



BTO 2015.030 | September 2015

BTO rapport

Prototype softsensor
voor het kalkafzettend
vermogen van
leidingwater

BTO

Prototype softsensor voor het kalkafzettend vermogen van leidingwater

BTO 2015.030 | September 2015

Opdrachtnummer

400554-021

Projectmanager

Stefan Kools

Opdrachtgever

BTO - Thematisch onderzoek - Nieuwe meetmethoden en sensing

Kwaliteitsborger(s)

Mirjam Blokker, Roberta Hofman-Caris

Auteur(s)

Dirk Vries, Benjamin van den Akker, Joost van Summeren

Verzonden aan

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

Jaar van publicatie
2015

Meer informatie
dr.ir. Dirk Vries
T 030 60 69 671
E dirk.vries@kwrwater.nl

Keywords
softsensor, TACC, SI, kalk

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl



BTO 2015.030 | September 2015 © KWR

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden veelevoudigd, 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.

Jaar van publicatie
2015

Meer informatie
dr.ir. Dirk Vries
T 030 60 69 671
E dirk.vries@kwrwater.nl

Keywords
softsensor, TACC, SI, kalk

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl



BTO 2015.030 | September 2015 © KWR

Alle rechten voorbehouden.

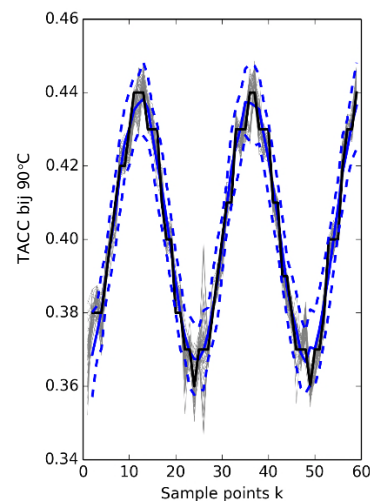
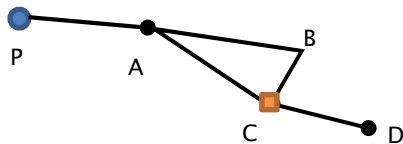
Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

BTO Managementsamenvatting

Prototype ontwikkeld van een softsensor voor bepaling van het kalkafzettend vermogen van leidingwater

Auteur(s) dr. ir. Dirk Vries

Softsensoren bieden de mogelijkheid om niet direct gemeten parameters realtime te bepalen. Softsensoren integreren (sensor)data en een model. Er is een prototype van een softsensor ontwikkeld voor het berekenen van theoretisch afzetbaar calciumcarbonaat (TACC) in leidingnetwerken, om in beeld te krijgen of kalk gaat neerslaan aan de wanden van de leidingen, drinkwaterinstallaties of huishoudelijke apparatuur. Onderdelen van die softsensor zijn een programma om chemische evenwichten en reacties te berekenen en een leidingnetmodel om het transport van water in een leidingnet te simuleren. Hiermee kunnen diverse (chemische) waterkwaliteitsparameters en de kalkafzettendheid in het gehele leidingnet worden bepaald. De twee modellen zijn aan elkaar gekoppeld en er is een statistische methode toegepast om de metingen van monsterpunten in het leidingnet optimaal te gebruiken. Resultaten met een klein voorbeeldnetwerk zijn veelbelovend, maar tonen ook aan dat de techniek doorontwikkeld dient te worden om de softsensor op full-scale leidingnetwerken in de praktijk te kunnen toepassen.



Voorspelling van de TACC bij 90°C in punt C (rechts) van het voorbeeldnetwerk (links), met veronderstelde werkelijke waarden (zwart) en de voorspelling (blauw) met een betrouwbaarheidsinterval (blauw gestippeld). In dit voorbeeld worden de waterkwaliteitsparameters bij de bron (P) bekend verondersteld en worden meetwaarden van temperatuur-, pH- en EGV-sensoren in punt C door de softsensor opgepikt.

Belang: robuuste methodiek voor bepaling kalkafzettend vermogen in het leidingnet en aan de tap
Rond de levering van water aan de tap staat momenteel de TACC sterk in de belangstelling: de parameter voor de mate waarin water

calciumcarbonaat (kalk) kan afzetten of opnemen. Deze parameter wordt berekend uit geanalyseerde waterkwaliteitsgegevens als pH, calcium en carbonaatconcentraties. Het is echter zeer kostbaar om al deze parameters op meerdere plekken in het

leidingnet te meten. Uit waterkwaliteitsgegevens af pompstation kan de waterkwaliteit - en dus de TACC - wel berekend worden, maar dan kan de waarde van de TACC lokaal afwijken van het werkelijke kalkafzettende vermogen omdat de waterkwaliteit verandert tijdens het transport door het leidingnet. Dit probleem is op te lossen door zogeheten softsensoren te gebruiken. Bij deze data-assimilatietechniek worden metingen van sensoren of meetprogramma's softwarematig vergeleken met modeluitkomsten. Door inzet van softsensoren zijn de kosten voor sensing aanzienlijk te beperken, terwijl toch real-time bepaling mogelijk is van de waterkwaliteit in de gehele keten van behandeling tot de tap. Onderzocht is of softsensoren te gebruiken zijn voor het bepalen van de TACC lokaal in het leidingnet.

Aanpak: statistische methode voor het updaten van modelberekeningen met metingen

Het kalkafzettend of -opnemend vermogen (TACC) van water is sterk afhankelijk van de waterkwaliteit af pompstation. Met een data-analyse is onderzocht wat de spreiding van de TACC is in een voorzieningsgebied en welke waterkwaliteitsparameters de meeste invloed hebben op veranderingen in de TACC. Een data-assimilatie-techniek is vervolgens uitgewerkt voor een prototype softsensor voor TACC-bepalingen. De techniek stuurt een transportmodel voor het leidingnet en een (PHREEQC-)model voor het bepalen van chemische evenwichten en reacties aan. In de praktijk zal de TACC ook beïnvloed worden door leidingnetmaterialen en opwerveling of neerslag van deeltjes, effecten die niet zijn meegenomen in het model. Deze niet-gemodelleerde effecten kunnen ondervangen worden indien de softsensor wordt gevoed met voldoende metingen.

Resultaten: prototype-softsensor toepasbaar op leidingnetwerken, maar vergt veel rekentijd

Uit de eerste testen met de softsensor in een fictief, klein leidingnetwerk blijkt dat de voorspellingen van de softsensor goed overeenkomen met de

werkelijke waarde van de TACC en andere waterkwaliteitsparameters. Een kanttekening hierbij is dat de softsensorbepalingen en de 'werkelijkheid' met hetzelfde model zijn gesimuleerd. Welke ruimtelijke resolutie en meetfrequentie nodig zijn, moet nog worden onderzocht.

Een ander aandachtspunt is de benodigde cpu-processortijd. Als voor alle knooppunten van een (middel)groot leidingnetwerk moeten worden gerekend, is de rekentijd nog te lang voor implementatie in de praktijk. Deze lange rekentijd wordt voor een groot deel veroorzaakt door het programma dat de chemische evenwichten berekent.

Implementatie: meer ontwikkel- en testtijd nodig om de softsensor in de praktijk te implementeren

De statistische methodiek met softsensoren lijkt bij uitstek toepasbaar voor waterkwaliteitsberekeningen voor grote leidingnetwerken. Doorontwikkeling is nodig om implementatie in de praktijk mogelijk te maken. Voor waterkwaliteitsberekeningen is het noodzakelijk een snel, chemisch rekenmodel te integreren of om het leidingnetmodel sterk te verdunnen. Er wordt daarom aanbevolen om bij de implementatie van een softsensor voor waterkwaliteitsparameters rekening te houden met een parallelle rekenimplementatie, ontwikkeltijd voor een snel, maar accuraat model voor waterkwaliteitsberekeningen en onderzoeksruimte voor de bepaling van optimale, voedende meetpunten voor de softsensor en validatie in de praktijk.

Rapport

Dit onderzoek is beschreven in rapport *Prototype softsensor voor het kalkafzettend vermogen van leidingwater* (BTO-2015.030) en *Soft Sensing The Potential Amount Of Calcium Carbonate Precipitate In Drinking Water Distribution Infrastructure And Warm Water Household Appliances*, International Conference on Hydroinformatics 2014.

Jaar van publicatie
2015

Meer informatie

dr.ir. Dirk Vries
T 030 60 69 671
E dirk.vries@kwrwater.nl

Keywords

softsensor, TACC, SI, kalk

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl



BTO 2015.030 | September 2015 © KWR

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden vervoelvoudigd, 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.

Jaar van publicatie
2015

Meer informatie
dr.ir. Dirk Vries
T 030 60 69 671
E dirk.vries@kwrwater.nl

Keywords
softsensor, TACC, SI, kalk

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl



BTO 2015.030 | September 2015 © KWR

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.

Voorwoord

Nieuwe ontwikkelingen op het gebied van sensing en data-analyse maken het mogelijk het waterkwaliteitsbeeld zowel in de tijd als in de ruimte continu te presenteren. Een techniek die steeds meer wordt toegepast is de real-time verwerking van (online) sensoren en modelberekeningen om real-time informatie te genereren over verschillende, ook niet direct gemeten parameters. Toepassing van deze, zogenaamde, softsensoren bieden een kans om het waterkwaliteitsbeeld met hoge ruimtelijke resolutie en op snelle wijze te presenteren.

Hiermee kunnen aanzienlijke kosten aan sensing voorkomen worden, terwijl monitoring van de waterkwaliteit in de gehele keten van behandeling tot de tap mogelijk is. In dit werk bestuderen we de methode achter de softsensor en passen deze toe voor een praktijkvoorbeeld: het bepalen van de parameter voor kalkafzettendheid.

Voorliggend werk is het resultaat van de projecten: 'Softsensors 2.0' en het verlengde van voorgaand project 'Softsensors voor de waterkwaliteit'. In het laatst genoemde project is verkend welke indicator gewenst is om een beeld van de waterkwaliteit te krijgen en is besloten om het kalkafzettend en -opnemend vermogen van leidingwater als indicator te kiezen. Vervolgens is een analyse uitgevoerd op parameters die dit kalkafzettend vermogen beïnvloeden en is een begin gemaakt met het ontwerp van de softsensor. Het tweede project omvatte de ontwikkeling van een (prototype) softsensor voor de bepaling van deze parameter.

In dit woord vooraf willen wij de leden van de begeleidingsgroep - Alex van der Helm (Waternet), Perry van der Marel (WLN), Ignaz Worm (toentertijd PWN, nu Isle Utilities), Rein Wuestman (Vitens), Luc Zandvliet (HWL), Luuk Rietveld (TU Delft) - bedanken voor hun inbreng en kritische blik op de tussenresultaten. Ook de BTO-themagroep Nieuwe meetmethoden & Sensing willen wij danken voor hun interesse in en evaluatie van dit onderzoek.

Inhoud

| | |
|--|-----------|
| Voorwoord | 3 |
| Inhoud | 4 |
| 1 Inleiding | 5 |
| 1.1 Belang van monitoring van de waterkwaliteit | 5 |
| 1.2 Doelstellingen en aanpak | 6 |
| 1.3 Leeswijzer | 6 |
| 2 Verkenning van mogelijkheden voor toepassing van softsensoren in de Nederlandse watersector | 7 |
| 2.1 Doel en aanpak van de verkenning | 7 |
| 2.2 Bevindingen verkenning | 7 |
| 2.3 Keuze indicatorparameter softsensor | 8 |
| 3 Parameters voor bepaling kalkafzettendheid en methodiek softsensor | 11 |
| 3.1 Kalk in leidingwater | 11 |
| 3.2 Data-analyse en bepaling van de TACC in een voorzieningsgebied | 12 |
| 3.3 Methodiek voor softsensor | 12 |
| 3.4 Instellingen simulatie – schema 2 (stress test) | 16 |
| 4 Resultaten | 18 |
| 4.1 Data-analyse van waterkwaliteitsparameters | 18 |
| 4.2 Prestatie van de softsensor | 20 |
| 4.3 Simulatie met de softsensor – schema 1 | 20 |
| 4.4 Simulatie met de softsensor – schema 2 | 23 |
| 5 Conclusies en aanbevelingen | 27 |
| 6 Literatuur | 30 |
| Bijlage I SI in het leidingnet van Amsterdam | 32 |
| Bijlage II Broncode ensemble Kalman Filter | 33 |

1 Inleiding

1.1 Belang van monitoring van de waterkwaliteit

Waterbedrijven behouden graag het vertrouwen van de kritische en snel ingelichte consument in een goede waterkwaliteit aan de tap, 24 uur per dag. Voor diverse waterkwaliteitsparameters bestaan wettelijke normen en vaak strengere bedrijfsnormen. Een door de waterbedrijven veelgebruikt middel om te voldoen aan de wettelijke verplichting te toetsen of de waterkwaliteit en de bedrijfsvoering (zuivering en conditie van het leidingnet) aan de doelstellingen voldoen, is het op regelmatige basis monsters nemen in het proces, van het uitgaande reine water af pompstation, en in het leidingnet, en deze te analyseren in het laboratorium. Daarnaast worden er ook on-line metingen toegepast. Echter, met de huidige bemonsteringsfrequentie (maandelijks tot wekelijks) is de kans aanwezig dat besmettingen of (kortdurende) perioden met een verminderde waterkwaliteit af pompstation, niet direct gedetecteerd en beoordeeld kunnen worden op verhoogde gezondheidsrisico's [1].

Real-time of hoog-frequente toetsing van de waterkwaliteit door inzet van sensoren wordt voornamelijk bemoeilijkt door:

- *Technische factoren*: niet alle waterkwaliteitsparameters kunnen on-line gemeten worden.
- *Kosten*: de aanschaf en met name het beheer van (vaak dure) sensoren brengen aanzienlijke kosten met zich mee. Het is (op dit moment) kostentechnisch en beheerstechnisch niet haalbaar om op elk uiteinde van het leidingnet sensoren te plaatsen.
- *Frequentie van monitoring*: alhoewel in de winning en in de zuivering het proces constant en online gemonitord wordt op kwaliteits- en procesparameters, gebeurt deze online monitoring op een beperkt aantal waterkwaliteitsparameters (met uitzondering van pH en temperatuur). Monitoring gebeurt veelal met sensoren die op industriële schaal in de proces- en (afval)water-industrie hun dienst bewijzen, zoals sensoren voor druk, pH, temperatuur en volumestroom. In het leidingnet wordt voornamelijk druk en volumestroom on-line gemeten, en slechts zeer beperkt waterkwaliteitsparameters. Alleen Vitens heeft op een aantal punten in het net (naast volumestroom- en druk-), conductiviteit- en temperatuursensoren in het leidingnet rondom Leeuwarden (Vitens Innovation Playground) geïnstalleerd. Daarnaast wordt op regelmatige basis het water op verscheidene tappunten in het leidingnet op een reeks waterkwaliteitsparameters in een laboratorium geanalyseerd.

