

Interactieve routeplanning van bulktransport:

Een praktijktoepassing.

Antoon Kolen, Meindert Snoei
Erasmus Universiteit Rotterdam

Samenvatting

Voor de fabriek in Zwolle van Hendrix' Voeders is op een IBM-PC een prototype van een interactief routeplanningssysteem ontworpen voor het bulktransport van veevoeder. Wij beschrijven de twee pijlers waarop dit systeem rust. De eerste pijler bestaat uit afstanden en rijtijden die geschat zijn uit gegevens over zowel de locatie van de klanten als over de in werkelijkheid verreden afstanden en tijden zoals deze uit tachograafschijven zijn afgelezen. De tweede pijler is een heuristisch routeplanning algoritme van het insertie type dat suggesties voor routes doorgeeft aan de planner, die deze vervolgens al dan niet kan overnemen.

Inleiding

Hendrix' Voeders vormt de voederdivisie van Hendrix' International. Hendrix' International b.v. is een zelfstandig opererende onderneming binnen de B.P.-groep en omvat een groep internationaal georiënteerde bedrijven met vestigingen binnen en buiten Europa. De activiteiten van deze vestigingen zijn ontwikkeling, fabricage en verkoop van produkten die behoren tot de veredelings- en levensmiddelensector. Hendrix' Voeders telt vier productievestigingen in Nederland. Nadat de grondstoffen opgeslagen zijn in silo's

correspondentieadressen:

dr. ir. A.W.J. Kolen
Econometrisch Instituut H2-22
Postbus 1738
3000 DR Rotterdam

drs. M. Snoei
BSO/Management Support b.v.
Postbus 8348
3503 RH Utrecht

worden zij in de mengerij in de juiste doses samengevoegd. Na het malen van de mengsels tot meel worden de produkten in diverse vormen geperst en opgeslagen in silo's. De bulkwagenbelading vormt de afsluiting van het produktieproces. Praktisch alle mengvoer wordt in bulkwagens afgeleverd, slechts een klein deel bestaat uit zakgoed. De expeditieplanner tracht een planning op te zetten die de kosten van het bulktransport minimaliseert. Naast het minimaliseren van het totaal aantal kilometers streeft hij naar minimale wachttijden bij de bulkbelading, een gelijke verdeling van het aantal werkuren van de chauffeurs over de week en een evenredige verdeling van het aantal bestellingen over het eigen en vreemd vervoer. Bovendien wordt het onderbezetten van de bulkwagens in hoge mate inefficiënt geacht. Bij het streven naar zijn doelstelling ziet de planner zich gesteld voor beperkingen van verschillende aard. Beperkingen die te maken hebben met:

1. Transport

- de beschikbaarheid van bulkwagens
- de capaciteit van de bulkwagens
- de compartimentenindeling van de bulkwagens
- de beschikbaarheid van chauffeurs
- de bereikbaarheid van de klant

2. Het bestelgedrag van de klant

- de te late bestellingen
- de termijnbestellingen

3. de afstemming van de expeditieplanning op de produktieplanning.

Gezien de vertrouwelijkheid van veel gegevens zullen wij ons bij de bespreking van deze beperkingen beperken tot die welke voor het begrijpen van het verhaal noodzakelijk zijn. Vertrouwend op inzicht en ervaring weet de planner, in een soort van samenspel met alle genoemde beperkingen, zijn doelstelling te bereiken. De middelen die hem hierbij ter beschikking staan, zijn beperkt. Aan de hand van proefbonnen, waarop voor elke bestelling de relevante gegevens staan vermeld, maakt de planner handmatig een planning. De vraag rees of automatisering van de routeplanning de taak van de planner kan verlichten. Aan de vakgroep Mathematische Besliskunde van de Erasmus Universiteit werd gevraagd dit te onderzoeken. Dit heeft geresulteerd in een prototype van een interactief routeplanningssysteem (waarvan een ruwe versie is geïmplementeerd op een IBM-PC). Resultaten met dit systeem zijn dermate bemoedigend dat besloten is om tot automatisering van de routeplanning over te gaan. Omdat binnen de fabriek in Zwolle alle genoemde beperkingen een rol spelen is het

