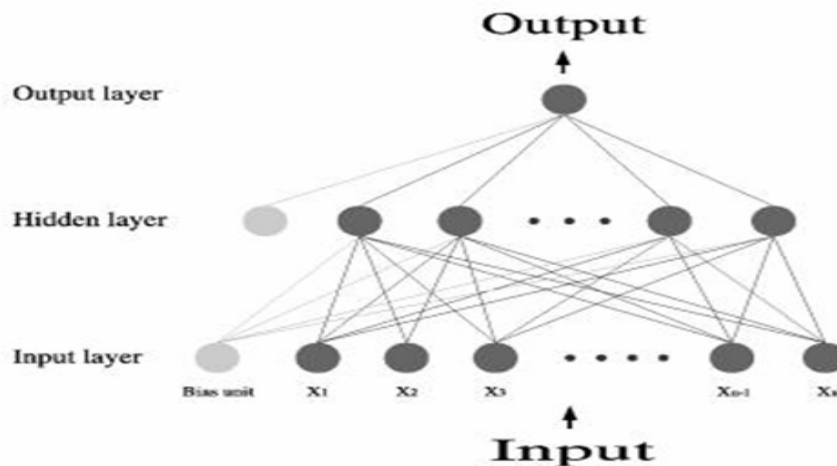
Zoals al eerder vermeld zijn artificiële neurale netwerken gebaseerd op de werking van de hersenen. Er worden een groot aantal vereenvoudigingen gemaakt. De basis bestaat uit neuronen. Schematisch ziet in een neurale netwerk een neuron er uit als weergegeven in figuur 3.1.



Figuur 3.1: Een enkel neuron (DeClaris en Roberts, 1997)

Figuur 3.1 laat zien dat een enkel neuron een invoer kan hebben van een willekeurig aantal vectoren X . Deze invoeren worden gewogen door evenzoveel vectoren W , de weegfactor. Al deze invoeren komen samen en worden gesommeerd en hierbij wordt een soort drempelwaarde, de bias, bij opgeteld (zie functie “a” in fig. 3.1). De uitkomst van deze functie gaat door een zogenaamde activatie functie, die de uitkomst van de functie “a” vergelijkt met de voorgedefinieerde “Threshold” waarde. Als “a” lager is dan de “Threshold”, zal het neuron geen signaal doorgeven. Is deze waarde hoger, dan zal het neuron een signaal doorgeven met een vooraf gedefinieerde waarde (DeClaris en Roberts, 1997).

Het meest gebruikte type neurale netwerk bestaat uit twee of meer lagen van neuronen (zie figuur 3.2). Dit type neurale netwerk bestaat uit ingangsneuronen en uitgangneuronen. Daarnaast zijn er neuronen in de lagen tussen de ingangs- en uitgangneuronen, welke “hidden units” worden genoemd. Het aantal neuronen in een laag kan variëren, afhankelijk van de situatie afhangen van bijvoorbeeld de hoeveelheid en soort data die gebruikt wordt en de gewenste uitkomst.



Figuur 3.2: Configuratie neurale netwerk (Gielen en Kappen, 1995)

De vraag is hoe een neurale netwerk geleerd kan worden een probleem op te lossen zonder het gebruik van een algoritme dat aangeeft hoe het probleem kan worden opgelost. Het antwoord op de vraag is redelijk eenvoudig. Wat nodig is, is een set voorbeelden waarbij voor een aantal ingangssignalen het daarbij behorende uitgangssignaal bekend is. Het uitgangssignaal van het neurale netwerk bij een bepaald ingangssignaal hangt af van de sterkte van de gewichten waarmee neuronen naar elkaar projecteren. Indien we het uitgangssignaal voorstellen door $y(x,w)$, waarbij x het ingangssignaal van het neurale netwerk voorstelt en w de gewichten in het neurale netwerk, dan wordt beoogd om voor elk ingangssignaal x het uitgangssignaal $y(x,w)$ van het neurale netwerk zo veel mogelijk gelijk te maken aan het bij x behorende uitgangssignaal y gewenst. Volgens recht-toe-recht-aan methoden uit de wiskunde is dan eenvoudig een formule af te leiden die voorschrijft hoe de gewichten in het neurale netwerk aangepast moeten worden om het gewenste gedrag te krijgen. Dit rekenwerk gebeurt meestal in de verborgen en in de uitvoerlaag. De invoerlaag registreert slechts het invoersignaal. Op deze manier wordt het neurale netwerk getraind en is het in staat te "leren" (Gielen en Kappen, 1995) (Mulder 2002).

3.3 Toepassingsmogelijkheden neurale netwerken

Het doel van het gebruik van neurale netwerken is om in staat te zijn datapatronen te voorspellen die te ingewikkeld zijn voor traditionele statistische modellen. De belangrijkste voordelen van neurale netwerken zijn (Computer Associates International, 1999) (Tillema, 2004) (Beale en Jackson, 1990):

- Grote hoeveelheden complexe informatie te analyseren en hieruit daadwerkelijk nieuwe informatie uit te halen wat resulteert in nieuwe inzichten en kennis, zonder inmenging van de modelleerder.
- Om te gaan met non-lineaire en/of discontinue functies, om te gaan met dynamische omgevingen en hier "realtime" voorspellingen over te doen. Dit komt omdat neurale netwerken sterk zijn in het herkennen van patronen en trends.
- Doordat het netwerk niet nadrukkelijk hoeft te worden geprogrammeerd is het zeer snel in staat tot een oplossing te komen voor een bepaald complex probleem. Er hoeft geen hercodering, simplificatie of andere interpretatie plaats te vinden, wat resulteert in tijdswinst.

- Het model is robuust. Het is niet gevoelig voor ruis of onbetrouwbare data.
- Nieuwe beschikbare data kan gemakkelijk aan het model worden toegevoegd wat mogelijk zelfs leid tot verbetering van het model. Daarnaast kan het model hierdoor aangepast worden aan veranderingen van het systeem in de tijd.

Daarnaast zijn er echter ook een aantal nadelen van neurale netwerken te noemen:

- Neurale netwerken hebben geschikte, grote hoeveelheden invoerdata nodig om zichzelf te trainen. Het trainen van een neurale netwerk is een tijdrovend proces, mede doordat het lastig is om de geschikte invoervariabelen en juiste mate van training van het systeem te vinden. Ook zijn er situaties waarbij geschikte trainingsdata niet beschikbaar is of dat er geen oorzaak-gevolg relatie aanwezig is.
- Neurale netwerken hebben een “black box” karakter. Dit betekent dat er geen inzicht te krijgen is hoe er tot de oplossing van het probleem gekomen is. Het geeft geen inzicht in het interne mechanisme wat leidt tot de uitkomst. Alleen het resultaat is te zien. Het kan hierdoor ook gebeuren dat het model tot onvoorspelbare uitkomsten leidt.

De bovenstaande opsomming van voor- en nadelen kan het best worden samengevat worden met de onderstaande woorden:

“Neural networks do not perform miracles. But if used sensibly they can produce some amazing results” (Stergiou en Siganos, 1996)

3.4 Vormgeving neurale netwerk

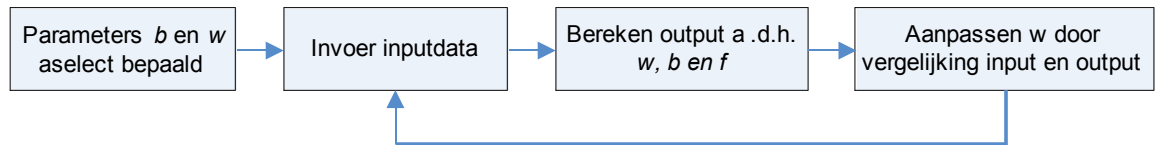
In de vorige paragrafen is duidelijk geworden wat de achterliggende principes zijn van de werking van neurale netwerken. Onduidelijk is echter hoe de data getransformeerd wordt binnen het neurale netwerk naar een gewenste output. In deze paragraaf wordt hierop een antwoord gegeven en duidelijk gemaakt hoe dit in de praktijk vormgegeven kan worden. Daarnaast wordt ingegaan op mogelijkheden om deze transformatie van data te beïnvloeden.

3.4.1 Kenmerken neurale netwerk in Matlab

Voor het modeleren van een neurale netwerk wordt in dit onderzoek gebruik gemaakt van het programma Matlab. Dit is een programma dat een interactieve technische rekenomgeving heeft, waar kan worden gewerkt aan algoritme ontwikkelingen, data visualisatie, data analyse en numerieke berekeningen (The Mathworks, 2005). Het heeft ook een zogenaamde “neural network toolbox”, waarmee neurale netwerken gemodelleerd kunnen worden. In deze “neural network toolbox” zijn de rekenkundige karakteristieken van een neurale netwerk al ingebouwd en is het mogelijk de specifieke kenmerken van een neurale netwerk te definiëren. Dit betekent dat hiermee ook een neurale LUTI model gemodelleerd kan worden.

In paragraaf 3.2 wordt het lerende vermogen van neurale netwerken al besproken door middel van het aanpassen van gewichten in de “hidden layer” om zo de gewenste uitvoer te krijgen. Om een neurale netwerk in staat te stellen te leren moet het getraind worden. In Matlab werkt het trainen van een neurale netwerk als volgt: in het neuron zijn de parameters “bias” en “gewicht”

aanpasbaar. Dit is een iteratief proces, dat als volgt werkt (met b=bias, w=gewicht en f=activatiefunctie):



De aanpassing van de parameters is de belangrijkste stap in dit proces. Matlab doet dit aan de hand van een backpropagation algoritme. Hierbij wordt de gewenste output waarde vergeleken met de door het neurale netwerk voorspelde uitkomst. De relatieve fout tussen beide waarden wordt bepaald met de “mean square error” (MSE) functie. Om deze relatieve fout zo klein mogelijk te krijgen, worden de oorspronkelijk toegekende gewichten aangepast. Het algoritme die aanpassing van het gewicht w bepaald ziet er in eenvoudige vorm als volgt uit (Russell en Norvig, 2003):

$$\Delta w_{ij} = \eta \delta_k x_{ij} \quad (3.1)$$

met:

Δw_{ij} aanpassing gewicht

η leersnelheid

x_{ij} input vector neuron

Waarbij δ_k bepaald wordt door de formule:

$$\delta_k = o_k (1 - o_k) (t_k - o_k) \quad (3.2)$$

met:

δ_k relatieve fout

o_k output vector neuraal netwerk

t_k output vector gewenst (target)

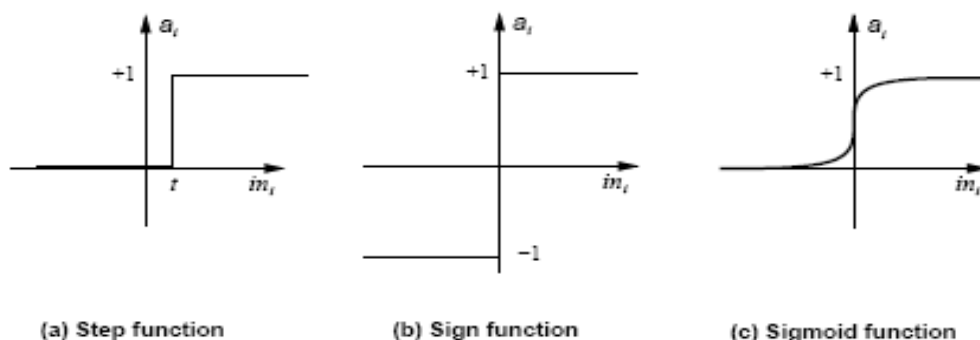
Bij een voldoende aantal iteraties zal op een gegeven moment de waarde van Δw_{ij} , en dus de relatieve fout, tot een acceptabel niveau gedaald zijn. De training is dan voltooid. Het neurale netwerk heeft dan een oplossing gevonden die stabiel is. Dit betekent dat de Δw_{ij} niet verder in een bepaalde richting aangepast kan worden zodat de relatieve fout nog kleiner wordt. Het leersnelheid η bepaalt de stapgrootte waarmee de gewichten worden aangepast. Het kan zijn dat de Δw_{ij} ervoor zorgt dat de gewichten naar een lokaal minimum van de functie (gewenste uitkomsten) divergeren. Om dit probleem op te lossen is er in Matlab een veelgebruikt verbetering ingebouwd, een zogenaamde “momentum term” die aan het algoritme toegevoegd is. Deze zorgt ervoor dat de stapgrootte van de iteratie groter wordt in de buurt van een minimum, zodat het neurale netwerk aan een lokaal minimum voorbij kan gaan.

3.4.2 Beïnvloeding kwaliteit neuraal netwerk

In voorgaande paragraaf is duidelijk geworden hoe een neuraal netwerk model getraind kan worden en in staat is te “leren”. Het gaat hierbij om datatransformatie. De kwaliteit van deze

datatransformatie kan op vele manieren beïnvloed worden, wat ook al gedeeltelijk in de voorgaande paragraaf naar voren is gekomen. Een gedeelte van dit onderzoek behelst het verbeteren van het bedrijfslocatiemodel. Dit wordt getracht door de inputdata te verbeteren, maar daarnaast kan dit ook bereikt worden door de transformatie van de inputdata zo optimaal mogelijk te laten verlopen. Hierbij is een gedegen kennis nodig van factoren van te voren aangepast kunnen worden en die daardoor de kwaliteit van dit proces, het trainen, beïnvloeden. Hieronder volgt dan ook een opsomming van een aantal van deze factoren.

- Normaliseren van data; door de manier waarop het neurale netwerk gemodelleerd is heeft het de neiging om relatief veel waarde te hechten aan inputvariabelen met grote waarden. Daarnaast zijn er veel iteratiestappen nodig als het neurale netwerk uit relatief kleine inputwaarden grote uitvoerwaarden dient te voorspellen. Hierdoor wordt het trainen een tijdrovend proces. De oplossing is om de data te normaliseren, zodat alle data binnen een bepaalde schaal vallen, bijvoorbeeld tussen de -1 en 1 of een schaalverdeling rond het gemiddelde met behulp van de standaard deviatie.
- Matlab beschikt over verschillende activatiefuncties. De functie hiervan staat in paragraaf 3.2 beschreven. Elke activatiefunctie heeft zijn eigen specifieke kenmerken. Een drietal veelgebruikte activatiefuncties staan in figuur 3.3 afgebeeld.



Figuur 3.3: Activatiefuncties

Op de x-as staat voor de mogelijke invoerwaarden. De y-as geeft de mogelijke uitvoer van de activatiefunctie weer. Het figuur maakt duidelijk dat de keuze van de activatiefunctie de aard van de uitvoer bepaald en mede de manier waarop de invoer verwerkt wordt. Dit betekent bijvoorbeeld dat voor het voorspellen van de leeftijd van een groep, de activatiefunctie (a) niet geschikt is. Uit figuur 3.3 blijkt dat met deze activatiefunctie slechts een uitkomst kan geven van 0 of 1.

- Met behulp van de waarde van de “mean square error” functie kan aangegeven worden wanneer de foutmarge voldoende laag is om geaccepteerd te worden. Het kan echter zijn dat dit niveau pas bereikt wordt na het uitvoeren van een zeer groot aantal iteratiestappen, of helemaal niet gehaald kan worden. De reductie van de foutmarge wordt gemiddeld bij meer iteratiestappen per stap steeds kleiner. Daarom kan ook het maximale aantal iteratiestappen begrensd worden, omdat de reductie in foutmarge nog slechts miniem is. Dit levert een grotere foutmarge op.

- Het aantal neuronen in de “hidden layer” en het aantal “hidden layers” zelf kan gevarieerd worden. Met meer neuronen en/of meer “layers” is het neurale netwerk in staat complexere verbanden te doorgronden. Bij teveel neuronen en/of layers ziet het neurale netwerk te complexe verbanden die in werkelijkheid niet bestaan. Het generaliserende vermogen neemt af doordat het ook de extremen in de data probeert te voorspellen. Het leert dan niet de wellicht eenvoudigere juiste verbanden te doorgronden. Daarnaast wordt bij een te groot netwerk de rekentijd die de computer nodig heeft te groot. Het is gebruikelijk door middel van trail en error een inschatting te maken van het juiste aantal neuronen en “layers”.

3.5 Conclusies neurale netwerken

Neurale netwerken maken het mogelijk om voorspellingen of data analyses uit te voeren, waarbij het probleem complex is en de interactie tussen elementen onbekend is. Het neurale netwerk is in staat patronen te herkennen in grote hoeveelheden data, zodat een heldere achterliggende theorie niet nodig is. Hierdoor is het neurale netwerk ook in staat om te gaan dynamische omgevingen en kunnen “real time” voorspellingen kunnen worden gedaan. Belangrijkste nadeel van neurale netwerken is het “black box” karakter, wat betekent dat het inzicht in de achterliggende verbanden tussen de elementen onduidelijk blijft.

Het programma Matlab maakt het mogelijk om een neurale netwerk te modelleren. Het lijkt erop dat hiermee ook een neurale LUT model goed gemodelleerd kan worden. De transformatie van de data naar de gewenste output kan op een aantal manieren beïnvloed worden. Dit betekent dat er goed moet worden nagedacht op welke manier de inputdata bewerkt wordt en op welke manier het neurale netwerk geconfigureerd moet worden. Het is duidelijk dat het normaliseren van de inputdata bijdraagt aan een betere training van het neurale netwerk. De activatiefunctie van een neuron in het neurale LUT model dient ten minste alle waarden tussen 0 en 1 in te kunnen nemen. Het gaat hier immers om het voorspellen van een willekeurig aantal activiteiten in een gebied.

Daarnaast zijn er een aantal configuratieopties waarvan door middel van “trail and error” moet worden uitgevonden wat de beste keuze voor het neurale netwerk, zoals het aantal iteratiestappen en het aantal neuronen in de “hidden layer”. Door deze vele opties is het vinden van de optimale configuratie erg lastig, zo niet onmogelijk. Hierdoor kan nooit gegarandeerd worden dat er gewerkt wordt met het optimale netwerk, maar moet worden volstaan met een benadering van het optimale netwerk.

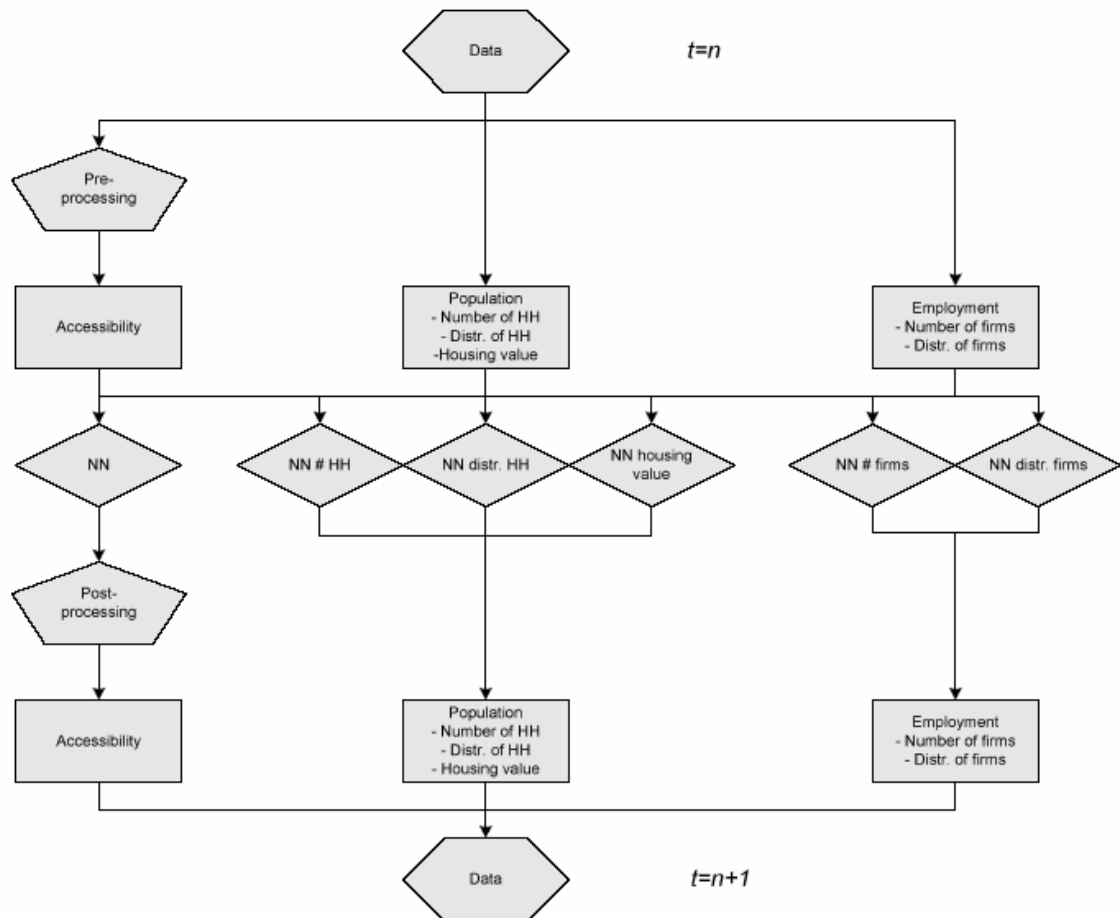
In bijlage B staat een voorbeeld van de configuratie van het neurale netwerk waarmee in Matlab gewerkt is. Hierin worden tevens de gemaakte keuzes betreffende de standaard configuratieonderdelen van het neurale netwerk toegelicht.



4 De bouw van het model

4.1 Achtergrond Tillema's bedrijfslocatiemodel

Het bedrijfslocatiemodel maakt onderdeel uit van het datagedreven LUTI model van Tillema. Zoals eerder aangegeven heeft dit model nog twee onderdelen, respectievelijk het bereikbaarheidsmodel, dat de potentiële bereikbaarheid voorspelt en het huishoudlocatiemodel, dat uitspraken doet over de populatiekenmerken (aantal huishoudens, distributie van huishoudens en woningwaarde) in een gebied. Het conceptuele model van Tillema staat hieronder afgebeeld.



Figuur 4.1: Tillema's conceptuele model

Tillema beschrijft in het kort de werking van het model als volgt: exogene data is input voor het model op tijdstip n ($t=n$). De data wordt naar de verschillende neurale netwerken geleid. De neurale netwerk modellen transformeren de data en deze getransformeerde data wordt gecombineerd op tijdstip $n+1$ ($t=n+1$). Dit kan aan de hand van een voorbeeld uitgelegd worden. Op het eerste tijdstip, $t=n$, beslist de modelleerder dat in een bepaald gebied een nieuw woongebied gesitueerd wordt. Dit houdt in dat de exogene data, de input van het model, verandert. Een nieuw gebied wordt ingevoerd met onder andere nieuwe huizen en nieuwe recreatiezones. Deze data geldt nu als input voor de verschillende neurale netwerken en de modelleerder kan zien hoe onder andere de bereikbaarheid en de locatiekeuzes van bedrijven reageert op dit nieuwe woongebied. De

uitkomsten worden gecombineerd tot een nieuwe dataset op tijdstip $t=n+1$. Dit onthult de situatie van een bepaald studiegebied op tijdstip $t=n+1$.

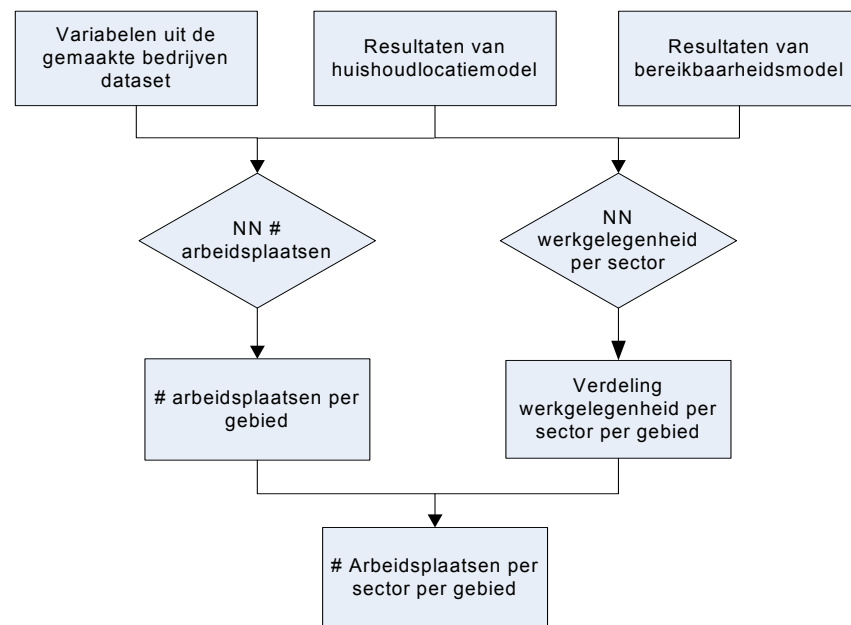
Op deze manier is het model van Tillema in staat voorspellingen te doen voor een toekomstige situatie. Opvallend is het feit dat het hele systeem opgebouwd is uit meerdere neurale netwerken. Tillema zegt hierover dat het gebruik van meer dan één neuraal netwerk het idee van een perfect sequentieel model teniet doet, maar dit niet opweegt tegen de praktische problemen, zoals de omgang met data en nauwkeurigheid. Nu de werking van het totale model globaal is toegelicht, is het nodig om dieper in te gaan op de werking en de specificaties van het bedrijfslocatiemodel.

4.2 Specificaties nieuwe bedrijfslocatiemodel

Het doel van het bedrijfslocatiemodel van Tillema is om het aantal bedrijven per gebied voor drie verschillende bedrijfssectoren te bepalen; industrie, commercie en andere sectoren. Tillema beoogt hiermee voorspellingen te kunnen doen over de werkgelegenheid in een gebied (zie ook figuur 4.1). Werkgelegenheid zorgt immers voor het aantrekken van huishoudens naar een bepaald gebied en visa versa. Om de werkgelegenheid in een gebied te definiëren gebruikt Tillema een variabele die hier indirect verband mee heeft, namelijk het aantal bedrijven in een gebied. Er valt echter een variabele aan te wijzen, waarvan het verband veel directer betrekking heeft op werkgelegenheid. Dit is het aantal arbeidsplaatsen in een gebied. Om deze reden is een aanpassing in het bedrijfslocatiemodel gemaakt en wordt als één van de twee outputs van het model het aantal arbeidsplaatsen in een gebied ingevoerd.

De andere output, de verdeling van de werkgelegenheid in 3 sectoren, respectievelijk industrie, commercie en andere sectoren, wordt tevens om min of meer dezelfde reden aangepast. In paragraaf 2.2.3 is namelijk gebleken dat de verschillen tussen sectoren erg duidelijk zijn als een indeling wordt gemaakt in vijf sectoren. Dit zijn respectievelijk de sectoren industrie- en productieondernemingen, verkeer en vervoersondernemingen, handelsbedrijven, dienstverlenende bedrijven en hoofdkantoren. In verband met beschikbaarheid van adequate data wordt de sector verkeer en vervoersondernemingen samengevoegd met de sector handelsbedrijven. Deze sectoren passen ook goed bij elkaar. Dit blijkt uit het feit dat in paragraaf 2.2.3 duidelijk is geworden dat de factoren waar handelsondernemingen specifiek veel belang aan hechten overeenkomen met die van verkeer en vervoersondernemingen. Hetzelfde wordt, ook in verband met de beschikbaarheid van adequate data, gedaan voor de sectoren dienstverlenende bedrijven en hoofdkantoren. Er zijn twee redenen aan te dragen voor de geldigheid van de samenvoeging van beide sectoren. De relatief kleine omvang van de sector hoofdkantoren ten opzicht van de sector dienstverlenende bedrijven heeft tot gevolg dat de invloed op de resultaten ook relatief klein is. Daarnaast blijkt in paragraaf 2.2.3 dat de voorkeuren van beide sectoren redelijk bij elkaar aansluiten.