Kortom, in hoeverre de frequentie van de huidige bemonsteringen in combinatie met laboratoriumanalyses en online metingen afdoende is om de effectiviteit van de zuiveringsstap en integriteit van het leidingnet te toetsen, blijft nog steeds een actuele vraag [1]. Daarnaast spelen er nog andere vragen, namelijk, welke online sensing technieken worden er in en buiten de Nederlandse drinkwatersector gebruikt? En kunnen deze dan effectiever worden ingezet voor monitoring van de waterkwaliteit?

Een oplossing voor de beperkte, realiseerbare inzet van sensoren en de wens grote relevante delen van het leidingnet te monitoren, is datafusie. Datafusie, of ook wel data-assimilatie genoemd, biedt namelijk de mogelijkheid om deze datastromen te combineren met

modelberekeningen [2, 3] en niet-gemeten parameters te bepalen. Door het gebruik van modelberekeningen (software) wordt de implementatie van de methodiek ook softsensor genoemd.

1.2 Doelstellingen en aanpak

Bovenstaande is aanleiding geweest om een overzicht van online sensing technieken te verkrijgen, om vervolgens de keus van indicatorparameter voor het waterkwaliteitsbeeld vast te leggen en de ontwikkeling van de sensor 'in silico'¹ te starten. In navolging van de termen "in vivo" en "in vitro", wordt tegenwoordig "in silico" gebruikt om aan te geven dat het hier een mathematische modellering betreft. De einddoelstelling in dit project is geformuleerd als: *Ontwikkel een methode om het waterkwaliteitsbeeld op elk willekeurig punt in het leidingnet te voorspellen, uit beschikbare meetdata in de zuivering (uitgaand water) en het leidingnet.*

In dit project is het onderzoek naar de softsensor uitgevoerd met de volgende stappen:

- **Verkenning** van welke indicator gewenst is om een beeld van de waterkwaliteit te krijgen. Bij deze verkenning is eveneens onderzocht welke online sensoren er anno 2014 in de waterinfrastructuur toegepast worden. Aan de hand van de verkenning en consultatie met de waterbedrijven, is gekozen om een softsensor te ontwikkelen voor een parameter die gerelateerd is aan de hardheid van water en kalkproblematiek. Specifiek is gekozen voor een sensor die de TACC (theoretisch afzetbaar calciumcarbonaat) bepaalt. Een indicatorparameter bepaalt waar de softsensor een beeld van moet geven.
- **Data-analyse.** Alvorens de ontwikkeling van softsensor is gestart, is bekeken wat de invloed is van diverse parameters op de TACC. Hiertoe zijn meetgegevens verstrekt door HWL gebruikt voor berekeningen van de TACC op verschillende punten in het leidingnet van Amsterdam.
- **Ontwikkeling** van een prototype softsensor voor de bepaling van TACC in het leidingnet.

1.3 Leeswijzer

De verkenning van de toepassing van online-sensoren en softsensoren in de Nederlandse drinkwatersector wordt toegelicht en besproken in Hoofdstuk 2. Voorafgaande aan de ontwikkeling van deze softsensor, is een data-analyse uitgevoerd welke wordt besproken in de eerste paragraaf van hoofdstuk 3. De hier ontwikkelde softsensor voorspelt de parameters TACC en SI voor elke willekeurige, door de gebruiker gespecificeerde locaties in het leidingnet *benedenstrooms* van de *online* invoer van de softsensor. Hoofdstuk 3 behandelt de methodiek achter de softsensor. In hoofdstuk 4 laten we de resultaten van een simulatie met de softsensor zien voor een fictief netwerk, als ook de voorstudie van TACC-bepalingen in een werkelijk leidingnet. In de bijlage is de code van de softsensor vermeld.

¹ Sensor op silicium, dat wil zeggen een softsensor die op een computerchip werkt.

2 Verkenning van mogelijkheden voor toepassing van softsensoren in de Nederlandse watersector

2.1 Doel en aanpak van de verkenning

Aan de start van dit onderzoek is verkend welke indicator gewenst is om een beeld van de waterkwaliteit te krijgen. Bij deze verkenning is eveneens onderzocht welke online sensoren er anno 2014 in de waterinfrastructuur toegepast worden. Dit hoofdstuk geeft een beeld van de uitkomsten van de verkenning middels een vragenronde, waarbij contactpersonen bij waterbedrijven via de themagroep Nieuwe Meetmethoden en Sensoring² zijn geconsulteerd.

Aan de hand van de verkenning en een discussie met de themagroep Nieuwe Meetmethoden en Sensoring is een indicatorparameter bepaald waar de softsensor een beeld van moet geven. Voor deze keuze zijn de volgende criteria gehanteerd:

- Een indicatorparameter voor de waterkwaliteit die goed bekend en beschreven is in (wetenschappelijke) literatuur, zodat er aan de ontwikkeling van de softsensor niet (te)veel ontwikkeltijd in modellering hoeft te worden gestoken;
- Een indicatorparameter die relevant is voor consument en waterbedrijf.

Ten behoeve van een gestructureerd overzicht van de bevindingen zijn de volgende categorieën gedefinieerd, met de eerste drie categorieën naar analogie met het Drinkwaterbesluit:

1. (Real-time) microbiologische parameters zoals E. coli, Enterococci, Cryptosporidium, Campylobacter, virussen, Giardia en bacteriofagen;
2. (Real-time) chemische parameters, zoals organische microverontreinigingen, zware metalen, nitraat en nitriet;
3. Indicatoren/bedrijfstechnische parameters:
 - a. Waterkwaliteitsparameters gericht op esthetische kenmerken en klantperceptie (geur, kleur, smaak, hardheid, deeltjes, waterstofcarbonaat, SI);
 - b. Waterkwaliteitsparameters die een indicatie geven van microbiologische activiteit of groei bevorderende factoren (zoals: Koloniegetal bij 22°C, bacteriën van de coligroep, Aeromonas bij 30 °C, DOC/TOC);
 - c. Waterkwaliteitsparameters die een indicatie geven van radioactiviteit;
 - d. Indicatorparameters die een indicatie geven van de algehele waterkwaliteit als zuurstof, zuurgraad, temperatuur en geleidingsvermogen.
4. 'Fingerprinting' van waterkwaliteit (zoals fotospectrometrische methoden).

2.2 Bevindingen verkenning

Uit de inventarisatie blijkt dat er, naast de sensortechnieken die een beeld presenteren van de waterkwantiteit, of verlies ervan (zie bijvoorbeeld het werk van Takadu bij Evides [4], de VLPV-

*VLPV-methode [5] of de toepassing van machine learning technieken beschreven in [6], er meerdere real-time*² *In deze themagroep zijn de volgende bedrijven vertegenwoordigd: WLN (tevens vertegenwoordiger WMD), PWN, HWL (tevens vertegenwoordiger Dunea en PWN), Vitens, Waternet, Evides en ALZ (tevens vertegenwoordiger BrabantWater).*

meerdere real-time sensortechnieken om de waterkwaliteit te monitoren en/of te sturen ontwikkeld ontwikkeld worden of zijn ontwikkeld. Een overzicht van de bevindingen van deze inventarisatie is inventarisatie is weergegeven in Tabel 2-1. Vanuit het oogpunt van monitoring van de waterkwaliteit op meerdere locaties, is effectievere benutting van deze sensormetingen door datafusie een uitgelezen kans. Uit de verkenning blijkt dan ook dat enkele waterbedrijven de handschoen hebben opgepakt en in consortia deze sensortechnieken hebben geïntegreerd met modellen om de (meng)waterkwaliteit te voorspellen, of om een andere parameter dan dat wat wordt gemeten in te schatten. Deze softsensoren worden voornamelijk daar toegepast waar sensoren beschikbaar zijn, namelijk in de winning en zuivering, bijvoorbeeld de hardheid-softsensor in de onthardingsstap bij de waterzuivering [7], de softsensor in oppervlaktewaterzuivering De Punt ten behoeve van een softsensor-gestuurde conditionering en de softsensoren in het Self-Learning Integrated Model Management-project (SLIMM) [8-10], zie

Tabel 2-2.

2.3 Keuze indicatorparameter softsensor

Aan de hand van de verkenning, de in Paragraaf 2.1 genoemde keuzecriteria en een discussie met de themagroep Nieuwe Meetmethoden en Sensoring is gekozen om een softsensor te ontwikkelen voor een indicatorparameter die gericht is op esthetische kenmerken en klantperceptie, namelijk de parameter die een indicatie geeft van de hardheid van water en kalkproblematiek (categorie 3a). Specifiek is gekozen voor een sensor die de TACC (theoretisch afzetbaar calciumcarbonaat) kan bepalen.

Tabel 2-1: reeds toegepaste real-time detectietechnieken (geen softsensoren). Gebruikte afkortingen: W: winning, Z: zuivering, D: distributie, WQ: waterkwaliteit. De nummering in de categorieën komt overeen met de categorisatie in paragraaf 2.1.

| Naam product/project | Meetgrootheid | Categorie | Waar | Methode | Uitvoerder/fabrikant | Waterbedrijf/waterschappen | Referentie |
|----------------------|--|-----------|------|----------------------|----------------------|----------------------------|------------|
| hydrochip | fytobenthos, fytoplankton | 1 | W | DNA-analyse | TNO | Vitens, Waternet | [11] |
| <diverse> | pH, EGV, troebelheid, zuurstof | 3 | W, Z | diverse | diverse | Alle | |
| SLIMM-project | deeltjesgrootte, redoxreacties, chloride | 3 | Z | diverse | diverse | Vitens, Waternet, WMD/WBG | [10] |
| Spectro::lyser | AOC en ozon | 2 en/of 4 | Z | UV-spectrofotometrie | s::can | Waternet | [12] |
| Eventlab | patroon brekingsindex | 4 | D | brekings-index | Optiqua | Vitens | [9] |

Tabel 2-2: reeds toegepaste soft-sensoren en soft-sensoren die in een ontwikkelingsstadium* zijn. Gebruikte afkortingen: W: winning, Z: zuivering, D: distributie. De nummering in de categorieën komt overeen met de categorisatie in paragraaf 2.1.

| Naam product | Doel & toepassing (W/Z/D) | Categorie | Methode | Uitvoerder/fabrikant | Waterbedrijf/waterschappen | Referentie |
|----------------------------------|--|-----------|--|----------------------|----------------------------|-------------|
| Sturing ontharding | Voorspelling TACC na ontharding (Z) | 3 | Model-gebaseerde sturing | RHDHV | Waternet | [13] |
| Sturing van pH | Beheersing SI, WPB De Punt (Z) | 3 | Model-gebaseerde sturing | | Waterbedrijf Groningen | [bron: WLN] |
| SLIMM-sensor mengwaterkwaliteit* | Voorspelling mengwaterkwaliteit putschakelingen (W) | 2 en 3 | Speciatie m.b.v. IPhreeqc | RHDHV | Vitens | [9] |
| SLIMM-sensor kalkgehalte* | Voorspelling waterkwaliteit ontharding (Z) | 3a | geen informatie | RHDHV | Vitens | [9] |
| i::scan i.c.m. black-boxmodel* | Adaptief geoptimaliseerde spoeling van snelfiltratie (Z) | 3 | Model-gebaseerde sturing i.c.m. UV-spectrofotometrie en troebelheidsmetingen | s::can | Waterbedrijf Groningen | [bron: WLN] |
| SLIMM-sensor ijzer* | Voorspelling snelfiltraatkwaliteit (Z) | 3 | Datagedreven-model | o.a. RHDHV | Vitens | [9] |

3 Parameters voor bepaling kalkafzettendheid en methodiek softsensor

3.1 Kalk in leidingwater

Calciumcarbonaat (CaCO_3), in de volksmond kalk genoemd, kan bij een bepaalde samenstelling van het water en temperatuur precipiteren en neerslaan op de wanden van leidingen of op de wanden van (huishoudelijke) apparatuur, zoals die van verwarmingselementen. Het is vooralsnog niet aangetoond [14] dat hoge of juist zeer lage concentraties van calcium en carbonaat, in Nederland tot gezondheidsproblemen leiden. De kalkafzettingen hinderen echter wel de warmteoverdracht en leiden daardoor tot een verhoogd energieverbruik, tot versnelde slijtage en tot een verhoogde frequentie van onderhoud bij huishoudelijke en industriële warmwatertoestellen. In koud water leidt kalkafzetting op den duur tot verstopte watermeters en perlatoren in kranen [15]. Daarnaast is er veel meer wasmiddel nodig bij hoge calciumcarbonaatgehalten, wat belastend is voor het milieu.

Het mechanisme van kalkafzetting is bekend en wordt kort in het navolgende herhaald. Neerslag van calciumcarbonaat ontstaat in enkele stappen:

1. door chemische samenstelling of temperatuur kan de oplosbaarheid van calciumcarbonaat overschreden worden. Het oplosbaarheidsproduct geeft de grens van verzadiging van calciumcarbonaat in een oplossing aan;
2. het transport van de calcium- en carbonaationen naar een oppervlak waar kristallisatie kan plaatsvinden;
3. De hechting van deze ionen aan dit oppervlak.