interactieve routeplanningssysteem daarvoor ontwikkeld. In dit verhaal beschrijven wij de twee pijlers waarop dit systeem rust. Allereerst wordt in Sectie 1 ingegaan op de berekening van de afstanden en rijtijden. In Sectie 2 wordt een heuristisch routeplanningssysteem van het insertie type beschreven dat suggesties voor routes doorgeeft aan de planner, die deze vervolgens al dan niet kan overnemen. In Sectie 3 wordt kort ingegaan op de schattingsmethoden zoals deze in Sectie 1 gebruikt zijn.

1. Berekening afstanden en rijtijden

De laatste jaren is veel aandacht geschonken aan de ontwikkeling van netwerk databases. In een netwerk database worden alle gegevens met betrekking tot de wegenstructuur van een bepaald geografisch gebied opgeslagen. Het achterliggende wegnetwerk kan gerepresenteerd worden in de vorm van een gerichte graaf, waarin de kanten wegen voorstellen en knopen corresponderen met kruispunten van wegen. Door nu vervolgens klanten toe te wijzen aan het dichtstbijzijnde kruispunt van wegen of klanten als knoop in het netwerk op te nemen, kan, gegeven afstanden en rijtijden van kanten, met behulp van een kortste pad algoritme een nauwkeurige benadering van de afstanden en rijtijden tussen klanten berekend worden. Rijkswaterstaat beschikt over een databestand waarin 52.008 knopen en 74.811 wegen zijn opgenomen. Per weg staat aangegeven of het hier een één- of tweerichtingsweg betreft zodat het aantal kanten bijna twee keer zo groot is als het aantal wegen. Bij de vakgroep Mathematische Besliskunde van de Erasmus Universiteit is men bezig dit databestand om te zetten in een database op de VAX-computer met daaraan gekoppeld de mogelijkheid om via een PC het netwerk voor een gegeven klantenbestand te definiëren. Klanten kunnen aan een reeds bestaande knoop worden toegekend of men kan nieuwe wegen introduceren. Hierbij wordt gebruik gemaakt van de postcode van een klant. De uiteindelijke berekening van afstanden en rijtijden gebeurt weer op de VAX. Dit database-systeem zal vermoedelijk eind maart 1986 operationeel zijn.

Naast de netwerk aanpak is er de meer op econometrische leest geschoeide methode die bestaat uit het schatten van de verreden afstand uit de Euclidische afstand. MADSEN (1984) geeft een overzicht van gebruikte econometrische methoden en beschrijft tevens het gebruik van netwerk databases. Wij zijn uitgegaan van de volgende relatie

$$y_i = \alpha + \beta x_i + u_i, \quad (1.1)$$

waarbij y_i = de te verklaren verreden afstand

x_i = de Euclidische afstand

u_i = de storingsterm

α, β = de regressiecoëfficiënten.

De tijdsduur van een rit is van belang in verband met zowel de werktijden van de chauffeurs als ook met het laden voor de volgende rit. Diverse ritten moeten tot een werkdag voor de chauffeurs worden gekoppeld zodanig dat de totale werktijd zo evenredig mogelijk over de chauffeurs is verdeeld. Verder is van belang dat de tijdsduren van ritten een zekere spreiding vertonen, dit om wachttijden bij herbelading te voorkomen. Wij zijn uitgegaan van de volgende relatie