Dit leidt ertoe dat het doel van het aangepaste bedrijfslocatiemodel is om het aantal arbeidsplaatsen per gebied voor drie verschillende bedrijfssectoren te bepalen; industrie & productie, handel & distributie en diensten. Deze aanpassingen sluiten aan op de doelstelling van dit onderzoek om het bedrijfslocatiemodel te verbeteren. Dit heeft geleid tot het aangepaste bedrijfslocatiemodel waarmee in dit onderzoek gewerkt zal worden en dat in figuur 4.2 staat geschematiseerd.



Figuur 4.2: Schema aangepaste bedrijfslocatiemodel

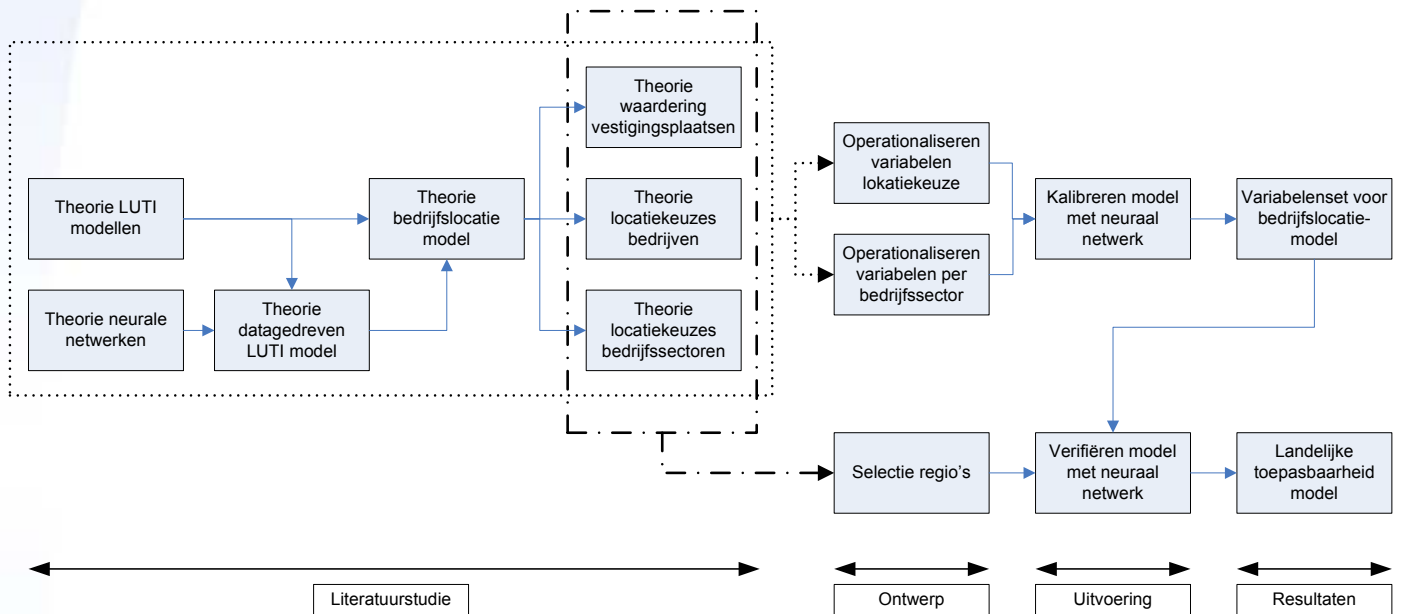
Naast de eerder genoemde aanpassingen kan worden gezien dat het model als input gebruik maakt van de resultaten van het huishoudlocatie onderdeel van het model, de resultaten van het bereikbaarheids onderdeel van het model en exogene variabelen waarvan verwacht wordt dat deze van invloed zijn op locatiekeuzes van bedrijven. De selectie van de variabelen wordt in hoofdstuk 6 nader toegelicht en uitgewerkt. De neurale netwerken worden, zoals eerder al uitgelegd, gebruikt om de data te transformeren. Hierdoor kan uiteindelijk een uitspraak gedaan worden over het aantal arbeidsplaatsen per sector per gebied op een toekomstig tijdstip.



5 Opzet onderzoek

5.1 Onderzoeksmodel

In het eerste hoofdstuk is onder meer een aantal onderzoeksvragen geformuleerd. Om deze vragen te kunnen beantwoorden is in hoofdstuk 2 en 3 ingegaan op de achterliggende theorie van dit onderzoek. In hoofdstuk 4 is daarna het onderzoeksobject, het bedrijfslocatiemodel, uitgebreid behandeld. Hiermee is de basis van het onderzoek uitgewerkt en kan het onderzoeksmodel geformuleerd worden. Het onderzoeksmodel is bedoeld om inzichtelijk te maken op welke manier in dit onderzoek de antwoorden op de onderzoeksvragen verkregen worden. In de volgende paragraaf wordt dit verder uitgewerkt.



Figuur 5.1: Het onderzoeksmodel

Het bovenstaande onderzoeksmodel kan als volgt verwoord worden:

De eerste stap is de *literatuurstudie*, die in hoofdstuk 2 staat beschreven. Met behulp van de theorie over LUTI modellen, neurale netwerken, het bedrijfslocatiemodel, factoren die bedrijfslocatiekeuzes beïnvloeden en regionale gebiedskarakteristieken kan het onderzoek vormgegeven worden, wat in figuur 5.1 aangegeven staat als het onderdeel *ontwerp*. Hierbij wordt getracht de tweeledige doelstelling te verwezenlijken. Het tweeledige karakter van de doelstelling zorgt er ook voor dat het ontwerp zich in tweeën splitst. Dit onderdeel levert de te toetsen invoervariabelen voor het bedrijfslocatiemodel op en een selectie van een aantal regio's waarmee de toepasbaarheid van het bedrijfslocatiemodel op elke willekeurige locatie in Nederland getoetst kan worden. Doordat het aangepaste bedrijfslocatiemodel uit twee onderdelen bestaat, onderdeel totaal aantal arbeidsplaatsen en onderdeel werkgelegenheidsaandeel van bedrijfssectoren, worden er voor beide onderdelen een te toetsen variabelenset geselecteerd. Hiermee wordt de volgende stap van het onderzoek gezet, *de uitvoering*, waarin eerst het aangepaste bedrijfslocatiemodel aan de hand van de nieuwe variabelenset gekalibreerd wordt en daarna geverifieerd aan de hand van de gekozen regio's. Dit levert *de resultaten* op, het laatste onderdeel in figuur 5.1. Met deze resultaten kan nagegaan worden of de doelstelling van het onderzoek verwezenlijkt is.

5.2 Onderzoeksstrategie

De contouren van het onderzoek zijn in de vorige paragraaf uitgezet. In deze paragraaf worden de hoofdlijnen verder uitgewerkt. Zoals ook uit het onderzoeksmodel af te leiden valt, wordt het onderzoek opgesplitst in twee onderdelen. Voor het eerste onderdeel, met als doelstelling een verbeterde variabelenset van het bedrijfslocatiemodel, zullen de volgende stappen uitgevoerd worden:

1. *Definiëren van een variabelenset die locatiekeuzes van bedrijven beschrijft*

Hiervoor zullen de variabelen gebruikt worden die uit de theorie in hoofdstuk 2 naar voren zijn gekomen. Hierbij worden het volgende criteria voor het gebruik van variabelen gehanteerd: de variabele moet omgezet kunnen worden in een meetbare variabele en geschikte data voor deze variabele dient eenvoudig te verkrijgen zijn.

1.1 *Variabelenset voor voorspellen aantal arbeidsplaatsen in bedrijfslocatiemodel*

Voor het bepalen van deze variabelenset wordt gebruik gemaakt van de vijf belangrijkste variabelen die in paragraaf 2.2.2 naar voren zijn gekomen uit het onderzoek van Reijs (2001) en wat bevestigd wordt door het onderzoek van Hanemaayer (1998). Deze variabelen zijn:

- bereikbaarheid van de locatie
- aanwezigheid parkeerfaciliteiten
- aanwezigheid van adequaat opgeleid personeel
- representativiteit van het gebouw
- telecommunicatie voorzieningen

Daarnaast worden de vier variabelen in paragraaf 2.2.4 gebruikt, die naar voren komen in het vestigingsplaats onderzoek van Meester (1999). Deze variabelen zijn:

- ligging
- agglomeratie-effecten
- kwaliteit van de infrastructuur
- ligging t.o.v. de externe markt.

1.2 *Variabelenset voor voorspellen werkgelegenheidsaandeel per bedrijfssector in bedrijfslocatiemodel*

Voor het bepalen van deze variabelenset wordt gebruik gemaakt van het onderzoek van Reijs (2001) en Everaars en Brand-van Tuijn (2001), waaruit in paragraaf 2.2.3 een aantal variabelen per bedrijfssector beschreven zijn. Deze zijn voor de industrie & productie sector:

- Uitbreidingsmogelijkheden
- laad en los capaciteit terrein
- aanwezigheid toeleverende bedrijven
- aanwezigheid logistieke dienstverlening.

Voor de sector handel & distributie:

- huurprijs/ grondprijsniveau
- aanwezigheid zee- of luchthaven
- internationale bedrijvigheid
- laad en los capaciteit terrein
- aanwezigheid logistieke dienstverlening

Voor de sector diensten:

- quality of life factoren

- representativiteit omgeving
- OV-bereikbaarheid

2. *Kalibreren van het bedrijfslocatiemodel*

Deze stap wordt voor beide onderdelen van het bedrijfslocatiemodel (stap 1.1 en 1.2) gedaan. Op deze manier worden beide variabelensets getoetst op hun kwaliteit.

2.1 *Kalibratie met dataset van geheel Nederland*

Het neurale netwerk wordt, nadat het goed geconfigureerd is, getraind met data op het niveau van vierposities postcodegebieden van geheel Nederland. De grote van een regio in het model heeft dus de grote van een vierposities postcodegebied, wat betekent dat Nederland in ongeveer 4000 gebieden wordt opgedeeld. Het netwerk probeert met behulp van de bij stap 1 gedefinieerde variabelen te leren zo goed mogelijk de gewenste uitvoer te verkrijgen. De manier waarop dit “leren” gebeurt staat in paragraaf 3.4.1 beschreven.

2.2 *Verificatie met dataset van geheel Nederland*

Het doel van de verificatie is om duidelijk te maken of het bij stap 2.1 gekalibreerde model ook in staat is de gewenste data te reproduceren. Hiervoor wordt gebruik gemaakt van hetzelfde databestand waarmee het model ook getraind is.

3. *Toetsing methodiek bedrijfslocatiemodel op gebiedsniveau*

Deze stap wordt net als stap 2 voor beide onderdelen van het bedrijfslocatiemodel (zie stap 1.1 en 1.2) gedaan. Op deze manier worden beide variabelensets getoetst op hun toepasbaarheid op gebiedsniveau. Het gebied dat voor deze test wordt gebruikt is de regio Twente.

3.1 *Kalibratie met dataset van geheel Nederland, behalve data testcase Twente*

Het neurale netwerk wordt opnieuw, nadat het goed geconfigureerd is, getraind met data op het niveau van vierposities postcodegebieden van geheel Nederland. Alleen de data uit de testcase Twente is deze keer uit de dataset gelaten. Het netwerk probeert met behulp van de bij stap 1 gedefinieerde variabelen te leren zo goed mogelijk de gewenste uitvoer te verkrijgen.

3.2 *Verificatie met dataset van testcase Twente*

Het doel van de verificatie is om duidelijk te maken of het bij stap 3.1 gekalibreerde model ook in staat is data te reproduceren op een lokaal niveau, waarmee het niet getraind is. Dit is namelijk één van de voorwaarden waaraan een goed LUTI model moet voldoen. Hiervoor wordt gebruik gemaakt van data uit Twente (tevens op niveau van vierposities postcodegebied).

4. *Beoordelen resultaten*

De kwaliteit van de resultaten die bij stap 2 en 3 verkregen zijn, dienen beoordeeld te worden.

4.1 *Bepaling afwijking t.o.v. werkelijke data*

Voor zowel de bij de kalibratie van het model als de bij de toetsing van de methodiek van het model op gebiedsniveau voorspelde uitkomsten van het neurale netwerk, wordt bepaald wat de afwijking is ten opzicht van de werkelijke geobserveerde data.

4.2 *Bepaling kwaliteit voorspellingen*

De laatste stap van het onderzoek behelst het bepalen van de kwaliteit van de nieuwe dataset in het bedrijfslocatiemodel. Hiervoor worden de uitkomsten van stap 4.1 gebruikt. Hoewel er vele mogelijkheden zijn om iets over de kwaliteit van de uitkomsten te zeggen, is gekozen om dezelfde methode als in Tillema's onderzoek te gebruiken, namelijk het percentage voorspelde data dat binnen een bepaalde range van de werkelijke data valt.

Hiermee is het goed mogelijk grote hoeveelheden uitkomsten van verschillende groottes te beoordelen op hun kwaliteit. Nadeel van deze methode is dat de kans relatief groot is voor kleine uitkomstwaarden om buiten een bepaalde range te vallen. Immers, de relatieve afwijking van 5 voorspelde arbeidsplaatsen is bij een geobserveerde werkelijkheid van 10 arbeidsplaatsen veel groter, namelijk 50%, dan bij een geobserveerde werkelijkheid van 100 arbeidsplaatsen, namelijk slechts 5%. Na een korte analyse van een aantal andere methoden bleek echter dat er aan deze methoden andere nadelen kleven die niet opwegen tegen de voordelen.

Voor het tweede onderdeel van het onderzoek, met als doelstelling het toetsen van de toepasbaarheid van het bedrijfslocatiemodel op alle locaties in Nederland, worden de volgende stappen uitgevoerd:

5. Bepaling toepasbaarheid bedrijfslocatiemodel op elke willekeurige locatie in Nederland

5.1 Selecteren van een zestal regio's

De selectie van de zes regio's zal gebaseerd zijn op een aantal specifieke kenmerken van deze gebieden. De regio's bestaan uit een beperkt aantal postcode-4 gebieden. De kenmerken die gezocht worden, worden bepaald op basis van het vestigingsplaatsonderzoek van Meester (1999) in paragraaf 2.2.4 en de algemene factoren waaraan bedrijven belang hechten bij het maken van locatiekeuzes, beschreven in paragraaf 2.2.2. Daarnaast wordt getracht ten minste één regio te selecteren per bedrijfssector waar betreffende sector bovengemiddeld vertegenwoordigd is. De mogelijkheid bestaat immers dat in een dergelijk regio de belangrijkste locatiefactoren die de locatiekeuzes van ondernemers beïnvloeden, afwijken van het landelijke gemiddelde. Hierdoor heeft het bedrijfslocatiemodel wellicht moeite met het doen van goede voorspellingen voor deze regio's. Dit betekent dat een regio met veel bedrijven uit de industrie & productie sector, een regio met veel bedrijven uit de handel & distributie sector en een regio met veel dienstverlenende bedrijvigheid geselecteerd wordt.

5.2 Toetsen bedrijfslocatiemodel met zes regio's

De geselecteerde regio's uit stap 5.1 worden gebruikt om duidelijk te maken of het model ook in staat is data te reproduceren van elke willekeurige locatie in Nederland. Allereerst wordt het model gekalibreerd met de data van heel Nederland, behalve de te toetsen regio. Daarna vindt de toetsing van het model plaats met behulp van de data van de geselecteerde regio. Dit proces wordt voor elk van de geselecteerde regio's afzonderlijk uitgevoerd.

5.3 Beoordelen resultaten

Deze stap wordt op dezelfde manier uitgevoerd als stap 4. Eerst wordt bepaald wat de afwijking is van de door het neurale netwerk voorspelde uitkomsten ten opzicht van de werkelijke geobserveerde data. Daarna wordt gekeken of het bedrijfslocatiemodel in staat is voorspellingen te doen voor gebieden in heel Nederland, door te bepalen welk percentage voorspelde data binnen een bepaalde range van de werkelijke data valt.

5.3 Onderzoeksmateriaal

Voor het verkrijgen van de benodigde achterliggende kennis over LUTI modellen, locatiekeuzes van bedrijven en neurale netwerken is gebruik gemaakt van literatuur. Daarnaast is het proefschrift van Tillema (2004) gebruikt om inzicht te krijgen in de werking van het datagedreven LUTI model.

Om het aangepaste bedrijfslocatiemodel, met als onderdeel het neurale netwerk, vorm te geven, is gebruik gemaakt van de “neural network toolbox” binnen het programma Matlab, wat al eerder in paragraaf 3.4 vermeld is (The Mathworks, 2005).

Daarnaast is er data gebruikt voor de invoer van de twee onderdelen van het model van Tillema, het bereikbaarheidsmodel en het huishoudlocatiemodel, die niet in het onderzoek zijn meegenomen, maar wel nodig zijn als input voor het bedrijfslocatiemodel. De belangrijkste bron van deze data is de database “Kerncijfers wijken en Buurten 2001” (CBS, 2005). De data die als uitvoer dient van het bedrijfslocatiemodel is afkomstig van de database “Nieuw Regionaal Model 2000” (AVV, 2003).

De bronnen van de invoerdata voor het bedrijfslocatiemodel zijn divers en worden bij de behandeling van de selectie van invoervariabelen voor het bedrijfslocatiemodel in bijlage C.1 en C.2 vermeld.



6 Uitwerking onderzoek

6.1 Operationaliseren variabelen

In voorgaand hoofdstuk zijn de hoofdlijnen van het onderzoek beschreven. In dit hoofdstuk vindt de uitwerking van deze hoofdlijnen in detail plaats. Ten eerste zijn de details van het te onderzoeken object, het bedrijfslocatiemodel, al uitvoerig in hoofdstuk 4 behandeld. Voor de input van het model zijn in hoofdstuk 2 een scala aan mogelijkheden aangereikt. Dit heeft een aantal variabelen opgeleverd waarvan verondersteld mag worden dat deze geschikt zijn als input voor het bedrijfslocatiemodel en de locatiekeuzes van bedrijven beschrijven. In deze paragraaf zullen deze variabelen geoperationaliseerd worden, zodat het model in staat is met deze variabelen om te gaan. Hierbij worden het volgende criteria voor het gebruik van variabelen gehanteerd: de variabele moet omgezet kunnen worden in een meetbare variabele en geschikte data voor deze variabele dient eenvoudig te verkrijgen zijn.

De variabelen die afkomstig zijn uit de twee andere modellen van het datagedreven LUTI model van Tillema (2004), het bereikbaarheidsmodel en het huishoudlocatiemodel, zijn echter al geoperationaliseerd. Zoals eerder al in figuur 4.2 uit paragraaf 4.2 naar voren kwam, worden deze variabelen voor zowel de voorspelling van het aantal arbeidsplaatsen, als voor de voorspelling van het werkgelegenheidsaandeel per bedrijfssector meegenomen. In tabel 6.1 staat een overzicht van deze variabelen.

Model	Variabelen	Toelichting
Locatie huishoudens	- Aantal huishoudens % 1-persoons huishoudens % huishoudens zonder kinderen % huishoudens met kinderen - WOZ-waarde	Gemiddelde waarde van een woning
Bereikbaarheid	Bereikbaarheid	Coëfficiënt voor de mate van bereikbaarheid via de weg

Tabel 6.1: Overzicht variabelen huishoudlocatie- en bereikbaarheidsmodel

Voor elke variabele geldt dat de data beschikbaar is op het niveau van een vierposities postcodegebied, tenzij in de toelichting anders staat vermeld. Hetzelfde zal in de volgende paragrafen gelden voor de variabelen voor de locatiekeuzes van bedrijven en bedrijfssectoren.

6.1.1 Operationalisatie variabelenset aantal arbeidsplaatsen

Het bedrijfslocatiemodel bestaat uit twee onderdelen. Daarom wordt er voor beide onderdelen afzonderlijk ook een variabelenset gecreëerd, waarmee het bedrijfslocatiemodel voorspellingen kan doen voor beide afzonderlijke onderdelen. Het eerste onderdeel doet voorspellingen over het aantal arbeidsplaatsen in een regio. Eerst volgt opnieuw een overzicht van de vijf belangrijkste variabelen voor locatiekeuzes van ondernemers die geselecteerd zijn. Dit zijn achtereenvolgens:

- Bereikbaarheid van de locatie
- Aanwezigheid parkeerfaciliteiten
- Beschikbaarheid van adequaat opgeleid personeel

- Representativiteit van het gebouw
- Kwaliteit van de aanwezigheid telecommunicatievoorzieningen

Deze vijf variabelen worden nu verder geoperationaliseerd. De variabele bereikbaarheid wordt niet apart in het bedrijfslocatiemodel meegenomen, omdat deze al in het onderdeel bereikbaarheid van het model van Tillema (2004) zit en dus automatisch al in het bedrijfslocatiemodel meegenomen wordt. Dan blijven er nog vier variabelen over. Deze vier variabelen zijn getoetst op de voorwaarde die in het vorige hoofdstuk gedefinieerd is, namelijk dat de variabele omgezet moet kunnen worden in een meetbare variabele en dat geschikte data voor deze variabele eenvoudig te verkrijgen moet zijn. Dit blijkt voor drie van de vier variabelen goed mogelijk te zijn. Voor de variabele aanwezigheid telecommunicatievoorzieningen blijkt het echter lastig om aan geschikte data te komen. Kabelbedrijven zijn niet in staat of niet bereid tot het verstrekken van bruikbare informatie. Informatie over telecommunicatieonderwerpen die wel gevonden worden, blijkt of niet bruikbaar of te algemeen van aard voor het onderzoek te zijn. Daarnaast valt het te betwijfelen of er tegenwoordig nog grote verschillen bestaan in telecommunicatievoorzieningen. Een aantal jaren geleden was er nog sprake van een beperkt aantal grootstedelijke gebieden waar nieuwe toepassingen zoals breedbandinternet en glasvezelverbindingen beschikbaar kwamen. Ondertussen zijn er grote investeringen gedaan en beschikken nagenoeg alle regio's in Nederland over redelijk tot goede internetinfrastructuur. De vraag valt dus te rechtvaardigen of deze variabele vandaag de dag nog steeds een zeer belangrijke rol speelt bij de locatiekeuzes van bedrijven. Om deze redenen is ervoor gekozen om deze variabele niet op te nemen in de uiteindelijke variabelenset.

Naast het kiezen van variabelen op basis van de literatuur is het ook belangrijk om naar het model zelf te kijken. Het is de bedoeling dat het model het aantal arbeidsplaatsen kan voorspellen. Een direct verband met het aantal arbeidsplaatsen in het gebied valt er aan te wijzen met het aantal bedrijven in een gebied. Een variabele die het aantal bedrijven in een gebied beschrijft lijkt dan ook voor de hand te liggen. Daarnaast wordt met de introductie van deze variabele ingespeeld op een ander probleem. In het bedrijfslocatiemodel wordt aangenomen dat bedrijven vrij in staat zijn zich te verplaatsen over geheel Nederland. Dit zal zeker het geval zijn voor middelgrote en grote bedrijven. Voor kleine bedrijven is dit echter wellicht niet altijd het geval. Het zou zo kunnen zijn dat een percentage relatief kleine ondernemers zich door hun lokale afzetgebied en hun lokale netwerk slechts kunnen richten op de regio's om hun heen. Dit type ondernemer overweegt dus absoluut niet om zijn onderneming naar de andere kant van het land te verplaatsen, ook al zouden daar de omstandigheden theoretisch veel gunstiger zijn. Zij zijn dan ook veel meer locatie of regiogebonden. Als de variabele "aantal bedrijfsvestigingen" ingevoerd wordt en daarbij aangenomen wordt dat er gemiddeld voor elke regio een evenredig groot aandeel van dit type ondernemers aanwezig is kan dit probleem ondervangen worden. Het model zou dan rekening moeten kunnen houden met het scenario dat, mocht er op een toekomstig tijdstip een aantal factoren verslechteren, dit voor een bepaald deel van de bedrijven in de regio geen aanleiding is om te verplaatsen. Dit betekent ook dat een deel van de arbeidsplaatsen niet verplaatst wordt. Om deze reden wordt het aantal bedrijven in een gebied ook als variabele meegenomen in de uiteindelijke variabelenset.

Er is nog een variabele toegevoegd aan de variabelenset. Dit is de variabele die de stedelijkheid van een gebied beschrijft. Belangrijkste reden is deze variabele indirect een toevoeging vormt op de vier variabelen die hierboven geselecteerd zijn. Zo zal de stedelijkheid van een gebied invloed hebben op het aantal parkeerplaatsen, en waarschijnlijk ook op de kans of er adequaat opgeleid

personeel aanwezig is. Ook zal de stedelijkheid van een gebied iets zeggen over mogelijke agglomeratie-effecten. Immers, waar een sterk stedelijk gebied aanwezig is, lijkt de kans aanwezig dat de omliggende gebieden ook matig tot sterk stedelijk zijn, zodat er gesproken kan worden van een agglomeratie.

De hoofdvariabelen die de locatiekeuzes van bedrijven in het model bepalen zijn hiermee bepaald. Er worden echter nog een viertal variabelen toegevoegd. Dit zijn de vier die Meester (1999) in zijn onderzoek behandelt en die specifiek ingaan op de kenmerken van een vestigingsplaats. Meester had al geconcludeerd dat deze variabelen goed meetbaar zijn en de benodigde data is verkrijgbaar. Dit alles zorgt voor de een set van variabelen die in tabel 6.2 weergegeven zijn. In bijlage C.1 staan de specificaties en de bronnen van deze variabelenset beschreven.

	Variabelen
Hoofdvariabelen	• Parkeerfaciliteiten • Aanwezigheid adequaat opgeleid personeel • Representatie gebouw • Aantal bedrijfsvestigingen • Stedelijkheid gebied
Vestigingsplaatsvariabelen	• Ligging in NL • Agglomeratie-effect • Infrastructuur • Oriëntatie t.o.v. externe markt

Tabel 6.2: Overzicht locatiekeuze variabelen arbeidsplaatsen

6.1.2 Operationalisatie variabelenset werkgelegenheidsaandeel bedrijfssectoren

Het tweede onderdeel van het bedrijfslocatiemodel doet voorspellingen over het werkgelegenheidsaandeel van bedrijfssectoren in een gebied. In paragraaf 4.2 is besproken welke bedrijfssectoren in het aangepaste bedrijfslocatiemodel gebruikt gaan worden, te weten industrie & productie, handel & distributie en diensten. Voor elk van deze sectoren zijn in hoofdstuk 5 een aantal specifieke variabelen geselecteerd, die hieronder opnieuw op een rijtje worden gezet. Opgemerkt dient te worden dat de hoofdvariabelen die in de voorgaande paragraaf genoemd zijn, ook voor de afzonderlijke bedrijfssectoren van belang zijn. Dit betekent dat de hieronder genoemde variabelen gezien moeten worden als een aanvulling op de hoofdvariabelen.

Industrie & productie sector

- Uitbreidingsmogelijkheden
- Laad en los capaciteit terrein
- Aanwezigheid toeleverende bedrijven
- Aanwezigheid logistieke dienstverlening

Handel & distributie sector

- Huurprijs/ grondprijsniveau
- Aanwezigheid zee- of luchthaven
- Internationale bedrijvigheid

- Laad en los capaciteit terrein
- Aanwezigheid logistieke dienstverlening

Diensten sector

- Quality of life factoren
- Representativiteit omgeving
- OV-bereikbaarheid

Deze variabelen worden nu verder geoperationaliseerd aan de hand van de aan het begin van deze paragraaf genoemde criteria. Dit blijkt voor meeste variabelen goed mogelijk te zijn. Er zijn echter een paar uitzonderingen. Voor de variabele laad en los capaciteit terrein en de variabele internationale bedrijvigheid lijkt geen geschikte data eenvoudig voor handen te zijn. Voor de hoeveelheid beschikbare laad en los ruimte op een terrein valt indirect de grondprijs als variabele aan te wijzen. Bij lage grondprijzen zijn bedrijven sneller geneigd meer te investeren in laad en los ruimte. Aangezien de variabele grondprijzen echter afzonderlijk al in de onderzoeken van Reijs (2001) en Everaars en Brand-van Tuijn (2001) behandeld wordt, is dit geen geschikte variabele voor de factor laad en los capaciteit op een terrein. Om deze reden wordt deze variabele en de al eerder genoemde variabele internationale bedrijvigheid niet opgenomen in de uiteindelijke variabelenset.

