

BTO 2006.022
mei 2006

Optimaliseren van ultrafiltratie met neurale netwerken

een verkenning

BTO 2006.022
mei 2006

Optimaliseren van ultrafiltratie met neurale netwerken

een verkenning

© 2006 Kiwa N.V.
Alle rechten voorbehouden.
Niets uit deze uitgave mag
worden verveelvoudigd,
opgeslagen in een
geautomatiseerd
gegevensbestand, of
openbaar gemaakt, in enige
vorm of op enige wijze,
hetzij elektronisch,
mechanisch, door
fotokopieën, opnamen, of
enig andere manier, zonder
voorafgaande schriftelijke
toestemming van de
uitgever.

Kiwa N.V.
Water Research
Groningenhaven 7
Postbus 1072
3430 BB Nieuwegein

Tel. (030) 606 95 11
Fax (030) 606 11 65
www.kiwa.nl

Colofon

Titel

Optimaliseren van ultrafiltratie met neurale netwerken

Projectnummer

11.1588.700

Projectmanager

Peter Wessels

Opdrachtgever

CvO

Kwaliteitsborger(s)

Peter Wessels

Auteur(s)

Raphaël van der Velde (Witteveen+Bos), Floris
Schulze (TU Delft en Kema, voorheen
Witteveen+Bos), Anneke Gijsbertsen (Kiwa)

Dit rapport is verspreid onder BTO-participanten en is openbaar.

Voorwoord

Voor u ligt het rapport van de verkennende studie naar toepassing van een kunstmatig neurale netwerk voor het vinden van relaties tussen waterkwaliteitsparameters en procesinstellingen bij ultrafiltratie enerzijds en de transmembraandruk anderzijds. Oorspronkelijk was het de bedoeling data van verschillende praktijk ultrafiltratie-installaties te gebruiken. Tijdens het project kwam echter naar voren dat slechts een beperkt aantal data met een voldoende hoge frequentie zijn gemeten. De verkennende studie met data van de UF-pilot van Waternet bood onvoldoende vertrouwen in de aanpak en het te bereiken resultaat, zodat besloten werd geen data-acquisitie te starten en het project op dit punt te staken.

Wij willen de volgende personen, die soms een niet geringe inspanning hebben moeten leveren om data beschikbaar te krijgen, bedanken: Eric Baars (Waternet), Hilde Prummel (WLN), Sjack van Agtmaal (Evides) en Stephan van de Wetering (Brabant Water). Tevens zijn wij Huub Savenije en Theo Olsthoorn (beiden TUDelft) erkentelijk voor de door hen geleverde “second opinion” over het uitvoeren van de verkennende studie.

Samenvatting

Met gegevens uit een eerder uitgevoerde inventarisatie van UF-installaties in Nederland konden geen verbanden worden gevonden tussen enerzijds procesinstellingen, zoals filtratietijd en flux, waterkwaliteitsparameters (bv. troebelheid en DOC-concentratie) en anderzijds de prestaties van de installaties. De reden hiervoor was het grote aantal parameters dat mogelijk een invloed heeft ten opzichte van de beperkte dataset; voor elke parameter was alleen een gemiddelde waarde bekend. In zo'n complexe situatie, met veel variabelen die onderling een verband kunnen hebben, kan met een kunstmatig neurale netwerk (KNN) toch een relatie worden gelegd tussen input (procesinstellingen en waterkwaliteitsgegevens) en output (transmembraandruk), zonder dat hiervoor de exacte fysische achtergrond bekend hoeft te zijn.

Het doel van dit onderzoek was om met een KNN algemeen geldende relaties te vinden tussen waterkwaliteitsparameters en procesparameters enerzijds en de transmembraandruk in praktijk UF-installaties anderzijds. In dit onderzoek zijn daarvoor data van verschillende praktijkinstallaties verzameld en is naast een KNN black-box model ook een KNN white-box model gebruikt. Met dit laatste model kunnen verbanden tussen input en outputparameters inzichtelijk worden gemaakt.

Na het inventariseren van de beschikbare data bleek bij alle praktijkinstallaties slechts een zeer beperkt aantal waterkwaliteitsparameters met voldoende hoge frequentie te zijn gemeten. Daarom is alleen met de data van de – inmiddels afgebroken – pilotinstallatie van Waternet een verkennende studie uitgevoerd. De betrouwbaarheid van de twee geteste Kunstmatige Neurale Netwerken is met 40% aanzienlijk hoger dan die van het lineaire regressiemodel (9%).

Op basis van de reeds bereikte betrouwbaarheid en de mogelijkheden tot verbetering, is de verwachting dat Kunstmatige Neurale Netwerken een bruikbaar hulpmiddel kunnen zijn bij het modelleren en optimaliseren van ultrafiltratie. Met een White Box Neural Network zullen de onderliggende (fysische) verbanden in relatief eenvoudige vergelijkingen uitgeschreven kunnen worden.

Na presentatie van de resultaten van deze verkennende studie is het project echter om twee redenen beëindigd:

- Gebrek aan geschikte datasets;
- Het ter discussie stellen van het doel, nl. het nut van het bepalen van “algemeen” geldende verbanden. Ook bij gebruik van data van verschillende installaties is de algemeenheid van de data beperkt, is het model bv. ook bruikbaar voor andere membraanvezels, of zelfs andere membraanmaterialen?

Bij een eventuele vervolgstudie zal eerst moeten worden besloten wat de doelstelling is, om vervolgens een geschikt meetprogramma op te stellen.

Inhoud

Voorwoord	1
Samenvatting	3
Inhoud	5
1 Inleiding	7
1.1 Aanleiding	7
1.2 Literatuur	7
1.3 Doel en aanpak van het onderzoek	8
1.4 Opbouw van dit rapport	8
2 Theorie statistische modellen	11
2.1 Lineaire regressie	11
2.2 Neurale netwerken	11
2.2.1 Black box (klassieke) neurale netwerken	12
2.2.2 White Box Neural Networks	12
2.3 kalibratie – validatie – verificatie	14
3 Data	15
3.1 Gewenste data	15
3.2 Analyse beschikbare data	16
3.3 Bewerken	17
3.3.1 Gegevens najaar 2003	17
3.3.2 Gegevens voorjaar 2003	18
4 Resultaten en discussie	21
4.1 Lineaire regressie	21
4.2 Multi-Layer Perceptron	21
4.3 White Box Neural Network	23
4.4 Discussie	24
5 Conclusie en aanbevelingen	25
6 Referenties	27
I Bijlage: beschikbare data	29

1 Inleiding

1.1 Aanleiding

Wereldwijd zijn er meer dan 400 micro- en ultrafiltratieinstallaties gebouwd, in Nederland meer dan 20. Uit een inventarisatie van deze installaties, uitgevoerd in 2004, blijkt dat er behoorlijke verschillen zijn in bedrijfsvoering tussen de verschillende installaties (Gijsbertsen, 2004). Zo varieert de filtratietijd van 15 tot 120 minuten en worden er bij chemische reiniging verschillende chemicaliën gebruikt, verschillende concentraties en inweektijden. Daarnaast lopen ook de prestaties uiteen: de netto fluxen verschillen nogal, evenals de waterverliezen door reinigingen. Het is echter niet gelukt verbanden te leggen tussen de waterkwaliteitsparameters en procesparameters enerzijds en de prestaties van de installatie anderzijds. In zo'n complexe situatie kan met een kunstmatig neurale netwerk (KNN) mogelijk toch een relatie worden gelegd tussen input (procesinstellingen en waterkwaliteitsgegevens) en output (flux en transmembraandruk), zonder dat hiervoor de exacte fysische achtergrond bekend hoeft te zijn.

1.2 Literatuur

Door verschillende onderzoeksgroepen is aangetoond dat met een getraind kunstmatig neurale netwerk het niet-lineaire gedrag tijdens ultrafiltratie nauwkeurig kan worden voorspeld (Cabassud *et al.*, 2002; Teodosiu *et al.*, 2000; Oh *et al.*, 2004). De modellen, de voorspelde parameters en de gebruikte data waren verschillend in deze onderzoeken.

