



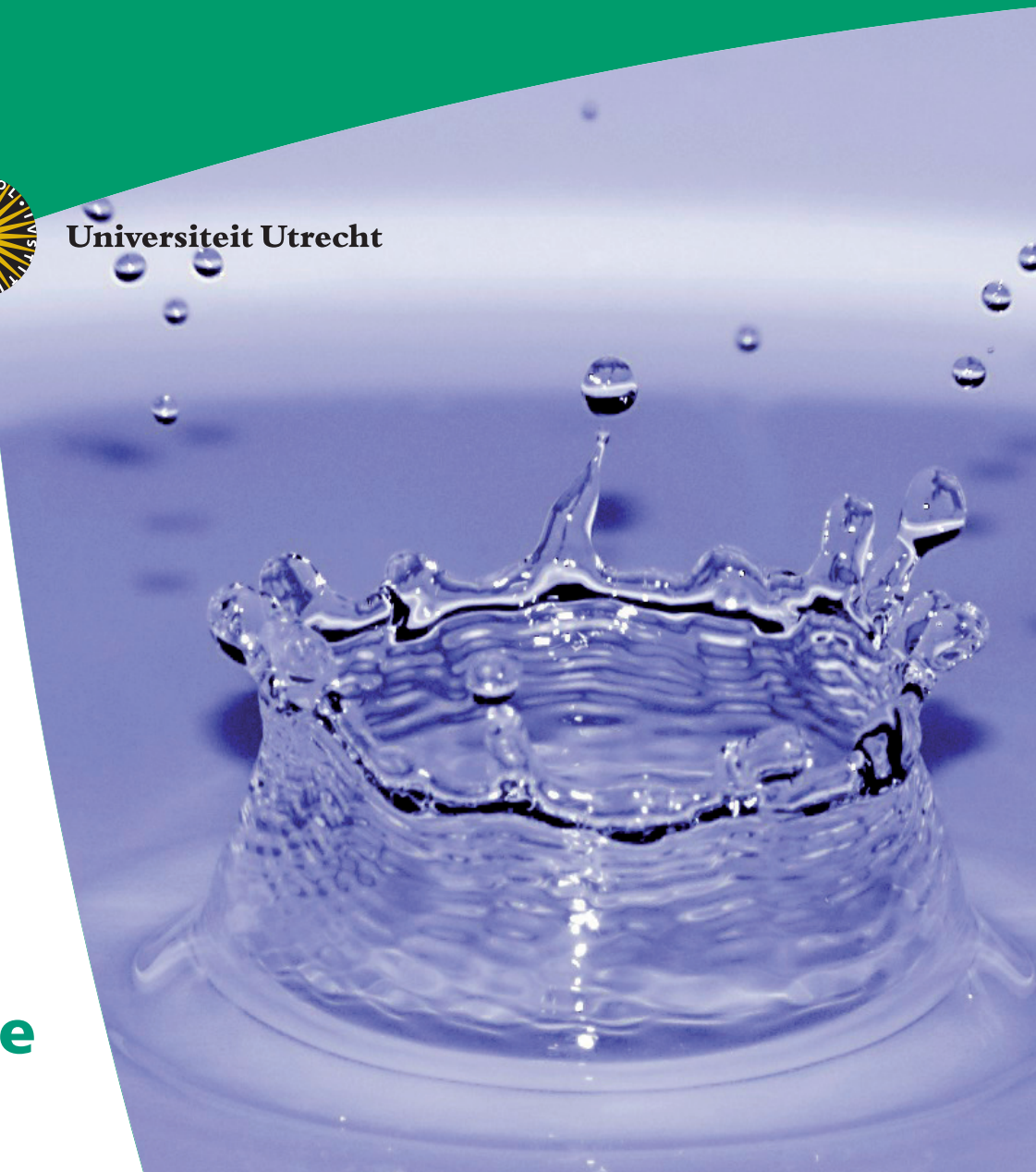
Zuiver drinkwater?!

Bronnenboek



Universiteit Utrecht

**Junior College
Utrecht**



Zuiver drinkwater?!

Procesmodellering vlokbehandeling

Bronnenboek

Versie 1.0



Universiteit Utrecht



FIsme

**Junior College
Utrecht**



Op de omslag

Foto door Jørn Nøding, Noorwegen

De module *Zuiver Drinkwater?!* is bestemd voor de lessen Natuur, Leven en Technologie. De module is op 4 juni 2009 gecertificeerd door de Stuurgroep NLT voor gebruik op het vwo in domein E (biofysica, biochemie en bioinformatica). Het certificeringsnummer is X224-041-VH.

De originele gecertificeerde module is in pdf-formaat downloadbaar via <http://www.betavak-nlt.nl>

De module is gemaakt in opdracht van het Freudenthal Instituut voor Didactiek van Wiskunde en Natuurwetenschappen en het Junior College Utrecht

Deze module is ontwikkeld door Gjalt Prins (Freudenthal Instituut) en Paul Drijvers (Freudenthal Instituut) in samenwerking met:

- Junior College Utrecht
 - Frans Teeuw
 - Rob Wiedemeijer
- Freudenthal Instituut/Sectie Chemiedidactiek
 - Astrid Bulte
 - Albert Pilot
 - Atie Wigmans
- Docentennetwerk Chemiedidactiek
 - Rens Bijma (Griffland College, Soest)
 - Jan de Vries (Oosterlicht College, Nieuwegein)
 - Jeannine Acampo (Bonifatius College, Utrecht)
 - Sanne Spijker (Lek & Linge, Culemborg)
 - Wijnand Rietman (Het Streek, Ede)
- Vormgeving
 - Krijn Kieviet
 - Communicatie en Vormgeving Faculteit Bètawetenschappen

© 2009. Versie 1.0

Voor deze module geldt een



Creative Commons Naamsvermelding-Niet-commercieel-Gelijk delen 3.0 Nederland Licentie

<http://creativecommons.org/licenses/by-nc-sa/3.0/nl>

Het auteursrecht op de module berust bij de Universiteit Utrecht, het Freudenthal Instituut voor Didactiek van Wiskunde en Natuurwetenschappen en het Junior College Utrecht.

Aangepaste versies van deze module mogen alleen verspreid worden, indien in de module vermeld wordt dat het een aangepaste versie betreft, onder vermelding van de naam van de auteur van de wijzigingen.

De auteurs hebben bij de ontwikkeling van dit materiaal gebruik gemaakt van materiaal van derden. Waar dat is gebeurd, is zo veel mogelijk de bron vermeld. De module is met zorg samengesteld. De Universiteit Utrecht aanvaardt geen enkele aansprakelijkheid voor enige schade voortkomend uit (het gebruik van) deze module.

Materialen die leerlingen nodig hebben bij deze module zijn beschikbaar via het vaklokaal NLT:

<http://www.digischool.nl/nlt>

Voorwoord

Dit bronnenboek hoort bij de lessenserie 'Zuiver drinkwater?! Procesmodellering vlokbehandeling'. In dit bronnenboek staat achtergrondinformatie in de vorm van documenten, artikelen en handleidingen.

Inhoud

Voorwoord	1
Inhoud	2
Bron 1. Definitiestudie Zuiveringsmodel Drinkwaterbereiding	3
Bron 2. Factsheet actieve koolfiltratie	6
Bron 3. Uitvoering van Vlokbehandeling	11
Bron 4. Coagulatie & Flocculatie	14
Bron 5. Procesmodellering	20
Bron 6. Het sommatieteken	22
Bron 7. Meetniveau en meetnauwkeurigheid	24
Bron 8. Centrum en spreiding in beeld	27
Bron 9. Samenhang en correlatie	30
Bron 10. Lineaire regressie	33
Bron 11. Machtsregressie	36
Bron 12. De derde wet van Kepler als voorbeeld van machtsregressie	38
Bron 13. Andere niet-lineaire regressiemodellen	40
Bron 14. Meervoudige regressie	42
Bron 15. Standaarddeviatie van populatie en steekproef	45
Bron 16. Correlatie en regressie met Excel	46
Bron 17. Correlatie en regressie met de TI83/84	54
Bron 18. Correlatie en regressie met de Casio 9850/9860	56

Bron 1. Definitiestudie Zuiveringsmodel Drinkwaterbereiding

Aangepaste versie van RIVM document nr. 734301011 'Definitiestudie Zuiveringsmodel Drinkwater', F.J. Kragt, F.W. van Gaalen, L.C. Rietveld, J.F. Schrijven, augustus 1996.

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU

Definitiestudie Zuiveringsmodel Drinkwaterbereiding

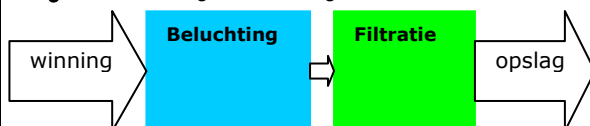
Beschrijving van het probleem

Drinkwater moet aan kwaliteitseisen voldoen vastgelegd in het Waterleidingbesluit (WLB). Om aan deze eisen te voldoen wordt het oppervlakte- en grondwater gezuiverd in een aantal opeenvolgende stappen. Hieronder zijn de belangrijkste zuiveringsstappen weergegeven.

Figuur 1. Zuiveringsketen voor oppervlaktewater



Figuur 2. Zuiveringsketen voor grondwater



Ondanks de doorlopen zuiveringsketen vinden er soms overschrijdingen van de norm plaats. Er zijn aanwijzingen dat dit (deels) opgelost kan worden door de invloed van de procesvariabelen te onderzoeken. Als de invloed bekend is, dan kan de zuivering beter afgestemd worden op de variërende kwaliteit van grond- en oppervlaktewater. Tevens kan vooraf berekend worden of het ongezuiverde water voldoende geschoond kan worden om te voldoen aan de eisen voor drinkwater.

Doel

Het doel van dit project is per zuiveringsstap en verontreiniging het verband te onderzoeken tussen de uitgaande hoeveelheid verontreiniging en procesvariabelen. Het verband wordt beschreven in een wiskundig model. Met het wiskundige model kan de kwaliteit van drinkwater worden voorspeld afhankelijk van (I) de kwaliteit van het grond- en

oppervlaktewater en (II) het zuiveringstraject.

Aanpak

In iedere zuiveringsstap wordt een beperkt aantal verontreinigingen verwijderd. In de volgende tabel staat een overzicht.

Zuiveringsstap	Verontreinigingen
Bekken(s)	Grotere zwevende deeltjes, algen, bacteriën, pesticiden, metaaloxiden (o.a. oxiden van ijzer, mangaan en nikkel).
Vlokbehandeling	Troebeling, metaaloxiden (o.a. oxiden van ijzer, mangaan en nikkel).
Snelle + langzame zandfiltratie	Bacteriën, pesticiden, nitraat, nitriet, ammonium, waterstofcarbonaat.
Actieve koolfiltratie	Organisch materiaal (o.a. broomdichloormethaan, trichloormethaan, dichloormethaan) en pesticiden.
Desinfectie	Bacteriën en micro-organismen.

De kwaliteit van het drinkwater wordt mede bepaald door de uitvoering van de diverse zuiveringstappen. Hieronder staat een beknopt overzicht van procesvariabelen per zuiveringsstap.

Zuiveringsstap	Procesvariabelen
Spaarbekken	o.a. diepte, beluchting, verblijftijd
Vlokbehandeling	o.a. hoeveelheid vlokmiddel, begintroebeling
Snelle + langzame zandfiltratie	o.a. verblijftijd, filtratielengte, textuur, structuur
Actieve koolfiltratie	o.a. contacttijd, adsorptieoppervlak, type kool, beginconcentratie
Desinfectie	o.a. dosis desinfectiemiddelen, UV licht

Wiskundig model

Per zuiveringsstap wordt de invloed van de procesvariabelen vastgesteld door (laboratorium) experimenten uit te voeren. Vervolgens wordt het verband beschreven m.b.v. een wiskundig model. De algemene vorm van het wiskundige model luidt:

$$C_{uit} = f(C_{in}, pv_1, pv_2, \dots, pv_n), \text{ waarbij}$$

C_{uit} = uitgangconcentratie van de verontreiniging

C_{in} = ingangconcentratie van de verontreiniging

pv_1, pv_2, \dots, pv_n = procesvariabelen 1, 2, ..., n.

Evaluatie

De wiskundige modellen worden beoordeeld op de passendheid, betrouwbaarheid en geldigheid. Daarnaast wordt de voorspellende waarde beoordeeld door bestaande bedrijfssituaties door te rekenen. Tot slot wordt geëvalueerd in hoeverre de ontwikkelde wiskundige modellen bijdragen aan het doel van dit project.

Reflectie en aanbevelingen

Na ieder (deel)onderzoek wordt gereflecteerd op de belangrijkste resultaten en werkwijze. Tevens wordt advies gegeven over eventueel vervolgonderzoek aan dezelfde en/of andere zuiveringsstappen. De rapportage vindt plaats in de vorm van factsheets.

Bron 2. Factsheet actieve koolfiltratie

Aangepaste versie van artikelen en RIVM documenten.

B. Pavoni, D. Drusian, A. Giacometti, & M. Zanette. (2006). Assessment of organic chlorinated compound removal from aqueous matrices by adsorption on activated carbon. *Water Research* **40**, 3571 – 3579.

P. Le Cloirec, Y. Andrès, C. Faur-Brasquet, & C. Gérente. (2003) Engineered biofilms for metal ion removal. *Reviews in Environmental Science and Bio/Technology* **2**: 177-192.

F.J. Kragt, F.W. van Gaalen, L.C. Rietveld, J.F. Schrijven. (1996) *Definitiestudie Zuiveringsmodel Drinkwater*, RIVM document nr. 734301011.

PROCESMODELLERING ACTIEVE KOOLFILTRATIE

Factsheet actieve koolfiltratie - Verwijdering van trichloormethaan

Inleiding & doel

In het kader van het project 'Definitiestudie Zuiveringsmodel Drinkwaterbereiding' is een onderzoek uitgevoerd naar de verwijdering van trichloormethaan door zuiveringsstap actieve koolfiltratie. Trichloormethaan is een koolwaterstof waarvoor incidentele overschrijding van de norm is geconstateerd. Er is een wiskundig model getoetst waarmee de concentratie van trichloormethaan in water kan worden voorspeld als functie van de tijd. Begrippen die cursief en vetgedrukt zijn weergegeven in de lopende tekst, worden nader toegelicht in de begrippenlijst.

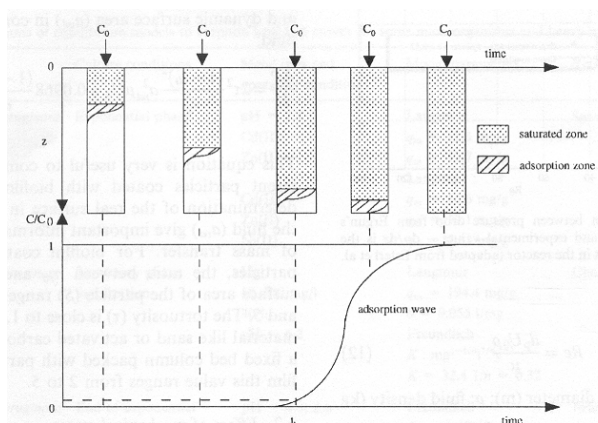
Werking van actieve koolfiltratie

Actieve koolfiltratie berust op **adsorptie** van verontreinigingen in water aan **actieve kool**. Actieve kool is speciaal behandelde koolstof die de eigenschap heeft allerlei stoffen aan zich te

kunnen binden. Actieve kool zit in een kolom waar het water van boven naar beneden wordt doorgepompt.

Het water heeft een continue doorstroom door de kolom. Na verloop van tijd raakt de actieve koolkolom verzadigd. In de figuur is schematisch de verzadiging van de actieve kool weergegeven in de tijd.

C_0 is de ingangconcentratie van een



verontreiniging in water, C de uitgangskoncentratie. Op een gegeven moment gaat de kolom 'leken'. Dit wordt het doorbraaktijdstip genoemd. Vanaf dit punt neemt de adsorptie van de kolom geleidelijk af tot nul. De uitgangskoncentratie C van de verontreiniging in het water neemt vanaf dat tijdstip dan geleidelijk toe.

Invloed van procesvariabelen

Er is een literatuurstudie uitgevoerd om een overzicht te verkrijgen van 1) de **procesvariabelen** die een rol spelen bij actieve koolfiltratie en 2) beschikbare wiskundige modellen voor het beschrijven van doorbraakcurves.

De verontreinigingen worden geadsorbeerd in de poriën van de actieve kool. De verdeling en grootte van de poriën hangt af van de manier waarop de kool is geproduceerd. Ieder type kool heeft een specifieke adsorptieconstante per verontreiniging. De ingangskoncentratie, temperatuur en de contacttijd (tussen verontreiniging en kool) hebben ook invloed op het adsorptieproces. In dit project is alleen de invloed van de procesvariabele *ingangskoncentratie* onderzocht.

In de literatuur zijn er verschillende wiskundige modellen beschikbaar voor beschrijven van doorbraakcurves. Voor alle modellen gelden de volgende aannames: 1) het systeem is in evenwicht (bijv. ingangskoncentratie is constant), 2) er vinden geen chemische reacties en/of biologische omzettingen plaats en 3) de temperatuur is constant. Voor het beschrijven van doorbraakcurves is gekozen voor een 'massa-overdrachts' model, omdat deze met succes is toegepast in vele andere onderzoeken. Het model is gebaseerd op 1) massabalans tussen de vloeistof en vaste fase, 2) lineaire massaoverdracht en 3) de Freundlich vergelijking. De onderliggende fysisch-chemische principes worden hier verder niet behandeld. Het model beschrijft de relatie tussen de uitgangskoncentratie, ingangskoncentratie en de tijd.

$$C(t) = \left[\frac{C_0^{n-1}}{1 + A * e^{-rt}} \right]^{1/n-1}, \text{ waarbij:}$$

$C(t)$ = uitgangskoncentratie (mg/L) op tijdstip t

C_0 = ingangskoncentratie (mg/L)

n = adsorptieconstante

A, r = experimenteel bepaalde fitparameters

Experimenteel onderzoek

Om de waarde van parameters A en r te bepalen zijn laboratoriumexperimenten uitgevoerd waarin de uitgangskoncentratie trichloormethaan is gemeten als functie van de tijd. Er zijn experimenten uitgevoerd met een ingangskoncentratie trichloormethaan in water van 6 mM. De in en uitgangskoncentraties trichloormethaan zijn vastgesteld m.b.v. gaschromatografie. Alle andere procesvariabelen zijn constant gehouden. Een overzicht:

- Ingangskoncentratie trichloormethaan in water: 6 mM.
- Tijd: 1 tot 150 min.