Bij bovenstaande stappen zijn allerlei factoren betrokken die invloed hebben op het kalkevenwicht of de vorming(snelheid) en aanhechting van kalk aan de omliggende materialen: watertype, temperatuur, affiniteit voor hechtoppervlak op leidingmateriaal, snelheid van neerslag- en kristallisatievorming en de hydrodynamische omstandigheden in de leiding of ruimte waarin het water zich bevindt. Pas wanneer er een model bestaat dat met *alle* factoren rekening houdt, zal de opname of hechting van kalk als mechanisme beschreven kunnen worden. Omdat een dergelijk model (nog) niet is beschreven, beperken we ons tot stap (1), zeker ook omdat de stappen 2 en 3 in de praktijk veel minder kritisch zijn. Stap (1) omvat de berekening van het kalkevenwicht en bepaalt, indien de verzadigingsindex is overschreden, de theoretische mogelijkheid dat een hoeveelheid kalk neerslaat [16]. De verzadigingsindex is een maat voor de mate van verzadiging: is er onderverzadiging ($SI < 0$), dan zal calciumcarbonaat niet neerslaan en blijven de calcium- en carbonaationen in oplossing. Bij oververzadiging ($SI > 0$) is er wel de mogelijkheid tot neerslagvorming, maar is nog niet bekend of neerslag *daadwerkelijk* wordt gevormd (ook afhankelijk van beschikbaar kristallisatie-oppervlak) en *hoeveel* kalk er neerslaat. Dit laatste aspect kan worden bepaald door de TACC, theoretisch afzetbaar calciumcarbonaat te bepalen. Vergelijkingen voor de SI en de TACC zijn beschreven in [17]. De PACC³ (in de

³ De PACCK wordt bepaald door het uitvoeren van een kookproef, zodat ook het effect van kristallisatiekiemen en -oppervlak wordt meegenomen.

Praktijk Afzetbaar CalciumCarbonaat) is een maat die aangeeft wat er daadwerkelijk aan kalk kan neerslaan [18], echter, een rekenmodel dat met alle relevante factoren rekening houdt, ontbreekt.

3.2 Data-analyse en bepaling van de TACC in een voorzieningsgebied

Ontwikkeling van een softsensor voor de real-time bepaling van TACC in het leidingnet heeft alleen nut indien er ook daadwerkelijk variatie van de TACC kan optreden door menging van verschillende watertypen, pendelzones van watertypen met een bepaalde samenstelling of andere factoren die een invloed hebben op de dynamiek in de TACC.

Voorafgaande aan de ontwikkeling van de softsensor, is daarom:

- onderzocht of enerzijds de TACC en SI reproduceerbaar kan worden bepaald door gebruik te maken van software die chemische speciatie kan bepalen (PHREEQC);
- verkend in welke mate de TACC kan variëren in de praktijk.

Waternet heeft daartoe een dataset aangeboden met meetgegevens afkomstig van verschillende meetlocaties in het leidingnetwerk van Amsterdam en bepaald door HWL. De beschikbare gegevens betroffen de parameters: carbonaat, sulfaat, zuurgraad (pH), temperatuur, chloride, nitraat, calcium, natrium, kalium, magnesium als ook EGV-metingen en bepalingen van de TACC. De TACC is aan de hand van deze gegevens door PHREEQC bepaald, zie Paragraaf 4.1.

3.3 Methodiek voor softsensor

3.3.1 Inleiding

Het doel van de softsensor is om de TACC bij hoge en lage temperaturen van het leidingwater te bepalen. Hiertoe worden met data-assimilatie (on-line) meetgegevens als volumestroom, pH en temperatuur van water gecombineerd met modelberekeningen. Net als bij weersvoorspellingen is het voor de beoordeling van de resultaten van belang om een onzekerheidsband weer te geven. Deze onzekerheidsband heeft twee componenten: de variabiliteit in de meetgegevens en de variatie in modelaannames. Veel algoritmes van data-assimilatie maken al gebruik van een methodiek om deze onzekerheidsband te berekenen.

Het bepalen van water(kwaliteits)parameters in een waterleidingnetwerk is echter geen gemakkelijke opgave: (1) het netwerk heeft typisch een complexe structuur met een omvang van meer dan 1000 knooppunten, (2) de neerslag- en evenwichtsreacties zijn niet-lineaire vergelijkingen en (3) de update van een berekening bij data-assimilatie bij de volgende tijdstap vereist het berekenen van de inverse van de covariantiematrix⁴. Deze laatste stap kost veel computerprocessorkracht, en dus rekentijd, indien de inverse van de covariantiematrix wordt berekend. In het bijzonder zal een dergelijke rekenmethode veel processorkracht [19] kosten vanwege de netwerkkarakteristieken⁵.

Een alternatief is om gebruik te maken van een andere methode om de onzekerheidsband te berekenen. Een data-assimilatietechniek hiervoor is ontwikkeld in weer- en klimaatonderzoek [20], [21], een werkveld waar typisch sprake is van een groot aantal modelvariabelen, een ruimtelijk (fijn) grid en slechts een beperkt aantal observatiepunten. Het achterliggende algoritme wordt *ensemble Kalman filtering* (enKF) genoemd, omdat er met behulp van *ensembles* een analoge manier van data-assimilatie wordt toegepast. Vergelijkbare problematiek speelt met een beperkt aantal bemonsteringslocaties in leidingnetwerken en de

⁴ De covariantiematrix geeft de mate van spreiding aan bij meet- of rekenresultaten.

⁵ Voor een netwerk met n_n knooppunten en n_s waterkwaliteitsparameters kost het gebruik van een extended Kalman Filter in de orde $(n_s \cdot n_n)^3$ operaties.

wens om over een totaalbeeld van de waterkwaliteit te beschikken. Het gebruik van enKF is een stuk minder processorintensief⁶ omdat de covariantie wordt geschat uit een groot aantal parallelle kopieën van het model, een ensemble set, gelijktijdig worden gesimuleerd en daaruit de onzekerheidsband kan worden bepaald.

In dit werk is de enKF geïmplementeerd om, gegeven de netwerkconfiguratie en een aantal punten in het netwerk waar metingen *continu* beschikbaar zijn, *continu* een zo goed mogelijke schatting van de verzadigingsindex (SI) van calciumcarbonaat en TACC in het leidingnetwerk te bepalen. Bij dit vraagstuk wordt uitgegaan van het volgende:

- een transport-model die het transport van water met een bepaalde samenstelling door een leidingnet simuleert;
- een chemisch reactie-model die effecten als menging, verzadigingsindex, neerslag en TACC kan berekenen,
- en tenslotte het enKF-algoritme dat een continue stroom metingen combineert met de modelberekeningen, een schatting van de SI en TACC geeft en de onzekerheidsband (betrouwbaarheidsinterval) van de soft-sensorresultaten bepaalt.

3.3.2 Lay-out en werkingsprincipe softsensor

De lay-out van de softsensor bestaat uit twee componenten:

1. een model dat het transport van water simuleert en eveneens de chemische speciatie en reacties in het leidingnetwerk berekent, en
2. het data-assimilatie-algoritme (het enKF) die
 - a. recursief de modelresultaten van de waterkwaliteitsparameters update op elk (gedefinieerd) knooppunt
 - b. de covariantie van alle berekende en gemeten waterkwaliteitsparameters bepaalt.

Het eerste component, de kern van de softsensor, wordt in dit werk berekend door een speciatie-transportmodel geschreven als EPANET-MSX code [22] en PHREEQC. De koppeling met PHREEQC dient om de kalkevenwichten te (her)berekenen en de eventueel opgenomen of neergeslagen kalk te bepalen, omdat de MSX module niet⁷ voldeed om deze berekeningen correct uit te voeren. Waterkwaliteit en de TACC worden in PHREEQC berekend volgens de richtlijnen als gespecificeerd in [17]. De volgende stappen zijn genomen in de softsensorwerking, waarbij EPANET-MSX wordt afgekort met EPANET.

De gebruiker specificeert:

- een instelwaarde of (ingeschat) drukpatroon op het distributienetwerk voor een bepaalde tijdshorizon $h_{tk} = [t_k, t_{k+h}]$, voor alle input- of bronknooppunten. Horizon h is typisch enkele uren tot enkele dagen. Bijvoorbeeld, met $h = 24u$ wordt indien de on-line bemonsteringsfrequentie 1 minuut is een raam van 24 uur (1440 minuten) steeds een minuut verschoven. Dus op tijdstip 0 rekent de softsensor in $h_{t_0} = [t_0, t_{4400}]$, dan 1 minuut later op t_1 : $h_{t_1} = [t_1, t_{4401}]$, et cetera.
- welke input- (of bron)knooppunten van het leidingnetwerkmodel de softsensor gaat gebruiken (set \hat{i}). Op deze locaties worden on-line metingen van

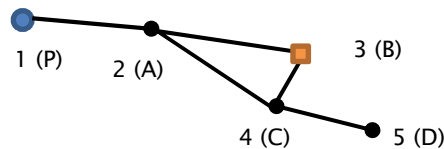
⁶ Voor een netwerk met n_n knooppunten en n_x waterkwaliteitsparameters kost het gebruik van een *ensemble* Kalman Filter in de orde $(q \cdot p \cdot n_x \cdot n_n)$ operaties, met q het aantal ensemble leden, p het aantal geobserveerde waterkwaliteitsparameters.

⁷ Resultaten van de MSX-berekeningen voor speciatie zijn niet gepresenteerd in dit rapport.

waterkwaliteitsparameters verricht, zoals pH, temperatuur, en mogelijk calcium en carbonaat.

- welke knooppunten van het leidingnetwerkmodel corresponderen met locaties waar met een zekere regelmaat enkele waterkwaliteitsparameters worden gemeten in het kader van monitoringsprogramma's, of vanwege de aanwezigheid van on-line sensoren. Deze set aan knooppunten duiden we aan met j .

De softsensor is getest met een klein voorbeeldnetwerk met 5 knooppunten, inclusief een bronknooppunt, zie Figuur 3-1. Het voorbeeldnetwerk is overgenomen van Shang et al. [22].



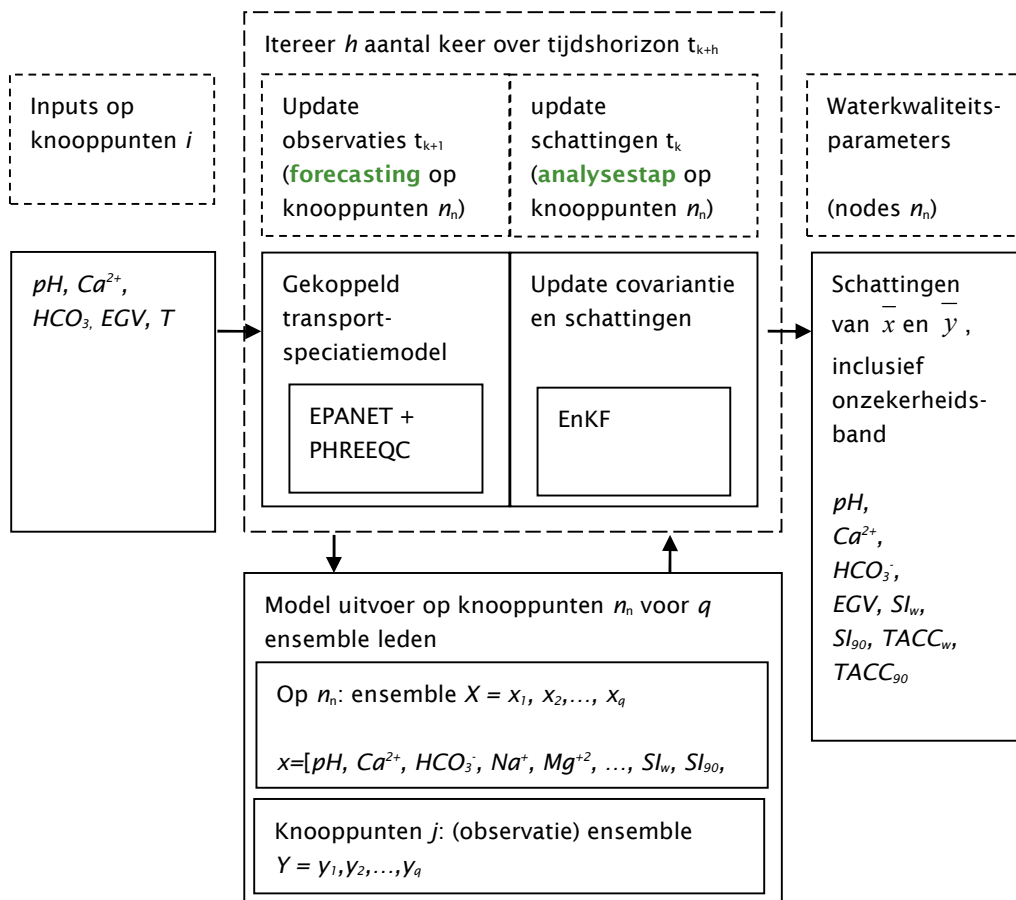
Figuur 3-1: Voorbeeld netwerkmodel, met bronknooppunt $i=1$ (blauwe cirkel P), knooppunten waar waterkwaliteitsparameters worden berekend $n \in N_n = \{2, 3, 4, 5\}$ en het observatieknooppunt waar (chemische) analyses worden uitgevoerd $j=3$ (oranje vierkant B).

Aan de hand van bovenstaande condities en een continue stroom aan meetgegevens, zal het Kalman Filter van de softsensor over de horizon h recursief:

1. het **transport van water** in het tijdsframe $[t_k, t_{k+1}]$ voor het gehele distributienetwerk **berekenen**, gebruikmakend van enerzijds geüpdatete **waterkwaliteitsparameters** op de (bron)knooppunten i en anderzijds een vorige *schatting* van de waterkwaliteitsparameters op alle overige knooppunten $n_n = N_t - i$, met N_t het totale aantal knooppunten. Dit is de modelpredictie-stap of **forecast**-stap [19][21]. Bij het ensemble Kalman filter wordt gebruik gemaakt van een set (*ensemble*) aan modeluitkomsten, waarbij het ensemble gegenereerd kan worden met modeluitkomsten en verschillende (additieve) ruistertermen en/of een set aan modeluitkomsten waarbij elke modeluitkomst correspondeert met een andere beginconditie. In dit werk is gekozen voor het genereren van verschillende modeluitkomsten door (random) ruistertermen toe te voegen aan de model- en gemeten signalen.
2. de **modeluitkomsten verzamelen**;
3. de vorige schattingen van **waterkwaliteitsparameters** (inclusief de SI en TACC) **updaten** – dit wordt de **analysestap** genoemd;
4. de **covariantiematrix** van alle parameters op alle knooppunten n_n **berekenen** en, tegelijkertijd de schattingen van de **waterkwaliteitsparameters** op de observatieknooppunten j **updaten**.

In stap 1 wordt EPANET aangeroepen, waarbij bij elk knooppunt apart de invoer voor PHREEQC wordt weggeschreven en vervolgens wordt ingelezen door PHREEQC. Na speciatie schrijft PHREEQC de resultaten weg, die dan door EPANET in worden ingelezen. De waterpakketjes worden getransporteerd gedurende een bepaalde, opgegeven tijdshorizon. Op de achtergrond vangt het Kalman Filter de resultaten op in stap 2 en stuurt de berekeningen aan.