Als laatste wordt de variabele huurprijs/ grondprijsniveau niet opgenomen in de uiteindelijke variabelenset. Dit om de eenvoudige reden dat voor de definitie van een van de hoofdvariabelen, de uitstraling van het gebouw, de variabele huurprijs/ grondprijs al is gebruikt. Deze variabele zit dus al in de variabelenset en hoeft niet opnieuw afzonderlijk te worden meegenomen voor de sector handel & distributie. Bij het onderzoek naar de afzonderlijke sectoren worden immers ook de hoofdvariabelen meegenomen. De bovenstaande afwegingen hebben geleid tot een keuze van de variabelen in tabel 6.3 weergegeven zijn. In bijlage C.2 staan de specificaties en de bronnen van deze variabelenset beschreven.

Bedrijfssector	Variabelen
Industrie & productie	• Uitbreidingsmogelijkheden • Toeleverende bedrijven • Logistieke dienstverlening
Handel & distributie	• Huurprijs- /grondprijsniveau • Aanwezigheid zee- of luchthaven • Logistieke dienstverlening
Diensten	• Quality of life factoren • Representatie omgeving • OV-bereikbaarheid

Tabel 6.3: Overzicht locatiekeuze variabelen per bedrijfssector

6.2 Toelichting output model

In hoofdstuk 4 is duidelijk gemaakt hoe de outputvariabelen van het bedrijfslocatiemodel eruit gaat zien. De eerste output van het model zijn het aantal arbeidsplaatsen in een gebied. De tweede output betreft het werkgelegenheidsaandeel van bedrijfssectoren in een gebied. Deze output is verdeeld over drie variabelen, te weten het aantal arbeidsplaatsen in respectievelijk de sector

industrie & productie, de sector handel & distributie en de sector diensten. Deze variabelen zijn afkomstig van het Nieuw Regionaal Model 2000 (AVV, 2003). Dit betekent dat het referentiejaar waarover het model uitspraken moet doen, het jaar 2000 is. Hoewel getracht is zoveel mogelijk gebruik te maken van data uit het referentiejaar, is er bij de creatie van de inputvariabelen ook data uit andere jaren gebruikt. Het gaat bij de meeste factoren die de inputvariabelen representeren over zaken die redelijk stabiel zijn en waarvan verwacht mag worden dat ze binnen één a twee jaar significant veranderd zijn. Daarnaast is in paragraaf 2.1.3 duidelijk geworden dat LUTI modellen een dynamisch karakter hebben, waarbij onduidelijk is op welk exact tijdstip factoren invloed uitoefenen op de locatiekeuzes van bedrijven. Om deze reden wordt aangenomen dat de kleine afwijkingen van de data van het referentiejaar niet van wezenlijke invloed is op de resultaten.

Een ander punt dat vermeld dient te worden is dat er één sector in zijn geheel buiten beschouwing wordt gelaten, ook bij de voorspelling van het totale aantal arbeidsplaatsen in een gebied. Dit is de agrarische sector. De reden hiervoor is dat deze sector een geheel eigen dynamiek heeft en dus ook specifieke locatievoorkeuren die absoluut niet overeenkomen met de andere sectoren. Hoewel het aandeel van deze sector niet zeer groot is ten opzichte van het totale aantal arbeidsplaatsen, valt te verwachten dat deze sector het algemene beeld van locatieprocessen van bedrijven alleen maar zal vertroebelen.

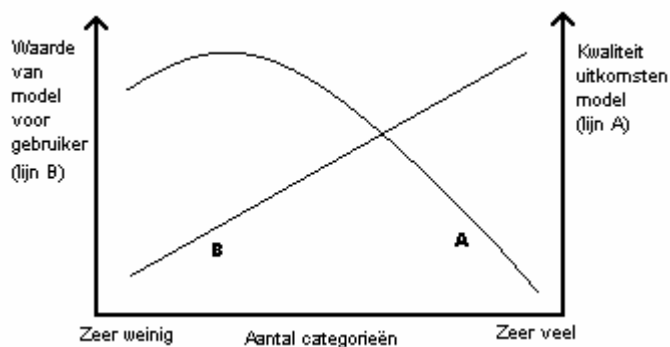
Er dient nog een lastige vraag beantwoord te worden. Het doel van het onderzoek is om het bedrijfslocatiemodel te verbeteren, door middel van de in de vorige paragraaf gedefinieerde variabelenset. Echter, om te bepalen wanneer het bedrijfslocatiemodel in staat is goede voorspellingen te doen over de locatiekeuzes van bedrijven, is een beoordeling van de kwaliteit van de output van het model nodig. In het vorige hoofdstuk is al een voorstel gedaan om dit te doen op dezelfde manier als in het onderzoek van Tillema (2004), door te bepalen welk percentage voorspelde data binnen een bepaalde range van de werkelijke data valt. Modelresultaten worden over het algemeen als goed geaccepteerd als 70% van de data nauwkeurig gereproduceerd wordt. De vraag wanneer data nauwkeurig voorspeld worden blijft altijd aanwezig, maar de door Tillema gebruikte ranges van 0.8-1.2 en 0.5-1.5 lijken acceptabel. Een ander probleem bij neurale netwerken is dat de kwaliteit van de kalibratie per run van het neurale netwerk nogal eens verschillende uitkomsten oplevert. Het kan zijn dat het neurale netwerk in een lokaal minimum blijft hangen, zoals eerder al uitgelegd in paragraaf 3.4. Om dit probleem te voorkomen, wordt elke onderzochte stap vijf keer uitgevoerd. Het beste resultaat wordt opgenomen in de resultaten. Op deze manier worden de kans op slechte uitkomsten kleiner en valt met redelijke zekerheid te zeggen dat de resultaten betrekking hebben op een optimaal gekalibreerd model. Hiermee is de vraag beantwoord op welke manier de kwaliteit van het bedrijfslocatiemodel beoordeeld wordt.

6.2.1 Categorisering uitkomstvariabelen

Er is nog een ander belangrijk aspect waarna gekeken moet worden. Zoals eerder al in hoofdstuk 3 is uitgelegd, leert het neurale netwerk verbanden te doorgronden doordat het getraind wordt. Hiervoor heeft het een aantal cases nodig, die dezelfde uitkomst hebben. Om dit te bereiken dienen de uitkomsten gecategoriseerd te worden, zodat in elke categorie een aantal cases zitten die dezelfde uitkomst hebben. Een categorie staat bijvoorbeeld voor een aantal arbeidsplaatsen tussen de 250 en 300. Stel dat dit categorie 10 is, dan betekent dit ook dat een uitkomst van 10, voorspeld door het neurale netwerk, betekent er tussen de 250 en 300 arbeidsplaatsen in een gebied door het neurale netwerk verwacht worden. Het totale aantal cases wordt evenredig verdeeld over de categorieën, waarbij de cases met de laagste aantallen arbeidsplaatsen in de laagste categorie vallen, en zo verder. Bij een lager aantal categorieën zijn er per categorie meer cases beschikbaar.

Met meer cases kan het neurale netwerk beter getraind worden, zodat het beter in staat is de gewenste output (in dit geval dus een categorie) te voorspellen.

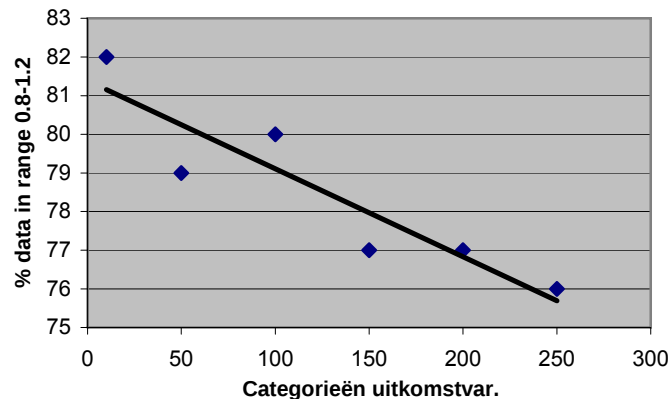
Nadeel is echter dat bij een lager aantal categorieën het aantal arbeidsplaatsen dat binnen deze categorie valt groter wordt. Dit zorgt voor twee problemen. De waarde voor de gebruiker van het model wordt kleiner, doordat uit de voorspelling van het model slechts in indicatie gehaald kan worden van het werkelijke aantal arbeidsplaatsen in een gebied. Daarnaast kan bij een te klein aantal categorieën ongewild een simplificatie van de werkelijkheid worden doorgevoerd die te groot is. Het zou bijvoorbeeld zo kunnen zijn dat gebieden met 2000 arbeidsplaatsen in dezelfde categorie gaan vallen als gebieden met 4000 arbeidsplaatsen. Het model beschouwt deze gebieden dan als uniform, terwijl een aantal variabelen wellicht compleet van elkaar verschillen. Dit kan tot gevolg hebben dat het model minder goed getraind wordt. De kwaliteit van de input neemt immers af. Dit zou dan ook kunnen leiden tot slechtere resultaten. De voor- en nadelen van het aantal categorieën komt goed tot uitdrukking in figuur 6.1.



Figuur 6.1: Voor en nadelen categorisering uitkomstvariabele

In figuur 6.1 wordt duidelijk dat getracht moet worden het aantal categorieën te benaderen, waarbij beide lijnen elkaar doorsnijden. Of in andere woorden, het is de bedoeling dat het aantal categorieën zo groot mogelijk wordt gemaakt, waarbij de kwaliteit van de uitkomsten niet te sterk zijn gedaald. Om dit punt te vinden is er een testcase opgesteld, waarbij de uitkomstvariabele die het totale aantal arbeidsplaatsen beschrijft in verschillende aantallen categorieën is opgedeeld. De kwaliteit van de categorieën wordt met behulp van een aantal runs met een neurale netwerk bepaald. Hierbij wordt als input de geselecteerde variabelen uit paragraaf 6.1 gebruikt. Aangenomen wordt dat de gevonden optimale categorisering van het aantal arbeidsplaatsen ook geschikt is voor het onderdeel dat het werkgelegenheidsaandeel van bedrijfssectoren in het model beschrijft. In bijlage D zijn de resultaten in tabelvorm gezet en wordt een toelichting gegeven. Om de resultaten overzichtelijk te maken zijn ze in figuur 6.2 neergezet.

Uit de uitkomsten blijkt dat de kwaliteit van de uitkomsten van het bedrijfslocatiemodel langzaam lijken af te nemen als het aantal arbeidsplaatsen in meer categorieën ingedeeld wordt. De eerder beweerde afname van de kwaliteit van het model bij zeer weinig categorieën valt niet op te maken uit de grafiek. Het kan zijn dat dit effect pas optreedt bij minder dan 10 categorieën. Er dient echter zeer voorzichtig met de interpretatie van de uitkomsten omgegaan te worden. De grootste oorzaak hiervan is de manier waarop de kwaliteit van de output van het model gemeten wordt. Dit is eerder in deze paragraaf al uitgelegd. Hier spelen twee tegenstrijdige zaken een rol.



Figuur 6.2: Kwaliteit uitkomsten bij categorisering uitkomstvariabele

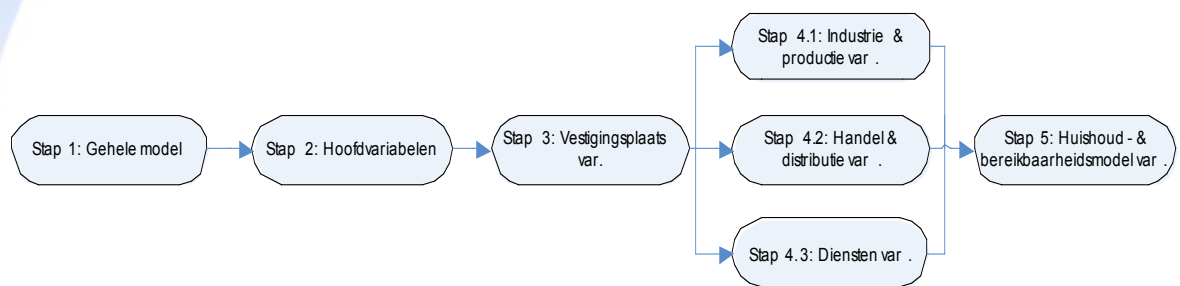
Ten eerste de beoordeling van kleine waarden, die min of meer aansluit op de tekortkoming van de methode die eerder in paragraaf 5.2 ter sprake zijn gebracht. Hierin wordt gezegd dat kleine outputwaarden eerder buiten een range vallen dan grote waarden en dus eerder relatief grote afwijkingen vertonen van de gewenste output. Des te meer categorieën, des te kleiner de kleinste categorieën en des te kleiner de waarden die het neurale netwerk dient te voorspellen. De kans dat de voorspelde uitkomst van het neurale netwerk relatief meer afwijkt van de gewenste uitkomst wordt dan dus groter. Dit zorgt voor een groter percentage data dat buiten de gewenste range van 0.8-1.2 valt. Meer categorieën leidt dus automatisch tot een verslechtering van de resultaten.

Meer categorieën leidt echter ook automatisch tot een positieve invloed op de kwaliteit van de uitkomsten. Stel dat het neurale netwerk exact het werkelijke aantal arbeidsplaatsen in een gebied voorspelt voor een bepaalde case. Het neurale netwerk is immers in staat als uitkomst alle waarden aan te nemen. Het voorspelt in dat geval niet exact de juiste categorie, die slechts een vereenvoudiging van de werkelijkheid is. In het geval van een groot aantal categorieën is de kans veel groter dat deze waarde in de buurt van de juiste categorie zit. De afstand tussen de afzonderlijke categorieën is namelijk veel kleiner.

Het is onduidelijk welk effect het sterkst is. Belangrijkste conclusie is dat zeer voorzichtig om dient te worden gegaan met de interpretatie van de uitkomsten. Er is gekozen om de uitkomstvariabelen in 100 categorieën in te delen. Uit de testcase bleek dat dit een goede kalibratie van het bedrijfslocatiemodel oplevert, want een relatief groot aantal van de voorspelde waarden waren juist voorspeld. Daarnaast geven de uitkomsten van het model met 100 categorieën voldoende nauwkeurige informatie voor de gebruiker over het aantal arbeidsplaatsen in een gebied.

6.3 Stappenplan testen variabelen voor bedrijfslocatiemodel

Het is nu duidelijk welke input- en outputvariabelen gebruikt gaan worden om te bepalen of het bedrijfslocatiemodel verbeterd kan worden. Daarnaast zijn de specificaties van elke variabele uitgewerkt. Ook is in vorige paragraaf duidelijk gemaakt op welke manier de kwaliteit van de resultaten gemeten wordt. De volgende stap is om een stappenplan uit te werken om te bepalen welke variabelen daadwerkelijk een bijdrage leveren aan de locatiekeuzes van bedrijven. In figuur 6.3 zijn deze stappen uitgewerkt.



Figuur 6.3: Stappenplan testen variabelen

Toelichting stappenplan:

Stap 1: Kalibratie gehele model

De eerste stap is om het bedrijfslocatiemodel te kalibreren met alle geselecteerde variabelen. Deze uitkomst zal als referentie worden gebruikt om te bepalen wat de kwaliteit is van de afzonderlijke groepen variabelen die zijn gedefinieerd. De kwaliteit van het model wordt tevens geverifieerd aan de hand van de data uit de testcase Twente.

Stap 2: Kalibratie model zonder hoofdvariabelen

Bij tweede stap worden de hoofdvariabelen weggelaten uit het bedrijfslocatiemodel. Het model wordt opnieuw gekalibreerd, nu zonder de hoofdvariabelen. Het valt te verwachten dat de kwaliteit van de voorspellingen zal afnemen. De mate van kwaliteitsdaling van de voorspellingen geeft een indicatie van de mate waarin de betreffende variabelen bijdragen aan de voorspellende kracht van het model en dus in hoeverre deze variabelen de locatiekeuzes van bedrijven beïnvloeden.

Stap 3: Kalibratie model zonder vestigingsplaatsvariabelen

In dit onderdeel wordt hetzelfde gedaan als bij stap 2, alleen nu zonder de vestigingsplaats variabelen. Het bedrijfslocatiemodel wordt dus gekalibreerd zonder de vestigingsplaats variabelen. Ook wordt het model gekalibreerd zonder zowel de hoofdvariabelen als de vestigingsvariabelen. Zo kan het belang van de variabelenset voor het aantal arbeidsplaatsen inzichtelijk worden gemaakt.

Stap 4: Kalibratie model zonder variabelen werkgelegenheidsaandeel van bedrijfssectoren

Deze stap bestaat uit drie onderdelen. De specifieke variabelen voor elke sector worden apart uit het bedrijfslocatiemodel weggelaten. Het bedrijfslocatiemodel wordt dan ook drie keer gekalibreerd, elke keer zonder één van de drie variabelensets voor een bepaalde sector.

Stap 5: Kalibratie model zonder externe variabelen

Bij deze laatste stap is het de bedoeling om te kijken in hoeverre de twee andere deelmodellen, het bereikbaarheidsmodel en het huishoudlocatiemodel van invloed zijn in het bedrijfslocatiemodel. Om deze reden worden de variabelen die beide onderdelen representeren gezamenlijk en afzonderlijk weggelaten uit het model, waarna het model opnieuw gekalibreerd wordt. Net als bij de andere stappen geeft de mate van kwaliteitsdaling van de voorspellingen een indicatie van de mate waarin de betreffende modellen bijdragen aan de voorspellende kracht van het bedrijfslocatiemodel en dus in hoeverre deze modelvariabelen de locatiekeuzes van bedrijven beïnvloeden.

Deze stappen zijn van toepassing op beide onderdelen van het bedrijfslocatiemodel. Er moet immers gekeken worden met welke variabelen het model in staat is het aantal arbeidsplaatsen in een gebied te voorspellen en met welke variabelen het model in staat is het werkgelegenheidsaandeel per bedrijfssector te voorspellen. Voor het modelonderdeel dat

voorspellingen voor het aantal arbeidsplaatsen doet, zullen alle geselecteerde variabelen ook apart worden onderzocht. Het is namelijk ook interessant om te bepalen welke rol de afzonderlijke variabelen in het model spelen. Hiermee kan dieper op de kwaliteit van de variabelenset ingegaan worden. Dit betekent ook dat de variabelen die specifiek voor de drie bedrijfssectoren geselecteerd worden al bij dit onderdeel meegenomen worden in het onderzoek. De reden hiervoor is dat, hoewel deze variabelen volgens de theorie in hoofdstuk 2 een geringere invloed uitoefenen op de locatiekeuzes van bedrijven dan de geselecteerde variabelen, er niet beweerd wordt dat ze in het geheel geen invloed uitoefenen. Aangezien de variabelen toch in variabelenset zijn opgenomen is het neurale netwerk bij uitstek geschikt om te bepalen of deze variabelen toch enige invloed uitoefenen op de locatiekeuzes van bedrijven.

Voor het modelonderdeel dat voorspellingen doet over het werkgelegenheidsaandeel van bedrijfssectoren worden alleen de variabelen die specifiek voor de drie bedrijfssectoren geselecteerd zijn afzonderlijk onderzocht.

6.4 Uitwerking toetsing toepasbaarheid model op alle locaties in NL

In het laatste onderdeel van dit onderzoek wordt de toepasbaarheid van het model op elke willekeurige locatie in Nederland onderzocht. Voor de regio Twente is dit al bekend, voor andere regio's in Nederland nog niet. In paragraaf 5.2 is redelijk gedetailleerd beschreven op welke manier dit gedeelte van het onderzoek gestalte krijgt. Van belang hieraan toe te voegen is dat als input van het model de totale variabelenset genomen wordt die al eerder in dit hoofdstuk gedefinieerd is. Met behulp van deze dataset zal geverifieerd worden of het model in staat is het aantal arbeidsplaatsen en het werkgelegenheidsaandeel van bedrijfssectoren goed kan voorspellen. Er wordt gewerkt met een configuratie van een neurale netwerk met 5 neuronen in de "hidden layer". Aangenomen wordt dat een optimale configuratie voor de regio Twente ook geldt voor andere regio's van vergelijkbare omvang.

6.4.1 Uitwerking selectie regio's

Om in staat te zijn uitspraken te doen over de toepasbaarheid van het model op alle locaties in Nederland worden er zes regio's geselecteerd, waarmee de hypothese dat het model op alle locaties in Nederland toegepast kan worden, getoetst zal worden. Hierbij is het van belang om een goede mix te maken van regio's, zodat de meest voorkomende verschillen tussen Nederlandse gebieden ook terugkomen in de gebiedskenmerken van de geselecteerde regio's. Om deze reden vindt de selectie van de regio's plaats op basis van een drietal uitgangspunten die regioafhankelijk zijn:

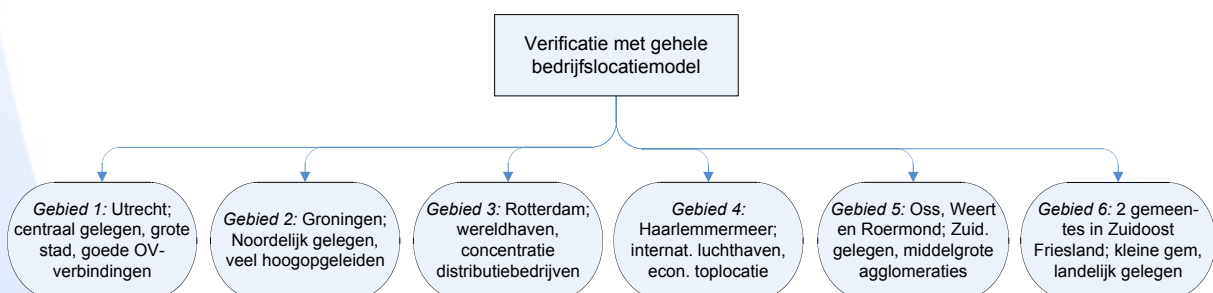
- *Vestigingsplaats kenmerken gedefinieerd door Meester (1999)*; gunstig zijn locaties die een centrale ligging in Nederland hebben, onderdeel van agglomeratie uitmaken, in de buurt van autosnelwegen liggen of een goede positie hebben ten opzichte van de externe markt.
- *De aantal belangrijke locatiefactoren die de locatiekeuze van bedrijven beïnvloeden*; bereikbaarheid van een gebied (in dit geval over de weg, met OV, via de zee of via de lucht), aanwezigheid adequaat opgeleid personeel, aanwezigheid van een zee- of luchthaven en huurprijs & grondprijsniveau.

- *Verschillen in keuzegedrag tussen bedrijven uit verschillende sectoren*; selectie van locatie met hoog percentage bedrijven uit industrie & productie sector, een locatie met hoog percentage bedrijven uit industrie & productie sector en een locatie met een hoog percentage bedrijven uit de diensten sector.

Deze uitgangspunten hebben geleid tot de selectie van de volgende regio's, waarbij in elke regio specifiek één of meerdere locatiekenmerken die in de uitgangspunten zijn gedefinieerd sterk aanwezig zijn:

1. *Regio Utrecht*; heeft een centrale ligging in Nederland, is een grote agglomeratie en heeft een goede bereikbaarheid, speciaal op het gebied van openbaar vervoer (NS knooppunt). Het aandeel bedrijven in de sector diensten is in deze regio hoog (AVV, 2000).
2. *Regio Groningen*; noordelijk gelegen in Nederland, wat betekent dat het een ongunstige ligging heeft ten opzichte van de externe markt (Bijv. Randstad, Roergebied). Daarentegen is het een regio met relatief veel hoogopgeleiden. Ruim 40 procent van de Groningse bevolking heeft minimaal een HBO opleiding (CBS, 2005). De huur- en grondprijzen zijn laag vergeleken met andere stedelijke regio's in het land (VROM, 2005).
3. *Regio Rotterdam*; is het op één na grootste havengebied ter wereld, dus aanwezigheid van een zeehaven in deze regio. Het aandeel bedrijven in de sector handel & distributie is relatief hoog in deze regio (AVV, 2000). Daarnaast is de regio Rotterdam een grote agglomeratie.
4. *Regio Haarlemmermeer*; aanwezigheid van luchthaven Schiphol, een grote internationale luchthaven. Komt in het onderzoek van Bureau Louter (2001) naar voren als de nummer 1 economische toplocatie van Nederland.
5. *Regio's Oss, Weert en Roermond*; alle drie zuidelijk gelegen in Nederland, liggen tussen externe markten in (Randstad, Roergebied en Vlaanderen). Zijn kleine tot middelgrote steden en behoren niet tot een agglomeratie.
6. *Regio Zuidoost Friesland*; zijn twee gemeentes gekozen, te weten Gaasterlân-Sleat en Wymbritseradiel. Het is een landelijk gelegen regio in een perifeer gebied. Het aandeel bedrijven in de sector industrie & productie is hoog in deze regio (AVV, 2000). Daarentegen zijn de huur- en grondprijzen relatief laag vergeleken met andere regio's in het land (VROM, 2005).

Hiermee zijn de zes geselecteerde regio's bekend. Het geheel is nog eens overzichtelijk samengevat in figuur 6.6. Dit kan tevens uitgelegd worden als een stappenplan voor het onderzoeken voor de verificatie van de toepasbaarheid van het bedrijfslocatiemodel op elke willekeurige locatie.



Figuur 6.6: Schema toetsing toepasbaarheid model op alle locaties in NL

7 Resultaten onderzoek

7.1 Inleiding

In dit hoofdstuk worden de resultaten van het onderzoek gepresenteerd. Hier wordt duidelijk of het vernieuwde bedrijfslocatiemodel daadwerkelijk de taak aankan waarvoor het gebouwd is, namelijk het voorspellen van de bedrijvigheid in een gebied. Dit kan pas worden gedaan als bekend is wat het optimale aantal neuronen in de “hidden layer” van het neurale netwerk zijn voor de voorspelling van het aantal arbeidsplaatsen en de voorspelling van de werkgelegenheidsverdeling per bedrijfssector. Om deze reden zal in paragraaf 7.2 hier eerst op worden ingegaan. Daarnaast zijn er twijfels over de kwaliteit van de variabele “huurprijs- /grondprijs niveau”. In paragraaf 7.3 staat het resultaat van toetsing van deze variabele, zodat definitief bepaald kan worden of deze variabele opgenomen dient te worden in de verklarende variabelenset. Nadat deze zaken zijn bekend zijn, kan onderzocht worden of het vernieuwde bedrijfslocatiemodel de bedrijvigheid in een gebied kan voorspellen en wat de invloed is van de afzonderlijke variabelen op de locatiekeuzes van ondernemers. In paragraaf 7.4 worden de resultaten getoond van de kalibratie van het model. In paragraaf 7.5 staan de resultaten van het onderzoek van de vraag of het bedrijfslocatiemodel toepasbaar is op gebiedsniveau. De resultaten van het onderzoek naar de toepasbaarheid van het model op elke willekeurige locatie in Nederland staan in paragraaf 7.6 gepresenteerd.

7.2 Configuratie neuraal netwerk in bedrijfslocatiemodel

In deze paragraaf is het de bedoeling om de optimale configuratie van het neurale netwerk te vinden. Hiermee wordt, waarvoor eerder al in paragraaf 3.4 de reden is aangegeven, een benadering van de optimale configuratie van het neurale netwerk bedoeld. In hoofdstuk 3 is verder duidelijk geworden dat een slechte configuratie zorgt voor een beperkt “lerend vermogen” van het neurale netwerk. Hierdoor kunnen de variabelen die de locatiekeuzes van bedrijven beïnvloeden niet goed door het neurale netwerk ontdekt worden.

Er wordt zowel voor de kalibratie van het model als voor de toetsing van de methodiek van het model op gebiedsniveau afzonderlijk bekeken wat de optimale configuratie van het neurale netwerk is. Dit betekent dat er in dit onderzoek twee verschillende neurale netwerken gebruikt worden om de kwaliteit van het bedrijfslocatiemodel te testen.