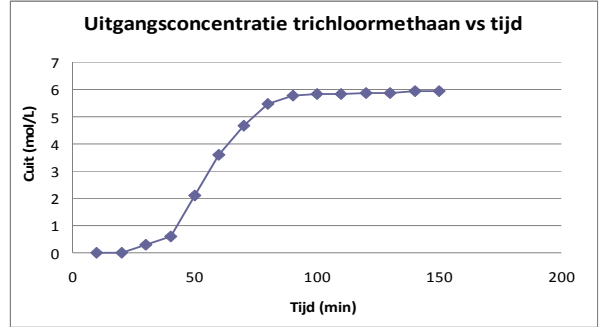
- Type actieve kool: Norit R® (diameter kooldeeltjes tussen 500 en 1000 μm)
- Contacttijd: 14,5 min (volume kolom 0,1 m^3 , waterdebiet door de kolom $6,9 \cdot 10^{-3} \text{ m}^3/\text{min}$)
- Adsorptieconstante r : 3,15 mg/g
- Temperatuur: 10°C

Meetresultaten

De resultaten zijn vastgelegd en verwerkt m.b.v. computerprogramma MS Excel.

De uitgangskoncentratie trichloormethaan (mM) uitgezet tegen de tijd (min).

Procescondities: zie onder Experimenteel onderzoek.



De experimenten zijn uitgevoerd zoals gepland. De metingen zijn nauwkeurig (geen uitschieters) en goed reproduceerbaar.

Wiskundig model

Het wiskundig model is **gefit** op de meetresultaten m.b.v. fitparameters A en r . In onderstaande kader zijn de resultaten weergegeven. In het diagram zijn de gemeten en door het model voorspelde uitgangskoncentraties trichloormethaan weergegeven.

Ingangskoncentratie trichloormethaan

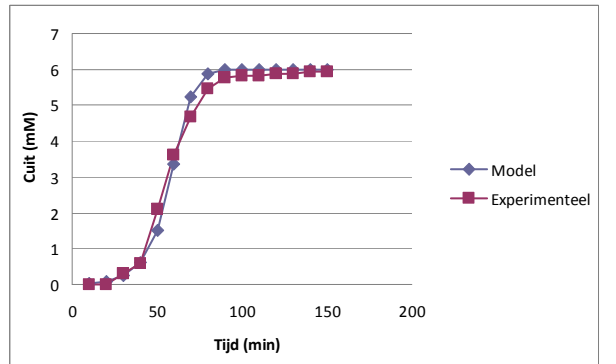
$$C_0 = 6 \text{ mM}$$

Waarde fit parameters:

$$A = 4,0 \cdot 10^5; r = 0,2$$

Formule:

$$C(t) = \left[\frac{C_0^{2,16}}{1 + 4,0 \cdot 10^5 * e^{-0,2*t}} \right]^{0,46}$$



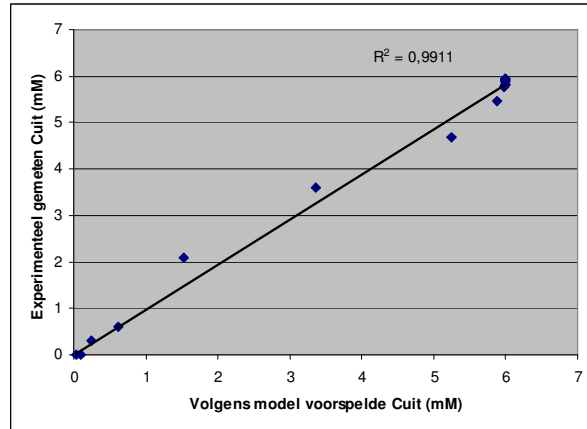
De **passendheid** van de fit op de experimentele data is bepaald m.b.v. een lineaire **regressie** van de gemeten uitgangskoncentraties op de door het model voorspelde uitgangskoncentraties trichloormethaan.

Ingangconcentratie trichloormethaan

$C_0 = 6 \text{ mM}$

Passendheid van het model:

$R^2 = 0,99$



De **geldigheidsvoorwaarden** van het model zijn:

Ingangconcentratie trichloormethaan in water: 6 mol/L.

Adsorptieconstante n : 3,15 mg trichloormethaan / g Norit R®

Contacttijd: 14,5 min.

Temperatuur: 10°C.

Tijd: 0 tot 150 min.

Evaluatie

Naar verwachting is het wiskundige model binnen de geldigheidsvoorwaarden voldoende betrouwbaar, mede gezien de goede kwaliteit van de metingen. Desondanks is er nog veel vervolgonderzoek nodig. Er zijn enkele aannames gemaakt en daarnaast zijn vele procesvariabelen nog niet onderzocht. Het wiskundige model dient verder getest te worden op de volgende aspecten:

- De grenzen van de geldigheidsvoorwaarden van het model, waaronder de ingangconcentratie trichloormethaan.
- De invloed van de temperatuur.
- Adsorptie van andere koolwaterstoffen.
- Testen in bedrijfssituaties.

Daarnaast verdient het aanbeveling ook andere wiskundige modellen, zoals beschreven in de literatuur, te testen. De wiskundige modellen moeten worden geëvalueerd op onderliggende fysisch-chemische principes teneinde te komen tot een solide, gefundeerde beschrijving van doorbraakcurves.

Reflectie en aanbevelingen

Het doel van dit project was het ontwikkelen van een wiskundig model om de uitgangconcentratie trichloormethaan te berekenen als functie van de tijd. M.b.v. een dergelijk model kan de uitgangconcentratie trichloormethaan worden voorspeld en kunnen mogelijk overschrijdingen van de norm worden voorkomen.

Tijdens het onderzoek is gebleken dat het opzetten en uitvoeren van de experimenten cruciaal is. Het verkrijgen van een goede en betrouwbare serie meetresultaten bepalen in hoge mate het eindresultaat.

Een eventueel vervolgproject zal zich moeten richten op de toepasbaarheid van het model onder andere procescondities. Daarnaast dient het model verder getest te worden bij bedrijfsinstallaties.

Tot slot bevelen wij aan dat ieder team zijn werkwijze (modelleerprocedure) op hoofdlijnen beschrijft. Dit kan nuttig zijn voor andere teams die soortgelijk onderzoek uitvoeren naar andere zuiveringsstappen en/of verontreinigingen.

Literatuur

- RIVM document nr. 734301011, *Definitiestudie Zuiveringsmodel Drinkwater*, F.J. Kragt, F.W. van Gaalen, L.C. Rietveld, J.F. Schrijven, augustus 1996
- B. Pavoni, Drusian, D., Giacometti, A., & Zanette, M. (2006). Assessment of organic chlorinated compound removal from aqueous matrices by adsorption on activated carbon. *Water Research* **40**, 3571 – 3579.
- P. Le Cloirec, Andrès, Y., Faur-Brasquet, C., & Gérente, C. (2003) Engineered biofilms for metal ion removal. *Reviews in Environmental Science and Bio/Technology* **2**, 177 – 192.

Begrippenlijst

Actieve kool: Speciaal behandelde koolstofsoort met een grote hoeveelheid fijn verdeelde poriën en een zeer groot inwendig porieoppervlak.

Adsorptie: Het opnemen en vasthouden van een oplosbare stof, bijv. het hechten van trichloormethaan uit water aan actieve kool.

Procesvariabelen: Variabelen die een invloed uitoefenen op een proces. Procesvariabelen kunnen tijdens het proces veranderen.

Fitten: Het 'passend maken' van een wiskundig model op de experimentele gegevens door systematisch de waarden van fitparameters te wijzigen.

Regressie: Een statistische methode om een verband tussen een serie meetresultaten zo goed mogelijk te beschrijven. Met zo goed mogelijk wordt bedoeld dat bij een modeltype (lineair, macht) de totale afwijking van de door het model berekende uitkomsten ten opzichte van de data minimaal is.

Passendheid: De passendheid van een wiskundig model bij een serie meetresultaten, bijv. uitgedrukt in de waarde van de coëfficiënt van determinatie R^2 .

Geldigheidsvoorwaarden: Conditie waarbinnen een model geldig is.

Bron 3. Uitvoering van Vlokbehandeling

Onbehandeld oppervlaktewater bevat kleideeltjes, geoxideerde metalen, micro-organismen en ander gesuspendeerd materiaal, die het water troebel maken. Deze deeltjes worden in de eerste twee stappen in het zuiveringstraject, spaarbekken en vlokbehandeling, verwijderd. Vlokbehandeling omvat het verwijderen van kleine gesuspendeerde colloïdale deeltjes die niet direct door bezinking of filtratie kunnen worden gescheiden. In dit artikel wordt de werking van de zuiveringsstap vlokbehandeling nader beschreven.

Onbehandeld oppervlaktewater bevat zwevende deeltjes die het water troebel maken. Bijvoorbeeld slib en kleine modderdeeltjes. De zwevende deeltjes die het water troebel maken kunnen ingedeeld worden in twee groepen: grotere deeltjes die vanzelf uitzakken, en zeer kleine deeltjes die niet uitzakken. In onderstaande figuur 1 staat een overzicht van deze indeling. Als de deeltjes in water groter zijn dan $1\ \mu\text{m}$, dan is er sprake van een *suspensie*. Deeltjes in suspensies zakken uiteindelijk vanzelf uit (sedimenteren). Deeltjes met een grootte tussen $10\ \mu\text{m}$ en $100\ \mu\text{m}$ zakken snel uit. Kleinere deeltjes tussen de $1\ \mu\text{m}$ en $10\ \mu\text{m}$ (bijvoorbeeld fijne klei) sedimenteren langzaam, maar zullen uiteindelijk ook onder invloed van de zwaartekracht neerslaan. Al deze deeltjes kunnen ook door een filter worden weggevangen.

Het opgepompte oppervlakte- en grondwater wordt opgevangen in een spaarbekken. Hier bezinken de deeltjes tot de orde grootte van $1\ \mu\text{m}$. In het water zitten dan nog de zwevende deeltjes met een grootte vanaf $0,01$ tot $1\ \mu\text{m}$ die niet uit zichzelf sedimenteren. Deze deeltjes vormen een zogenaamd *colloïdaal systeem*. Om colloïdale deeltjes te verwijderen wordt vlokbehandeling toegepast. Vlokbehandeling is vaak de tweede stap in het zuiveringstraject van oppervlaktewater. Tijdens vlokbehandeling wordt het colloïdale systeem 'gedestabiliseerd' waardoor de deeltjes samenklitten en uitgroeien tot vlokken. Deze vlokken kunnen vervolgens wel door filtratie worden verwijderd. Het proces vlokbehandeling bestaat uit drie stappen: coagulatie, flocculatie en sedimentatie of filtratie.

$100\ \mu\text{m}$	$10\ \mu\text{m}$	$1\ \mu\text{m}$	$0,1\ \mu\text{m}$ μm	$0,01$	$0,001\ \mu\text{m}$ 1\AA
Suspensies			colloïdale systemen		Oplossingen
Deeltjes zichtbaar met microscoop			zichtbaar met elektronenmicroscoop		
snelle sedimentatie		trage sedimentatie	geen sedimentatie		
tegengehouden door filtreerpapier			niet tegengehouden door filtreerpapier		

Figuur 1: Overzicht van grootte van deeltjes in water

Bezinksnelheden

De sedimentatiesnelheden is afhankelijk van de grootte van de deeltjes. Deeltjes die kleiner zijn dan 1 μm hebben sedimenteren niet of nauwelijks. In onderstaande Tabel 1 staat een overzicht van sedimentatiesnelheden.

Grootte deeltje (diameter mm)	Deeltje	Sedimentatie Tijd over 30 cm
10	Gravel	0.3 seconden
1	Grof zand	3 seconden
0.1	Fijn zand	38 seconden
0.01	Slib	33 minuten
0.0001	Klei	230 dagen
0.00001	Colloïden	63 jaar

Tabel 1: Relatie tussen grootte van het deeltje en de sedimentatiesnelheid

Coagulatie

De eerste stap bij vlokbehandeling betreft coagulatie. Coagulatie is het proces waarbij door het toevoegen van chemicaliën de elektrostatische afstoting tussen colloïdale gesuspendeerde deeltjes wordt verlaagd of geheel verdwijnt. Onder natuurlijke omstandigheden zijn colloïdale deeltjes in water negatief geladen. Hierdoor blijven de deeltjes stabiel in een waterfase. Dit wordt doorbroken door inbrengen van een

vlokmiddel: een zout met een sterk positief geladen ion, zoals FeCl_3 of $\text{Al}_2(\text{SO}_4)_3$. De positief geladen ionen Fe^{3+} of Al^{3+} fungeren als coagulant. De colloïdale deeltjes worden instabiel en klonteren samen tot grotere deeltjes.

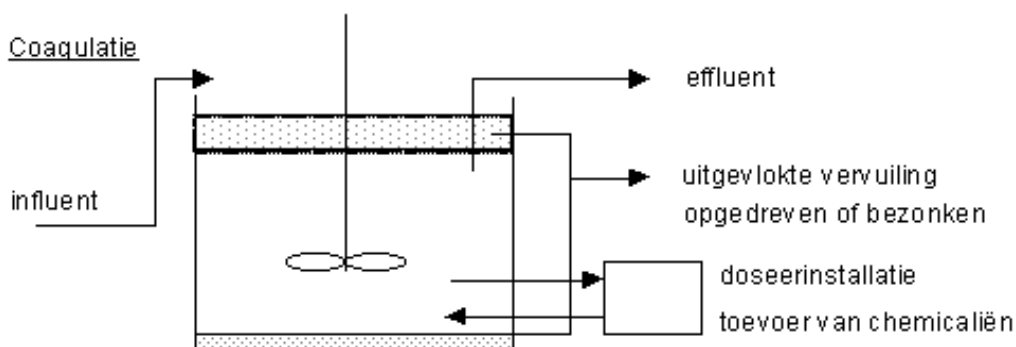
Coagulatie is een proces dat zich in korte tijd afspeelt. Het is van belang dat de coagulant in korte tijd volledig door het water gemengd wordt, zodat overal in de vloeistof coagulatie kan plaatsvinden. Snelle en volledige menging wordt bereikt door in een tank tijdens toevoegen van het vlokmiddel hard te roeren (hoge turbulentie). Zie onderstaande figuur 2.

Flocculatie

Tijdens coagulatie groeien de deeltjes in korte tijd tot vlokkernen van circa 1 μm . Vanaf deze vloggrootte kunnen de deeltjes alleen nog maar doorgroeien door *rustig* te roeren. Dit proces wordt flocculatie genoemd en vindt doorgaans plaats in een tweede tank met een langere verblijftijd en lagere roersnelheid.

Sedimentatie of filtratie

De samengeklonterde deeltjes na flocculatie kunnen door bezinking of filtratie worden verwijderd. Het restmateriaal (de vlokken) wordt als slib afgevoerd. Het slib wordt ontwaterd tot droge stof en vervolgens afgevoerd.



Figuur 2: Schematische weergave van de uitvoering van vlokbehandeling in een tank.

Procesvariabelen

Voor elk type water dient proefondervindelijk de optimale coagaliatie- en flocculatiecondities te worden vastgesteld. Er zijn vele procesvariabelen die een invloed hebben op vlokbehandeling. Een drinkwaterbedrijf zal constant moeten inspelen op de variërende omstandigheden. Het bedrijf kan dit doen door de hoeveelheid vlokmiddel aan te passen aan de omstandigheden.

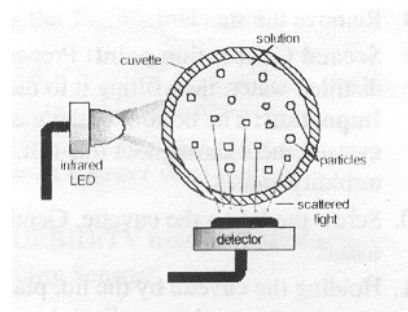
Grootheid troebeling en eenheid NTU

De troebeling van water wordt veroorzaakt door colloïdale en andere gesuspendeerde deeltjes in water. De grootheid troebeling is een maat voor de helderheid van water en een belangrijke indicator van de waterkwaliteit. Water met hoge troebeling is mistig, terwijl water met lage troebeling helder is. De mistigheid wordt veroorzaakt door licht dat wordt gereflecteerd door aanwezige deeltjes in het water. Daarom: hoe meer deeltjes in het water, hoe troebeler het water.

Troebelingsmetingen zijn gebaseerd op de optische eigenschap van water om licht te verstrooien en te absorberen i.p.v. recht door te laten (transmissie). Zelfs volledig gezuiverd water heeft een zekere mate van troebeling.

Er bestaan vele technieken om de troebeling van water te meten. Een tegenwoordig gangbare techniek voor troebelingsmetingen is *nephelometrie*. Nephelometrische troebelingsmeters meten de intensiteit van licht verstrooid onder een hoek van 90° met de ingaande bundel licht. Zie figuur 3 voor een schematisch overzicht.

De eenheid voor de mate van troebeling is NTU. Deze afkorting staat voor de Engelse term *Nephelometric Turbidity Unit*. In Nederland wordt de eenheid FTE gebruikt, een afkorting voor Formazine Troebeling Eenheid. De eenheden NTU en FTE hebben gelijke waarden. Een troebelingsmeter wordt geijkt met een troebele referentie vloeistof van 100 NTU en demiwater (2 NTU).



Figuur 3: Meten van intensiteit van licht verstrooid onder een hoek van 90°

De totale intensiteit van verstrooid licht door een troebel watermonster is afhankelijk van aantal, grootte, vorm en kleur van aanwezige deeltjes en gebruikte golflengte van het licht. Er bestaat een grove relatie tussen de gemeten troebelheid in NTU en het aantal deeltjes in water. Onderstaande Tabel 2 geeft een indruk van het effect van het aantal deeltjes op de gemeten troebeling.

