



Optimalisatie meetprogramma E. coli in distributienet

BTO 2009.008
Maart 2009



Watercycle Research Institute

Optimalisatie meetprogramma E. coli in distributienet

BTO 2009.008

Maart 2009

© 2009 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.

Colofon

Titel

Optimalisatie meetprogramma E. coli in distributienet

Projectnummer

B111595.100

Projectmanager

W.J.M.K. Senden

Opdrachtgever

CvO

Kwaliteitsborger

G.J. Medema

Auteurs

E.J.M. Blokker, A.J. Vogelaar

Verzonden aan

Dit rapport is verspreid onder BTO-participanten en is openbaar

Samenvatting

In 2004 zijn berekeningen uitgevoerd om vast te stellen wat de detectiekans is van *E. coli* in het distributienet van Almelo met het monsternamprogramma van 2002. De conclusie was toen dat de detectiekans erg laag was (rond de 5%). Een aantal voorstellen werd gedaan om monsternamprogramma te optimaliseren zoals de meetfrequentie op pompstation en in reservoirs te verhogen, de monsternam te combineren met klachten en storingen en een betere verdeling van de monsterlocaties over het leidingnet.

Doel van voorliggend onderzoek was het vergroten van de detectiekans van *E. coli* in het distributienet. De eerste vraag die daarbij van belang is, is wat het doel van het (verbeterde) meetprogramma is. Bijvoorbeeld “Voldoen aan de VROM-inspectienorm”; “Detecteren van verontreinigingen (meten van prestatie)” of “Detecteren van verontreinigingen en erger voorkomen”. Met het huidige meetprogramma voldoen de waterbedrijven reeds aan de (huidige) VROM-inspectienorm. Het detecteren en passende actie ondernemen lijkt op dit moment nog een stap te ver. Het doel van het huidige onderzoek is derhalve om met een verbeterd meetprogramma de prestatie beter vast te kunnen stellen.

Uit de analyse van de detectiekans met een standaard monsterprogramma in Almelo blijkt dat de kans dat een verontreiniging wordt gedetecteerd gemiddeld 5,0% is en op zijn best 7,7 %. Het monsterprogramma waarbij 7,7% detectiekans wordt gerealiseerd is echter niet van te voren vast te stellen. Door een monsterprogramma te kiezen op basis van locaties met relatief lange verblijftijden kan de gemiddelde detectiekans worden verhoogd van 5,0% tot 6,3%. Door de monsterfrequentie op te voeren kan ook de detectiekans worden verhoogd. Een driemaal zo hoge monsterfrequentie leidt bijvoorbeeld tot een toename van de detectiekans met meer dan 100% (van 5,0% gemiddeld naar 10,9% gemiddeld).

Een andere mogelijkheid om de detectiekans te verhogen is het plaatsen van sensoren voor continue bemonstering. Door verschillende onderzoeksgroepen wordt gerekend aan een optimale plaatsing van sensoren. Daarbij buigen zij zich over verschillende aspecten van de modellering zoals o.a. “wat zijn de beste optimalisatieparameters (aantal sensoren / kostprijs; tijd tot detectie; het aantal consumenten dat verontreinigd water drinkt; het volume geconsumeerd verontreinigd water; etc.)”; “hoe rekening te houden met reactietijd (van detectie tot kookadvies of het spuien van het net)”; “wat is de invloed van het niet perfect kennen van de volumestromen in het netwerk (dus de afnamepatronen)”.

In voorliggend onderzoek zijn voor het bepalen van de beste sensorlocaties twee optimalisatietechnieken op basis van genetische algoritmen nader onderzocht, namelijk NSGA-II (Non dominated Sorting Genetic Algorithm, extension) en mMOSA (modified Multi-Objective Simulated Annealing). Door te kiezen voor een (semi)continue monitoring met behulp van sensoren kan de detectiekans aanzienlijk worden verhoogd ten opzichte van monsternam. Door sensorlocaties te bepalen met NSGA-II of mMOSA kan reeds met 5 sensoren een detectiekans van 40% tot 60% worden behaald.

Met de huidige meetinspanning blijft de detectiekans van een verontreiniging van het distributienet met rioolwater gering, ook als het meetprogramma geoptimaliseerd is en zelfs als er vaker wordt gemeten dan de wet vraagt. Dat betekent dat waterbedrijven ook alert moeten zijn op andere signalen (klachten van consumenten) die zouden kunnen aangeven dat er een besmetting is opgetreden. Als *E. coli* wordt aangetroffen met het meetprogramma is dat een signaal dat er besmetting is opgetreden waar snel op gereageerd moet worden.

Met sensoren komt een duidelijke verhoging van de detectiekans en tijdig kunnen reageren (“het detecteren en passende actie ondernemen”) in zicht. Dit is vooralsnog theorie, want echt gevoelige sensoren voor *E. coli* zijn nog niet op de markt.

Inhoud

Samenvatting	1
Inhoud	3
1 Inleiding	5
2 Bepaling van de optimale meetlocaties	7
2.1 Literatuuronderzoek	7
2.1.1 Locaties voor steekproeven	7
2.1.2 Locaties voor continue metingen	8
2.2 Toegepaste methodes voor de bepaling van meetlocaties	12
2.2.1 Meetlocaties steekproeven op basis van verblijftijd	12
2.2.2 Meetlocaties met poortwachterfunctie	14
2.2.3 Meetlocaties (semi)continue metingen op basis van mMOSA	15
2.2.4 Meetlocaties (semi)continue metingen op basis van NSGA-II	15
3 Bepaling van de detectiekans	17
3.1 Beschrijving simulatie	17
3.1.1 Voorzieningsgebied Almelo	17
3.1.2 Handmatige ALEID-simulaties van verontreinigingen	17
3.1.3 ALEID-dll-simulaties van verontreinigingen	17
3.1.4 Beschrijving meetprogramma's	18
3.1.5 Berekening van de detectiekans	19
3.2 Detectiekans bij verschillende scenario's	20
3.2.1 Monstername in het leidingnet – willekeurig over 24 uur	21
3.2.2 Monstername in het leidingnet – willekeurig over de werkdag	22
3.2.3 Monstername in het leidingnet – opvoeren van de monsterfrequentie	23
3.2.4 Continue sensoren in het leidingnet	23
3.2.5 Semi-continue sensoren in het leidingnet	24
4 Discussie en conclusies	27
4.1 Detectiekans met monsterprogramma	27
4.2 Detectiekans met verzwaard monsterprogramma	27
4.3 Detectiekans met sensoren	27
5 Referenties	29

BIJLAGE 1 Paper Exeter

BIJLAGE 2 Vergelijk ALEID-simulatie handmatig en automatisch (dll)

1 Inleiding

Een belangrijk kenmerk van de Nederlandse drinkwatervoorziening is de distributie van drinkwater zonder of met zeer lage concentraties desinfectiemiddel. De integriteit van de infrastructuur, de waterdruk en hygiëneprotocollen bij werkzaamheden zijn daarom de enige barrière tegen verontreiniging. De waterkwaliteit wordt bewaakt met het meetnet voor *E. coli*. Verontreinigingen van het drinkwater in het leidingnet met ziekteverwekkende micro-organismen moeten zo vroeg mogelijk worden gedetecteerd om snel maatregelen te kunnen nemen.

Een oriënterend onderzoek in 2004 (Lieverloo et al., 2004) heeft aangegeven dat met het huidige meetprogramma de detectiekans van een substantiële verontreiniging met rioolwater in een distributienet gering (0-12%) is. Ook zijn uit het onderzoek aanwijzingen gekomen dat de detectiekans verhoogd zou kunnen worden door “slimmer” te meten.

Het doel van dit onderzoek is vast te stellen of en hoe de kans op detectie van een fecale verontreiniging van het drinkwater in het distributienet kan worden verhoogd. Daarbij wordt in eerste instantie gekeken naar gericht meten (slimme keuze van meetlocatie en tijd op basis van hydraulica van het net) en pas in tweede instantie naar anders meten (hogere meetfrequentie tot aan bijvoorbeeld continu meten met sensoren in het leidingnet). In het geval van het toepassen van sensoren worden internationaal verschillende optimalisatietechnieken toegepast. Een afstudeerder aan de universiteit van Exeter heeft een tweetal technieken toegepast op het netwerk van Almelo, dat in deze studie als modelnetwerk is gebruikt.

In hoofdstuk 2 worden methodes beschreven om de optimale locaties voor monstername en sensoren te bepalen; een uitgebreide beschrijving van de methodes van de universiteit van Exeter is opgenomen in bijlage I. In hoofdstuk 3 worden verschillende monsterprogramma's met elkaar vergeleken op basis van de detectiekans van gesimuleerde verontreinigingen in het distributienet van Almelo. In hoofdstuk 4 volgt een discussie en aanbevelingen.

2 Bepaling van de optimale meetlocaties

De optimale meetlocatie hangt samen met het doel van de meting en de meetmethode. Wanneer monsternamen plaatsvindt volgens het huidige Waterleidingbesluit, d.w.z. kleine volumes op roulerende locaties waarmee steekproefsgewijs de waterkwaliteit wordt gecontroleerd, zijn de optimale meetlocaties die locaties waar de grootste detectiekans optreedt. Omdat besmettingen doorgaans kortdurend of intermitterend van aard zijn, zou de detectiekans van monsternamen kunnen worden vergroot door te bemonsteren op locaties met lange verblijftijden. Bij een minimalisatie van de tijd tot detectie (in theorie mogelijk door continu te meten met sensoren) worden de optimale meetlocaties bepaald door de leidingen waar het water van zoveel mogelijk verschillende locaties (mogelijke verontreinigingsbronnen) passeert.

Haas (1993) heeft reeds vastgesteld dat het voor de microbiologische waterkwaliteitsbewaking beter is veel (vaak) kleine samples te nemen dan af en toe grote samples vanwege het variabele karakter van deze verontreinigingen. In voorliggend onderzoek wordt alleen naar de optimale meetlocaties gekeken; er wordt geen uitspraak gedaan over de optimale hoeveelheid monsters en de grootte van de monsters. Janke e.a. (2005) hebben op basis van kosten en aantal te verwachten infecties een afweging gemaakt tussen drie type monsternamen, namelijk dagelijkse, twee-dagelijkse en continue (een per uur) monsternamen. Het verschil in het gemiddelde aantal geïnfecteerde personen tussen een dagelijkse en twee-dagelijkse bemonstering is slechts 7% (in het voordeel van de dagelijkse bemonstering). Een bemonstering ieder uur met een verwerkingstijd van 4 uur levert een extra 60% beperking in het aantal geïnfecteerde personen op. Hoe dit afgewogen kan worden tegen kosten is onduidelijk, maar duidelijk is dat het opvoeren van de monsterfrequentie de detectiekans verhoogt.

In §2.1 worden op basis van een literatuuronderzoek methodes besproken voor de bepaling van de beste steekproeflocaties en de beste sensorlocaties van continue metingen. In §2.2 worden de methodes besproken die in voorliggend onderzoek zijn toegepast voor de bepaling van de beste meetlocaties voor steekproeven en continue metingen. De toegepaste methodes zijn deels handmatige bepaling van locaties (ondersteund door een leidingnetmodel) en deels een bepaling van locaties met behulp van wiskundige optimalisatietechnieken.

2.1 Literatuuronderzoek

2.1.1 Locaties voor steekproeven

Speight e.a. (2004) hebben een voorstel gedaan om de willekeurige monsterlocaties gestratificeerd te bepalen. Stratificatie vindt plaats op basis van leidingdiameter, leidingmateriaal, afstand tot pompstation. Het aantal samples wordt bepaald op basis van de te verwachten variantie en betrouwbaarheidsniveau. Binnen de strata zijn de monsterlocaties willekeurig. De reden voor deze aanpak ligt veel meer in het evenredig verdelen van de monsterlocaties over het distributienet (en zo niet slechts een lokaal beeld te verkrijgen) dan in de verbetering van de detectiekans.

Lee en Deininger (2002) bepalen de meetlocaties op basis van verbruik op knopen en de dekkingsgraad ten opzichte van bovenstrooms gelegen knopen. Getracht is om dit analytisch op te lossen.

Chastain (2005) geeft een voorbeeld van een heuristische methode voor de optimalisatie van meetlocaties. Daarbij wordt een (klein) distributienetwerk vele malen doorgerekend met iedere mogelijke knoop als besmettingslocatie en detectielocatie. De knopen waar de meeste besmettingen worden gedetecteerd zijn de optimale meetlocaties. Voor grotere netwerken zal dit echter te veel rekentijd kosten.

Boccelli en Hart (2007) kiezen monsterlocaties als representatieve set van het netwerk op basis van de verblijftijden. N.B. in de USA zijn verblijftijden waarschijnlijk belangrijker dan in Nederland in verband met de afname van chloor in het distributienet.

Geen van genoemde onderzoeken doen een uitspraak over de hoogte van de detectiekans of de verbetering van de detectiekans door een bepaalde methode te volgen.

2.1.2 Locaties voor continue metingen

Sinds 11 september 2001 wordt er veel onderzoek gedaan naar online monitoring van waterkwaliteit in verband met mogelijke terroristische aanvallen op het drinkwaternet. Het doel van detectie is tweeledig. Ten eerste is een snelle detectie van belang voor de volksgezondheid (*early warning systems - EWS*), ten tweede is het een hulpmiddel bij het bepalen van de locatie van de besmettingsbron (*source location inversion - SLI*).

Er zijn verschillende methoden onderzocht om de optimale meetlocatie te bepalen. Uber (2005) heeft met behulp van lineaire algebra aangetoond dat bij een beperkt aantal meetlocaties de detectiekans laag is en de identificatie van de bronlocatie lang niet altijd mogelijk is doordat de opgestelde matrices 'rank-deficient' zijn, vooral doordat er veel 0-en in voorkomen. Dit betekent dat er dus (vele) besmettingen mogelijk zijn die niet langs de sensoren stromen.

Wanneer er vanuit wordt gegaan dat iedere locatie een even grote kans op besmetting heeft is de detectiekans weliswaar laag, maar kunnen wel optimale monsterlocaties worden vastgesteld. De meest gebruikte methode bij het vaststellen van locaties ten behoeve van EWS is *integer linear programming* (Propato et al., 2005), waarbij de relatie tussen besmettingslocaties en sensorlocaties in lineaire formules wordt uitgedrukt en bijvoorbeeld de tijd tot detectie kan worden geminimaliseerd. Berry e.a. (2005b) concluderen dat de veronderstellingen van het model de uitkomst beïnvloeden, zoals de duur van de besmetting of het tijdstip op de dag. Ze stellen verder vast dat het model geoptimaliseerd kan worden op het minimaliseren van de impact op de gezondheid (aantal geïnfecteerden) of op de snelheid van detectie. McKenna e.a. (2005) hebben gekeken naar de invloed van de stochasticiteit van afnamepatronen op de mogelijkheid van SLI. Er wordt geen uitspraak gedaan over het effect van stochastische afnamepatronen op de totale detectiekans, maar wanneer een besmetting is gedetecteerd is de invloed van de stochasticiteit op het lokaliseren van de bron beperkt. Een andere keuze of randvoorwaarde is de minimalisatie van kosten. Berry e.a. (2005a) bekijken bijvoorbeeld de kosten per sensor en de kosten voor het plaatsen afhankelijk van de locatie; een sensor in een leiding die opgegraven moet worden en waar stroom naartoe gebracht moet worden kost meer dan een sensor op een pompstation. De voorbeelden die al deze auteurs geven zijn oplossingen voor kleine distributienetwerken. Een aantal auteurs heeft zich ook gebogen over het opschalen naar grote netwerken (Watson et al., 2005); (Berry et al., 2005c).

Een andere benadering is dat rekening wordt gehouden met de kans op besmettingen per locatie. De locaties die het grootste besmettingsgevaar opleveren moeten in elk geval door de sensorlocaties worden afgedekt. Tidwell e.a. (2005) gebruiken bijvoorbeeld *Markov Latent Effects modeling* om de dreigingen in te schatten en combineert dit met een EPANET netwerkmodel waarmee de effecten worden berekend. De hoogste 'gevaar'-score werd in hun voorbeeld behaald door de (opzettelijke) biologische besmetting van een reinwaterkelder door een medewerker van het waterbedrijf. Grayman e.a. (2005) hebben in de vorm van een spel getracht de kans op terroristische besmettingen per locatie in beeld te brengen. Het 'rode team' had beperkte informatie over het leidingnet en koos de besmettingslocaties; het 'blauwe team' had alle informatie over het leidingnet beschikbaar en moest zich zo goed mogelijk wapenen tegen besmettingen door op de juiste locaties de waterkwaliteit te monitoren.

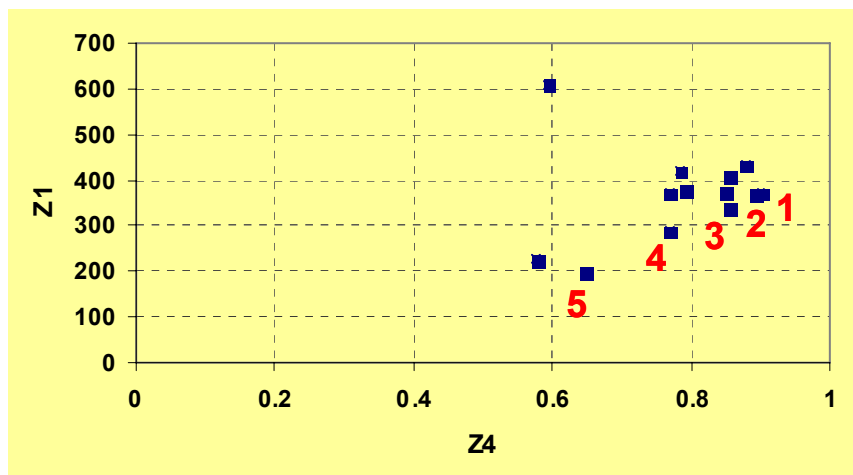
Tijdens de *8th annual symposium on Water Distribution System Analysis* is een wedstrijd gehouden in het vaststellen van de beste sensorlocaties (oftewel *the battle of the water sensor networks*). Een definitieve publicatie wordt verwacht eind 2007. De beschrijving hierna is gebaseerd op de publicaties van het symposium en de samenvattende presentatie.

