
De combinatie van een Fuzzy C-Means clustering met de Stuyfzand-classificatie bij het uitvoeren van een hydrogeochemische systeemanalyse van de IJsselvallei

Steef van Baalen
Hannie de Ruiter
Pauline van Gaans

Bij de drinkwaterwinning in Deventer vormt optrekkend zout water een probleem. Om de problemen die met dit optrekken gepaard gaan te kunnen duiden en oplossingen aan te dragen, hebben we een hydrogeochemische systeemanalyse op regionale schaal uitgevoerd. Op basis van de resultaten van deze analyse kon bepaald worden of het noodzakelijk was om een meer gedetailleerde hydrologische modellering uit te voeren.

Het doel van de hydrogeochemische systeemanalyse is om, aan de hand van een kwalitatieve analyse, verschillende typen grondwater en hun oorsprong, te definiëren en te karakteriseren. De classificatie van watertypen is daarbij het belangrijkste thema. Deze classificatie is uitgevoerd met behulp van Fuzzy C-Means clustering en Non-Linear Mapping, gecombineerd met de classificatie van watertypen volgens Stuyfzand. Aan de hand van de resultaten van deze classificatie kon het regionale patroon van watertypen goed beschreven worden. Tevens was het mogelijk om de oorzaak van de verzilting vast te stellen.

Het gebruik van Fuzzy C-Means clustering in combinatie met de Stuyfzand-classificatie wordt hier ten eerste aanbevolen bij de analyse van hydrochemische data, omdat deze combinatie het mogelijk maakt om een goed en vooral makkelijk te interpreteren beeld te krijgen van data die over veel dimensies verspreid zijn, zoals in dit geval een chemische dataset.

Steef van Baalen was tijdens het uitvoeren van deze studie student aan de Universiteit Utrecht, vakgroep Fysische Geografie. Zijn huidige werkadres is: RIZA, Postbus 17, 8200 AA Lelystad, telefoon (0320) 298384, fax (0320) 298379, e-mail: s.vbaalen@riza.rws.minvenw.nl

Hannie de Ruiter is werkzaam bij de Waterleiding Maatschappij 'Overijssel', Postbus 10005, 8000 GA Zwolle, telefoon (038) 4276245, fax (038) 4276276

Pauline van Gaans is werkzaam bij de Universiteit Utrecht, Faculteit Ruimtelijke Wetenschappen, vakgroep Fysische Geografie, Postbus 80.115, 3508 TC Utrecht, telefoon (030) 2537453, fax (030) 2540604, e-mail: p.gaans@frw.ruu.nl

Inleiding

Een hydrologische systeem analyse, zoals beschreven door Engelen en Jones [5], is een uitstekende manier om inzicht te verkrijgen in een regionaal hydrologisch systeem. Het voornaamste doel van een hydrologische systeem analyse is het in kaart brengen van geneste grondwatersystemen van verschillende orde, waarbij elk afzonderlijk systeem een voedingsgebied verbindt met één of meerdere kwelgebieden. Bij het uitvoeren van een hydrologische systeemanalyse wordt gebruik gemaakt van verschillende, onderling gerelateerde kennisvelden, zoals geologie (formaties, sedimentologie), (geo)hydrologie (hydrologische eigenschappen, stroombaanpatronen), topografie (hoogtekaart, landgebruik), geochemie (bodem) en hydrochemie (grondwaterkwaliteit).

In dit artikel ligt de nadruk op de geohydrochemische interpretatie van watermonsters. Bij hydrochemische data kan het een probleem zijn om de eigenschappen van één monster weer te geven op een eenvoudige en toch zinvolle manier. Matthes [7] noemt 10 verschillende manieren om dit probleem op te lossen. Door Stuyfzand [10] wordt er echter op gewezen dat elke methode is ontwikkeld voor een specifiek doel en derhalve specifieke voordelen biedt, maar dat het bij alle methoden ontbreekt aan een meer 'holistische visie op waterkwaliteit'. Stuyfzand ontwikkelde daarom een methode om de belangrijkste factoren, die de regionale variaties in waterkwaliteit bepalen, te karteren en te analyseren. Bij deze methode speelt een classificatie-systeem van watertypen een belangrijke rol.

Watermonsters kunnen ook geclassificeerd worden door middel van statistische methoden. Als er a priori kennis aanwezig is over de processen die verantwoordelijk zijn voor de inhomogeniteit van de dataset, dan kunnen conventionele multivariate statistische methoden, zoals discriminant analyse, gebruikt worden [3, 12]. In het geval dat dergelijke a priori kennis ontbreekt, worden meestal conventionele clustermethoden gebruikt om de dataset te herschikken in homogene groepen die achteraf geïnterpreteerd kunnen worden [6, 12].

Zowel discriminant-analyse als conventionele clustertechnieken hebben het nadeel dat ze star zijn, d.w.z. dat ze een monster ondubbelzinnig aan één groep of cluster toewijzen [6]. De grenzen tussen verschillende watertypen zijn in het algemeen gradueel of enigszins vaag. Dit betekent dat er behoefte is aan een bepaalde mate van 'overlap' tussen de groepen om het hydrochemische patroon op een bevredigende wijze te beschrijven. Een methode die gebruikt maakt van 'fuzzy' of continue klassen ligt dus meer voor de hand, omdat in dat geval 'overlap' tussen klassen mogelijk is [4, 12].