- Oh *et al.* (2004) trainde een KNT neurale netwerk met data verzameld met een ultrafiltratie pilotinstallatie die gedurende 7 maanden met grondwater werd gevoed. (KNT staat voor Kalman neuro training). In tegenstelling tot wat gebruikelijk is bij ultrafiltratie, werd de druk ingesteld en de resulterende flux gemeten. Verder werden naast de tijdsduur van de filtratie ook de waterkwaliteitsparameters troebelheid, temperatuur en UV₂₅₄ gemeten. Troebelheid en UV₂₅₄ waren laag en bleken geen significante invloed op de permeaatflux te hebben.
- Cabassud *et al.* (2002) bouwde twee gekoppelde kunstmatig neurale netwerken, waarbij het ene model de membraanweerstand aan het eind van de filtratieperiode voorspelt aan de hand van de bedrijfsvoeringsparameters tijdens filtratie, de waterkwaliteitsparameters en de membraanweerstand aan het begin van de filtratieperiode. Het tweede model voorspelt de membraanweerstand aan het eind van de backwash (dat is begin van de filtratieperiode) aan de hand van de membraanweerstand aan het eind van de filtratieperiode, de bedrijfsvoeringsparameters tijdens terugspoelen en de waterkwaliteitsparameters. De dagelijks gemeten waterkwaliteitsparameters van het oppervlaktewater (rivier de Seine) waren troebelheid, temperatuur, UV₂₅₄, pH, de opgeloste zuurstofconcentratie, TOC (total organic carbon) en de redoxpotentiaal. De laatste twee parameters, die overigens ook het lastigst zijn te meten, leverden voor dit oppervlaktewater geen significante bijdrage aan de membraanweerstand (Delgrange-Vincent *et al.*, 2000). Van de procesparameters bleken de filtratieflux,

terugspoeldruk, filtratietijd en de chloorconcentratie in het terugspoelwater een significante invloed te hebben op de berekende membraanweerstand.

- Teodosiu *et al.* (2000) koppelden op dezelfde manier twee kunstmatig neurale netwerken aan elkaar, echter tijdens de filtratie kan op elk willekeurig moment de flux worden voorspeld. De fluxen kunnen echter maar over een periode van 2500 seconden worden voorspeld.

De beschreven modellen worden gepresenteerd als “black boxes”, waardoor geen inzicht wordt verkregen in het verband tussen de proces- en waterkwaliteitsparameters en de outputparameter (flux, membraanweerstand of transmembraandruk). Bovendien zijn de modellen beperkt toepasbaar omdat ze zijn getraind en getest op data gegenereerd met één watertype.

1.3 Doel en aanpak van het onderzoek

Het doel van dit onderzoek was om met een KNN de belangrijkste relaties te vinden tussen waterkwaliteitsparameters en procesparameters enerzijds en de transmembraandruk in praktijk UF-installaties anderzijds. In dit onderzoek zijn daarvoor data van verschillende praktijkinstallaties verzameld en is naast een black-box model ook een white-box model gebruikt. Met dit laatste model kunnen verbanden tussen input en outputparameters inzichtelijk worden gemaakt.

De volgende activiteiten zijn uitgevoerd:

1. Selectie van UF-installaties met vergelijkbare configuratie;
2. Gegevens verzamelen
 - definiëren van het type data (input en output parameters), kwaliteit en hoeveelheid per dataset/aantal datasets;
 - contact met eigenaren van UF-installaties en membraanleverancier over mogelijkheid tot aanleveren gegevens;
3. Gegevens controleren en in gewenste vorm omzetten;
4. Ontwikkelen kunstmatig neurale netwerk
 - programmeren;
 - calibreren;
 - valideren inclusief gevoeligheidsanalyse.

Na activiteit 2 bleek dat bij alle geselecteerde praktijkinstallaties slechts een zeer beperkt aantal waterkwaliteitsparameters met voldoende hoge frequentie is gemeten. De data van de – inmiddels afgebroken – pilotinstallatie van Waternet leken wel bruikbaar om een model te ontwikkelen en te testen. Dit is in een verkennende studie onderzocht (activiteiten 3 en 4). Na deze studie is echter besloten het project niet voort te zetten en de resultaten tot dan toe te rapporteren.

1.4 Opbouw van dit rapport

In hoofdstuk 2 wordt kort ingegaan op de statistische methoden, waarvan het ontwikkelde “white box” neurale netwerk er één is. De gekozen uitgangspunten bij het verzamelen van data van praktijkinstallaties en een overzicht van de verzamelde data worden in hoofdstuk 3 gegeven. Daarnaast wordt in dat hoofdstuk het prepareren van de data van de Waternet UF-pilot voor de verkennende studie toegelicht. Vervolgens wordt in hoofdstuk 4 per type model beschreven hoe goed de invoerparameters (waterkwaliteit- en procesparameters) de uitvoerparameter (drukverschil) verklaren. De resultaten van de in dit rapport beschreven

verkennende studie waren minder goed dan verwacht, maar er zijn mogelijkheden tot verbetering. In hoofdstuk 5 wordt een aantal aanbevelingen gedaan die in de toekomst wellicht kunnen leiden tot, of gebruikt kunnen worden in, een nieuw project.

2 Theorie statistische modellen

Een regressieanalyse wordt uitgevoerd om inzicht te verkrijgen in de specifieke samenhang tussen verschillende data. In feite wordt getracht om een afhankelijke variabele (output, bijvoorbeeld de drukopbouw gedurende filtratie) te voorspellen op basis van een set verklarende variabelen (inputs, bijvoorbeeld troebelheid, watertemperatuur, zuurstofgehalte).

De belangrijkste beperking van regressiemodellen is dat ze alleen inzicht geven in de statistische verbanden, maar dat de achterliggende causale verbanden niet met zekerheid vastgesteld kunnen worden.

2.1 Lineaire regressie

Bij (meervoudig) lineaire regressie is de te voorspellen variabele lineair afhankelijk van de verklarende variabelen volgens onderstaande formule:

$$Y = \beta_0 + \beta_1 x_1 + \dots \beta_m x_m$$

met:

Y	onafhankelijke variabele
x_1 t/m x_m	verklarende variabelen
β_0 t/m β_m	modelcoëfficiënten

Als de relatie tussen twee variabelen niet lineair, maar wel bekend is (bijvoorbeeld logaritmisch), kan voorafgaand aan het kalibreren (zie paragraaf 2.3) een transformatie op de variabelen uitgevoerd worden.

2.2 Neurale netwerken

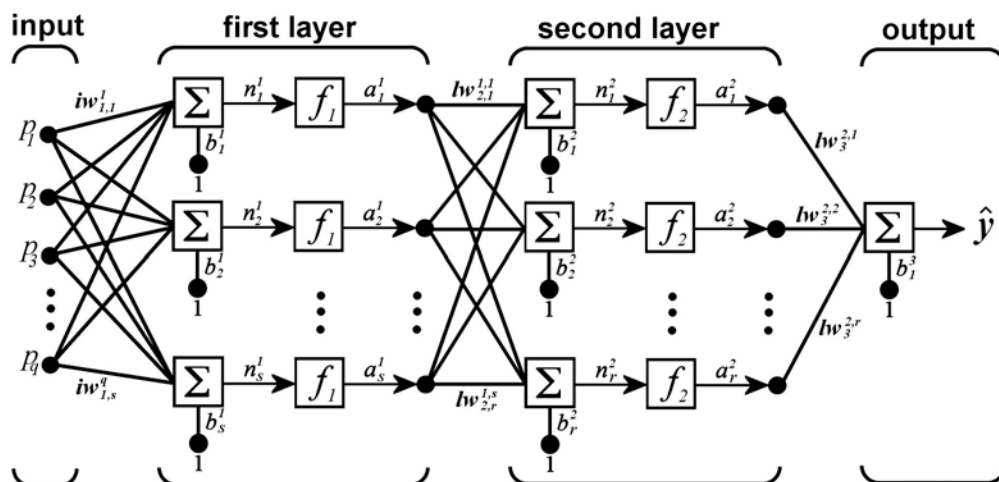
Kunstmatige neurale netwerken (KNN) zijn in staat om complexe, lineaire en niet-lineaire processen te modelleren, zonder het systeem zelf in beschouwing te nemen. Een KNN is een aaneenschakeling van relatief eenvoudige mathematische functies. Deze basiselementen van het netwerk worden, naar analogie van het menselijk brein, neuronen genoemd.