Voor 2 netwerken (een kleine, ca. 100 knopen, en een grote, ca. 12.000 knopen) en een aantal verschillende verontreinigingsscenario's (een verontreiniging van 230 g/l met een volumestroom van

125 l/h gedurende 2 (scenario A) of 10 uur (scenario B), op één injectiepunt of simultaan op twee locaties (scenario D), zonder en met een vertraging in detectie (scenario C)) moesten 5 en 20 optimale meetlocaties worden gegeven. Overige randvoorwaarden waren een optimale detectielimiet, bekende afnamepatronen en alle knopen hebben een gelijke kans als injectiepunt te dienen. Een viertal optimalisatiedoelen werd gesteld waarop de deelnemers werden beoordeeld:

- Z_1 : het minimaliseren van de tijd tot detectie;
- Z_2 : het minimaliseren van het aantal getroffen mensen (voor detectie);
- Z_3 : het minimaliseren van het verontreinigde geconsumeerde volume (voor detectie);
- Z_4 : het maximaliseren van de detectiekans.

Bij het optimaliseren van de vier (niet geheel onafhankelijke) parameters is de beste oplossing veeleer een paretofront waarbij veel goede oplossingen mogelijk zijn maar een afweging kan worden gemaakt welke parameter het belangrijkste wordt gevonden (zie bijvoorbeeld figuur 2-1). Bij het beoordelen van de verschillende oplossingen is dan ook gekeken naar de paretofronts van Z_1 , Z_2 en Z_3 ten opzichte van Z_4 . Een winnaar in de eerste categorie kan minder goed scoren in de andere categorieën. Zodoende was er niet slechts één winnaar aan te wijzen. In tabel 2-1 is voor alle deelnemers samengevat met welke methode ze welke score hebben behaald.



figuur 2-1 Voorbeeld van een set oplossingen. Oplossingen voor netwerk 1, scenario A, 20 sensorlocaties. De oplossingen 1-5 vormen het paretofront bij de afweging tussen Z_1 en Z_4 . [1 = Wu and Walski, 2 = Eliades and Polycarpou, 3 = Guan et al., 4 = Berry et al., 5 = Ostfeld and Salomons].

tabel 2-1 Overzicht van oplossing voor de battle of the water sensor networks plus score (bij welk scenario was de oplossing deel van het paretofront, hoe meer scenario's hoe algemener toepasbaar, dus betere methode), afkortingen in noten¹.

auteurs	titel	beschrijving	score
Berry et al.	A facility location approach to sensor placement optimization	Gebruiken GRASP + local search Optimaliseren op slechts Z_2 en Z_4 . Scenario D (simultane verontr.) is niet doorgerekend.	N1S5Z2 N1S5Z3 N1S20Z1 N1S20Z2 N1S20Z3 N2S5Z1 N2S5Z2 N2S5Z3 N2S20Z1 N2S20Z2 N2S20Z3
Dorini et al.	An efficient algorithm for sensor placement in WDS	Gebruiken zogenaamd Noisy Cross-Entropy Sensor Location algoritme	N1S5Z1 N1S5Z2 N1S5Z3 N1S20Z2 N2S5Z2 N2S5Z3
Eliades and Polycarpou	Iterative deepening of pareto solutions in water sensor networks	Zoeken niet de beste oplossing, maar 'good enough'. Eerst wordt de beste sensorlocatie (SL) gezocht (paretofront van oplossingen), dan de beste SL van deze + een 2 ^e etc. Kiezen een vijfde extra parameter: optimaliseren van 'demand coverage'.	N1S5Z2 N1S20Z1 N1S20Z2 N1S20Z3
Ghimire and Barkdoll	Heuristic method for the BWNS demand-based approach	Plaatsen sensoren op knopen met de hoogste verbruiken	N1S5Z1
Guan et al.	Optimization model and algorithm for design of water sensor placement in WDS	Gebruiken verbeterde GA (eerst lokaal optimum) en vormen multi-objective om naar single-objective door de 4 parameters in een enkele index te vatten. Verlengen de rekenstap (30 minuten ipv 5), rekenen voor N1 20 verontreinigingsscenario's per knoop door, voor N2 300 scenario's op de knopen met hoogste verbruik	N1S5Z2 N1S5Z3 N1S20Z1 N1S20Z3
Gueli	Predator-prey model for discrete sensor placement	Gebruikt standaard GA ??	X
Huang et al.	Multi-objective optimization for monitoring sensor placement in WDS	Gebruiken GA. Model vereenvoudigd: slechts gekeken naar leidingen > 200 mm.	N1S5Z1

¹ NaSbZc: Netwerk a (1, 2) – alleen scenario A; b aantal Sensoren (5, 20); criterium Z_c (c = 1, 2 of 3) ten opzichte van Z_4

BWSN: battle of the water sensor networks

WDS: water distribution systems

GA: genetic algorithm – NSGA-II: non-dominated sorted GA – II

SLI: source location identification

auteurs	titel	beschrijving	score
Krause et al.	Optimizing sensor placements in WDS using submodular function maximization	Gebruiken Greedy algoritme. Hebben middels het koppelen van PC's (distributed computation) alle mogelijke scenario's volledig doorgerekend	X
Ostfeld and Salomons	Sensor network design proposal for BWSN	Gebruiken NSGA-II. Onduidelijk welke vereenvoudigingen zijn aangenomen	N1S5Z1 N1S20Z1
Preis and Ostfeld	Multiobjective sensor design for WDS security	Gebruiken NSGA-II. Optimaliseren niet Z_2 en Z_3 , maar voeren andere parameters op naast Z_1 en Z_4 (namelijk maximalisatie van redundancy en maximaliseren van SLI). Naast 2D paretofronts (p.f.) laten ze ook 3D p.f. zien en beschrijven 4D p.f. (in database)	N1S5Z1
Propato and Piller	Battle of the water sensor networks	Gebruiken MILP (mixed-integer linear program). Rekenstap vergroot om rekentijd te bekorten. Hebben tevens onnauwkeurigheid berekend.	N1S5Z2 N1S5Z3
Trachtman	A "strawman" common sense approach for water quality sensor site selection	Kiest met de hand locaties op basis van hydraulica (verbruik) en speciale klanten (bijv. dialysecentra)	N1S5Z1 N1S5Z2
Wu and Walski	Multiobjective optimization of sensor placement in WDS	Vormen multi-objective om naar single-objective door 3 parameters in een enkele index te vatten (Z_2 wordt als afhankelijke parameter verder niet meegenomen). Reduceren het aantal knopen (verdunnen model) en reduceren het aantal mogelijke verontreinigingslocaties.	N1S5Z1 N1S20Z1 N1S20Z3 N2S5Z1 N2S5Z2 N2S5Z1 N2S20Z1 N2S20Z2 N2S20Z3

Wat opvalt is dat de oplossingen van Berry et al., Dorini et al. and Wu en Walski vaak hoog scoren, terwijl soortgelijke oplossingen slechter scoren (zoals Propato en Piller). Eenvoudige oplossingen zoals van Trachtman en Ghimire and Barkdoll horen soms (vooral bij N1S5) toch ook bij de beste oplossingen.

De beste twintig sensorlocaties bevatten niet persé de beste vijf sensorlocaties. Dit blijkt uit de oplossingen die alle partijen hebben aangeleverd. Uit de oplossingsrichting van Eliades en Polycarpou blijkt dat deze stelling vooral bij een relatief klein aantal sensoren geldt (bij N1S5 en N2S5 scoren zij niet goed, bij N1S20 scoren zij wel goed).

Een aantal kanttekeningen werd geplaatst bij de 'wedstrijd'. Het aantal doorgerekende scenario's in netwerk 2 was bijvoorbeeld maar een deelverzameling (1%); een nieuwe berekening kan tot andere winnaars leiden. Opgemerkt werd ook dat veel deelnemers met hun eigen berekeningen beter scoorden dan in de wedstrijd. Blijkbaar waren de modellen toch niet 100% vergelijkbaar.

Een aantal verontreinigingen leidt bij een bepaalde set sensoren tot geen detectie – op dit moment zijn deze oplossingen niet meegewogen in het eindoordeel terwijl dat natuurlijk wel zou moeten (en het eindoordeel negatief beïnvloedt)². Tevens werd opgemerkt dat het ook mogelijk is om de vier parameters middels een 4D paretofront tegen elkaar af te wegen. Er zal dan een duidelijker onderscheid tussen 'winnaars' en 'verliezers' optreden.

² Wanneer $Z_4 = 0$ kunnen Z_1 , Z_2 , Z_3 niet worden bepaald. Eigenlijk zouden deze dan bijv. gemaximaliseerd moeten worden. In het eindoordeel zijn deze nu gewoon buiten beschouwing gelaten.

2.2 Toegepaste methodes voor de bepaling van meetlocaties

2.2.1 Meetlocaties steekproeven op basis van verblijftijd

In deze paragraaf wordt een door Kiwa Water Research ontwikkelde methode beschreven om, bij gegeven monsterfrequentie, een hoge detectiekans te bewerkstelligen door monsterlocaties te kiezen op basis van een leidingnetberekening. De detectiekans wordt verondersteld toe te nemen met de verblijftijd van (verontreinigd) water in het distributienet. Bemonstering op een locatie waar verontreinigd water lang verblijft vergroot de detectiekans.

De methode maakt gebruik van de volgende veronderstellingen:

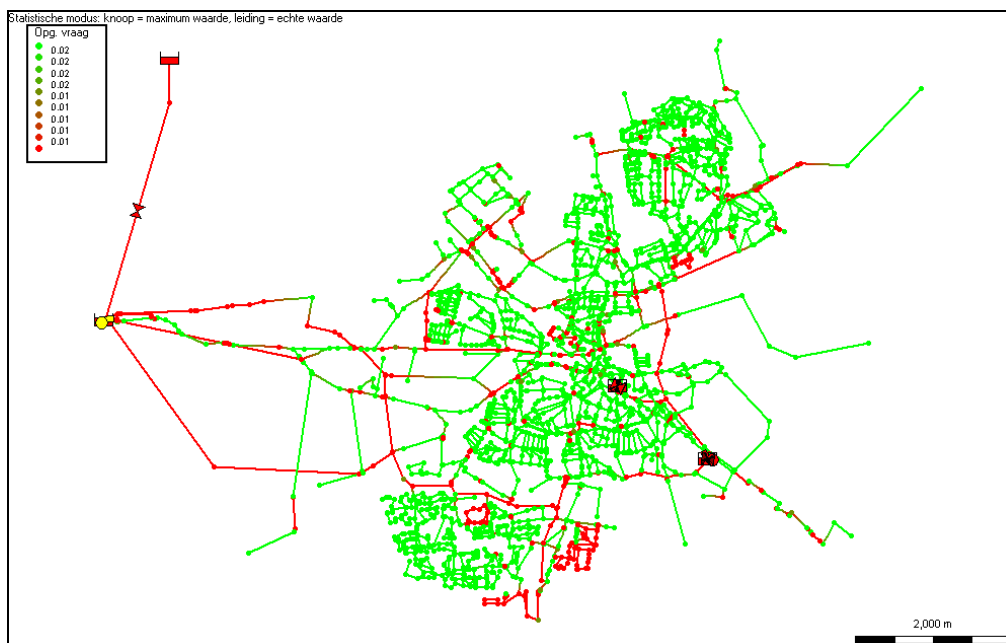
- een verontreiniging vanaf de productielocatie zal alle verbruiksknoppen in haar voorzieningsgebied bereiken (binnen 40 dagen);
- een verontreiniging wordt gesimuleerd als een conservatieve stof (geen afsterving);
- afhankelijk van de leidingconfiguratie en berekende stroomsnelheden zal in bepaalde delen van het net (uitlopers en pendelzones) een relatief lange verblijftijd heersen waardoor er in het leidingnet verschillen ontstaan in de tijdsduur waarin de detectielimiet wordt overschreden (en bij steekproefname de detectiekans toeneemt).

De aanpak bestaat uit drie stappen die hieronder nader worden toegelicht.

N.B. In het model zijn de verontreinigingsconcentratie en de detectielimiet vrij te kiezen variabelen. Verhoging van de verontreinigingsconcentratie of verlaging van de detectielimiet leiden beide tot een verhoging van de detectiekans. In stap 1 wordt de verontreinigingsconcentratie gekozen en in stap 2 wordt de detectielimiet zodanig gekozen dat een werkbare uitkomst volgt voor stap 3. De detectielimiet in deze voorbereidingsfase (namelijk het kiezen van meetlocaties) heeft hier niets te maken met een werkelijke detectielimiet.

Stap 1. Gereedmaken van model voor waterkwaliteitsberekening

Het netwerkmodel (in dit geval van Almelo, zie ook §3.1.1) wordt gereed gemaakt voor een waterkwaliteitsberekening. Doodlopende leidingen met inwendige diameter < 1000 mm én een drukval < 0,001 mwk worden verwijderd met het programma Verdunner®. Een verontreiniging (gedurende 16 uur een dosering van 10 L/u met 10^7 E.coli-bacteriën/L) op de uitgaande leiding van de productielocatie wordt ingebouwd. Dit levert het basismodel (figuur 2-2). De verspreiding van de verontreiniging wordt voor 240 uur doorgerekend (met een rekentijdstap van 1 uur).

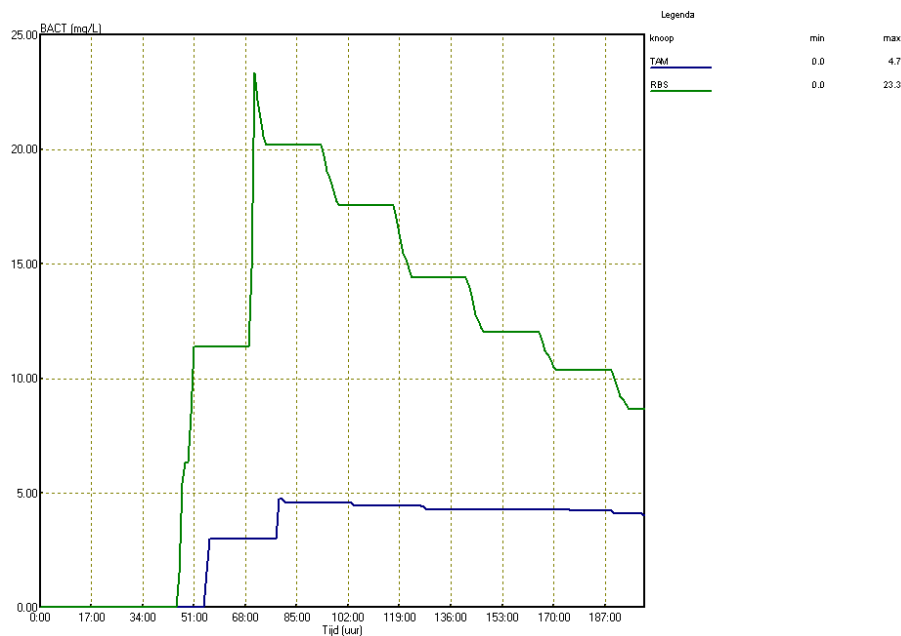


figuur 2-2 Basismodel (verdund) met verontreiniging op productielocatie.

Stap 2. Analyse van het berekende concentratieverloop

a. Keuze van een maatgevende detectielimiet in het model

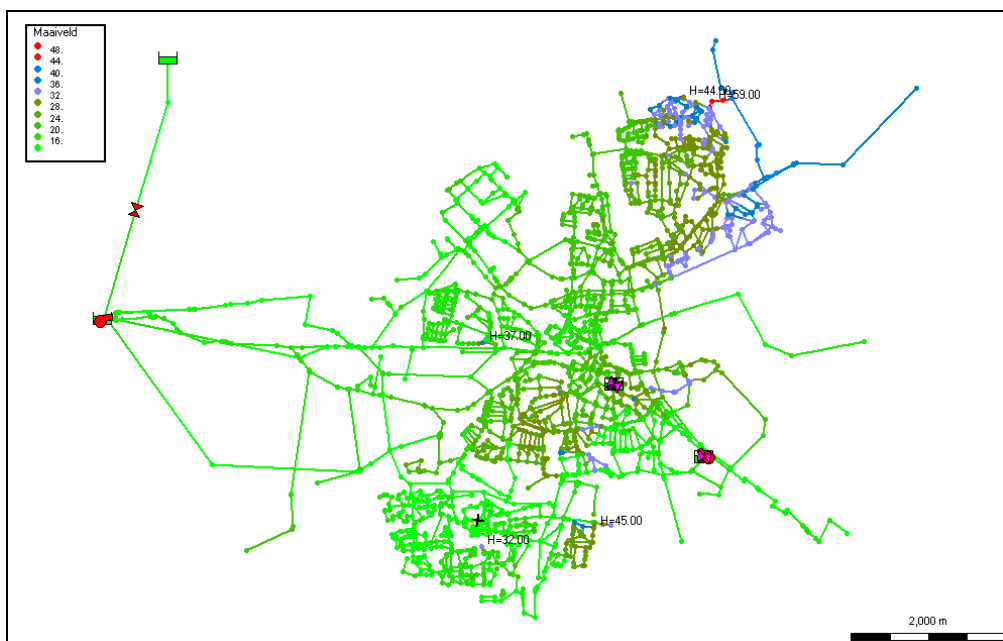
Aan de hand van het concentratieverloop in de reservoirs wordt een maatgevende detectielimiet voor het model vastgesteld. Voor de situatie te Almelo is deze vastgesteld op 25 KVD/L; iets hoger dan de maximale concentratie gemeten in de reservoirs. De potentiële optimale monsterpunten dienen namelijk een meerwaarde te hebben ten opzichte van de toch al frequente meting in de reservoirs, zie figuur 2-3.