In dit artikel zullen we een vergelijking maken tussen het classificatie systeem zoals voorgesteld door Stuyfzand [8, 9, 10] en de Fuzzy C-Means clustering methode in combinatie met Non-Linear Mapping zoals gebruikt door Vriend et al. [12] en Frapporti et al. [6]. Het ligt niet in onze bedoeling om één van deze methoden als 'beste koop' aan te duiden, maar we willen aantonen hoe ze elkaar kunnen aanvullen waardoor het resultaat beter is dan bij het gebruik van slechts één van deze methoden.

Na een korte introductie in de theoretische achtergronden van beide methoden zal aan de hand van een case study in het IJsseldal de toepasbaarheid van deze combinatie getoond worden. In de slotdiscussie zal nader ingegaan worden op enige kanttekeningen die gemaakt zijn bij het gebruik van statistische methoden in hydrochemisch onderzoek.

Theoretische Achtergronden

In dit artikel wordt kort de relevante informatie over beide methoden gegeven. Een compleet overzicht kan gevonden worden in de originele publicaties en de handboeken die genoemd zijn in de referenties.

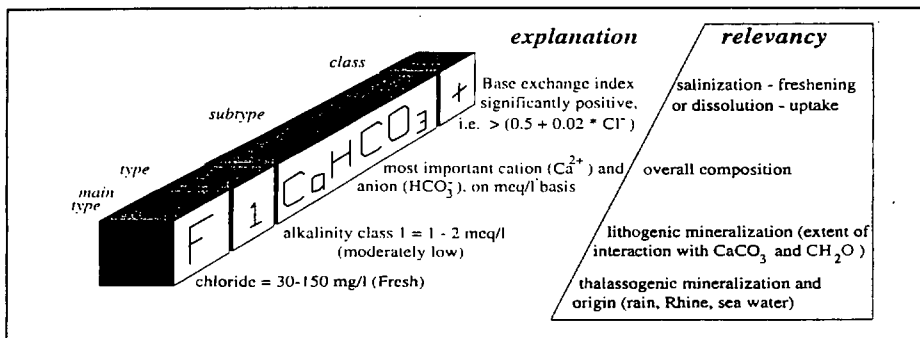
De classificatie van watertypen volgens Stuyfzand

Het classificatie-systeem combineert vier wezenlijke kenmerken van een watermonster in één codering (zie Figuur 1): de saliniteit (hoofdtype), alkaliteit (type), belangrijkste kation en anion (subtype) en de uitwisselingscoëfficiënt (klasse). Door het bepalen van deze vier kenmerken kan een watermonster op een hiërarchische manier geïnclassificeerd worden.

De belangrijkste voordelen van dit systeem zijn de mogelijkheid om het optreden van verschillende chemische processen vast te stellen en het feit dat een watermonster altijd op dezelfde manier geïnclassificeerd wordt vanwege de starre grenzen. Dit betekent echter ook dat het scheidend vermogen in principe altijd gelijk is, vanwege diezelfde starre grenzen. Stuyfzand erkent de starheid van zijn classificatie en stelt verschillende manieren voor om binnen het systeem associaties of differentiaties te maken. Associatie vindt plaats door het samenvoegen van twee of meer hoofdtypen of door het weglaten van klassen en/of subtypen. Differentiatie kan uitgevoerd worden door het verder onderscheid maken op bepaalde niveaus of door het toevoegen van nieuwe parameters. Deze flexibilisering blijft echter gebaseerd op vooronderstellingen, die soms remmend kunnen werken.

Fuzzy C-Means clustering gecombineerd met Non-Linear Mapping

Deze clustermethode is een krachtig gereedschap voor het scheiden van een dataset in een aantal homogene groepen, zonder enige *a priori* kennis. Hierbij wordt de totale dataset gesplitst in een aantal clusters, waarbij de verschillen tussen de clusters *in een orde-grootte relevant voor de dataset* zijn. Dat wil zeggen dat het verschil tussen onder- en bovengrens



Figuur 1: Het hydrochemische classificatiesysteem van Stuyfzand (9), weergegeven in een codering van 9 posities. Het voorbeeld is een 'zoet, matig laag alkalisch, calciumbicarbonaat watertype, met een positieve uitwisselingscoëfficiënt' (10).

per cluster gerelateerd is aan het verschil tussen het minimum en het maximum (range) van de hele dataset: hoe groter de range, hoe groter de clusters. De techniek staat een bepaalde mate van vaagheid toe in de beschrijving van het clustermodel. De overeenkomst van een monster met een clustergroep wordt aangeduid door een continue functie ('membership') tussen nul (volledig verschillend) en één (exact gelijk). De som van de memberships voor alle clusters van een specifiek monster is gelijk aan één. Een monster kan dus aan meer dan één cluster toegewezen worden. Dit biedt enige flexibiliteit bij de interpretatie van de dataset. Deze flexibiliteit is het belangrijkste verschil met zogenaamde 'harde' statistische technieken.

Een belangrijk punt bij het uitvoeren van deze clustering is het kiezen van het aantal clusters. In eerste instantie worden er verschillende runs gemaakt [11], waarbij elke keer een verschillend aantal clusters wordt onderscheiden. Het uiteindelijke aantal wordt gekozen op grond van verschillende criteria. In deze studie werden de regionale distributie, de hydrochemische interpreteerbaarheid en de 'structuur', zoals weergegeven in een 2-dimensionale Non-Linear Mapping plot, als criteria gebruikt [6].