De set-up van het model is schematisch weergegeven in Figuur 3-2.



Figuur 3-2: schematisch overzicht van data-assimilatie door softsensor, waarbij meetwaarden binnenkomen op knooppunten i en j . De SI en $TACC$ worden voor alle knooppunten berekend.

3.3.3 Instellingen simulatie – schema 1

De eerste simulatie met de softsensor is uitgevoerd door een aantal gesimuleerde instellingen: aan de bron worden calcium, waterstofcarbonaat en temperatuur gevarieerd met een sinusvormige trend gedurende een duur van 24 uur. De pH wordt constant verondersteld op een waarde van 8.15. Aan de ensembles voor diverse concentraties, temperatuur en pH worden ruissignalen⁸ toegevoegd, dus ook aan de bron (P). Alle andere, afgeleide signalen (EGV, SI en TACC) worden niet verstoord. De ensemble set wordt gevuld met $q=16$ leden. Verder wordt aangenomen dat aan de bron (P) *alle* waterkwaliteitsparameters bekend zijn (deze kunnen in de praktijk bijvoorbeeld door een procesinformatiesysteem worden doorgegeven), en dat deze parameters een bepaalde meetonzekerheid hebben. De meetonzekerheid in dit schema (schema 1) is even groot gekozen als de variatie van het ruissignalen op het ensemble.

Een overzicht van de simulatie-instellingen is weergegeven in Tabel 2-1.

⁸ Het ruissignaal heeft de volgende eigenschappen: het heeft een normaalverdeling N met gemiddelde rondom nul en een vooraf gespecificeerde standaarddeviatie (σ).

TABEL 2-1. INSTELLINGEN VOOR DE SIMULATIE VAN DE SOFTSENSOR, SCHEMA 1: DE SIGNALEN BIJ DE BRON, PARAMETERWAARDEN VOOR DE RUISSIGNALEN EN DE GEMETEN SIGNALEN OP EEN KNOOPPUNT.

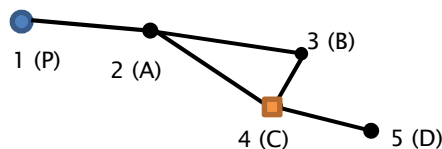
| Waterkwaliteitsparameter | Ruissignaal | Signaal bij de bron (P) (on-line) | Gemeten signaal (B) op tijdstippen t_k (u.) |
|---|---------------------|-----------------------------------|---|
| T | $N(0,\sigma=0.5)$ | $20 + 5\sin(\pi t_k/12 - \pi/2)$ | $t_k=[0,6,12,18]$ |
| pH | $N(0,\sigma=0.001)$ | 8.15 | n.v.t. |
| Ca^{+2} | $N(0,\sigma=0.2)$ | $40 + 4\sin(\pi t_k/12 - \pi/2)$ | $t_k=[0,6,12,18]$ |
| HCO_3^- | $N(0,\sigma=0.1)$ | $180 + 3\cos(\pi t_k/12 - \pi/2)$ | n.v.t. |
| Specifieke elektrische geleidbaarheid (EGV) | Geen | n.v.t. | n.v.t. |
| SI | Geen | n.v.t. | n.v.t. |
| TACC | Geen | n.v.t. | n.v.t. |

Numerieke notatie van knooppunten zal in het navolgende niet worden gebruikt, in plaats daarvan worden de letters P, A, B, C en D toegepast.

De resultaten met dit schema zijn opgenomen in de congresbijdrage [23] en in Hoofdstuk 4.

3.4 Instellingen simulatie – schema 2 (stress test)

Het doel dit schema is om de softsensor op de pijnbank te leggen: de ruis is opgehoogd met een factor 10 voor calcium en waterstofcarbonaat, een factor 100 voor de pH, en ook is een ruisterm toegevoegd voor de EGV. Voor de afgeleide (berekende) parameters SI en TACC is op punt P ook enige ruis toegevoegd. De pH wordt wederom constant verondersteld op een waarde van 8.15. De ensembles voor diverse concentraties, temperatuur en pH worden met een ruissignaal⁹ verstoord, maar nu gevuld met $q=50$ leden en hebben in tegenstelling tot schema 1 niet alle dezelfde onzekerheidsband als het signaal bij de bron. De softsensor berekent de waterkwaliteitsparameters voor een periode van 2,5 dag (60 uur). De ‘echte’ metingen komen nu binnen vanuit punt C, waarbij aangenomen is dat daar on-line pH-, EGV- en temperatuursensoren zijn. De softsensor is getest met hetzelfde voorbeeldnetwerk met 5 knooppunten, zie Figuur 3-3, maar nu met een ander meetpunt C.



Figuur 3-3: Voorbeeld netwerkmodel, met bronknooppunt $i=1$ (blauwe cirkel P) en het observatieknooppunt waar pH, EGV en temperatuur wordt gemeten $j=4$, oranje vierkant (C).

⁹ Het ruissignaal heeft de volgende eigenschappen: het heeft een gemiddelde rondom nul en een vooraf gespecificeerde standaarddeviatie (σ), waarbij de waarschijnlijkheidsverdeling normaal verdeeld is $N(0,\sigma)$.

Er wordt opnieuw ervan uitgegaan dat aan de bron alle waterkwaliteitsparameters bekend zijn. Een overzicht van de instellingen staat in Tabel 2-2.

TABEL 2-2. INSTELLINGEN VOOR DE SIMULATIE VAN DE SOFTSENSOR, SCHEMA 2: DE SIGNALLEN BIJ DE BRON, PARAMETERWAARDEN VOOR DE RUISSIGNALLEN EN DE GEMETEN SIGNALLEN OP EEN KNOOPPUNT.

| Waterkwaliteitsparameter | Ruissignaal op ensemble bij P | Ruissignaal op alle knooppunten of alleen P | Signaal bij de bron (P) (on-line) | Gemeten signaal (B) |
|--------------------------|-------------------------------|---|-----------------------------------|---------------------|
| <i>T</i> | | $N(0, \sigma=0.5)$ | $20 + 5\sin(\pi t_k/12 - \pi/2)$ | $t_k=[0,6,12,18]$ |
| <i>pH</i> | | $N(0, \sigma= 0.1)$ | 8.15 | n.v.t. |
| Ca^{+2} | $N(0, \sigma= 5.0)$ | $N(0, \sigma=1.0)$ | $40 + 4\sin(\pi t_k/12 - \pi/2)$ | $t_k=[0,6,12,18]$ |
| HCO_3^- | $N(0, \sigma= 5.0)$ | $N(0, \sigma=1.0)$ | $180 + 3\cos(\pi t_k/12 - \pi/2)$ | n.v.t. |
| <i>EGV</i> | | $N(0, \sigma=10)$ | n.v.t. | n.v.t. |
| <i>SI</i> | $N(0, \sigma= 0.5)$ | $N(0, \sigma=0.01)^*$ | n.v.t. | n.v.t. |
| <i>TACC</i> | $N(0, \sigma= 0.5)$ | $N(0, \sigma=0.01)^*$ | n.v.t. | n.v.t. |

Resultaten van simulaties met dit schema worden in Hoofdstuk 4 weergegeven en bediscussieerd.

4 Resultaten

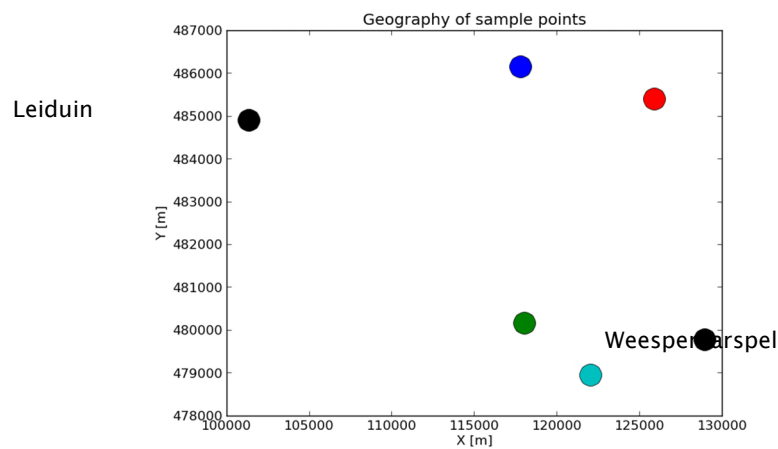
4.1 Data-analyse van waterkwaliteitsparameters

Waternet heeft een dataset beschikbaar gesteld met daarin meetgegevens afkomstig van verschillende meetlocaties in het leidingnetwerk van Amsterdam, die bepaald zijn door HWL en on-line sensoren. De gegevens betreffen de parameters: waterstofcarbonaat, sulfaat, zuurgraad, temperatuur, chloride, nitraat, calcium, natrium, kalium, magnesium als ook EGV-metingen en bepalingen van de TACC. De resultaten van de data-analyse worden in deze sectie gepresenteerd met als doel de variabiliteit van de TACC in een werkelijk leidingnet te tonen.

De meetlocaties zijn:

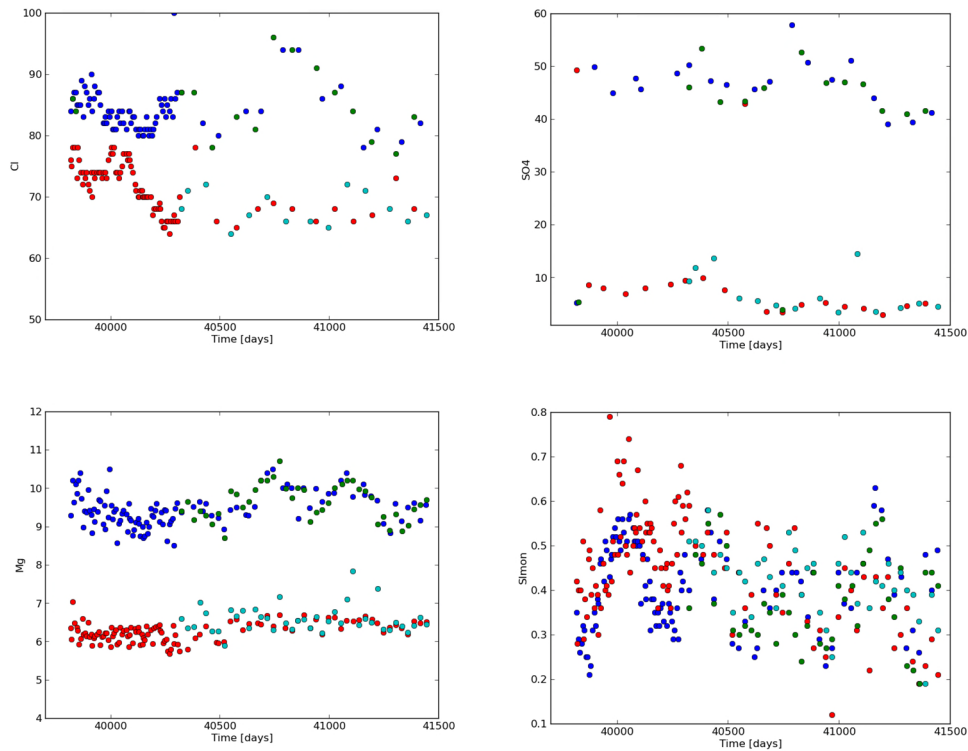
- uitgaand Wpb Leiduin;
- uitgaand Wpb Weesperkarspel;
- vier monsterpunten in het distributiegebied voorzien van de volgende coderingen van Waternet: DAM-DN-*, DAM-MO-*, DAV-DN-* en DOA-DN-*,

en worden als gekleurde punten weergegeven in Figuur 4-1.



Figuur 4-1: geografische locatie van diverse meetpunten, getoond in XY-coördinaten. De zwarte bolletjes geven Leiduin en Weesperkarspel aan. De kleurcoderingen blauw, rood, mintgroen en groen worden ook gebruikt in Figuur 4-2.

Met PHREEQC zijn speciatie-evenwichten berekend met de invoer van de diverse gemeten waterkwaliteitsparameters, en zijn de SI en TACC berekend. De tijdreeksen van chloride, magnesium, sulfaat en de SI worden weergegeven in Figuur 4-2. Ter illustratie is de berekende SI voor verschillende concentraties van waterkwaliteitsparameters getoond in Figuur I- in Bijlage I.



Figuur 4-2: van linksboven naar rechtsonder: tijdreeksen van chloride- (mg/l), sulfaat- (mg/l), en magnesiumconcentraties (mg/l) en de SI van calciumcarbonaat, waarbij de kleurcoderingen overeenkomen met de meetlocaties als in Figuur 4-1.

Uit de resultaten blijkt dat:

- met de chloride-, magnesium- en sulfaatconcentraties te achterhalen is waar het water vandaan komt: van productielocatie Leiduin of Weesperkarspel. De meetpunten waarbij het water (grotendeels) van Leiduin komt, hebben significant hogere concentraties aan magnesium, sulfaat en chlorides dan het water van Weesperkarspel. Het Weesperkarspelse water toont relatief gezien een minder variabel patroon in deze ionconcentraties dan het Leiduinse water. De verschillen in waterkwaliteit zijn minder goed zichtbaar in de berekende verzadigingsindex van calciumcarbonaat.

En wordt bevestigd dat:

- er een spreiding is in de SI (zie Figuur I-1 in Bijlage I) en TACC (hier niet getoond) als gevolg van variabiliteit in waterkwaliteit, waarbij hoge waarden van de SI bij 90° C boven de 1.0 uitkomen, en de TACC bij deze temperatuur in de orde grootte is van 5 mmol CC/l;
- de spreiding in SI en TACC is sterk afhankelijk is van calcium- en waterstofcarbonaatconcentraties en verandert met de temperatuur en pH van het water. Andere ionen, zoals nitraat en sulfaat, blijken de SI en TACC slechts zeer beperkt – hoogstens 3 tot 5 procent – te beïnvloeden.