Troebeling (NTU)	Geschatte aantal deeltjes in 10 ml vloeistof
5	200.000
1	60.000
0.5	10.000
0.1	200

Tabel 2: Relatie tussen troebeling in NTU en aantal aanwezige deeltjes

Literatuur

- Papacosta, K. (2002) *Turbidity calibration standards evaluated from a different perspective*. Paper presented at the *Turbidity and Other Surrogates Workshop*, Reno, NV.
- Vlok, C., K. Bell, E. Ibrahim, D. Verges, G. Amy and M. Lechevallier. (2000). *Impact of enhanced and optimized coagulation on removal of organic matter and its biodegradable fraction in drinking water*. In *Wat. Res.*, Vol. 34(2) pp. 3247-3257.
- Health Canada (1986) *Turbidity*. In *Water Quality Control Annual Report, Great Vancouver Region District*, 2005

Bron 4. Coagulatie & Flocculatie

Het verwijderen van gesuspendeerde deeltjes uit water tijdens vlokbehandeling is één van de belangrijkste doelen tijdens waterzuivering. Het belang van vlokbehandeling wordt extra onderstreept door het feit dat betrouwbare desinfectie, doorgaans de laatste stap in de zuiveringsketen, afhangt van het slagen van de vlokbehandeling. Alle reden dus om het proces van coagulatie en flocculatie nader te bestuderen en te begrijpen. In dit artikel volgt een chemisch fysische beschouwing van colloïdale systemen en de principes waarop zuiveringsmethode vlokbehandeling berust.

Inleiding

Gesuspendeerde deeltjes in water veroorzaken troebelheid. Deze deeltjes variëren in grootte van 0.01 tot 100 μm . De grotere deeltjes van 1 tot 100 μm kunnen gemakkelijk verwijderd worden door sedimentatie of filtratie. Echter, de kleine deeltjes van 0.01 tot 1 μm zijn moeilijk te verwijderen. Deze deeltjes hebben een oneindig trage sedimentatiesnelheid en worden niet tegengehouden door filtratie. De zuiveringsstap vlokbehandeling is erop gericht deze kleinere, colloïdale deeltjes te verwijderen. Om dit proces te begrijpen is kennis nodig over interacties tussen colloïdale deeltjes.

Onder natuurlijke omstandigheden zijn colloïdale deeltjes in water negatief geladen. Hierdoor stoten de deeltjes elkaar af. Dit voorkomt het samenklitten van de colloïdale deeltjes tot grotere deeltjes die wel sedimenteren onder invloed van de zwaartekracht.

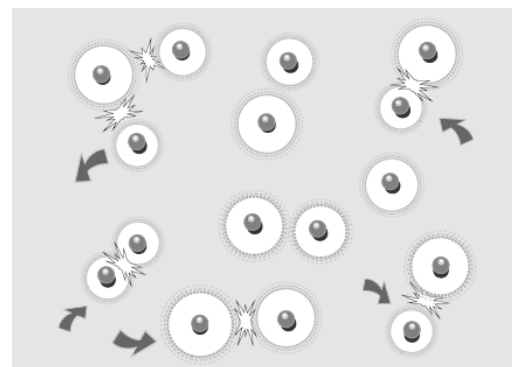
Om de colloïdale deeltjes te verwijderen is het nodig de netto negatieve lading te verminderen. Vlokbehandeling berust op dit principe. Dit is mogelijk door het toevoegen van een vlokmiddel, bijvoorbeeld een zout met een sterk positief geladen ion (zoals Fe^{3+} of Al^{3+}). Hierdoor kunnen de deeltjes wel samenklitten en langzaam uitgroeien tot bezinkbare vlokken.

Coagulatie is een complex proces waarin meerdere variabelen een rol spelen. In dit artikel wordt nader ingegaan op de

moleculaire mechanismen van coagulatie en flocculatie.

Colloïden: geladen deeltjes in water

Het bedrag van colloïdale deeltjes in water wordt sterk beïnvloed door de lading van het deeltje. Natuurlijk voorkomende colloïdale deeltjes hebben over het algemeen allemaal een negatieve lading. Door deze gelijke lading stoten de deeltjes elkaar af met de zgn. Coulombkracht (Figuur 1).



Figuur 1: Geladen deeltjes stoten elkaar af

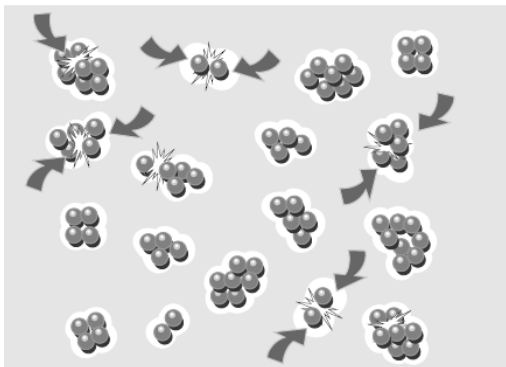
Naast de afstotende Coulombkracht werken ook de aantrekkende Van der Waals krachten op de deeltjes in. Echter, de aantrekkende Van der Waals krachten zijn zwakker dan de afstotende Coulombkracht. Dit voorkomt dat de deeltjes samen kunnen klonteren.

In het water bewegen en botsen de colloïdale deeltjes constant tegen elkaar en tegen watermoleculen. Deze voortdurende beweging wordt de Brownse beweging

genoemd. De voortdurende beweging van de deeltjes, ofwel de kinetische energie van de deeltjes, zorgt ervoor dat de deeltjes niet neerslaan ondanks de werking van de zwaartekracht.

Het resultaat van de afstotende Coulombkrachten en aantrekkende Van der Waals krachten enerzijds, en de Brownse beweging en zwaartekracht anderzijds, is dat de colloïdale deeltjes in het water blijven zweven. Er is dan sprake van een stabiel colloïdaal systeem.

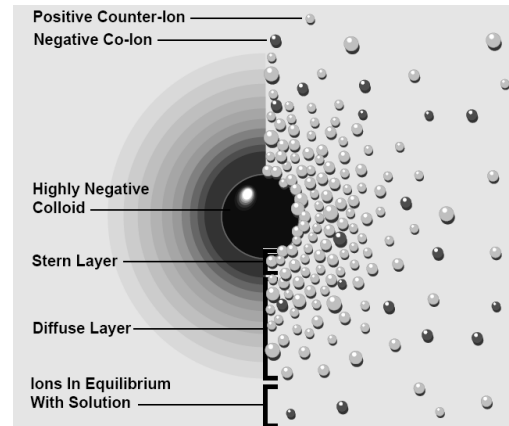
Echter, als de negatieve lading op de deeltjes wordt verminderd of geneutraliseerd, dan worden de afstotende Coulombkrachten kleiner. Op dat moment zijn de aantrekkende Van der Waals krachten wel voldoende groot en klonteren de colloïdale deeltjes samen. Eerst worden kleine kernen gevormd, die langzaam uitgroeien tot vlokken. Deze vlokken hebben meer massa (grotere zwaartekracht) en bewegen langzamer (minder kinetische energie; Brownse beweging neemt af). De ontstane vlokken slaan wel spontaan neer (sedimenteren). Dit proces is schematisch weergegeven in figuur 2.



Figuur 2: Ongeladen deeltjes klonteren samen

Elektrostatische krachten op een colloïdaal deeltje

Om de moleculaire processen rond een colloïdaal te visualiseren maken we gebruik van het 'dubbele laag' model. Dit model maakt het gemakkelijker te begrijpen wat er gebeurt in de directe omgeving van een negatief geladen colloïdaal deeltje als de omringende mantel van ionen wordt verstoord.



Figuur 3: Dubbele laag model op twee manieren: links de ladingsdichtheid rondom een colloïdaal deeltje, rechts de verdeling van de positieve en negatieve ionen om het geladen colloïdale deeltje

De positieve ionen in het water vormen een eerste 'ring' rondom het oppervlak van een negatief geladen deeltje. Deze ring wordt de *Stern* ring genoemd. De positieve ionen in de Stern ring zijn gebonden aan het oppervlak van het deeltje. Andere positieve ionen in de oplossing worden ook aangetrokken door de negatieve lading van het deeltje, maar worden tegelijk ook afgestoten door al aanwezige positieve ionen in de Stern ring. Hierdoor ontstaat een tweede diffuse ring van positieve ionen rondom de Stern ring. De concentratie positieve ionen in deze tweede diffuse ring is hoog dicht bij het colloïdale deeltje en neemt geleidelijk af tot evenwicht concentratie van de positieve ionen in oplossing.

De gebonden positieve ionen in de Stern ring en de diffuse ring worden samen de 'dubbele laag' genoemd. De dikte van de dubbele laag hangt af van de concentratie ionen in de oplossing. Bij een hoge concentratie ionen in de oplossing zijn er veel positieve ionen beschikbaar om het colloïdale deeltje te neutraliseren. Er zitten veel ionen in de Stern en diffuse ring. De lading van het colloïde deeltje wordt goed afgeschermd. Op korte afstand van het colloïde deeltje is de negatieve lading al geneutraliseerd. We spreken dan van een *dunne* 'dubbele laag'. Bij een lage zoutconcentratie is er sprake van het omgekeerde effect. Er zijn weinig positieve ionen in de Stern en diffuse ring. De lading van het colloïde deeltje wordt niet goed afgeschermd. Pas op lange afstand van het colloïde deeltje is de lading geneutraliseerd. Dit wordt aangeduid met een *dikke* 'dubbele laag'.

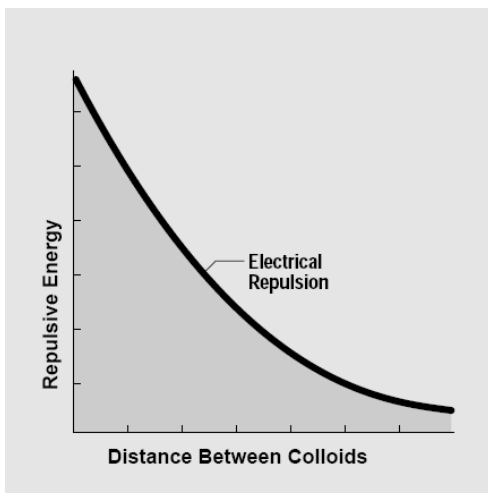
Het type positief ion is zeer van invloed. Bijvoorbeeld, ijzerionen (Fe^{3+}) zullen veel effectiever zijn in het neutraliseren van de negatieve lading dan natriumionen (Na^+) in gelijke concentratie. Het verschil in lading is de oorzaak. Ook de grootte van het ion is van belang. Er kunnen veel meer kleine positieve ionen in de Stern ring opgenomen worden dan grootte, met als resultaat dat de negatieve lading meer wordt afgeschermd. De conclusie is dat ijzerionen effectiever zijn in het afschermen van de negatieve lading van het colloïdale deeltje dan natriumionen. Natriumionen worden daarom ook wel aangeduid met 'inerte' ionen. Dat houdt in dat een verschil in concentratie natriumionen in water slechts een klein effect heeft op het afschermen van de negatieve lading van het colloïdale deeltje.

Balanceren tussen tegengestelde krachten

De DLVO theorie (genoemd naar Derjaguin, Landau, Verwey en Overbeek) geeft een klassieke beschrijving van interacties tussen deeltjes. De theorie gaat uit van twee tegengestelde krachten: de elektrostatische afstoting en de aantrekkende Van der Waals krachten.

Afstoting

De afstotende kracht (Coulombkracht) tussen twee deeltjes neemt toe als de afstand tussen twee deeltjes afneemt (vergelijk met twee magneten).

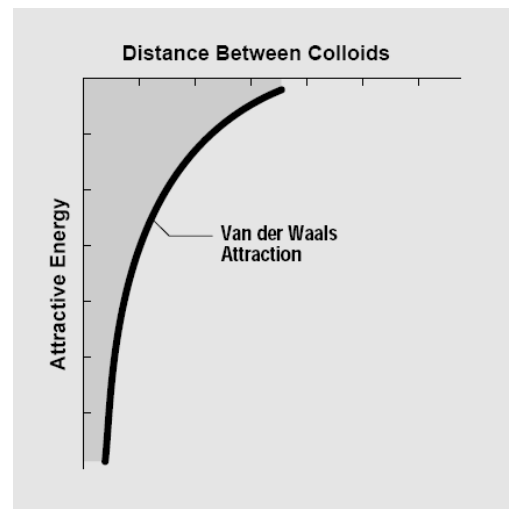


Figuur 4: Afstotende energie t.g.v. elektrostatische krachten

Op moleculaire schaal gaan beide 'dubbele lagen' overlappen. Figuur 4 geeft een *indicatie* van de energie die nodig is om deeltjes samen te brengen afhankelijk van de afstand tussen twee deeltjes.

Aantrekking

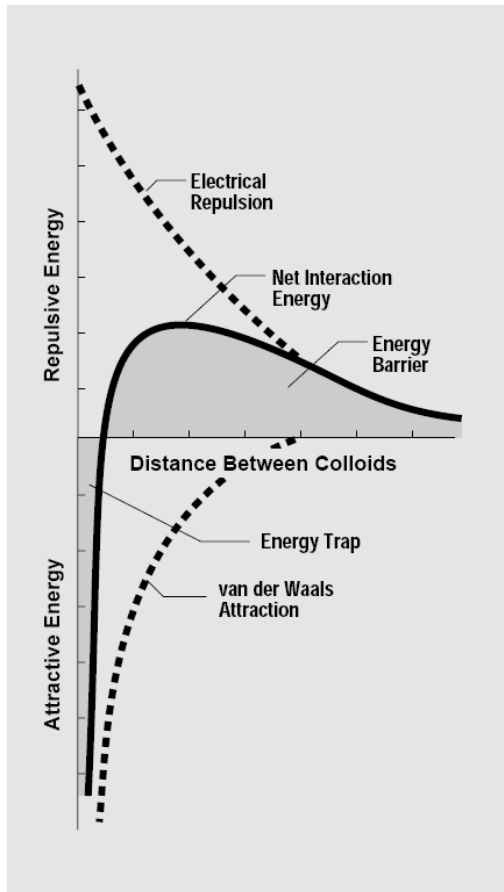
De Van der Waals aantrekking tussen twee deeltjes komt door de krachten tussen aparte moleculen in elk deeltje. Het effect stapelt zich op. Ieder molecuul in deeltje 1 oefent Van der Waals krachten uit op alle moleculen in deeltje 2. Dit geldt voor alle moleculen in zowel deeltje 1 en 2. Ook dit effect kan weergegeven worden in een diagram. In Figuur 5 staat de aantrekkende energie als functie van de afstand tussen deeltjes.



Figuur 5: Van der Waals aantrekking tussen colloïdale deeltjes

De energiebarrière

De DLVO theorie combineert de afstotende elektrostatistische met de aantrekkende Van der Waals krachten. In Figuur 6 is het netto effect weergegeven als functie van de afstand tussen colloïdale deeltjes.



Figuur 6: Netto afstoting en aantreking tussen twee colloïdale deeltjes als functie van de afstand

Het gedeelte waarin de afstoting groter is dan de aantreking wordt aangeduid met de *energiebarrière*. Om het coagulatie proces op gang te brengen moet de energiebarrière overwonnen worden. Hoe hoger de temperatuur van het water, hoe sneller de deeltjes bewegen en heftiger de botsingen zijn. Bij hogere temperaturen zijn meer botsingen effectief en overwinnen meer deeltjes de energiebarrière. Als de barrière wordt overwonnen, dan is het netto effect aantreking. De deeltjes klitten samen tot groeikernen.

Verlagen van de energiebarrière

Voor een effectieve coagulatie is het nodig de energiebarrière zoveel mogelijk te verlagen, of tot nul te reduceren zodat er altijd een netto aantrekkend effect is. Dit kan worden bereikt door de negatieve lading van het colloïde deeltje te verminderen, zodat de deeltjes elkaar dichter kunnen naderen. Dit wordt ook wel 'dubbele laag compressie' genoemd.

Er zijn feitelijk twee manieren waarop de negatieve lading van colloïde deeltjes kan worden verminderd: door 'dubbele laag compressie' en 'neutralisatie'. Beide worden hieronder kort beschreven.

Dubbele laag compressie

Dubbele laag compressie kan worden bereikt door de concentratie positieve ionen in oplossing te verhogen. Als de concentratie stijgt, dan wordt de 'dubbele laag' dunner en neemt de energiebarrière af. Deeltjes klonteren gemakkelijk samen omdat ze elkaar dichter kunnen naderen en minder kinetische energie nodig hebben om de barrière te overwinnen. Dit effect wordt ook wel uitzouten genoemd.

Verlagen van oppervlakte lading

Door toevoegen van een coagulant wordt de totale oppervlakte lading op een colloïdaal deeltje verlaagd. Een coagulant hecht zich op het oppervlakte van het colloïde deeltje, dus in de Stern ring. Ook hierbij geldt dat de deeltjes elkaar dan dichter kunnen naderen en minder kinetische energie nodig hebben om de barrière te overwinnen.

Mechanismen voor vlokbehandeling

Tijdens vlokbehandeling worden de termen coagulatie en flocculatie vaak door elkaar heen gebruikt. Het is verstandig beide termen apart te houden uitgaande van hun functie.

Met coagulatie wordt bedoeld het verlagen van de energiebarrière. Deze fase wordt ook wel aangeduid met destabiliseren van een colloïdaal systeem door toevoegen van een coagulant.

Met flocculatie wordt bedoeld het verder samenklonteren van groeikernen in een gedestabiliseerd colloïdaal systeem. Flocculatie kan worden gestimuleerd door langzaam te roeren zodat deeltjes met elkaar

in contact komen. Aanwezige groeikernen groeien dan snel door tot zichtbare vlokken. Verder kan flocculatie worden gestimuleerd door toevoegen van een flocculant.

Er is geen strikte grens tussen coagulatie en flocculatie. Veel coagulanten, zoals FeCl_3 , fungeren ook als flocculant.

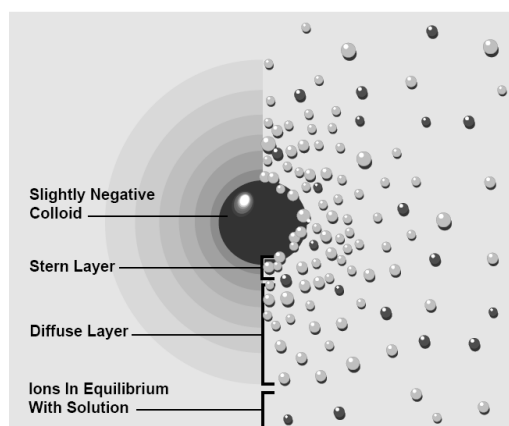
Hieronder wordt verder ingegaan op drie mechanismen voor coagulatie:

1. Verlagen van oppervlakte lading
2. Dubbele laag compressie
3. Colloïden 'inkapselen'

In de praktijk zal, wanneer over coagulatie en flocculatie wordt gesproken, altijd sprake zijn van twee of meer gelijktijdig verloopende mechanismen. Bij het opzetten van experimenten om vlokbehandeling te testen is het verstandig vooraf na te gaan van welk(e) mechanisme(n) sprake zal zijn.