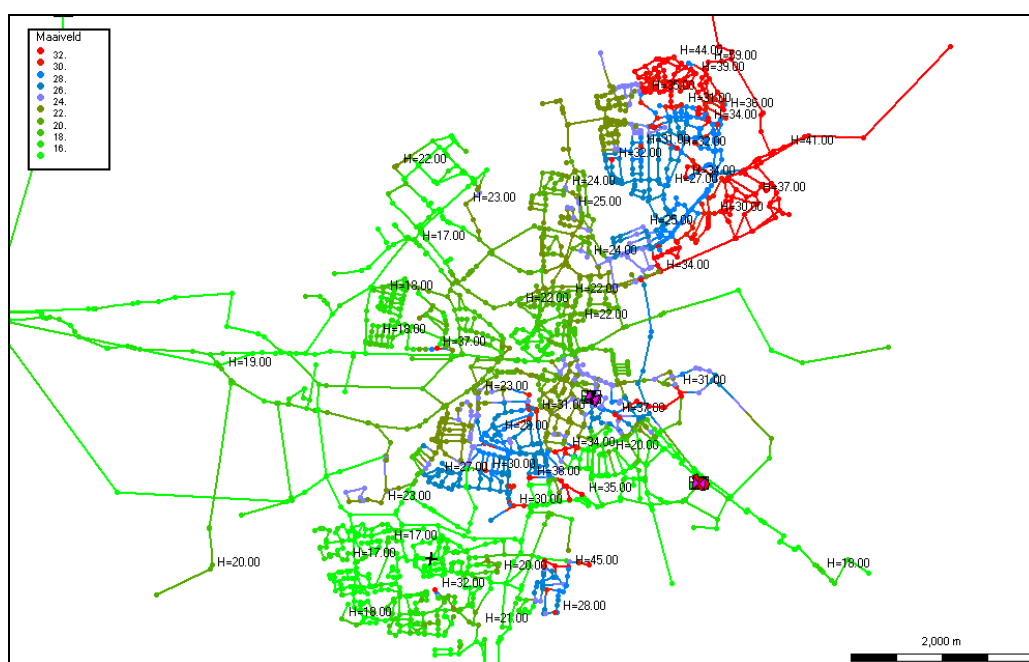
figuur 2-3 Concentratieverloop in reservoirs (TAM en RBS)

b. Berekening van de overschrijdingsduur van de detectielimiet op elke knoop

Voor alle knopen wordt bepaald gedurende hoeveel uur (tussen 0 en 240 uur) de concentratie boven de detectielimiet uitkomt. Per wijk worden de monsterlocaties bepaald door de knopen te kiezen met het hoogste aantal positieve monsters. Dit kan worden gedaan middels een grafische weergave van het aantal positieve monsters in het leidingnet en in te zoomen door de stapgrootte te variëren, zie figuur 2-4 en figuur 2-5 (N.B. er is gebruik gemaakt van het ALEID-dataveld 'maaiveld-hoogte' als hulpmiddel om het aantal positieve monsters grafisch weer te geven).



figuur 2-4 Resultaat 1; aantal positieve monsters >25 KVD/L met legenda stapgrootte 4.



figuur 2-5 Resultaat 2; aantal positieve monsters >25 KVD/L met legenda stapgrootte 2.

Stap 3. Aanwijzen van monsterknopen

Vervolgens zijn de monsterknopen geselecteerd; Figuur 3-2 op pagina 18 geeft een overzicht van de geselecteerde optimale monsterlocaties.

2.2.2 Meetlocaties met poortwachterfunctie

In deze paragraaf wordt een methode beschreven om in geval van (semi-) continue metingen een hoge detectiekans te bewerkstelligen door monsterlocaties (handmatig) te kiezen op basis van hun poortwachterfunctie. Verondersteld wordt dat de optimale locaties gekenmerkt worden door een relatief

hoge volumestroom in betreffende leidingen en een korte reistijd hebben vanaf de productielocatie ten opzichte van de rest van de wijk.

Bij de selectie van geschikte knooppunten is eerst het totale modelgebied intuïtief verdeeld in circa 15 tot 20 wijken. Vervolgens zijn (aan de hand van berekende concentratiepatronen voor verschillende tijdstappen als beschreven onder stap 1 in §2.2.1) circa 20 monsterpunten geselecteerd; zie Figuur 3-2 op pagina 18.

2.2.3 Meetlocaties (semi)continue metingen op basis van mMOSA

Aan de universiteit van Exeter heeft een student met het mMOSA-algoritme (modified Multi-Objective Simulated Annealing) de optimale meetlocaties gekozen waarbij het geconsumeerde volume verontreinigd drinkwater is beperkt. Om de rekentijd te beperken heeft hij gebruik gemaakt van een *Importance based Sampling Method*, d.w.z. dat niet alle mogelijke verontreinigingsscenario's worden doorgerekend maar een (willekeurige) selectie op basis van geconsumeerd volume op knoop n en tijd t . Voor een uitgebreide beschrijving, zie bijlage I.

De gebruikte optimalisatie is niet per sé gelijk aan de optimalisatie van de detectiekans, maar de beide optimalisaties zijn niet volledig onafhankelijk omdat zonder detectie er ook geen beperking van de consumptie is.

De optimale samenstelling van meetlocaties met 5, 14 en 30 sensoren zijn in voorliggend onderzoek beoordeeld. Zie Figuur 3-2 op pagina 18 voor de meetlocaties.

2.2.4 Meetlocaties (semi)continue metingen op basis van NSGA-II

Daarnaast heeft de student aan de universiteit van Exeter met het NSGA-II-algoritme (Non dominated Sorting Genetic Algorithm, extension) de optimale meetlocaties gekozen waarbij het geconsumeerde volume verontreinigd drinkwater is beperkt. Om de rekentijd te beperken heeft hij gebruik gemaakt van een *Importance based Sampling Method*. Voor een uitgebreide beschrijving, zie bijlage I.

De gebruikte optimalisatie is niet per sé gelijk aan de optimalisatie van de detectiekans, maar de beide optimalisaties zijn niet volledig onafhankelijk omdat zonder detectie er ook geen beperking van de consumptie is.

De optimale samenstelling van meetlocaties met 5, 15 en 25 sensoren zijn in voorliggend onderzoek beoordeeld. Zie Figuur 3-2 op pagina 18 voor de meetlocaties.

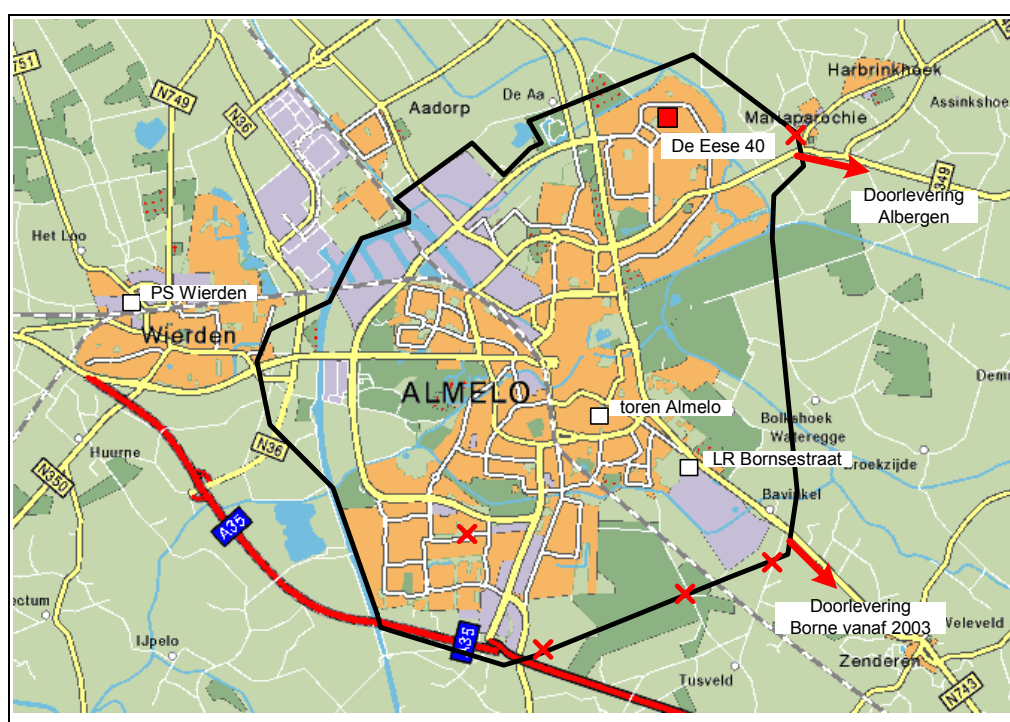
3 Bepaling van de detectiekans

3.1 Beschrijving simulatie

3.1.1 Voorzieningsgebied Almelo

In 2004 (Lieverloo et al., 2004) is Almelo willekeurig (maar niet formeel aselekt) gekozen uit de beschikbare gedetailleerde leidingnetmodellen. Ditzelfde gebied (en hetzelfde leidingnetmodel) is weer gebruikt in voorliggend onderzoek.

Het gebied wordt voorzien vanuit één drinkwaterproductiebedrijf (Wierden), de watertoren Almelo en het laagreservoir Bornestraat. In figuur 3-1 is het voorzieningsgebied aangegeven. Zie voor een uitgebreide beschrijving (Lieverloo et al., 2004).



figuur 3-1 Voorzieningsgebied Almelo en omgeving

3.1.2 Handmatige ALEID-simulaties van verontreinigingen

Zie bijlage II.

3.1.3 ALEID-dll-simulaties van verontreinigingen

Omdat bleek voor een voldoende statistisch betrouwbare uitkomst meer verontreinigingen door gerekend moesten worden dan handmatig mogelijk was, is gebruik gemaakt van de in Exeter ontwikkelde en gebruikte ALEID.dll (zie ook bijlage II). Om de resultaten van voorliggend onderzoek te kunnen vergelijken met de resultaten uit 2004 zijn een aantal instellingen in de berekening aangepast en afgestemd op de resultaten van 2004.

Voor de simulaties is gebruik gemaakt van een verontreiniging met een continue concentratie van 10^3 KVD *E. coli* per liter drinkwater, die gedurende 16 uur in het leidingnet aanwezig is, gestart om 00.00 uur (12.00 uur 's nachts). De berekeningen zijn met intervallen van 1 uur uitgevoerd en per uur

gerapporteerd voor 40 dagen na de start van de verontreiniging. Per verontreiniging zijn de concentratieverlopen van 960 uurmetingen op alle (meer dan 1700) mogelijke monsterlocaties (incl. productiebedrijf, reservoirs en de 44 woningen van het monsterprogramma 2002) berekend.

De verontreinigingen zijn op de volgende plaatsen gesimuleerd:

- Op 50 distributieleidingen, waarvan 9 op gelijke knopen als in 2004, 11 knopen zijn handmatig gekozen en de overige 30 zijn willekeurig in het leidingnet gekozen.

3.1.4 Beschrijving meetprogramma's

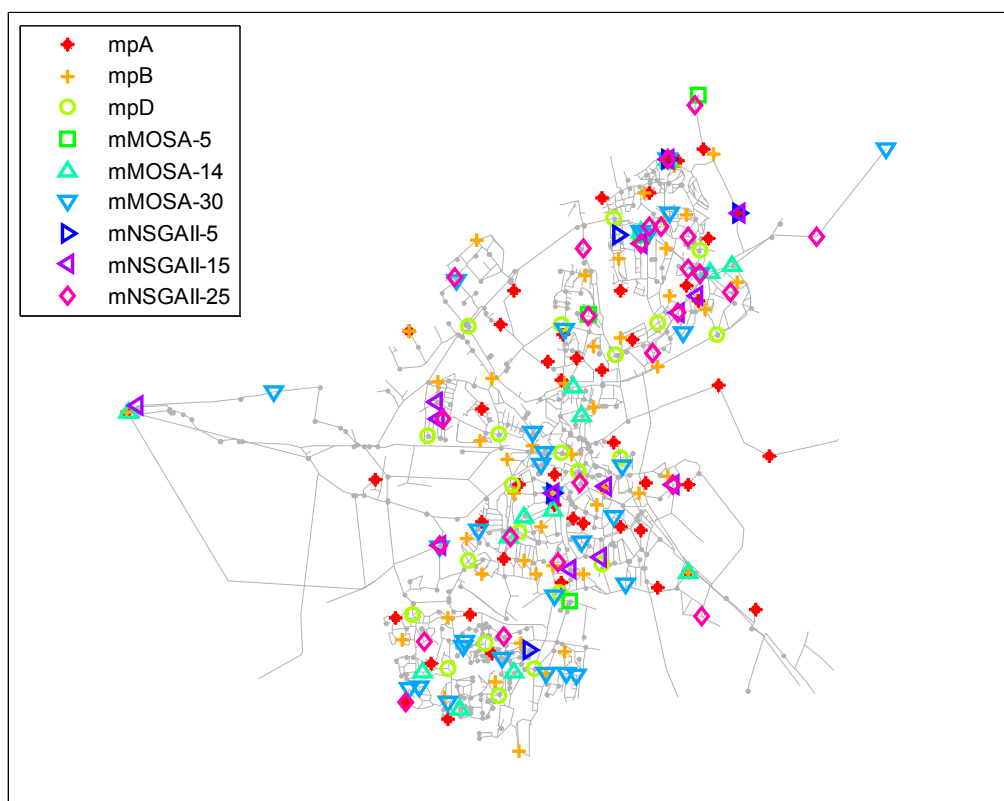
Vier verschillende meetprogramma's zijn met elkaar vergeleken, zie tabel 3-1:

- Monsterprogramma **mpA** is het monsterprogramma van pompstation en reservoirs en 44 locaties in het net, dat Vitens in 2002 hanteerde, zie (Lieverloo et al., 2004).
- Monsterprogramma **mpB** is het geoptimaliseerde monsterprogramma van pompstation en reservoirs en 44 locaties, bepaald volgens §2.2.1.
- Monsterprogramma **mpC** is een volledig willekeurig monsterprogramma dat geanalyseerd wordt met behulp van een Monte-Carlo-simulatie.
- Meetprogramma **mpD** vervult de poortwachtersfunctie, 22 locaties bepaald volgens §2.2.2.

Daarnaast is een aantal sensorlocaties bepaald met als doel de consumptie van verontreinigd drinkwater zo veel mogelijk te beperken. De voor dit doel optimale sensorlocaties zullen waarschijnlijk ook een hoge detectiekans leveren:

- Meetprogramma **mMOSA** (zie §2.2.3)
- Meetprogramma **NSGA-II** (zie §2.2.4)

De verschillende meetlocaties zijn in onderstaande figuur weergegeven.



Figuur 3-2 Meetlocaties van de verschillende meetprogramma's.

Naast variatie in meetlocaties is ook gekeken naar de invloed van de monsterfrequentie. Dat betekent het opvoeren van de monsterfrequentie in het netwerk zonder al over te gaan naar (semi)continue sensoren.

tabel 3-1 Beschrijving monsterprogramma's.

			§ 3.2.1	§ 3.2.2	§ 3.2.3	§ 3.2.4	§ 3.2.5		
type meting			bemonstering 24 h	bemonstering overdag	bemonstering 24 h	continue sensoren	semi-continue sensoren		
tijdstip meting			0:00 - 23:59	8:00 - 17:00	0:00 - 23:59	0:00 - 23:59	0:00 - 23:59		
meetfrequentie (# per jaar)	pompstation		52			(365 x 24)	(52 x 7 x 24)		
	hoogreservoir		12						
	laagreservoir		12						
	rest		4						
	rest netwerk		variabel						
			aantal locaties			44		variabel: 5 - 45	
locaties in netwerk verschillende meet- programma's	mpA		locaties 2002						
	mpB		locaties op basis van verblijftijd (§2.1.2)						
	mpC		locaties op basis van Monte-Carlo-simulatie						
	mpD		nvt			poortwachterlocaties (§2.2.2)			
	mMOSA					locaties o.b.v. optimalisatie met mMOSA (§2.2.3)			
	NSGA-II					locaties o.b.v. optimalisatie met NSGA-II (§2.2.4)			
# verontreinigingslocaties in het netwerk dat is doorgerekend			50			50			

3.1.5 Berekening van de detectiekans

De uitkomst van de ALEID-simulaties van verontreinigingen (vorige paragraaf) is een concentratie over de tijd op alle mogelijke meetlocaties. De detectiekans wordt vervolgens bepaald met de volgende formules.

para-meter	omschrijving	waarde
t	teller voor tijdstip	uur 1 t/m 960
v	teller voor verontreiniging	1 t/m 50
l	teller voor monsternamen- / sensorlocatie	Monsternamen: 1 t/m 47 (44 in net, 1 op PS, 2 op reservoirs) Sensoren: variabel 1 t/m 5, 15, 25, 35 en 45
$C_{t,v,l}$	concentratie op tijdstip t ten gevolge van verontreiniging v op locatie l	Uitkomst waterkwaliteitsberekening ALEID
$D_{t,v,l}$	wel/geen detectie op tijdstip t ten gevolge van verontreiniging v op locatie l	$D_t = \begin{cases} 1, & C_t \geq \text{detectielimiet} \\ 0, & C_t < \text{detectielimiet} \end{cases}$ detectielimiet = 5 bij de berekening beschreven in §3.1.3 (resultaten in §3.2).
$Q_{t,l}$	kans op monsternamen op tijdstip t en locatie l	$Q_{t,l}^{uur} = \begin{cases} 1/24, & \text{monsternamen over 24 h} \\ 1/10, & \text{overdag} \\ 0, & \text{'s nachts} \end{cases}$ monsternamen tijdens werktijden $Q_{t,l}^{dag} = \begin{cases} 1, & \text{(semi)continue sensoren} \\ 0, & \text{geen sensoren/monsternamen} \end{cases}$ $Q_{t,l}^{dag} = \begin{cases} 4/365, & \text{monsternamen in het net} \\ 12/365, & \text{monsternamen in reservoirs} \\ 52/365, & \text{monsternamen op PS} \\ 1, & \text{(semi)continue sensoren} \\ 0, & \text{geen sensoren/monsternamen} \end{cases}$
$P_{v,l}$	kans op detectie van verontreiniging v op locatie l	$P_{v,l} = \sum_{t=1}^{960} D_{t,v,l} \cdot Q_{t,l}^{uur} \cdot Q_{t,l}^{dag}$

para- meter	omschrijving	waarde
P_v	kans op detectie van verontreiniging v	$P_v = \begin{cases} 1, & \sum_l P_{v,l} > 1 \\ \sum_l P_{v,l}, & \sum_l P_{v,l} \leq 1 \end{cases}$
P	kans op detectie van een willekeurige verontreiniging	$P = \frac{1}{50} \sum_{v=1}^{50} P_v$

In de volgende paragrafen zijn de volgende grafieken geplot:

- Bij monsternamen: een box-whisker-plot van alle P_v
 - de rode horizontale lijn is de mediaan;
 - de blauwe 'box' geeft de 25- tot 75-percentielen weer;
 - de 'whiskers' (snorharen) geven de waarden aan die binnen een bepaalde grens liggen;
 - buiten de whiskers zijn mogelijk nog enige uitschieters weergegeven;
 - de grijze punten zijn alle waarden (N.B. gelijke waarden vallen over elkaar heen);
 - de groene cirkel is het gemiddelde.
- Bij sensoren: een box-whisker-plot van alle P
 - de rode horizontale lijn is de mediaan;
 - de blauwe 'box' geeft de 25- tot 75-percentielen weer;
 - de 'whiskers' (snorharen) geven de waarden aan die binnen een bepaalde grens liggen;
 - buiten de whiskers zijn mogelijk nog enige uitschieters weergegeven;
 - extra punten geven de resultaten weer van de individueel vastgestelde (optimale) meetlocaties (de meetprogramma's mpA, mpB, mpD, mMOSA en NSGA-II).