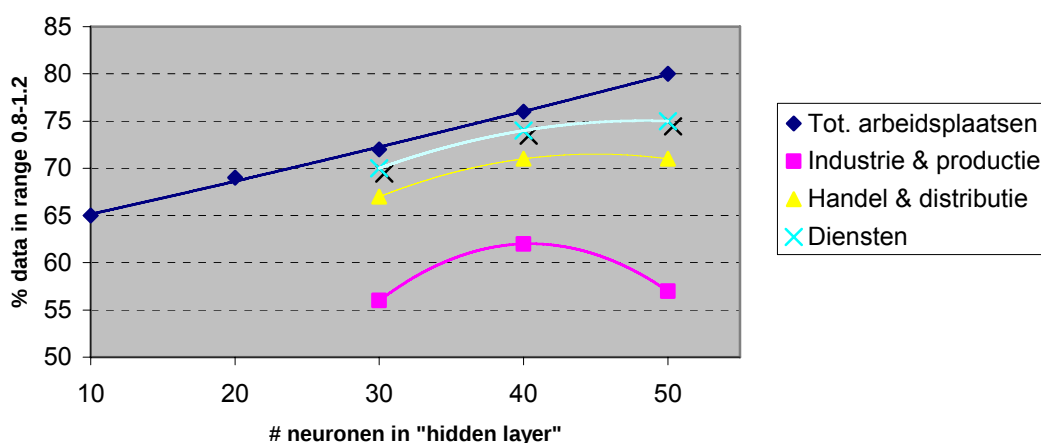
Er zijn een aantal configuratieopties van het neurale netwerk die in alle onderdelen van het onderzoek hetzelfde zijn. Zoals eerder al vermeld, staat in bijlage B een voorbeeld van de configuratie van het neurale netwerk in Matlab. Hierin worden tevens de gemaakte keuzes betreffende de standaard configuratieonderdelen van het neurale netwerk toegelicht. De optimale configuratie van het aantal neuronen in de “hidden layer” is niet bepaald in bijlage B. Dit wordt in de volgende twee subparagrafen uitgewerkt.

7.2.1 Configuratie voor kalibratie neuraal netwerk

Het bedrijfslocatiemodel bestaat uit twee afzonderlijke onderdelen, zodat voor beide onderdelen naar het optimale aantal neuronen in de “hidden layer” gezocht zal worden. De reden hiervoor is dat er wellicht per bedrijfssector andere verbanden bestaan die de locatiekeuzes van de betreffende bedrijfssector voorspellen. Om deze verbanden juist te doorgronden is wellicht een

ander aantal neuronen in de “hidden layer” nodig. Ten tweede verschilt de outputvariabele van beide onderdelen van het bedrijfslocatiemodel in kwantiteit. Het totale aantal arbeidsplaatsen per bedrijfssector bedraagt namelijk globaal genomen slechts een derde van het totale aantal arbeidsplaatsen. Hierdoor wordt het absolute onderscheid tussen de 100 gebruikte categorieën kleiner. De grootte van de datarange binnen één categorie wordt immers minder groot. De kans dat gebieden met een ongeveer evenredige grote bedrijfssector (en dus met dezelfde kenmerken) in andere categorieën komen te liggen wordt groter. Een logische stap om dit effect op te lossen lijkt het verminderen van het aantal categorieën voor een bedrijfssector. Het blijkt echter na een korte testcase dat dit geen noemenswaardig effect heeft op de kwaliteit van de uitkomsten. Blijkbaar ondervindt het neurale netwerk geen hinder van het relatief kleinere onderscheid tussen categorieën.

Voor de kalibratie van het model met verschillende aantallen neuronen in de “hidden layer” zijn de resultaten in figuur 7.1 neergezet. In bijlage E staat een uitgebreidere beschrijving van de resultaten. Er is geen onderzoek gedaan naar de prestaties van neurale netwerken met meer dan 50 neuronen. De reden hiervoor is, dat de rekentijd die een computer nodig heeft om een iteratie van het neurale netwerk uit te voeren bij meer dan 50 neuronen eenvoudigweg te hoog wordt.



Figuur 7.1: Bepaling optimale aantal neuronen voor kalibratie model

Uit figuur 7.1 blijkt dat het onderdeel van het model dat voorspellingen doet over het totale aantal arbeidsplaatsen in een gebied de beste resultaten geeft als een “hidden layer” met 50 neuronen gebruikt wordt. Het lijkt erop dat de resultaten met meer neuronen in de “hidden layer” nog verder zullen verbeteren. Doordat maximaal 50 neuronen gebruikt zullen worden, is 50 het aantal neuronen waarmee het neurale netwerk gekalibreerd zal worden voor de voorspellingen van het aantal arbeidsplaatsen in een gebied.

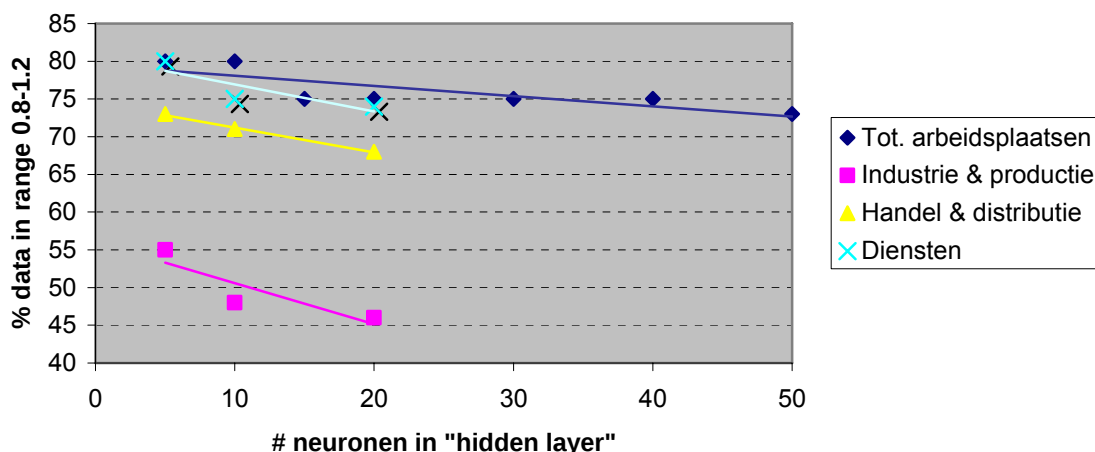
Het tweede onderdeel, dat voorspellingen doet over het werkgelegenheidsaandeel van bedrijfssectoren, is opgedeeld in drie onderdelen (de drie gedefinieerde bedrijfssectoren) die alle drie apart gekalibreerd worden en dus ook een apart geconfigureerd neuraal netwerk nodig hebben. Voor de kalibratie van het onderdeel industrie & productie is duidelijk dat een “hidden layer” met 40 neuronen de beste resultaten geeft. Voor de sector handel & distributie en de sector diensten is dit minder duidelijk. Het optimale aantal neuronen in de “hidden layer” ligt hier ergens tussen de 40 en 50. Om hier een beter beeld van te krijgen is ook het gemiddelde bekeken van de 5 runs die voor het verkrijgen van elk resultaat zijn gemaakt. Zoals eerder al in dit hoofdstuk is

aangegeven worden voor elke stap immers 5 runs uitgevoerd, om zo zekerheid te hebben over een optimaal gekalibreerd model. Uit de gemiddelden blijkt dat voor de sector handel & distributie het gemiddelde van een “hidden layer” met 40 neuronen veel hoger ligt dan een van een “hidden layer” met 50 neuronen. Voor de sector diensten zijn deze gemiddelden echter even hoog. Om deze reden zal het neurale netwerk voor de sector handel & distributie en de sector diensten met respectievelijk 40 en 45 neuronen in de “hidden layer” gekalibreerd worden.

7.2.2 Configuratie voor toetsing methodiek model op gebiedsniveau

Het is nu duidelijk met welke aantallen neuronen in de “hidden layer” het neurale netwerk gekalibreerd wordt. In deze paragraaf wordt duidelijk gemaakt wat de optimale configuraties zijn voor de toetsing van de methodiek van het model op gebiedsniveau, waarvoor de testcase Twente wordt gebruikt. Dezelfde overwegingen die uiteengezet zijn in de vorige paragraaf betreffende mogelijke verschillen tussen de twee onderdelen van het bedrijfslocatiemodel, zijn ook hier van toepassing.

Voor de bepaling van de optimale configuratie is het model getest met verschillende aantallen neuronen in de “hidden layer”. De resultaten zijn in figuur 7.2 neergezet. In bijlage E staat een uitgebreidere beschrijving van de resultaten.



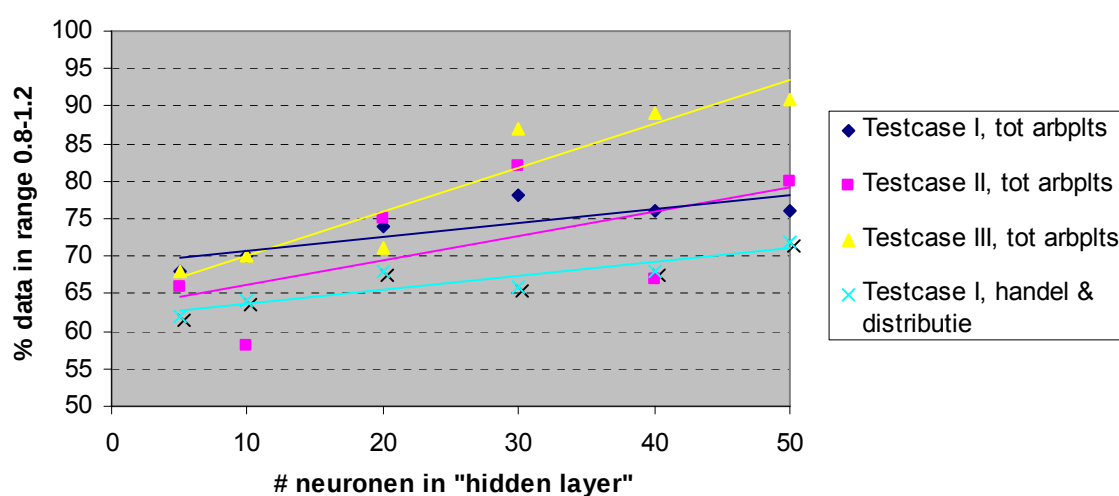
Figuur 7.2: Bepaling optimale aantal neuronen voorspelling testcase Twente

Uit figuur 7.2 blijkt dat er tussen de vier onderdelen waarvoor het bedrijfslocatiemodel voorspellingen doet, geen verschillen bestaan betreffende de optimale configuratie van het aantal neuronen in de “hidden layer”. De resultaten van het model met de testcase Twente levert in alle gevallen de beste resultaten op met 5 neuronen in de “hidden layer”. Het gebruik van minder dan 5 neuronen in de “hidden layer” is geen optie. Het neurale netwerk belandt dan snel op het punt waar het slechts zeer eenvoudige relaties kan herkennen.

Opvallend is dat het model voor voorspellingen op gebiedsniveau de beste prestaties lijkt te leveren met een veel lager aantal neuronen in de “hidden layer” dan bij de kalibratie van het model. Dit zou verklaard kunnen worden door de eigenschappen van de dataset van Twente, die wellicht een groot aantal cases bevat die een afwijkend beeld vertonen in vergelijking met het landelijke gemiddelde. Met andere woorden, de dataset van Twente bevat veel ruis. Hierdoor worden de

voorspellingen beter als het neurale netwerk alleen de eenvoudige verbanden schat (door het gebruik van weinig neuronen), zodat deze ruis in de data genegeerd wordt.

Dit zou betekenen dat een vergelijkbaar test met een dataset van vergelijkbare grootte die uit cases bestaat die willekeurig uit de dataset van Nederland zijn getrokken een ander beeld op zou moeten leveren. Hierbij zouden de resultaten wel moeten overeenstemmen met de resultaten van de kalibratie van het model in figuur 7.1. Om deze reden zijn drie testcases geselecteerd die uit 50 gebieden (cases) bestaan. Deze gebieden zijn willekeurig uit de dataset getrokken. Voor deze testcases zijn opnieuw het optimale aantal neuronen in de “hidden layer” bepaald, voor de voorspelling van het aantal arbeidsplaatsen in een gebied. Daarnaast is voor de eerste testcase hetzelfde gedaan voor de voorspelling van het aantal arbeidsplaatsen in de sector handel & distributie.



Figuur 7.3: Bepaling optimale aantal neuronen voorspelling verschillende testcases

Figuur 7.3 bevestigt de veronderstelling dat de dataset van Twente de nodige ruis bevat, wat ervoor zorgt dat het model voor Twente de beste voorspellingen doet met 5 neuronen in de “hidden layer”. De resultaten van de testcases in figuur 7.3 komen namelijk wel overeen met de resultaten bij de kalibratie van het model. Een neurale netwerk met 50 neuronen in de “hidden layer” zorgt ervoor dat het model de beste voorspellingen kan doen op gebiedsniveau. Voor de testcase Twente blijft gewerkt worden met een neurale netwerk met 5 neuronen in de “hidden layer”. Met deze testcase wordt immers ook gekeken hoe de afzonderlijke variabelen zich gedragen op gebiedsniveau. Dit kan het best onderzocht worden aan de hand van een optimaal gekalibreerd neurale netwerk voor dit specifieke gebied. In verband met de beschikbare tijd in dit onderzoek zal ook voor de toetsing van het model in de zes geselecteerde regio's volstaan worden met een eenvoudig netwerk van 5 neuronen.

7.3 Bepaling definitieve variabelenset

In paragraaf 6.1 zijn de variabelen die als input dienen voor het bedrijfslocatiemodel geoperationaliseerd en daarbij zijn om verschillende redenen al een aantal variabelen afgefallen. Ook de variabele “huurprijs- /grondprijsniveau” is geoperationaliseerd. Deze variabele wordt gebruikt om twee factoren uit te drukken. Dit zijn de factoren representatie gebouw en huurprijs-/grondprijsniveau in een gebied. Toch zijn er twijfels over de bruikbaarheid van deze variabele in

het onderzoek. De reden is de beschikbare hoeveelheid data van deze variabele. De data van deze variabele is slechts voor een derde van de cases bekend. Dit betekent dat als deze variabele meegenomen wordt er slechts een derde van de cases overblijft. Bij het ontbreken van dergelijke hoeveelheden data is het neurale netwerk niet meer in staat dit te negeren. In dit geval is de enige oplossing om deze cases uit de dataset te verwijderen. Dit heeft tot gevolg dat de overgebleven cases niet meer representatief zijn voor heel Nederland. Dit is juist een belangrijke voorwaarde voor het model. Kortom, de beperkte beschikbaarheid van de data zorgt voor een dilemma, daar het wel zeer interessant is om de invloed van deze variabele op de locatiekeuzes van bedrijven te onderzoeken.

Om het probleem op te lossen is ervoor gekozen deze variabele uit het totale onderzoek te lichten en apart te bekijken. Dit houdt in dat er kortstondig met een beperkte, aangepaste dataset gewerkt gaat worden. Hierin blijven de cases over waarvoor de hoogte van het huurprijs- /grondprijsniveau wel bekend is. Met deze dataset kan wel onderzocht worden of de variabele invloed heeft op de locatiekeuzes van bedrijven. Het gaat hier dan nog om de locatiekeuzes van bedrijven uit een beperkt aantal gebieden in Nederland. In figuur 7.4 stellen de donkere stippen op de kaart van Nederland de gebieden voor waar wel gegevens van bekend zijn. Dit zijn overwegend stedelijke gebieden.

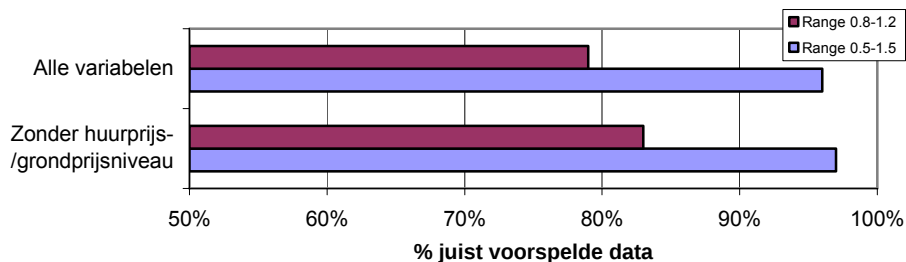


Figuur 7.4: Overzicht herkomst resterende data

Het neurale netwerk zal eerst gekalibreerd worden met de totale geselecteerde variabelenset. Als output waarde wordt het totale aantal arbeidsplaatsen in een gebied getracht te voorspellen.

De configuratie van het neurale netwerk staat beschreven in bijlage B. Alleen het aantal neuronen in de “hidden layer” wijkt af van het eerder vastgestelde aantal voor de kalibratie van dit onderdeel van het model. Er is gekozen om maximaal 10 neuronen te gebruiken in de “hidden layer”, omdat bij meer neuronen het gevaar bestaat dat de verhouding tussen het geringe aantal cases en het aantal vrijheidsgraden evenredig wordt. Hierdoor ontstaat er in het neurale netwerk voor elke case een vergelijking. Dit betekent dat het neurale netwerk niet meer de achterliggende verbanden doorgrondt.

In de tweede stap wordt het neurale netwerk opnieuw gekalibreerd, dit keer met een variabelenset waar de variabele “huurprijs- /grondprijsniveau” is uitgehaald. De mate waarin de kwaliteit van de uitkomsten verandert, geeft een indicatie over de invloed van deze variabele op de locatiekeuzes van bedrijven. De resultaten van deze twee stappen zijn in figuur 7.5 neergezet.



Figuur 7.5: Toetsing variabele huurprijs- /grondprijsniveau

Uit figuur 7.5 blijkt dat de kwaliteit van de voorspellingen van het bedrijfslocatiemodel zonder de variabele “huurprijs- /grondprijsniveau” niet verslechtert. Integendeel, de kwaliteit lijkt zelfs iets beter te worden. Wellicht geeft de variabele “huurprijs- /grondprijsniveau” een vertekend beeld, waardoor een dataset zonder deze variabele het neurale netwerk beter in staat stelt verbanden tussen de overige variabelen en de locatiekeuzes van bedrijven te ontdekken.

Het is duidelijk dat deze variabele geen bijdrage levert aan de voorspelling van het aantal arbeidsplaatsen in een gebied. Door de duidelijke resultaten wordt ook aangenomen dat dit het geval is bij het tweede onderdeel, het werkgelegenheidsaandeel van bedrijfssectoren. De variabelen “representatie gebouw” en “huurprijs- /grondprijsniveau” in een gebied worden niet in de definitieve variabelenset opgenomen.

7.3.1 Overzicht definitieve variabelenset

De afwegingen in hoofdstuk 6 en de bevindingen die eerder in deze paragraaf zijn gedaan zorgen voor de keuze van een definitieve variabelenset zoals beschreven in tabel 7.1. Deze variabelen dienen als input voor het bedrijfslocatiemodel.

Onderdeel	Variabelen
Hoofdvariabelen	• Parkeerfaciliteiten • Aanwezigheid adequaat opgeleid personeel • Aantal bedrijfsvestigingen • Stedelijkheid gebied
Vestigingsplaatsvariabelen	• Ligging in NL • Agglomeratie-effect • Infrastructuur • Oriëntatie t.o.v. externe markt
<i>Sectorspecifieke variabelen</i>	
Industrie & productie	• Uitbreidingsmogelijkheden • Toeleverende bedrijven • Logistieke dienstverlening
Handel & distributie	• Aanwezigheid zee- of luchthaven • Logistieke dienstverlening
Diensten	• Quality of life factoren • Representatie omgeving • OV-bereikbaarheid
<i>Externe modelvariabelen</i>	
Locatie huishoudens	• Aantal huishoudens • % 1-persoons huishoudens • % huishoudens zonder kinderen • % huishoudens met kinderen • WOZ-waarde
Bereikbaarheid	• Bereikbaarheid

Tabel 7.1: Overzicht inputvariabelen bedrijfslocatiemodel

De invloed van deze variabelen wordt in de volgende paragrafen met behulp van het bedrijfslocatiemodel onderzocht. Er is echter een groot aantal combinaties mogelijk van variabelen die onderzocht kunnen worden. Om het onderzoek gestructureerd te houden, zijn er in tabel 7.1 zes categorieën gecreëerd. Naast het afzonderlijk onderzoeken van elke variabele zullen deze categorieën onderzocht worden. Uitgangspunt bij het beoordelen van de variabelen zal de totale variabelenset zijn. Dit betekent dat bij het weglaten van een variabele of categorie uit de variabelenset, steeds gekeken wordt naar de verandering van kwaliteit in de resultaten ten opzichte van de totale variabelenset. In bijlage F1 en bijlage F2 staan overzichten van data die verkregen is bij het uitvoeren van het onderzoek.

De figuren 7.6 tot en met 7.16 in de volgende twee paragrafen zijn gebaseerd op deze data. In de figuren is voor elke onderzochte variabelenset het percentage door het model voorspelde data uitgezet, dat binnen een bepaalde range van de werkelijke uitkomsten ligt. In elke grafiek stellen de onderste twee balken de resultaten voor van een variabelenset waarin alle variabelen zijn opgenomen. Dit wordt, zoals eerder al aangegeven, als referentie genomen. Om in staat te zijn de resultaten beter te kunnen vergelijken, is een doorzichtige balk verticaal over de resultaten heen geplaatst. Dit stelt het gebied voor waarbij de afwijking van de kwaliteit ten opzichte van de kwaliteit van de resultaten van de totale variabelenset dusdanig gering is, dat niet kan worden bewezen dat er aanwijsbare verschillen bestaan. Dit betekent dat de betreffende variabele of categorie geen bijdrage levert aan de voorspelling van het aantal arbeidsplaatsen in het bedrijfslocatiemodel. De reden dat er een klein gebied genomen wordt waarop dit van toepassing is, volgt uit de eigenschappen van het neurale netwerk. In de kwaliteit van de uitkomsten van een neurale netwerk zit bij elke run immers een kleine fluctuatie, zelfs als er voor elke test van een variabele of categorie vijf runs worden gemaakt.

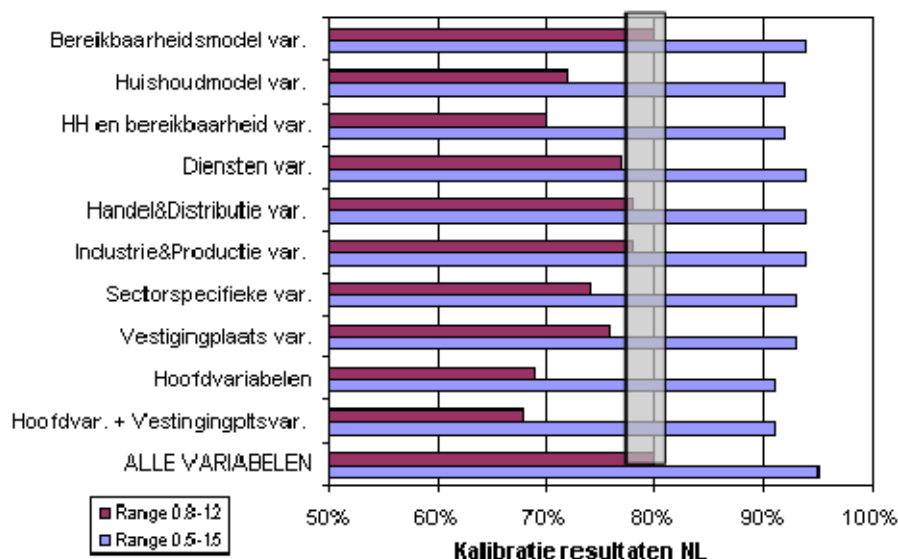
7.4 Resultaten kalibratie model

In dit deel van het onderzoek wordt het bedrijfslocatiemodel gekalibreerd met de in paragraaf 7.3 gedefinieerde variabelenset. Hiervoor worden de cases uit heel Nederland gebruikt. De kwaliteit van de voorspellingen wordt getoetst aan de hand van de afwijking van de voorspelde uitkomsten door het model ten opzichte van de werkelijke uitkomsten van dezelfde cases uit heel Nederland. De kalibratie van het model vindt plaats voor zowel het voorspellen van het aantal arbeidsplaatsen als voor het voorspellen van het werkgelegenheidsaandeel per bedrijfssector.

7.4.1 Resultaten voorspelling aantal arbeidsplaatsen

In figuur 7.6 zijn de resultaten per categorie neergezet voor de voorspelling van het aantal arbeidsplaatsen door het gekalibreerde model.

Uit de resultaten blijkt dat 80 procent van de data binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Zoals verwacht blijken de hoofdvariabelen de meeste invloed te hebben op de kwaliteit van de uitkomsten. De vestigingsplaats variabelen leveren een bijdrage, zij het gering, aan de prestaties van het model. Dit is niet in overeenstemming met de theorie van Meester (1999), waarin een veel groter belang van vestigingsplaats variabelen wordt verondersteld. De variabelen die specifiek voor de bedrijfssectoren zijn geselecteerd voldoen aan de verwachtingen en hebben slechts een geringe invloed op de kwaliteit van de voorspellingen van het totale aantal arbeidsplaatsen.



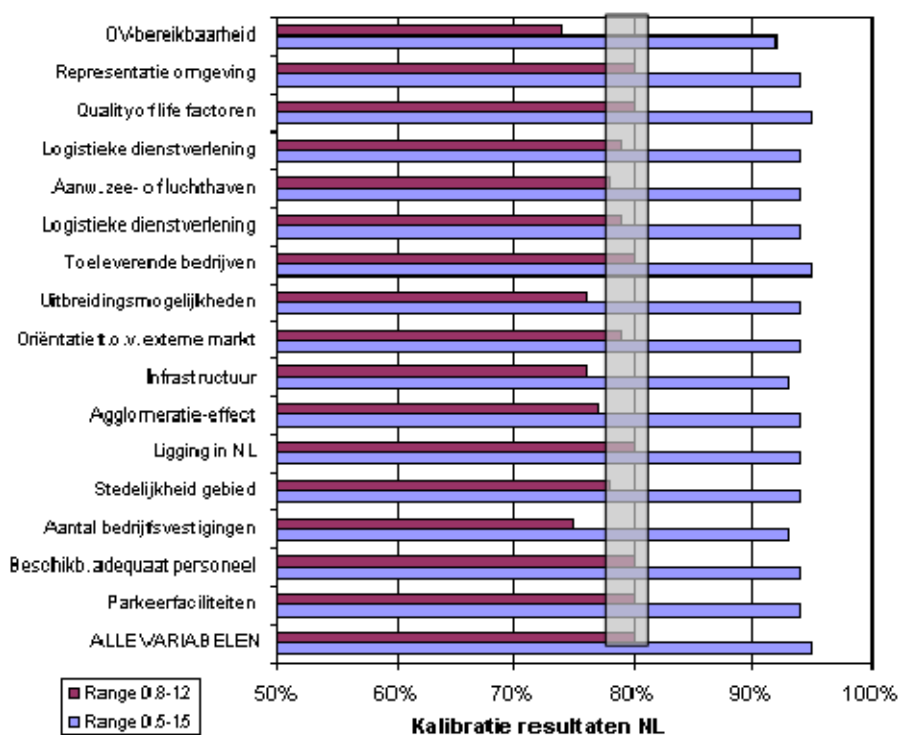
Figuur 7.6: Kwaliteit voorspellingen aantal arbeidsplaatsen (deel 1)

Als laatste zijn de resultaten van de twee andere externe modellen in figuur 7.6 opgenomen. De resultaten tonen de invloed aan van het huishoudmodel in het bedrijfslocatiemodel. De resultaten laten namelijk zien dat de voorspellingen van het bedrijfslocatiemodel minder goed zijn als de variabelen uit het huishoudmodel weg worden gelaten. Dit is een bevestiging van de interactie tussen huishoudens en bedrijven. Hiermee wordt de theorie achter LUTI modellen onderbouwd, waarin onder meer een dergelijke interactie voorspeld wordt. Het resultaat van het bereikbaarheidsmodel laat echter een heel ander beeld zien. De resultaten tonen aan dat het bereikbaarheidsmodel geen invloed heeft in het bedrijfslocatiemodel. Dit is in tegenspraak met de gangbare theorieën, waarin de bereikbaarheid juist als belangrijkste factor voor locatiekeuzes van bedrijven wordt gezien. Het sluit echter wel aan bij de twijfels die Tillema (2004) uit in zijn onderzoek over de bereikbaarheidsvariabele. De manier waarop de bereikbaarheid in het model van Tillema is uitgedrukt, in potentiële (auto)bereikbaarheid, sluit het best aan bij de locatiekeuzes van huishoudens. Volgens Tillema blijkt uit literatuur namelijk dat bedrijven en huishoudens door verschillende vormen van bereikbaarheid beïnvloed worden. De resultaten bevestigen dat locatiekeuzes van bedrijven niet beïnvloed worden door de manier waarop bereikbaarheid in het model van Tillema is uitgedrukt.