Een neuron is opgebouwd uit twee belangrijke elementen, een sommering Σ en een transferfunctie f . Een aantal mogelijke inputs p_1 t/m p_q wordt eerst vermenigvuldigd met ieder een eigen 'gewicht' w_1 t/m w_q en tezamen met een vast niveau b gesommeerd met als uitkomst n . Deze waarde dient als input voor de transferfunctie f met als uitkomst a (zie onderstaande vergelijking).

$$a = f(n) = f\left(\left(\sum_{i=1}^q p_i w_i\right) + b\right) \quad (\text{vergelijking van een enkele neuron})$$

Een netwerk van neuronen wordt een Kunstmatig Neuraal Netwerk genoemd. Binnen de KNN-modellen is echter een groot aantal mogelijkheden voor het koppelen van neuronen.

Voor een meer gedetailleerde beschrijving van KNN wordt verwezen naar *Neural, Novel & Hybrid Algorithms for Time Series Prediction* (Masters, 1995) en *Gestructureerd modelontwerp met Kunstmatige Neurale Netwerken* (Van der Wielen, 2001).



Figuur 1: Schematische weergave Multi Layer Perceptron (Hagan *et al.*, 1996).

Het Multi Layer Perceptron is waarschijnlijk de meest toegepaste netwerkstructuur op het gebied van niet-lineaire regressie (zie Figuur 1).

2.2.1 Black box (klassieke) neurale netwerken

Zoals reeds aangegeven is een Kunstmatig Neuraal Netwerk een wiskundige rekentechniek die in staat is complexe systemen te modelleren zonder dat de achterliggende processen zelf in beschouwing (hoeven te) worden genomen. Een vereiste is wel dat er voldoende meetgegevens beschikbaar zijn die representatief zijn voor het probleem dat moet worden opgelost.

Tot nu toe zijn vooral klassieke KNN toegepast; de methoden en toepassingen zijn uitgebreid in de literatuur beschreven. De beperkingen die daarbij zichtbaar zijn geworden, zijn: de onoverzichtelijkheid van de gevonden wiskundige verbanden; de benodigde waakzaamheid bij extrapolatie van het opgestelde model en het risico op overparameterisatie (waarbij het verminderen van het aantal beschrijvende variabelen leidt tot betere resultaten). Een belangrijk nadeel van de met Kunstmatige Neurale Netwerken gevonden complexe statistische relaties, is dat ze door hun aard weinig inzicht verschaffen in achterliggende processen en causale verbanden. Kunstmatige Neurale Netwerken worden dan ook veelal aangeduid als Black Box modellen. Gevoeligheidsanalyses, die over het algemeen veel tijd kosten, kunnen het inzicht enigszins vergroten. Het “black-box” achtige karakter is overigens geen diskwalificatie van dit type Kunstmatige Neurale Netwerken, want dit soort methoden kan nuttig worden toegepast als model voor complexe problemen waarbij snelheid en nauwkeurigheid belangrijk zijn en waarvan voldoende representatieve gegevens beschikbaar zijn. In dergelijke situaties zijn veelal de achterliggende verbanden minder relevant en gaat het voornamelijk om het modelresultaat in termen van voorspelnauwkeurigheid, snelheid en flexibiliteit.

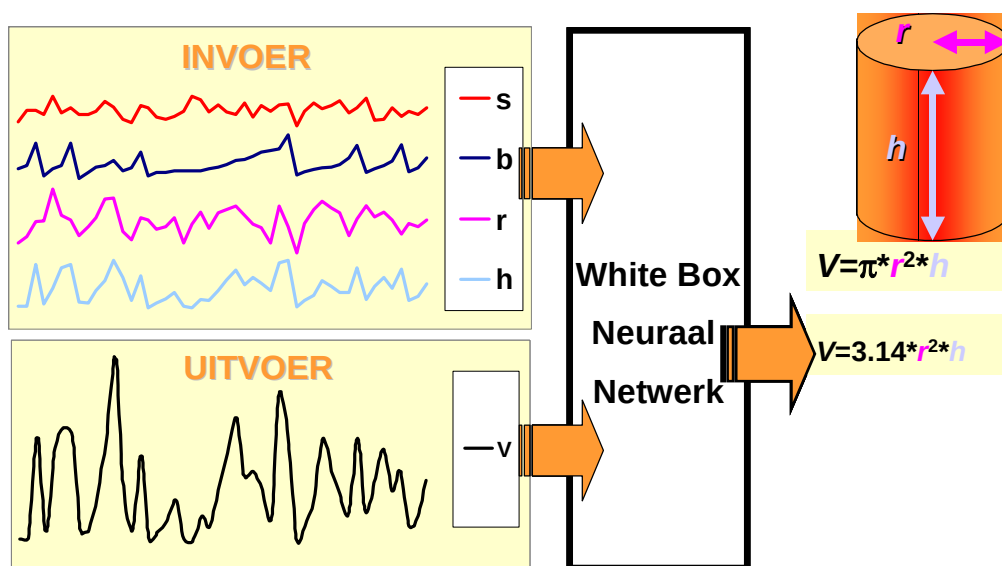
2.2.2 White Box Neural Networks

In het kader van een promotieonderzoek van de heer Schulze aan de Technische Universiteit van Delft wordt een White Box Neural Networks (WBNN) methodiek ontwikkeld. Deze is in staat achterliggende processen te extraheren en is toch praktisch toepasbaar want flexibel, snel en nauwkeurig. De causale verbanden kunnen ongeacht de aard van de problematiek worden geëxtraheerd. Deze

methodiek combineert in feite de positieve eigenschappen van de ‘traditionele’ Kunstmatige Neurale Netwerken (snelheid, nauwkeurigheid, flexibiliteit, inzetbaarheid, bruikbaarheid) met positieve eigenschappen van fysisch georiënteerde methodieken (systeemkennis, proceskennis) (Brinke *et al.*, 2004; Oost *et al.*, 2002; Schulze en Zijderfeld, 1999).

Het principe van de WBNN methodiek is vergelijkbaar met die van Multi Layer Perceptrons in Figuur 1. Echter, er worden andere transferfuncties en optimaliseroutines gehanteerd. Deze zijn in staat de neuronen te beschouwen als zogenaamde hoofdcomponenten van het te modelleren proces. De achterliggende vergelijkingen kunnen voor elk neuron afzonderlijk worden uitgeschreven. Dit maakt het mogelijk na te gaan of bepaalde neuronvergelijkingen bepaalde specifieke processen vertegenwoordigen. Omdat de specifieke relaties duidelijker naar voren komen is extrapolatie waarschijnlijk veel nauwkeuriger dan bij ‘traditionele’ KNN. Zowel het aantal invoervariabelen als het aantal neuronen wordt zodanig geoptimaliseerd dat uiteindelijk een optimum wordt gevonden tussen complexiteit van het model en de bijbehorende nauwkeurigheid. Ofwel, het streven is een zo eenvoudig mogelijk model met een zo hoog mogelijke nauwkeurigheid. Dit proces wordt Pruning genoemd (Brinke *et al.*, 2004; Oost *et al.*, 2002; Schulze en Zijderfeld, 1999). Door specifieke transferfuncties te gebruiken in combinatie met een aangepast pruning algoritme is het mogelijk eenvoudige en overzichtelijke modelvergelijkingen te genereren van het proces. Ook kan vooraf proceskennis worden ingebracht die vervolgens wordt meegenomen.