De detectie van verontreinigingen op pompstation en reservoirs is niet meegenomen in de analyse. De detectiekans is namelijk 100% bij monsternaming op deze locaties. Omdat reeds bij productie (of reservoirs) een verontreiniging is gedetecteerd hoeft verder in het net niet gemeten te worden, omdat dan al een kookadvies is uitgegaan (of andere maatregelen zijn getroffen). De statistische analyse is gedaan op basis van 50 verontreinigingslocaties.

De invloed van de concentratie van de besmetting respectievelijk de invloed van de detectielimiet op de detectiekans is reeds in 2004 bepaald. In dit project wordt met de automatische simulatie met de ALEID-dll gerekend met een verontreinigingsconcentratie van 10^3 KVD/L en met een detectiegrens van 5 KVD/L. Dit levert vergelijkbare resultaten op als het onderzoek in 2004 met een verontreiniging van 16 liter ofwel een vaste detectiegrens van 10 KVD/L (figuren 42 en 43 op pagina 39 van (Lieverloo et al., 2004)). In bijlage II zijn de beide methoden met elkaar vergeleken. N.B. de simulatie (met de ALEID-dll) en een verontreinigingsconcentratie van $2 \cdot 10^3$ KVD/L en met een detectiegrens van 10 KVD/L levert exact dezelfde resultaten op.

De resultaten van 2004 en voorliggend onderzoek zijn niet één-op-één te vergelijken, omdat er enkele kleine verschillen in de berekening zitten.

3.2 Detectiekans bij verschillende scenario's

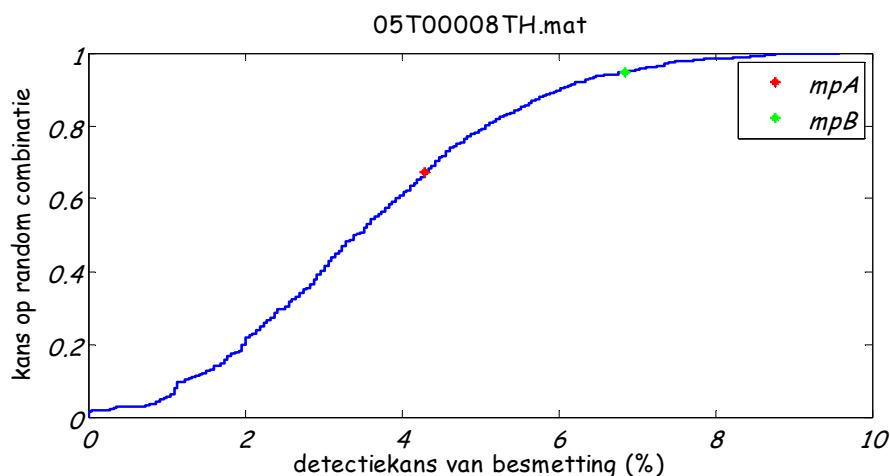
In dit project worden de verschillende monsterlocaties en wijze van monsternaming vergeleken op basis van de detectiekans. De verschillende scenario's die worden bekeken zijn:

- monsternamen
 - verschillende tijden van monsternamen, willekeurig binnen een dag van 24 uur (§3.2.1);
 - verschillende tijden van monsternamen, willekeurig binnen een werkdag (overdag, §3.2.2 – hierbij is geen rekening gehouden met de dag van de week);
 - verschillende tijden van monsternamen met een verhoging van de monsterfrequentie, willekeurig binnen een dag van 24 uur (§3.2.3);
- continue monitoring (§3.2.4);
- semi-continue monitoring (mengmonsters van 5 of 8 dagen, §3.2.5).

Locaties voor continue monitoring kunnen naast een grotere detectiekans ook een snellere detectie opleveren. Op detectiesnelheid zijn de verschillende sensorlocaties niet met elkaar vergeleken.

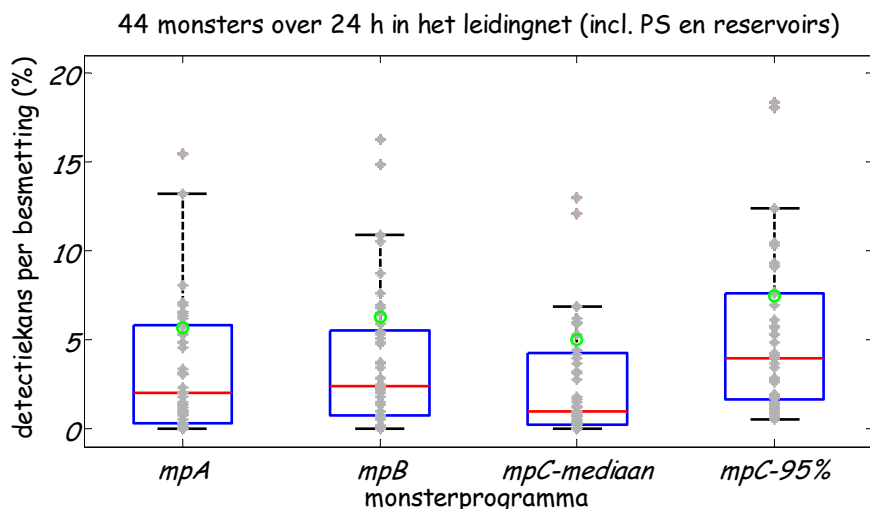
3.2.1 Monsternamen in het leidingnet – willekeurig over 24 uur

Voor ieder verontreinigingsscenario is voor 40 * 24 uur doorgerekend wat de concentratie is op ieder knooppunt. Vervolgens is voor verschillende monsterprogramma's (zie tabel 3-1) de detectiekans bepaald. Ter illustratie is in figuur 3-3 de detectiekans voor verontreiniging op knoop T00008TH (loc. 5) weergegeven; de doorgetrokken lijn is (alle mogelijke uitkomsten van) monsterprogramma C (bepaald met een Monte-Carlo-simulatie), de rode en groene '+' geven respectievelijk monsterprogramma A en B weer. Uit de figuur blijkt dat in ca. 95% van de willekeurige combinaties de detectiekans kleiner is dan 7%; de mediaan is 3,4%. Met het monster-programma van 2002 (mpA) is er ca. 4,3% kans op detectie, het geoptimaliseerde monsterprogramma (mpB) doet het in dit geval beter: 6,8% detectiekans.



figuur 3-3 Detectiekans van besmetting op locatie 5 bij verschillende monsternamenprogramma's (mpA: 2002, mpB: 2007 en mpC: Monte Carlo) gedurende 24 uur.

Om de detectiekans van alle 50 verontreinigingen met elkaar te vergelijken zijn deze per meetprogramma in een box-whisker-plot uitgezet. Voor mpC is de detectiekans bij de mediaan en bij het 95-percentiel van de willekeurige combinaties bepaald en in twee box-whisker-plots weergegeven. N.B. bij mpC worden telkens andere willekeurige meetlocaties gekozen. Het resultaat is weergegeven in figuur 3-4.



figuur 3-4 Detectiekans van besmetting op 50 verschillende locaties bij verschillende monsternamenprogramma's (mpA: 2002, mpB: 2007 en mpC: Monte Carlo) gedurende 24 uur.

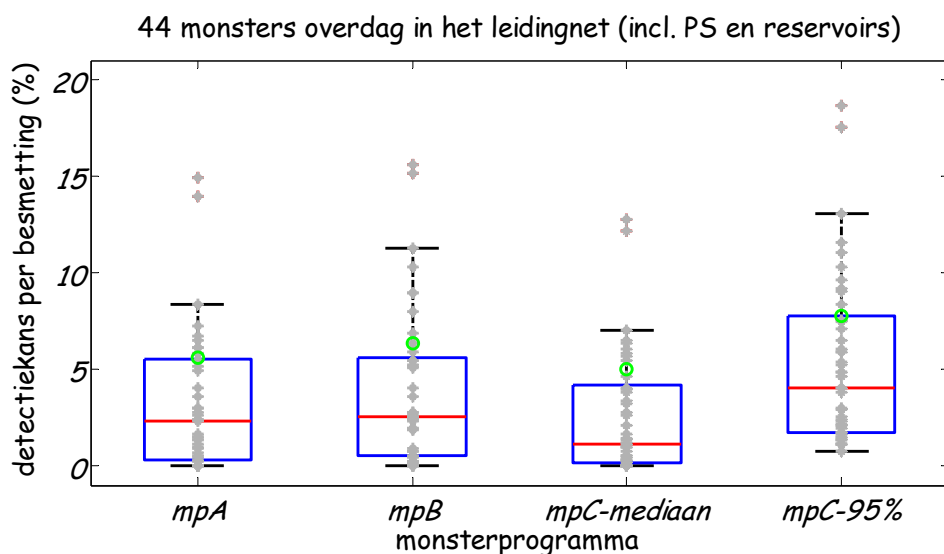
Uit figuur 3-4 blijkt dat wanneer tijdens een verontreiniging toevallig het beste monsterprogramma wordt toegepast (mpC-95%) de gemiddelde detectiekans 7,7% is (groene cirkel in de figuur). Wanneer een volledig willekeurig monsterprogramma (mpC-mediaan) wordt toegepast is de gemiddelde detectiekans 5,0%; bij mpA (dat een manifestatie is van mpC) is dat 5,6%. Bij mpB is de gemiddelde detectiekans 6,3%. Dit is een verbetering van 26% ten opzichte van de willekeurige monsternamen, maar de detectiekans blijft erg klein. Daarnaast blijkt uit de figuur dat in 50% van de gevallen (de mediaan) de geoptimaliseerde monsterlocaties (mpB) iets beter zijn dan de huidige (mpA) en ook beter dan de mediaan van de willekeurige locaties (mpC-mediaan). Er zijn per besmetting betere monsterlocaties te vinden, maar in 95% van de gevallen (mpC-95%) is er maximaal 13% kans op ontdekking van de besmetting. Deze 13% is exclusief enkele uitschieters; er zijn gevallen die voor een verontreiniging op een bepaalde locatie een detectiekans van meer dan 80% opleveren (in figuur 3-4 is de y-as slechts tot 20% weergegeven).

Het feit dat mpA het beter doet dan de mpC-mediaan is niet toevallig. Bij mpA zijn de monsterlocaties verspreid over het leidingnet, terwijl het bij mpC ook voorkomt dat locaties toevallig dicht bij elkaar liggen en daardoor een minder effectief monsterprogramma opleveren.

3.2.2 Monsternamen in het leidingnet – willekeurig over de werkdag

Dezelfde analyse als in §3.2.1 kan gedaan worden met een monsternamen die niet willekeurig over 24 uur plaatsvindt, maar willekeurig tussen 8.00 's ochtends en 17.00 's avonds. De resultaten staan in figuur 3-5.

Uit de vergelijking van figuur 3-4 en figuur 3-5 blijkt dat door alleen overdag te meten de detectiekans vrijwel gelijk blijft. Voor mpC-95% is de gemiddelde detectiekans met 7,4% iets kleiner dan wanneer willekeurige over 24 uur wordt gemeten.

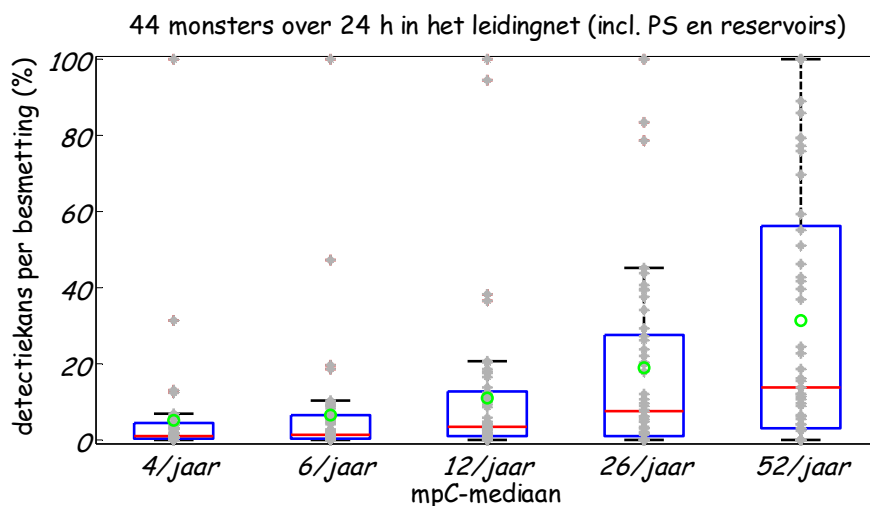


figuur 3-5 Detectiekans van besmetting op 50 verschillende locaties bij verschillende monsternamenprogramma's (mpA: 2002, mpB: 2007 en mpC: Monte Carlo) gedurende een werkdag.

Door alleen overdag te meten blijft de detectiekans vrijwel gelijk. Het is daarom aannemelijk dat de starttijd van de verontreiniging van beperkte invloed is op de detectiekans. De tijd tot detectie zal wel samenhangen met de starttijd van de verontreiniging. Voor een waterbedrijf is het prettig dat monsternamen tijdens kantooruren de detectiekans niet negatief beïnvloed. N.B. in deze analyse is de dag van de week niet meegenomen; bij een verontreiniging die vlak voor het weekend start en de mogelijke monsternamen pas na het weekend plaatsvindt kan de detectiekans wel negatief beïnvloed worden door alleen tijdens werkdagen te bemonsteren. Hierdoor kan de gemiddelde detectiekans potentieel met 2/7 (bijna 30%) afnemen.

3.2.3 Monsternamen in het leidingnet – opvoeren van de monsterfrequentie

Een andere mogelijkheid om de detectiekans te verhogen is niet op de meest optimale locaties te meten, maar vaker te gaan meten. Daartoe is in figuur 3-6 de monsterfrequentie van mpC-mediaan in het distributienet opgevoerd. Dat wil zeggen dat de monsterfrequentie op het pompstation 1 maal per week blijft, in de reservoirs blijft de frequentie 1 maal per maand en de frequentie in het distributienet gaat van 4 maal per jaar (eens per 3 maanden) op 44 locaties naar steeds hogere frequenties: eens per 2 maanden, eens per maand, eens per 2 weken en eens per week (in de figuur aangeduid als 4/jaar, 6/jaar, 12/jaar, 26/jaar en 52/jaar).



figuur 3-6 Detectiekans van besmetting op 50 verschillende locaties bij monsternamenprogramma C (Monte Carlo methode) gedurende een werkdag met verschillende monsternamenfrequenties op 44 locaties in het leidingnet.

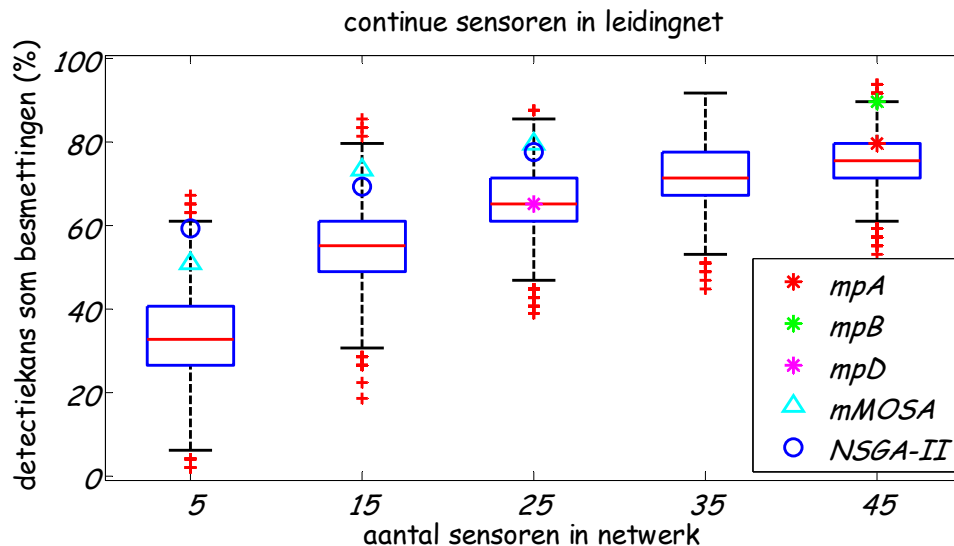
Met de toename van de monsterfrequentie neemt ook de detectiekans toe. Bij een monsterfrequentie van een maal per week is de gemiddelde detectiekans toegenomen tot 31,3%. Wanneer de relatie tussen monsterfrequentie en detectiekans wordt geëxtrapoleerd blijkt dat voor een detectiekans van gemiddeld 40% er ca. 75 maal per jaar op 44 locaties monsters genomen zouden moeten worden.

3.2.4 Continue sensoren in het leidingnet

In de analyse van de simulatie van §3.2.1 betekent een continue monitoring dat eens per uur wordt gekeken of er sprake is van een verontreiniging. Met de combinatie van meerdere sensoren neemt de detectiekans toe. Een Monte-Carlo-simulatie is gedaan met 5, 15, 25, 35 en 45 sensoren willekeurig in het leidingnet geplaatst. Voor iedere combinatie van sensoren is bepaald welk percentage van de 50 verontreinigingen is gedetecteerd (dit wijkt af van de monsternamen, waar de detectiekans van telkens een willekeurige andere set locaties is geanalyseerd). De resultaten zijn in een box-whisker-plot weergegeven (figuur 3-7).

De monsternaming kan worden vergeleken met de continue meting (figuur 3-5 en figuur 3-7). Bij 5 sensoren is reeds een duidelijke verbetering in detectiekans te zien. Met 5 sensoren op de locaties van NSGA-II wordt ca. 60% van alle verontreinigingen gedetecteerd; dit is een aanzienlijke verbetering ten opzichte van de gemiddelde detectiekans van 6,3% met monsternamen volgens mpB. Een detectiekans van 100% zal niet gemakkelijk te realiseren zijn.

Ook de detectiekans van continue sensoren op de locaties van mpA en mpB is weergegeven in figuur 3-7: mpA lijkt een redelijke keuze (75-percentiel van mpC), mpB is een betere keuze. Het poortwachtersmeetprogramma (mpD) geeft een detectiekans gelijk aan de mediaan van de willekeurige locaties; voor de 50 verontreinigingslocaties zijn dus betere sensorlocaties te kiezen. De meetlocaties bepaald met mMOSA en NSGA-II doen het telkens veel beter dan het 75-percentiel van de willekeurige meetlocatiesets.