Non-Linear Mapping maakt een 2- of 3-dimensionaal plaatje van een X-dimensionale puntenwolk, zodanig dat de afstanden tussen de punten onderling zo goed mogelijk weergegeven worden [12].

Hoewel de overlap die er bestaat tussen de verschillende clusters een voordeel is bij het interpreteren van de clusters, is voor een goede presentatie een zekere 'hardheid' van de clusters noodzakelijk. Deze 'defuzzificatie' wordt uitgevoerd aan de hand van het criterium dat een monster aan een zeker cluster wordt toegewezen als het hoogste membership van het betreffende monster minstens 1.67 maal zo hoog is als het op één na hoogste membership.

Vergelijking van de Stuyfzand-classificatie en Fuzzy C-Means clustering

Zoals met alle classificatiesystemen hebben beide methoden hun voor- en nadelen.

Het voordeel van de *Stuyfzand-classificatie* is dat het resulterend watertype van een bepaald monster altijd hetzelfde is vanwege de starre grenzen en dat de vaststelling van verschillende chemische processen mogelijk is. Als belangrijkste nadeel wordt hier de inflexibiliteit, ten gevolge van diezelfde starre grenzen, genoemd.

Het voordeel van *Fuzzy C-Means clustering* is dat deze methode het mogelijk maakt om zonder a priori kennis homogene clusters binnen de dataset te onderscheiden, waarbij de verschillen tussen de clusters in een *orde-grootte relevant voor de dataset* zijn. Dit betekent dat de klassen die gevonden worden met behulp van de clustering soms breder en soms smaller zijn dan de klassen die gevonden worden aan de hand van de *Stuyfzand-classificatie*. De belangrijkste nadelen van de clustermethode zijn het onvermogen om een cluster direct te interpreteren en de afhankelijkheid van een specifieke dataset: clusters van verschillende datasets zijn niet onderling vergelijkbaar.

Aan de hand van de volgende case study willen we aantonen dat beide methoden elkaar aanvullen, zodat de nadelen van elke afzonderlijke methode worden opgeheven, terwijl de voordelen behouden blijven.

Case study: Hydrogeochemische Systemanalyse van de IJsselvallei

Inleiding

De Waterleiding Maatschappij Overijssel (WMO) heeft twee puttenvelden in Deventer overgenomen van de gemeentelijke nutsbedrijven. Beide puttenvelden hebben al geruime tijd problemen met een toenemend chloride-gehalte van het gewonnen water. In de periode tussen 1970 en 1993 is het chloride-gehalte gemiddeld van 86 mg/l tot 117 mg/l toegenomen.

Bij het bepalen van een nieuwe, optimale inrichting van het puttenveld wilde de WMO meer weten over:

- de positie van het puttenveld in het regionale hydrologische systeem;
- de processen die verantwoordelijk zijn voor de toenemende verzilting van het puttenveld.

Wij hebben gekozen voor het uitvoeren van een regionale hydrogeochemische systemanalyse, waarbij de nadruk lag op geologie, (geo)hydrologie en hydrochemie. Geologische en (geo)hydrologische data zijn verzameld uit de literatuur. Met nadruk wordt hier gewezen op het feit dat binnen deze studie nog geen hydrologische modellering is uitgevoerd. De resultaten van deze studie hebben wel geleid tot een lijst met randvoorwaarden voor de hydrologische modellering van het gebied, die op dit moment uitgevoerd wordt.

Tot dusver is voornamelijk gebruik gemaakt van reeds bestaande data, waarmee een kwalitatieve systemanalyse is uitgevoerd, met de nadruk op de classificatie van watertypen.