In figuur 7.7 zijn de resultaten per variabele neergezet voor de voorspelling van het aantal arbeidsplaatsen door het gekalibreerde model. Deze resultaten zijn een verdere uitwerking van de resultaten in figuur 7.6. Hiermee worden voor elke categorie de afzonderlijke variabelen belicht.

Uit de resultaten blijkt dat van de afzonderlijke hoofdvariabelen alleen de variabele “aantal bedrijfsvestigingen” invloed heeft op de kwaliteit van de voorspellingen van het model. De variabelen “parkeerfaciliteiten” en “beschikbaarheid adequaat opgeleid personeel” hebben geen invloed op de resultaten. Dit betekent dat deze variabelen geen invloed hebben op de locatiekeuzes van bedrijven in het model. Dit is in tegenstelling met de literatuur waarin verondersteld wordt dat deze variabelen significante invloed hebben op de locatiekeuzes van ondernemers. In de categorie vestigingsplaats variabelen hebben de variabelen “agglomeratie-

effect” en “infrastructuur” invloed op de resultaten van het model. Eerdere resultaten in figuur 7.6 lieten een geringere invloed van de vestigingsplaats variabelen zien dan op basis van de theorie van Meester (1999) mocht worden verondersteld. Het feit dat ten minste twee van de vier variabelen bijdragen aan de kwaliteit van de voorspellingen van het model betekent dat de invloed van de vestigingsplaats variabelen niet onderschat mag worden en wel degelijk bestaat.



Figuur 7.7: Kwaliteit voorspellingen aantal arbeidsplaatsen (deel 2)

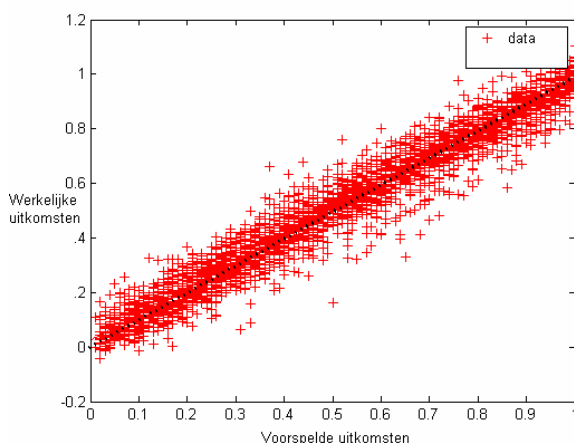
De variabele “uitbreidingsmogelijkheden” valt niet in de categorie hoofdvariabelen en ook niet in de categorie vestigingsplaats variabelen. Het is dus niet één van de belangrijke variabelen die volgens de literatuur de locatiekeuzes van ondernemers beïnvloeden. Toch heeft deze variabele invloed op de voorspelling van het totale aantal arbeidsplaatsen. Dit betekent dat op basis van de resultaten blijkt dat deze variabele invloed heeft op de locatiekeuzes van ondernemers. Het lijkt erop dat de definitie van deze variabele, die uitgedrukt wordt in de beschikbaarheid van bedrijfsterrein, niet alleen de uitbreidingsmogelijkheden op een terrein beschrijft, maar ook de laad- en loscapaciteit op een terrein. Deze twee factoren vertonen namelijk sterke overeenkomsten. Uit de literatuur volgt daarnaast dat de factor laad- en loscapaciteit van een terrein een belangrijke locatiefactor is waarop ondernemers hun keuze baseren, net iets minder belangrijk dan de factoren die geselecteerd zijn bij de categorie hoofdvariabelen. Deze twee zaken gecombineerd vormen een verklaring voor het belang van de variabele “uitbreidingsmogelijkheden” in het model.

Uit de resultaten blijkt dat de bereikbaarheid in het model toch een rol speelt. Uit figuur 7.6 was al gebleken dat de bereikbaarheidsvariabele uit het bereikbaarheidsmodel geen rol speelt in het bedrijfslocatiemodel. Hierbij werd Tillema (2004) aangehaald die al veronderstelde dat bedrijven wellicht hun locatiekeuzes laten afhangen van een andere vorm van bereikbaarheid. De resultaten in figuur 7.7 bevestigen deze veronderstelling. Er zijn twee variabelen in het model die (een gedeelte van) bereikbaarheid anders uitdrukken dan in het bereikbaarheidsmodel. Dit zijn de variabelen “infrastructuur” en “OV-bereikbaarheid”. Uit de resultaten blijkt dat deze variabelen

invloed hebben op de kwaliteit van de voorspellingen van het model. Dit betekent dat de bereikbaarheid wel degelijk invloed uitoefent op locatiekeuzes van ondernemers.

7.4.2 Een deelresultaat uitgelicht

Een andere manier om de resultaten inzichtelijk te maken is om de voorspellingen van het model uit te zetten tegen het werkelijke aantal arbeidsplaatsen die per gebied aanwezig zijn. Hiermee kan onder andere een goed beeld worden verkregen van de extremen in de voorspellingen van het model. Op de x-as zijn de voorspelde uitkomsten neergezet, op de y-as de werkelijke uitkomsten. Liggen de punten exact op de lijn $y=x$ dan heeft het model de werkelijkheid juist voorspeld. Opgemerkt dient te worden dat de uitkomsten genormaliseerd zijn. Dit betekent dat alle waarden tussen de 0 en de 1 liggen. In figuur 7.8 zijn de resultaten afgebeeld van een run van het neurale netwerk waarbij het model gekalibreerd is met alle variabelen.



Figuur 7.8: Uitkomsten gekalibreerde model met alle variabelen

In figuur 7.8 wordt duidelijk dat de meeste uitkomsten zich rond de lijn $y=x$ bevinden. Dit komt overeen met de eerder geziene resultaat van 80 procent binnen een range van 0.8-1.2 en 95 procent binnen een range van 0.5-1.5. Toch zijn er ook een aantal uitkomsten te zien die een redelijke afwijking vertonen van de optimale lijn $y=x$. Het is interessant om te kijken welke cases deze extremen veroorzaken en welke overeenkomsten in deze cases hebben.

Als extremen zijn alle case geselecteerd die buiten de range 0.5-1.5 vallen. In dit geval betreft dit 5 procent van de data. De volgende overeenkomsten kunnen bij deze cases geconstateerd worden:

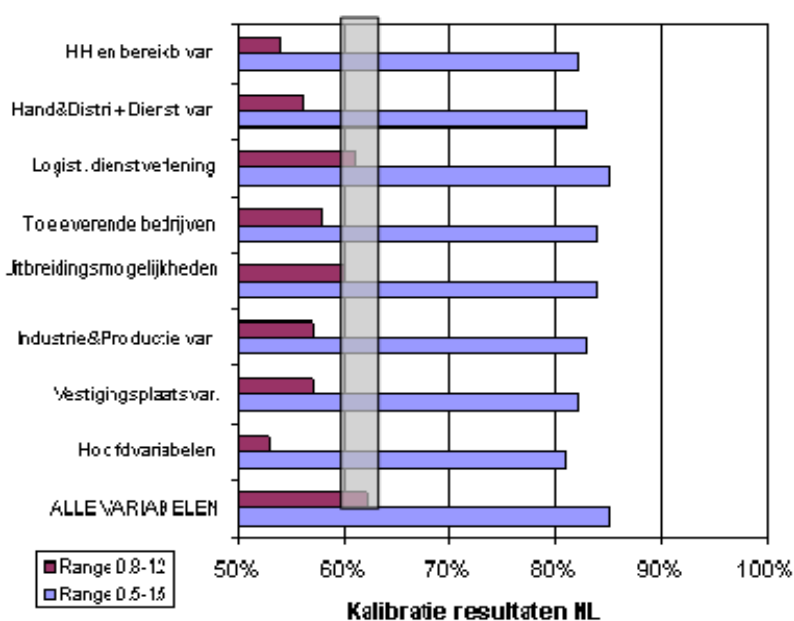
- Ruim 90 procent van de cases heeft een aantal arbeidsplaatsen in de cat. 20 of lager.
- Bijna 70 procent van de cases heeft een aantal arbeidsplaatsen in de cat. 10 of lager.
- Van de cases komt ruim 65 procent uit Noord Nederland (+ kop van Overijssel).
- Van de cases heeft 98 procent een stedelijkheidsgraad van vijf, wat betekent dat het een niet stedelijk gebied betreft.
- Het percentage gebieden met stedelijkheidsgraad vijf en dat in Noord Nederland (+ kop van Overijssel) ligt bedraagt ongeveer 65 procent.

Er zijn veel cases die als outputwaarde een lage categorie hebben. Immers, de output van het model, het aantal arbeidsplaatsen in een gebied, is opgedeeld in 100 categorieën. In hoofdstuk 6 is al betoogd dat kleine outputwaarden eerder buiten een range vallen dan grote waarden en dus

eerder relatief grote afwijkingen vertonen van de gewenste output. Het hoge aantal cases uit niet stedelijke gebieden lijkt hiermee verband te houden. In landelijke gebieden ligt over het algemeen het aantal arbeidsplaatsen lager dan in stedelijke gebieden. Daarnaast is opvallend dat het aantal cases uit Noord Nederland zeer hoog is. Dit valt zeer waarschijnlijk voor het grootste gedeelte te verklaren uit het overwegend landelijke karakter van dit gebied. Het sluit dan ook aan op de hierboven gedane veronderstelling. Wellicht speelt de theorie van Meester (1999) ook een rol, die beweert dat verschillen in waardering ook afhankelijk zijn van de regio waaruit de ondernemer afkomstig is. De mogelijkheid bestaat dat de locatievoorkeuren van ondernemers uit het Noorden afwijkende locatievoorkeuren hebben. Het is echter niet mogelijk om dit op basis van deze analyse van extremen te bewijzen.

7.4.3 Resultaten voorspelling werkgelegenheidsaandeel per bedrijfssector

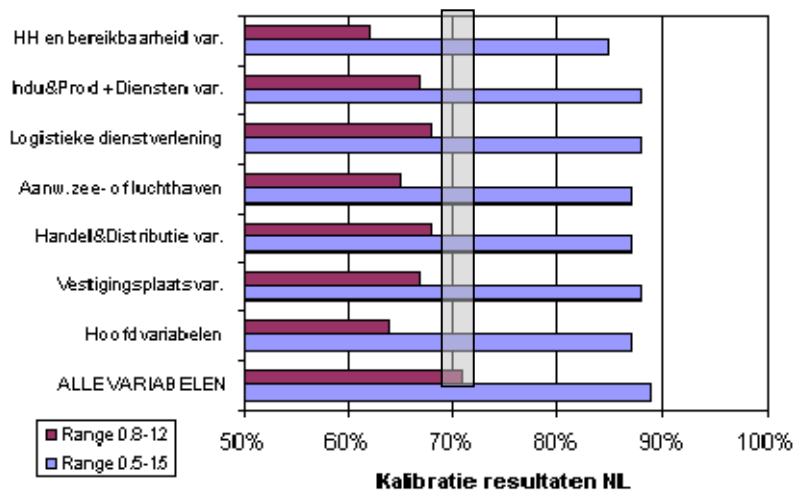
In het tweede onderdeel van het model worden de werkgelegenheidsaandelen per bedrijfssector voorspeld. Hiervoor dient het model opnieuw gekalibreerd te worden. Zoals eerder al in hoofdstuk 6 aangegeven, zijn nu slechts de categorieën variabelen en de sectorspecifieke variabelen onderzocht. De resultaten zijn in drie figuren uitgezet. In figuur 7.9 staan de resultaten van het onderzoek voor de sector productie & industrie.



Figuur 7.9: Kwaliteit voorspellingen sector industrie & productie

Uit de resultaten blijkt dat ruim 60 procent van de data binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Dit percentage ligt onder de in paragraaf 6.2 gestelde grens van 70 procent. Dit was percentage vanaf waar gesproken mocht worden van een goed modelresultaat. Dit betekent dat uit de resultaten blijkt dat het bedrijfslocatiemodel niet in staat is goede voorspellingen te doen voor de sector industrie & productie. Dit betekent dat de locatiekeuzes van ondernemers uit de sector industrie & productie verschilt van de algemene locatiekeuzes. Het bedrijfslocatiemodel is immers op basis van dezelfde variabelenset wel in staat voorspellingen te doen voor het totale aantal arbeidsplaatsen. Doordat het model niet in staat is goede voorspellingen te doen, zijn de resultaten van de afzonderlijk variabelen niet betrouwbaar genoeg om uitspraken over te kunnen doen.

In figuur 7.10 staan de resultaten van het onderzoek voor de sector handel & distributie.



Figuur 7.10: Kwaliteit voorspellingen sector handel & distributie

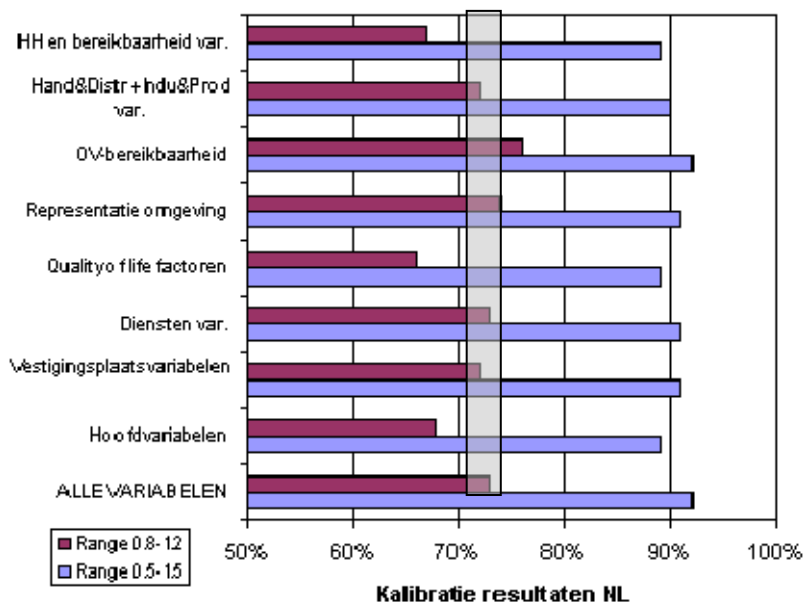
Uit de resultaten blijkt dat ruim 70 procent van de data binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Het blijkt dat voor de sector handel & distributie zowel de hoofdvariabelen als de vestigingsvariabelen van invloed zijn op de kwaliteit van de voorspellingen van het model. Dit komt overeen met de bevindingen uit de literatuur. Daarnaast is ook de input van de externe modellen, het huishoudmodel en het bereikbaarheidsmodel van invloed op de kwaliteit van de resultaten. Dit is opnieuw een bevestiging van de interactie tussen huishoudens en/of transport en bedrijven en sluit aan op de theorie achter LUTI modellen. Dit keer gaat het specifiek om de interactie tussen bedrijven uit de handel & distributie sector en huishoudens en/of transport. De invloed van de bereikbaarheid (transport) is niet afzonderlijk onderzocht. Er kunnen dan ook geen uitspraken over worden gedaan over de invloed van bereikbaarheid op de locatiekeuzes van ondernemers uit de sector handel & distributie.

Reijs (2001) en Everaar en Brand-van Tuijn (2001) tonen aan dat bedrijven uit de sector handel & distributie specifiek veel waarde hechten aan onder meer de variabele “aanwezigheid zee- of luchthaven” en “logistieke dienstverlening”. De resultaten uit figuur 7.10 bevestigen deze conclusie. De kwaliteit van de voorspellingen van het bedrijfslocatiemodel zijn voor de sector handel & distributie namelijk beter als deze variabelen opgenomen zijn in de variabelenset. Dit betekent dat ondernemers uit de sector handel & distributie meer waarde hechten aan deze twee variabelen bij het maken van hun locatiekeuzes dan ondernemers uit andere sectoren. De invloed van deze variabelen bleek immers niet uit de resultaten van figuur 7.7, waar de locatiekeuzes van alle bedrijfssectoren zijn onderzocht.

In figuur 7.11 staan de resultaten van het onderzoek voor de sector handel & distributie.

Uit de resultaten blijkt dat bijna 75 procent van de data binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Het belang van de hoofdvariabelen wordt ook bevestigd door de resultaten van het bedrijfslocatiemodel voor de sector diensten. De vestigingsplaats variabelen hebben echter geen invloed op de resultaten van het model. Dit is niet in overeenstemming met de theorie van Meester (1999), al bleek al eerder uit resultaten dat de invloed van deze variabelen minder groot is dan op basis van de conclusies van Meester mag worden aangenomen. Blijkbaar

hechten ondernemers uit de diensten sector meer waarde aan andere locatiekeuze factoren en laten ze zich niet leiden door de vestigingsplaats. Wellicht komt dit omdat ondernemers uit deze sector minder afhankelijk zijn van transportkosten, welke erg afhankelijk zijn van de vestigingsplaats. De transportkosten maken namelijk vaak maar een klein deel uit van de totale kosten van een geleverde dienst. Zo zal het voor bijvoorbeeld callcenters weinig uitmaken of zij een opdracht uitvoeren voor een klant die in de buurt zit of aan de andere kant van het land.



Figuur 7.11: Kwaliteit voorspellingen sector diensten

De input van de externe modellen, het huishoudmodel en het bereikbaarheidsmodel zijn van invloed op de kwaliteit van de resultaten. Hierbij kunnen opnieuw de opmerkingen geplaatst worden die voor de sector handel & distributie gedaan zijn. In het kort betekent dit dat de theorie achter LUTI modellen en specifiek de interactie tussen bedrijven uit de diensten sector en huishoudens en/of transport bevestigd wordt.

De theorie van Reijs (2001) en Everaar en Brand-van Tuijn (2001) over specifieke variabelen waaraan bedrijven uit de sector diensten extra belang hechten bij het maken van locatiekeuzes, wordt slechts beperkt bevestigd. Alleen het weglaten van de variabele “quality of life factoren” uit de variabelenset zorgt voor mindere resultaten van het model en heeft daarom invloed op locatiekeuzes van ondernemers in het model. Voor de variabele “representatie omgeving” is dit niet het geval. Dit betekent dat deze variabele de theorie van Reijs (2001) en Everaar en Brand-van Tuijn (2001) niet bevestigd. De resultaten voor de variabele “OV-bereikbaarheid” zijn opvallend. Deze variabele heeft een negatieve invloed op de resultaten van het model. Eerder bleek bij de voorspelling van het totale aantal arbeidsplaatsen nog dat deze variabele een gunstige invloed heeft op de resultaten. De vraag is waarom dit voor de sector diensten, die volgens de theorie relatief veel waarde hecht aan bereikbaarheid met het openbaar vervoer, niet het geval is. Voor het beantwoorden van deze vraag is aanvullend onderzoek nodig. Zo zouden bijvoorbeeld de voorspellingen van het bedrijfslocatiemodel in de stedelijke centra kunnen worden onderzocht, omdat hier veel dienstverlenende bedrijven gevestigd zijn die beweren veel waarde te hechten aan goede OV-verbindingen.

De resultaten van deze paragraaf laten zien dat het bedrijfslocatiemodel het aantal arbeidsplaatsen (en daarmee de bepaling van het werkgelegenheidsaandeel) van twee van de drie bedrijfssectoren goed kan voorspellen. Het model kan alleen geen goede voorspellingen doen voor de sector industrie & productie. Wellicht kan de sector industrie & productie indirect voorspeld worden uit de vergelijking:

- $\text{Arbplts sector industrie \& productie} = \text{Totaal arbplts} - \text{Arbplts sector handel \& distributie} - \text{Arbplts sector diensten}.$

Uit de resultaten is gebleken dat alle parameters achter het “=” teken goed door het model voorspeld kunnen worden en hiermee is het wellicht mogelijk alsnog het aantal arbeidsplaatsen in de sector industrie & productie te bepalen. Deze hypothese is onderzocht aan de hand van de resultaten van de voorspellingen van het model met de dataset van heel Nederland. De resultaten laten zien dat minder dan de helft van de berekende uitkomsten binnen de range van 0.8-1.2 van de werkelijke uitkomsten vallen. Dit betekent dat deze hypothese verworpen dient te worden. Doordat het model alleen categorieën voorspelt en hierbij tevens een foutmarge wordt gehanteerd, heeft dit tot gevolg dat er significante afwijkingen van de werkelijkheid ontstaan bij het terugrekenen naar werkelijke aantallen arbeidsplaatsen. Deze afwijking geldt voor alle drie parameters en zorgt ervoor dat het niet goed mogelijk is op deze manier indirect voorspellingen te doen voor de sector industrie & productie.

7.5 Resultaten toetsing methodiek model op gebiedsniveau

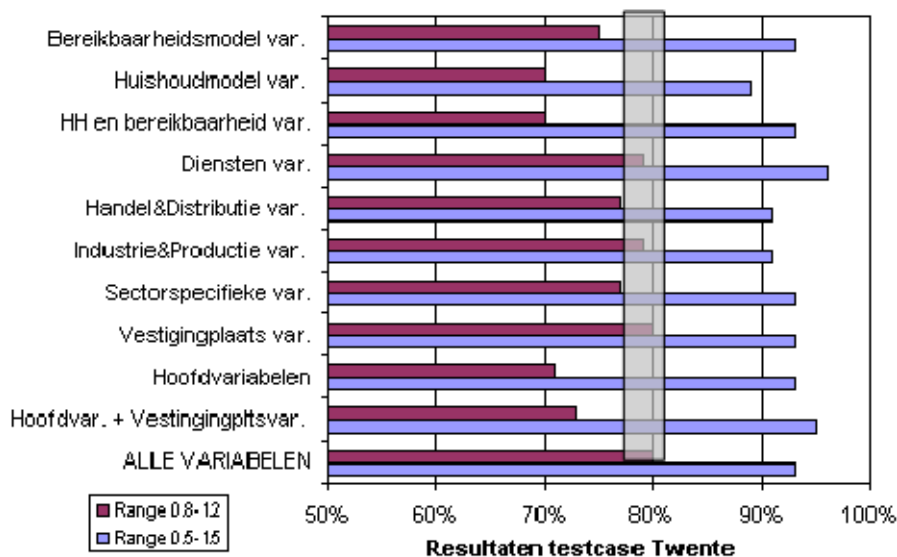
In dit deel van het onderzoek wordt de methodiek van het bedrijfslocatiemodel op gebiedsniveau getest met behulp van de in paragraaf 7.3 gedefinieerde variabelenset. Hiervoor wordt het model opnieuw gekalibreerd met de dataset van heel Nederland, behalve uit de regio Twente. Dit keer wordt er met een neurale netwerk gebruikt met 5 neuronen in de “hidden layer”, waarmee zo is in paragraaf 7.2 gebleken, voor de testcase Twente de beste resultaten worden bereikt. Aan de hand van de cases uit de dataset van Twente wordt de kwaliteit van de voorspellingen getoetst door de afwijking van de voorspelde uitkomsten door het gekalibreerde model ten opzichte van de werkelijke uitkomsten te bepalen. De methodiek van het model op gebiedsniveau wordt zowel getoetst voor het voorspellen van het aantal arbeidsplaatsen als voor het voorspellen van het werkgelegenheidsaandeel per bedrijfssector.

7.5.1 Resultaten voorspelling aantal arbeidsplaatsen

In figuur 7.12 zijn de resultaten per categorie neergezet voor de voorspelling van het aantal arbeidsplaatsen in de regio Twente door het gekalibreerde model.

Uit de resultaten blijkt dat 80 procent van de data van de regio Twente binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Het weglaten van de hoofdvariabelen uit het model zorgt voor een duidelijke afname in de kwaliteit van de uitkomsten van het model. Ook in de regio Twente zijn deze variabelen van invloed op de locatiekeuzes van ondernemers. De vestigingsplaats variabelen hebben daarentegen geen invloed op de resultaten van het model. In Twente hechten ondernemers blijkbaar geen waarde aan de vestigingsplaats. De resultaten in Twente bevestigen de theorie van Meester (1999) over het belang van vestigingsvariabelen dus niet. Dit kan niet verklaard worden uit het aandeel van bedrijven uit de sector diensten. In de vorige paragraaf is namelijk gebleken dat ondernemers uit de sector diensten weinig waarde hechten aan de vestigingsplaats variabelen. Het aandeel van de sector diensten ligt in Twente echter onder het

landelijke gemiddelde. Een andere verklaring zou kunnen zijn dat de verbanden tussen vestigingsplaatsen en locatiekeuzes redelijk complex zijn en dat een eenvoudig neurale netwerk, waarmee hier gewerkt is, niet in staat is deze verbanden te doorgronden. Het doorgronden van de verbanden kan bemoeilijkt worden door de ruis in de data, die eerder al in paragraaf 7.2 is vastgesteld.



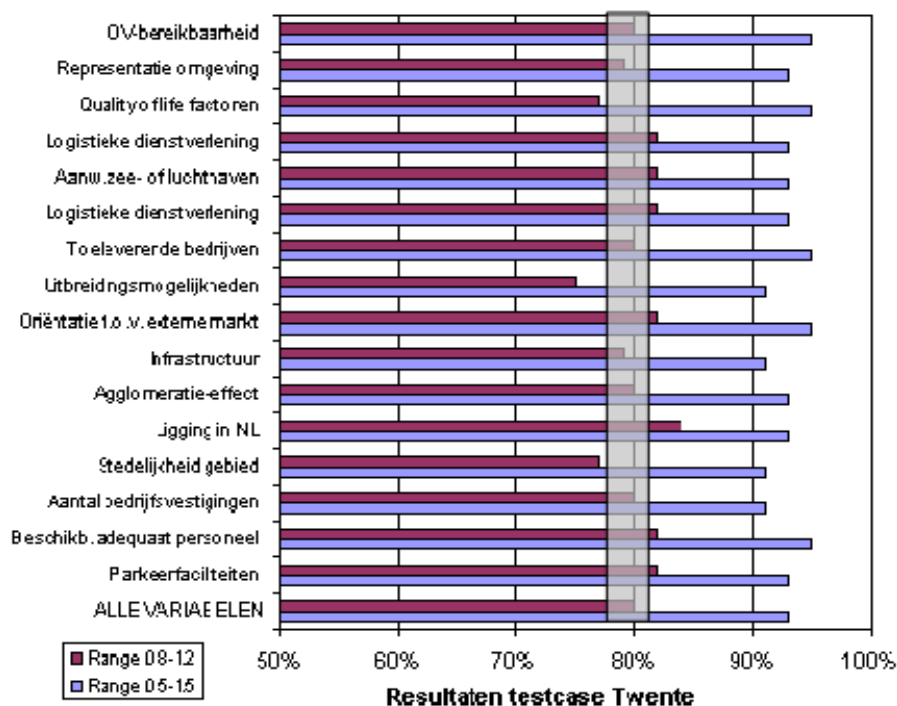
Figuur 7.12: Kwaliteit voorspellingen aantal arbeidsplaatsen in Twente (deel 1)

De variabelen die in de sectorspecifieke categorieën vallen, hebben geen duidelijke invloed op de kwaliteit van de uitkomsten van het model. Dit is in overeenstemming met de resultaten van de kalibratie van het model met de data uit heel Nederland in paragraaf 7.4. Ook de resultaten voor de variabelen van het huishoudmodel sluiten hier bij aan. De theorie achter LUTI modellen over de interactie tussen huishoudens en bedrijven wordt ook op gebiedsniveau bevestigd. De resultaten van het model voor de bereikbaarheidsvariabele laten zien dat deze van invloed is op de locatiekeuzes van ondernemers. In de resultaten uit paragraaf 7.4 kwam naar voren dat deze variabele op landelijk niveau geen invloed had op de locatiekeuzes van ondernemers. Blijkbaar waarderen ondernemers in Twente de variabele bereikbaarheid anders dan ondernemers uit de rest van het land. Bij de bespreking van de resultaten van figuur 7.13 wordt nader ingegaan op de mogelijke oorzaken hiervan.