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \quad (1.2)$$

waarbij y_i = de te verklaren rijtijd van een rit

x_{1i} = de Euclidische verreden afstand in een rit

x_{2i} = het aantal losadressen

x_{3i} = het tonnage van de bulkwagen

u_i = de storingsterm

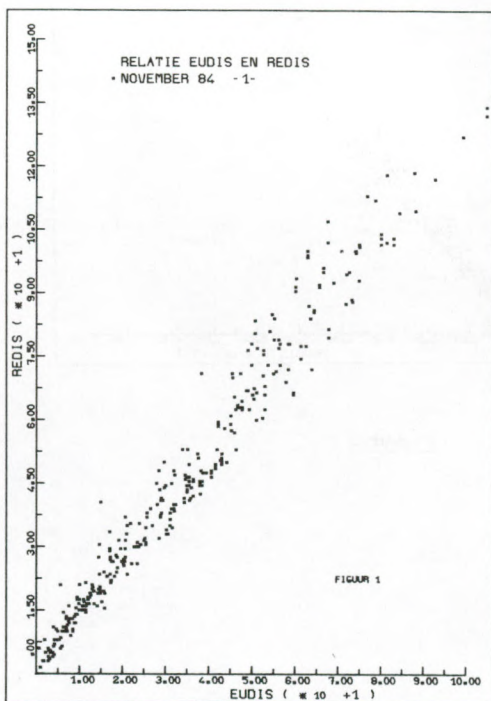
$\alpha, \beta_1, \beta_2, \beta_3$ = de regressiecoëfficiënten.

Teneinde de Euclidische afstanden te kunnen berekenen is voor iedere klant een coördinaatsaanduiding bepaald. Hiervoor is gebruik gemaakt van de bij de Rijksplanalogische Dienst beschikbare programmatuur, waarbij elk adres dat voorzien is van zowel een geldende postcode als een bestaand huisnummer, binnen een internationaal assenstelsel met middelpunt Lyon (Frankrijk), coördinaten krijgt toegewezen. De aanduiding is niet exact omdat elk adres wordt geplaatst in de linker benedenhoek van het gebied van 500 x 500 meter waarin het zich bevindt. Bij het bepalen van werkelijk verreden afstanden en rijtijden hadden wij de beschikking over de dagstaten en tachograafschijven van november en december 1984. Op de dagstaten staat o.a. vermeld welke klanten zijn bezocht en in welke volgorde. De tachograaf gegevens vermelden de verreden afstand en rijtijd. Door koppeling van dagstaat- en tachograafgegevens komen gegevens over verreden afstand en rijtijd tussen klanten

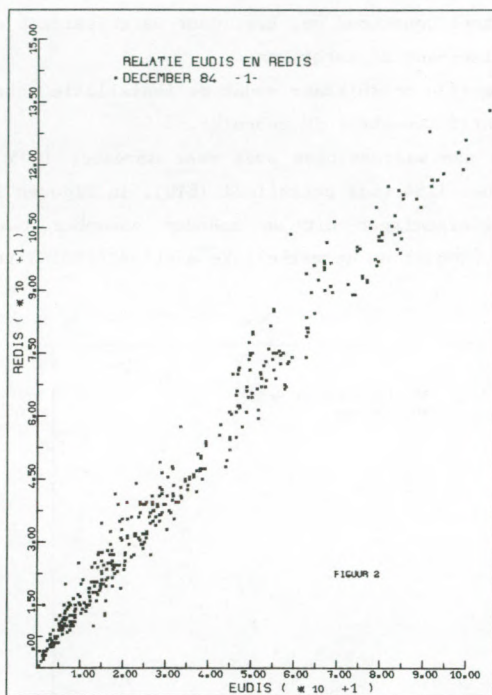
beschikbaar. Bij het screenen van het databestand kwamen wij de volgende problemen tegen

- postcode klant is niet bekend of correspondeert niet met afleveradres
- er treedt incidenteel oponthoud op, b.v. door materiaalpech of koffiepauzes
- dagstaten zijn onleesbaar of ontbreken
- tachograafgegevens zijn onbruikbaar omdat de installatie onbetrouwbaar is of omdat dezelfde schijf twee keer is gebruikt.