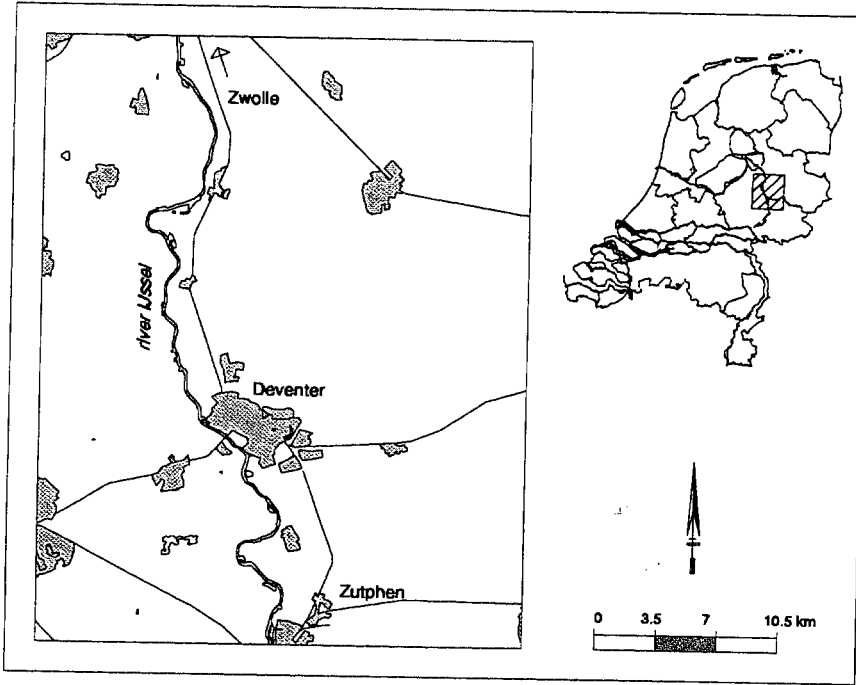
Gebiedsbeschrijving

Het studiegebied betreft de IJsselvallei tussen Zwolle en Zutphen, en meet 32 km van west naar oost en 40 km van noord naar zuid. De noord- en zuidgrens zijn zodanig gekozen dat ze op een flinke afstand van Deventer liggen; de west- en oostgrens zijn gekozen aan de hand van de waterscheidingen aan beide zijden van de IJsselvallei, respectievelijk gelegen op de Veluwe en de Sallandse Heuvelrug (zie figuur 2). Deze twee stuwwallen vormen de begrenzing van het regionale grondwatersysteem van de IJsselvallei.

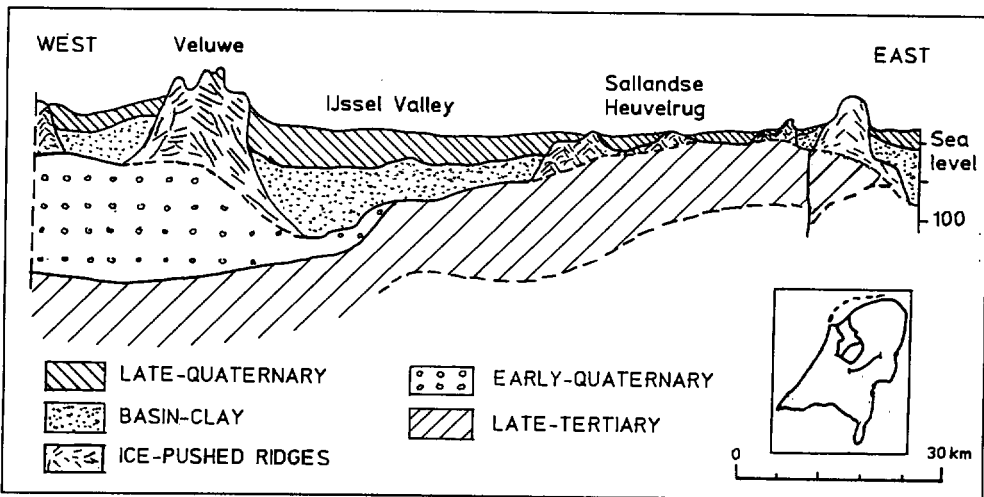
Geologie

De formaties die in het studiegebied van belang zijn, zijn allen in het laat-Tertiair (vanaf het Mioceen) of het Kwartair afgezet (figuur 3). De oudste, diepst gelegen formaties bevatten brak-zout water; de jongere formaties bevatten zoet water.

De IJsselvallei is gevormd tijdens de Saale-ijstijd, waarin het ijs in Nederland haar grootste verbreiding kende. In het studiegebied bewoog een gletsjertong zich zuidwaarts, waarbij de Tertiaire en Kwartaire sedimenten rondom opgestuwd werden. Op deze manier werd een 120 m diep glaciaal bekken gevormd, geflankeerd door stuwwallen met een hoogte tot 100 m. Tijdens het afsmelten van het landijs ontstond in dat bekken, aan de zuidzijde van het ijs, een meer. Hierin sedimenteerde een dik pakket fluvioglaciaal bekkenklei, de huidige Formatie van Drente. Deze bekkenklei werd tijdens het laat-Kwartair begraven onder dikke pakketten fluviaatiele en eolische afzettingen (Formaties van Kreftenheye, Asten en Twente).



Figuur 2: Ligging van het studiegebied



Figuur 3: Geologische west-oost doorsnede van de IJsselvallei

Geohydrologie

De twee stuwwallen functioneren als infiltratiegebieden, terwijl de bekkenklei de scheiding vormt tussen ondiepe en diepe watervoerende pakketten in de IJsselvallei. Er vindt vrijwel geen transport plaats tussen het ondiepe grondwater, dat onder antropogene invloed staat, en het zeer oude (tot 20.000 jaar) diepe grondwater. Het zoet-zout grensvlak (150 mg/l Cl⁻) bevindt zich onder Deventer op ongeveer 130 m –NAP.

In deze studie heeft de nadruk gelegen op het diepe watervoerende pakket, aangezien de onttrekking plaatsvindt in dit pakket en omdat een systeemanalyse van het ondiepe watervoerende pakket reeds is uitgevoerd.

Werkwijze

We hebben, zoals reeds in de inleiding ter sprake is gekomen, voornamelijk gebruik gemaakt van reeds bestaande data. Informatie over de geologie en de (geo)hydrologie is verkregen uit de literatuur. De noodzakelijke hydrochemische gegevens zijn ontleend aan het On Line Grondwater Archief (OLGA), dat onderhouden wordt door TNO-GG, en interne WMO-archieven.

Allereerst is een inventarisatie gemaakt van alle in archieven beschikbare analyses in het studiegebied. Vervolgens is een selectie gemaakt op basis van de volgende criteria:

- de gegevens mochten niet van voor 1970 dateren;
- de analyses moesten in ieder geval Na⁺, Cl⁻, Ca²⁺, HCO₃⁻, Mg²⁺, SO₄²⁻– Fe²⁺, NO₃²⁻, pH en EC bevatten;
- er is alleen gebruik gemaakt van gegevens van peilputten met tenminste één filter dieper dan 50 m –NAP (= de minimale diepte van de onderkant van de bekkenklei).