In figuur 7.13 zijn de resultaten per variabele neergezet voor de voorspelling van het aantal arbeidsplaatsen in de regio Twente door het gekalibreerde model. Deze resultaten zijn een verdere uitwerking van de resultaten in figuur 7.12. Nu worden voor elke categorie de afzonderlijke variabelen uitgelicht.

Uit de resultaten blijkt dat van de afzonderlijke hoofdvariabelen alleen de variabele “stedelijkheid gebied” invloed heeft op de kwaliteit van de voorspellingen van het model. De variabelen “parkeerfaciliteiten” en “beschikbaarheid adequaat opgeleid personeel” hebben net als in de rest van Nederland geen invloed op de resultaten in de dataset van Twente. Dit betekent dat deze variabelen ook in Twente geen invloed hebben op de locatiekeuzes van bedrijven in het model. De resultaten van de afzonderlijke variabelen die de vestigingsplaats beschrijven bevestigen het

ontbreken van invloed van de vestigingsplaats op de locatiekeuzes van ondernemers in Twente, wat ook al in figuur 7.12 is vastgesteld.

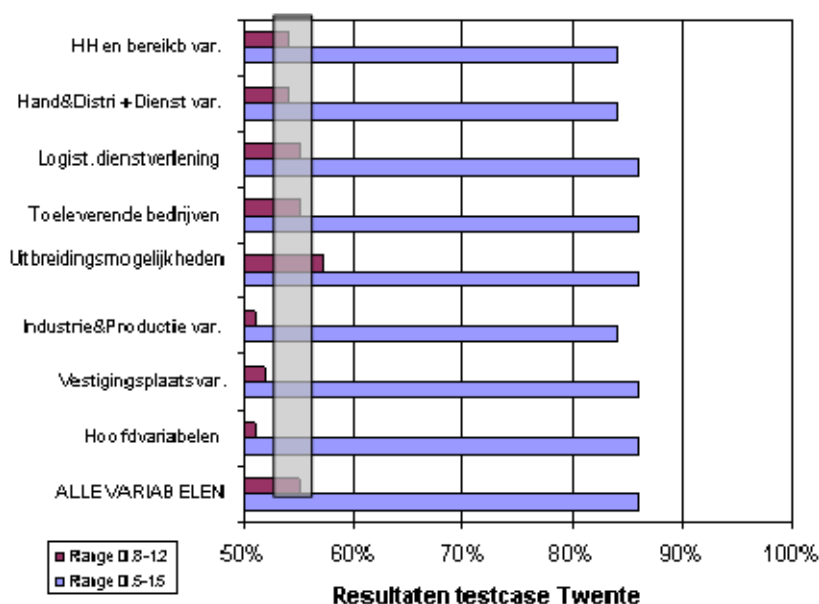


Figuur 7.13: Kwaliteit voorspellingen aantal arbeidsplaatsen in Twente (deel 2)

Het belang van de variabele “uitbreidingsmogelijkheden” wordt ook op gebiedsniveau bewezen. Net als in paragraaf 7.4 blijkt uit de resultaten dat deze variabele van invloed is op de resultaten van het bedrijfslocatiemodel. Daarnaast blijkt uit de resultaten dat de variabele “quality of life factoren” in Twente ook een bescheiden invloed heeft op de kwaliteit van de uitkomsten van het model. In de bespreking van de resultaten van figuur 7.16 wordt hier nog kort op ingegaan. Als laatste dienen de resultaten van de variabelen “OV-bereikbaarheid” en “infrastructuur” besproken te worden. Uit de resultaten blijkt dat deze variabele juist geen invloed hebben op de locatiekeuzes van ondernemers in Twente. Uit de resultaten in paragraaf 7.4 is gebleken deze variabelen in de rest van Nederland wel van invloed zijn op de locatiekeuzes van ondernemers. In figuur 7.13 is al gebleken dat de bereikbaarheidsvariabele in de regio Twente wel van invloed is op de resultaten in het model. Hieruit blijkt dat ondernemers in Twente meer belang hechten aan de auto-bereikbaarheid (uitkomst bereikbaarheidsmodel) en minder belang aan andere vormen van bereikbaarheid (bijvoorbeeld openbaar vervoer) dan de rest van Nederland. In Twente is het aandeel van bedrijven uit de sector industrie & productie relatief hoog en het aandeel van de dienstensector daarentegen relatief klein. Uit het onderzoek van Hanemaayer (1998) blijkt dat de sector industrie & productie de bereikbaarheid over de weg hoger waardeert in vergelijking met de sector diensten en de bereikbaarheid met het openbaar vervoer lager. Dit vormt een verklaring voor de verschillen tussen Twente en de rest van Nederland in de waarde die aan de verschillende vormen van bereikbaarheid wordt gegeven.

7.5.2 Resultaten voorspelling werkgelegenheidsaandeel per bedrijfssector

In het tweede onderdeel van het model worden de werkgelegenheidsaandelen per bedrijfssector in de regio Twente voorspeld. Hiervoor dient het model, dat een neurale netwerk gebruikt met 5 neuronen in de “hidden layer”, opnieuw gekalibreerd te worden met de dataset van heel Nederland, behalve uit de regio Twente. Met behulp van de cases uit de dataset van Twente wordt de kwaliteit van de voorspellingen van het model getoetst. Zoals eerder al in hoofdstuk 6 aangegeven, zijn nu slechts de categorieën variabelen en de sectorspecifieke variabelen onderzocht. De resultaten zijn in drie figuren uitgezet. In figuur 7.14 staan de resultaten van het onderzoek voor de sector productie & industrie.

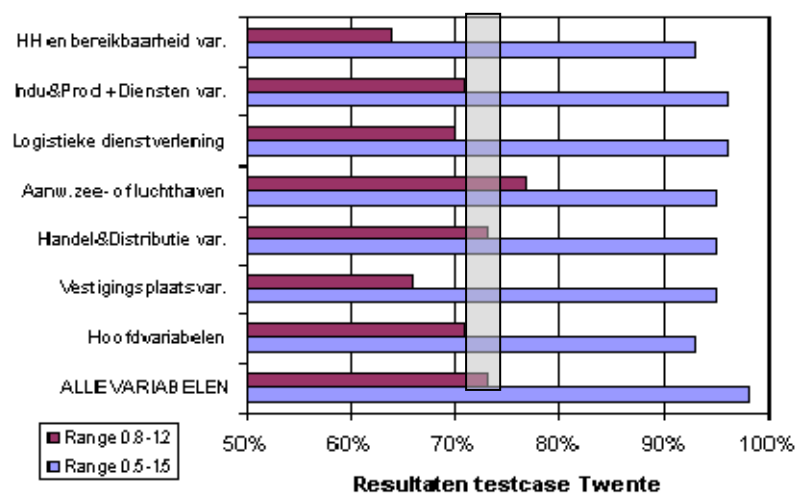


Figuur 7.14: Kwaliteit voorspellingen sector industrie & productie in Twente

Uit de resultaten blijkt dat slechts 55 procent van de data binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Dit percentage ligt net als in paragraaf 7.4 ver onder de grens van 70 procent, waarbij mag worden gesproken van een goed modelresultaat. Dit betekent dat het model ook op gebiedsniveau niet in staat is goede voorspellingen te doen voor de sector industrie & productie.

In figuur 7.15 staan de resultaten van het onderzoek voor de sector handel & distributie.

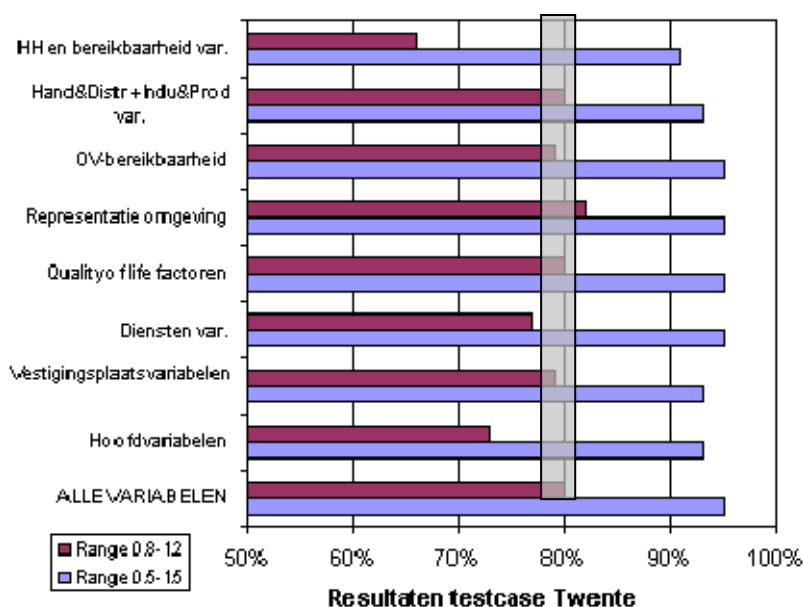
Uit de resultaten blijkt dat ruim 70 procent van de data in de testcase Twente binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Figuur 7.15 laat een beeld zien dat overeenkomstig is met de resultaten voor de sector handel & distributie in paragraaf 7.4, waar het model geïntegreerd is met de data van heel Nederland. Dit betekent dat het model ook geschikt is om voorspellingen te doen voor de sector handel & distributie op gebiedsniveau.



Figuur 7.15: Kwaliteit voorspellingen sector handel & distributie in Twente

Er is echter één resultaat dat opvallend is te noemen. De variabele “aanwezigheid zee- of luchthaven” heeft namelijk een negatieve invloed op de resultaten. Als deze variabele weg wordt gelaten uit de dataset, blijkt de kwaliteit van de voorspellingen van het model beter te worden. Een verklaring hiervoor zou de aanwezigheid van een (klein) vliegveld in de regio Twente kunnen zijn. Dit ziet het neurale netwerk als een gunstige factor voor het maken van een locatiekeuze. Wellicht hechten Twentse ondernemers echter nauwelijks of geen waarde aan dit vliegveld. Hierdoor vertroebelt deze variabele de factoren waarop ondernemers in Twente in werkelijkheid hun locatiekeuzes baseren. Dit zorgt ervoor dat het weglaten van deze variabele uit de variabelenset tot gevolg heeft dat de resultaten van het model beter worden.

In figuur 7.16 staan de resultaten van het onderzoek voor de sector handel & distributie.



Figuur 7.16: Kwaliteit voorspellingen sector diensten in Twente

Uit de resultaten blijkt dat 80 procent van de data in de testcase Twente binnen een range van 0.8-1.2 goed voorspeld wordt door het model. Figuur 7.15 laat een beeld zien dat overeenkomstig is met de resultaten voor de sector diensten in de rest van Nederland uit paragraaf 7.4. Dit betekent dat het model ook geschikt is om voorspellingen te doen voor de sector diensten op gebiedsniveau. Opnieuw valt het op dat de vestigingplaats variabelen en de variabele “OV-bereikbaarheid” geen invloed hebben op de voorspellingen van het model voor de sector diensten.

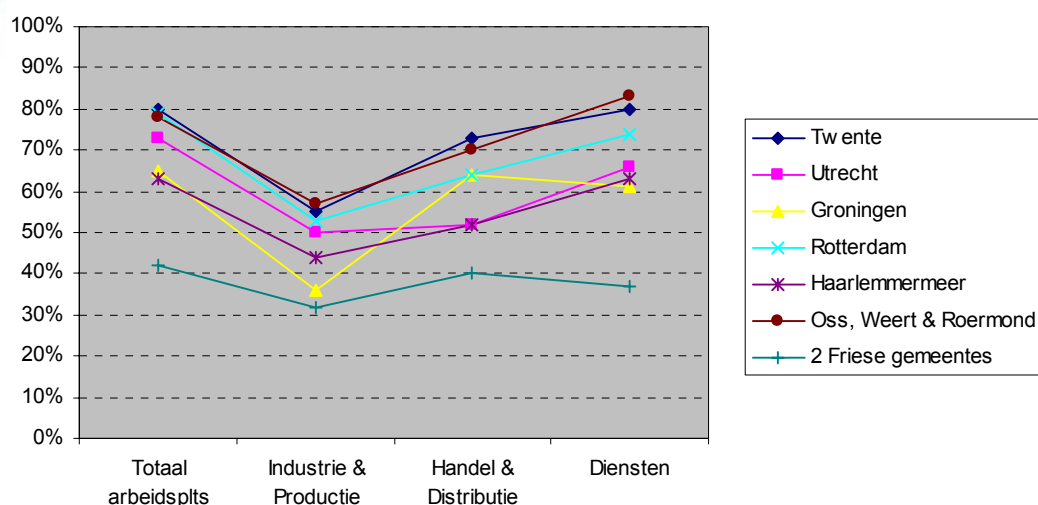
Uitzondering is de variabele “quality of life factoren”, die in Twente niet bij lijkt te dragen aan de kwaliteit van de resultaten. In paragraaf 7.4 lieten de resultaten nog de invloed zien van deze variabele op de sector diensten in Nederland. Daarnaast is eerder in deze paragraaf op basis van figuur 7.13 een bescheiden invloed geconstateerd van deze variabele op de voorspelling van het totale aantal arbeidsplaatsen in de regio Twente. Deze schijnbare tegenstrijdigheid is niet eenvoudig te verklaren uit deze resultaten. De meest aannemelijke verklaring hiervoor is de ruis in de dataset van Twente, die eerder al in paragraaf 7.2 is vastgesteld. Hierdoor bestaat de mogelijkheid dat de verbanden tussen vestigingsplaatsen en locatiekeuzes van ondernemers uit de sector diensten vertroebeld worden, waardoor het neurale netwerk niet in staat is deze verbanden te doorgronden.

7.6 Resultaten toetsing toepasbaarheid model op alle locaties in NL

De methodiek van het bedrijfslocatiemodel is in de voorgaande paragraaf uitgebreid getoetst op de regio Twente. Het blijkt dat het model in staat is goed het aantal arbeidsplaatsen te voorspellen en in staat is het werkgelegenheidsaandeel van de sector handel & distributie en de sector diensten goed te voorspellen. Wel blijkt dat de invloed van enkele afzonderlijke variabelen in deze regio af te wijken van het landelijke gemiddelde. In deze paragraaf staan de resultaten van de verificatie van het model met zes andere geselecteerde regio's. Op deze manier kan bepaald worden of het model toepasbaar is op elke willekeurige locatie in Nederland. De uitgangspunten van dit onderzoek zijn beschreven in paragraaf 6.4. In de eerste stap wordt het model gekalibreerd met alle cases, uitgezonderd het gebied dat voor de verificatie dient. Dit is voor elke geselecteerde regio gedaan en zoals verwacht blijkt dit dezelfde resultaten op te leveren die eerder al bij de kalibratie van de modelonderdelen in de vorige paragrafen naar voren zijn gekomen. Immers, de totale dataset verandert nauwelijks als er één enkele regio uit wordt gehaald. De resultaten staan in bijlage H. Hier zijn ook de resultaten van de verificatie van de regio's neergezet. Om deze resultaten gaat het in deze paragraaf in feite. Figuur 7.7 geeft een overzicht van de deze resultaten, waarbij gekeken wordt naar de resultaten binnen de range van 0.8-1.2.

De resultaten voor de zes regio's bevestigen het beeld dat het bedrijfslocatiemodel ook op gebiedsniveau het aantal arbeidsplaatsen goed voorspellen kan, moeite heeft met voorspellingen voor de sector industrie & productie en redelijk in staat is voorspellingen te doen voor de sectoren handel & distributie en diensten.

De resultaten van de afzonderlijke regio's zijn echter niet eenduidig. Het model kan naast de regio Twente goed voorspellingen doen voor de regio's Utrecht, Rotterdam en de regio Oss, Weert & Roermond. Het bedrijfslocatiemodel heeft moeite met de regio's Groningen en Haarlemmermeer. De kwaliteit van de resultaten voor de twee gemeentes in Zuidoost Friesland zijn niet goed. Het model kan geen goede voorspellingen doen voor deze regio.



Figuur 7.7: Resultaten bepaling toepasbaarheid model in zes regio's

Op basis van deze resultaten in figuur 7.7 kan niet worden bevestigd dat het model toepasbaar is op alle locaties in Nederland. Daarvoor zijn de resultaten voor een paar regio's niet goed genoeg, als wordt vastgehouden aan de norm van het nauwkeurig voorspellen van 70 procent van de data om de modelresultaten als goed te kwalificeren.

Toch dienen de resultaten nader te worden uitgelegd. Ten eerste wordt er voor elke regio met een klein aantal cases gewerkt (tussen de 30 en 50 cases). Dit betekent dat één of twee unieke kenmerken in een regio al kunnen zorgen voor een aantal uitschieters, welke veel invloed hebben op de resultaten. Daarnaast is het interessant om te kijken naar een specifieke regio en vast te stellen wat de oorzaak is van de slechte resultaten van het model voor de betreffende regio. Hiervoor wordt de regio Zuidoost Friesland gekozen, waarvoor de modeluitkomsten slecht zijn. Na de analyse van de data van deze cases worden de volgende opvallende zaken geconstateerd:

- De regio bevat 38 cases, waarvan 12 cases met een aantal arbeidsplaatsen in de categorie 10 of lager. Dit is bijna 30 procent van totaal.
- De regio bevat 38 cases, waarvan 20 cases met een aantal arbeidsplaatsen in de categorie 20 of lager. Dit is ruim 60 procent van totaal.
- Het aandeel van de sector industrie & productie is ruim 40 procent, tegenover een landelijk gemiddelde van 24 procent.

Dit betekent ten eerste dat er relatief weinig arbeidplaatsen aanwezig zijn in de regio Zuidoost Friesland, wat tot uitdrukking komt in het hoge percentage van lage categorieën (er zijn in totaal 100 categorieën waarin het aantal arbeidsplaatsen zijn uitgedrukt) in deze regio. In deze regio zijn dus veel lage outputwaarden. In hoofdstuk 6 is al betoogd dat kleine outputwaarden eerder buiten een range vallen dan grote waarden en dus eerder relatief grote afwijkingen vertonen van de gewenste output. Daarnaast is al gebleken bij de sectoranalyse dat het model niet goed in staat is voorspellingen te doen voor de sector industrie & distributie. Het relatief grote aandeel van deze sector in deze regio zorgt ervoor dat dit probleem zwaarder meeweegt in de bepaling van het totale aantal arbeidsplaatsen en dus een sterkere negatieve invloed heeft op de resultaten.

Het gaat te ver om voor elke regio specifiek de oorzaken te vinden voor kleine afwijkingen in kwaliteit van de resultaten, maar het ligt in de lijn der verwachting dat er overeenkomstige oorzaken gevonden kunnen worden, zij het minder duidelijk dan in het geval bij de regio Zuidoost Friesland. Zo ligt de regio Groningen ook in het Noorden, waarvan eerder al geconcludeerd werd, dat de kans groter is dat het model hiermee moeite heeft. Als deze bevindingen in overweging worden genomen, kan geconcludeerd worden dat het model wel in staat is toegepast te worden op een willekeurige locatie in Nederland, mits er rekening wordt gehouden met een paar factoren. Twee van die factoren zijn al gevonden. Dit zijn:

- De beperkte voorspellingskracht van het model voor de sector industrie & productie
- Het nadelige effect van de meetmethode in gebieden met lage aantallen arbeidsplaatsen

Doordat er slechts beperkt onderzoek gedaan is naar oorzaken voor mindere voorspellingen van het model in bepaalde regio's, bestaat de mogelijkheid dat er meerdere factoren zijn waarmee rekening moet worden gehouden. Het gaat te ver om deze oorzaken binnen dit onderzoek verder uit te werken.



8 Conclusies en aanbevelingen

8.1 Conclusie

Dit onderzoek is een voortzetting geweest van het promotiewerk van Tillema. Hierin werd ondermeer aangetoond dat het mogelijk is om met behulp van een neurale netwerk een LUTI model te bouwen. De doelstelling van dit onderzoek is om een onderdeel van dit model, het bedrijfslocatiemodel, te herdefiniëren op basis van het ontwerp van een variabelenset die de locatiekeuzes van bedrijven voorspelt. Daarbij dient tevens onderzocht te worden of dit nieuwe bedrijfslocatiemodel op elke willekeurige locatie in Nederland toepasbaar is. Als uitgangspunt is genomen dat het nieuwe model voorspellingen zou moeten kunnen doen over twee zaken; het aantal arbeidsplaatsen in een gebied en het werkgelegenheidsaandeel van bedrijfssectoren in een gebied.

8.1.1 Ontwerp variabelenset voor locatiekeuzes bedrijven

Er is uitgebreide literatuur voor handen over de locatiekeuzes van bedrijven. Op basis van de belangrijkste bevindingen uit deze literatuur is een nieuwe variabelenset gedefinieerd, waarmee de factoren die locatiekeuzes van bedrijven uitdrukken beter worden beschreven dan in de oude variabelenset van Tillema. Het lastigste onderdeel bij het definiëren van deze nieuwe variabelenset blijkt het kwantificeren van de factoren die de locatiekeuzes van ondernemers beschrijven. Dit komt deels doordat sommige van deze factoren lastig uit te drukken zijn in meetbare variabelen, deels doordat de acquisitie van benodigde data niet altijd mogelijk was in dit onderzoek. Dit bevestigt één van de al eerder genoemde nadelen van het werken met een neurale netwerk, de noodzaak van grote hoeveelheden geschikte invoerdata. De vier hoofdvariabelen die opgenomen zijn in de variabelenset zijn:

- Parkeerfaciliteiten
- Aanwezigheid adequaat opgeleid personeel
- Aantal bedrijfsvestigingen
- Stedelijkheid gebied

Ten tweede zijn er een viertal variabelen geselecteerd die specifiek de vestigingsplaats beschrijven. Onderzoek heeft namelijk aangetoond dat de vestigingsplaats een duidelijke invloed heeft op de locatiekeuzes van ondernemers. Daarnaast blijkt er een aantoonbaar verschil te bestaan tussen de locatiekeuzes van ondernemers uit drie verschillende bedrijfssectoren, te weten de sector industrie & productie, de sector handel & distributie en de sector diensten. Om deze reden is voor elk van deze sectoren een aantal variabelen gedefinieerd die volgens de literatuur speciaal van belang zijn voor de locatiekeuzes van ondernemers uit de betreffende sector. Een compleet overzicht van deze variabelenset staat in tabel 7.1.

8.1.2 Kalibratie bedrijfslocatiemodel

Voorspelling aantal arbeidsplaatsen

Aan de hand de nieuwe variabelenset is het bedrijfslocatiemodel gekalibreerd voor voorspellingen over bedrijvigheid. Het gekalibreerde bedrijfslocatiemodel blijkt goed in staat te zijn voorspellingen te doen over het aantal arbeidsplaatsen in een gebied. Het model voorspelt 80 procent van de data binnen de range van 0.8-1.2 van de werkelijke data. Vooral gebieden met lage aantallen arbeidsplaatsen blijken vaak buiten de range te vallen. Opvallend is dat de twee hoofdvariabelen

“parkeerfaciliteiten” en “aanwezigheid van adequaat opgeleid personeel” geen invloed hebben op de modelresultaten. Dit betekent dat uit dit onderzoek blijkt dat deze variabelen niet, zoals in de literatuur wordt verondersteld, een significantie invloed hebben op locatiekeuzes van ondernemers. De invloed van de variabele “uitbreidingsmogelijkheden” is daarentegen groter in het bedrijfslocatiemodel dan op basis van de literatuur verwacht zou mogen worden. Het lijkt erop dat de genoemde locatiefactoren door ondernemers in “stated preference” onderzoeken niet altijd overeenstemmen met de werkelijke locatiefactoren waarop zij hun keuzes baseren.

Daarnaast blijkt de input van het bereikbaarheidsmodel de resultaten van het bedrijfslocatiemodel niet te beïnvloeden. Hiermee wordt het vermoeden van Tillema bevestigd, die betwijfelt of de gekozen definitie van bereikbaarheid in zijn model ook geschikt is voor het voorspellen van locatiekeuzes van ondernemers en niet slechts voor huishoudens. Dat ondernemers hun locatievoorkeur laten beïnvloeden door een vorm van bereikbaarheid lijkt echter evident. Dit komt ook tot uitdrukking in de resultaten van dit onderzoek. De variabelen “aanwezigheid infrastructuur” en “OV-bereikbaarheid” hebben namelijk wel invloed op de resultaten. Dit betekent dat de bereikbaarheid wel degelijk invloed uitoefent op de locatiekeuzes van ondernemers.

Als laatste blijkt dat de invloed van de vestigingsplaats op locatiekeuzes van ondernemers minder groot is, dan op basis van de theorie van Meester (1999) zou mogen worden verwacht. Dit betekent niet dat de invloed van de vestigingsplaats genegeerd mag worden. Het onderzoek toont namelijk aan dat twee van de vier variabelen die de vestigingsplaats beschrijven wel degelijk (een bescheiden) invloed hebben op de locatiekeuzes van ondernemers. Dit zijn de variabelen “agglomeratie-effect” en “aanwezigheid infrastructuur”.

Voorspelling werkgelegenheidsaandeel per bedrijfssector

Het gekalibreerde bedrijfslocatiemodel blijkt eveneens goed in staat te zijn voorspellingen te doen over het werkgelegenheidsaandeel van de sector handel & distributie en de sector diensten. Het model voorspelt in beide gevallen ruim 70 procent van de data binnen de range van 0.8-1.2 van de werkelijke data. Voor de sector industrie & productie kan het model geen goede voorspellingen doen. Dit betekent dat de locatiekeuzes van ondernemers uit deze sector afwijken van die van ondernemers uit andere sectoren. Ook zijn de variabelen die volgens de literatuur specifiek de locatiekeuzes van deze sector beschrijven, niet goed genoeg om voorspellingen mee te doen. Dit zijn de variabelen “uitbreidingsmogelijkheden”, “aanwezigheid toeleverende bedrijven” en “logistieke dienstverlening”.

De resultaten van het onderzoek bevestigen het in de literatuur genoemde belang van de variabelen “aanwezigheid zee- of luchthaven” en “logistieke dienstverlening” voor de sector handel & distributie. Hetzelfde geldt voor de sector diensten betreffende de variabele “quality of life factoren”. Opvallend is dat de vestigingsplaats variabelen geen invloed hebben op de voorspellingen van het model voor de sector diensten. Dit kan verklaard worden doordat ondernemers in deze sector minder afhankelijk zijn van transportkosten, die op hun beurt sterk afhankelijk zijn van de vestigingsplaats. Hierdoor is de sector diensten minder afhankelijk van de vestigingsplaats.

8.1.3 Toetsing methodiek bedrijfslocatiemodel op gebiedsniveau

Het model blijkt in staat voorspellingen te doen op gebiedsniveau. Dit volgt uit het feit dat 80 procent van de data van de testcase Twente door het model binnen een range van 0.8-1.2 goed

wordt voorspeld. Daarnaast komen de resultaten van de bedrijfssectoren en de afzonderlijke variabelen in grote lijnen overeen met de resultaten van de kalibratie van het model die in de vorige paragraaf zijn besproken. Dit toont aan dat het model ook gebruikt kan worden om voorspellingen te doen in een bepaald gebied. Hiermee is voldaan aan een belangrijke voorwaarde van het model.

De testcase Twente laat echter wel enkele verschillen zien met het landelijke gemiddelde. Het model wordt bij dit onderdeel namelijk geconfigureerd met een veel eenvoudiger neurale netwerk dan bij het vorige onderdeel. De reden hiervoor is, dat de data van Twente veel ruis blijkt te bevatten. Dit zorgt ervoor dat een eenvoudig netwerk, die in staat is deze ruis te negeren, de beste resultaten oplevert. Een voorbeeld van deze ruis is de negatieve invloed van de variabele “aanwezigheid zee- of luchthaven” op de resultaten voor de sector handel & distributie in Twente. Landelijk gezien is gebleken dat deze variabele wel van belang is voor ondernemers uit deze sector. De verklaring hiervoor is de aanwezigheid van een klein vliegveld in Twente. Het model ziet dit als een gunstige factor, terwijl Twentse ondernemers hier nauwelijks waarde aan hechten. Daarnaast blijken ondernemers in Twente relatief meer belang te hechten aan de bereikbaarheid met de auto en minder belang aan andere vormen van bereikbaarheid, zoals het openbaar vervoer.