4.2 Prestatie van de softsensor

Tijdens de simulaties van de softsensor is bijgehouden hoeveel rekentijd van de computerprocessor werd gependend aan de verschillende deelprocessen (sub-functies) van het Kalman Filter. Onderstaande tabel vat de timing samen:

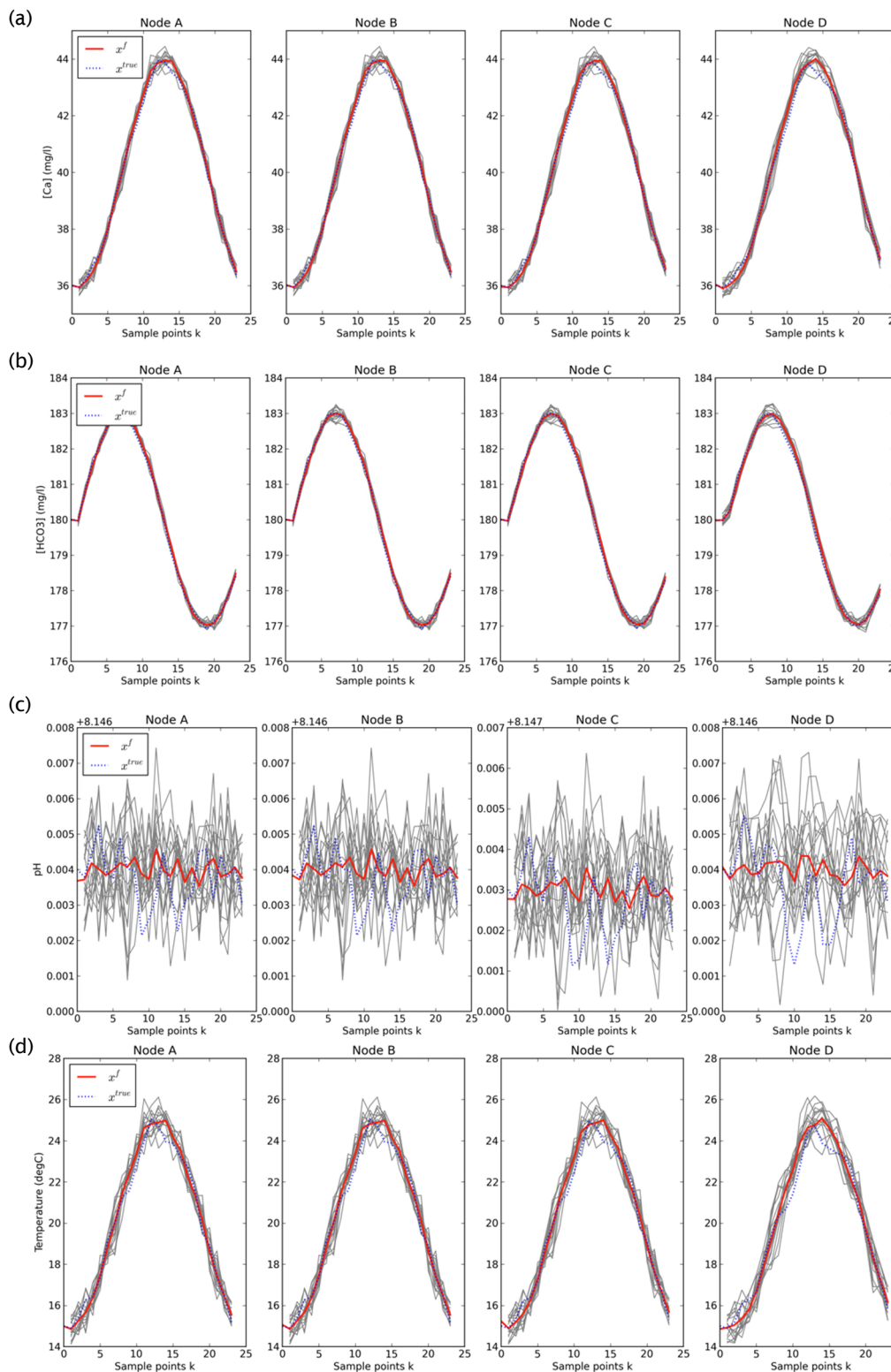
| Deelproces | CPU-tijd (ms) |
|--------------------|---------------|
| Aanroep PHREEQC | 9.0 |
| Aanroep EPANET-MSX | 0.3 |

Met 5 knooppunten is de totale rekentijd van iets meer dan 9 ms te overzien, maar zelfs dan wordt voor een set met 50 signalen met steeds een iets andere ruister toegevoegd (dit wordt ensemble genoemd) en een predictieraam van 12 tijdstappen vooruit¹⁰ de rekentijd 28 seconden. Met eenzelfde type PC en onder dezelfde randvoorwaarden (12 stappen vooruit, 50 ensemble-leden) zou het netwerk slechts maximaal 643 knooppunten kunnen bevatten om binnen het uur een uur vooruit te rekenen. Bij een groter netwerk zou deze softsensor met de gebruikte PC de on-line metingen niet meer kunnen bijbenen.

4.3 Simulatie met de softsensor – schema 1

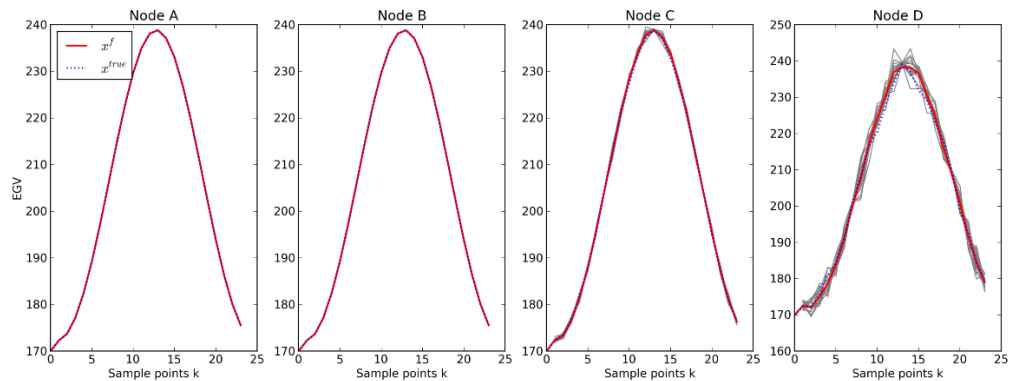
Met de instellingen als getoond in Paragraaf 3.3.3 is de softsensor getest voor het voorbeeldnetwerk als in Figuur 3-1. Resultaten van temperatuur, pH, calcium- en carbonaatconcentraties zijn getoond in Figuur 4-3. Merk op dat er in dit voorbeeld slechts een kleine verblijftijd (enkele minuten) is tussen knooppunt A en D.

¹⁰ Indien een (bemonsterings)frequentie van on-line sensoren (of van een procesbesturingsysteem) van 5 minuten wordt aangehouden, kan met 12 stappen een uur vooruit worden voorspeld.



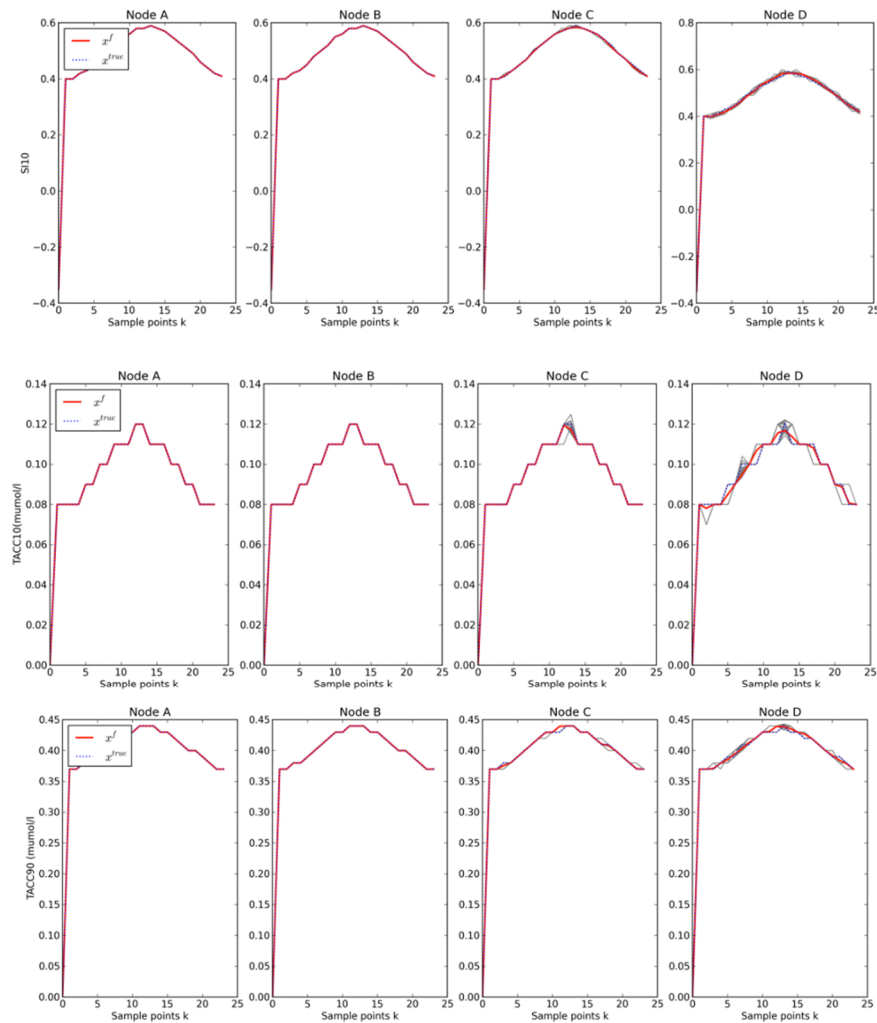
Figuur 4-3: Van bovenste (a) tot onderste paneel (d): Ca^{+2} (mg/l), HCO_3 (mg/l), pH (-) en temperatuur (°C) met 'werkelijke' waarden (blauwe onderbroken lijn), voorspelde waarden (rood) en waarden van het ensemble (grijs), berekend over 24 uur. Van links naar rechts: waarden voor de verschillende knooppunten A tot en met D.

De specifieke elektrische geleidbaarheid (EGV) is getoond in Figuur 4-4. Opvallend is dat de variatie in EGV als gevolg van de verstoringssignalen lijken te worden versterkt met de stroom mee.



Figuur 4-4: EGV op de knooppunten A tot en met D, met werkelijke waarden (blauwe onderbroken lijn), voorspelde waarden (rood) en waarden van het ensemble (grijs), berekend over 24 uur.

De SI en de TACC zijn weergegeven in Figuur 4-5. Hetzelfde verschijnsel als bij de EGV volgt uit de simulatieresultaten, verstoringen worden met name bij knooppunt C en D zichtbaar door de additie van de verstoring bij observatiepunt C. Wat verder opvalt is dat bij de TACC er een trapvormig signaal ontstaat. Dit trapvormige signaal is vermoedelijk het gevolg van het inlezen van afgeronde getallen, waardoor er discretisatiefouten zijn ontstaan. Deze artefacten kunnen bij doorontwikkeling van de softsensor eenvoudig worden opgelost door aandacht te besteden aan een correcte verwerking (met eventuele afronding) van de getallen.



Figuur 4-5: van bovenste (a) naar onderste paneel (d): SI bij 10°C (-), de TACC bij 10 en 90°C in mmol/l met werkelijke waarden (blauwe onderbroken lijn), voorspelde waarden (rood) en waarden van het ensemble (grijs), berekend over 24 uur.

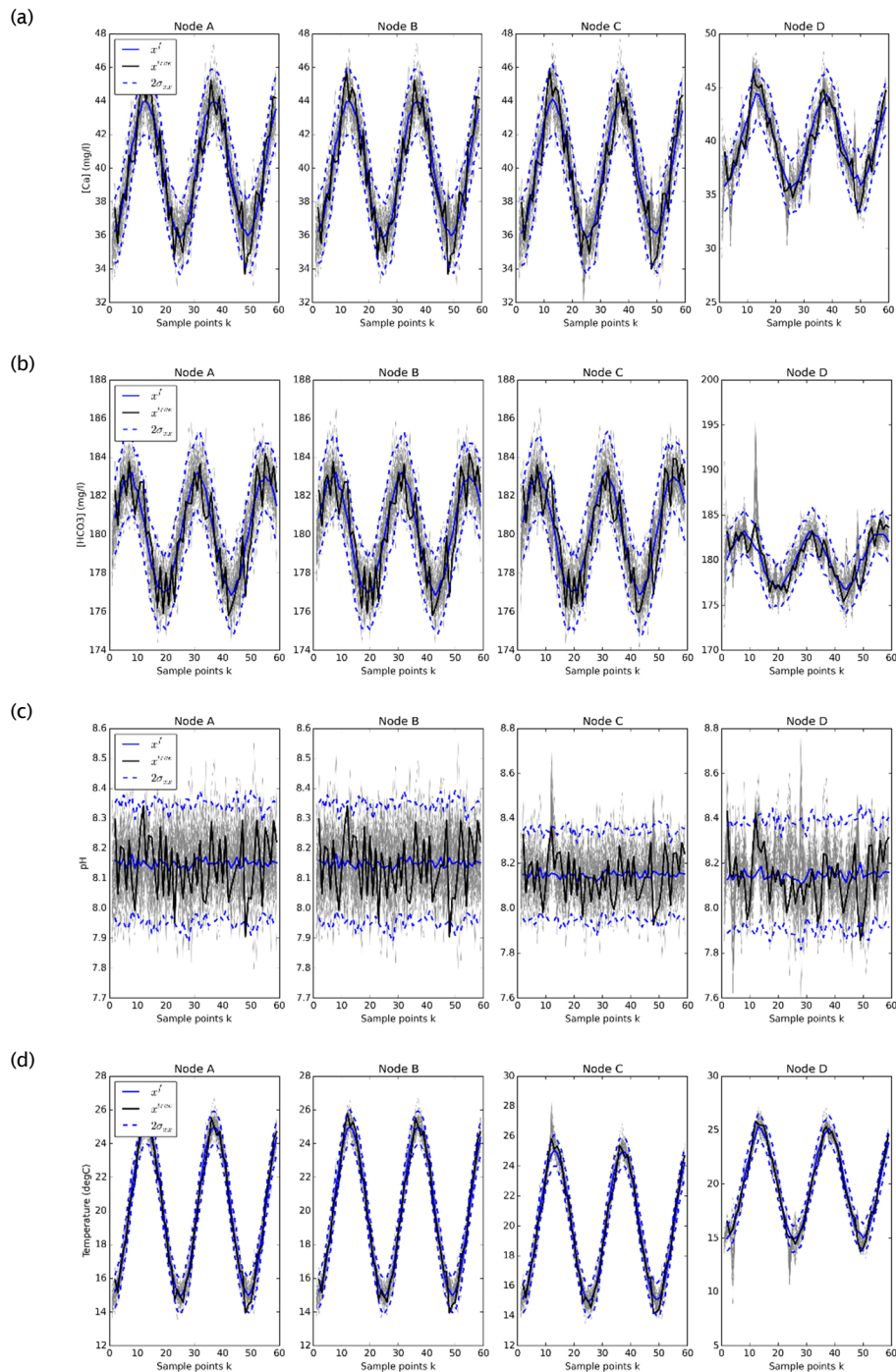
4.4 Simulatie met de softsensor – schema 2 (stress test)

Met de instellingen als getoond in Paragraaf 3.4 is de softsensor getest voor het voorbeeldnetwerk als in Figuur 3-3. Ten opzichte van schema 1, is er nu meer ruis toegevoegd aan het ensemble en het werkelijke signaal.