Het resultaat van deze selectie was een dataset van 189 analyses. Hieraan zijn 17 nieuwe analyses toegevoegd, zodat we in totaal de beschikking hadden over 206 analyses, verdeeld over 33 locaties. Drie monsters zijn niet gebruikt in de analyse, omdat er aanwijzingen waren dat ze niet betrouwbaar waren.

Het watertype volgens Stuyfzand is bepaald met behulp van het programma ChemProc v. 4.0 [1]. De clustering is uitgevoerd met behulp van het programma FNX [11].

Resultaten en discussie

Normaliteit van de gebruikte variabelen is voor het uitvoeren van de clustering geen voorwaarde. Het resultaat is echter wel beter als er gebruik gemaakt wordt van normaal verdeelde variabelen. Om normaliteit te bereiken bleek het noodzakelijk om op alle variabelen, exclusief pH, een logtransformatie uit te voeren. De EC is niet gebruikt bij het uitvoeren van de clustering, omdat dit een somparameter is van een aantal reeds gebruikte variabelen.

De clustering kan in principe uitgevoerd worden met een onbeperkt aantal variabelen, maar ten gevolge van onderlinge correlatie is het resultaat niet altijd beter als er gebruik gemaakt wordt van veel variabelen. Het weglaten van de variabelen met het laagste scheidend vermogen is zinvol omdat dan in feite een deel van de 'ruis' wordt weggelaten; de clusters worden hierdoor beter definieerbaar.

Uiteindelijk is gekozen voor een clustering op basis van Na^+ , Cl^- , Ca^{2+} , HCO_3^- , Mg^{2+} , pH en K^+ . Op basis van de criteria, zoals genoemd in de theoretische achtergronden, is gekozen voor een clustermodel met 7 verschillende clusters. In tabel 1 worden de cluster-centra van de 'fuzzy' clusters en de geometrische gemiddelden van de 'verharde' clusters weergegeven. De clusters zijn gerangschikt naar toenemend chloride-gehalte, d.w.z. dat cluster 1 het laagste chloride-gehalte heeft en cluster 7 het hoogste. De Non-Linear Mapping plot van Figuur 4 toont de samenhang tussen de gevormde clusters. De monsters die aangeduid worden met het cijfer '10' zijn intermediair; ze konden niet 'hard' gemaakt worden vanwege hun lage membership voor alle clusters.

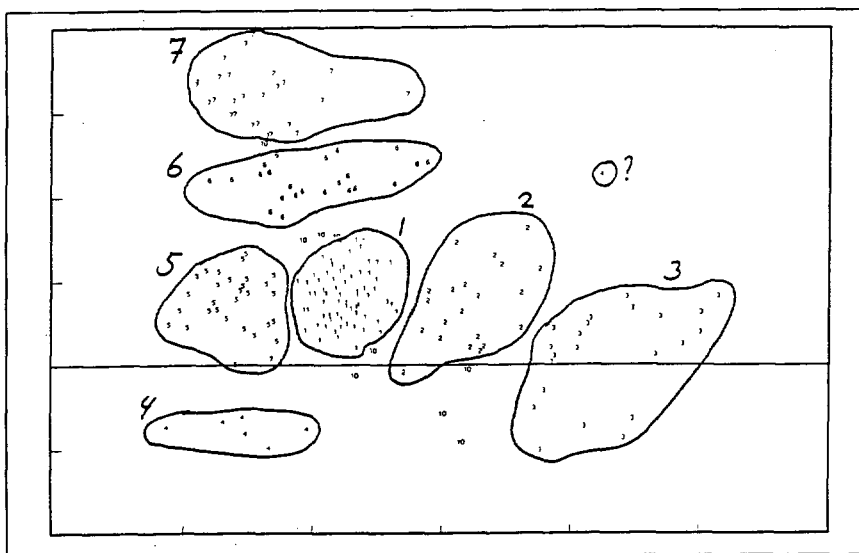
Tabel 1: Cluster-centra (CC) van de 'fuzzy' clusters en geometrische gemiddelden (GM) van de 'verharde' clusters voor het gekozen cluster model. Negen monsters konden niet 'hard' gemaakt worden. De verschillen tussen de CC's en de GM's zijn het resultaat van het effect van uitbijters (monsters met een laag membership). Bij het berekenen van de CC wordt het membership gebruikt als weegfactor, zodat het effect van uitbijters op de CC veel minder is dan op de GM. Alle variabelen hebben als eenheid mg/l, behalve pH. Het aantal monsters in de tabel is gelijk aan 194 (206 minus 3 onbetrouwbare monsters en 9 intermediaire monsters).