Uiteindelijk bleven aan waarnemingen over voor november 1095 (was potentieel 1592) en voor december 1254 (was potentieel 1819). In Figures 1 en 2 zijn voor een groot aantal waarnemingen uit de maanden november resp. december de Euclidische afstand (EUDIS) en de werkelijke afstand (REDIS) tegen elkaar



Figuur 1



Figuur 2

uitgezet. Uit de puntenwolk is af te lezen dat het redelijk is te veronderstellen dat er een lineair verband bestaat tussen de te verklaren en de verklarende variabele. Het valt verder op dat naarmate de Euclidische afstand toeneemt de spreiding tussen de waarnemingen groter wordt. De variantie van de storingen verbonden aan deze waarnemingen zijn niet voor elke waarde van de verklarende variabele identiek. Deze observatie duidt op de aanwezigheid van heteroscedasticiteit. In Sectie 3 wordt de gebruikte schattingsprocedure van Theil uiteengezet. Wij volstaan hier met het vermelden van de resultaten (tussen haken staat de standaard fout vermeld).

November

$$\alpha = 1,2216 \quad (0,0973)$$

$$\beta = 1,3546 \quad (0,0116)$$

$$R^2 = 0,9513$$

December

$$\alpha = 1,2149 \quad (0,0890)$$

$$\beta = 1,3460 \quad (0,0108)$$

$$R^2 = 0,9488$$

Een voorspelling van de afstanden in de maand december met de geschatte parameters uit de maand november levert een absolute voorspelfout op van 4,27 km per gereden afstand (i.e., een voorspelfout van 9,09%); omgekeerd is dit 4,05 km (9,07%). Wat betreft de rijtijden per rit gegeven door de relatie (1.2) hadden wij geen reden om af te wijken van de standaard veronderstellingen betreffende het lineaire model, d.w.z. $E(u_1) = 0$ en $\text{Var}(u_1) = \sigma^2$. Alvorens de regressie uit te voeren voor (1.2) worden de verklarende variabelen gestandaardiseerd naar de dimensie tijd. Voor de gemiddelde snelheid van de bulkwagens staat 64 km/uur, voor het lossen een gemiddelde van 24 ton/uur en voor elke klant een extra lostijd van 10 minuten. Model (1.2) is geschat met 291 waarnemingen van de maand november. De resultaten zijn

$$\alpha = 0,6595 \quad (0,1115)$$

$$\beta_1 = 1,0307 \quad (0,0274)$$

$$\beta_2 = 1,6859 \quad (0,1252)$$

$$\beta_3 = 0,6216 \quad (0,1205)$$

$$R^2 = 0,8009 \quad \sigma^2 = 0,1104$$

Een voorspelling voor de tijdsduur van de 363 ritten in december met het geschatte model van november geeft een absolute voorspelfout van 0,3160 uur/rit (i.e. een voorspelfout van 10%). Vooruitlopend op de resultaten van Sectie 2 kunnen wij stellen dat de relaties (1.1) en (1.2) bijzonder goede resultaten hebben gegeven in de gebruikte routeringsheuristiek.

2. De routeringsheuristiek

Het vervoer bij Hendrix' Voeders geschiedt met behulp van verschillende bulkwagens die elk bestaan uit een aantal compartimenten van 4 of 2 ton. Omdat bulkwagens niet in staat zijn voeder af te wegen neemt de lading voor een klant noodzakelijkerwijs een geheel aantal compartimenten in beslag. Dit geeft complicaties als wij ons de vraag stellen of een gegeven klantenbestand I door een gegeven bulkwagen kan worden bediend. Nodige voorwaarden voor het fysiek kunnen inplannen van klantenbestand I worden gegeven door

$$\sum_{i \in I} q_i \leq Q, \quad (2.1)$$

en

$$\sum_{i \in I} \alpha_i \leq A, \quad (2.2)$$

waarbij q_i de vraag van klant i is uitgedrukt in tonnen, α_i het minimaal aantal compartimenten dat nodig is om q_i te vervoeren, $i \in I$, Q is het laadvermogen van de bulkwagen en A het beschikbare aantal compartimenten. Enkele voorbeelden mogen dit verduidelijken.