1. Verlagen van oppervlakte lading

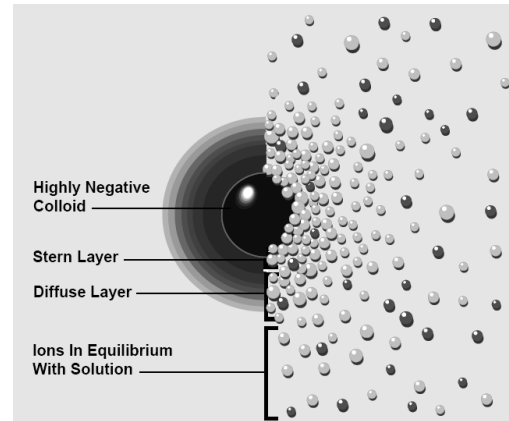
Vele coagulanten werken door het verlagen van de oppervlakte lading op een colloïdaal deeltje. Bekende coagulanten zijn ijzerchloride en aluminiumsulfaat. Het werkzame bestanddeel is het positief geladen ion. Verlagen van oppervlakte lading houdt in het hechten van een positieve ion op het negatief geladen oppervlak van een colloïdaal deeltje (Stern ring). Bij voldoende positieve ionen op het oppervlak wordt de netto lading van een colloïdaal deeltje nul. Dit mechanisme staat schematisch weergegeven in Figuur 8.



Figuur 8: Verlagen van oppervlakte lading op een colloïdaal deeltje.

2. Dubbele laag compressie

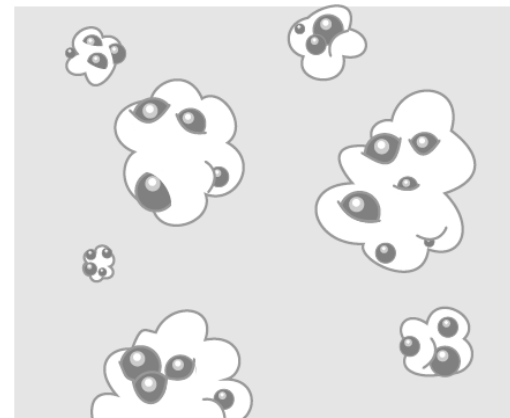
Dubbele laag compressie houdt in het toevoegen van grote hoeveelheden van een 'inert' zout, zoals NaCl . Met 'inert' wordt bedoeld dat het positief ion onveranderd blijft en zich niet hecht aan het oppervlak van het colloïdale deeltje. Dit proces wordt ook wel *uitzouten* genoemd. Dit mechanisme heeft alleen effect op de 'dubbele laag', en niet op de netto negatieve lading van het colloïdale deeltje zelf. Dubbele laag compressie staat schematisch in Figuur 9.



Figuur 9: Dubbele laag compressie.

3. Colloïden 'inkapselen'

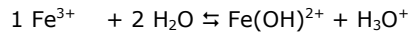
Colloïden 'inkapselen' slaat op het toevoegen van relatief grote hoeveelheden van een coagulant. Veel positief geladen metaal ionen slaan neer met aanwezige OH^- ionen in het water en vormen onoplosbare oxiden. Tijdens dit neerslag proces worden veel colloïdale deeltjes letterlijk ingekapseld. Dit is schematisch weergegeven in Figuur 10. De hoeveelheid toegevoegde coagulant is echter veel groter dan benodigd in mechanisme 2 (verlagen van oppervlakte lading).



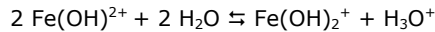
Figuur 10: Colloïden 'inkapselen'.

De zuurgraad van het water (pH)

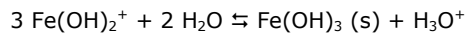
Vele metaalionen oxideren in water en slaan neer als metaaloxiden. Bijvoorbeeld, in water oxideert het ion Fe^{3+} tot het slecht oplosbare zout $\text{Fe}(\text{OH})_3$. Het oxidatieproces verloopt via een aantal evenwichtsreacties:



$$K = 6.8 \cdot 10^{-3}$$



$$K = 2.6 \cdot 10^{-5}$$



$$K = 1 \cdot 10^{-6}$$

De zuurgraad van het water (*pH*) heeft invloed op de ligging van de evenwichten. In een zuur milieu (lage *pH*), dan is er relatief veel Fe^{3+} , $\text{Fe}(\text{OH})^{2+}$ en $\text{Fe}(\text{OH})_2^+$ aanwezig. Onder basische condities (hoge *pH*), dan slaat er relatief veel $\text{Fe}(\text{OH})_3$ neer.

Welk(e) coagulatie mechanisme(n) plaatsvinden, hangt sterk af van de aanwezigheid van de deeltjes Fe^{3+} , $\text{Fe}(\text{OH})^{2+}$, $\text{Fe}(\text{OH})_2^+$ en $\text{Fe}(\text{OH})_3$. Bij lage *pH* zullen hoofdzakelijk de mechanismen *neutralisatie* en *inkapselen* plaatsvinden, bij hoge *pH* is *inkapselen* het dominante mechanisme.

In overzicht:

<i>pH</i>	Coagulatie mechanisme(n)
laag	Combinatie <i>neutralisatie</i> / <i>inkapselen</i>
hoog	<i>Inkapselen</i>

Conclusie

De zuiveringsmethode vlokbehandeling is gebaseerd op het principe van het verstoren van de elektrostatische krachten in een colloïdaal systeem. Daartoe wordt een zogenaamd vlokmiddel toegevoegd. Als vlokmiddel kunnen kleine, sterk geladen ionen gebruikt worden, bijvoorbeeld Fe^{3+} . In dit artikel is een overzicht gegeven van de chemie achter vlokbehandeling. Er vindt veel onderzoek plaats naar het gedrag van colloïdale systemen. Het is mogelijk dat er in de toekomst nieuwe innovatieve technieken ontwikkeld worden voor verwijderen van colloïden uit water.

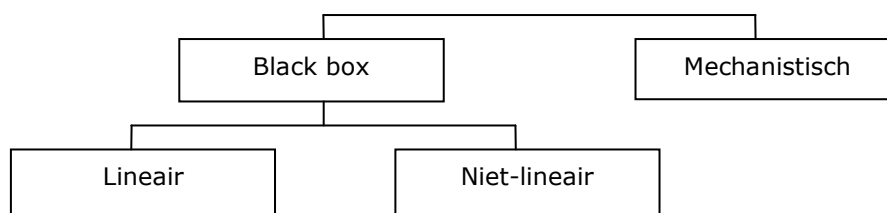
Literatuur

1. Everett, D.H. (2007). *What are Colloids?* In *Basic principles of Colloid Science*, chapter 1, pp. 1-15. Royal Society of Chemistry Paperbacks.
2. Everett, D.H. (2007). *Why are Colloidal Dispersions Stable? I Basic Principles.* In *Basic principles of Colloid Science*, chapter 2, pp. 16-29. Royal Society of Chemistry Paperbacks.
3. Everett, D.H. (2007). *Why are Colloidal Dispersions Stable? II Interparticle Forces.* In *Basic principles of Colloid Science*, chapter 3, pp. 30-53. Royal Society of Chemistry Paperbacks.
4. Figuren uit *Everything you want to know about Coagulation & Flocculation.* (1993). Zeta-Meter, Inc.

Bron 5. Procesmodellering

In deze module wordt een wiskundig model ontwikkeld dat de relatie beschrijft tussen de eindtroebeling en enkele procesvariabelen bij zuiveringsstap vlokbehandeling. Het ontwikkelde model kan gebruikt worden om het effect van veranderingen in één van de procesvariabelen op de eindtroebeling door te rekenen. De uitkomsten zijn bedoeld om de dosering van de stuurparameter dosis vlokmiddel aan te passen aan de omstandigheden.

Het algemene doel van procesmodellering is het verkrijgen van inzicht en controle over een (industriële) proces. Een belangrijke eis is dat het model in staat is het effect van veranderingen in de procescondities op het eindresultaat te voorspellen. Ruwweg kunnen twee typen modellen worden onderscheiden: mechanistische en 'black box' modellen. Black box modellen worden verder onderverdeeld in lineaire en niet-lineaire typen. In onderstaande figuur is de classificatie van schematisch weergegeven.



Mechanistische modellen

Als het onderliggende mechanisme en karakteristieken van een proces tot in detail bekend zijn, dan kan het gedrag van een proces worden beschreven door een mechanistisch model (in veel gevallen door een set van differentiaal vergelijkingen). Een mechanistisch model is vaak afgeleid van onderliggende natuurkundige en/of chemische principes. Een mechanistisch model geeft een navolgbare en exacte beschrijving van de werking en gedrag van een proces. Het model is toetsbaar door experimenten uit te voeren (empirie). Het ontwikkelen van een mechanistisch model kost veel tijd (en dus geld). Dit geldt vooral voor processen die zeer complex zijn en/of processen waarover weinig bekend is. Soms is het ontwikkelen van een mechanistisch model niet mogelijk, terwijl het wel van belang is inzicht te verkrijgen in het gedrag van een proces. In dat soort omstandigheden worden 'black box' modellen ontwikkeld.

Black box modellen

Black box modellen beschrijven de relatie tussen ingaande en uitgaande variabelen. De factoren in deze modellen hebben vaak geen fysische betekenis en/of relatie met procesvariabelen, zoals temperatuur of zuurgraad. Dit is een nadeel van dit soort modellen vergeleken met mechanistische modellen. Echter, als het gaat om het beschrijven van trends in het gedrag van een proces, dan zijn black box modellen net zo effectief. Daarom wordt in de praktijk regelmatig gebruik gemaakt van black box modellen. Een belangrijk aandachtspunt bij black box modellen is de kwaliteit en hoeveelheid benodigde empirie om de trends met voldoende betrouwbaarheid te beschrijven. Met andere woorden: hoeveel empirie is nodig om je model als 'betrouwbaar' te bestempelen. Het ontwikkelen van een black box model kost weinig tijd in vergelijking met een mechanistisch model.

Procesmodellering vlokbehandeling

Bij vlokbehandeling spelen diverse variabelen een rol. Op basis van de beschikbare fysisch-chemische kennis over coagulatie mechanismen en praktijkervaringen wordt hieronder voor variabelen dosering vlokmiddel, begintroebeling, zuurgraad (pH), temperatuur en zoutconcentratie de verwachte invloed op de eindtroebeling beschreven. Het effect van de menging valt buiten het bestek van deze module en wordt hier niet besproken.

Dosering vlokmiddel

Tussen de eindtroebeling en de dosering vlokmiddel wordt een lineair verband aangenomen, zolang er voldoende deeltjes in het water zijn en de dosering van het vlokmiddel laag is. Echter, bij toenemende dosering vlokmiddel nadert de eindtroebeling asymptotisch naar 0 NTU. Er wordt aangenomen dat het verband tussen de eindtroebeling en de dosering vlokmiddel beschreven kan worden met een machtsfunctie, mits alle andere variabelen constant zijn.

Begintroebeling

Naast de dosering vlokmiddel is de begintroebeling een belangrijke variabele. Voor water met een hoge begintroebeling is er *relatief* minder vlokmiddel nodig dan voor water met een lage begintroebeling. Doorgaans wordt een lineair verband aangenomen tussen eindtroebeling en begintroebeling, onder voorwaarde dat alle andere variabelen constant zijn.

Zuurgraad (pH)

Verandering van de zuurgraad (pH) heeft invloed op het mechanisme van coagulatie. Bij hoge pH waarden is inkapselen het dominante mechanisme, terwijl bij lage pH waarden voornamelijk neutralisatie plaats vindt. Vaak is er sprake van een optimale pH waarde, mede afhankelijk van het type water, die proefondervindelijk moet worden vastgesteld.

Temperatuur

Bij lage temperatuur verloopt het coagulatie en flocculatie proces langzamer. Echter, door de toenemende viscositeit van het water en bij gelijk blijvende mengintensiteit, neemt de energie-inbreng toe en dientengevolge ook de shear-rate. Uiteindelijk kan dat zowel positief als negatief uitpakken op het eindresultaat.

Zoutconcentratie

Een toenemende zoutconcentratie leidt tot dubbele laag compressie. De deeltjes kunnen elkaar dichter naderen en gemakkelijker coaguleren. Echter, de natuurlijke variatie in de zoutconcentratie in zoet water is te klein om enig significant effect te verwachten. Ter vergelijking, als zoet water in aanraking komt met zeewater, treedt vaak spontaan coagulatie op. De totale zoutconcentratie in zeewater bedraagt echter ca. 35 g/L, terwijl deze in zoet water schommelt rond ca. 150 mg/L. Er is dus een enorme hoeveelheid zout nodig voordat het mechanisme dubbele laag compressie effect heeft op de eindtroebeling.

Invloed van alle variabelen gelijktijdig

In de praktijk spelen alle bovenstaande variabelen gelijktijdig een rol. Wanneer alle variabelen in beschouwing worden genomen, moeten we ons allereerst afvragen of de variabelen onderling onafhankelijk zijn of niet. Indien onafhankelijk, dus als de variabelen elkaar onderling niet beïnvloeden, dan is de verwachting dat het verband tussen de eindtroebeling en de procesvariabelen beschreven kan worden met een additief model. Indien de variabelen elkaar onderling wel beïnvloeden, dus afhankelijk van elkaar zijn, dan lijkt een multiplicatief model geschikter.

Op het eerste gezicht lijken de dosering vlokmiddel, begintroebeling en temperatuur onderling onafhankelijk. Echter, toevoegen van vlokmiddel $FeCl_3$ heeft invloed op de zuurgraad (pH) van het water, en dus op het coagulatie mechanisme. Het is vraag of deze invloed verwaarloosbaar klein is of niet.

De pH en zoutconcentratie hebben invloed op het coagulatie mechanisme dat daadwerkelijk plaats vindt. Als ook het effect van de pH en/of de zoutconcentratie op de eindtroebeling wordt onderzocht, dan beschouwen we de procesvariabelen als onderling afhankelijk.

In activiteit 16 komen we hier nader op terug.

Bron 6. Het sommatieteken

In de wiskunde worden regelmatig grote aantallen getallen opgeteld. Om zo'n optelling of sommatie kort te noteren, wordt het sommatieteken Σ gebruikt. Dat is de Griekse hoofdletter Sigma, de S van sommeren. Het wordt ook wel het sigmateken genoemd.

Enkele voorbeelden:

$$\sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5$$

$$\sum_{k=1}^{10} k^2 = 1^2 + 2^2 + \dots + 10^2$$

$$\sum_{n=3}^6 a \cdot n = 3a + 4a + 5a + 6a$$

$$\sum_{j=1}^{10} 5 = 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5$$

Probeer het zelf maar. Schrijf zonder sommatieteken:

a. $\sum_{i=1}^{100} 3i + 1 =$

b. $\sum_{k=3}^5 k^2 =$

c. $(\sum_{k=3}^5 k)^2 =$

d. $a \cdot \sum_{n=3}^6 n =$

e. $\sum_{n=0}^{10} a =$

Schrijf met sommatieteken:

$$2 + 4 + 6 + 8 + 10 + 12 + 14 + 16 + 18 + 20$$

$$(3-7)^2 + (5-7)^2 + (7-7)^2 + (9-7)^2 + (11-7)^2$$

$$\frac{(3-7)^2 + (5-7)^2 + (7-7)^2 + (9-7)^2 + (11-7)^2}{5}$$

$$\frac{1}{5} \cdot ((3-7)^2 + (5-7)^2 + (7-7)^2 + (9-7)^2 + (11-7)^2)$$

De som van de derdemachten van de eerste tien natuurlijke getallen

De derdemacht van de som van de eerste tien natuurlijke getallen

Als het goed is, ben je nu enigszins gewend aan het sommatieteken. Het komt bij correlatie- en regressierekening nog veelvuldig van pas.

Bron 7. Meetniveau en meetnauwkeurigheid

In het waterzuiveringsexperiment heb je een aantal metingen gedaan. Bij het analyseren daarvan is het goed je rekenschap te geven van twee aspecten die daarbij een rol spelen, namelijk het meetniveau en de meetnauwkeurigheid.

Meetniveau

In de statistiek wordt onderscheid gemaakt tussen vier niveaus waarop een variabele kan worden gemeten: nominaal, ordinaal, interval en ratio meetniveau.

Wanneer een variabele op nominaal meetniveau wordt gemeten, bestaan de waarnemingen uit namen of woorden. Denk bijvoorbeeld aan de naam van de plaats waar je watermonsters hebt genomen of aan de naam van een proefpersoon bij een inspanningstest. Deze variabelen kun je niet ordenen en je kunt er ook niet mee rekenen.

Wanneer een variabele op ordinaal meetniveau wordt gemeten, kunnen de waarnemingen op volgorde worden gezet; ze zijn op grootte te ordenen. Als de antwoordmogelijkheden bij een enquête bijvoorbeeld bestaan uit de vijf alternatieven sterk oneens / beetje oneens / neutraal / beetje eens / sterk eens, dan zit er een duidelijke ordening in de waarnemingen. Je kunt de verschillen echter niet kwantificeren. Het is niet zinvol om te zeggen: 'sterk eens is 3 meer dan beetje eens'.

Waarnemingen van nominaal of ordinaal niveau zijn dus geen getallen waarmee je kunt rekenen. Bij variabelen die gemeten worden op interval meetniveau geldt dat wel. Denk bijvoorbeeld aan het tijdstip zoals je dat afleest op een klok. Je kunt tijdstippen op volgorde zetten, maar ook het verschil kwantificeren: het tijdsverschil tussen 21:00 en 23:00 uur is even groot als dat van 6:40 en 8:40. Waarnemingen gemeten op intervalniveau zijn dus getallen die geordend kunnen worden, maar waarbij ook de intervallen tussen de waarden betekenis hebben.

Datzelfde geldt voor een variabele lichaamslengte. Een leerling van 160 cm is niet alleen langer dan iemand van 155 cm, maar ook weten we hoeveel langer, namelijk 5 cm. Bovendien kunnen we zeggen dat een leerling van 165 cm lang anderhalf keer zo lang is als een leerling van 110 cm. De verhouding tussen de twee lengtes is $3 : 2$. In het geval dat verhoudingen tussen waarnemingen betekenis hebben, is sprake van ratio meetniveau. Voor de tijdstippen die worden afgelezen op een klok hebben verhoudingen geen betekenis. Rekenkundig is 12 twee keer zo veel als 6, maar de uitspraak '12 uur is twee keer zo laat als 6 uur' heeft geen betekenis. Temperatuur in graden Celsius wordt gemeten op intervalniveau, omdat 20 graden Celsius energetisch gesproken niet twee maal zo warm is als 10 graden Celsius. Bij temperatuurmeting in graden Kelvin geldt dit laatste wel, en dan is temperatuur dus op ratio meetniveau gemeten.