In Figuur 2 is een voorbeeld gegeven hoe het proces van WBNN te werk gaat en hoe deze als formule kan worden uitgeschreven. Een vergelijking zoals de inhoud van een cilinder kan exact met een WBNN worden weergegeven. Dit in tegenstelling tot de klassieke neurale netwerken, die slechts een benadering van de exacte oplossing zouden geven. Een WBNN is zo opgezet dat uit complexe data de belangrijkste factoren worden geëxtraheerd, en dat die als een overzichtelijke



Figuur 2: Principe van een White Box Neuraal Netwerk

formule worden weergegeven. Een ander voordeel van WBNN is dat met kleine datasets vanaf zo'n 100 waarden modelleren mogelijk is. Met traditionele KNN is een dataset met minstens 200 waarnemingen gewenst (referentie: Neural Computing, Neural Netware Inc, 1990).

2.3 kalibratie – validatie – verificatie

Na het kiezen van de netwerkstructuur en de toe te passen transferfuncties, wordt het netwerk 'getraind'. Trainen of kalibreren is het schatten van de gewichten van de neuronen. Bij lineaire regressie worden de lineaire modelcoëfficiënten geschat. Tijdens de validatie van het model wordt onderzocht of het gekalibreerde model de output van een onafhankelijke dataset voldoende nauwkeurig kan beschrijven. De verificatie geeft antwoord op de vraag of een model de werkelijkheid min of meer realistisch weergeeft. Dit wordt gedaan door het modelgedrag te vergelijken met de reeds kwalitatief bekende relaties: geeft het model bij een grotere vuillast een grotere drukopbouw als uitkomst?

3 Data

3.1 Gewenste data

Het opstellen van een goed kunstmatig neurale netwerk staat of valt met de beschikbaarheid van consistente data. Het onderzoek is vanuit dat perspectief beperkt tot één membraantype (X-Flow Xiga UFC met capillairdiameter 0.8 mm) en (voorgezuiverd) oppervlaktewater als bron.

Voor het verkrijgen van goede relaties is het van belang dat er variatie is in de waterkwaliteitsgegevens en de procesinstellingen. Van elke installatie is daarom gevraagd naar data uit een “slechtste” en een “beste” maand qua waterkwaliteit en/of procesinstellingen uit bij voorkeur 3 opeenvolgende jaren.

Van elke installatie werden de onderstaande gegevens gevraagd over de periode waarover de meetgegevens werden aangeleverd.

Algemene gegevens - aangenomen is dat deze constant zijn gedurende de beschouwde periode.

- Datum ingebruikname modules
- Aantal modules
- Totaal membraanoppervlak
- Dosering voorafgaand aan UF: welke chemicaliën
- Concentratie(s) chemicaliën (of in tabel waterkwaliteit vermelden als actuele concentratie varieert)
- Enkelzijdige of dubbelzijdige afvoer bij filtratie?
- Enkelzijdige of dubbelzijdige afvoer bij backwash?
- Filtratietijd (min)
- Filtratieflex (l/m²h)
- Backwashtijd (min)
- Backwashflux (l/m²h)
- Enhanced backwash: totale tijdsduur (min)
- Enhanced backwash: type chemicaliën
- Enhanced backwash: concentratie of pH
- Enhanced backwash: doseertijd (min)
- Enhanced backwash: flux tijdens doseren (l/m²h)
- Enhanced backwash: inweektijd (min)

In de onderstaande tabellen (Tabel 1 en Tabel 2) staan de waterkwaliteitsparameters en procesinstellingen vermeld die in het model opgenomen zouden moeten worden. Daarbij werd een meetfrequentie van tenminste 1 maal per uur wenselijk geacht. Aangegeven diende te worden wanneer een terugspoeling of enhanced backwash (chemische reiniging) werd toegepast.

Tevens is informatie (logboekgegevens) gevraagd over bijzonderheden in de bedrijfsvoering in de gerapporteerde periode. Hierbij valt te denken aan stroomstoringen, incidentele reinigingen, buiten werking zijn van de installatie en reparatie van gebroken vezels.

Tabel 1: Ingaande Waterkwaliteit

datum	tijd	zuurgraad	Troebeling	DOC	UV-ext 254nm	temperatuur	geleidbaarheid	chlorofyl -a	Zwevende stof	doserings
dd-mm-jj	uu:mm	pH	FTU/NTU	mg/l	m ⁻¹	°C	µs/cm	mg/l	mg/l	mg/l

Tabel 2: Procesinstellingen

datum	tijd	TMD	Flux	Filtratie	Backwash	Enhanced backwash
dd-mm-jj	uu:mm	bar	l/m ² .h			

3.2 Analyse beschikbare data

In de tabel in bijlage I staat vermeld welke van de oorspronkelijk gevraagde data van welke installaties beschikbaar waren, of beschikbaar konden worden gesteld door de waterleidingbedrijven. Daarbij is tevens aangegeven of er meer dan 250 waarden beschikbaar zijn (wenselijk voor het uitvoeren van een betrouwbare KNN-analyse) en de frequentie waarmee de data zijn gemeten.

Uit de analyse van de geleverde data blijkt het volgende:

1. In geen enkel geval worden parameters die een maat zijn voor organische stof online gemeten (bv. DOC en UV-extinctie).
2. Van de off-line gemeten parameters is niet bekend wanneer deze exact bepaald zijn. Alleen de monsterdatum is gemeld, niet het tijdstip.
3. Van slechts enkele kwaliteitsparameters zijn meer dan 250 gegevens voorhanden. Van een aantal installaties bleek dat van geen enkele waterkwaliteitsparameter voldoende gegevens voorhanden waren.
4. Alleen Temperatuur en pH zijn online beschikbaar van 2 installaties. Daarnaast zijn EGV, troebelingsgraad en zuurstofgehalte van slechts 1 installatie beschikbaar.

Uit deze analyse komt naar voren dat er te weinig aan een tijdstip toe te wijzen gegevens van waterkwaliteitsparameters van meerdere installaties beschikbaar zijn om algemeen geldende verbanden met KNN af te kunnen leiden. Met name het ontbreken van tijdspecifieke gegevens van organische parameters wordt als een probleem gezien. De conclusie is dat er onvoldoende waterkwaliteitsdata zijn om het doel van het project te realiseren. Alleen online meting van de verschillende waterkwaliteitsparameters gedurende een bepaalde tijd op een aantal installaties, zou voldoende data kunnen opleveren. De meetperiode zou bij voorkeur enkele seizoenen moeten duren om de invloed van variaties in de ruwwaterkwaliteit (o.a. temperatuur, pH, algengroei) van het water op de filtreerbaarheid mee te nemen.

Met de dataset van de proefinstallatie van Waternet is een vingeroefening uitgevoerd om vast te stellen of de aanpak met neurale netwerken werkt en om de daarbij naar voren komende knelpunten vast te stellen. De bewerking van deze dataset wordt hierna toegelicht.

3.3 Bewerken

De haalbaarheid om het ultrafiltratieproces te modelleren met behulp van KNN is onderzocht aan de hand van meetgegevens van de UF1-installatie die verkregen zijn tijdens een pilot meetproject uitgevoerd in 2003 door Waterleidingbedrijf Amsterdam (thans Waternet). De procesparameters en influentgegevens zijn gedurende het voor- en najaar gemeten met een frequentie van eens per 10 seconden.