Variabele	Cluster 1 (n = 66)		Cluster 2 (n = 21)		Cluster 3 (n = 21)		Cluster 4 (n = 7)	
	CC	GM	CC	GM	CC	GM	CC	GM
Na^+	21,5	23,5	14,9	17,9	11,8	30,9	20,5	52,7
Cl^-	13,5	16,3	13,6	13,7	16,3	46,9	25,4	32,1
Ca^{2+}	36,3	36,8	15,4	15,4	9,2	10,8	51,0	54,0
HCO_3^-	169,8	172,9	66,7	74,4	14,8	14,7	210,6	269,7
Mg^{2+}	4,4	4,6	2,7	2,8	1,4	3,0	7,7	8,4
pH	7,7	7,7	7,1	7,3	6,4	6,3	7,5	7,5
K^+	2,6	3,1	1,5	1,6	1,1	1,5	0,0	0,0

Variabele	Cluster 5 (n = 30)		Cluster 6 (n = 23)		Cluster 7 (n = 26)	
	CC	GM	CC	GM	CC	GM
Na^+	23,7	24,4	202,4	231,4	1152,0	1290,0
Cl^-	32,6	35,3	257,0	304,0	1822,9	2068,0
Ca^{2+}	87,8	92,5	28,0	30,7	37,9	50,4
HCO_3^-	262,6	277,6	258,4	276,6	391,0	397,0
Mg^{2+}	8,9	9,3	11,6	12,3	52,0	58,0
pH	7,3	7,3	7,9	7,9	8,0	8,0
K^+	1,9	2,3	16,4	17,6	50,3	53,0

Bij de typering en interpretatie van de verschillende clusters is gebruik gemaakt van de volgende methoden:

- statistische typering per variabele per cluster (geometrisch gemiddelde, mediaan, kwartielen, minimum en maximum);
- Stiff-diagrammen van de monsters met de hoogste memberships per cluster;
- de Stuyfzand-classificatie van elk monster;
- de ruimtelijke spreiding van elk cluster.



Figuur 4: Door Non-Linear Mapping berekende afstanden tussen de monsterpunten, weergegeven met behulp van de met Fuzzy C-Means clustering gevonden indeling. Non-Linear Mapping geeft punten in een X-dimensionale ruimte weer in een 2D-plaatje. De waarden op de x- en y-as hebben hierbij geen fysische betekenis; van belang is de onderlinge positie van de verschillende punten.

Omdat het buiten het kader van dit artikel valt om het watertype van elk cluster uitgebreid te analyseren, zal hier enkel de uiteindelijke interpretatie gegeven worden. Deze interpretatie is gebaseerd op de hiervoor genoemde methoden, samen met informatie uit de literatuur. Naast het interpreteren van elk cluster, is ook de ruimtelijke spreiding van elk cluster bepaald. Hierbij is gebruik gemaakt van horizontale en verticale dwarsdoorsneden. Bij het lokaliseren van grenzen tussen de clusters zijn de memberships van alle monsters gebruikt, samen met geologische gegevens en hydrologische (model)resultaten uit de literatuur.

In tabel 2 wordt van elk cluster een korte omschrijving gegeven van de interpretatie per watertype, samen met het overheersende watertype volgens de Stuyfzand-classificatie. Deze tabel toont dat in deze studie de combinatie van de twee methoden resulteert in een associatie van de watertypen volgens de Stuyfzand-classificatie. Deze associatie is dus uitgevoerd zonder enige *à priori* kennis.

Tabel 2: Interpretatie van de verschillende clusters, samen met een korte omschrijving en het overheersende watertype volgens de Stuyfzand-classificatie. Merk op dat de watertypen volgens de Stuyfzand-classificatie allemaal uniek zijn voor elk Fuzzy C-Means (FCM) cluster, behalve g2-CaHCO₃+.

(De afkortingen zijn overgenomen uit de originele, Engelstalige publikatie.)

FCM-cluster	Interpretatie	Omschrijving	Stuyfzand-classificatie
1	Oud Infiltraat (regen)water (OI)	Water in diepe watervoerende pakket, beïnvloed door processen zoals reductie en kalkoplossing	g2-CaHCO ₃ +
2	Jong Infiltraat (regen)water (YI)	Weinig beïnvloed regenwater in infiltratie gebieden, gesitueerd boven type 1	g1-CaHCO ₃ + g1-NaHCO ₃ +
3,4,5	Ondiep, Antropogeen water (SM)	Water boven de bekkenklei, sterk beïnvloed door geïnfiltrteerd rivierwater en landbouwkundige activiteiten	g*-CaMClO, g*-CaMISO (3) F2-CaHCO ₃ o (4) g2-CaHCO ₃ + (4) F3-CaHCO ₃ o/+ (5) g3-CaHCO ₃ +/o (5)
6	Relict Brak water (RB)	Grenzone tussen diep zoet water en marien zout water	f3-NaCl+ B3-NaCl+ /o
7	Relict Zout water (RS)	Marien zout water, vastgelegd in Tertiaire mariene sedimenten	b3-NaClO b4-NaClO