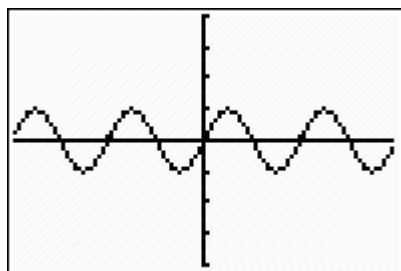
MEETNIVEAU	Betekenisvolle ordening?	Betekenisvolle verschillen?	Betekenisvolle verhoudingen?
Nominaal	Nee	Nee	Nee
Ordinaal	Ja	Nee	Nee
Interval	Ja	Ja	Nee
Ratio	Ja	Ja	Ja

De bovenstaande tabel vat de verschillen tussen de vier meetniveaus samen. Het belangrijkste onderscheid is dat tussen de nominaal enerzijds, waar geen sprake is van getallen, en interval en ratio anderzijds, waar wel met getallen gerekend kan worden. In veel toepassingen uit de bètavakken is het meetniveau interval of ratio.

Meetnauwkeurigheid

Bij het meten speelt vanzelfsprekend de meetnauwkeurigheid een rol. Dat is niet alleen een kwestie van de precisie van je meetapparatuur; als het gaat om processen in de loop van de tijd, speelt ook de frequentie van meting een rol. Uit onderstaande opgave blijkt dat de intervalbemonstering van processen die continu verlopen tot een verkeerd beeld kan leiden.

Op een grafische rekenmachine van het type TI84 ziet de grafiek van $x \rightarrow \sin(50x)$ er in het Zoom Trig kijkvenster zo uit:



Als je dit plaatje op je GR namaakt en vervolgens inzoomt op de grafiek, zul je zien dat die veel meer schommelingen heeft dan in het oorspronkelijke beeld het geval lijkt te zijn. Omdat de GR, met het beperkte aantal pixels, maar 'af en toe' de sinuswaarden 'meet', mis je in het oorspronkelijke plaatje veel van de fluctuatie.

Als je een chemische eigenschap van een voorwerp of van een stof enige keren meet, dan zie je dat je niet steeds dezelfde meetuitkomst verkrijgt. De meetwaarden vertonen spreiding. Waarschijnlijk heb je dat zelf ook ondervonden tijdens de uitvoering van de experimenten vlokbehandeling. Een dergelijke spreiding kan worden veroorzaakt door variatie in de uitvoering van het experiment. Tijdens het experiment vlokbehandeling is het bijvoorbeeld mogelijk dat:

- de metingen door verschillende personen zijn uitgevoerd;
- de vloeistofhoeveelheden tussen twee schaalstreepjes op een maatcilinder of pipet *geschat* zijn;
- de Fe^{3+} oplossing snel of langzaam is toegevoegd;
- er misschien af en toe een halve druppel van de Fe^{3+} oplossing aan de pipet blijft hangen;
- er niet altijd even hard of zacht is geroerd;
- er niet altijd even lang is geroerd;
- er verschillen zijn tussen de gebruikte roervlootjes;
- de cuvet voor meting van de eindtroebelheid vervuild is met vet- of stofdeeltjes;
- het bekerglas waarin het experiment uitgevoerd is niet schoon was.

Wanneer er een ruwere meting van de eindtroebelheid had plaats gevonden, zouden er misschien wel steeds dezelfde waarden gemeten zijn. Dat is geen goed teken: als er bij herhaalde meting van een continue grootheid steeds dezelfde uitkomst wordt gevonden, dan is er waarschijnlijk te grof gemeten! Samengevat is spreiding in meetuitkomsten niet te vermijden, en zijn de oorzaken van deze spreiding velerlei en in het algemeen niet allemaal te achterhalen.

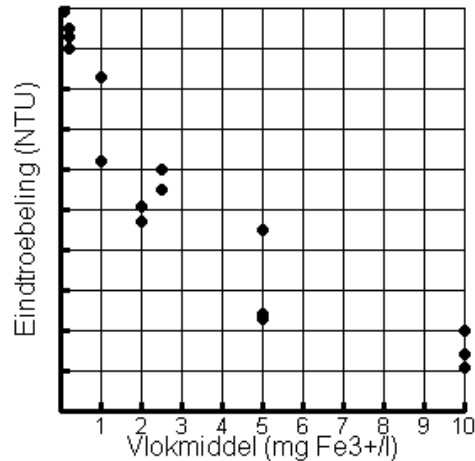
De aanwezigheid van spreiding in meetuitkomsten is een zeer normale zaak. De oorzaak ligt in oncontroleerbare fluctuaties in de meetapparatuur, de meetomstandigheden, de monsters, de waarnemer. Dat wil dus ook zeggen dat we gewoonlijk spreken over relatief kleine schommelingen rond de gemiddelde waarde. Het komt echter helaas toch af en toe voor dat een meetuitkomst totaal onverwacht en onverklaarbaar veel afwijkt van de overige meetwaarden. In zo'n geval spreken we van een uitschieter. Alle vreemde afwijkingen ten gevolge van geknoei, onervarenheid, onoplettendheid en niet traceerbare vreemde uitschieters in de meetuitkomsten worden uit de reeks metingen geschrapt. Chemici gaan echter eerst zorgvuldig en eerlijk na of de meetuitkomst toch niet te "herstellen" is, d.w.z. de oorzaak van de afwijking te achterhalen en de meetuitkomst te corrigeren is. In een laboratoriumjournaal vermelden chemici deze afwijkingen doorgaans wel, omdat je van dit soort 'fouten' vaak veel leert en je wilt voorkomen dat je dezelfde 'fout' nogmaals maakt.

Bron 8. Centrum en spreiding in beeld

In deze bron wordt beschreven hoe je deze gegevens goed in beeld kunt brengen. We richten ons op het waterzuiveringsexperiment waaraan je zelf een aantal metingen hebt gedaan, en letten daarbij alleen op de dosis vlokmiddel Fe^{3+} en de eindtroebeling van het water. Als meetgegevens nemen we de data uit de volgende tabel, die afkomstig zijn van experimenten op een andere school. We nemen aan dat de waarden van de andere procesvariabelen constant zijn gebleven tijdens deze metingen.

Fe^{3+} (mg/L)	E (NTU)	Fe^{3+} (mg/L)	E (NTU)
0,1	99	2,5	55
0,2	90	2,5	55
0,2	95	2,5	60
0,2	112	5	23
0,2	93	5	24
1	83	5	24
1	83	5	45
1	119	10	11
1	62	10	14
2	47	10	20
2	51		

Hierboven staan de gegevens van 21 metingen in een tabel. Hieronder zijn ze weergegeven in een zogenaamd spreidingsdiagram, in het Engels scatterplot. Zo'n scatterplot geeft een indruk van de ligging van de punten, die samen een 'puntenwolk' vormen. Voorwaarde voor het maken van een scatterplot is dat de metingen van interval of ratio meetniveau zijn (zie bron @@). Meestal wordt de onafhankelijke variabele (hier: dosis vlokmiddel) op de horizontale as gezet en de afhankelijke (hier: eindtroebeling) op de verticale.



Aan een spreidingsdiagram kun je in een oogopslag twee belangrijke kenmerken van je dataset of puntenwolk zien, namelijk het centrum en de spreiding.

Het centrum van de puntenwolk is als het ware het zwaartepunt. De eerste coördinaat daarvan is de gemiddelde waarde van het vlokmiddel, en de tweede is de gemiddelde waarde van de eindtroebelings. Als we V en E als afkortingen voor Vlokmiddel en Eindtroebelings gebruiken, en de streep erboven voor de gemiddelden, dan zijn de coördinaten van het zwaartepunt:

$$(\bar{V}, \bar{E}) = \left(\frac{1}{21} \sum_{i=1}^{21} V_i, \frac{1}{21} \sum_{i=1}^{21} E_i \right)$$

Dan de spreiding, die aangeeft 'hoe ver de punten gespreid liggen'. De spreiding van de puntenwolk wordt net als bij de gemiddelden apart voor de horizontale en de verticale richting bepaald. Hoewel er verschillende maten voor spreiding zijn, is de standaardafwijking het meest gangbaar. Deze wordt ook wel standaarddeviatie genoemd en genoteerd met de Griekse kleine letter s : σ . De σ , bijvoorbeeld die van V , bereken je zo:

In woorden:	In formule:
Neem voor elke waarneming de afwijking tot het gemiddelde	$V_i - \bar{V}$
Kwadrateer elke afwijking	$(V_i - \bar{V})^2$
Neem het gemiddelde van die kwadraten. Dit heet de variantie.	$\frac{1}{21} \sum_{i=1}^{21} (V_i - \bar{V})^2$
Neem van dit gemiddelde de wortel. De uitkomst is de standaardafwijking of standaarddeviatie, aangeduid met σ .	$\sigma = \sqrt{\frac{1}{21} \sum_{i=1}^{21} (V_i - \bar{V})^2}$

Nu zul je misschien zeggen: waarom eerst kwadrateren en dan op het laatst weer de wortel trekken? Het kwadrateren in de tweede regel gebeurt om te voorkomen dat de negatieve afwijkingen ten opzichte van het gemiddelde de positieve afwijkingen opheffen. Het worteltrekken op het laatst zorgt ervoor dat de eenheid van σ gelijk is aan die van V zelf. En dat worteltrekken maakt het eerdere kwadrateren niet ongedaan, omdat je tussendoor nog hebt opgeteld en gedeeld.

In praktijk wordt een standaardafwijking niet met de hand berekend, maar met een of ander rekengereedschap. Hieronder zie je bijvoorbeeld de uitvoer van de TI 84 en van Excel, waaruit je de standaardafwijking van V kunt aflezen. Die van Excel wijkt iets af van die van de GR, omdat Excel deelt door 20 in plaats van door 21. Zie voor een verdere bespreking hiervan bron 14.

```

1-Var Stats
x̄=3.161904762
Σx=66.4
Σx²=430.92
Sx=3.323924817
σx=3.243818472
↓n=21

```

AVERAGE(A1:A21)	3,161904762
STDEV(A1:A21)	3,323924817

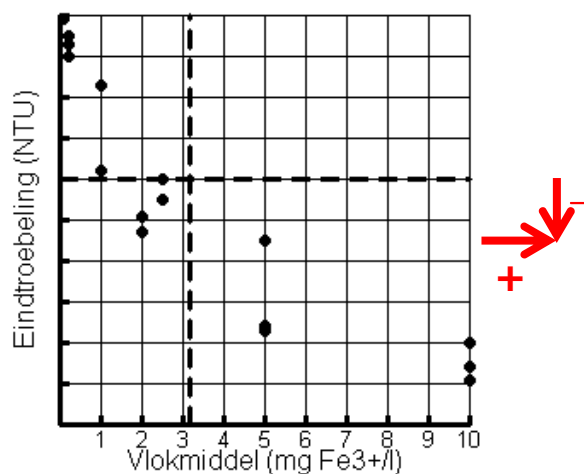
Zorg ervoor dat je zelf ook een standaardafwijking kunt uitrekenen, en bereken zelf bij wijze van oefening bijvoorbeeld de standaardafwijking van E. Als het goed is, zie je dat de standaardafwijking van E ongeveer 10 keer zo groot is als die van V. Dat lijkt wel redelijk, als je naar de gegevens in de tabel kijkt.

Bron 9. Samenhang en correlatie

In de situatie van de waterzuivering is de vraag in hoeverre de eindtroebeling afhangt van andere variabelen, zoals bijvoorbeeld de dosis vlokmiddel. Je hebt een aantal metingen gedaan en de vraag is hoe je de veronderstelde afhankelijkheid daaruit kunt onderbouwen of weerleggen. Bij het beantwoorden van deze vraag ga je statistische methoden gebruiken, en met name correlatie- en regressierekening. Dit zijn methoden om verbanden tussen grootheden te kwantificeren en om na te gaan hoe grootheden een andere grootheid beïnvloeden. De grootheden die kunnen worden ingesteld worden de stuurvariabelen of onafhankelijke variabelen genoemd. In ons geval is bijvoorbeeld de dosis vlokmiddel een stuurvariabele. De grootheid die hierdoor wordt beïnvloed heet de afhankelijke variabele, en dat is in ons geval de eindtroebeling. Daarnaast kan nog een aantal externe factoren of procesvariabelen een rol spelen. Als er één onafhankelijke variabele is, wordt die in een grafiek op de horizontale as uitgezet en de afhankelijke variabele op de verticale.

Er bestaat een verschil tussen een kwalitatieve en een kwantitatieve beschrijving van samenhang of afhankelijkheid. Een kwalitatieve beschrijving is bijvoorbeeld: "Hoe groter de kracht die op een veer werkt, hoe groter de uitrekking van de veer". Of: "Hoe hoger de temperatuur, hoe sneller diffusieprocessen verlopen". Deze beschrijvingen zijn kwalitatief, omdat ze iets zeggen over hoe de afhankelijke grootheid verandert. Een kwantitatieve beschrijving zegt ook iets over hoeveel groter of kleiner de afhankelijke grootheid precies wordt. Een kwantitatieve beschrijving is bijvoorbeeld: "Als de kracht die op de veer werkt twee keer zo groot wordt, wordt de uitrekking van de veer ook twee keer zo groot". Of: "Als de temperatuur 10 graden stijgt, verlopen diffusieprocessen twee keer zo snel."

Het spreidingsdiagram bij de gegevens uit bron 7 geeft behalve informatie over zwaartepunt en spreiding ook een indruk van de samenhang van twee variabelen. In kwalitatieve zin wekt het spreidingsdiagram de indruk dat grote waarden van V samen gaan met kleine waarden van E . Anders gezegd: hoe groter V , hoe kleiner E . Dat noemen we een negatieve vorm van samenhang. De vraag is nu hoe deze kwalitatieve uitspraak gekwantificeerd kan worden.



Hierboven zijn in het spreidingsdiagram een horizontale en een verticale lijn getrokken door het zwaartepunt van de puntenwolk. Die verdelen het vlak in vier kwanten, ofwel kwadranten. Punten die in de kwadranten rechtsboven en linksonder liggen, wijzen op een positieve samenhang; punten linksboven en rechtsonder op een negatieve. Daarbij maakt het natuurlijk ook uit hoe ver

de punten van de twee lijnen door het zwaartepunt afliggen. Voor het punt (5, 45) zijn de afwijkingen in de figuur aangegeven met pijlen.

Om de spreiding te kwantificeren, vermenigvuldigen we nu de afwijkingen van V en E ten opzichte van de lijnen door het zwaartepunt met elkaar. Immers, rechtsboven en linksonder is dat product positief, en linksboven/rechtsonder negatief. Bovendien telt een punt dat ver van het zwaartepunt ligt sterk mee, omdat de afstanden en hun product groot zijn. Op die manier berekenen we de zogeheten covariantie van V en E.

In woorden:	In formule:
Bepaal voor elke waarneming de horizontale en de verticale afwijking	$V_i - \bar{V}, E_i - \bar{E}$
Vermenigvuldig deze afwijkingen met elkaar	$(V_i - \bar{V}) \cdot (E_i - \bar{E})$
Neem het gemiddelde van deze producten. Dit heet de covariantie	$\frac{1}{21} \sum_{i=1}^{21} (V_i - \bar{V}) \cdot (E_i - \bar{E})$

De covariantie is dus een getal dat groot wordt bij een positieve samenhang, klein bij een negatieve samenhang, en dicht bij nul ligt als er geen samenhang is. Maar wat is 'groot' en wat is 'klein'? Dat is moeilijk te zeggen. Dat komt door het nadeel dat de covariantie heeft: als je van eenheid verandert, bijvoorbeeld je meet V niet in mg/L maar in g/l, dan worden alle waarden van V 1000 keer zo klein, en daarmee de covariantie van V en E ook. Dat terwijl we vinden dat de maat van samenhang onafhankelijk moet zijn van de gekozen eenheid.

Om dat op te lossen, wordt de covariantie gedeeld door de standaardafwijking van V en door die van E. Dat geeft de zogeheten correlatiecoëfficiënt R:

$$R_{V,E} = \frac{\text{cov}_{V,E}}{\sigma_V \cdot \sigma_E} = \frac{\frac{1}{21} \sum_{i=1}^{21} (V_i - \bar{V}) \cdot (E_i - \bar{E})}{\sigma_V \cdot \sigma_E}$$

Net als bij de standaarddeviatie en bij de covariantie zul je een correlatiecoëfficiënt zelden met de hand berekenen. Zorg ervoor dat je weet hoe je dit met de grafische rekenmachine of met Excel kunt doen. Zie eventueel bronnen @@ en @@ en ga bijvoorbeeld na dat $R_{V,E} = -0.86$.

De correlatiecoëfficiënt R heeft de volgende eigenschappen, die we hier niet bewijzen:

R is symmetrisch: $R_{V,E} = R_{E,V}$

R is minimaal gelijk aan -1 en maximaal 1

R = 1 precies wanneer alle punten in de puntenwolk op dezelfde stijgende lijn liggen

R = -1 precies wanneer alle punten in de puntenwolk op dezelfde dalende lijn liggen

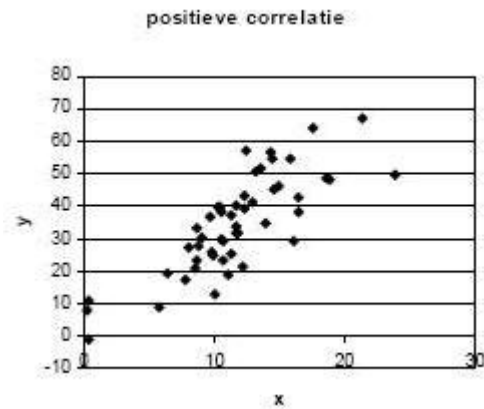
De correlatiecoëfficiënt kwantificeert dus de mate van samenhang tussen twee variabelen, zoals gemeten en weergegeven in een spreidingsdiagram. De sterkte van R wordt als volgt geïnterpreteerd:

Correlatie...	Negatief	Positief
... is verwaarloosbaar	$-0,1 < R \leq 0$	$0 \leq R < 0,1$
... is zwak	$-0,3 < R < -0,1$	$0,1 < R < 0,3$
... is matig	$-0,5 < R < -0,3$	$0,3 < R < 0,5$
... is sterk	$-1 \leq R < -0,5$	$0,5 < R \leq 1$

Hieronder drie plaatjes van positieve, negatieve en geen correlatie.