Er vallen dus een aantal gebiedsspecifieke karakteristieken in Twente aan te wijzen, die voor ruis in de data zorgen. Het valt te verwachten dat soortgelijke afwijkingen van het landelijke gemiddelde ook in andere gebieden voorkomen. De resultaten tonen aan dat het bedrijfslocatiemodel ondanks afwijkingen in de data, toch in staat is goede voorspellingen te doen op gebiedsniveau.

8.1.4 Toetsing toepasbaarheid model op alle locaties in NL

Uit de resultaten blijkt dat het model, naast de regio Twente, ook voor andere regio's met verschillende kenmerken, in staat is goede voorspellingen te doen over het aantal arbeidsplaatsen en de werkgelegenheidsaandeel van de bedrijfssectoren. Net zoals bij de kalibratie van het model, is dit alleen niet het geval voor de sector industrie & productie. Daarnaast blijkt het model ook niet voor alle regio's goede voorspellingen te kunnen doen. Hiervoor zijn twee oorzaken aan te wijzen:

- De beperkte voorspellingskracht van het model voor de sector industrie & productie
- Het nadelige effect van de meetmethode in gebieden met lage aantallen arbeidsplaatsen

Dit betekent dat als rekening wordt gehouden met deze twee factoren, geconcludeerd kan worden dat het model toepasbaar is op elke willekeurige locatie in Nederland.

8.2 Reflectie

Met het formuleren van de conclusies blijven er twee belangrijke vragen open die beantwoord dienen te worden. Ten eerste moet worden besproken in hoeverre het nieuwe bedrijfslocatiemodel bijdraagt aan een verbetering van het model van Tillema. Daarna zal kort worden ingegaan op de betekenis van dit onderzoek binnen de theorie over de interactie tussen landgebruik en transport.

Zoals eerder al is vastgesteld, hangt de kwaliteit van een datagedreven model af van de kwaliteit van de input die het krijgt. Dit onderzoek heeft een nieuwe, veel uitgebreidere variabelenset opgeleverd, waarvan in de literatuur is aangetoond dat deze de locatiekeuzes van bedrijven beschrijven. Daarnaast heeft dit onderzoek aangetoond dat met behulp van deze dataset de interactie tussen het landgebruik van bedrijven en transport voorspeld kan worden. Hiermee zijn de basiselementen van de interactie tussen bedrijvigheid en transport nader gedefinieerd en getest op

hun prestaties. Dit betekent dat twee onderdelen die Tillema in zijn onderzoek noemt voor het ontwikkelen van een goed datagedreven LUTI model verder uitgewerkt zijn. Daarnaast is het onderdeel toepasbaarheid van het model, door middel van de toetsing van het model in zes verschillende regio's, verder uitgewerkt.

Ten tweede is de conceptuele modelstructuur van het bedrijfslocatiemodel aangepast. Dit is ook een van de onderdelen die Tillema benoemt voor het ontwikkelen van een goed datagedreven model. In dit onderzoek is niet aangetoond of de nieuwe outputvariabelen een betere input vormen voor de andere twee modellen van Tillema. Uit de theorie over LUTI blijkt echter dat de interactie tussen huishoudens en bedrijven plaatsvindt op ondermeer de arbeidsmarkt. De nieuwe output van het model, het aantal arbeidsplaatsen in een gebied, beschrijft de toestand op de arbeidsmarkt beter dan de oude output, het aantal bedrijven in een gebied. Daarnaast is ook duidelijk dat deze nieuwe output de vraag op de transportmarkt ook beter representeert. Het hoeveelheid woon-werk verkeer, die het grootste deel van de vraag naar transport veroorzaakt, valt immers direct af te leiden uit het aantal arbeidsplaatsen in een gebied. Dit betekent dat de verbanden uit de LUTI theorie beter beschreven worden met de nieuwe modelstructuur. Daarmee mag worden aangenomen dat de nieuwe output een betere input vormt voor het huishoudmodel en het bereikbaarheidsmodel.

Een derde punt dat als verbetering kan worden genoemd, is dat met behulp van het nieuwe bedrijfslocatiemodel aangetoond is dat de definitie van bereikbaarheid uit het bereikbaarheidsmodel niet geschikt is voor het voorspellen van locatiekeuzes van ondernemers. Daarnaast is van een tweetal andere vormen van bereikbaarheid bewezen dat zij wel geschikt zijn voor deze taak. Deze bevindingen zijn een prima basis om te gebruiken in verder onderzoek voor het verbeteren van de kwaliteit van het bereikbaarheidsmodel.

Hiermee zijn de belangrijkste punten genoemd waarop het bedrijfslocatiemodel verbeterd is. Dit betekent niet dat het nieuwe model helder inzicht verschaft in de interactie tussen landgebruik en transport. Hiervoor zijn de relaties in de black box neurale netwerken te onduidelijk. Wel worden er in dit onderzoek een aantal theoretische verbanden bevestigd. Zo wordt de invloed van huishoudens op de activiteiten van bedrijvigheid meerdere malen in dit onderzoek bevestigd en wordt ook bevestigd dat ondernemers zich door andere vormen van bereikbaarheid laten beïnvloeden dan huishoudens. Dit komt overeen met eerder bevindingen in LUTI onderzoek. Daarnaast geeft dit onderzoek een onderbouwing van bedrijfslocatietheorieën betreffende een aantal factoren die locatiekeuzes van ondernemers beïnvloeden. Doordat achterliggende verbanden door het black box karakter van het neurale netwerk verborgen blijven is het in dit onderzoek niet mogelijk om een optimale variabelenset voor locatiekeuzes van ondernemers aan te wijzen en van een aantal factoren de invloed op locatiekeuzes van ondernemers uit te sluiten.

Als laatste dient genoemd te worden, wat ook door Tillema ter sprake wordt gebracht, dat het verbeterde bedrijfslocatiemodel het mogelijk maakt beter de gevolgen in te schatten van het stimuleren van economische activiteiten, het aanleggen van woongebieden of het verbeteren van de transportinfrastructuur, voor de activiteiten van bedrijven in een gebied.

8.3 Aanbevelingen

De belangrijkste eigenschap van het bedrijfslocatiemodel is wellicht wel dat dit datagedreven is. Dit betekent dat de kwaliteit van de voorspellingen volledig afhankelijk is van de kwaliteit van de data die het als input krijgt. Er zijn tijdens dit onderzoek vele concessies en aanpassingen gedaan aan

de datakwantiteit om deze geschikt te maken voor invoer in het neurale netwerk. Ook zijn om deze redenen een aantal aannames gedaan, waarvan de gevolgen voor de representatie van de werkelijkheid niet altijd duidelijk zijn. Daarnaast bleek zoals eerder al aangegeven de meest geschikte data soms wel aanwezig, maar niet beschikbaar voor dit onderzoek. Dit alles zorgt ervoor dat er veel winst valt te halen in het verbeteren van de dataset. Dit kan gerealiseerd worden door meer geld beschikbaar te stellen voor kostbare data, of nog beter te kijken of welke manier bepaalde factoren te kwantificeren zijn.

In dit onderzoek zijn de bedrijven onderverdeeld in sectoren. Er zijn echter ook andere indelingen te bedenken, die wellicht ook een goede aanvulling vormen op de input van het bedrijfslocatiemodel in de twee andere modellen van Tillema's LUTI model. Zo zou ook een indeling naar karakter van bedrijven kunnen worden gemaakt. Hierbij kan bijvoorbeeld gedacht worden aan een indeling in een type traditioneel hiërarchisch geleide, formele bedrijfssector en type bedrijfssector met een platte, informeel geleide structuur. Het lijkt aannemelijk dat beide types bedrijven verschillend ten opzichte van elkaar reageren op veranderingen in hun omgeving.

Het onderzoek naar de toepasbaarheid van het model op elke willekeurige locatie in Nederland laat nog een aantal vragen open. Hoewel bewezen is dat het model toepasbaar is op elke locatie in Nederland, zijn de factoren waarmee hier rekening moet worden gehouden slechts beperkt onderzocht. Hiervoor zou het model opnieuw in een aantal regio's geverifieerd moeten worden. Op basis hiervan zou getracht moeten worden de exacte factoren die de kwaliteit van de voorspellingen in bepaalde regio's beïnvloeden, te vinden. Hierdoor zouden de voorspellingen van het model op elke willekeurige locatie in Nederland met een veel grotere nauwkeurigheid geïnterpreteerd kunnen worden.

In hoofdstuk 2 van dit rapport is al beschreven dat de relaties in LUTI modellen dynamisch van aard zijn. De exacte tijdsspanne tussen bepaalde processen in deze modellen is echter niet bekend. Ook in dit onderzoek is gebruik gemaakt van statische data. Door het introduceren van dynamische data valt echter nog wel een duidelijke verbetering in de dataset te brengen. De vraag is echter welke perioden gebruikt zouden moeten worden. Wellicht kan met behulp van het neurale netwerk bepaald worden welke perioden de beste resultaten opleveren. Het verschil tussen de invoerperiode en het tijdstip waarop de uitvoerdata is gemeten geeft inzicht in de tijd die bepaalde processen in het LUTI model nodig hebben om van invloed te zijn.

Het laatste punt waar op dient te worden ingegaan is het neurale netwerk. Het belang van een configuratie voor de kwaliteit van de resultaten is duidelijk. De vraag blijft of het optimale neurale netwerk gevonden is, wat zeer waarschijnlijk niet het geval is. De resultaten varieerden namelijk behoorlijk in het onderzoek. Door meerdere runs voor hetzelfde onderdeel uit te voeren, werd getracht om te gaan met dit probleem. Wellicht zijn er betere manieren te bedenken om dit probleem op te lossen. Daarnaast valt het aan te bevelen om de verschillende types neurale netwerken en configuratie opties nader te bestuderen. Hiermee wordt meer inzicht gekregen in de werking van een specifiek neurale netwerk en daarmee kan ook meer inzicht worden gekregen in het black box karakter van het neurale netwerk.



Referenties

- AVV (2003). Adviesdienst Verkeer en Vervoer,
<http://www.rijkswaterstaat.nl/regiooverzicht/organisatie/adviesdienstvoorverkeerenvervoer/>
- Beale, R. and Jackson, T. (1990). *Neural Computing: An Introduction*, Department of Computer Science, University of York, Institute of Physics Publishing, Bristol and Philadelphia.
- Brits ministerie van Transport, (2003). *Land-Use / Transport Interaction Models*,
http://www.webtag.org.uk/webdocuments/3_Expert/1_Modelling/3.1.3.htm#3_5.
- CBS (2005). Centraal Bureau voor Statistiek, www.cbs.nl
- Computer Associates International, inc (1999). *An introduction to Neural Computing*, Islandia, New York.
- DeClaris, J.W. en Roberts, J. (1997). *An introduction to neural networks*,
<http://www.ee.umd.edu/medlab/neural/nn1.html>
- Dicken, P. (1986). *Global Shift, industrial change in a turbulent world*, Harper & Row Ltd., London.
- Everaars, J. en Brand-van Tuijn, J.J. (2001). *Locatieontwikkeling: Relatiemodel tussen ruimte, economie en verkeer en de invloed van actoren hierop in beeld gebracht*, proefschrift, Universiteit Twente.
- Geurs, K.T. and Ritsema van Eck, J.R. (2001). *Accessibility Measures: Review and applications, evaluation of accessibility impacts of land-use transport scenarios, and related social and economic impacts*, Rijksinstituut voor volksgezondheid en milieu, rapport 408505 006, Bilthoven.
- Gielen, S. en Kappen, B. (1995). *Neurale netwerken en hun toepassingen*. In: I&I nr 1.
- Hanemaayer, D. en Rekkers, P. (1998). *Vestigingsplaatsfactoren: belang, waardering en knelpunten*, B&A groep, Den Haag.
- Hayter, R. (2004). *The Dynamics of Industrial Location: The Factory, the Firm and the Production System*, Department of Geography, Simon Frasier University, Burnaby.
- Lee, D.B. (1973). *Requiem for Large-Scale Models*, *Journal of the American Institute of Planners*, Vol. 39, blz. 163-178.
- Maanen, van T. Verroen, E. en Heerema, P. (1992). *Mobiliteitsprofielen revisited, nadere analyse v/d samenhang tussen bedrijf-, lokatie- en mobiliteitskenmerken*, INRO-TNO, Delft.
- Martin, P. en C-A Rogers, "Industrial Location and Public Infrastructure," *Journal of International Economics*, 1995, 39(3-4), 335-51.
- Meester, W.J. (1999). *Subjectieve waardering van vestigingsplaatsen door ondernemers*, proefschrift, Rijksuniversiteit Groningen.

Ministerie van Economische zaken (2003). *Van Gennip zoekt dialoog met buitenlandse bedrijven*, persbericht, Den Haag.

Minsky, M. and Papert, S. (1969). *Perceptrons, An introduction to computational geometry*, MIT press, expanded edition.

Mooij, de R. et al, (2004). *In de slag om bedrijvigheid, theorie en praktijk van vestigingsplaatsconcurrentie*, preadvies, CPB.

Mulder, B. (2002). *Fonotactisch leren met neurale netwerken*, Ph.D. thesis, University of Groningen.

Reijs, T. et al. (2001). Core Centra en vestigingsplaatskeuze. *Hoe kunnen bedrijventerreinen op duurzame wijze ontwikkelingen in logistieke dienstverlening, ketenvorming en ICT faciliteren? Een 'state of the art' betreffende vestigingsplaatskeuze theorieën*, TNO Inro Afdeling R.O., Delft.

Rietveld, P. en Koonen, E. (2004). *De "discrete choice" theorie voor ruimtegebruikmodellering; de economische achtergrond van de Ruimtescanner*, LUMOS Symposium, RIVM.

Russell, S.J. en Norvig, P. (2003). *Artificial Intelligence, A modern approach*, 2^e editie, Prentice Hall.

Stergiou, C. en Siganos, D. (1996). *Neural Networks*, Department of Computing, Department of Electrical and Electronic Engineering, Imperial College London, SURPRISE 96, volume 4.

Stoianov, I. (2000). *Connectionist Lexical Processing*, Doctoral thesis, University of Groningen.

Teunissen, B. (1992). *Hoe gunstig zijn bedrijfsverplaatsingen passend in het locatiebeleid?; Een casestudy in Amsterdam*. In: Rooilijn, jrg.25, nr.4, p.116-120.

Tillema, F (2004). *Development of a data driven land use interaction model*, proefschrift, Universiteit Twente.

Timmermans, H. (2003). *The Saga of Integrated Land Use – Transport Modelling: How Many More Dreams Before We Wake Up?*, keynote paper, 10^{de} Internationale conferentie over reisgedrag onderzoek, Lucerne.

The MathWorks (2005). *Matlab*, versie 7

Veen, van der H.C.J. (2004). *Het effect van milieubeleid op locatiebeslissingen van bedrijven: verplaatsen, sluiten, of innoveren?*, proefschrift, Rijksuniversiteit Groningen.

Van de Vooren, F.W.C.J. en Pauwels, T. (2002). *MOBILEC: De wisselwerking tussen mobiliteit en economie gemodelleerd: beleidseffecten in Vlaanderen*, onderzoeksrapport, Universiteit Antwerpen.

Van Wee, G.P. (1993). *Locatiebeleid en Ruimtelijke Ordening: de effecten op verkeer en vervoer*, Rijksinstituut voor Volksgezondheid en Milieu, Bilthoven.

Bijlagen

BIJLAGE A: OVERZICHT LOCATIEKEUZEFACTOREN.....	79
BIJLAGE B: CONFIGURATIE NEURAAAL NETWERK.....	80
BIJLAGE C.1: SPECIFICATIES VARIABELEN VOORSPELLING ARBEIDSPLAATSEN	82
BIJLAGE C.2: SPECIFICATIES VARIABELEN VOORSPELLING WERKGELEGEN- HEIDSAANDEEL BEDRIJFSSECTOREN.....	85
BIJLAGE D: RESULTATEN BEPALING OPTIMALE CATEGORISERING OUTPUT.....	88
BIJLAGE E: BEPALING OPTIMALE AANTAL NEURONEN IN NEURAAAL NETWERK	89
BIJLAGE F.1: RESULTATEN VOORSPELLING AANTAL ARBEIDSPLAATSEN	91
BIJLAGE F.2: RESULTATEN VOORSPELLING WERKGELEGENHEIDSAANDEEL BEDRIJFSSECTOREN	92
BIJLAGE G: RESULTATEN TOETSEN MODEL IN ZES REGIO'S.....	94



Bijlage A: Overzicht locatiekeuzefactoren

In onderstaande tabel staan naar bedrijfstypen het belang dat aan de afzonderlijke aspecten van de bedrijfsomgeving wordt toegekend (score van 0 (onbelangrijk) tot 10 (zeer belangrijk)). Afkomstig uit het onderzoek van Hanemaayer (1998).

Aspect/cluster	industrie/productie	handel	verkeer/vervoer	dienstverlening	totaal
GEBOUW EN TERREIN	5,3	5,4	5,6	5,8	5,5
parkeermogelijkheden personeel/bezoekers	7,3	7,4	7,1	8,0	7,5
representativiteit gebouw	6,1	6,4	6,5	7,3	6,6
huurprijs- of grondprijsniveau	5,2	5,4	6,0	5,3	5,4
representativiteit omgeving	4,8	5,0	5,3	5,6	5,2
uitbreidingsmogelijkheden	5,1	4,6	5,0	4,1	4,7
omgevingskwaliteit	3,2	3,5	3,7	4,2	3,6
BEREIKBAARHEID	6,7	6,8	6,6	5,4	6,3
bereikbaarheid over weg	8,6	8,6	8,9	7,7	8,3
laad- en losmogelijkheden op terrein	7,3	7,6	7,0	3,0	5,9
bereikbaarheid via openbaar vervoer	4,3	4,3	4,1	5,5	4,7
INTERNATIONALE ORIËNTATIE	2,0	2,2	3,8	1,2	2,0
nabijheid internationale bedrijvigheid	2,0	2,0	4,3	1,9	2,3
nabijheid luchthaven	2,1	2,5	3,5	1,4	2,1
nabijheid douane-entrepot	2,0	1,9	3,5	0,4	1,6
OVERIGE VERVOERMOGELIJKHEDEN	1,6	1,2	2,7	0,5	1,3
combinatiemogelijkheden vervoer	2,0	1,6	2,9	0,9	1,7
nabijheid zeehaven	1,9	1,7	3,9	0,5	1,6
nabijheid vaarwater	1,3	1,1	1,6	0,4	1,0
goederenvervoermogelijkheden trein	1,0	0,6	2,2	0,4	0,9
TELECOMMUNICATIE	5,1	5,3	6,0	4,9	5,2
Telecommunicatievoorzieningen	6,0	6,4	6,9	6,1	6,2
nabijheid dienstverlening telecomm.	4,1	4,2	5,1	3,7	4,1
TOELEVERENDE BEDRIJVEN	3,3	2,9	3,4	1,6	2,6
gespecialiseerde bedrijven toelevering	4,1	2,9	4,3	2,2	3,3
nabijheid logistieke dienstverlening	4,1	4,1	3,8	1,9	3,3
nabijheid distributiecentrum	2,8	2,8	2,9	0,7	2,1
nabijheid zelfde bedrijvigheid	2,0	1,6	2,6	1,8	1,9
KENNISPOTENTIEEL	4,1	3,1	3,7	3,9	3,8
beschikbaarh. adequaat opgeleid personeel	7,2	6,2	6,9	7,4	7,0
nabijheid scholings- en onderwijsinstituten	3,1	1,8	2,7	2,2	2,5
nabijheid kenniscentra	2,1	1,4	1,6	2,0	1,9

Bijlage B: Configuratie neuraal netwerk

In deze bijlage wordt de standaard configuratie van het neurale netwerk beschreven, zoals die is gebruikt in het onderzoek. Het neurale netwerk is gemodelleerd in het programma Matlab, in het onderdeel "Neural network toolbox", dat eerder al in hoofdstuk 3 beschreven is. In figuur 9.1 staat de syntax beschreven, waarmee in Matlab het neurale netwerk is gedefinieerd waarmee in dit onderzoek gewerkt is.

<i>KALIBRATIE</i>	
<code>x=load('Data_NLzonderTwente.txt');</code>	<i>Invoer kalibratiedata</i>
<code>X=x(:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21]);</code>	<i>Inputvariabelen</i>
<code>S=X/10000;</code>	
<code>I=S';</code>	
<code>Y=x(:,30);</code>	<i>Outputvariabele</i>
<code>T=Y/10000;</code>	
<code>Z=T';</code>	
<code>ANN=newff(minmax(I),[10 1],{'logsig' 'purelin'},'trainlm');</code>	<i>Configuratie Neuraal netwerk</i>
<code>ANN.trainparam.epochs=250;</code>	<i>Aantal iteraties</i>
<code>Calibrate=train(ANN,I,Z);</code>	<i>Trainen Neuraal Netwerk</i>
<i>VALIDATIE</i>	
<code>a=load('Data_Twente.txt');</code>	<i>Invoer validatiedata</i>
<code>A=a(:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21]);</code>	<i>Inputvariabelen</i>
<code>B=A/10000;</code>	
<code>TESTSET =B';</code>	
<code>D=a(:,[30]);</code>	<i>Outputvariabele</i>
<code>E=D/10000;</code>	
<code>F=E';</code>	
<code>ResultatenNN=sim(Calibrate,TESTSET);</code>	<i>Validatiedata ingevoerd in gekalibreerd neuraal netwerk</i>

Figuur 9.1: Syntax in Matlab, voor voorspelling aantal arbeidsplaatsen in Twente

De invoerdata, zowel voor de kalibratie als de validatie, wordt eerst genormaliseerd. De voordelen van een genormaliseerde input in een neuraal netwerk zijn in hoofdstuk 3 duidelijk gemaakt. Normalisatie van de data betekent in dit geval dat voor elke variabele alle data teruggebracht worden tussen de 0 en 1. De onderlinge afstand tussen de waarden drukt het werkelijke verschil uit. Er zijn twee kleine testcases gemaakt waaruit duidelijk bleek dat zoals verwacht genormaliseerde data betere resultaten oplevert dan gewone data.

Nadat de kalibratiedata is ingevoerd (let op: er worden 21 inputvariabelen en 1 uitputvariabele ingevoerd), wordt de configuratie van het neurale netwerk beschreven. Het hierboven beschreven neurale netwerk bestaat uit één “hidden layer” met 10 neuronen en één “output layer”. Het aantal neuronen dat gebruikt wordt in de “hidden layer” varieert in dit onderzoek. In paragraaf 7.1 en paragraaf 7.2 wordt voor elk deel van het onderzoek onderzocht welke hoeveelheid neuronen in de “hidden layer” het meest geschikt is. Er worden twee activatiefuncties gebruikt. De eerste is de zogenaamde “logsig” activatiefunctie, die alle input in de “hidden layer” verwerkt tot outputwaarden tussen de 0 en 1. Deze waarden gaan naar de “output layer”, die een zogenaamde “purelin” activatiefunctie heeft, die de input in alle mogelijke waarden als output om kan zetten. Het trainen van het netwerk, en dus het aanpassen van de gewichten- en biaswaarden gebeurt via de Levenberg-Marquardt optimalisatie methode. Dit is een geschikte methode om neurale netwerken van normale grote redelijk efficiënt te trainen. Het maximale aantal iteratiestappen (epochs) is ingesteld op 250. Dit is voldoende om zeker te zijn dat bij meer trainingsstappen de kwaliteit van de resultaten nauwelijks meer zal verbeteren. Hierna stopt de training.

In de laatste stap wordt de inputvariabelen in het gedefinieerde neurale netwerk gestopt, waarmee het getraind wordt de ingevoerde outputvariabele te voorspellen. Hiermee is het neurale netwerk gekalibreerd. De validatie van het netwerk vindt op dezelfde manier plaats. De validatiedata wordt in het gekalibreerde neurale netwerk gestopt. Hiermee doet het neurale netwerk een voorspelling over de outputdata.

Bijlage C.1: Specificaties variabelen voorspelling arbeidsplaatsen

Beschrijving van de specificaties en de bronnen van de geselecteerde variabelen voor het voorspellen van het aantal arbeidsplaatsen in het bedrijfslocatiemodel.

Overzicht hoofdvariabelen

Parkeerfaciliteiten	Bron: Postcodekaart 2001, PTT Post; Wegenkaart NL; NS
De bepaling van deze variabele is indirect afgeleid uit 3 anderen factoren. Harde cijfers over het aantal beschikbare parkeerplaatsen per regio zijn niet beschikbaar. De 3 factoren zijn: • Ligging ten opzichte van hoofdwegenet; ligt een gebied in de buurt van een hoofdweg dan wordt aangenomen dat er ook veel parkeerplaatsen aanwezig zijn, omdat in dit type omgeving veel bedrijven zich vestigen die goed met de auto bereikbaar willen zijn en diens gevolg ook veel parkeerplaatsen tot hun beschikking nodig hebben. • Aanwezigheid van een station; ligt er in het gebied een station dan is aangenomen dat dit betekent dat er weinig parkeerplaatsen beschikbaar zijn. Twee redenen hiervoor, ten eerste is er alternatief, openbaar vervoer beschikbaar en ten tweede omdat stations zich doorgaans in stedelijke gebieden bevinden waar de ruimte schaars is en hierdoor ook het aantal beschikbare parkeermogelijkheden. • Stedelijkheid van het postcodegebied; van gebieden met de stedelijkheid in de klassen 5 en 4, respectievelijk "niet stedelijk" en "weinig stedelijk" wordt aangenomen dat er relatief veel ruimte beschikbaar is en daardoor ook voldoende parkeervoorzieningen beschikbaar zijn. Er worden 3 stappen gemaakt bij het toedelen van gebieden in kwaliteit parkeerfaciliteiten. Achtereenvolgens, in welke stedelijkheid valt het betreffende gebied, ligt er in de buurt een hoofdweg en als derde of er in de buurt een station aanwezig is. Toedeling gebieden (van klasse 1 (veel parkeerfaciliteiten) tot klasse 5 (zeer weinig parkeerfaciliteiten)): • Klasse 1: Landelijk of weinig stedelijk gebied en in de buurt van hoofdwegenet • Klasse 2: Landelijk gebied • Klasse 3: Weinig stedelijk gebied • Klasse 4: Matig tot zeer sterk stedelijk gebied • Klasse 5: Matig tot zeer sterk stedelijk gebied en in de buurt van station	
Aanwezigheid adequaat opgeleid personeel	Bron: CBS
Percentage werklozen in 2001 met minimaal HBO opleiding op niveau COROP gebied. Het niveau COROP gebied is gekozen, omdat bij de selectie van personeel bedrijven over het algemeen mensen in dienst nemen, woonachtig binnen een straal van 20-30 km. Dit komt wel overeen met de grootte van een Coropgebied. Probleem is dat locaties op de rand van een coropgebied ook beïnvloed wordt door de parameters van het aangrenzende coropgebied. Dit effect wordt verwaarloosd.	
Representatie gebouw	Bron: Tabellenboek werklocaties 2003, VROM
Deze variabele valt lastig uit te drukken in een meetbare eenheid die ook nog eenvoudig verkrijgbaar was. Er is gekozen om de koopprijs of huurprijs van een bedrijfsterrein of kantoorgebouw te nemen. De gemiddelde prijzen op een bepaalde locatie zijn gebruikt en daarnaast is er om kantoorlocaties en bedrijventerreinlocaties met elkaar te kunnen vergelijken een omrekenfactor gebruikt. Deze is gebaseerd op de beide gemiddelden van heel Nederland en hierna zijn alle bedrijfsterrein prijzen vermenigvuldigd met het relatieve verschil tussen beide prijzen. Deze data was alleen per woonplaats beschikbaar. Veel prijzen van bedrijventerreinen dan wel kantoorlocaties waren echter niet voor handen. Gevolg is dat voor veel plaatsen	

geen data beschikbaar is en voor alle overige plaatsen de uiteindelijke waarde is opgebouwd uit locaties waar wel data van beschikbaar was, wat tussen de 10% en 50 % van het totaal schommelt. In het "Tabellenboek Werklocaties 2003" zijn geen gegevens uit Noord Holland opgenomen. Om dit te ondervangen zijn de prijzen van de plaatsen Amsterdam en Haarlem, bron Meetlat 50.000+ (VSO), alsnog handmatig toegevoegd.