Voorbeeld 2.1. Een bulkwagen met compartimentenindeling 4-2-2-2, twee klanten met vraag vijf respectievelijk zes. Wij hebben $q_1 = 5$, $q_2 = 6$; $\alpha_1 = 2$ en $\alpha_2 = 2$ immers in beide gevallen volstaan de compartimenten 4-2. Wij zien dat aan (2.2) is voldaan, maar niet aan (2.1).

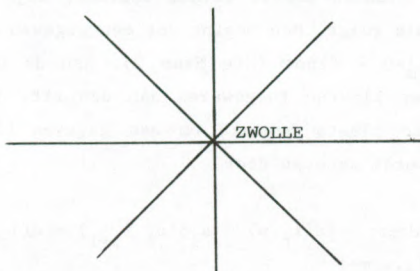
Voorbeeld 2.2. Een bulkwagen met compartimentenindeling 4-4-4-4-4-2-2-2, zes klanten met vraag vijf elk. Wij hebben $q_i = 6$, $\alpha_i = 2$, $i=1,2,3,4,5$. Wij zien dat aan (2.1) is voldaan, maar niet aan (2.2).

Voorbeeld 2.3. Een bulkwagen met compartimentenindeling 4-2-2-2, twee klanten met vraag vijf elk. Wij hebben $q_i = 5$, $\alpha_i = 2$, $i=1,2$. Wij zien dat aan (2.1) en (2.2) is voldaan, maar tevens dat het niet mogelijk is de lading fysiek in te plannen.

Uit de Voorbeelden 2.1 en 2.2 volgt dat geen van beide voorwaarden redundant is. Uit Voorbeeld 2.3 volgt dat voorwaarden niet voldoende zijn. In tegenstelling tot dit voorbeeld is het bij Hendrix' Voeders alleen mogelijk even bestellingen te plaatsen. Wij zullen in Stelling 2.1 aantonen dat in geval van even bestellingen voorwaarden (2.1) en (2.2) tevens voldoende voorwaarden zijn.

De routingheuristiek doorloopt de volgende fasen

1. Indeling van het vervoersgebied in naar keuze 1, 4, 8 of 16 sectoren. Middelpunt van deze sectorindeling is Zwolle (Zie Figuur 3)



Figuur 3 Sectorindeling met 8 sectoren.

2. Keuze van de sector, die samen met de beide aangrenzende sectoren het gebied vormt waarbinnen een toewijzing van klanten aan een rit wordt gerealiseerd. Op het scherm verschijnt informatie over deze klanten zoals bestelnummer, plaatsnaam, dagcode (geeft aan in welke rit van de dag deze klant gepland kan worden), bereikbaarheidscode (geeft aan welke bulkwagens deze klant kunnen bedienen), α_1 en q_1 , afstand tot Zwolle. De planner kan te allen tijde een overzicht van de gemaakte planning, nog beschikbare bulkwagens en nog te beleveren klanten opvragen.
3. Binnen het gekozen gebied krijgt de planner de mogelijkheid de bulkwagen en een of meerdere klanten vast te leggen. Kiest de planner geen klanten dan wordt de verste klant gekozen.
4. Uitgaande van de gemaakte keuzes maakt de (nog te beschrijven) insertie heuristiek een toewijzing van klanten aan een rit, uiteraard rekening houdend met de capaciteitsbeperkingen (2.1) en (2.2).
5. Indien de toewijzing niet goed geacht wordt door de planner kan hij deze wijzigingen of afwijzen in welk geval wij terug gaan naar 2.

Naarmate het aantal nog te beleveren klanten per sector afneemt zal de sectorindeling grover worden. Gebleken is dat deze sectorindeling door de planner als bijzonder plezierig is ervaren en hem het overzicht verstrekte dat de handmatige sortering hem niet kan leveren. Bewust gekozen is voor een interactieve planning om optimaal van de inzichten van de planner te profiteren. Zo is het soms mogelijk de bulkwagen te overladen of door