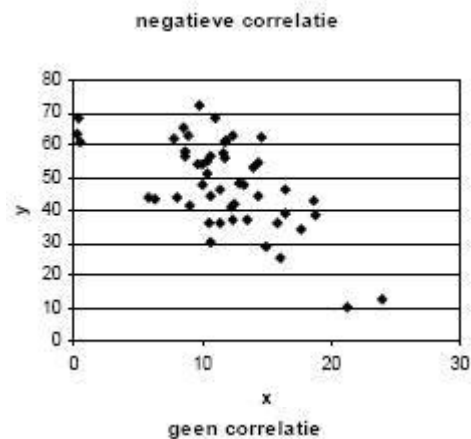
Positieve correlatie

Als de correlatiecoëfficiënt positief is, is er sprake van positieve correlatie tussen de twee gemeten grootheden. Hoewel de significantie van R mede van de steekproefomvang afhangt, kun je als vuistregel zeggen dat $R > 0.5$ op een behoorlijke tot sterke positieve correlatie wijst. Grote waarden van de ene variabele gaan dan gepaard met grote waarden van de andere. In de puntenwolk zie je een positieve trend, zoals hiernaast afgebeeld.



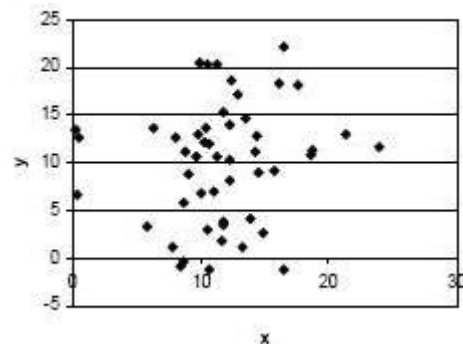
Negatieve correlatie

Als de correlatiecoëfficiënt negatief is, is er sprake van negatieve correlatie tussen de twee gemeten grootheden. Hoewel de significantie van R mede van de steekproefomvang afhangt, kun je als vuistregel zeggen dat $R < -0.5$ op een behoorlijke tot sterke negatieve correlatie wijst. Grote waarden van de ene variabele gaan dan gepaard met kleine waarden van de andere, en andersom. In de puntenwolk zie je een negatieve trend, zoals hiernaast afgebeeld.



Geen correlatie

Als de correlatiecoëfficiënt dicht bij 0 ligt, is er geen sprake van correlatie tussen de twee gemeten grootheden. Anders dan bij positieve of negatieve correlatie is in dit geval in de puntenwolk geen trend te ontdekken, zoals hiernaast afgebeeld.



Als slot nog de opmerking dat je aan de waarde van de correlatiecoëfficiënt in het algemeen geen oorzaak-gevolg conclusies kunt verbinden. Wanneer twee grootheden positief correleren, is het niet vanzelfsprekend dat één van de twee de verandering van de andere veroorzaakt: er is bijvoorbeeld een positieve correlatie tussen het aantal ooievaars dat per maand wordt waargenomen en het geboortecijfer van die maand. Maar om nu te spreken van een oorzakelijk verband

Bron 10. Lineaire regressie

In bron 8 is de correlatiecoëfficiënt ingevoerd als maat voor lineaire samenhang tussen twee gemeten variabelen. Omdat de formule van de correlatiecoëfficiënt symmetrisch is in de twee variabelen, maakt het voor de correlatierekening niet uit welke variabele je op de horizontale en welke je op de verticale as uitzet. Omdat we vermoeden dat de dosis vlokmiddel wel van invloed is op de eindtroebeling en zeker weten dat andersom niet het geval is, staat het vlokmiddel op de horizontale as en de eindtroebeling op de verticale as.

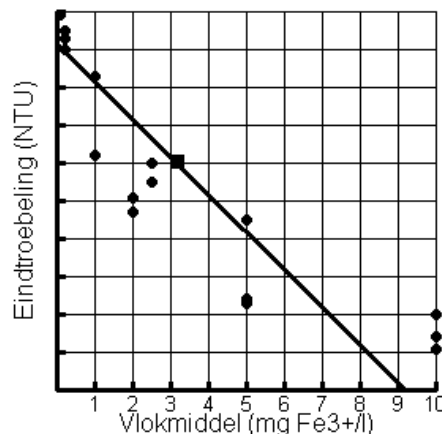
In bron 8 staat dat $R_{V,E} = -0.86$, dus dat er een sterk negatieve correlatie is tussen V en E. Omdat E van V afhangt, zoeken we nu een lineair verband waarmee we uit de waarde van V die van E kunnen voorspellen. In het spreidingsdiagram komt dat erop neer dat we de 'best passende' lijn door de puntenwolk zoeken. De vraag is welke lijn dat is.

Om deze vraag te beantwoorden nemen we aan dat de best passende lijn door het zwaartepunt van de puntenwolk gaat. Als we de coördinaten daarvan voorstellen als \bar{V} en \bar{E} , dan hebben lijnen die aan die voorwaarde voldoen de volgende vergelijking:

$$y = c \cdot (x - \bar{V}) + \bar{E}$$

Ga maar na, als $x = \bar{V}$, dan geldt dat $y = \bar{E}$.

In het spreidingsdiagram hieronder is zo'n lijn door het zwaartepunt getekend. Het vierkantje geeft het zwaartepunt weer.



In de vergelijking van de lijn stelt c de helling voor. In het plaatje is die gelijk aan -10 gekozen. Maar misschien geeft een andere waarde van c wel een beter model voor de afhankelijkheidsrelatie. Dat zou betekenen dat we de lijn vanuit het zwaartepunt een beetje moeten draaien. De vraag is dus welke waarde van c de afhankelijkheid het beste weergeeft, en dus het beste past bij de puntenwolk.

Zonder bewijs geven we de formule voor de best passende lijn:

$$y = \frac{\sigma_E}{\sigma_V} R_{V,E} \cdot (x - \bar{V}) + \bar{E}$$

Hierin hebben de symbolen de volgende betekenis:

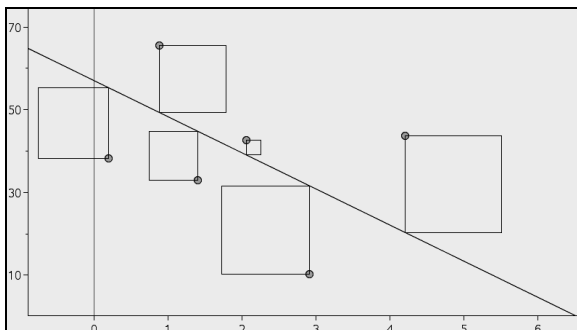
- x: de ingestelde waarde van V
 y: de door het lineaire model voorspelde waarde van E
 \bar{V} : het gemiddelde van de ingestelde waarden van V in het experiment
 \bar{E} : het gemiddelde van de gemeten waarden van E in het experiment
 σ_V : de standaardafwijking van de ingestelde waarden van V in het experiment
 σ_E : de standaardafwijking van de gemeten waarden van E in het experiment
 $R_{V,E}$: de correlatiecoëfficiënt van V en E zoals berekend uit de experimentele waarden

De hierboven beschreven lijn heet de regressielijn van E op V.

Hoe goed past die lijn nu eigenlijk en hoe weten we dat? De regressielijn geeft de voorspelde E-waarde bij een ingestelde dosis V van vlokmiddel. Deze voorspelling bij een V-waarde van V_i is dus $y(V_i)$. De fout die bij deze voorspelling gemaakt wordt, zeg F_i , is het verschil met de werkelijke waarde E_i . Kort gezegd:

$$F_i = E_i - y(V_i)$$

Voor punten die boven de lijn liggen, is deze fout positief, en voor punten eronder negatief. Om te voorkomen dat deze positieve en negatieve fouten elkaar opheffen bij optelling, nemen we als criterium voor de best passende lijn de eis dat de som van de kwadraten van de fouten minimaal is. In de figuur hieronder worden de kwadraten van de fouten voor enkele punten als oppervlakten voorgesteld. Het idee is nu dat we de lijn zoeken met de kleinste totale oppervlakte, de kleinste kwadraatfout.



Als we de gemiddelde totale kwadraatfout K^2 noemen, dan geldt:

$$K^2 = \frac{1}{21} \sum_{i=1}^{21} F_i^2 = \frac{1}{21} \sum_{i=1}^{21} (E_i - y(V_i))^2$$

De waarde van K^2 is minimaal als de lijn de vergelijking heeft die op de vorige pagina is gegeven.

Samengevat geeft de correlatiecoëfficiënt geeft aan of er lineaire samenhang is tussen twee variabelen. Als we ervan uitgaan dat E afhangt van V, weten we ook welke lijn het best bij de puntenwolk past: dat is de regressielijn met de vergelijking

$$y = \frac{\sigma_E}{\sigma_V} R_{V,E} \cdot (x - \bar{V}) + \bar{E}$$

De gemiddelde kwadraatfout is dan gelijk aan

$$K^2 = \frac{1}{21} \sum_{i=1}^{21} F_i^2 = \frac{1}{21} \sum_{i=1}^{21} (E_i - y(V_i))^2$$

Tussen de gemiddelde kwadraatfout van de best passende lijn en de correlatiecoëfficiënt bestaat het volgende verband:

$$R^2 = 1 - \frac{K^2}{\sigma_E^2}$$

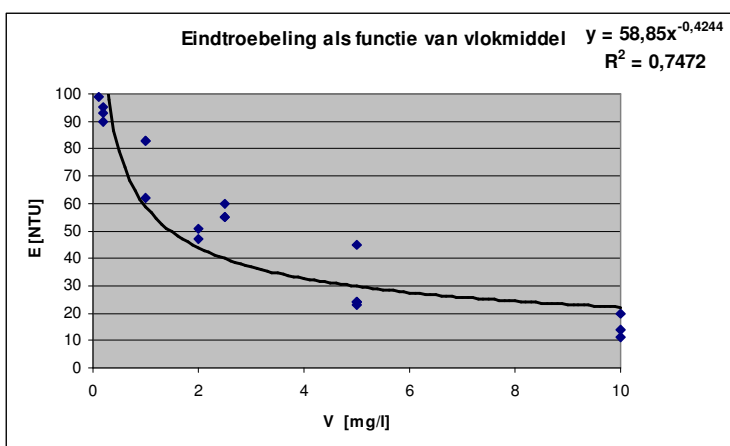
Als R dus dicht bij 1 of -1 ligt, zal K² dicht bij 0 liggen. Dat is mooi, want bij een goed passende lijn moet dat allebei het geval zijn!

Bron 11. Machtsregressie

In bron 9 zijn we ervan uitgegaan dat de eindtroebeling lineair afhangt van de dosis vlokmiddel. In veel situaties is er reden om van een niet-rechthoekig verband uit te gaan (zie bijvoorbeeld bron 11). Een voorbeeld van een niet-rechthoekig verband is het machtsverband. Het machtsverband heeft de volgende algemene vorm:

$$E = a \cdot V^b$$

De vraag is dan, net als bij de lineaire regressie, welke waarden van a en b de best passende kromme geven. Om hierop antwoord te geven, wordt in het algemeen ook de kleinste kwadraten methode gebruikt. Omdat dat met de hand moeilijk uitvoerbaar is, nemen we weer onze toevlucht tot hulpmiddelen zoals de grafische rekenmachine of Excel. In bronnen 15 – 17 staat hoe je die kunt gebruiken.



In de afbeelding hierboven zie je het resultaat van een machtsregressie op de 21 gegevens uit bron 7, zoals berekend door Excel. Het best passende machtsverband tussen vlokmiddel V en eindtroebeling E is dus op basis van deze metingen:

$$E = 58,85 \cdot V^{-0,4244}$$

Deze formule kun je gebruiken om bijvoorbeeld de eindtroebeling te voorspellen bij een dosis vlokmiddel van 4 mg/l.

Excel geeft als uitvoer van de machtsregressie ook nog 'R2 = 0,7472'. Hoe moeten we dat opvatten? Deze R2 wordt berekend via de in bron 9 gegeven formule

$$R^2 = 1 - \frac{K^2}{\sigma_E^2}$$

$$K^2 = \frac{1}{n} \sum_{i=1}^n (E_i - a \cdot V_i^b)^2$$

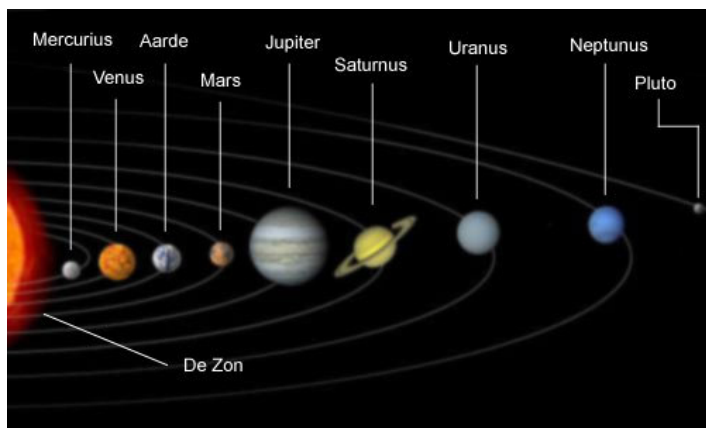
Hierin is K2 de gemiddelde kwadraatfout:

σ_E^2 is het kwadraat van de standaardafwijking van E.

Als de waarde van R^2 dicht bij 0 ligt, geeft dit aan dat de kwadraatfout groot is. Een kleine kwadraatfout geeft aan dat R^2 dicht bij 1 ligt. Net zoals bij de correlatiecoëfficiënt dus, maar er zit een addertje onder het gras: omdat er geen sprake is van lineaire samenhang, heeft de correlatiecoëfficiënt geen betekenis. Bij niet-lineaire regressie spreken we dus niet van R maar van R^2 , en dat wordt wel de coëfficiënt van determinatie genoemd.

Bron 12. De derde wet van Kepler als voorbeeld van machtsregressie

De machtsregressie wordt in deze bron geïllustreerd aan de hand van een voorbeeld uit de sterrenkunde.



Onderstaande tabel geeft van elke planeet van ons zonnestelsel de omlooptijd en de afstand tot de zon. Pluto wordt tegenwoordig niet meer als planeet beschouwd, dus die laten we weg.

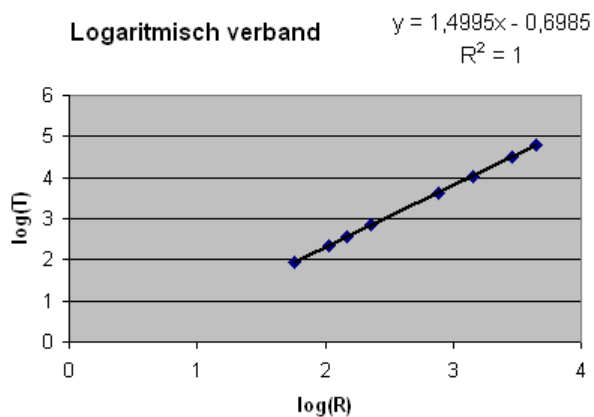
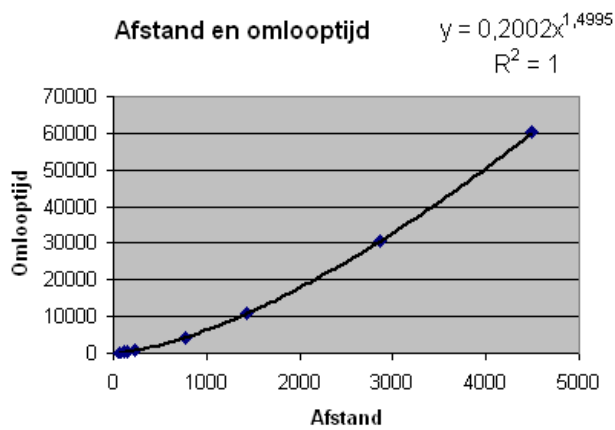
Planeet	Gemiddelde afstand tot de zon R (x 10 ⁶ km)	Omlooptijd T (dagen)
Mercurius	57.9	88
Venus	108.2	225
Aarde	149.6	365
Mars	227.7	687
Jupiter	778.3	4329
Saturnus	1427.0	10753
Uranus	2870.0	30660
Neptunus	4497.0	60150

Volgens de derde wet van Kepler, geformuleerd in 1619, is het kwadraat van de omlooptijd van een planeet evenredig met de derde macht van haar gemiddelde afstand tot de zon. Er zijn dus in dit geval theoretische redenen om niet een lineair maar een machtsmodel te kiezen.

In de figuren op de volgende pagina zie je hoe de machtsregressie in Excel is uitgevoerd. De eerste grafiek geeft het machtsmodel. We herkennen daarin de macht 1,4995 als de 3/2 die je op grond van de wet van Kepler zou verwachten. De coëfficiënt van determinatie R² is afgerond 1, dus er is sprake van een vrijwel perfecte 'fit'.

In de Excel tabel zijn vervolgens de logaritmen van afstand en omlooptijd genomen. De onderste grafiek laat zien dat daar een lineair model vrijwel perfect van toepassing is. Weer herkennen we de 1,4995. De constante -0,69 is de logaritme van 0,2.

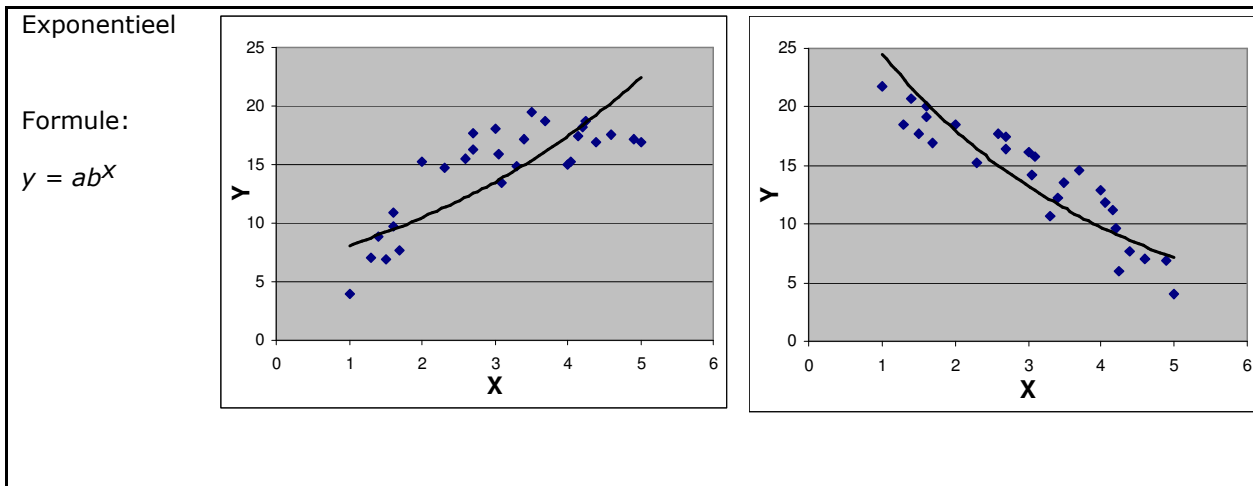
Planeet	Afstand R	Tijd T	log(R)	log(T)
Mercurius	57,9	88	1,76268	1,94448
Venus	108,2	225	2,03423	2,35218
Aarde	149,6	365	2,17493	2,56229
Mars	227,7	687	2,35736	2,83696
Jupiter	778,3	4329	2,89115	3,63639
Saturnus	1427	10753	3,15442	4,03153
Uranus	2870	30660	3,45788	4,48657
Neptunus	4497	60150	3,65292	4,77924



Bron 13. Andere niet-lineaire regressiemodellen

Behalve het lineaire model en het machtsmodel (zie bron 9, 10 en 11) zijn er nog andere vormen van regressie. Hieronder vind je een aantal veelgebruikte regressiemodellen.