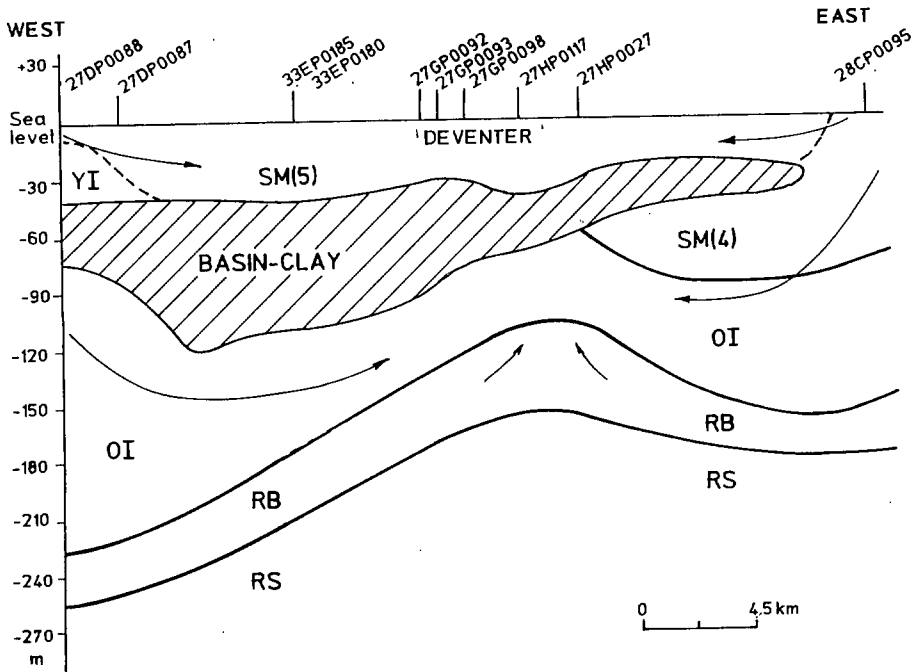
Cluster 1 omvat het watertype g2-CaHCO₃+: een oligohalien (regenwater-achtig), matig alkalisch, calciumbicarbonaat water met een lage positieve uitwisselingscoëfficiënt. Het gaat hier om regenwater dat minstens 2500 jaar geleden is geïnfiltrteerd [2]. Ten gevolge van een complex van bodemvormende processen (reductie van ijzer en sulfaat, verwerking van silicaten, oplossing van calciumbicarbonaat) is de samenstelling van dit water veranderd, maar het is nog steeds herkenbaar als Oud Infiltraat (regen)water (OI).

Cluster 2 omvat de watertypen g1-CaHCO₃+ and g2-NaHCO₃+: oligohalien, matig laag alkalisch, calcium/natriumbicarbonaat water met een lage positieve uitwisselingscoëfficiënt. Het gaat hier om regenwater dat is geïnfiltrteerd tussen nu en 2500 jaar geleden [2]. Vanwege de zeer lage concentraties macro-ionen wordt dit water geïnterpreteerd als Jong Infiltraat (regen)water (YI).

Clusters 3 tot 5 omvatten een mengeling van verschillende watertypen, variërend van zeer kalkarm tot zeer kalkrijk water, maar allemaal met hoge concentraties nitraat en/of sulfaat. De maximale diepte waarop deze typen gevonden worden is 70 m -NAP. Antropogene invloeden ten gevolge van geïnfiltrteerd rivierwater en landbouwkundige activiteiten domineren deze clusters. Ze worden daarom allemaal geïnterpreteerd als Ondiep, Antropogeen water (SM).

Cluster 6 omvat de watertypen f3-NaCl+ and B3-NaCl+ /o: zoet-brak tot brak, matig hoog alkalisch, natriumchloride water met een variabele uitwisselingscoëfficiënt. Onder Deventer duiden de laagste uitwisselingscoëfficiënten op verzilting. Net buiten Deventer duiden de hoge uitwisselingscoëfficiënten van dit watertype juist op een verzoeting. Dit patroon duidt op een relatief snel, lokaal optrekken van zout water onder Deventer. Deze monsters worden geïnterpreteerd als Relict Brak water (RB).

Cluster 7 omvat de watertypen b3/b4-NaClO: een brak-zout, matig hoog tot hoog alkalisch, natriumchloride water, met een zeer lage uitwisselingscoëfficiënt. Dit water wordt



Figuur 5: Verticale dwarsdoorsnede van de IJsselvallei, ter hoogte van Deventer. Zie tabel 2 voor een verklaring van de afkortingen.

geïnterpreteerd als Relict Zout water (RS) zonder dat er sprake is van verzoeting of verzilting.

Figuur 5 toont een voorbeeld van een west-oost georiënteerde verticale dwarsdoorsnede van de IJsselvallei. In deze figuur zijn de verschillende watertypen makkelijk te herkennen. Tevens is zichtbaar dat de verzilting veroorzaakt wordt door optrekkend relict zout water.

Conclusies

Bij beschouwing van tabel 2 en figuur 5 kunnen de volgende conclusies getrokken worden:

- door gebruik te maken van hydrochemische data samen met geologische en (geo)hydrologische data uit de literatuur is het mogelijk gebleken om een goed gestructureerd plaatje te maken van de verschillende soorten water die aanwezig zijn in de diepe watervoerende pakketten van de IJsselvallei;
- de oorzaak en omvang van de verzilting in de puttenvelden van Deventer zijn duidelijk zichtbaar in de horizontale en verticale doorsneden;
- tabel 2 laat zien dat een combinatie van een Fuzzy C-Means clustering met de Stuyfzand-classificatie goede resultaten geeft. De homogene clusters die gevonden zijn in de clustering hebben onderlinge verschillen in een orde-grootte die relevant is voor de betreffende dataset. De watertypen volgens de Stuyfzand-classificatie maken het mogelijk

om het voorkomen van verschillende chemische processen vast te stellen, en om de gevonden clusters te vergelijken met de resultaten van andere studies.

Discussie en algemene conclusies

In dit artikel hebben we geprobeerd om de toepasbaarheid van de combinatie van Fuzzy C-Means clustering met de Stuyfzand-classificatie aan te tonen. Het resultaat is een eenvoudig te interpreteren hydrochemische classificatie op het juiste schaalniveau, zoals is getoond in de case study.