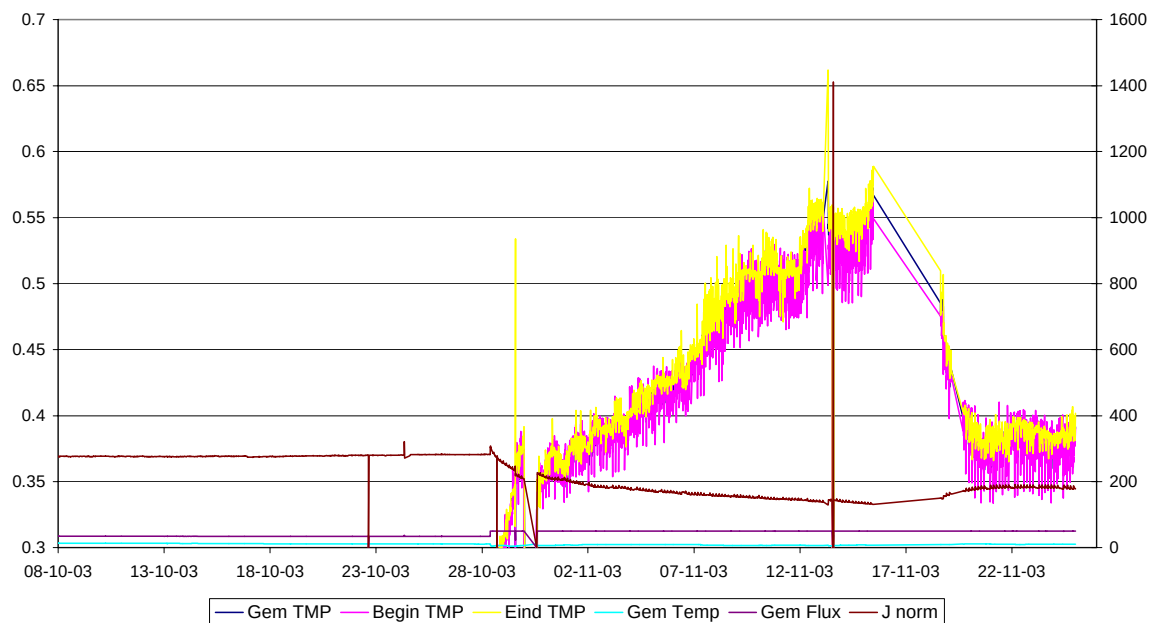
Tijdens het meetproject is de flux gestuurd op een constante waarde. De transmembraandruk is daardoor een variabele. De drukopbouw tijdens filtratie was echter dermate traag, dat deze op 10 seconden of zelfs op minuutbasis in de meeste gevallen niet waarneembaar was. Daarom is de drukopbouw tussen twee opeenvolgende hydraulische (dTMP) reinigingen gekozen als outputvariabele. Meestal bedraagt de tijd tussen twee hydraulische reinigingen ongeveer 11 minuten, maar dit bleek geen vast interval. Daarom is de filtratieduur één van de inputvariabelen.

3.3.1 Gegevens najaar 2003

De meetgegevens van het najaar zijn behoorlijk compleet. Omdat in de maand oktober drinkwater als influent is gebruikt, vond in deze maand geen noemenswaardige drukopbouw plaats (Figuur 3). De modellen zijn gekalibreerd met gegevens van 1 november t/m 2 december. Tabel 3 geeft een overzicht van de meegenomen input gegevens. Dit zijn dus de variabelen die mogelijk een invloed hebben op de outputvariabele, de drukopbouw over het membraan (dTMP) tussen twee hydraulische reinigingen.

Tabel 3: Input gegevens voor modellen op basis van gegevens november 2003.

parameter	beschrijving	eenheid	range
duur	duur van de ultrafiltratie tussen twee reinigingen	min	2 - 20
Temp	temperatuur	°C	5 - 9
Flux	flux door filter tijdens normale bedrijfsvoering	l/m ² h	45 - 51
Flush	Terugspoeling	m ³	0.0 - 0.2
pH	zuurgraad		6.1 - 9.4
Gel	geleidbaarheid		10 - 538
O2	zuurstofgehalte		1.0 - 10.1
NTU	troebelheid		0.0 - 35.3



Figuur 3: TMP gedurende november 2003.

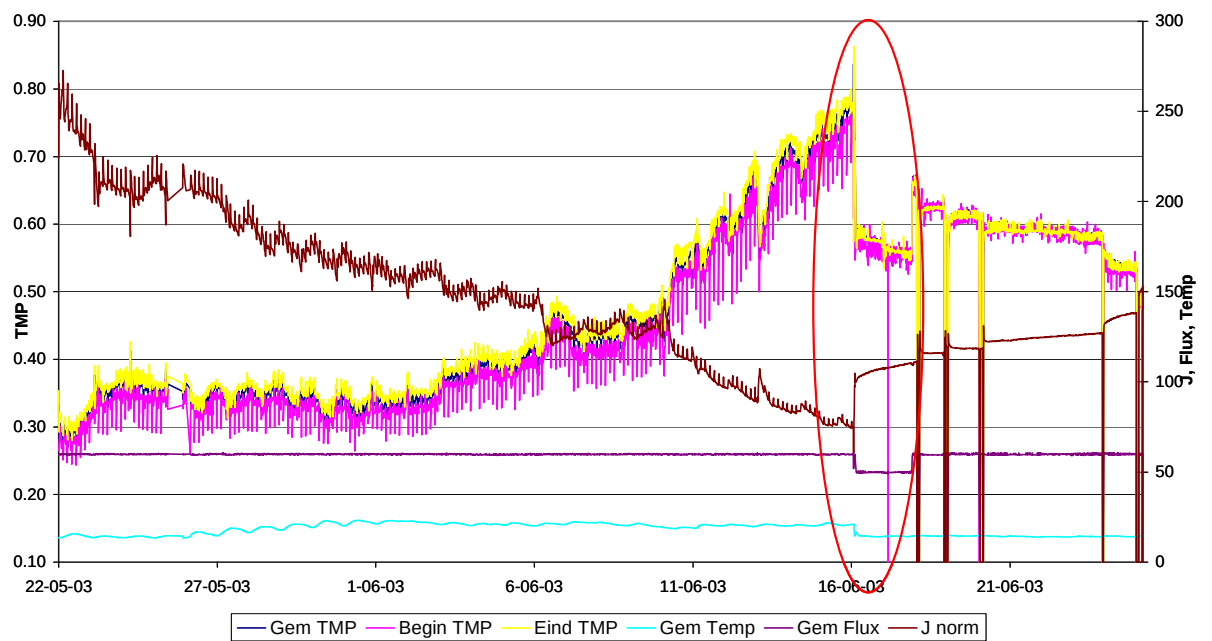
Door de vele (stroom)storingen is de representativiteit van de gegevens twijfelachtig, aangezien de normale bedrijfsvoering, met de daarbij horende relatie tussen enerzijds waterkwaliteit en procesinstellingen en anderzijds drukopbouw, regelmatig is verstoord. Bovendien valt het op dat er in een aantal gevallen sprake is van een drukafname tussen twee reinigingen in plaats van een te verwachten druktoename.

3.3.2 Gegevens voorjaar 2003

Omdat de meetgegevens van november een twijfelachtige betrouwbaarheid hebben, zijn ook de meetgegevens van half mei t/m half juni 2003 bewerkt (weergegeven in Figuur 4). Van deze periode is geen logboek beschikbaar en de troebelheidsmeting ontbreekt. Omdat de terugspoelparameters en de filtratieduur slechts matig bijdragen aan de modelnauwkeurigheid en het relatief veel tijd kost om deze parameters uit de data te halen, worden ook de terugspoelparameters en tijdsduur weggelaten uit de analyse. Tabel 4 geeft een overzicht van de meegenomen input gegevens. De gemeten TMP tussen 25 mei 12.00 uur en 26 mei 14.00 uur en van na 16 juni lijken onrealistisch en worden niet gebruikt bij de modelkalibratie en -validatie (Figuur 4).