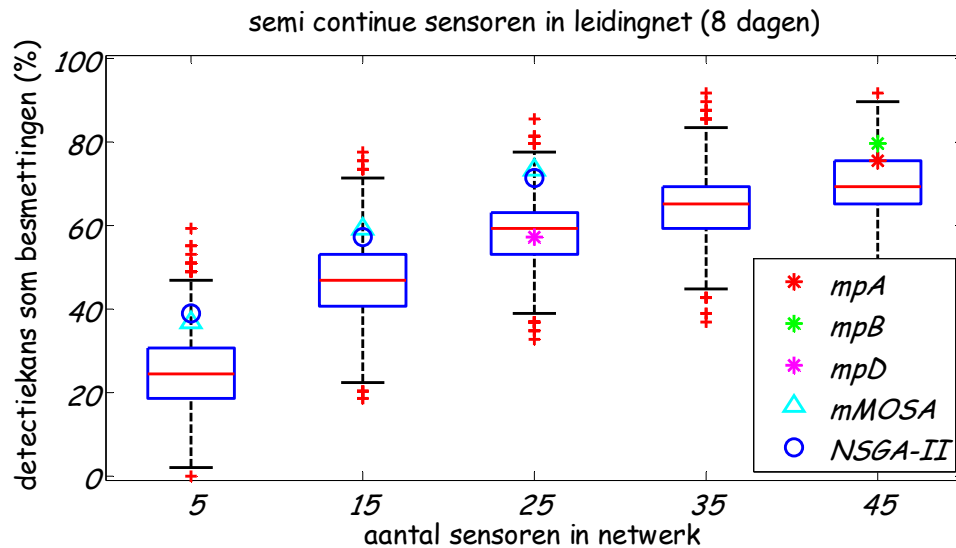
figuur 3-7 Detectiekans van 50 verontreinigingen met verschillende aantallen sensoren en sensorlocaties (continue monitoring), bepaald met Monte Carlo methode en mpA (2002), mpB (2007), mpD (poortwachter), mMOSA en NSGA-II (Exeter).

3.2.5 Semi-continue sensoren in het leidingnet

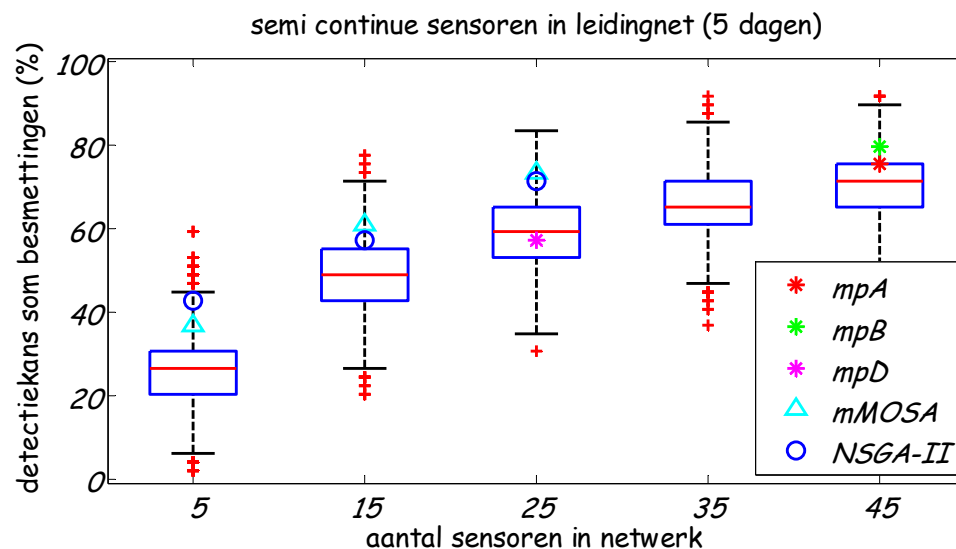
In de analyse van de simulatie van §3.2.1 betekent een semi-continue monitoring dat eens per 8 (of 5) dagen wordt gekeken of er sprake is van een verontreiniging in een mengmonster is samengesteld uit uurmonsters. Het mengmonster wordt in de analyse gesommeerd over de 8*24 (of 5*24) uurmonsters en gedeeld door het totaal aantal monsters, zodat een gemiddelde concentratie ontstaat. Met de combinatie van meerdere sensoren neemt de detectiekans toe. Een Monte-Carlo-simulatie is gedaan met 5, 15, 25, 35 en 45 sensoren willekeurig in het leidingnet geplaatst. Het resultaat staat in figuur 3-8 en figuur 3-9 voor resp. 8 en 5 dagen.

Het onderscheid tussen mengmonster van 5 en 8 dagen is minimaal. Dezelfde uitspraken als voor de continue monitoring gelden:

- Bij 5 sensors is de verbetering ten opzichte van monsternamen reeds significant.
- mpA lijkt is (toevallig) een redelijk goede keuze, mpB is nog beter.
- mpD geeft een detectiekans gelijk aan de mediaan van de willekeurige locaties; voor de 50 verontreinigingslocaties zijn dus betere sensorlocaties te kiezen.
- mMOSA en NSGA-II geven zeer goed de beste sensorlocaties aan.



figuur 3-8 Detectiekans van 50 verontreinigingen met verschillende aantallen sensoren en sensorlocaties (semi-continue monitoring per 8 dagen), bepaald met Monte Carlo methode en mpA (2002), mpB (2007), mpD (poortwachter), mMOSA en NSGA-II (Exeter).



figuur 3-9 Detectiekans van 50 verontreinigingen met verschillende aantallen sensoren en sensorlocaties (semi-continue monitoring per 5 dagen), bepaald met Monte Carlo methode en mpA (2002), mpB (2007), mpD (poortwachter), mMOSA en NSGA-II (Exeter).

4 Discussie en conclusies

In 2004 zijn berekeningen uitgevoerd om vast te stellen wat de detectiekans is van E.coli in het distributienet met het huidige monsternamprogramma. De conclusie was toen dat de detectiekans erg laag was (rond de 5%). Een aantal voorstellen werd gedaan om monsternamprogramma te optimaliseren zoals de meetfrequentie op pompstation en in reservoirs te verhogen, de monsternam te combineren met klachten en storingen en een betere verdeling van de monsterlocaties over het leidingnet.

4.1 Detectiekans met monsterprogramma

Doel van voorliggend onderzoek was het vergroten van de detectiekans van E.coli in het distributienet. De eerste vraag die daarbij van belang is, is wat het doel van het meetprogramma is. Bijvoorbeeld

- “Voldoen aan de VROM-inspectienorm”;
- “Detecteren van verontreinigingen (meten van prestatie)”;
- “Detecteren van verontreinigingen en erger voorkomen (waarbij bijvoorbeeld geoptimaliseerd kan worden op het beperken van het verontreinigde geconsumeerde volume of de hoeveelheid geïnfecteerde personen)”.

Met het huidige meetprogramma voldoen de waterbedrijven reeds aan de (huidige) VROM-inspectienorm. Het detecteren en passende actie ondernemen lijkt op dit moment nog een stap te ver.

Nader gespecificeerd is het doel van het huidige onderzoek dus om met een verbeterd meetprogramma de prestatie beter vast te kunnen stellen. Concreet betekent dit dat

- ofwel alle verontreinigingen gedetecteerd moeten worden
- ofwel dat goed bekend is wat de detectiekans is zodat voor misdetectie gecorrigeerd kan worden.

Uit de uitgebreide analyse van de detectiekans met een standaard monsterprogramma blijkt dat op z'n best er een kans van 7,7 % is dat een verontreiniging wordt gedetecteerd. De gemiddelde detectiekans is 5,0%. Door een monsterprogramma te kiezen op basis van locaties met relatief lange verblijftijden kan de gemiddelde detectiekans worden verhoogd tot 6,3%. Het detecteren van alle verontreinigingen met behulp van een monsterprogramma van de huidige omvang en frequentie is derhalve onmogelijk. Uit de analyse blijkt verder dat de detectiekans voor iedere verontreiniging zeer variabel is. Bijvoorbeeld bij het geoptimaliseerde meetprogramma – mpB – is de detectiekans gemiddeld 6,3%, minimaal 0% en maximaal 12 % met enige uitschieters naar meer dan 80% en de mediaan rond de 2%. Met een klein aantal positieve monsters per jaar betekent deze variatie in de detectiekans en de afhankelijkheid van verontreinigingsconcentratie en –duur in combinatie met de detectielimiet dat ook correctie voor misdetectie lastig is.

4.2 Detectiekans met verzaamd monsterprogramma

Door de monsterfrequentie op te voeren kan de detectiekans worden verhoogd. Een driemaal zo hoge monsterfrequentie leidt bijvoorbeeld tot een toename van de detectiekans met 100% (van 5,0% gemiddeld naar 10,9% gemiddeld).

4.3 Detectiekans met sensoren

Het punt “waar neem je de monsters/plaats je sensoren?” staat ook internationaal erg in de belangstelling. Sinds 11 september 2001 wordt namelijk steeds meer gekeken naar de beveiliging (security) en bewaking (safety) van de waterkwaliteit in het leidingnet. Er wordt door verschillende onderzoeksgroepen gerekend aan een optimale plaatsing van sensoren (continue bemonstering). In augustus 2006 vond er zelfs een heuse *battle of the water sensor networks* plaats. Overigens zonder onovertroffen winnaar. Verschillende onderzoeksinstituten buigen zich over (de geldigheid van de aannames in) de modellering zoals

- wat zijn de beste optimalisatieparameters, zoals
- aantal sensoren / kostprijs;
- tijd tot detectie;

- het aantal consumenten dat verontreinigd water drinkt;
- het volume geconsumeerd verontreinigd water;
- etc.
- hoe rekening te houden met reactietijd (van detectie tot kookadvies of het spuien van het net);
- hoe rekening te houden met het feit dat sensoren niet perfect zijn;
- wat is de invloed van het niet perfect kennen van de volumestromen in het netwerk (dus de afnamepatronen);
- wat is de kans op een (opzettelijke) verontreiniging op verschillende locaties in het net?
- kan het leidingnetmodel de waterkwaliteit goed genoeg modelleren (zoals menging op kruispunten, dispersie en interactie met het leidingmateriaal) of moet het model worden aangepast?

In voorliggend onderzoek zijn twee optimalisatietechnieken voor het bepalen van de beste sensorlocaties nader onderzocht, namelijk NSGA-II (Non dominated Sorting Genetic Algorithm, extension) en mMOSA (modified Multi-Objective Simulated Annealing). Uit de literatuur blijkt dat deze beide methoden kandidaat zijn voor de meest geschikte technieken, maar er zijn er meer. In dit onderzoek is geoptimaliseerd op ten eerste de kans op de detectie en ten tweede het beperken van het volume geconsumeerd verontreinigd water (bij de veronderstelling dat na detectie geen water meer wordt gedronken). Ten derde is ook gekeken naar het aantal sensoren zodat een afweging gemaakt kan worden tussen investering en detectie(kans).

Door te kiezen voor een (semi)continue monitoring met behulp van sensoren kan de detectiekans aanzienlijk worden verhoogd. Door sensorlocaties te bepalen met behulp van een optimalisatietechniek op basis van genetische algoritmen (zoals NSGA-II en mMOSA) kan reeds met 5 sensoren een detectiekans van ca. 40% (bij semi-continue monitoring) tot 60% (bij continue monitoring) worden behaald. Met sensoren komt “het detecteren en passende actie ondernemen” in zicht. Dit is vooralsnog theorie, want echt gevoelige sensoren voor *E. coli* zijn nog niet op de markt.

De semi-continue monitoring op 5 locaties geeft een detectiekans van 40% bij een analyse van 5 locaties * 1 mengmonster per week = ca. 250 monsters per jaar. Om met verhoogde monsternamen aan een gemiddelde detectiekans van 40% te halen zijn ca. $44 \times 75 = 3300$ monsters per jaar nodig. Een afweging tussen kosten van monsternamen en het doen van semi-continue metingen en de detectiekans van beide methoden kan gunstig uitpakken voor de semi-continue monitoring wanneer deze beschikbaar komt.

N.B. de hier genoemde cijfers zijn van toepassing op de doorgerekende situatie, d.w.z. het leidingnet van Almelo met een verontreiniging van 16 uur, startende om 0:00 's nachts en bij een gegeven detectiekans.

5 Referenties

- Berry, J. W., Hart, W. E., Philips, C. A., Uber, J. G., and Walski, T. M., "Water Quality Sensor Placement in Water Networks With Budget Constraints" (2005a). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 60.
- Berry, J. W., Hart, W. E., Philips, C. A., Uber, J. G., and Watson, J. P., "Validation and Assessment of Integer Programming Sensor Placement Models" (2005b). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 55.
- Berry, J. W., Hart, W. E., Philips, C. A., and Watson, J. P., "Scalability of Integer Programming Computations for Sensor Placement in Water Networks" (2005c). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 281.
- Berry, J. W., Hart, W. E., Philips, C. A., Uber, J. G., and Watson, J. P. (2006), "Sensor Placement in Municipal Water Networks with Temporal Integer Programming Models", *Journal of water resources planning and management*, 218-224.
- Boccelli, D.L. and W.E. Hart (2007), "Optimal Monitoring Location Selection for Water Quality Issues", World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat.
- Chastain, J. R., "A Heuristic Methodology for Locating Monitoring Stations to Detect Contamination in Water Distribution Systems" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 22.
- Grayman, W. M., Ostfeld, A., and Salomons, E., "Red Team-Blue Team Exercise for Locating Monitors in Distribution Systems" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 42.
- Haas, C. N. (1993), "Microbial sampling: is it better to sample many times or use large samples?", *Water science and technology*, 27(3-4), 19-25.
- Janke, R., Murray, R., Uber, J., and Allgeier, S., "An Evaluation of System Architectures for Contamination Warning Systems" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 18.
- Lee, B. H. and Deininger, R. A. (1992), "Optimal locations of monitorin stations in water distribution system", *Journal of Environmental Engineering*, 118(1), 4-16.
- Lieverloo, J. H. M. van, Mesman, G. A. M., Baggelaar, P. K., Hamed, A., and Bakker, G. L. (2004). *Detectiekans van fecale verontreiniging in drinkwaterdistributiesystemen*, Kiwa N.V., Nieuwegein, BTO 2004.063.
- McKenna, S. A., van Bloemen Waanders, B., Laird, C. D., Buchberger, S. G., Li, Z., and Janke, R., "Source Location Inversion and the Effect of Stochastically Varying Demand" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 47.
- Propato, M., Piller, O., and Uber, J. G., "A Sensor Location Model to Detect Contaminations in Water Distribution Networks" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 45.
- Speight, V. L., Kalsbeek, W. D., and DiGiano, F. A. (2004), "Randomized stratified sampling methodology for water quality in distribution systems", *Journal of water resources planning and management*, 330-338.
- Tidwell, V. C., Silva, C. J., and Jurado, S., "An Integrated Approach to Vulnerability Assessment" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 29.
- Uber, J. G., "Identifiability of Contaminant Source Characteristics in Steady-State and Time-Varying Network Flows" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005).

Watson, J. P., Hart, W. E., and Berry, J. W., "Scalable High-Performance Heuristics for Sensor Placement in Water Distribution Networks" (2005). *Impacts of Global Climate Change; 2005 World water and environmental resources congress*, (American Society of Civil Engineers, Reston, VA, USA, 2005), 282.

BIJLAGE 1 Paper Exeter

Almelo Project - Report #2

Martin Weickgenannt, Zoran Kapelan, and Dragan Savic

July 19, 2007

Abstract

This report presents a method of optimising sensor locations to detect contaminations in a water distribution network as quickly as possible. The optimisation task is formulated as two-objective problem, taking (1) the number of sensors and (2) the risk of contamination into account. The latter is a new formulation of risk which incorporates the likelihood that a set of sensors cannot detect a scenario and the amount of polluted water that is consumed prior to detection. The task of scenario selection is tackled by an Importance-based Sampling Method, which is presented here for the first time. The solution of the optimisation problem is conducted by two optimisation algorithms, namely *NSGA-II* and a modified version of *MOSA*. A case study is carried out for the water distribution system of the Dutch city Almelo. The results indicate that both algorithms show similar performance. The estimated Pareto front implies that no more than 10 sensors are used to provide reasonable contamination protection.

1 Introduction

Water Distribution Networks (WDN) are vital parts of our cities. Due to their size and complexity it is obvious that water contaminations can occur at any time, be it through attacks or accidents. Therefore, measures have to be taken to detect contaminations in WDS as fast as possible to prevent the population from being affected. But a trade-off has to be made between safety (which can never be total) and the costs to provide this safety. That means that not every junction within a WDS can be equipped with a contamination sensor. Therefore, sensor locations within a WDS have to be chosen carefully.

In recent years there has been a number of publications dealing with the problem of optimal contamination sensor location with respect to various objective functions. Publication [1] uses a single-objective formulation as a basis for sensor location optimisation. The aim consists in finding a set of sensors which can provide a given *level of service*. The level of service is defined as the amount of polluted water which is allowed to be consumed prior to the detection of the contamination. The problem is addressed by creating a pollution matrix which indicates if a node can detect the contamination before the given level of service is reached. The optimisation problem is solved by finding the minimum number of sensors needed to provide a maximum coverage of the pollution matrix (i.e. finding a minimal set of sensors which can detect all possible contamination events before the level of service is reached). The solution is reached by formulating and solving a set covering problem. [2] uses a similar approach and solves the maximum coverage problem by a Genetic Algorithm. The results of [1] are discussed in [3].

Being a contestant at the *Battle of the Water Sensor Networks 2006* (BWSN), [4] proposes an optimisation algorithm using a two-stage *Noisy Cross-Entropy Sensor Locator algorithm* (nCESL), which uses the cross-entropy method. The optimisation task is formulated as a four-objective problem, including the detection possibility, various risk measurements and the time to detection. The optimisation is carried out for 5 and 20 sensors and renders very good results compared to other contestants in the battle. The results of another BWSN contestant are published in [5] which also formulates the optimisation task as a four-objective problem and obtains optimal sensor locations by using the *NSGA-II* algorithm.

In [6], the optimal sensor placement under parameter uncertainties is examined. The uncertainties

are not only the place of the attack (i.e. the contamination node), but also the demand of water at the time of the attack. By changing the demands, flow directions can change. Thus, initially unaffected nodes can be contaminated if the demands change slightly. The authors of that paper consider demand changes of $\pm 25\%$. The objective function for optimisation stated in [7] is modified to consider different demand pattern scenarios. The sum of objective functions (i.e., risk of contamination) over all samples (samples differ in demand pattern) is to be minimised. For each sample, a hydraulic and a quality analysis has to be performed because the hydraulic conditions change. This re-solving of the hydraulic model requires a huge amount of computation time, so the proposed approach is only suitable for rather small networks. As both the hydraulic and quality analysis have to be performed for every demand pattern sample, the solution of the optimisation problem is only suitable for relatively small networks.