Model	Positief	Negatief
Lineair Formule: $y = ax + b$		
Logaritmisch Formule: $y = a \log(x) + b$		
Macht Formule: $y = ax^b$		



Net als bij lineaire regressie wordt de vergelijking van de niet-lineaire regressiekromme berekend aan de hand van het kleinstekwadraten criterium. De manier waarop dat gebeurt, is in het algemeen weinig doorzichtig en voert voor dit moment te ver.

Als maat voor de passendheid van de regressiekromme bij de puntenwolk wordt weer de coëfficiënt van determinatie gehanteerd:

$$R^2 = 1 - \frac{K^2}{\sigma_E^2}$$

De waarde van R^2 geeft weer hoe goed de variatie van de afhankelijke variabele verklaard kan worden uit de verandering van de onafhankelijke variabele. Als $R^2 = 1$ gaat de kromme precies door de meetpunten en is de 'fit' perfect. Bij een waarde van 0 is er geen verband van de veronderstelde vorm tussen de afhankelijke en onafhankelijke variabele. Hoe dichterbij 1, hoe beter de formule het verband beschrijft. $R^2 > 0.8$ geldt als een goede fit, boven de 0.9 als zeer goed.

Een waarschuwing is hier op zijn plaats. De waarde van R^2 is een indicatie van de 'goodness of fit', maar de betrouwbaarheid daarvan hangt af van het aantal metingen of waarnemingen. Je kunt je voorstellen dat je bij een klein aantal metingen wel een goede fit kunt vinden, maar dat het toeval daarin zo'n grote rol heeft gespeeld dat de fit bij een herhaald experiment minder goed is. Daarom is eenzelfde waarde van R^2 meer significant naarmate het aantal metingen groter is.

Bron 14. Meervoudige regressie

In bronnen 8 – 12 is het uitgangspunt steeds geweest dat de eindtroebeling van het water alleen afhangt van de dosis vlokmiddel Fe³⁺. Dat is vermoedelijk een onterechte vereenvoudiging van de situatie: wellicht hangt de eindtroebeling ook af van procesvariabelen zoals begintroebeling, zuurgraad, temperatuur en zoutconcentratie van het water. De vraag is hoe we de invloed van deze verschillende variabelen in één formule kunnen weergeven. Omdat er sprake is van meer dan één onafhankelijke variabele, heet dit meervoudige of multiple regressie.

Het idee is dat we de enkelvoudige regressiemodellen uit de vorige bronnen combineren tot één meervoudig regressiemodel. Voorwaarde hiervoor is allereerst dat het om dezelfde afhankelijke variabele gaat, in ons geval de eindtroebeling. Als voorbeeld gaan we een meervoudig regressiemodel ontwikkelen van eindtroebeling E op vlokmiddel V en begintroebeling B . Daarbij gaan we uit van onderstaande meetgegevens.

Vlokmiddel V (mg/L)	Begintroebeling B (NTU)	Eindtroebeling E (NTU)
0,1	160	99
0,2	160	93
1	77	35
1	105	46
1	158	70
1	185	106
1	210	132
2	160	47
2	160	51
2,5	160	55
2,5	160	55
5	160	45
10	160	11

Een meervoudig lineair model

Het eerste meervoudige regressiemodel is het lineaire model, dat de volgende vorm heeft:

$$E = m_1 \cdot V + m_2 \cdot B + b$$

De waarden van de coëfficiënten m_1 , m_2 en b kunnen worden berekend met Excel op de manier zoals beschreven in bron 15. Dat geeft de volgende, wat moeilijk leesbare uitvoer:

0,67699	-8,71086	-20,3002	#N/A
0,13048	1,60058	20,6662	#N/A
0,83981	14,6687	#N/A	#N/A
26,2125	10	#N/A	#N/A

In de eerste kolom staan achtereenvolgens m_2 , m_1 en b . Het model wordt dus:

$$E = -8,7 \cdot V + 0,7 \cdot B - 20,3$$

In de derde rij staat de waarde van R^2 : 0,84. Dat is een hoge waarde, die wijst op een goede fit.

De andere waarden in de tabel doen er op dit moment niet toe.

Een tweede, minder gangbare manier om een meervoudig lineair regressiemodel te maken is door de twee enkelvoudige modellen 'in elkaar te vlechten' of, anders gezegd, 'in elkaar te knutselen'.

De resultaten van de enkelvoudige regressies zijn:

$$E = -8,1 \cdot V + 83,2 \quad \text{en} \quad E = 0,6 \cdot B - 31,3$$

We combineren die twee, maar wel in dezelfde verhouding:

$$E = a \cdot (-8,1 \cdot V + 0,6 \cdot B) + b$$

en de vraag is weer hoe groot a en b zijn. Dit kun je berekenen door in Excel een nieuwe kolom, zeg Z , te definiëren waarin $-8,1 \cdot V + 0,6 \cdot B$ staat en vervolgens een enkelvoudige lineaire regressie van E op Z uit te voeren.

Deze manier verdient niet de voorkeur, maar is wel aardig om uit te proberen omdat je dan wat meer ziet wat je doet dan wanneer je de 'Black-Box-methode' van Excel gebruikt.

Een meervoudig machtsmodel

Een vergelijkbare procedure kan worden gevolgd bij het combineren van twee enkelvoudige machtsmodellen. We zoeken een meervoudig machtsmodel van de vorm

$$E = a \cdot V^{b_1} \cdot B^{b_2}$$

Omdat Excel hiervoor geen standaardprocedure heeft, moeten we wel de 'knutselmethode' volgen van het combineren van de twee enkelvoudige machtsmodellen. Die luiden:

$$E = 61,2 \cdot V^{-0,37} \quad \text{en} \quad E = 0,4 \cdot B^{0,98}$$

We vermenigvuldigen V en B nu, elk met hun eigen macht, en noemen het resultaat weer Z :

$$Z = V^{-0,37} \cdot B^{0,98}$$

De waarde van a vinden we door een lineaire regressie van E op Z uit te voeren, waarbij we als extra eis stellen dat de lijn door de oorsprong gaat.

Tot slot van deze bron nog enkele overwegingen bij deze meervoudige regressiemodellen.

Ten eerste merken we op dat er vele parallellen zijn met de enkelvoudige regressie, maar dan alles een dimensie hoger. Waar de enkelvoudige lineaire regressie leidde tot een lijn in het vlak die door het zwaartepunt gaat, zo leidt de meervoudige lineaire regressie tot een vlak in de ruimte, dat door het zwaartepunt $(\bar{V}, \bar{B}, \bar{E})$ gaat. En net als bij de enkelvoudige modellen is de gemiddelde kwadraatfout een maat voor de passendheid, en kun je die eventueel omschrijven tot R^2 via het verband

$$R^2 = 1 - \frac{K^2}{\sigma_E^2}$$

Een tweede opmerking betreft de variabelen die je betreft in de meervoudige regressieanalyse. Een voor de hand liggende vuistregel is: als een variabele bij enkelvoudige regressie niet tot een hoge waarde van R^2 leidt, dan nemen we die niet mee in de meervoudige analyse. Het resultaat van de enkelvoudige analyse bepaalt dus welke procesvariabelen van belang zijn voor een meervoudig model.

Ten slotte een opmerking over het verschil tussen het meervoudige lineaire model en het meervoudige machtsmodel. In het meervoudige lineaire model tellen we in feite de twee enkelvoudige lineaire modellen op. We spreken dan van een *additief model*. Een aanname daarbij is dat er tussen de twee onafhankelijke stuurvariabelen geen onderlinge beïnvloeding bestaat. In een *multiplicatief model*, zoals het meervoudige machtsmodel, worden de variabelen met elkaar vermenigvuldigd, en dat geeft aan dat de waarde van de ene variabele het enkelvoudige regressiemodel van de andere beïnvloedt.

Bron 15. Standaarddeviatie van populatie en steekproef

In bron 7 is als spreidingsmaat van een dataverzameling van 21 metingen de standaardafwijking of standaarddeviatie gedefinieerd door:

$$\sigma = \sqrt{\frac{1}{21} \sum_{i=1}^{21} (V_i - \bar{V})^2}$$

In woorden: de standaarddeviatie is de "wortel van de gemiddelde kwadraatfout van de waarnemingen ten opzichte van het gemiddelde".

Tegelijkertijd zien we iets vreemds: de schermen laten zien dat de TI84 niet alleen de σ_x berekent volgens de definitie hierboven, maar ook een S_x die iets groter is. Het Excel-commando STDEV geeft dezelfde waarde als de S_x van de TI84. Wat is hier aan de hand?

De definitie van σ zoals die hierboven is gegeven, is in feite die van de zogeheten *populatiestandaarddeviatie*. Dat is de spreidingsmaat die je kunt gebruiken als je alle elementen van de hele populatie in je berekening hebt betrokken. In ons geval is dat niet de situatie, want we hadden best nog meer metingen kunnen doen dan 21. En we zullen nooit in staat zijn alle mogelijke metingen te doen. Dat betekent dat we een feite de populatiestandaarddeviatie gaan schatten op basis van onze 21 metingen. Onze uitkomst is dan ook eigenlijk niet meer dan een *steekproefstandaarddeviatie*, waarvan we hopen dat die dicht ligt bij de onbekende populatiestandaarddeviatie.

Op grond van statistische theorie, die te ver gaat om hier te bespreken, kan worden bewezen dat de beste schatting van een onbekende populatiestandaarddeviatie wordt verkregen door de totale kwadraatfout niet te delen door het aantal metingen 21, maar door het aantal metingen min 1, dus in dit geval 20. Een beetje tegen de eerste intuïtie in misschien, maar toch is het zo.

De definitie van de steekproefstandaarddeviatie wordt dus:

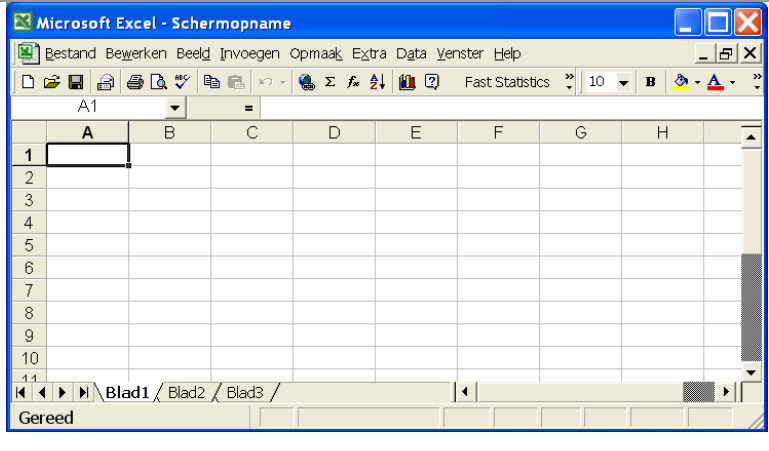
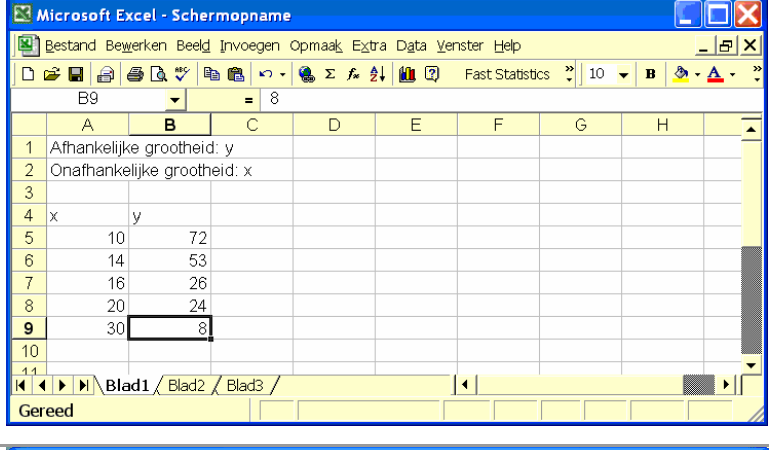
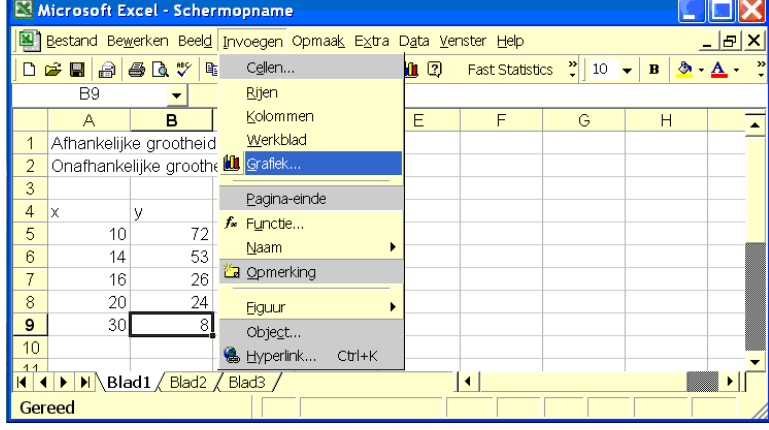
$$S = \sqrt{\frac{1}{20} \sum_{i=1}^{21} (V_i - \bar{V})^2}$$

Wat Excel berekent met STDEV is dus in feite de steekproefstandaarddeviatie en niet die van de populatie, terwijl de TI84 beide resultaten geeft.

Overigens is voor grote aantallen waarnemingen het verschil tussen de twee uitkomsten klein, zodat het onderscheid dan minder belangrijk is.

Bron 16. Correlatie en regressie met Excel

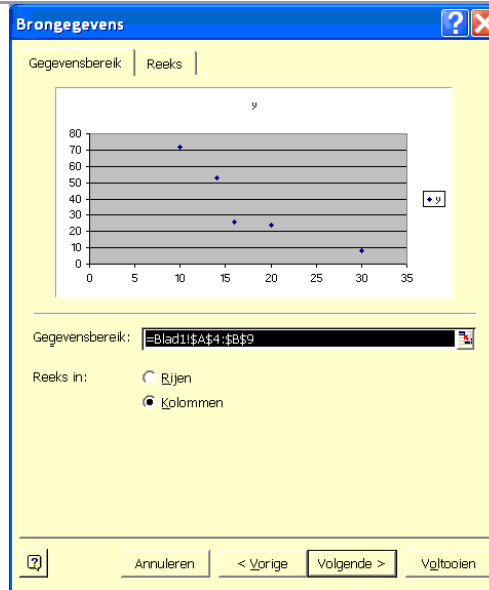
(i) Spreidingsdiagrammen in Excel

<p>Open een MS Excel werkblad.</p>	
<p>Voer de gegevens voor de grafiek in. Gebruik twee kolommen.</p>	
<p>Laat MS Excel een spreidingsgrafiek maken van de afhankelijke variabele y (y-as) tegen de onafhankelijke variabele x (x-as).</p> <p>Klik op de optie Invoegen > Grafiek.</p>	

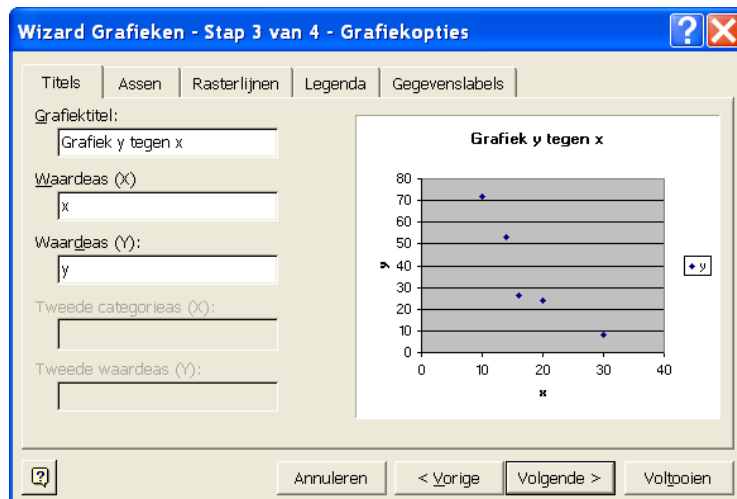
Kies vervolgens voor Spreiding.



Selecteer de gegevens voor de grafiek en plaats de grafiek vervolgens in het werkblad.



Geef de grootheid aan bij de x- en y-as en vermeld de eenheid.



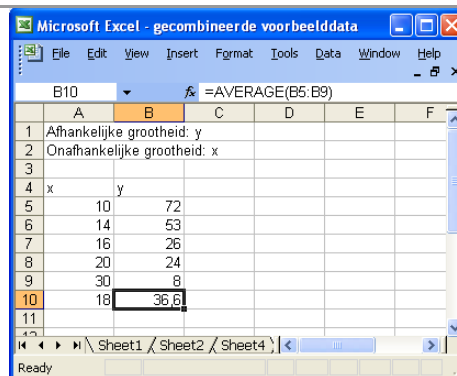
Plaats de grafiek in het werkblad via Voltoeien.