Stuyfzand noemt een aantal mogelijke problemen die zich kunnen voordoen bij de toepassing van een cluster analyse voor de classificatie van watermonsters. Hij benadrukt dat:

- 1 de grenzen van de clusters niet goed gedefinieerd zijn en dat aangrenzende kaarten onderling niet goed vergelijkbaar zijn;
- 2 (log)normaliteit en statistische onafhankelijkheid verondersteld moeten worden;
- 3 uitbijters uitgesloten moeten worden;
- 4 een discriminant analyse nodig is om het relatieve belang van alle hydrochemische variabelen te bepalen.

Er zit natuurlijk een kern van waarheid in deze punten, waardoor het werken met Fuzzy C-Means clustering niet altijd even zinvol is, met name als er reeds veel bekend is over het betreffende gebied. Maar volgens ons heft de combinatie van de clustering met de classificatie een deel van deze problemen op, en is daarom zeer bruikbaar in een nieuw gebied met onbekende processen.

ad 1 Enige vaagheid van de grenzen tussen verschillende clusters is een goede weergave van de geleidelijke overgangen die in de werkelijkheid aanwezig zijn, en wordt daarom zeer op prijs gesteld door geografen. Aangrenzende kaarten zijn wel onderling vergelijkbaar als de watertypen volgens de Stuyfzand-classificatie worden gebruikt.

ad 2 Multivariate (log)normaliteit en statistische onafhankelijkheid zijn geen vereiste voor de toepassing van een cluster analyse [12], maar het is niet onverstandig om deze aspecten in gedachten te houden bij de keuze van de variabelen waarmee de clustering uitgevoerd gaat worden.

ad 3 Uitbijters dragen in gelijke mate bij aan alle clusters en vormen geen probleem bij het uitvoeren van de clustering. De schaal van een Non-Linear Mapping plot wordt echter wel sterk beïnvloed door uitbijters [12], en daarom blijft het nodig om uitbijters kritisch te bekijken.

ad 4 Als gebruik gemaakt wordt van watertypen volgens de Stuyfzand-classificatie is geen discriminant analyse nodig om de clusters te interpreteren.

Onze conclusie luidt daarom dat het grote voordeel van een statistische techniek zoals Fuzzy C-Means clustering gelegen is in de mogelijkheid om een dataset te verdelen in homogene groepen op het juiste schaalniveau. De combinatie met de Stuyfzand-classificatie maakt het mogelijk om de clusters goed te interpreteren.

Tot slot

Voor meer informatie over het programma FNX kunnen geïnteresseerden zich wenden tot Pauline van Gaans, Universiteit Utrecht.

Met dank aan Jan Borst (WMO) en Bas v.d. Pas (RIZA) voor hun hulp bij het vervaardigen van de figuren 2, 3 en 5.

Referenties

- [1] **Biesheuvel, A. (1991)** ChemProc v. 4.0; Vrije Universiteit Amsterdam, The Netherlands.
- [2] **Broks (1992)** Geohydrologische onderzoek winbare hoeveelheid 'Salland-diep'; projectno. 11.250, 4 delen.
- [3] **Davis, J.C. (1986)** Statistics and data analysis in geology; John Wiley & Sons, New York, 2nd ed.
- [4] **De Gruijter, J.J. & McBratney, A.B. (1988)** A modified fuzzy k-means method for predictive classification; Classification and Related Methods of Data Analysis, H.H.Bock (ed.), Elsevier Science Publishers, The Netherlands, pp. 97-104.
- [5] **Engelen, G.B. & Jones, G.P. (1986)** Developments in the analysis of ground water flow systems; IAHS Publ. 163, 365 pp.
- [6] **Frapporti, G., Vriend, S.P. & Gaans, P.F.M. van (1993)** Hydrogeochemistry of the shallow dutch groundwater: interpretation of the National Groundwater Quality Monitoring Network; Water Resources Research, Vol. 29, No. 9, pp. 2993-3004.
- [7] **Matthess, G. (1982)** The properties of groundwater; John Wiley & Sons, New York, 406 pp.
- [8] **Stuyfzand, P.J. (1986)** Een nieuwe hydrochemische classificatie van watertypen, met Nederlandse voorbeelden van toepassing; H₂O, Vol. 19, no. 23, pp. 562-568.
- [9] **Stuyfzand, P.J. (1989)** A new hydrochemical classification of watertypes; IAHS Publ. 182, pp. 89-98.
- [10] **Stuyfzand, P.J. (1993)** Hydrochemistry and Hydrology of the Coastal Dune area of the Western Netherlands; KIWA-report I11, Nieuwegein, The Netherlands, 366 p., ook gepubliceerd als these Vrije Universiteit, Amsterdam, The Netherlands.
- [11] **Gaans, P.F.M. van en Vriend, S.P. (1995)** FNX, a program for Fuzzy C-Means clustering and Non-Linear Mapping; Universiteit Utrecht, The Netherlands.
- [12] **Vriend, S.P., Gaans, P.F.M. van, Middelburg, J. en De Nijs, A. (1988)** The application of fuzzy c-means cluster analysis and non-linear mapping to geochemical datasets: examples from Portugal; Applied Geochemistry, Vol. 3, pp. 213-224.

Dit artikel is eerder (in het Engels) gepubliceerd in:

14th Salt Water Intrusion Meeting (SWIM 96), Rapporten och meddelanden nr. 87, Geological Survey of Sweden (1996)