[8] also addresses the effect of various kinds of uncertainties concerning the optimal sensor placement in water distribution systems (WDS). The problem is tackled by formulating objective functions which are robust against uncertainties as mixed-integer programs. The resulting objective functions depend on uncertainty assumptions. The robustness of the proposed approaches is shown in theory.

In [9], an interesting topic is raised, namely the impact of a sensor detection limit on the number of sensors needed for contamination detection. The problem is formulated as a two-objective optimisation task, with the number of sensors and the number of detected scenarios as objectives. Subsequent optimisation runs for various sensor detection limits are performed. The results show that if the detection limit is lower than 10% of the average node concentration, the effect of lowering the limit on detection possibility is rather small.

The optimal sensor location task is formulated as a mixed-integer program in [7] and [10]. The single objective is the fraction of the population being affected by the contamination. It considers the density of the population at each node and the node's attack probability. A case study is conducted by using real world distribution networks. [10] solves the MIP formulation and carries out a sensitivity analysis based on changes in the nodal demands.

Here, the optimisation task is formulated as a two-objective problem. The two conflicting objectives are *number of sensors* and *risk of contamination*. The risk of contamination is formulated in terms of detection likelihood and demand-based contamination and it is assumed that the sensors are not optimal, i.e. they have a lower detection limit. The hydraulic analysis as well as the quality analysis is performed by using the simulation software **ALEID**, which is based on **EPANET**, Version 1.0. Subsequent runs of quality analyses, each representing a different contamination scenario, are carried out by using **Matlab** and a *Dynamic Link Library* (DLL) for **ALEID**, which was developed for this project. The samples for the quality analysis (i.e. time and node of intrusion) are chosen via the newly introduced *Importance-based Sampling Method*. The solution of the optimisation task is obtained independently by two optimisation algorithms: *Non-Dominated Sorted Genetic Algorithm-II* (*NSGA-II*) and *Modified Multi-objective Simulated Annealing* (*mMOSA*). The two solution methods are compared with respect to objectives, sensor locations, and time needed for optimisation.

The aim of this project is to optimise sensor locations for the Dutch city *Almelo*, which is located in the east of the Netherlands.

This report is organised as follows: In Section 1, the problem formulation is stated as a two-objective optimisation task. An efficient way of calculating the objectives in **Matlab** during the optimisation is presented. Section 2 presents the *Almelo* water distribution system (WDS). In Section 3, the term *scenario* is defined and the *Importance-based Sampling Method* for scenario selection is presented. The topic of multiple quality analysis runs, using an **ALEID** DLL, is addressed in Section 4. Sections 6 and 7 present the two optimisation algorithms used within the scope of this project. Finally, in Section 8 optimisation results for *NSGA-II* and *mMOSA* are presented and compared with each other. sectionOptimisation task and problem formulation The task of this paper is to find a set of sensor locations to detect a contamination occurring in a water distribution system as quickly as possible. Having a network with N nodes and a maximum number of S sensors, the following two objectives are to be minimised:

- number of sensors, and
- risk of contamination.

As the two objectives clearly compete against another, the optimisation solution will consist of a set of optimal compromise solutions, the so-called *Pareto Front* [11]. Mathematically, the optimisation problem is formulated as follows:

$$\text{Minimise } f(\underline{S}) = \begin{pmatrix} S \\ R(\underline{S}) \end{pmatrix},$$

where $\underline{S}$ is a set (or vector) of sensor locations, S is the number of elements in $\underline{S}$, and $R(\underline{S})$ is the contamination risk.

The risk of contamination is defined as follows.

Definition The risk of contamination $R(\underline{S})$ is the product of

- the likelihood that a set of sensors cannot detect a contamination, and
- the average amount of polluted water consumed prior to detection.

In a mathematical formulation, this can be written as:

$$R(\underline{S}) = \left(1 - \frac{1}{m} \sum_{l=1}^{l=m} \max_{\underline{S}_p \in \underline{S}} \xi(\underline{S}_p, K_l) \right) \cdot \left(\frac{1}{k} \sum_{l=1}^{l=m} \min_{\underline{S}_p \in \underline{S}} \chi(\underline{S}_p, K_l) \right),$$

where $\underline{S}$ is a vector of sensor locations, $\underline{S}_p$ is one element of this vector (i. e. a node having a sensor attached), m is the number of contamination scenarios (see Section 3), and K_l is the l^{th} element of the scenario set $\mathbb{K}$.

The detection function $\xi(\underline{S}_p, K_l)$ is a binary function, determining whether scenario K_l can be detected by the sensor $\underline{S}_p$ ($\xi = 1$), or not ($\xi = 0$). The function $\chi(\underline{S}_p, K_l)$ is the amount of polluted water being consumed prior to contamination detection at Sensor $\underline{S}_p$.

Detection function $\xi(\cdot)$

The calculation of the detection function is based on the following definition of the term *detection*.

Definition A scenario $K(N_i, T_j)$ is detected by a sensor attached to node N_l , iff the maximal nodal concentration $C(t)_m$ of the contaminant E exceeds a threshold value C_t within the time $[T_0, T]$.

Therefore, the detection function $\xi(\cdot)$ is calculated by examining the concentration maximum at the corresponding node.

Contamination function $\chi(\cdot)$

The amount D_c of polluted water consumed in the network prior to detection at node N_l is calculated as

$$D_c(N_l) = \int_{t_0}^{t_{det}} \left(\sum_{i=1}^{i=N} (D(t, K_i) \cdot h(C(t, N_i), C_t)) \right) dt,$$

Component	Number of elements in the network
Junctions	1758
Pipes	2307
Reservoirs and Tanks	4
Pumps	2
Valves	6

Table 1: Components in the Almelo WDS network

where N_l is the node for which the contamination function is calculated, t_{det} is the time to detection at that node, $C(t, N_i)$ is the concentration at node N_i , and $D(t, N_i)$ is the water demand at node N_i at time t . The function $h(C(t, N_i), C_t)$ is the Heaviside Step function and is here defined as:

$$h(C(t, N_i), C_t) = \begin{cases} 1 & \text{if } C(t, N_i) > C_t \\ 0 & \text{otherwise} \end{cases}$$

In the scope of this report the sensor location optimisation is carried out using two different algorithms: *NSGA-II* and *mMOSA* (based on *MOSA*). These algorithms are particularly suitable for the solution of large-scale multi-objective optimisation tasks, such as the sensor location optimisation. The algorithms belong to the field of combinatorial optimisation which in turn is a branch of the field of global optimisation.

2 Almelo water distribution network

In the scope of this project the sensor location optimisation is carried out for the water distribution network of the Dutch city Almelo which is operated by the company *Vitens* (www.vitens.nl). Almelo is located in the east of the Netherlands. Almelo's water distribution network is middle-sized, the number of components is as in Table 1. Figure 1 shows the water distribution network of the city Almelo as it is modelled in the hydraulic simulation software *ALEID*. The figure shows that the Almelo WDN consists of three major parts, the one in the upper-right, the one in the lower-left, and the central part. The results of the sensor location optimisation (Section 8) show that sensors should be placed in each of these three areas to keep the risk of contamination for the population low.

The colour coding shows the elevation of each junction. The lowest junction is at sea level, the highest is about 14m above sea level. The water source for the network is located on the left side where the pump is located. Through this pump water from a resource enters the network. Pipes with large diameters then conduct the water into the city. This can be seen in Figure 2 where the WDS is shown. Thicker pipe lines indicate larger pipe diameters. One can see that the main parts of the networks consist of a network of many pipes small in diameter which are connected to the source by only a few pipes having large diameters. The water demand is also concentrated in the three areas of the network. A particular high water demand is located at the node which is on the middle-right. Various demand patterns are defined to represent the time-varying demands. Patterns are defined for water demand of the population (having maxima at 10 AM and 7 PM) as well as for industrial use (having maxima at 10 AM and high demand all day). The water consumption is also split up into various categories. These categories present a possibility to

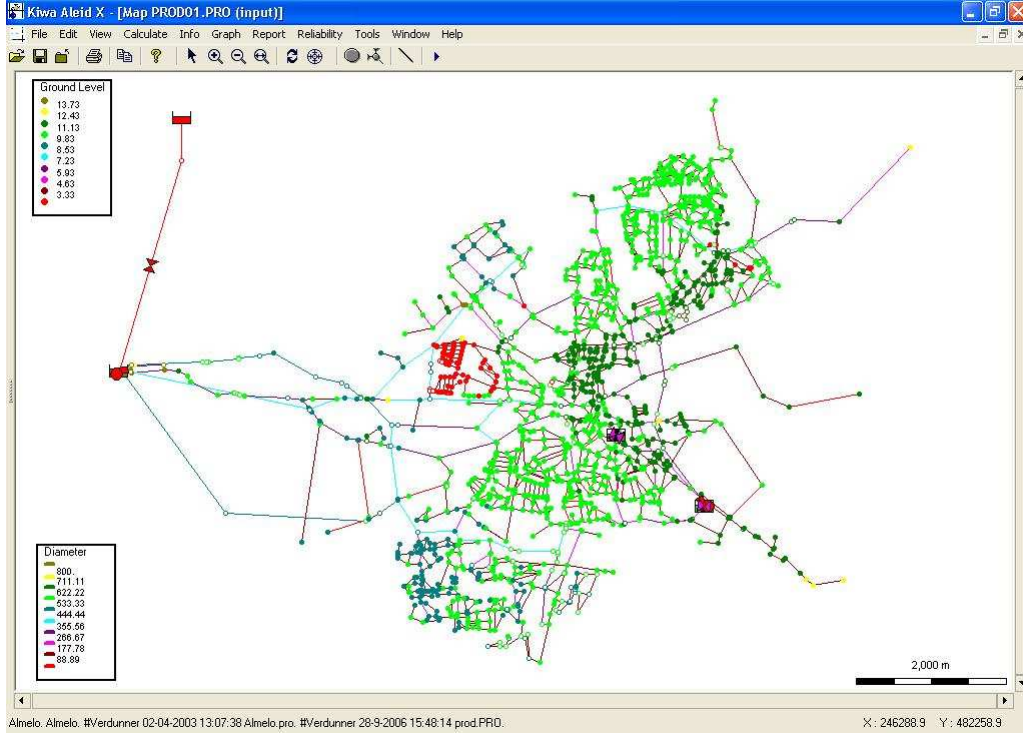


Figure 1: Almelo water distribution system

perform the sensor location optimisation specifically to protect a certain consumer category, such as population or industry. Here, these categories have not been taken into account because the aim was to protect all elements of the water distribution network, not only the population or industry.

There are also controls attached to the network model. These are used to control the flow within the network and to ensure that a tank located in the lower-right of the network does not exceed its maximal water level value.

3 Scenario definition and sampling design

3.1 Scenario definition

For the calculation of the objective functions described in Section 1 detailed information about the propagation of a contaminant in the network has to be obtained. Here, the water quality analysis is performed by using the hydraulic simulation program **ALEID** which is based on **EPANET** (Version 1.0). An intrusion at node N_i is modelled as a constant concentration C_i of a chemical at that node.

As important parameters of the network are time-varying (e.g. Demands), the propagation of a chemical in the network does not only depend on the intrusion node but also on the intrusion time. Therefore, a contamination scenario (or event) can be defined as follows.

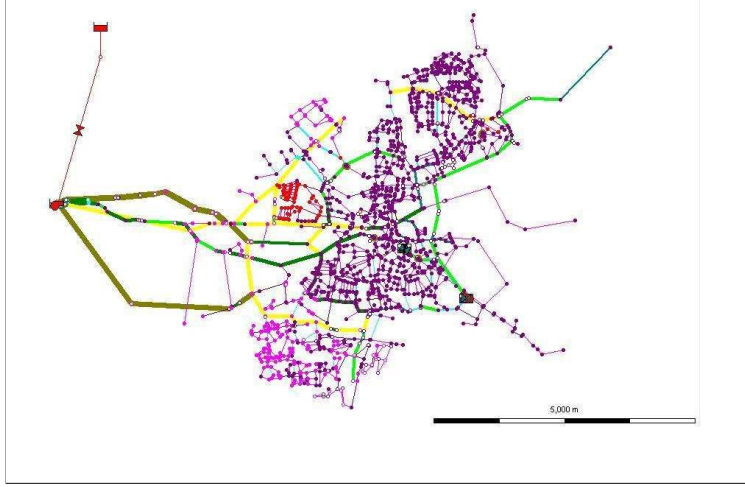


Figure 2: Pipe diameters in the Almelo WDS

Definition A contamination scenario $K(N_i, t_j)$ is the combination of time t_j and place N_i where a contaminant E enters a water distribution network. N_i $i = 1, \dots, N$ is the intrusion node and $t_j \in [T_{i,0}, T_{i,end}]$ is the time of intrusion. N is the number of junctions in the network, $T_{i,0}$ and $T_{i,end}$ are the earliest and latest possible time of intrusion, respectively.

Remark Other parameters, such as

- duration of intrusion
- contaminant E
- amount of contaminant E entering the network

can also be seen as part of a contamination scenario. In this work we abstain from doing this because these parameters are the same for all contamination events.

Assuming that a contamination event can occur every 10 minutes within 24 hours in the *Almelo* water distribution network (1758 junctions), the number of possible scenarios is

$$\|K\| = 288 \cdot 1758 = 253,152.$$

A water quality analysis of all possible scenarios would take almost 6 days and would produce more than 12GB of data for the calculation of objective matrices. This amount of time and data is not feasible for an optimisation. Therefore, a set of sample scenarios have to be chosen for which a quality analysis is performed.

3.2 Scenario selection by an importance-based sampling method

If the time of intrusion is discretised with a step width of ΔT_i the set $\mathbb{K}$ of all possible contamination scenarios can be represented by a two-dimensional matrix:

$$\mathbf{K} = \begin{pmatrix} (N_1, t_0) & (N_1, t_0 + \Delta t) & \dots & (N_1, t_J) \\ \vdots & \vdots & \vdots & \vdots \\ (N_i, t_0) & (N_i, t_0 + \Delta t) & \dots & (N_i, t_J) \\ \vdots & \vdots & \vdots & \vdots \\ (N_N, t_0) & (N_N, t_0 + \Delta t) & \dots & (N_N, t_J) \end{pmatrix},$$

with N_i , $i = 1, \dots, N$ and $t_j = t_0 + (j - 1)\Delta t$, $j = 0, \dots, J$, $J = \frac{t_{end} - t_0}{\Delta t}$. A weight I_{ij} can be assigned to each scenario K_{ij} representing the importance of the event. The importance of a contamination event is defined as the amount of water which is contaminated by this event within a time period ΔT . After assigning importance values to all contamination events, the importance matrix $\mathbf{I}$ can be written as

$$\mathbf{I} = \begin{pmatrix} I(1, 1) & \dots & I(1, J) \\ \vdots & \ddots & \vdots \\ I(N, 1) & \dots & I(N, J) \end{pmatrix}.$$

From this matrix a normalised importance vector $\overline{Iv}$ can be calculated by writing the columns of the matrix one below another and normalising the resulting vector. From this vector a set of M contaminations can be chosen randomly. The selection likelihood for every scenario is given by its normalised importance value. For the selection algorithm, see Procedure *Importance-based scenario selection*.

```

Set found = 0;
while found < M do
    Create random number  $r \in [0, 1]$ ;
    Search j, such that:
        
$$\sum_{i=1}^{j-1} < r \leq \sum_{i=1}^j$$

        if j not yet chosen then
            found=found+1;
            Mark j as chosen
        else
            _

```

Procedure Importance-based scenario selection

Figure 3 shows the importance matrix $\mathbf{I}$ of the *Almelo* WDS. The red crosses show the selected scenarios for $M = 1000$.

4 Water Quality Analysis using ALEID

After selecting the scenarios, a quality analysis is performed. For a list of parameters see the Appendix C. As a large number of scenarios has to be analysed to get accurate optimisation results, the process of quality analysis is very time consuming. Therefore, the quality analysis of every scenario has to be performed in minimal time. To achieve this it is desirable to be able to run multiple quality runs without having to repeat the hydraulic analysis. This is supported by the software ALEID. Another way of decreasing the simulation time for a quality analysis run is

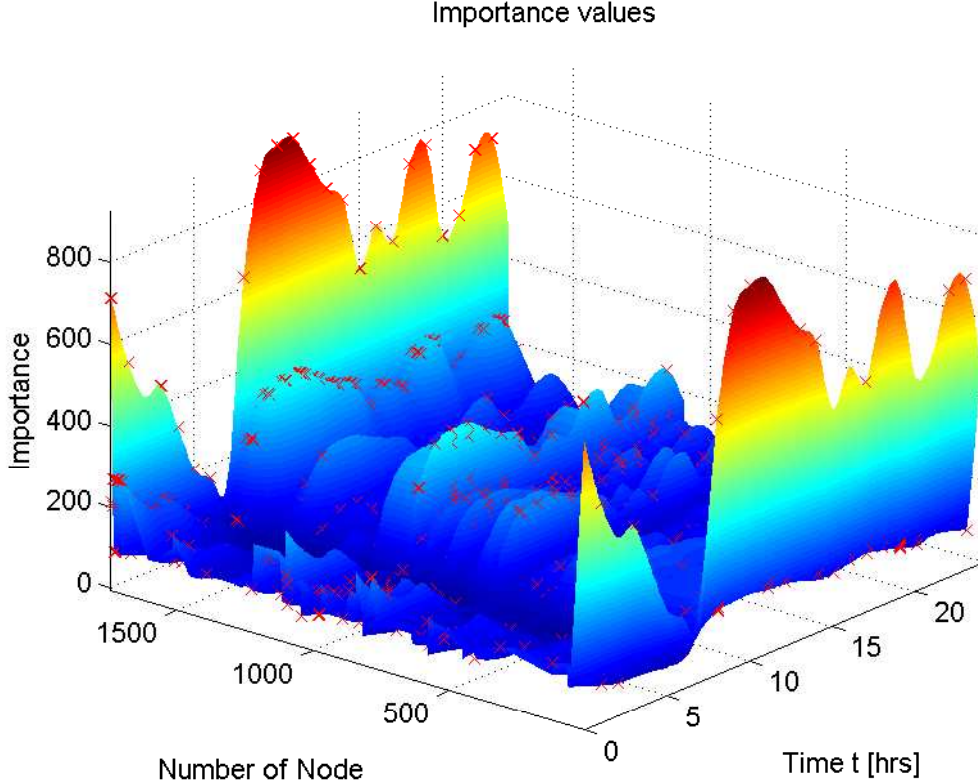


Figure 3: Importance matrix $\mathbf{I}$ of the Almelo WDN

the use of *Dynamic Link Libraries*. These *DLLs* can be loaded into the computer memory and, thus, the analysis functions can be accessed much faster than with a static program which would have to be called as a console application. The process of setting the parameters for a scenario, running the quality analysis and analysing the results takes about 1.5 seconds if a *DLL* is used. The use of a console application would increase this time to 4 to 5 seconds.