Tabel 4: Input gegevens voor modellen op basis van gegevens voorjaar 2003.

parameter	beschrijving	eenheid	range
Temp	temperatuur	°C	13.5 – 23.3
Flux	flux door filter tijdens normale bedrijfsvoering	l/m ² /h	59 - 61
pH	zuurgraad		9.37 – 9.44
Gel	geleidbaarheid		0 - 30
O2	zuurstofgehalte		0.0 – 5.0

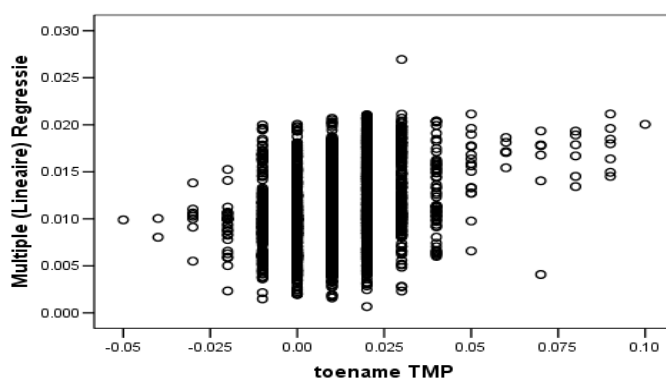


Figuur 4: TMP gedurende half mei t/m half juni 2003.

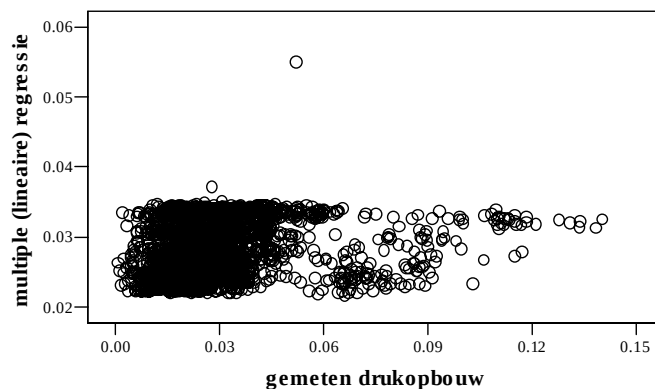
4 Resultaten en discussie

4.1 Lineaire regressie

Met behulp van lineaire regressie is het niet mogelijk de drukopbouw tijdens de ultrafiltratie goed te voorspellen. Dit blijkt uit Figuur 5 en Figuur 6, waarin de voorspelde druktoename is uitgezet tegen de gemeten druktoename. De nauwkeurigheid van de lineaire modellen bedraagt 5 à 9 % (zie ook scatterplots). De nauwkeurigheid geeft aan in welke mate de variatie van de variabele verklaard wordt door het model. Vanwege de lage nauwkeurigheid wordt een modelvalidatie niet zinvol geacht.



Figuur 5: Scatterplot lineaire regressiemodel drukopbouw op basis van novembergegevens.

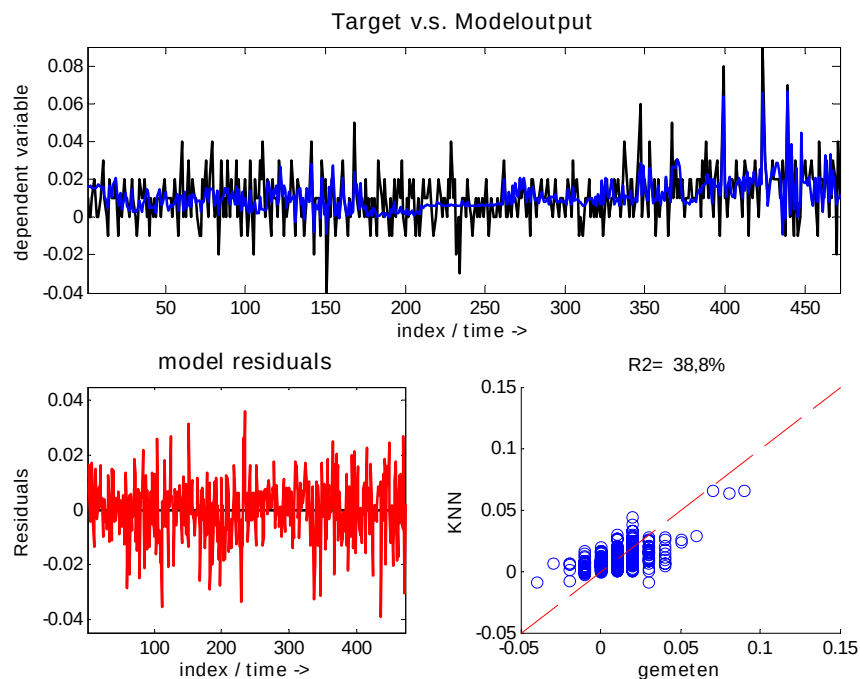


Figuur 6: Scatterplot lineaire regressiemodel drukopbouw op basis van gegevens half mei t/m half juni.

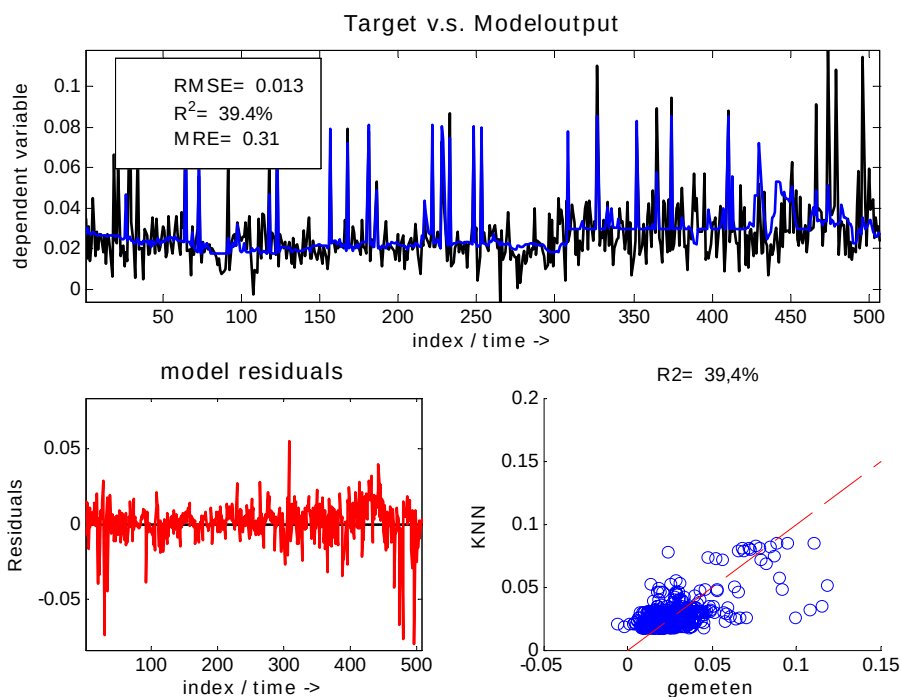
4.2 Multi-Layer Perceptron

De beschikbare dataset is opgedeeld in een kalibratieset (85% van de data) en een validatieset (15% van de data). Het neurale netwerk is vele malen opnieuw getraind op basis van steeds dezelfde kalibratieset en gevalideerd met steeds dezelfde validatieset.

Met de zogenaamde Multi-Layer Perceptron (MLP) wordt een nauwkeurigheid gehaald van ongeveer 39% op de validatieset (Figuur 7 en Figuur 8). Vanwege de lage nauwkeurigheid die behaald is, lijkt het niet zinvol om de modellen uitgebreid te verifiëren. Wat opvalt is dat een aanzienlijk aantal pieken in de drukopbouw redelijk tot goed wordt voorspeld.



Figuur 7: Modelprestaties MLP op basis van gegevens november.