telefonisch contact met een klant ervoor te zorgen dat zijn bestelling wordt aangepast om de capaciteit van de bulkwagen optimaal te benutten. De insertieheuristiek, waarin overigens eenvoudige tijdvensters, d.w.z. tijdsintervallen waarbinnen klanten moeten worden bediend, zijn in te bouwen (Zie SOLOMON (1983)), werkt als volgt. Men begint met een gegeven rit $(i_1, i_2, \dots, i_m)$ met $i_1 = i_m = 0$ = depot (Zie Stap 2). Aan de hand van twee verschillende criteria worden klanten toegewezen aan een rit. In het eerste criterium c_1 wordt de beste plaats bepaald om een gegeven klant u in te plannen. Deze beste plaats wordt gegeven door

$$c_1(i_u, u, i_{u+1}) = \underset{p=1, \dots, m-1}{\text{minimum}} \{d(i_p, u) + a_1 d(u, i_{p+1}) - d(i_p, i_{p+1})\}, \quad (2.3)$$

waarbij $d(i, j)$ de afstand tussen klant i en klant j aangeeft en a_1 een nog te kiezen parameter. Nemen wij $a_1 = 1$ dan voegt criterium c_1 klant u in op de plaats die het minste omrijden vergt. De in te voegen klant u^0 wordt bepaald door het criterium c_2 . Deze beste klant wordt gegeven door

$$c_2(i_{u^0}, u^0, i_{u^0+1}) = \underset{u}{\text{maximum}} \{a_2 d(o, u) - c_1(i_u, u, i_{u+1})\}, \quad (2.4)$$

waarbij a_2 een te kiezen parameter aangeeft. Nemen wij $a_2 = 2$ dan kan $2d(o, u) - c_1(i_u, u, i_{u+1})$ gezien worden als de besparing die men krijgt door klant u in de bestaande rit in te plannen ten opzichte van klant u rechtstreeks vanuit het depot te bedienen. Omdat wij hoge prioriteit willen geven aan een volledige benutting van de capaciteit van een bulkwagen hebben wij 9 parameterkeuzes van a_1 en a_2 ingebouwd. Wordt de rit van een gegeven parameterkeuze afgewezen dan wordt een volgende parameterkeuze bekeken. De parameter a_1 heeft de waarden 1, 2 en 4, a_2 de waarden 0, 0,5 en 2.

Het programma is door de planner getest tegenover een handmatige planning op een probleeminstantie van 76 klanten en 13 beschikbare bulkwagens. Hierbij kwamen de volgende voordelen van de computerplanning ten opzichte van de handmatige planning naar voren: de snelheid, betere spreiding van de aankomsten op het depot, gelijkmatigere werklast van chauffeurs, minder bulkwagens werden ingezet (12 tegenover 13), alle toewijzingen waren toegelaten (de planner houdt alleen rekening met (2.1) en niet met (2.2)). Dit resultaat was

voor Hendrix' Voeders aanleiding de rittenberekening te gaan automatiseren. Tenslotte komen wij nog terug op het bewijs dat in geval van even bestellingen voorwaarden (2.1) en (2.2) voldoende zijn. Daar zowel bestellingen als inhoud van de compartimenten even zijn, mogen wij zonder verlies van algemeenheid aannemen dat de compartimenten een inhoud van 2 of 1 ton hebben en bestellingen alle mogelijke waarden kunnen aannemen.

Stelling 2.1. Gegeven een bulkwagen met compartimenten van 2 of 1 ton en gegeven een klantenbestand I die aan (2.1) en (2.2) voldoet. Dan is het fysiek mogelijk de lading van het klantenbestand I in te plannen.

Bewijs: Laat k het aantal compartimenten van 2 ton zijn en l het aantal compartimenten van 1 ton.