Resultaten van temperatuur, pH, calcium- en carbonaatconcentraties zijn getoond in Figuur 4-6. Naast het werkelijke signaal dat verstoord is met ruis, wordt ook de voorspelling en een 95%-betrouwbaarheidsinterval rondom de voorspelde waarde weergegeven (blauwe onderbroken lijnen).

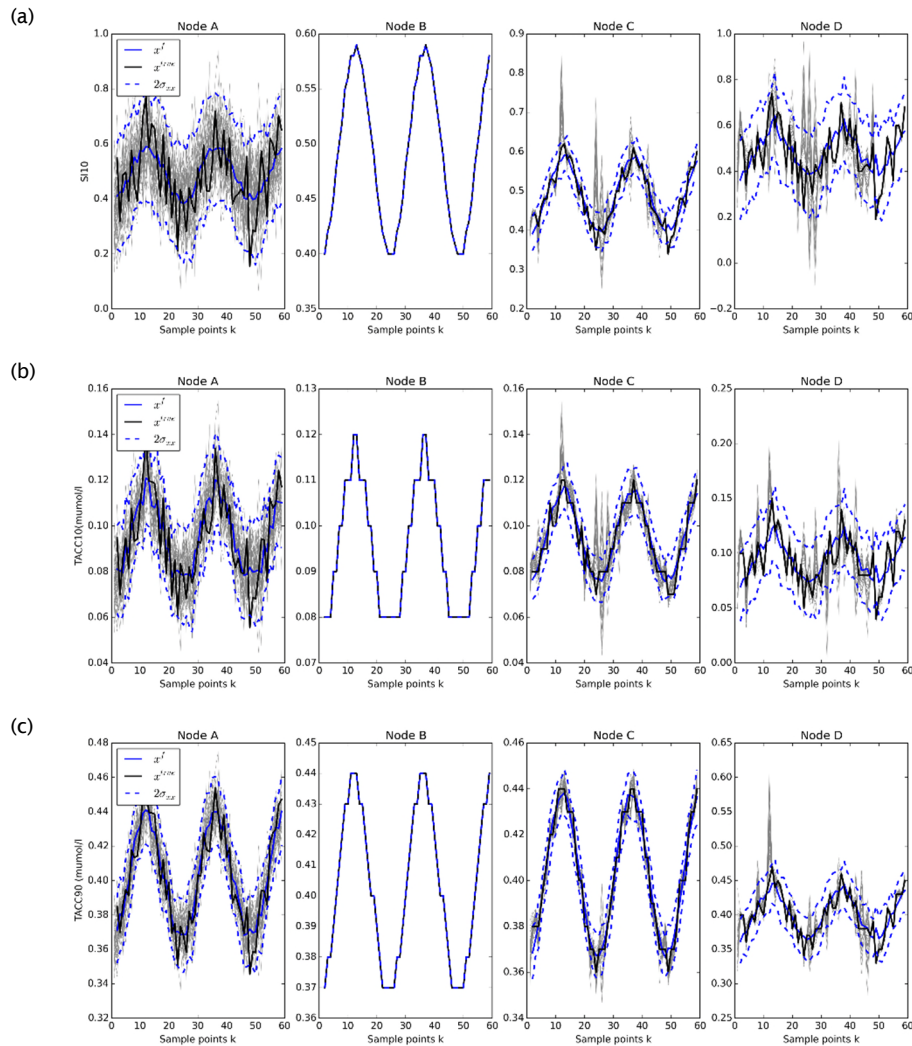
Het is opvallend dat op knooppunt C en D enkele modelvoorspellingen in het ensemble zeer grote uitschieters tonen, en buiten de 95%-betrouwbaarheidsintervallen vallen. Het is aannemelijk dat doordat op punt C twee (gesimuleerde) waterkwaliteiten samenkomen, er op dat moment twee waterkwaliteiten met ‘toevallig’ een ruisterm met hetzelfde teken (positief of negatief) op worden opgeteld, zie bijvoorbeeld het signaal op $t=12\text{u}$. We zien dat het werkelijke signaal in punt A ook een (kleinere) uitschieter heeft op dit moment. De softsensor werkt met bronsignalen waar ruis op is gegenereerd, dit zal ook effecten hebben

op de voorspellingen benedenstrooms. Omdat het ensemble uit 50 leden bestaat, heeft dit echter nauwelijks effect op de uiteindelijke voorspelde waarde van de waterkwaliteitsparameters als ook nauwelijks effect op de SI, EGV en TACC (Figuur 4-7). Het is met de enKF-techniek van belang om het ensemble met genoeg leden te vullen, anders zal een (additief) ruissignaal wel ongewenste effecten hebben.



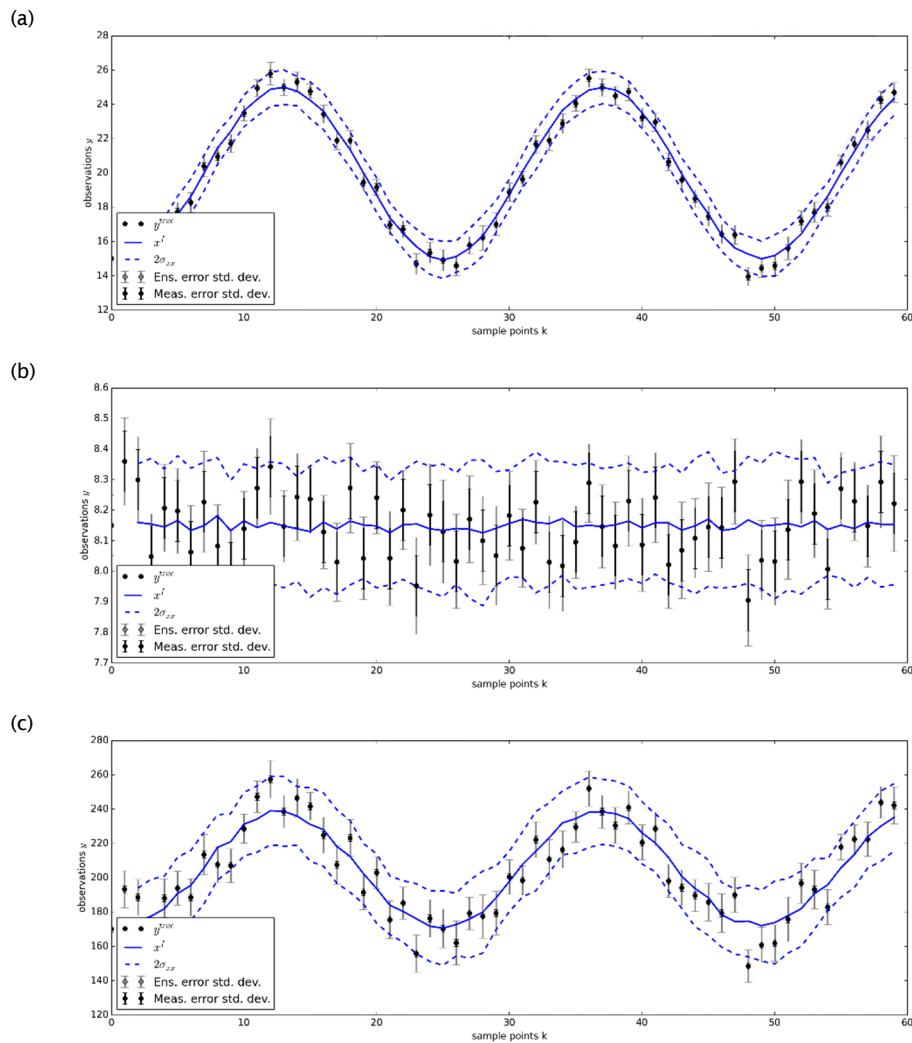
Figuur 4-6: Van bovenste (a) tot onderste paneel (d): Ca^{2+} (mg/l), HCO_3^- (mg/l), pH (-) en temperatuur ($^{\circ}C$) met 'werkelijke' waarden (zwarte lijn), voorspelde waarden (blauw) en waarden van het ensemble (grijs), berekend over 24 uur. Van links naar rechts: waarden voor de verschillende knooppunten A tot en met D.

Merk verder op dat de afgeleide parameters zijn berekend aan de hand van de voorspelling van de waterkwaliteitsparameters en dat alleen een ruisterm op het ensemble in punt P en op de geobserveerde signalen in punt C (T, pH en EGV) is toegevoegd. Dit verklaart waarom het signaal in punt B gladder is. Het dient nog te worden uitgezocht waarom het signaal in punt B zo glad is dat alle signalen van het ensemble hetzelfde zijn. In punt D zien we wel een effect van versterking van de uitschieters die in punt C zijn gegenereerd.



Figuur 4-7: van bovenste (a) naar onderste paneel (d): SI bij 10° C (-), de TACC bij 10 en 90° C in mmol/l met werkelijke waarden (zwarte lijn), voorspelde waarden (blauw) en waarden van het ensemble (grijs), berekend over 60 uur.

De gemeten signalen van temperatuur, pH en EGV in het (on-line) meetpunt C zijn weergegeven in Figuur 4-8. Ter vergelijking zijn error bars met de standaarddeviaties van de ensembles tezamen met de error bars met de meetvariëties rondom de metingen weergegeven. Het ensemble heeft een grotere standaarddeviatie, wat direct het gevolg is van de ingestelde ruistermen. In het algemeen is het aan te bevelen deze ruistermen groter te nemen dan de standaarddeviatie op een meting, zodat er voldoende (model)variëtie is om de werkelijkheid te kunnen beschrijven.



Figuur 4-8: van bovenste (a) naar onderste paneel (c): T ($^{\circ}\text{C}$), de pH en de EGV. Blauwe doorgetrokken lijn zijn de voorspelde waarden, error bars zijn weergegeven voor de meetpunten (zwarte punten), met standaarddeviatie van de meting (zwart, dikke error bars) en van het ensemble (zwart, dunne bars). Ook is de 95%-betrouwbaarheidsinterval weergegeven (blauwe onderbroken lijn).

De eerste (simulatieve) testen van de softsensor laten zien dat de ‘werkelijke’ waarde van de TACC, maar ook van andere waterkwaliteitsparameters, heel dicht overeenkomen met de voorspelling van het ensemble Kalman filter, zelfs bij een relatief klein ensemble (16 leden). Een kanttekening hierbij is dat softsensorbepalingen en ‘werkelijkheid’ met hetzelfde model gesimuleerd zijn, waarbij de werkelijkheid verstoord is met Gaussianse (witte) ruis. In de praktijk zal het potentieel om kalk af te zetten of op te nemen ook bepaald worden door leidingnetmaterialen en opwerveling of neerslag van deeltjes. Hoe groot de invloed is van deze effecten is op dit moment echter niet goed bekend.

5 Conclusies en aanbevelingen

In deze studie is een softsensor voor het bepalen van theoretisch afzetbaar kalk in leidingnetwerken ontwikkeld. Daarbij is gebruik gemaakt van beschikbare tools: PHREEQC, een programma om chemische evenwichten en reacties te berekenen, en EPANET, een programma om het transport van water in een leidingnet te simuleren.

De volgende concluderende opmerkingen en constatering worden gemaakt:

- Uit een data-analyse wordt bevestigd dat, overeenkomstig de theorie, kalkvorming met name afhankelijk is van calcium- en carbonaationen, pH en temperatuur, en veel minder (minder dan 5%) afhankelijk is van sulfaat-, chloride- en magnesiumionen.
- Voor het voorzieningsgebied van Waternet is het vermoeden bevestigd dat uit de chloride-, magnesium- en sulfaatconcentraties is te achterhalen waar het water in een bepaald punt in het net vandaan komt: uit productielocatie Leiduin of uit Weesperkaspel.
- PHREEQC slokt in de huidige implementatie, met berekeningen over alle knooppunten, veel (meer dan 95%) processortijd ten opzichte van de berekeningen in het transportmodel en het data-assimilatiefilter.
- Een eerste test met de softsensor voor een fictief leidingnetwerk laat zien dat de voorspelling van het ensemble Kalman filter, zelfs bij een relatief klein ensemble (16 leden), met de werkelijke waarde van de TACC, maar ook van andere waterkwaliteitsparameters dicht benadert.
- Het algoritme (een ensemble Kalman filter) blijkt robuust voor berekeningen benedenstrooms en opstapeling van meetfouten/-ruis, maar meer onderzoek(stijd) is nodig om grotere netwerken uit te testen en te toetsen of divergentie door opstapeling van ruis kan worden voorkomen of beperkt.

Uit deze studie volgen de volgende aanbevelingen:

- *Algemeen*
 - Uit een snelle scan is gebleken dat ontwikkeling van softsensoren toekomst heeft voor het real-time representeren van een representatief beeld bij parameters die op dit moment op een lage bemonsteringsfrequentie worden gemeten, maar wel behoefte naar is. Deze parameters variëren in het type informatie van indicatorparameters die een afdruk (fingerprint) van de waterkwaliteit bieden, tot softsensoren die een beeld geven van esthetische parameters (kleur, smaak, hardheid) tot softsensoren die van specifieke metingen een microbiologische karakterisatie (bijna) real-time kunnen bepalen.
 - Het verdient aanbeveling om kritische parameters, waar geen directe metingen van voorhanden zijn, te volgen door middel van softsensoren. Toepassingen kunnen enkelvoudig (bijvoorbeeld monitoring van een waterkwaliteitsparameter) of zelfs meervoudig zijn (bijvoorbeeld monitoring van waterkwaliteit, maar ook input voor lekdetectietools).
- *Implementatie van een softsensor voor TACC of andere waterkwaliteitsbepalingen in een leidingnet*
 - Voor leidingnetwerken met veel knooppunten geniet het gebruik van het ensemble Kalman Filter of verwante particle filters *in theorie* de voorkeur boven andere data-assimilatiemethodieken waarbij de inverse van een covariantie-

matrix wordt gebruikt. Een uitgebreidere toets met een implementatie van deze methode voor een werkelijk leidingnet is noodzakelijk. Een nadeel van het enKF is dat de berekening van de onzekerheidsband minder exact is.