Aantal bedrijfsvestigingen

Bron: Kerncijfers wijken en Buurten 2001, CBS

Aantal bedrijfsvestigingen ingedeeld in 9 klassen:

Klasse Omschrijving

- | Klasse | Omschrijving |
|--------|-------------------------------------|
| 1 | 0 tot 10 bedrijfsvestigingen |
| 2 | 10 tot 20 bedrijfsvestigingen |
| 3 | 20 tot 50 bedrijfsvestigingen |
| 4 | 50 tot 100 bedrijfsvestigingen |
| 5 | 100 tot 200 bedrijfsvestigingen |
| 6 | 200 tot 500 bedrijfsvestigingen |
| 7 | 500 tot 1 000 bedrijfsvestigingen |
| 8 | 1 000 tot 2 000 bedrijfsvestigingen |
| 9 | 2 000 of meer bedrijfsvestigingen |

Stedelijkheid gebied

Bron: Nieuw Regionaal Model 2000, AVV

Stedelijkheid ingedeeld in 5 klassen:

Klasse Omschrijving Omgevingsadressendichtheid

- | Klasse | Omschrijving | Omgevingsadressendichtheid |
|--------|----------------------|---|
| 1 | zeer sterk stedelijk | > = 2500 omgevingsadressen per km ² |
| 2 | sterk stedelijk | 1500 - < 2500 omgevingsadressen per km ² |
| 3 | matig stedelijk | 1000 - < 1500 omgevingsadressen per km ² |
| 4 | weinig stedelijk | 500 - < 1000 omgevingsadressen per km ² |
| 5 | niet stedelijk | < 500 omgevingsadressen per km ² |

Overzicht vestigingsvariabelen

Ligging in Nederland

Bron: Nieuw Regionaal Model 2000, AVV

Bepaalt op basis van de afstand tot het middelpunt van Nederland, waarvoor de Domtoren in Utrecht is gebruikt. De afstand van een gebied ten opzichte van dit punt is bepaald met behulp van gebiedscoördinaten en de volgende formule:

$$\text{Afstand} = \sqrt{(X_{\text{utrecht}} - X_i)^2 + (Y_{\text{utrecht}} - Y_i)^2}$$

waar X en Y de x en y coördinaten voorstellen van betreffende locatie en met $X_{\text{utrecht}} = 136$ $Y_{\text{utrecht}} = 455$

Agglomeratie-effect

Bron: CBS

Er is gekozen om de indeling aan te houden die het CBS ook hanteert om aan te geven of een gebied binnen een agglomeratie valt. Dit betekent dat deze variabele eenvoudigweg twee opties heeft; een gebied valt wel binnen een agglomeratie of niet.

Infrastructuur	Bron: Postcodekaart 2001, PTT Post; Wegenkaart NL
Deze variabele is op een eenvoudige wijze bepaald. Op een kaart met niveau van drie positie postcodegebied is gekeken welke gebieden aansluiting hebben op het hoofdwegennet en deze hebben een hoge score gekregen. Gebieden die een te grote omvang hadden zijn niet meegerekend, evenals gebieden die een stadshart herbergen. Laatstgenoemde gebieden zijn op vier positie postcodegebied opnieuw, meer gedetailleerd, uitgewerkt. Hierbij hebben de stadsranden die aan een snelweg grenzen en waarvan aangenomen wordt dat ze een goede aansluiting hebben op het hoofdwegennet een hoge score gekregen. De stadscentra daarentegen hebben een lagere score gekregen. Deze variabele kan gezien worden als een variant op de variabele (potentiële) bereikbaarheid die al wordt meegenomen in het bedrijfslocatiemodel. Tillema (2004) concludeert echter al dat uit literatuur blijkt dat voor locatiekeuzes van ondernemers wellicht een andere vorm van bereikbaarheid geïntroduceerd dient te worden. Deze variabele kan gezien worden als een alternatieve variabele voor bereikbaarheid. Het is dan ook de moeite waard om te onderzoeken welke invloed deze bereikbaarheidsvariabele heeft op de locatiekeuzes van bedrijven.	
Oriëntatie t.o.v. externe markt	Bron: Nieuw Regionaal Model 2000, AVV
Voor deze variabele zijn de gebiedscoördinaten gebruikt. Deze zijn ingesteld als onafhankelijke variabele.	

Bijlage C.2: Specificaties variabelen voorspelling werkgelegenheidsaandeel bedrijfssectoren

Beschrijving van de specificaties en de bronnen van de geselecteerde variabelen voor het voorspellen van het werkgelegenheidsaandeel van bedrijfssectoren in het bedrijfslocatiemodel.

Overzicht variabelen industrie & productie sector

Uitbreidingsmogelijkheden	Bron: Tabellenboek werklocaties 2003, VROM
Deze variabele is bepaald op basis van de hoeveelheid hectaren bedrijfsterrein die beschikbaar zijn (zowel direct uitgeefbaar als indirect uitgeefbaar). Ook dient opgemerkt te worden dat de data hier op “woonplaatsniveau” beschikbaar is. Daarnaast dient te worden opgemerkt dat voor een betere vergelijking het aantal beschikbare hectares bedrijventerrein gedeeld is door het aantal vierposities postcodegebieden waaruit de betreffende woonplaats bestaat. Op deze manier wordt een betere afspiegeling gekregen van de werkelijke hoeveelheid beschikbaar bedrijventerrein, omdat anders grote woonplaatsen altijd door hun grote oppervlak relatief veel terreinen beschikbaar zouden hebben.	
Toeleverende bedrijven	Bron: Nieuw Regionaal Model 2000, AVV
Is dezelfde variabele voor gebruikt als de variabele “ligging in Nederland”. Het is een lastige factor om uit te drukken in een meetbare variabele. Elk afzonderlijk bedrijf heeft unieke leveranciers, die niet in bepaalde sectoren vallen te definiëren. Om toch een uitspraak te kunnen doen voor deze factor is aangenomen dat bedrijven die in het centrum van het land liggen relatief meer kans hebben om hun leveranciers van producten en diensten bij hun in de buurt te hebben, simpelweg door de gunstige ligging.	
Logistieke dienstverlening	Bron: CBS
Het aantal vestigingen van bedrijven in de sector Vervoer, Opslag en Communicatie is in een variabele vastgelegd. Deze variabele is ingedeeld in 9 klassen, waarbij klasse 1 voor een gebied staat waar zeer weinig logistieke bedrijven aanwezig zijn en klasse 9 voor een gebied waar een zeer groot aantal logistieke bedrijven aanwezig is. Het aggregatieniveau van de data is gemeenteniveau. Dit niveau lijkt hier voldoende doordat bedrijven die transportbedrijven nodig hebben, niet noodzakelijkerwijs een bedrijf uit de eigen wijk selecteren, maar wel nadrukkelijk in de nabije regio, vandaar op gemeenteniveau.	

Overzicht variabelen handel & distributie sector

Huurprijs- /grondprijsniveau	Bron: Tabellenboek werklocaties 2003, VROM
Voor deze factor is dezelfde variabele genomen die ook bij de factor representatie gebouw is gebruikt in bijlage 2.1. Dit houdt in dat de variabele gebruikt wordt die de koopprijs of huurprijs van een bedrijfsterrein of kantoorgebouw uitdrukt. Deze is precies van toepassing op deze factor. Voor de specificaties van deze variabele wordt verwezen naar de beschrijving bij de factor representatie gebouw in bijlage 2.1.	
Aanwezigheid zee- of luchthaven	Bron: Ministerie Verkeer en Waterstaat

Er bestaan geen data die deze variabele direct beschrijft. Om deze reden is de data op basis van een aantal aannames gecreëerd. De belangrijkste luchthavens en zeehavens in Nederland worden geselecteerd en ingedeeld in klassen. Daarna wordt aan elk gebied een kengetal toegekend, gebaseerd op de ligging ten opzichte van deze punten. Hierbij krijgen alle gebieden in een bepaalde straal rond de betreffende lucht of zeehaven dezelfde waarde toegekend. Gebieden kunnen meerdere waarden toegekend krijgen van verschillende lucht of zeehavens bij hun in de buurt, maar de som ligt nooit hoger dan 1.

De volgende zee en luchthavens zijn geselecteerd:

Wat betreft luchthavens zijn Schiphol en de regionale luchthavens geselecteerd, dit zijn:

Schiphol	(39.000.000 passagiers, 400.000 vliegtuigbewegingen in 2003)
Den Helder	(53.000 passagiers, 10.000 vliegbewegingen in 2001)
Eindhoven Airport	(280.000 passagiers in 2001)
Groningen Eelde	(148.000 passagiers, 67.000 vliegtuigbewegingen in 2002)
Lelystad	
Maastricht Aachen	(369.000 passagiers, 41.000 vliegtuigbewegingen in 2003)
Rotterdam	(825.000 passagiers, 93.000 vliegtuigbewegingen in 2001)
Twente Enschede	

Het gewicht dat is toegekend aan de luchthavens is gebaseerd op de hoeveelheid passagiers en vliegtuigbewegingen die verwerkt worden. De reikwijdte (straal) van de impact is gebaseerd op de uitstraling van de betreffende luchthaven.

Luchthaven	Grootteklasse	Gewicht	Reikwijdte/impact	Coördinaten (X,Y)
Schiphol	Groot	1	20 km	(112, 481)
Den Helder	Klein	0.25	10 km	(114, 545)
Eindhoven	Middelgroot	0.5	15 km	(155, 385)
Groningen	Klein	0.25	10 km + Groningen	(235, 572)
Lelystad	Klein	0.25	10 km + Lelystad	(157, 495)
Maastricht	Middelgroot	0.5	15 km	(183, 327)
Rotterdam	Redelijk groot	0.75	15 km	(090, 441)
Twente	Klein	0.25	10 km	(258, 476)

Wat betreft zeehavens zijn de volgende locaties geselecteerd: Rotterdam-Rijnmond, Eemsmond, Vlissingen & Terneuzen en Amsterdam & IJmuiden. De coördinaten zijn bepaald door een inschatting te maken van het middelpunt van het betreffende havengebied en hiervan de coördinaten te noteren. Doordat er met reikwijdtes (stralen) wordt gewerkt, wordt het gehele gebied meegenomen. Van gebieden die verder weg gelegen zijn dan de helft van de afstand van de maximale reikwijdte hebben is hun gewicht gehalveerd. Op deze manier wordt bereikt dat gebieden dichtbij een zee en/of luchthaven aantrekkelijker zijn dan veraf gelegen gebieden. Voor de havens van Amsterdam & IJmuiden en Vlissingen & Terneuzen zijn twee coördinaten genomen waarbij zodat twee cirkels het hele gebied bestrijken.

Zeehaven	Grootteklasse	Gewicht	Reikwijdte/impact	Coördinaten (X,Y)
Rotterdam-Rijnmond	Groot	1	30 km	(077, 433)
Eemsmond	Klein	0.25	15 km	(253, 601)
Vlissingen & Terneuzen	Klein	0.25	10 km	(030, 386) & (047, 372)
Amsterdam & IJmuiden	Middelgroot	0.5	15 km & 10 km	(116, 491) & (103, 499)

Opmerking: het is erg lastig om de reikwijdte van de impact van een bepaalde zee- of luchthaven aan te geven. De aannames die hierbij zijn gedaan zijn betwifelbaar. Er is in twee gevallen (Groningen en Lelystad) een extra gebied toegevoegd, omdat het mogelijk is dat een bepaalde luchthaven in 1 bepaalde richting meer invloed heeft. Bijvoorbeeld vliegveld Eelde, waarvan verondersteld wordt dat de impact richting het Noorden, richting Groningen verder reikt dan naar bijvoorbeeld het Oosten.

Logistieke dienstverlening

Bron: CBS

Het aantal vestigingen van bedrijven in de sector Vervoer, Opslag en Communicatie is in een variabele vastgelegd. Deze variabele is ingedeeld in 9 klassen, waarbij klasse 1 voor een gebied staat waar zeer weinig logistieke bedrijven aanwezig zijn en klasse 9 voor een gebied waar een zeer groot aantal logistieke bedrijven aanwezig is. Het aggregatiegraad van de data is gemeenteniveau. Dit niveau lijkt hier voldoende doordat bedrijven die transportbedrijven nodig hebben, niet noodzakelijkerwijs een bedrijf uit de eigen wijk selecteren, maar wel nadrukkelijk in de nabije regio, vandaar op gemeenteniveau.

Overzicht variabelen diensten sector

Quality of life factoren

Bron: VSO 2003

Er is gebruik gemaakt van een onderzoek waarbij elke gemeente een rangnummer toegewezen kreeg op basis van een aantal kwaliteitsindicatoren. Er werden in het onderzoek 8 criteria gebruikt: wonen, voorzieningen, beschikbaarheid van zorg, veiligheid, bereikbaarheid, economie, mogelijkheden voor besteding van vrije tijd en de samenstelling van de bevolking. Deze criteria sluiten prima aan bij de definitie van quality of life factoren. Het niveau van de data op gemeentegrootte voldoet prima, doordat de quality of life factoren bedoeld zijn om te geven of een regio aantrekkelijk is voor potentiële werknemers om te wonen.

Representatie omgeving

Bron: Bureau Louter 2001

Er is gebruik gemaakt van een onderzoek waarbij alle gemeenten een rangnummer krijgen op basis van hun economische prestaties, ontwikkeling van economische kerngebieden en sterkte van kantorenmarktlocaties.

Aangenomen is dat gemeenten die hoge positieve score hebben in dit onderzoek ook door ondernemers worden gezien als economische toplocaties. Dit is een omgeving waarin bedrijven zich graag vestigen. Het niveau van de data op gemeentegrootte lijkt redelijk, doordat de variabele informatie geeft over het feit of er in een gemeente al dan niet aantrekkelijke economische locaties aanwezig zijn. Dit betekent dat de regio voor een bedrijf al dan niet uitstraling heeft.

OV-bereikbaarheid

Bron: NS; Postcodekaart 2001, PTT Post

Bij de bepaling van deze variabele is uitsluitend gekeken naar openbaar vervoergedeelte dat door de NS verzorgd wordt. Voor alle stations in Nederland is bepaald in welk driepositie postcodegebied zij zich bevinden. Deze gebieden hebben de waarde 0,5 gekregen. De waarde 1 is toegekend aan gebieden waarin zich een station vindt dat door de NS aangeduid wordt als "belangrijk station". De data is omgezet naar vierpositie postcodegebieden.

Bijlage D: Resultaten bepaling optimale categorisering output

Overzicht resultaten kalibratie bedrijfslocatiemodel met indeling uitkomstvariabele in oplopend aantal categorieën. Voor de kalibratie van het bedrijfslocatiemodel is een neurale netwerk gebruikt met 50 neuronen in de hidden layer. Daarnaast zijn alle input- en outputvariabelen genormaliseerd. Deze aanpassingen zijn pas later in het onderzoek naar voren gekomen als geschikte configuratie voor het bedrijfslocatiemodel. De eerste keer dat deze toetsing uitgevoerd werd gebeurde dan ook met een andere configuratie. Met deze configuratie werden dezelfde resultaten geboekt, zij het in een andere ordergrootte. Om zeker te zijn dat voor een optimaal geconfigureerd netwerk dezelfde conclusies getrokken konden worden, is de toetsing opnieuw uitgevoerd en opgenomen in deze bijlage.

Aantal arbeidsplaatsen	Kwaliteit gekalibreerd model
10 categorieën	- <i>Range 0.8-1.2</i> : 82% - <i>Range 0.5-1.5</i> : 96%
50 categorieën	- <i>Range 0.8-1.2</i> : 79% - <i>Range 0.5-1.5</i> : 93%
100 categorieën	- <i>Range 0.8-1.2</i> : 80% - <i>Range 0.5-1.5</i> : 95%
150 categorieën	- <i>Range 0.8-1.2</i> : 77% - <i>Range 0.5-1.5</i> : 92%
200 categorieën	- <i>Range 0.8-1.2</i> : 77% - <i>Range 0.5-1.5</i> : 93%
250 categorieën	- <i>Range 0.8-1.2</i> : 76% - <i>Range 0.5-1.5</i> : 92%

Bijlage E: Bepaling optimale aantal neuronen in neuraal netwerk

In deze bijlage wordt het optimale aantal neuronen in “hidden layer” van het neurale netwerk onderzocht, zodat de kwaliteit van de uitkomsten voor de twee onderdelen van het bedrijfslocatiemodel zo hoog mogelijk zijn. Daarnaast wordt onderscheidt gemaakt tussen de kalibratie van het model en de toepasbaarheid op gebiedsniveau van het model. In totaal levert dit dus vier onderzochte configuraties van het neurale netwerk op.

Optimaal aantal neuronen voor kalibratie neuraal netwerk

Bij de kalibratie van het neurale netwerk voor de voorspellingen van het totale aantal arbeidsplaatsen wordt al duidelijk dat slechts een groot aantal neuronen zorgt voor een optimale configuratie van het neuraal netwerk. Om deze reden is voor het gedeelte dat het werkgelegenheidsaandeel van bedrijfssectoren voorspelt, slechts gekeken naar een neuraal netwerk met 30, 40 of 50 neuronen in de “hidden layer”. De reden dat in twee gevallen ook naar het gemiddelde van 5 runs van het neurale netwerk wordt gekeken, is omdat de uitkomsten in deze gevallen nagenoeg gelijk zijn.

Aantal neuronen	Totaal arbeidsplaatsen		Industrie/ Productie		Handel/ Distributie		Diensten	
	Range	Range	Range	Range	Range	Range	Range	Range
	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5
10	65	90						
20	69	91						
30	72	92	56	82	67	88	70	90
40	76	93	62	85	71	89	74	91
Gem. van 5 runs					68,6	88,2	69,5	89
50	80	95	57	82	71	90	75	92
Gem. van 5 runs					62,8	86,2	67,3	89,2

Optimaal aantal neuronen voor bepaling toepasbaarheid model op gebiedsniveau

Bij de testen van het optimale aantal neuronen voor de bepaling van de toepasbaarheid van het model op gebiedsniveau, door het voorspellen van het aantal arbeidsplaatsen in de testcase Twente, wordt al duidelijk dat slechts een klein aantal neuronen zorgt voor een optimale configuratie van het neurale netwerk. Om deze reden is voor het gedeelte dat het werkgelegenheidsaandeel van bedrijfssectoren voorspelt, slechts gekeken naar een neurale netwerk met 5, 10 of 20 neuronen in de "hidden layer".

Aantal neuronen	Totaal arbeidsplaatsen		Industrie/ Productie		Handel/ Distributie		Diensten	
	Range	Range	Range	Range	Range	Range	Range	Range
	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5
5	80	91	55	86	73	98	80	95
10	80	93	48	86	71	95	75	95
15	75	91						
20	75	89	46	84	68	95	75	95
30	75	93						
40	75	91						
50	73	91						

Optimaal aantal neuronen voor testcases

De test van het optimale aantal neuronen voor de bepaling van de toepasbaarheid van het model op gebiedsniveau is uitgebreid, door 3 testcases te selecteren met elk 50 willekeurig gekozen cases. Hierbij is alleen gekeken naar de voorspellingen van het model voor het aantal arbeidsplaatsen en daarnaast voor testcase I ook naar de voorspellingen voor de sector handel & distributie.

Aantal neuronen	Testcase I, tot. aantal arbeidspkts		Testcase II, tot. aantal arbeidspkts		Testcase III, tot. aantal arbeidspkts		Testcase I, sector handel & distributie	
	Range	Range	Range	Range	Range	Range	Range	Range
	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5	0.8 -1.2	0.5 -1.5
5	68	90	66	91	68	95	62	74
10	70	92	58	93	70	95	64	84
20	74	88	75	91	71	96	68	84
30	78	92	82	93	87	100	66	82
40	76	92	67	93	89	100	68	86
50	76	94	80	91	91	100	72	90

Bijlage F.1: Resultaten voorspelling aantal arbeidsplaatsen

In deze bijlage staat een overzicht van de resultaten van de kalibratie van het model en de resultaten van de bepaling van de toepasbaarheid van het model op gebiedsniveau met behulp van de testcase Twente. In deze bijlage gaat het hierbij om de voorspellingen van het bedrijfslocatiemodel voor het aantal arbeidsplaatsen in een gebied. In de tabel betekent het noemen van een variabele dat deze *niet* is meegenomen in de kalibratie van het neurale netwerk (uitzondering 1^e stap; toetsing met alle variabelen). Deze methode houdt in dat de totale variabelenset als basis is genomen en daarna steeds één of meerdere variabelen uit de totale variabelenset weg zijn gelaten. De stappen die gedaan worden zijn beschreven in paragraaf 6.3.

	Kalibratie		Testcase Twente	
	Range 0.8-1.2	Range 0.5-1.5	Range 0.8-1.2	Range 0.5-1.5
Stap1: Alle variabelen	80	95	80	93
<i>Stap 2: Hoofdvariabelen</i>	69	91	71	93
- Parkeerfaciliteiten	80	94	82	93
- Beschikbaarheid adequaat opgeleid personeel	80	94	82	95
- Aantal bedrijfsvestigingen	75	93	80	91
- Stedelijkheid gebied	78	94	77	91
Stap 3: Hoofdvar. + Vestigingsvar.	68	91	73	95
<i>Vestiging plaatsvariabelen</i>	76	93	80	93
- Ligging in NL	80	94	84	93
- Agglomeratie-effect	77	94	80	93
- Infrastructuur	76	93	79	91
- Oriëntatie t.o.v. externe markt	79	94	82	95
Stap 4: Sectorspecifieke variabelen	74	93	77	93
<i>Industrie/Productie var.</i>	78	94	79	91
- Uitbreidingsmogelijkheden	76	94	75	91
- Toeleverende bedrijven	80	95	80	95
- Logistieke dienstverlening	79	94	82	93
<i>Handel/Distributie var.</i>	78	94	77	91
- Aanwezigheid zee- of luchthaven	78	94	82	93
- Logistieke dienstverlening	79	94	82	93
<i>Diensten var.</i>	77	94	79	96
- Quality of life factoren	80	95	77	95
- Representatie omgeving	80	94	79	93
- OV-bereikbaarheid	70	92	80	95
Stap 5: Zonder HH en bereikbaarheid	70	92	70	93
<i>Huishoudmodel var.</i>	72	92	70	89
<i>Bereikbaarheidsmodel var.</i>	80	94	75	93

Bijlage F.2: Resultaten voorspelling werkgelegenheidsaandeel bedrijfssectoren

In deze bijlage staat een opnieuw een overzicht van de resultaten van de kalibratie van het model en de resultaten van de bepaling van de toepasbaarheid van het model op gebiedsniveau met behulp van de testcase Twente. Deze keer worden de resultaten weergegeven van de voorspellingen van het bedrijfslocatiemodel voor het werkgelegenheidsaandeel van bedrijfssectoren in een gebied. Net als in bijlage G.1 betekent het noemen van een variabele dat deze *niet* is meegenomen in de kalibratie van het neurale netwerk (uitzondering 1^e stap; toetsing met alle variabelen). Deze methode houdt in dat de totale variabelenset als basis is genomen en daarna steeds één of meerdere variabelen uit de totale variabelenset weg zijn gelaten. Dezelfde stappen die in bijlage G.1 zijn uitgevoerd, worden ook hier uitgevoerd. Uitgezonderd de stap waarbij ook elke variabele afzonderlijk wordt bekeken. Dit wordt alleen gedaan voor de sectorspecifieke variabelen. Dit alles is terug te vinden in de beschrijving van de stappen voor het testen van de variabelen in paragraaf 6.3.

Overzicht resultaten industrie & productie sector

	Kalibratie		Testcase Twente	
	Range 0.8-1.2	Range 0.5-1.5	Range 0.8-1.2	Range 0.5-1.5
Alle variabelen	62	85	55	86
<i>Hoofdvariabelen</i>	53	81	50	86
<i>Vestigingsplaatsvariabelen</i>	57	82	52	86
<i>Industrie/Productie var.</i>	57	83	50	84
- Uitbreidingsmogelijkheden	60	84	57	86
- Toeleverende bedrijven	58	84	55	86
- Logistieke dienstverlening	61	85	55	86
<i>Handel/Distributie & Diensten var.</i>	56	83	54	84
<i>HH en bereikbaarheid var.</i>	54	82	54	84

Overzicht resultaten handel & distributie sector

	Kalibratie		Testcase Twente	
	Range 0.8-1.2	Range 0.5-1.5	Range 0.8-1.2	Range 0.5-1.5
Alle variabelen	71	89	73	98
<i>Hoofdvariabelen</i>	64	87	71	93
<i>Vestigingsplaatsvariabelen</i>	67	88	66	95
<i>Handel/Distributie var.</i>	68	87	73	95
- Aanwezigheid zee- of luchthaven	65	87	77	95
- Logistieke dienstverlening	68	88	70	96
<i>Industrie/Productie & Diensten var.</i>	67	88	71	96
<i>HH en bereikbaarheid var.</i>	62	85	64	93

Overzicht resultaten diensten sector

	Kalibratie		Testcase Twente	
	Range 0.8-1.2	Range 0.5-1.5	Range 0.8-1.2	Range 0.5-1.5
Alle variabelen	73	92	80	95
<i>Hoofdvariabelen</i>	68	89	73	93
<i>Vestigingsplaatsvariabelen</i>	72	91	79	93
<i>Diensten var.</i>	73	91	77	95
- Quality of life factoren	66	89	80	95
- Representatie omgeving	74	91	82	95
- OV-bereikbaarheid	76	92	79	95
<i>Handel/Distributie & Industrie/Productie var.</i>	72	90	80	93
<i>HH en bereikbaarheid var.</i>	67	89	66	91

Bijlage G: Resultaten toetsen model in zes regio's

In deze bijlage staan de resultaten van de toetsing van de vraag of het bedrijfslocatiemodel toepasbaar is op elke willekeurige locatie in Nederland. De in hoofdstuk 6 geselecteerde regio's worden afzonderlijk getoetst in het bedrijfslocatiemodel. Daarnaast wordt er gekeken naar de kwaliteit van de voorspellingen van het totaal aantal arbeidsplaatsen, het aantal arbeidsplaatsen in de sector industrie & productie, het aantal arbeidsplaatsen in de sector handel & distributie, het aantal arbeidsplaatsen in de sector diensten in een regio.

De kwaliteit van de kalibratie wordt getoetst met de data van heel Nederland, behalve de te toetsen regio. Daarna vindt de verificatie plaats door in hetzelfde netwerk de data van de geselecteerde regio in te voeren. De voorspelde uitkomsten door het model worden vergeleken met de geobserveerde uitkomsten. Dit proces wordt voor elk van de geselecteerde regio's afzonderlijk uitgevoerd.

		Totaal arbeidsplaatsen		Industrie & Productie		Handel & Distributie		Diensten	
		<i>0.8 -1.2</i>	<i>0.5 -1.5</i>	<i>0.8 -1.2</i>	<i>0.5 -1.5</i>	<i>0.8 -1.2</i>	<i>0.5 -1.5</i>	<i>0.8 -1.2</i>	<i>0.5 -1.5</i>
Utrecht	Kalibratie	80	94	54	82	68	88	74	92
	Verificatie	73	86	50	84	52	80	66	86
Groningen	Kalibratie	79	94	62	86	70	89	76	91
	Verificatie	65	90	36	68	64	97	61	87
Rotterdam	Kalibratie	81	94	60	85	69	89	75	92
	Verificatie	79	93	53	85	64	85	74	90
Haarlemmermeer	Kalibratie	80	94	62	85	70	88	74	91
	Verificatie	63	82	44	70	52	89	63	85
Oss, Weert & Roermond	Kalibratie	78	94	61	84	71	89	75	91
	Verificatie	78	91	57	87	70	96	83	96
2 Friese gemeentes	Kalibratie	80	95	60	85	71	90	75	91
	Verificatie	42	74	32	66	40	68	37	82