Definieer $I_1 = \{i \in I | q_i \text{ even, } q_i \leq 2k\}$, $I_2 = \{i \in I | q_i \text{ oneven, } q_i \leq 2k\}$, en $I_3 = \{i \in I | q_i > 2k\}$. Herschrijven van (2.1) geeft

$$\sum_{i \in I_1} q_i + \sum_{i \in I_2} q_i + \sum_{i \in I_3} q_i \leq 2k + l. \quad (2.5)$$

Wij onderscheiden de volgende gevallen

1. Stel $I_3 \neq \emptyset$ en $j \in I_3$.

Wij stellen voor om voor q_j k compartimenten van 2 ton en $q_j - 2k$ compartimenten van 1 ton te gebruiken. Alle overige bestellingen gaan in compartimenten van 1 ton. Het aantal benodigde compartimenten van 1 ton wordt gegeven door

$$\sum_{i \in I_1} q_i + \sum_{i \in I_2} q_i + \sum_{i \in I_3, i \neq j} q_i + (q_j - 2k) \quad (2.6)$$

Uit (2.5) volgt dat (2.6) kleiner of gelijk is aan l, m.a.w. er zijn voldoende compartimenten van 1 ton beschikbaar.

2a. $I_3 = \emptyset$ en $\sum_{i \in I_1} q_i \geq 2k$.

Wij stellen voor om met klanten in I_1 alle compartimenten van 2 ton vol te plannen en verder nog $\sum_{i \in I_1} q_i - 2k$ compartimenten van 1 ton. Alle overige bestellingen gaan in compartimenten van 1 ton. In totaal worden

$$\sum_{i \in I_1} q_i - 2k + \sum_{i \in I_2} q_i \quad (2.7)$$

compartimenten van 1 ton gebruikt. Volgens (2.5) is dit aantal kleiner of gelijk aan l.

$$2b. \quad I_3 = \emptyset, \quad \sum_{i \in I_1} q_i = 2k_1 < 2k \text{ en } \sum_{i \in I_2} (q_i - 1)/2 > k_2 = k - k_1.$$

Wij plannen alle klanten in I_1 in compartimenten van 2 ton. Indien wij alle lading van een klant in I_2 opsplitsen in een aantal ladingen van 2 ton en één restant van 1 ton geeft dit voor lading q_i $(q_i - 1)/2$ compartimenten van 2 ton. Uit het gegeven volgt dat alle resterende compartimenten van 2 ton op deze manier kunnen worden opgevuld. Er reteren nog $\sum_{i \in I_2} (q_i - 1)/2 - k_2$ ladingen van 2 ton en $|I_2|$ ladingen van 1 ton. Hiervoor zijn benodigd

$$2 \left\{ \sum_{i \in I_2} (q_i - 1)/2 - k_2 \right\} + |I_2| \quad (2.8)$$

compartimenten van 1 ton. Herschrijven van (2.8) geeft de benodigde ongelijkheid

$$\sum_{i \in I_2} q_i - 2k_2 \leq l. \quad (2.9)$$

Uit (2.5) volgt dat hieraan voldaan is daar $\sum_{i \in I_1} q_i = 2k_1$ en $k_1 + k_2 = k$.

$$2c. \quad I_3 = \emptyset, \quad \sum_{i \in I_1} q_i = 2k_1 < 2k \text{ en } \sum_{i \in I_2} (q_i - 1)/2 \leq k_2 = k - k_1.$$

Merk op dat $a_i = q_i/2$, $i \in I_1$ en $a_i = (q_i + 1)/2$, $i \in I_2$.

Herschrijven van (2.2) geeft

$$\sum_{i \in I_1} q_i/2 + \sum_{i \in I_2} (q_i + 1)/2 \leq k + l. \quad (2.10)$$

Wij plannen de klanten behorende tot I_1 in k_1 compartimenten van 2 ton. Van klant $i \in I_2$ wordt $q_i - 1$ in compartimenten van 2 ton gestopt en 1 in compartimenten van 1 ton. Het gebruikte aantal compartimenten van 2 ton is gelijk aan

$$\sum_{i \in I_1} q_i/2 + \sum_{i \in I_2} (q_i - 1)/2. \quad (2.11)$$

Het aantal resterende ladingen van 1 ton is $|I_2|$. Het hiervoor beschikbare aantal compartimenten is

$$\{k - \sum_{i \in I_1} q_i/2 - \sum_{i \in I_2} (q_i - 1)/2\} + l. \quad (2.12)$$

De eerste term is het aantal beschikbare 2 ton compartimenten die slechts voor de helft gevuld kunnen worden. Dat (2.12) groter of gelijk is aan $|I_2|$ volgt

uit (2.10) en uit $\sum_{i \in I_2} (q_i - 1)/2 + |I_2| = \sum_{i \in I_2} (q_i + 1)/2$.