- In de praktijk zullen de TACC en SI ook beïnvloed worden door leidingnetmaterialen en opwerveling of neerslag van deeltjes en temperatuurvariaties, terwijl deze effecten niet zijn meegenomen in het model. Deze niet-gemodelleerde effecten kunnen worden gecompenseerd indien de softsensor wordt gevoed door voldoende metingen. Temperaturen van leidingnetmaterialen kunnen goed gemodelleerd worden. Welke ruimtelijke resolutie en meetfrequentie benodigd zijn om voldoende nauwkeurigheid in de voorspellingen te verkrijgen, is een punt van verder onderzoek.
- Voor een goede werking van de enKF-softsensor dienen er genoeg modelsimulaties (bij voorkeur >100) binnen een ensemble te worden uitgevoerd met voldoende variatie op de leden van het ensemble.
- Een validatie in de praktijk is benodigd om de TACC die de softsensor bepaalt op bepaalde knoop- of tappunten ook daadwerkelijk te toetsen aan metingen van pH, temperatuur, calcium en carbonaatconcentraties op deze locaties. Uit de validatie moet blijken of een calibratie van het hydraulisch model (EPANET) noodzakelijk is. Door middel van tracertesten (zoals eerder gedaan in kleine distributienetten als Benthuizen en Zandvoort) kunnen deze kalibratie en validatie worden uitgevoerd.
- De koppeling van het leidingnetmodel (EPANET) is nu handmatig gekoppeld met het speciatiemodel (PHREEQC) gekoppeld, evenals allerlei instellingen van het enKF. Aanbevolen wordt om een automatische koppeling te realiseren, bijvoorbeeld door het implementeren van een (gebruikers)interface waarbij met een commando of een paar klikken aangegeven kan worden:
 - Op welke knooppunten de (online) invoersignalen zijn, inclusief meetfrequentie;
 - Op welke knooppunten waterkwaliteitswaarden van laboratoriumanalyses kunnen worden ingevoerd;
 - Wat de specifieke instellingen zijn van het enKF (o.a. hoe groot het ensemble moet worden gekozen; optimale waarden voor de ruisinstellingen);
 - Hoe ver vooruit in het leidingnetwerk moet worden gerekend;
 - Wat de simulatietijd moet zijn (een stap van 5 minuten vooruit of bijvoorbeeld een dag vooruit rekenen);
 - Op welke knooppunten de bepaling van de TACC in elk geval gewenst is.
- In plaats van PHREEQC, zou een snelle rekenimplementatie om het kalkevenwicht te berekenen, sterk de voorkeur genieten. Mogelijkheden zijn:
 - Gebruikmaken van de rekenmethode van het koolzuurevenwicht (als bijvoorbeeld in AquaCalc) en deze in C te compileren met EPANET-MSX;
 - Verbeteren van de communicatie tussen EPANET-MSX en PHREEQC. De huidige implementatie gaat via inlezen en wegschrijven van input en output naar de harde schijf, wat veel tijd kost. Een andere optie is om de communicatie op (dynamisch) geheugenniveau te laten plaatsvinden. Of deze implementatie voldoende snel is, is nog onzeker.
 - PHREEQC wél te gebruiken, maar deze gekoppeld met EPANET-MSX parallel te laten draaien in een cloud-omgeving. Hierbij dient te worden opgemerkt dat het ook voor een cloud-omgeving van belang is de communicatie op geheugenniveau te laten plaatsvinden; zie hierboven.

- Het Pitzer-model te gebruiken voor het bepalen van concentraties van ionen zodat nauwkeurig de TACC kan worden bepaald. Vervolgens dit model eveneens in C te compileren tezamen met EPANET-MSX, of;
- PHREEQC allerlei mogelijke combinaties doorrekenen die als mal dienen om een parametrisch zwarte-doos-model (bijvoorbeeld een neurale netwerk) te trainen.
- Een alternatief voor een snelle rekenimplementatie om het kalkevenwicht te berekenen, is om gebruik te maken van geautomatiseerde methodes om het leidingnet te verdunnen, om op die manier het aantal knooppunten sterk te verkleinen.

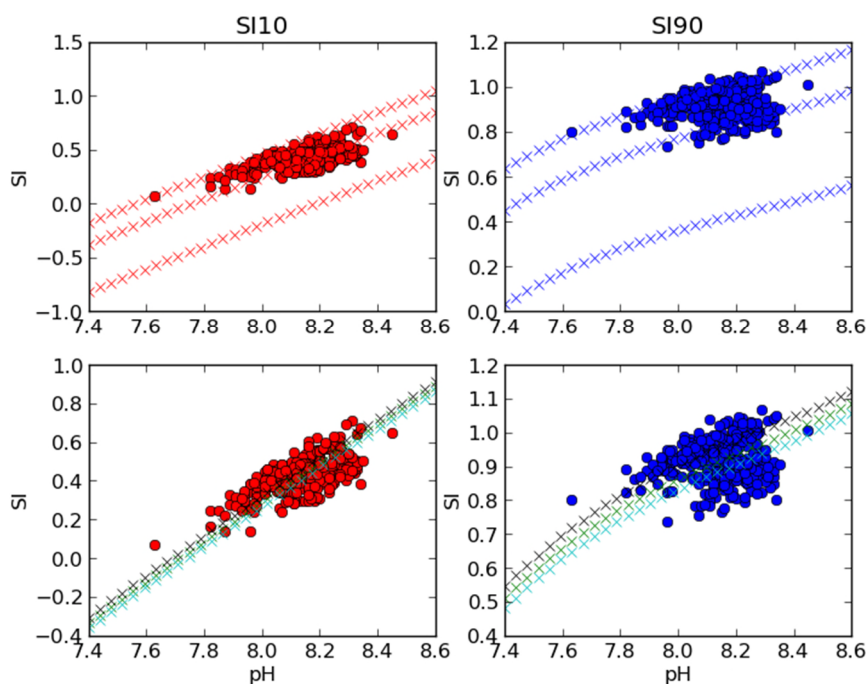
Een groot voordeel van het gebruik van het ensemble Kalman Filter als data-assimilatietechniek is dat de modelberekeningen parallel kunnen worden uitgevoerd. Aanbevolen wordt om bij softsensoren voor een werkelijk leidingnetwerk een parallelle rekenimplementatie uit te werken.

6 Literatuur

- [1] P. Smeets, L. Rietveld, J. van Dijk, and G. Medema, "Practical applications of quantitative microbial risk assessment (QMRA) for water safety plans," 2010.
- [2] J. H. Andersen and R. S. Powell, "Implicit state-estimation technique for water network monitoring," *Urban Water*, vol. 2, no. 2, pp. 123-130, 2000.
- [3] R. G. Hanea, G. J. Velders, and A. Heemink, "Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model," *Journal of Geophysical Research: Atmospheres (1984-2012)*, vol. 109, no. D10, 2004.
- [4] VEWIN, "Smart Water Networkⁱ is weer een stapje dichterbij," in *Waterspiegel*, vol. 15, no. 1, 2012.
- [5] P. Van Thienen, "A method for quantitative discrimination in flow pattern evolution of water distribution supply areas with interpretation in terms of demand and leakage," *Journal of Hydroinformatics*, vol. 15, no. 1, pp. 86-102, 2013.
- [6] D. Vries, B. v. d. Akker, S. Velickov, W. d. Jong, and J. v. Summeren, "Application of machine learning techniques to predict anomalies in water supply networks," in *2nd IWA New Developments in IT & Water Conference*, 2015.
- [7] K. Van Schagen, R. Babuška, L. C. Rietveld, and A. M. J. Veersma, "Model-Based Dosing Control of a Pellet Softening Reactor," in *Advanced Control of Chemical Processes*, 2009, vol. 7, no. 1, pp. 267-272.
- [8] B. de Graaf, F. Williamson, M. K. Koerkamp, J. W. Verhoef, R. Wuestman, B. Bajema, E. Trietsch, and W. van Delft, "Implementation of an innovative sensor technology for effective online water quality monitoring in the distribution network," *Water Practice & Technology*, vol. 7, no. 4, 2012.
- [9] R. Wuestmann, "Personal communication." 2013.
- [10] Eureka, "Eureka project: Self Learning Integrated Model-Based Management." .
- [11] H. Cremer, F. Schuren, B. de Graaf, and B. v. d. Wal, "Eerste resultaten DNA-gebaseerde bio-monitoring kiezelalgen," *H2O*, vol. 45, no. 1, p. 4, 2012.
- [12] J. Van den Broeke, P. S. Ross, A. W. C. van der Helm, E. T. Baars, and L. C. Rietveld, "Use of on-line UV/Vis-spectrometry in the measurement of dissolved ozone and AOC concentrations in drinking water treatment," *Water Sci Technol*, vol. 57, no. 8, pp. 1169-1175, 2008.
- [13] K. van Schagen, R. Babuška, L. C. Rietveld, and A. M. J. Veersma, "Model-Based Dosing Control of a Pellet Softening Reactor," in *Advanced Control of Chemical Processes*, 2009, vol. 7, no. 1, pp. 267-272.
- [14] L. J. Leurs, L. . Schouten, M. N. Mons, R. Alexandra Goldbohm, and P. A. van den Brandt, "Relationship between tap water hardness, magnesium, and calcium concentration and mortality due to ischemic heart disease or stroke in the Netherlands," *Environmental health perspectives*, vol. 118, no. 3, p. 414, 2010.
- [15] P. G. G. Slaats and C. H. M. M.A. en Hofman-Caris Meerkerk, "Conditioning: de optimale samenstelling van drinkwater," KWR Watercycle Research Institute, KWR 2013.069, 2013.
- [16] C. H. M. Hofman-Caris, "Berekening van kalk-koolzuurevenwichten in drinkwater; Ontwikkelingen in de afgelopen twintig jaar," KWR Watercycle Research Institute, Nieuwegein, BTO 2013.206(s), 2013.
- [17] P. J. De Moel, A. W. C. Van der Helm, M. van Rijn, J. C. Van Dijk, and W. G. J. Van der Meer, "Assessment of calculation methods for calcium carbonate saturation in drinking water for DIN 38404-10 compliance," *Drinking Water Engineering and Science Discussions*, vol. 6, no. 2, pp. 167-198, 2013.
- [18] L. Keltjens and H. Brink, "Met nieuwe kookproef snel informatie over kalkafzetting - Alternatief voor PACC-meting," in *H2O*, vol. 38, no. 7, KNW, 2005, pp. 30-32.
- [19] S. Gillijns, O. B. Mendoza, J. Chandrasekar, B. L. R. De Moor, D. S. Bernstein, and A. Ridley, "What is the ensemble Kalman filter and how well does it work?," in *American Control Conference, 2006*, 2006, p. 6-pp.
- [20] J. L. Anderson, "An ensemble adjustment Kalman filter for data assimilation," *Monthly weather review*, vol. 129, no. 12, pp. 2884-2903, 2001.
- [21] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean dynamics*, vol. 53, no. 4, pp. 343-367, 2003.
- [22] F. Shang, J. Uber, and L. Rossman, "EPANET multi-species extension software," *US Environmental Protection Agency EPA/600/C-10*, vol. 2, 2008.

- [23] D. Vries, B. v. d. Akker, J. v. Summeren, A. W. C. v. d. Helm, G. I. M. Worm, and P. v. Thienen, "Soft Sensing The Potential Amount Of Calcium Carbonate Precipitate In Drinking Water Distribution Infrastructure And Warm Water Household Appliances," in *11th International Conference on Hydroinformatics*, 2015, vol. HIC 2014 Proceedings.

Bijlage I SI in het leidingnet van Amsterdam



Figuur I-1: SI berekend voor waterkwaliteitsparameters gemeten in het voorzieningsgebied Amsterdam (data Waternet), de panelen links SI bij 10°C en rechts bij 90°C. Rode stippen in paneel linksboven tonen SI-waarden gecorrigeerd naar een temperatuur van 10°C en pH-waarden op de horizontale as. De berekende SI (kruisjes) in de bovenste twee panelen is uitgerekend voor het ionproduct $[Ca^{+2}][HCO_3^-]$ met waarde 1.0, 3.0 en 5.0 mmol/l², gezien van boven naar beneden. Deze waarden dekken een groot gedeelte van het bereik van verschillende watertypen. De onderste panelen laten de invloed van gemeten Cl^- , CO_2 , Mg^{+2} , NO_3^- , en SO_4^{2-} concentraties zien op de berekende SI (kruisjes), met een factor 0x deze concentraties (cyaan), factor 1x (groen) en een factor 2x (zwart). In de onderste panelen zijn de Ca^{+2} en HCO_3^- concentraties gemiddeld over alle metingen.