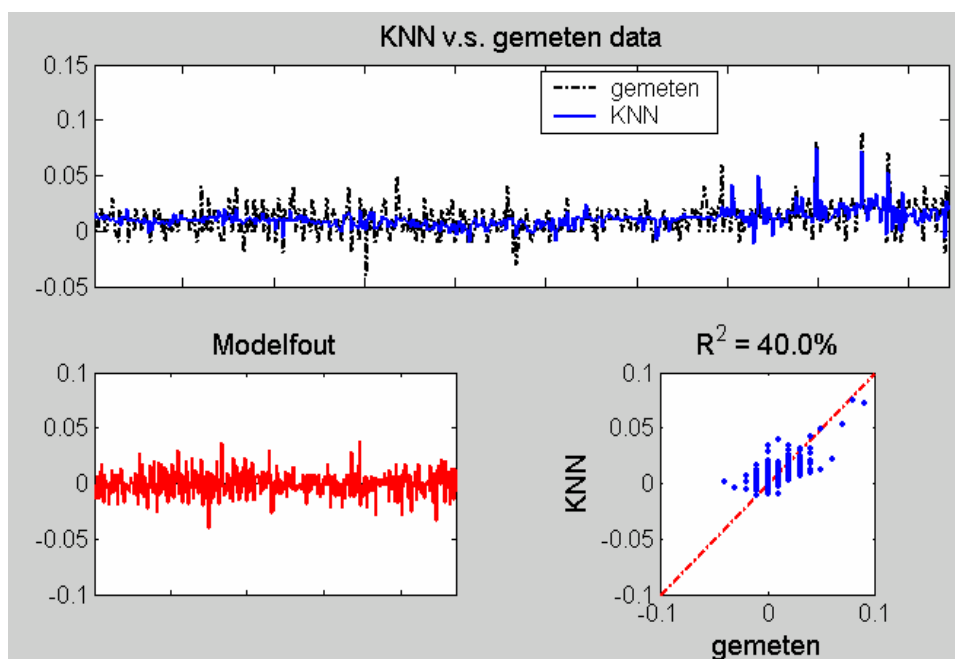


Figuur 8: Modelprestaties MLP op basis van gegevens half mei t/m half juni.

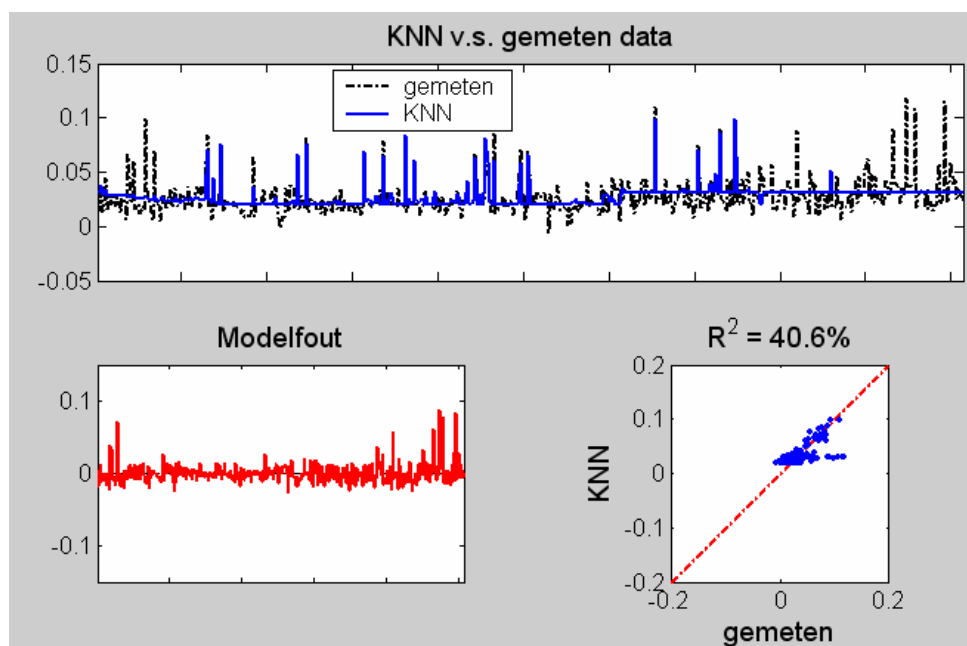
4.3 White Box Neural Network

Net als bij de MLP is de beschikbare dataset opgedeeld in een kalibratieset (85% van de data) en een validatieset (15% van de data). Het neurale netwerk is vele malen opnieuw getraind op basis van steeds dezelfde kalibratieset en gevalideerd met steeds dezelfde validatieset.

Met de zogenaamde WBNN wordt een nauwkeurigheid gehaald van ongeveer 40% op de validatieset (Figuur 9 en Figuur 10). Het resultaat is vergelijkbaar met die van een MLP.



Figuur 9: Modelprestaties WBNN op basis van gegevens november.



Figuur 10: Modelprestaties WBNN op basis van gegevens half mei t/m half juni.

Wanneer het ruwe WBNN wordt weergegeven in wiskundige termen wordt een model verkregen dat voldoet aan onderstaande vergelijking. Dit is slechts een tussenresultaat omdat het model nog niet is geoptimaliseerd. Interpretatie van deze vergelijking dient met terughoudendheid te geschieden. Het is enkel bedoeld als voorbeeld van een vergelijking verkregen met een WBNN.

$$\begin{aligned}
 &5.01 - 61.3 * [\text{TEMP}]^{\wedge-} \\
 &7.69 * [\text{FLUX}]^{\wedge 2.47} * [\text{PBEGIN}]^{\wedge 0.197} * [\text{PH}]^{\wedge 1.71} * [\text{GEL}]^{\wedge 0.0095} * [\text{NTU}]^{\wedge 0} \\
 &.193 + 4.02\text{e}+003 * [\text{TEMP}]^{\wedge-} 7.95 * [\text{FLUX}]^{\wedge 2.46} * [\text{PBEGIN}]^{\wedge-} \\
 &0.271 * [\text{VOLFILT}]^{\wedge-} 1.4\text{e}-006 * [\text{PH}]^{\wedge 0.527} * [\text{GEL}]^{\wedge 0.00329} * [\text{O2}]^{\wedge-} \\
 &0.0383 * [\text{NTU}]^{\wedge 0.219} + 7.97\text{e}+003 * [\text{TEMP}]^{\wedge-} \\
 &8.5 * [\text{FLUX}]^{\wedge 2.47} * [\text{PBEGIN}]^{\wedge-} 1.13 * [\text{O2}]^{\wedge-} 0.116 * [\text{NTU}]^{\wedge 0.258} - \\
 &1.79\text{e}+004 * [\text{TEMP}]^{\wedge-} 8.19 * [\text{FLUX}]^{\wedge 2.47} * [\text{PBEGIN}]^{\wedge-} 0.665 * [\text{O2}]^{\wedge-} \\
 &0.073 * [\text{NTU}]^{\wedge 0.239}
 \end{aligned}$$

Op basis van bovenstaande vergelijking zouden troebelheid en temperatuur als belangrijke verklarende variabelen kunnen worden gezien.

4.4 Discussie

De nauwkeurigheid varieert van 9% (lineair model) tot 40% (beide typen neurale netwerken).

Ondanks een uitgebreide controle van de gegevens op juistheid bleek achteraf toch twijfel te bestaan omtrent de kwaliteit van bepaalde gegevens. De verwachting is dat de nauwkeurigheid aanzienlijk zal toenemen wanneer de beschikbare gegevens verder worden ontdaan van onjuiste niet-representatieve gegevens in de invoer en in de uitvoer. Deze vertroebelen de uiteindelijke voorspelbaarheid van de drukopbouw aanzienlijk. Voor een vervolgfase is het dan ook aan te bevelen een procedure te ontwikkelen om meetgegevens continu of regelmatig te controleren op juistheid en representativiteit.

Daarnaast blijkt (Tabel 3 en 4) dat sommige inputvariabelen slechts zeer beperkt variëren. Bij het verzamelen van data is het van belang dat metingen worden verricht in het gehele bereik waarbinnen het proces kan opereren.

5 Conclusie en aanbevelingen

Het doel van dit onderzoek was de belangrijkste relaties te vinden tussen waterkwaliteitsparameters en procesparameters enerzijds en de transmembraandruk in praktijk ultrafiltratieinstallaties anderzijds. De verwachting was dat in de loop van de jaren genoeg data waren gegenereerd met de bestaande UF-installaties. Gebleken is echter dat in veel gevallen waterkwaliteitsdata niet online worden gemeten, of slechts een beperkt aantal parameters. De enige dataset die voldeed aan de criteria, was er een uit 2003, gemeten op een proefinstallatie van Waternet. Het doel, algemeen geldende relaties te vinden, kon met deze ene dataset niet worden bereikt. De set is echter gebruikt om na te gaan of de aanpak met neurale netwerken werkt en om daarbij naar voren komende knelpunten vast te stellen.