(ii) Gemiddelde en standaarddeviatie met Excel

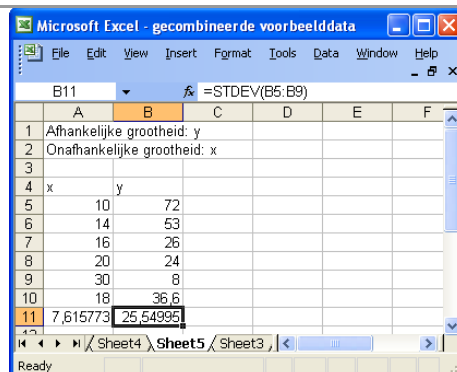
Bereken bijvoorbeeld het gemiddelde van de y -waarden door in te voeren

Average(B5:B9)



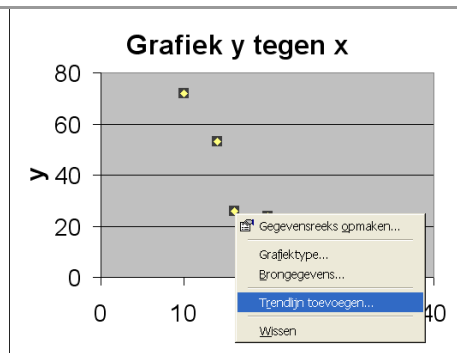
Bereken bijvoorbeeld de standaardafwijking van de y -waarden door in te voeren

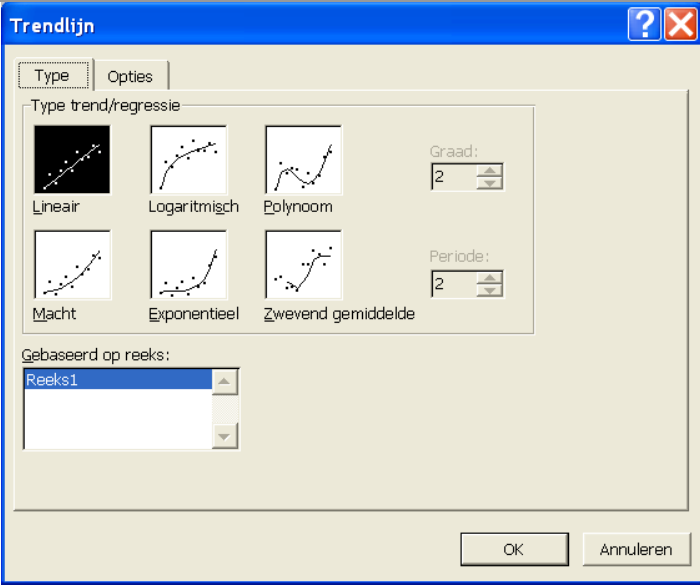
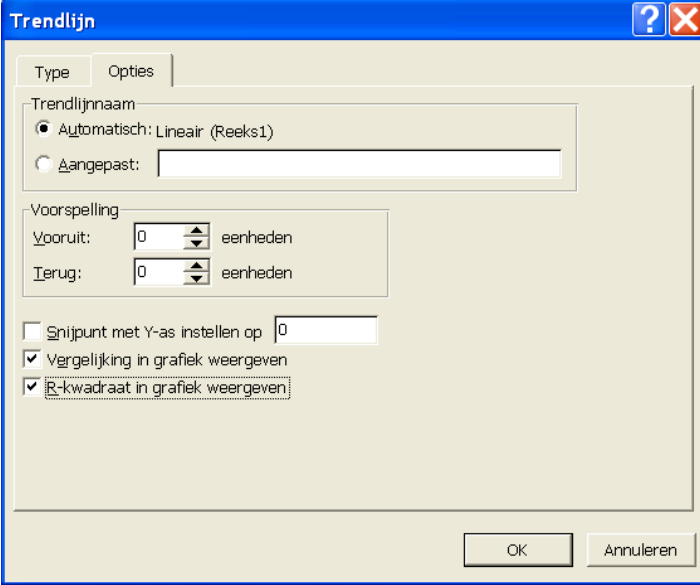
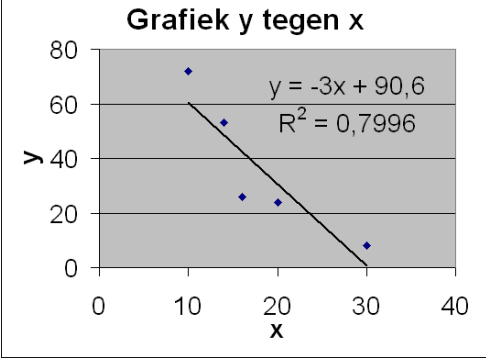
Stdev(B5:B9)



(iii) Enkelvoudige regressierekening in Excel

Klik in de grafiek op 1 van de punten met de *rechtermuisknop* en selecteer Trendlijn toevoegen.



<p>Kies vervolgens uit verschillende typen regressie:</p> <ul style="list-style-type: none"> ○ Lineair ○ Logaritmisch ○ Macht ○ Exponentieel ○ Polynoom ○ Zwevend gemiddelde <p>Kies de juiste trend/regressie (polynoom en zwevend gemiddelde worden meestal niet gebruikt).</p> <p>Kies daarna voor tabblad 'Opties'.</p>	
<p>In tabblad 'Opties' aanvinken:</p> <ul style="list-style-type: none"> ○ 'Vergelijking in grafiek weergeven' ○ 'R-kwadraat in grafiek weergeven' <p>Bevestig vervolgens met OK.</p>	
<p>Excel trekt nu een lijn door de meetpunten en toont de bijbehorende formule van die lijn met de R-kwadraat.</p> <p>In dit geval is getoond een lineaire trend/regressie.</p> <p>De waarde van R^2 geeft de betrouwbaarheid van de lijn aan.</p>	

(iv) Meervoudige lineaire regressie in Excel

Voer de metingen in. Doe dit in kolommen. Hiernaast is een voorbeeld gegeven met 1 afhankelijke variabele y en twee onafhankelijke variabelen x_1 en x_2 .

Selecteer een gebied van ongeveer 16 hokjes op een vrije plaats in het werkblad.

Microsoft Excel - Schermopname

Bestand Bewerken Beeld Invoegen Opmaak Extra Data

A11 =

	A	B	C	D	E
1	Afhankelijke grootte: y				
2	Onafhankelijke grootte: x1 en x2				
3					
4	x1	x2	y		
5	10	150	72		
6	14	136	88		
7	16	128	32		
8	20	200	88		
9	30	150	8		
10					
11					
12					
13					
14					
15					
16					
17					
18					

Typ in de commando regel:

=LIJNSCH(

(LINEST in de engelse versie van Excel)

Selecteer vervolgens de y-waarden (waarden van de afhankelijke variabele)

Typ een ` ; '

Microsoft Excel - Schermopname

Bestand Bewerken Beeld Invoegen Opmaak Extra Data Venst

SOM X ✓ = =LIJNSCH(C5:C9)

	A	B	C	D	E
1	Afhankelijke grootheid: y				
2	Onafhankelijke grootheden: x1 en x2				
3					
4	x1	x2	y		
5		10	150	72	
6		14	136	88	
7		16	128	32	
8		20	200	88	
9		30	150	8	
10					
11	=LIJNSCH(C5:C9)				
12					
13					
14					
15					
16					
17					
18					

Selecteer vervolgens alle x-waarden (waarden van de onafhankelijke variabelen)

Microsoft Excel - Schermopname

Bestand Bewerken Beeld Invoegen Opmaak Extra Data Venster

SOM X ✓ = =LIJNSCH(C5:C9;A5:B9)

	A	B	C	D	E
1	Afhankelijke grootheid: y				
2	Onafhankelijke grootheden: x1 en x2				
3					
4	x1	x2	y		
5		10	150	72	
6		14	136	88	
7		16	128	32	
8		20	200	88	
9		30	150	8	
10					
11	=LIJNSCH(C5:C9;A5:B9)				
12					
13					
14					
15					
16					
17					
18					

Typ daarna:
;WAAR;WAAR

en sluit af met `) `

	A	B	C	D	E	F	G
1	Afhankelijke grootheid: y						
2	Onafhankelijke grootheden: x1 en x2						
3							
4	x1	x2	y				
5	10	150	72				
6	14	136	88				
7	16	128	32				
8	20	200	88				
9	30	150	8				
10							
11	=LIJNSCH(C5:C9;A5:B9;WAAR;WAAR)						
12							
13							
14							
15							
16							
17							
18							

Voer de berekening uit door
tegelijk op **Ctrl - Shift - Enter**
te drukken.

Nu wordt een aantal waarden
zichtbaar in het geselecteerde
gebied.

	A	B	C	D	E	F	G
1	Afhankelijke grootheid: y						
2	Onafhankelijke grootheden: x1 en x2						
3							
4	x1	x2	y				
5	10	150	72				
6	14	136	88				
7	16	128	32				
8	20	200	88				
9	30	150	8				
10							
11	0,767518	-3,7651	8,095089	#N/B			
12	0,447352	1,645983	68,70274	#N/B			
13	0,770249	24,37303	#N/B	#N/B			
14	3,352536	2	#N/B	#N/B			
15	3983,111	1188,089	#N/B	#N/B			
16	#N/B	#N/B	#N/B	#N/B			
17							
18							

Uitgaande van het model $y = m_1 * x_1 + m_2 * x_2 + b$ is de meest rechtse waarde in de tabel de waarde van b , een plaats naar links de waarde van m_1 en de eerste waarde is die van m_2 .

m_2	m_1	b
#	#	#
R^2	#	

De derde waarde in de eerste kolom is die voor R^2 . De andere getallen kunnen gebruikt worden voor andere statistische toetsen, maar die zijn voor ons niet van belang.

(v) Meervoudige exponentiële regressie in Excel

Een exponentiële regressie kan worden uitgevoerd met de functie LOGSCH (LOGEST in de engelse versie van Excel). Volg de procedure zoals beschreven bij de meervoudige lineaire regressie, maar gebruik nu het commando LOGSCH in plaats van LIJNSCH. De resulterende exponentiële formule heeft de vorm $y = b * (m_1^{x_1}) * (m_2^{x_2})$.

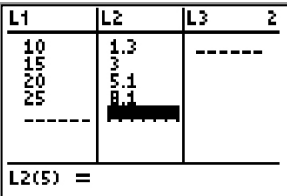
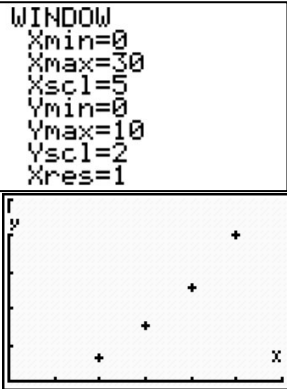
De waarden van m_1 , m_2 , b en R^2 vind je op dezelfde wijze als bij LIJNSCH.

Bron 17. Correlatie en regressie met de TI83/84

Zorg dat de rekenmachine ingesteld staat op zelf afronden (door [MODE], dan bovenaan [NORMAL] en op de tweede rij [FLOAT]) of op de wetenschappelijke notatie (door [MODE], dan bovenaan [SCI] en bijvoorbeeld op de tweede rij 2 om totaal drie significante cijfers te krijgen).

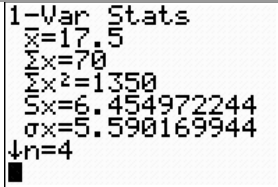
(i) Spreidingsdiagrammen met de TI83/84

Om een spreidingsdiagram te maken moeten de metingen in een tabel geplaatst worden. Daarna kan met [STAT PLOT] een spreidingsdiagram worden gemaakt.

<p>Punten invoeren in een tabel:</p> <p>Gebruik [STAT] – [EDIT] om de lijsten L1 en L2 te vullen. Lijst L1 betreft de x-waarden, L2 de y-waarden. Eventueel schoonmaken van de lijsten kan via [2ND] [MEM] – [CLRALLLISTS].</p>	
<p>Punten in een assenstelsel zetten:</p> <p>Kies in [STAT PLOT] (via [2nd] [STAT PLOT] voor bijvoorbeeld [PLOT1])</p> <p>Instellingen voor [PLOT1]:</p> <p>zet de grafiek [ON]</p> <p>Type: kies het eerste type (een grafiek met losse punten)</p> <p>voor [XLIST] kies je [L1] (x-waarden)</p> <p>voor [YLIST] kies je [L2] (y-waarden)</p> <p>markering een plusteken.</p> <p>Kies in [WINDOW] de maximale en minimale waarden voor X en Y. Met [GRAPH] worden de punten in beeld gebracht. Zorg ervoor dat de <i>oorsprong</i> in beeld is.</p>	

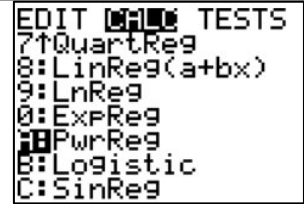
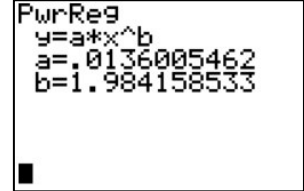
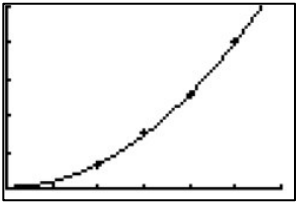
(ii) Gemiddelde en standaarddeviatie met de TI83/84

<p>Gebruik [STAT] – [CALC] optie 1:</p> <p>1-VAR STATS</p>	
--	---

<p>Vul het commando aan tot:</p> <p>1-VAR STATS L1</p> <p>Bovenaan verschijnt het gemiddelde, en σ staat bijna onderaan.</p>	
--	---

(iii) Enkelvoudige regressierekening met de TI83/84

De grafische rekenmachine heeft (onder [STAT] - [CALC]) een aantal regressiemethoden. Elke methode berekent de best passende functie van het gekozen type. In het onderstaande voorbeeld wordt een machtsregressie (PwrReg) uitgevoerd. Verder zijn o.a. ook beschikbaar een lineaire (LinReg), exponentiële (ExpReg) en logaritmische regressiemethode (LogReg).

<p>Kies in [STAT] - [CALC] voor [PWRREG] (of voor een andere <i>passende</i> regressiemethode).</p> <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;"> <p>PwrReg: $y=a \cdot x^b$</p> </div>	
<p>Voer in: PwrReg L₁, L₂, Y₁</p> <p>Voer L1 (x-waarden) in via [2ND] - [L1]</p> <p>Voer L2 (y-waarden) in via [2ND] - [L2]</p> <p>De variabele [Y₁] staat bij [VARS] onder [Y-VARS]. Door de toevoeging Y₁ komt de gevonden formule automatisch in de functielijst Y₁, en dus automatisch in de grafiek ([GRAPH]).</p>	
<p>Resultaten van regressie bekijken:</p> <p>Na bevestiging met [ENTER] wordt het resultaat van PwrReg zichtbaar.</p> <p>Via [GRAPH] wordt het resultaat zichtbaar als een lijn door de meetpunten.</p>	

Ook de correlatiecoëfficiënten (R en R^2) kunnen worden berekend. Deze waarden worden echter niet samen met de resultaten van het regressiemodel weergegeven. Om deze gegevens zichtbaar te maken moet de weergavemodus worden ingesteld via DiagnosticOn of DiagnosticOff via [CATALOG]:

Kies [2nd] - [CATALOG]:

Scroll door de (alfabetische) lijst naar beneden naar functie [DIAGNOSTICON]

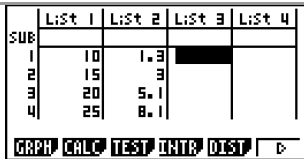
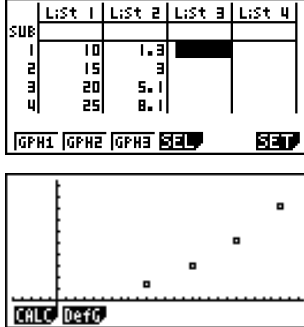
Bevestig met [ENTER]

Voer hierna de regressie opnieuw uit.

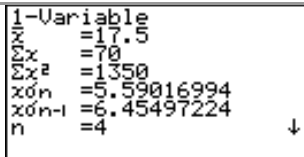
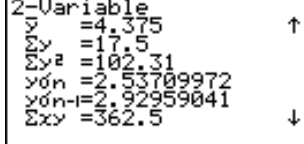
Bron 18. Correlatie en regressie met de Casio 9850/9860

(i) Spreidingsdiagrammen met de Casio 9850/9860

Om een spreidingsdiagram te maken moeten de metingen in een tabel geplaatst worden.

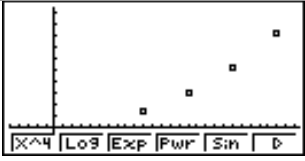
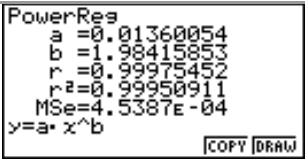
<p>Punten invoeren in een tabel:</p> <p>Kies [STAT] om de lijsten List 1 en List 2 te vullen. Lijst List 1 betreft de x-waarden, List 2 de y-waarden. Eventueel schoonmaken van de lijsten kan via F6 > en vervolgens [DEL-A].</p>	
<p>Punten in een assenstelsel zetten:</p> <p>Kies [GRPH] met functietoets F1.</p> <p>Kies [GPH1]</p> <p>Gebruik eventueel [SETUP] (via Shift Menu) om de instelling van het kijkvenster op 'maul' te zetten en pas dat dan met [V-WIN] of [ZOOM] aan. Zorg ervoor dat de oorsprong in beeld is.</p>	

(ii) Gemiddelde en standaarddeviatie met de Casio 9850/9860

<p>Ga met [EXIT] terug naar het scherm met de lijsten en kies voor [CALC] en dan voor [1-VAR]. Je kunt ook eerst [SET] gebruiken om de lijst aan te geven waarin de data staan.</p> <p>Dat geeft de gegevens van List 1. Met het pijltje naar beneden kun je nog meer gegevens in beeld krijgen.</p>	
<p>Je kunt ook kiezen voor [2-VAR]</p> <p>Dan krijg je na enig bladeren ook de gegevens van Lijst 2 in beeld.</p>	

(iii) Enkelvoudige regressierekening met de Casio 9850/9860

De Casio grafische rekenmachine heeft een aantal regressiemethoden. Elke methode berekent de best passende functie van het gekozen type. In het onderstaande voorbeeld wordt een machtsregressie (PwrReg) uitgevoerd. Verder zijn o.a. ook beschikbaar een lineaire (LinReg), exponentiële (ExpReg) en logaritmische regressiemethode (LogReg).

<p>Kies in het spreidingsdiagram voor [CALC] en zorg met F6 dat [PWR] in beeld komt.</p>	
<p>Kies voor [PWR]. Je krijgt de vergelijking van de machtsregressiekromme in beeld, evenals de waarde van R en R^2.</p>	
<p>Resultaten van regressie bekijken: Met [DRAW] wordt het resultaat van PwrReg zichtbaar als een lijn door de meetpunten.</p>	