In de volgende sectie beschrijven wij de gebruikte schattingsprocedures.

3. Het lineaire model

Het lineaire model heeft de volgende gedaante

$$y = X\beta + u \quad (3.1)$$

waarin y = (nx1)-vector van de te verklaren variabelen

X = (nxk)-matrix van verklarende variabelen

u = (nx1)-vector van storingen

β = (kx1)-vector van regressiecoëfficiënten.

De eerste kolom van de X -matrix bestaat alleen uit enen, hiermee aangevend dat het eerste element van de β -vector een constante term is. Onder de klassieke veronderstellingen van het lineaire model, namelijk $E(u) = 0$ en $\text{Var}(u) = \sigma^2 I$ met σ^2 een onbekende constante, kunnen de parameters met de kleinste kwadraten (KK) methode geschat worden. Minimalisatie van het gekwadraterde verschil tussen de waarde van de waargenomen en de waarde van de geschatte te verklaren variabele geeft de volgende geschatte waarde b voor de parametervector β .

$$b = (X^T X)^{-1} X^T y \quad (3.2)$$

De KK schatter in het klassieke lineaire model heeft aantrekkelijke statische eigenschappen. Binnen de klasse van lineaire zuivere schatters heeft de KK schatter de kleinste variantie. Indien heteroscedasticiteit optreedt, d.w.z. de variantie van de storingen is niet voor elke waarde van de variabele identiek (de variantie-covariantie matrix betreffende storingen wordt nu een diagonaalmatrix met op de diagonaal de variantie van elk van de storingen) heeft de KK schatter binnen de klasse van lineaire zuivere schatters niet meer de kleinste variantie. De gewogen kleinste kwadraten (WKK) schatter heeft, omdat zij aan storingen met een grotere variantie een kleiner gewicht toekent, men neemt namelijk als gewicht de reciproke van de variantie, een kleinere variantie dan de KK schatter. Probleem hierbij is wel dat de varianties onbekend zijn en dus geschat moeten worden. Wij bespreken hier de methode van Theil die een oplossing geeft voor dit probleem.

THEIL (1971) doet de volgende suggestie

$$\sigma^2 = \sigma(E(y_i))^2, \quad i=1, \dots, n \quad (3.2)$$

Echter $E(y_i)$ hangt af van de te schatten parameters in het model, hetgeen een directe toepassing van de WKK methode niet mogelijk maakt. Theil geeft met behulp van een iteratieve procedure een benaderende oplossing voor dit probleem. In de eerste iteratie wordt de parameter β met een normale KK regressie geschat. Dit resulteert met vergelijking (3.2) in een schatting van de varianties en dientengevolge in een WKK schatting van de parameters in het model. Met deze gewijzigde parameterschattingen kunnen echter andermaal schattingen van de variantie van de storingen gemaakt worden, die op hun beurt weer tot herziening van de parameters aanleiding geven. Dit proces wordt herhaald tot de b -vector een gewenste precisie heeft bereikt. Bij onze data werd gestopt indien de geschatte waarde van de b -vector in twee opeenvolgende iteraties niet meer dan 0,0001 toe- of afneemt. Voor de maanden november en december waren respectievelijk 6 en 7 iteraties nodig.

Referenties

- MADSEN, O.B.G. (1984), On the use of road network databases in the planning of locational and routing decisions, Reserach report 8, IMSOR, Lyngby, Denemarken.
- SOLOMON, M.M. (1983), Vehicle routing and scheduling with time window constraints: models and algorithms, Working paper Department of Decision Sciences, University of Pennsylvania.
- THEIL, H. (1971), Principles of econometrics, John Wiley & Sons.