De betrouwbaarheid van de twee geteste Kunstmatige Neurale Netwerken is met 40% aanzienlijk hoger dan die van het lineaire regressiemodel (9%). De neurale netwerken kunnen verder worden geoptimaliseerd door meer data van verschillende installaties mee te nemen en een uitgebreide optimalisatie van het neurale netwerk zelf uit te voeren. Het gebruik van meer data kan leiden tot een beter model, omdat er een gedegen optimalisatie kan worden uitgevoerd, waarbij op onderbouwde wijze de significante variabelen kunnen worden gescheiden van de minder en niet-significante variabelen. Op basis van de reeds bereikte betrouwbaarheid en de mogelijkheden tot verbetering, is de verwachting dat Kunstmatige Neurale Netwerken een bruikbaar hulpmiddel kunnen zijn bij het modelleren en optimaliseren van ultrafiltratie. Het toepassen van White Box Neural Networks leidt in het algemeen tot relatief overzichtelijke en goed interpreteerbare relaties in vergelijking met de Multi Layer Perceptron. De nauwkeurigheid van deze laatste kan echter in de praktijk iets hoger liggen. Om relaties te vinden tussen waterkwaliteit, procesparameters en de prestaties, ligt het voor de hand een White Box Neural Network te gebruiken.

Op basis van presentatie van de resultaten aan de PBC Waterbehandeling op 9 februari 2006 is besloten dit project stop te zetten. Aanbevelingen bij het uitvoeren van een eventuele vervolgstudie zijn:

- Stel een meetprogramma op; hierbij moet rekening worden gehouden met:
 - o online meten van alle gewenste waterkwaliteitsparameters;
 - o mogelijke variaties van de waterkwaliteits- en procesparameters; De periode moet bv. zo lang zijn dat de seizoensinvloed (watertemperatuur) wordt meegenomen.
 - o meetfrequentie.
- Neurale netwerken kunnen op twee manieren worden toegepast:
 - o Als optimalisatie tool van één specifieke installatie;
 - o als generieke optimalisatie tool voor het proces in zijn algemeenheid. Er moeten gegevens beschikbaar zijn van meerdere installaties. De gegevens hoeven niet zo gedetailleerd te zijn als wanneer één specifieke installatie wordt gemodelleerd.
- Aandacht besteden aan een protocol voor het verwijderen van incorrecte data.

6 Referenties

- W.B.M. ten Brinke, F.H. Schulze, P. van der Veer, „Sand exchange between groyne-field beaches and the navigation channel of the Dutch Rhine: The impact of navigation versus river flow. River research and applications Vol. 20, 899-928, John Wiley & Sons, Ltd. (2004)
- M. Cabassud, N. Delgrange-Vincent, C. Cabassud, L. Durand-Bourlier, J.M. Lainé, Neural networks: a tool to improve UF plant productivity, *Desalination* 145 (2002) 223-231.
- N. Delgrange-Vincent, C. Cabassud, M. Cabassud, L. Durand-Bourlier, J.M. Lainé, Neural networks for long term prediction of fouling and backwash efficiency in ultrafiltration for drinking water production, *Proceedings of the Conference on Membranes in Drinking and Industrial Water Production* 1 (2000) 179-188.
- A. Gijsbertsen, Inventarisatie UF-installaties in Nederland – state-of-the-art, Kiwa BTO 2004.016.
- M.T. Hagan, H.B. Demuth, M. Beale, Neural Network Design, Boston: PWS Publishing Company (1996).
- T. Masters, Neural, Novel & Hybrid Algorithms for Time Series Prediction, Chichester: John Wiley & Sons (1995).
- H.K. Oh, M.J. Yu, E.M. Gwon, J.Y. Koo, S.G. Kim, A. Koizumi, KNT-artificial neural network model for flux prediction of ultrafiltration membrane producing drinking water, *Water Science and Technology* 50 (2004) 103-110.
- E.M. Oost, S. ten Hagen, F. Schulze, Extracting multivariate power functions from complex datasets. Proceedings of BNAIC'02, Amsterdam The Netherlands (2002).
- F.H. Schulze, A. Zijderveld, 'A White Box Neural Network to relate erosion of groyne fields sediments to the characteristics of passing vessels'. Proceedings applications of artificial neural networks in integrated watermanagement, Delft The Netherlands (1999).
- C. Teodosiu, O. Pastravanu, M. Macoveanu, Neural network models for ultrafiltration and backwashing, *Water Research* 34 (2000) 4371-4380.
- J.M.L. van der Wielen, Gestructureerd modelontwerp met Kunstmatig Neurale Netwerken, Deventer: Witteveen+Bos en Wageningen Universiteit (2001).

I Bijlage: beschikbare data

installatie + bedrijf	periode	parameter	aantal	frequentie	on/offline	opmerkingen
pilot WLB Amsterdam (Thans Waternet)	mei/juni 2003; nov. 2003	T	>>250	6 x per minuut	on	
		pH	>>250	6 x per minuut	on	
		EGV	>>250	6 x per minuut	on	
		NTU	>>250	6 x per minuut	on	
		O ₂	>>250	6 x per minuut	on	
		DOC	-	-	-	
		Uvext	-	-	-	
		SS	-	-	-	
		Chl-a	-	-	-	
		TMD	>>250	6 x per minuut	on	
		Flux	>>250	6 x per minuut	on	
		enhanced bw backwash				
Baanhoek, Evides	mei 2005 – heden (toegezegd); eerdere data kan ook (omzetten van data nodig)	T	<100 (92)	16 x per dag	on	
		T	>>250	1 x per week	off	
		pH	<100 (92)	16 x per dag	on	
		pH	>>250	1 x per week	off	
		EGV	<100 (86)	16 x per dag	off	
		NTU	>>250	1 x per week	off	
		O ₂	>>250	1 x per week	off	
		DOC	<100 (39)	1 x per maand	off	
		Uvext	<100 (85)	1 x per maand	off	
		SS	>>250	1 x per week	off	
		Chl-a	>>250	1 x per week	off	
		TMD				
		Flux				
		enhanced bw backwash				
DAF, Brabant Water	september 2005	T	?	?	on	
		pH	?	?	off	
		EGV	-	-	-	
		NTU	?	?	on	
		O ₂	-	-	-	
		DOC	-	-	-	
		Uvext	-	-	-	
		SS	-	-	-	
		Chl-a	-	-	-	
		TMD	?	?		
		Flux	?	?		
		enhanced bw	?	?		
		backwash	?	?		
Norit WbG	april en juli 2004	dosering voor UF	PAC			
		T	8		off	8 metingen
		pH	8		off	8 metingen
		EGV	8	-	off	8 metingen
		NTU	8		off	8 metingen
		O ₂	-	-	-	
		DOC	2	-	off	2 metingen
		Uvext	8	-	off	8 metingen
		SS	2	-	off	2 metingen

installatie + bedrijf	periode	parameter	aantal	frequentie	on/offline	opmerkingen
		Chl-a	3	-	off	3 metingen
		TMD	>>250	1 x per minuut	on	
		Flux	>>250	1 x per minuut	on	
		enhanced bw		elke dag 1 uur		
		backwash		elke 20/ 40 minuten circa 1 minuut		uit data volgt circa 4/5 minuten
Klazienaveen, WMD	augustus en november 2004	dosering voor UF	FeSO ₄			
		T	8	-	off	
		pH	1	-	off	
		EGV	1	-	off	
		NTU	8	-	off	
		O ₂	-	-		
		DOC	-	-		
		Uvext	8	-	off	
		SS	1	-	off	
		Chl-a	5	-	off	
		TMD	>>250	1 x per minuut	on	
		Flux	>>250	1 x per minuut	on	
		enhanced bw		elke dag 1 uur		
		backwash		elke 20/ 40 minuten circa 1 minuut		uit data volgt circa 4/5 minuten