As the original version of *ALEID* does not come with a *DLL*, one was developed within the scope of this project. The *DLL* has Input/Output functions which allow the user to set arbitrary network and simulation parameters. Therefore, also parameters concerning a contamination scenario can be set very quickly. The quality analysis is performed and the results are stored in a quality file. This file contains only the necessary demand- and concentration information and, thus, can be read and analysed rather quickly.

5 Multi-objective optimisation

In recent years there have been a number of publications dealing with the design of algorithms for multi-objective optimisation using Genetic Algorithms, see [12], [13], and [14]. In [15] a new algorithm, *NSGA-II*, was introduced and it was shown that the algorithm performs exceptionally well on high-order optimisation problems.

The aim of a dominance-based multi-objective optimisation algorithm is to find the so-called Pareto Front F , a set of optimal solutions of the optimisation problem

$$\text{Minimise } y = \underline{f}(\underline{x}) = (f_1(\underline{x}), \dots, f_D(\underline{x})),$$

where D is the number of objectives. As the problem formulation is multi-objective, a mere comparison of the objective values is not sufficient to distinguish if a new solution proposal is better or worse than another one. In fact, the terms "better" and "worse" are - in the context of dominance-based optimisation - replaced by the term domination. A solution f is said to dominate a solution g (in the case of minimisation) if all objective of f are smaller or equal than the objectives of g and at least one objective of f is smaller than the corresponding objective of g . Written $f \prec g$, the mathematical formulation of dominance is: $f \prec g$ iff

$$\begin{aligned} f_i &\leq g_i \quad \forall i = 1, \dots, D \quad \text{and} \\ f_i &< g_i \quad \text{for at least one } i. \end{aligned}$$

If neither $f \prec g$, nor $g \prec f$, the solutions f and g are called *mutually non-dominant*. The Pareto Front F is a set of mutually non-dominant solutions:

$$a \not\prec b \quad \forall a, b \in F$$

That means that all elements of F are optimal compromise solutions of the optimisation problem.

5.1 Design of the decision variables

Each element of the population p consists of two elements: the decision variables $\underline{x}$ and the objective functions $\underline{f}(\underline{x})$. The decision variables form a vector which is split up into two segments of same length:

1. Sensor locations ($L \in [1, N]$), and
2. Sensor switches ($W \in [0, 1]$).

A node N_i is equipped with a sensor iff

$$\begin{aligned} L(j) &= i, \quad 1 \leq j \leq S \text{ and} \\ W(j) &= 1. \end{aligned}$$

The length of the decision vector $\underline{x}$ is $2 \cdot S$, where S is the maximal number of sensors available. A decision vector could look like the following:

$$\underline{x} = [\underbrace{1, 4, 1756, 269, 185, 697, 27, 183}_{\text{Sensor locations}}, \underbrace{1, 0, 0, 1, 1, 0, 1, 1}_{\text{Sensor switches}}]'$$

In this constellation S would be 8, and nodes 1, 269, 185, 27, and 183 would be equipped with a contamination sensor.

6 Optimisation by NSGA-II

This section reviews the *NSGA-II* algorithm which was introduced in [15]. It is an extension of *NSGA* (see [12]) solving various problems of this algorithm, such as high computation time.

The *NSGA-II* algorithm basically uses six steps for optimisation [15].

Creation of an initial population

The first step of the algorithm is the creation of an initial population of solution proposals. The parent population generation is based on a random process which chooses a value for each decision variable for each individual. The initial population serves as starting point of the optimisation. All subsequently generated solution proposals (i. e. individuals) emanate from this parent generation by means of recombination and mutation.

Assigning a fitness value to each individual

The second step marks the first step of the optimisation iteration. After initialising the parent population, each individual has fitness values assigned. The fitness values corresponds with the objectives of the optimisation formulation.

Population sorting based on non-domination

The third step of the *NSGA-II* algorithm is the non-dominated sorting. This step was, being a part of *NSGA*, criticised by many authors for it needed high computational complexity ($O(MN^3)$, with M as the number of objectives and N the population size). The non-dominated sorting aims to assign a rank to each individual of the population which indicates how "good" or "bad" an individual's fitness is compared to the whole population. The rank of an individual is the number of elements in the population which dominate this individual. Thus, it is obvious that a rank of 0 marks the best individual.

After ranking all individuals, they are sorted into several Pareto groups, each group containing elements with the same rank. The elements in the group with the smallest rank are obviously the best estimation $\tilde{F}$ of the Pareto front. The complexity of the non-dominated sorting is reduced for *NSGA-II* namely to a value of $O(M(2N)^2)$ instead of $O(MN^3)$ in *NSGA*.

Sorting each Pareto group according to crowding distance

As the optimisation algorithm aims for finding the best estimation of the Pareto front, the individuals with the highest rank have to be preserved and cannot be exposed to recombination (crossover) and mutation. Therefore, the upper half of individuals (according to the assigned rank) are preserved and not mutated.

Figure 4 (same as in [15]) shows a problem that can occur when omitting the lower half of the individuals: it might not be possible to select exactly half of the population for preservation if only whole groups are taken into consideration. To overcome this difficulty, the elements in each Pareto group (all having the same rank) are sorted according to their *crowding distance*, a measure of how far the objectives of this individual are away from other individuals in the same group (distance in an Euclidean sense). The elements are then sorted according to their crowding distance (in descending order). This ensures a wide coverage of solutions with respect to the objectives.

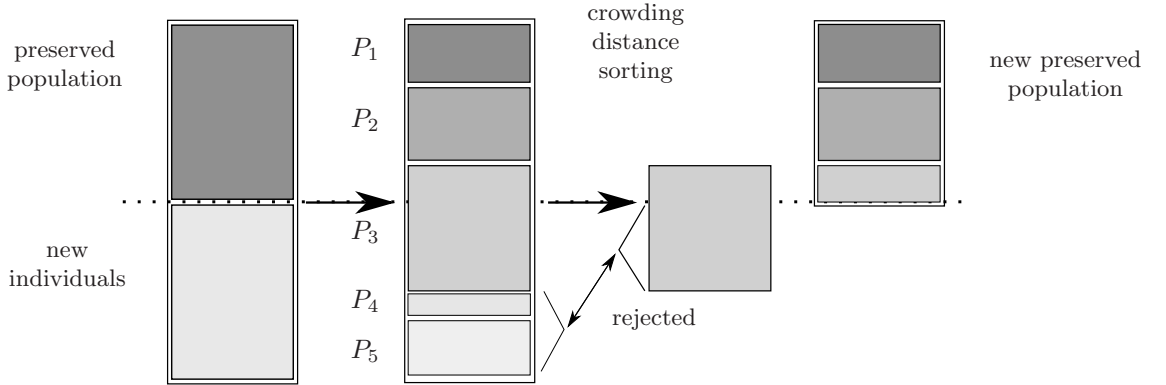


Figure 4: Sketch of the *NSGA-II* main loop (P_i are Pareto groups)

Constraint handling

This step is important if the decision variables have constraints. It searches the parent population for constraint violations. If a chromosome violates a constraint, another non-violating chromosome is preferred, even if its rank or crowding distance has an inferior value. This step helps to ensure

that only feasible solutions are contained within the best Pareto groups. In the scope of this project the binary coded decision variables are not constraint, but the number of bits for each sensor location is chosen such that the corresponding integer value cannot exceed a certain value. By multiplication with a certain value, the value set of the decoded binaries is compressed to be the value set of sensor location indices.

Generation of new individuals

In this step of the optimisation algorithm new solution proposals are generated from the best individuals of the population which have been chosen in the previous step. The reproduction of the best individuals renders an offspring population by means of crossover and mutation. The parent and offspring populations are then combined to form a new population for the next iteration cycle of the optimisation.

After this step is finished, the iteration jumps to the second step, *Assigning a fitness value to each individual*.

Re-assignment of fitness values after all iterations are completed

If any decision variable violated a constraint in the constraint-handling-step, a penalty has to be added to the corresponding objective. This ensures that only feasible solutions are contained in the Pareto front.

Heeding these six optimisation steps, the optimisation algorithm can be outlined as shown in Procedure *NSGA-II*.

```

Initialise parent population;
while  $i < \text{maxIterations}$  do
    Assign fitness to each individual of the population;
    Perform the non-dominated sorting and sort individuals into Pareto groups;
    Sort the elements of each Pareto group according to crowding distance;
    Generate a new population;
    Handle constraints;
    if a stopping criterion is fulfilled then
        | optimisation is finished, break the loop;
    else
        | do next iteration;
Re-assignment of fitness values;

```

Procedure NSGA-II

7 Optimisation by mMOSA

This section introduces the second algorithm which is, within the scope of this project, used to optimise sensor locations within the *Almelo*-WDS with respect to the optimisation formulation in Section 1. We call the algorithm *modified Multi-Objective Simulated Annealing* (mMOSA). The design of the decision variables is the same as for *NSGA-II*, see Section 5.1 for the definition.

The algorithm is strongly based on the dominance-based MOSA algorithm which was introduced in [16] and aims for finding the true Pareto front of the optimisation problem. A detailed description of the algorithm can be found in [16]. In this report only the changes are described which were made to the algorithm to ensure good performance in the field of sensor location optimisation.

Our algorithm is called *modified* MOSA because some changes have been made to improve the performance of a sensor location optimisation. One of these changes consists in the use of a population p . While [16] only deals with one element searching for a solution, *mMOSA* uses P

elements that are searching for solutions simultaneously. This increases the speed of optimisation and the quality of the results when a large search space is assumed. Major changes have also been made to the perturbation function to ensure a wide coverage of the searching population p over the whole search space.

7.1 Perturbation

The perturbation in *mMOSA* basically consists of two steps:

1. Perturbation of the number of sensors.
2. Perturbation of the sensor locations.

Perturbation: Number of sensors

In the first step, the number of sensors used for contamination detection is altered within each individual. However, this change of number of sensors is not completely randomised. For example, if the number of sensors used by an individual increases, it cannot decrease for B iterations of the algorithm (but it can increase again). This provides the individual with time to find an optimal location for the new sensor. The value for B has to be chosen carefully to prevent that all individuals approach the upper and lower boundaries for the number of sensors.

Perturbation: Sensor locations

In the second step, the location of the sensors is perturbed. As the node index does not correspond to a certain area within the WDS, small changes can have a large effect on the contamination risk, and large changes in the node index can have a small effect. Therefore, the sensor locations are not perturbed in the usual sense, but are randomly chosen by a uniform random function. In each iteration, only one sensor location is changed. It is made sure that only sensor locations of sensors that are switched on are changed.

In addition to perturbation, which affects all elements of the population, new solution proposals are generated from the most successful individuals in the population. An individual is said to be successful, if it is included in the archive, i.e. the estimated Pareto Front $\tilde{F}$. If an individual is not included in the archive at a certain iteration, its success rate is decreased. In each iteration a certain percentage of the most successful individuals is copied twice into the population and overwrites the least successful individuals. The number of sensors for each copy of the individual is changed, one copy has a sensor removed, and the other copy has a sensor added. This method helps the algorithm to find new optimal solutions by using elements of the Pareto Front. This functionality of forced perturbation should only affect a small part of the population's individuals.

8 Optimisation results

This section presents results for the sensor location optimisation task stated in Section 1. Figures 5 and 6 show the objectives (i.e. number of sensors and risk of contamination) for both *NSGA-II* and *mMOSA*. The number of sensors is plotted against the x-axis, the risk of contamination is plotted against the y-axis. The unit of the latter is m^3 and is the average amount of water that is consumed before a contamination is detected. The markers show points where the corresponding optimisation algorithm has actually found an optimal solution. It can be seen that the contamination risk is almost similar for both both algorithms.

Another interesting thing is that both algorithms found optimal solutions for $S = 1$ to $S = 10$ sensors, i.e. the coverage in this area is very good. Multiple optimisation runs with different random seeds (i.e. different parent populations) show good coverage for both algorithms. The weaker coverage in the area of more than 10 sensors implies that the algorithms have difficulties to find better solutions for more than $S = 10$ sensors. The solutions that have been found (for $S > 10$) only differ slightly in their contamination risk. This again implies that the risk of

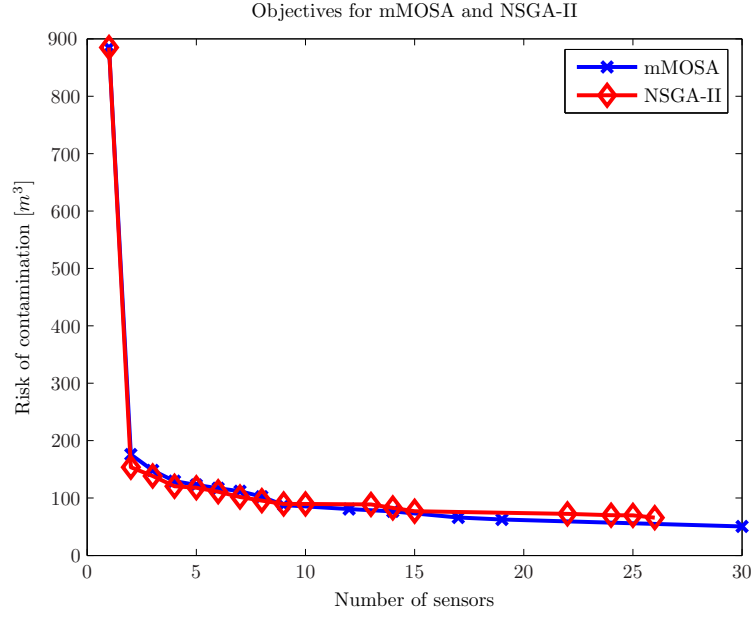


Figure 5: Optimisation results for *NSGA-II* and *mMOSA*

contamination cannot be decreased much further with a reasonable number of sensors. Therefore, it is recommended to use a maximal number of $S = 10$ sensors for contamination detection.

Tables 4 and 3 show the sensor locations for each mark in Figure 5. To show which sensors are particularly important, in Figure 7 a histogram of the sensor locations is plotted. This histogram shows the number of occurrences of each sensor location (i. e. node) in the optimal set for *NSGA-II* and *mMOSA*. Table 2 shows this histogram as a table; The ten most important sensor locations are shown for both two algorithms. Figures 8 and 9 show the locations of the most important sensor locations within the network (colour in the table equals the colour in the network graph) for *NSGA-II* and *mMOSA*, respectively.

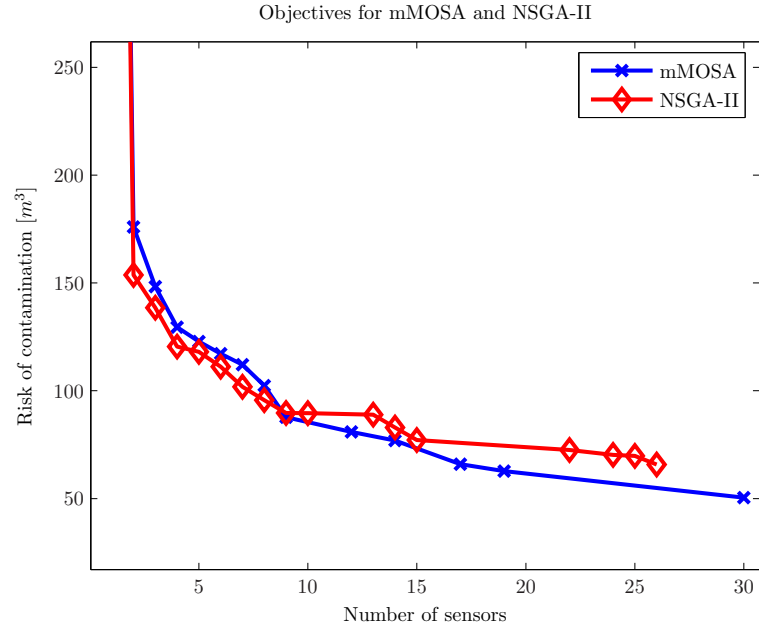


Figure 6: Optimisation results (cutout) for *NSGA-II* and *mMOSA*

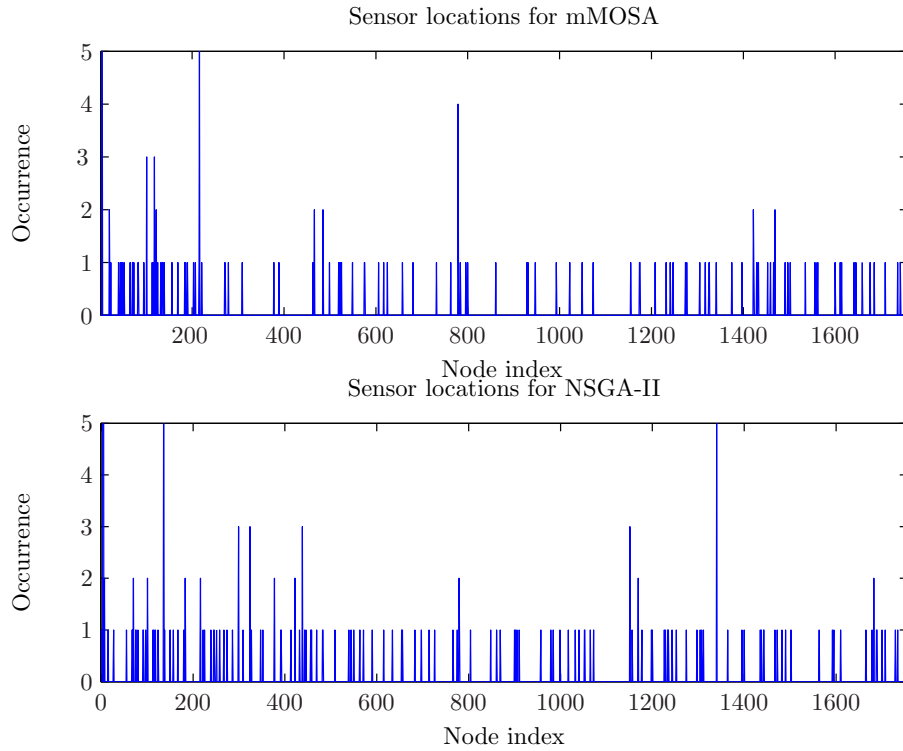


Figure 7: Histogram of sensor locations

<i>NSGA-II</i>		<i>mMOSA</i>	
node index	occurrences in the Pareto set	node index	occurrences in the Pareto set
4	19	4	7
5	9	216	5
136	6	779	4
1340	6	101	3
299	3	118	3
324	3	20	2
438	3	122	2
1151	3	466	2
7	2	485	2
70	2	1422	2