Bijlage II Broncode ensemble Kalman Filter

```
# -*- coding: utf-8 -*-

"""

Created on Mon Mar 24 20:39:42 2014
Ensemble Kalman Filter algorithm from: S Gillijns et. al., "What Is the Ensemble Kalman
Filter and How Well Does it Work?", Proceedings of the 2006 American Control Conference,
Minneapolis, Minnesota, USA, June 14-16, 2006, pp 4448-4453.
This program uses the ensemble Kalman filter to estimate a system's state. The state is
 $x_{new}=f(x,u)+w$ , where  $u$  some input,  $w$  the Gaussian distributed process noise, and  $f$  is a
nonlinear function. The measurement is  $y_{new}=h(x)+v$  where  $h$  is a nonlinear function and  $v$ 
Gaussian distributed measurement noise.
@author: Dirk Vries, KWR, 2014/03/24
"""

import numpy as np
from numpy.matlib import randn, concatenate, ones, mean, matrix, diag, empty, reshape
from numpy.linalg import inv
from time import time
### FUNCTION CALLS
# Define function calls here. This can be a script to call EpaNET-MSX, dependent on
(measured) states and (measured) inputs. In the case of a soft sensor for the calcium
carbonate precipitation potential, the state vector would comprise of:
# Ca2+, HCO3-, pH, T, ... at time instant t_k
# which from we would like to create an ensemble set.
# and flow_patterns_in, flow_patterns_out at time instant t_k
#
# NB: if possible, run EPANET-MSX in parallel
#####
# physical model: Calculates evolution over one time step using EPANET-MSX.
def fl(x,xvec_src,nodenames,linknames,allphyspar):
    import os
    physpar=["T","H","CA","CO3","EGV"]
    # --- Construct msx file from fin_msx.
    construct_msx(nodenames,linknames,physpar,allphyspar,x,xvec_src)
    # --- Run epanet.
    tstart = time()
    os.system("nice -19 epanetmsx_phrq example.inp example.msx example.rpt > out")
    # --- Get new state vector from msx report file.
    x_upd=extract_rpt(nodenames,linknames,allphyspar)
    return x_upd
#####
# observation function
def h1(x,idx_obs):
    y = x[idx_obs,:]
    return y
#####
### DATA ASSIMILATION # # implementation of ensemble Kalman Filter
def ensemblefilter2
(x_true,x_initial,w,z,N_hor,idx_obs,rho,alpha,delta_t,k_meas,nodenames,linknames,allphyspar
):

    # Defines the array of observations corresponding to a generated ensemble.
    # Observation = (measurements x nodes j) x (ensemble size q)
    ## Dimensions and initial calculations
    q = np.size(x_initial,1) # number of members
    n1 = np.size(x_initial,0) # number of states
    m1 = np.size(h1(x_true,idx_obs),0) # length of observation vector
    m2 = np.size(k_meas,0)

    xvec = empty((n1,N_hor)); xvec[:,0]=x_true;
    yvec = empty((m1,m2))
    y_e = empty((m1*N_hor, q))
    X_f = empty((n1,q)) # forecast ensemble (states) matrix
    X_a_tot = empty((n1*(N_hor-1),q)) #
    Y_f = empty((m1,q)) # forecast ensemble (observation) matrix
    Y_tp = empty((m1,q)) # true observation ensemble set
    Y_fp = empty((m1,q)) #
    x_f_vec = empty((n1,N_hor)); x_f_vec[:,0] = mean(x_initial,1) # forecast
    state_vector
    X_a = x_initial # initial ensemble
```

```

Pyy_mat      = empty((m1,N_hor))

#Zcov = diag(z)**2
#Delta = z**2
Delta = 1e-16
#Delta = 0
k_obs = 0

# --- Construct epanet inp file.
construct_inp() #k,nennodes,"fin_inp.txt")
tstart = time()
for k in range(0,N_hor):
    tstart_step = time()
    print "ef: k = "+str(k)
    # --- Construct epanet inp file.
    #construct_inp() #k,nennodes,"fin_inp.txt")

    # --- Generate time-dependent quantities at source node.
    xvec_src=gen_source(k)

    ### Noise generation
    W = w*randn(1,q)          # process noise matrix
    Z = z*randn(1,q)          # measurement noise matrix

    ### Physics/time Update
    #x_t_upd = f1(x_true,xvec_src,nodenames,linknames,allphyspar) + w*randn(n1,1)
# x_{k+1}   JvS
    x_t_upd = f1(x_true,xvec_src,nodenames,linknames,allphyspar) + w*randn(1,1)
# x_{k+1}
    y_t_upd = h1(x_true,idx_obs)          # y_k = h(x_k)

    #sys.exit()

    for j in range(0,q):
        #print " ef: j = "+str(j)
        ### Data assimilation step (generation of ensembles)
        Y_f[:,j] = h1(X_a[:,j],idx_obs)    # in fact, h(x_a) = h(x_f_{k})
        if k==k_meas[k_obs]:
            # measurement update
            Y_tp[:,j] = y_t_upd + Z[:,j]   # true (perturbed) observation
ensemble
            Y_fp[:,j] = Y_f[:,j] + Z[:,j]   # perturbed forecast observations

            ### Forecast steps
            X_f[:,j] = f1(X_a[:,j],xvec_src,nodenames,linknames,allphyspar) + W[:,j]
# forecast step (3.8) (calculate ensemble at once)
            #print "ef: Y_f[:,j] ", Y_f[:,j]

            X_f_bar = mean(X_f,1)
            Y_fp_bar = mean(Y_fp,1)

            E_x = (X_f - X_f_bar)          # ensemble error matrix (3.2)
            E_y = (Y_fp - Y_fp_bar)        # matrix of innovation errors (y-y_hat)
(3.2)
            Pxy = E_x*E_y.transpose()/(q-1) # ensemble cov matrix Pxy (3.4), here
Delta is added in case num_members=2 or Pyy is nearly singular
            Pyy = E_y*E_y.transpose()/(q-1)+Delta # forecast error cov Pyy (3.4), here
Delta is added in case num_members=2 or Pyy is nearly singular

            ### Data analysis step
            ###
            if k==k_meas[k_obs]:
                # measurement update
                K = Pxy*inv(Pyy)          # determine filter gain (3.7)
                X_a = X_f + K*(Y_tp - Y_f) # analysis equation (3.5)
            else:
                X_a = X_f

            ### Store in arrays
            ###
            if k != N_hor-1:
                xvec[:,k+1] = x_t_upd
                x_f_vec[:,k+1] = X_f_bar
                for j in range(n1):
                    X_a_tot[k+j*(N_hor-1),:] = X_a[j,:]
            for j in range(m1):
                y_e[k+j*N_hor,:] = Y_fp[j,:]
            x_true = x_t_upd
            Pyy_mat[:,k] = diag(Pyy)[:,None] #BvdA
            if k==k_meas[k_obs]:
                yvec[:,k_obs] = y_t_upd
                k_obs = min(k_obs+1,m2-1)
            print 'Time (step): ', time() - tstart_step
            print 'Time (total): ', time() - tstart
            return x_true,X_f_bar,Y_fp_bar,xvec,x_f_vec,yvec,y_e,Pyy_mat,X_a_tot;

```



```
#####
def construct_inp(): #itime,nennodes,file_inp):
    # --- Read relevant data for constructing input file (not yet used).
    #finp = open(file_inp,"rw")
    #for iennodes in range(nennodes):
    #    strwhole=finp.readline()
    #    strprt=strwhole.split(" ")
    #finp.close()
    # --- Write epanet input file.
    fout = open("example.inp","w")
    fout.write("[TITLE]\n");
    fout.write("EPANET-MSX Example Network\n");
    fout.write("[JUNCTIONS]\n");
    fout.write(";ID      Elev   Demand  Pattern\n");
    fout.write(" A      0      41\n");
    fout.write(" B      0      34\n");
    fout.write(" C      0      55\n");
    fout.write(" D      0      75\n");
    fout.write("[RESERVOIRS]\n");
    fout.write(";ID      Head  Pattern\n");
    fout.write(" Source 100\n");
    fout.write("[PIPES]\n");
    fout.write(";ID      Node1  Node2  Length  Diameter  Roughness\n");
    fout.write(" 1      Source A      1000   200     100\n");
    fout.write(" 2      A      B      800    150     100\n");
    fout.write(" 3      A      C      1200   200     100\n");
    fout.write(" 4      B      C      1000   150     100\n");
    fout.write(" 5      C      D      2000   150     100\n");
    fout.write("[TIMES]\n");
    fout.write(" Duration          1:00\n");
    fout.write(" Hydraulic Timestep 1:00\n");
    fout.write(" Quality Timestep  0:05\n");
    fout.write(" Pattern Timestep  1:00\n");
    fout.write(" Pattern Start     0:00\n");
    fout.write(" Report Timestep   1\n");
    fout.write(" Report Start      0\n");
    fout.write(" Statistic         NONE\n");
    fout.write("[OPTIONS]\n");
    fout.write(" Units             CMH\n");
    fout.write(" Headloss          H-W\n");
    fout.write(" Quality           NONE\n");
    fout.write("[END]\n");
    fout.close()
#####
def construct_msx(nodenames,linknames,physpar,allphyspar,vector,xvec_src):
    nnode=len(nodenames)
    nlinks=len(linknames)
    nen=nnode+nlinks
    nphyspar=len(physpar)
    allnphyspar=len(allphyspar)

    pH_global = 8.15
    H_global = pow(10,-pH_global)

    fout = open("example.msx","w")
    fout.write("[TITLE]\n Inert Ca and CO3 transport. \n\n");
    fout.write("[OPTIONS]\n AREA_UNITS M2\n RATE_UNITS HR\n SOLVER RK5\n TIMESTEP 300\n\n");
    fout.write("[SPECIES]\n BULK T MG\n BULK H MG\n BULK CA MG\n BULK CO3 MG\n BULK CL MG\n
    BULK MG MG\n BULK CO2 MG\n BULK NO3 MG\n BULK SO4 MG\n BULK SI10 MG\n BULK SI90 MG\n BULK
    TACC10 MG\n BULK TACC90 MG\n BULK EGV MG\n\n");
    fout.write("[PIPES]\n RATE T 0\n RATE H 0\n RATE CA 0\n RATE CO3 0\n RATE CL 0\n RATE MG
    0\n RATE CO2 0\n RATE NO3 0\n RATE SO4 0\n RATE SI10 0\n RATE SI90 0\n RATE TACC10 0\n RATE
    TACC90 0\n RATE EGV 0\n\n");

    # --- Initial conditions.
    fout.write("[QUALITY];Init.cond. (mg/l).\n ");

    fout.write("GLOBAL "+str(allphyspar[1])+" "+str(pH_global)+"\n")
    # Exception for obscure fail on H in quality-block.
    skip = [1]
    #for m in range(nphyspar-1):
    for m in range(4): # loop over physical params.
        #if m in skip:
        #    continue
        # --- Source IC.
        if m==1:
            fout.write("NODE Source "+str(allphyspar[m])+" "+str(pow(10.,float(-
            xvec_src[m])))+"\n")
        else:
            fout.write("NODE Source "+str(allphyspar[m])+" "+str(xvec_src[m])+"\n")
    # --- Nodes IC.
    for l in range(nnode):
        p = m*nen + l
        if (m==1):
```

```

        fout.write("NODE "+str(nodenames[l])+" "+str(allphyspar[m])+" "+str(pow(10.,-
float(vector[p])))+"\n") # pH -> [H+].
    else:
        fout.write("NODE "+str(nodenames[l])+" "+str(allphyspar[m])+"
"+str(float(vector[p])))+"\n")
    # --- Links IC.
    for n in range(nlinks):
        p = m*nen + (n+nnode)
        if (m==1):
            fout.write("LINK "+str(linknames[n])+" "+str(allphyspar[m])+" "+str(pow(10.,-
float(vector[p])))+"\n") # pH -> [H+].
        else:
            fout.write("LINK "+str(linknames[n])+" "+str(allphyspar[m])+"
"+str(float(vector[p])))+"\n")

    # --- Set t-dep. patterns on source and observation nodes.
    # All other nodes are evaluation nodes.
    fout.write("\n[SOURCES];Patterns.\n")
    #for m in range(nphyspar-1): # Minus 1, because EGV is a free parameter.
    for m in range(4):
        if (m==1):
            fout.write(" SETPOINT Source "+str(allphyspar[m])+" "+str(pow(10.,float(-
xvec_src[m])))+" PAT_DFLT \n")
        else:
            fout.write(" SETPOINT Source "+str(allphyspar[m])+" "+str(xvec_src[m])+" PAT_DFLT
\n")

    # --- Define patterns.
    fout.write("\n[PATTERNS]\n PAT_DFLT 1\n\n")

    # --- Report.
    fout.write("\n[REPORT]\n")
    fout.write(" NODES")
    for l in range(nnode):
        fout.write(" "+str(nodenames[l]))
    fout.write(" \nLINKS")
    for n in range(nlinks):
        fout.write(" "+str(linknames[n]))
    for m in range(allnphyspar):
        if (m==1):
            fout.write("\n SPECIE "+str(allphyspar[m])+" " YES 16")
        else:
            fout.write("\n SPECIE "+str(allphyspar[m])+" " YES")
    fout.close()

#####
def extract_rpt(nodenames,linknames,allphyspar):
    import sys
    import math
    import numpy as np

    skiplines=6
    # --- Check for starting lines of NODES.
    iline_node=[0,0,0,0]
    fin=open("example.rpt","rb")
    iline = 0
    for line in fin:
        if "Node" in line:
            for l in range(len(nodenames)):
                if nodenames[l] in line:
                    iline_node[l]=iline
            iline+=1
    fin.close()
    # --- Check for starting lines of LINKS.
    iline_link=[0,0,0,0,0]
    fin=open("example.rpt","rb")
    iline = 0
    for line in fin:
        if "Link" in line:
            for n in range(len(linknames)):
                if linknames[n] in line:
                    iline_link[n]=iline
            iline+=1
    fin.close()

    # --- Get state vector data from msx report file.
    ivector=[0. for x in range(len(allphyspar)*(len(nodenames)+len(linknames)))]
    vector = np.matlib.matrix(ivector).transpose()
    fin=open("example.rpt","rb")
    filelist = fin.readlines()
    for m in range(len(allphyspar)):
        for l in range(len(nodenames)):
            p=m*(len(nodenames)+len(linknames)) + l
            if m==1:

```

```

        #vector[p]=min(9.,max(7.,-
math.log10(float(filelist[iline_node[l]+skiplines].split()[m+1]))) # [H+] -> pH.
        vector[p]=-math.log10(float(filelist[iline_node[l]+skiplines].split()[m+1])) #
[H+] -> pH.
        else:
            vector[p]=float(filelist[iline_node[l]+skiplines].split()[m+1])
            #print m, p,vector[p]
        for n in range(len(linknames)):
            p=m*(len(linknames)+len(nodenames)) + n + len(nodenames)
            if m==1:
                #vector[p]=min(9.,max(7.,-
math.log10(float(filelist[iline_link[n]+skiplines].split()[m+1]))) # [H+] -> pH.
                vector[p]=-math.log10(float(filelist[iline_link[n]+skiplines].split()[m+1])) #
[H+] -> pH.
                else:
                    vector[p]=float(filelist[iline_link[n]+skiplines].split()[m+1])
                    #print m, p,vector[p]
            #print "vector: \n",vector
            fin.close()

        return(vector)

#####
def gen_source(k):
    import math
    T_m=20.; T_f=5.
    T=T_m+T_f*math.sin((float(k)*math.pi/12.)-(math.pi/2.))
    pH=8.15
    Ca_m=40.; Ca_f=4.
    Ca=Ca_m+Ca_f*math.sin((float(k)*math.pi/12.)-(math.pi/2.))
    C_m=180.; C_f=3.
    C=C_m+C_f*math.cos((float(k)*math.pi/12.)-(math.pi/2.))
    EGV=170.
    xvec_src=[T,pH,Ca,C,EGV]

    return(xvec_src)

```