Table 2: The ten most important sensor locations

Number of Sensors	Sensor locations
1	101
2	118, 203
3	4, 122, 1685
4	4, 118, 485, 1556
5	46, 101, 135, 658, 1429
6	82, 118, 136, 207, 466, 1433
7	48, 65, 101, 221, 780, 1422, 1676
8	40, 83, 122, 216, 575, 1022, 1241, 1341
9	20, 23, 44, 139, 271, 466, 779, 1232, 1247
12	4, 95, 115, 125, 216, 272, 522, 681, 763, 931, 1326, 1600
14	4, 70, 71, 216, 606, 861, 929, 1155, 1325, 1491, 1560, 1645, 1736, 1742
17	4, 20, 73, 156, 169, 184, 216, 389, 576, 625, 779, 1073, 1175, 1453, 1466, 1502, 1535
19	4, 113, 186, 190, 215, 463, 499, 519, 525, 549, 779, 796, 993, 1049, 1274, 1305, 1397, 1469, 1497
30	4, 52, 132, 216, 279, 309, 378, 485, 617, 732, 779, 784, 800, 947, 1174, 1208, 1277, 1317, 1375, 1422, 1423, 1432, 1459, 1469, 1562, 1611, 1614, 1641, 1659, 1709

Table 3: Sensor locations (mMOSA)

Number of Sensors	Sensor locations
1	101
2	4, 55
3	4, 70, 1243
4	4, 70, 1340, 1665
5	4, 5, 509, 779, 1472
6	4, 5, 299, 1340, 1400, 1682
7	4, 5, 113, 299, 377, 848, 1340
8	4, 5, 7, 216, 438, 615, 979, 1340
9	4, 118, 299, 377, 443, 550, 1340, 1437, 1442
10	4, 5, 67, 136, 352, 634, 861, 957, 1274, 1340
13	1, 4, 4, 5, 136, 324, 539, 590, 656, 775, 1151, 1297, 1482
14	4, 4, 5, 97, 150, 216, 239, 438, 984, 1169, 1228, 1311, 1443, 1688
15	4, 4, 5, 136, 286, 324, 422, 766, 779, 1053, 1065, 1151, 1304, 1395, 1734
22	4, 76, 136, 183, 225, 246, 274, 347, 438, 483, 571, 683, 726, 869, 900, 1040, 1169, 1235, 1252, 1595, 1610, 1700
24	4, 5, 15, 27, 80, 91, 138, 180, 245, 432, 456, 469, 563, 804, 905, 910, 1017, 1072, 1198, 1226, 1233, 1467, 1502, 1729
25	4, 7, 68, 75, 115, 136, 167, 183, 251, 258, 267, 324, 391, 422, 482, 544, 654, 902, 1052, 1151, 1200, 1364, 1435, 1563, 1682
26	4, 92, 101, 124, 136, 157, 222, 309, 327, 392, 413, 446, 458, 697, 714, 780, 999, 1032, 1156, 1177, 1307, 1490, 1592, 1678, 1689, 1707

Table 4: Sensor locations (NSGA-II)

Almelo WDS, Sensor locations NSGA-II

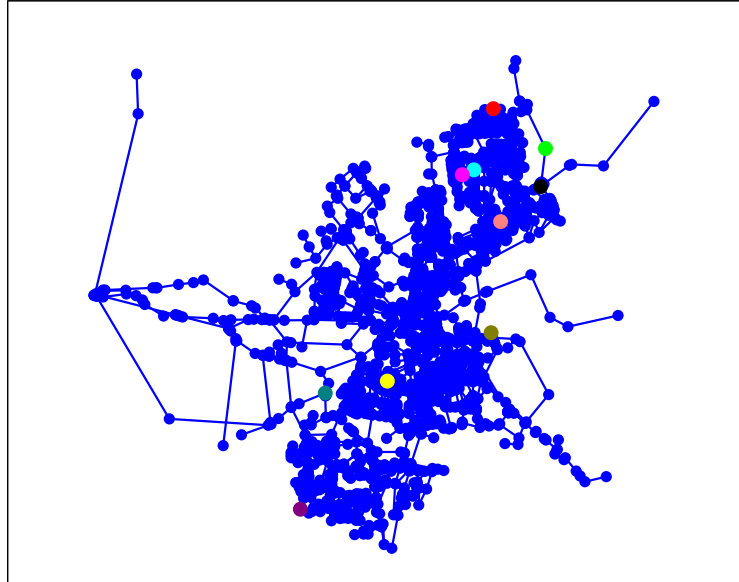


Figure 8: Important sensor locations for *NSGA-II*

Almelo WDS, Sensor locations mMOSA

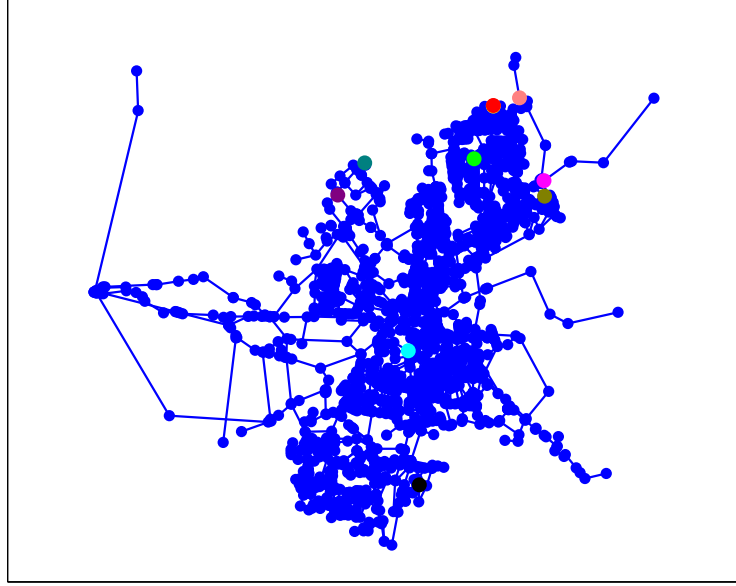


Figure 9: Important sensor locations for *mMOSA*

Figures 8 and 9 clearly show that there are two important parts of the *Almelo* WDN where sensors have to be placed, namely the upper-right corner of the network (north-east), and its centre. These are the areas where the highest amount of water is consumed. In the case where only one sensor is available for contamination detection, the optimal solution consists in placing that sensor in the upper-right corner of the WDN. This is shown in Figure 10. Even though large parts of the network seem to be unprotected (i. e. without sensor), the location in the North-East of the network minimises the risk of contamination. An exhaustive search showed that this point actually lies on the true Pareto front. For more than two sensors it is not possible (in reasonable time) to conduct an exhaustive search, because the number of possibilities increases rapidly. While the number of possible sensor locations is equal to the number of junctions in the case of one sensor, i. e. 1758, in the case of two sensors there are more than 1.5 million possibilities, for three sensors there are more than 900 million possibilities. As the exhaustive search in the case of two sensors took about 45 minutes, a search for three sensors would take almost 19 days (27,000 minutes). Figure 11 shows the optimal sensor locations for $S = 2$. The risk of contamination for the optimal sensor locations is 150.79 and thus very close to the result produced by *NSGA-II*.

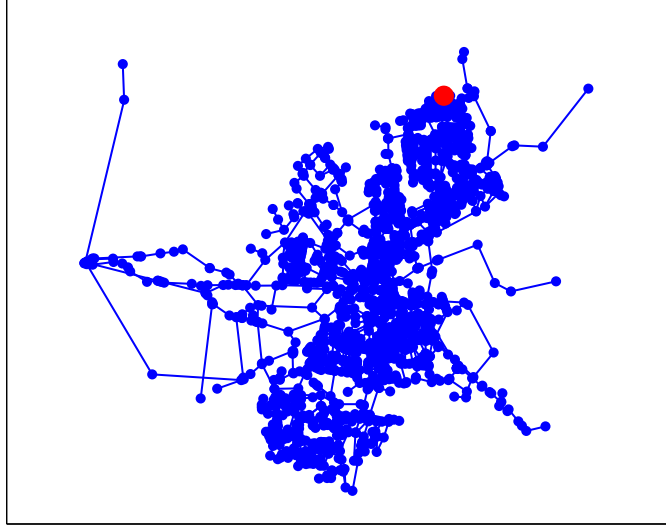


Figure 10: Optimal location for one sensor

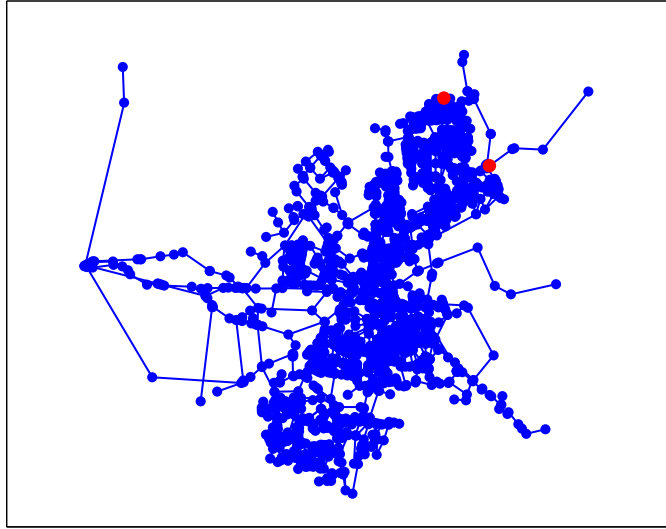


Figure 11: Optimal locations for the use of two sensors

9 Conclusions

An overall concept of solving the optimal sensor location problem for contamination detection in water distribution systems is presented here. Starting with a hydraulic analysis, scenarios are chosen by a new algorithm, called the *Importance-based sampling method*. After choosing the scenarios, quality data for each event are calculated by performing an analysis using the hydraulic simulation program *ALEID* and a *DLL* which was developed in the scope of this project.

The optimisation task is formulated as a two-objective problem. The first objective is the number of sensors used for contamination detection, the second objective is the risk of contamination which is the average amount of water that is consumed prior to contamination detection.

The optimisation is carried out for the Dutch city Almelo and is performed by two algorithms:

NSGA-II and *mMOSA*. While *NSGA-II* is well known, *mMOSA* is a new algorithm which is based on *MOSA*. The optimisation results show similar performance for both algorithms. Even though the sensor indices used for contamination detection are not the same for both algorithms (except for the use of one sensor), three major areas of the water distribution network can be identified. The majority of the sensors is located in one of these areas.

A Table of symbols

b	binary variable (GA)
C	concentration of contaminant at a node N_l
C_i	intrusion concentration
C_t	threshold value for intrusion detection (sensor parameter)
d	pipe diameter
D	nodal demand
E	contaminant (e. g., chemical)
f	objective function
$\mathbb{F}$	objective function space
F	true Pareto front
$\hat{F}$	augmented Pareto front
$\tilde{F}$	estimated Pareto front
h	headloss in a pipe
H	nodal head
I	importance value of a scenario
$\mathbf{I}$	importance matrix
K	contamination event
L	pipe length
$\underline{L}$	sensor locations
$\mathbb{K}$	set of possible scenarios
$\mathbb{K}_c$	set of scenarios chosen for optimisation ($\mathbb{K}_c \subset \mathbb{K}$)
M	number of samples
N	number of junctions in the network
N_i	intrusion node
P	number of individuals (NSGA-II and mMOSA)
Q	flow in a pipe
r	random variable in $(0, 1)$
R	contamination risk for the population
S	number of sensors used for contamination detection
$\underline{S}$	vector of sensors used to localise contamination
S_m	maximum number of available sensors
S_n	minimum number of sensors
δt_q	time step for the quality analysis
t_{det}	time of intrusion detection at a node N_l
ΔT	duration of an intrusion
ΔT_i	intrusion resolution (interval between possible intrusions)
T	maximal simulation time for quality analysis
$T_{i,0}$	latest, earliest possible time of intrusion
t_j	intrusion time
V	pipe volume
W	energy of a solution candidate
$\underline{W}$	sensor switches
x	decision variable
$\underline{x}$	decision vector

$\xi(\cdot)$	detection function
ξ	detection matrix
$\chi(\cdot)$	contamination function
χ	contamination matrix

B Abbreviations

CE	Cross-entropy method
EA	Evolutionary Algorithms
EPS	Extended Period Simulation
GA	Genetic Algorithms
KCL	Kirchhoff's Current Law
mMOSA	Modified Multi-objective Simulated Annealing
MOSA	Multi-objective Simulated Annealing
NSGA-II	Non-dominated Sorting Genetic Algorithm II
TSP	Travelling Salesman Problem
USEPA	U.S. Environmental Protection Agency, Water Supply and Water Resources Division
WDS	Water distribution system
WDN	Water distribution network

C Quality analysis parameters

Name	Value	Unit	Description
C_t	100	$[mg]$	Detection threshold for the contamination sensors
C_i	10^7	$[mg/L]$	Concentration of the contaminant at the intrusion node (fixed)
M	10000 ¹	[1]	Number of scenarios which are selected for a quality analysis
δt_q	600	[s]	Time step for the quality analysis (10 minutes)
T	259200	[s]	Simulation time for the quality analysis (72 hours)
ΔT_i	1200	[s]	intrusion resolution (interval between possible intrusions, 20 minutes)
T_i	86400 and 0 ²	[s]	Time in which a contamination event can occur (first 24 hours into the simulation)
ΔT	57600	[s]	Duration of the intrusion (16 hours)

¹ $M = 1758$ if intrusion only at 0 hrs

²Optimisation runs were carried out for both parameter values

References

- [1] A. Kessler, A. Ostfeld, and G. Sinai, “Detecting accidental contaminations in municipal water networks,” *Journal of Water Resources Planning and Management*, vol. 124, pp. 192–198, 1998.
- [2] A. Ostfeld and E. Salomons, “Optimal layout of early warning detection stations for water distribution systems security,” *Journal of Water Resources Planning and Management*, vol. 130, pp. 377–385, 2004.
- [3] A. Kumar, M. L. Kansal, and G. Arora, “Detecting accidental contaminations in municipal water networks - discussion,” *Journal of Water Resources Planning and Management*, pp. 308–310, 1999.
- [4] G. Dorini, P. Jonkergouw, Z. Kapelan, F. di Pierro, S. Khu, and D. Savic, “An efficient algorithm for sensor placement in water distribution systems,” in *8th Annual Water Distribution System Analysis Symposium, Cincinnati, Ohio, USA*, 2006.
- [5] A. Preis and A. Ostfeld, “Multiobjective sensor design for water distribution systems security,” in *8th Annual Water Distribution System Analysis Symposium, Cincinnati, Ohio, USA*, 2006.
- [6] Y. Shastri and U. Diwekar, “Sensor placement in water networks: A stochastic programming approach,” *Journal of Water Resources Planning and Management*, vol. 132, pp. 192–203, 2006.
- [7] J. W. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J.-P. Watson, “Sensor placement in municipal water networks,” *Journal of Water Resources Planning and Management*, pp. 237–243, 2005.
- [8] R. D. Carr, H. J. Greenberg, W. E. Hart, and C. A. Phillips, “Addressing modeling uncertainties in sensor placement for community water systems,” in *World Water Congress 2004 - Critical Transitions in Water and Environmental Resources Management*, 2004.
- [9] S. A. McKenna and D. B. H. and Lane Yarrington, “Impact of sensor detection limits on protecting water distribution systems from contamination events,” *Journal of Water Resources Planning and Management*, vol. 132, pp. 305–309, 2006.
- [10] J. Berry, W. E. Hart, C. A. Phillips, J. G. Uber, and J.-P. Watson, “Sensor placement in municipal water networks with temporal integer programming models,” *Journal of Water Resources Planning and Management*, vol. 132, pp. 218–224, 2006.
- [11] V. Pareto, *Manuale di Economia Politica*. Piccola Biblioteca Scientifica, Milan, 1906. Translated into English by Ann S. Schwier (1971), *Manual of Political Economy*, MacMillan, London.
- [12] N. Srinivas and K. Deb, “Multiobjective optimization using nondominated sorting in genetic algorithms,” *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248, 1994.
- [13] C. M. Fonseca and P. J. Fleming, *Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms—Part I: A Unified Formulation*, vol. 28. 1998.
- [14] C. M. Fonseca and P. J. Fleming, “An overview of evolutionary algorithms in multiobjective optimization,” *Evolutionary Computation*, vol. 3, no. 1, pp. 1–16, 1995.
- [15] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, “A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii,” 2000.
- [16] K. I. Smith, R. M. Everson, J. E. Fieldsend, C. Murphy, and R. Misra, “Dominance-based multi-objective simulated annealing,” *IEEE Transactions On Evolutionary Algorithms*.

BIJLAGE 2 Vergelijk ALEID-simulatie handmatig en automatisch (dII)

Handmatige ALEID-simulatie

In vergelijking met 2004 zijn de waterkwaliteitsberekeningen iets verbeterd:

Het ALEID-model is verdund (met het programma Verdunner®, zie ook §2.2.1) waardoor de berekeningen sneller zijn.

In dit project is gerekend met een verontreiniging met continu 10^7 KV/L, die gedurende 16 uur in het leidingnet is binnengedrongen met een volumestroom van 10 L/h, gestart om 00.00 uur (12.00 uur 's nachts). In 2004 is gerekend met een verontreiniging met continu 10^8 KV/L, die gedurende 16 uur in het leidingnet is binnengedrongen met een volumestroom van 1 L/h. Hierdoor is ten opzichte van 2004 in het model sprake van een betere balans tussen nauwkeurigheid en een stabiele berekening en daardoor treden minder afrondingsfouten op.

Omdat met name bij verontreinigingen op leidingen met een kleine volumestroom de mengverhouding tussen schoon en verontreinigd water verandert wanneer niet 1 L/h, maar 10 L/h wordt toegevoegd is ook de detectiegrens van het model (zie §3.1.5) aangepast (van 10 KVD/L naar 100 KVD/L) zodat er vergelijkbare resultaten met 2004 zijn te verwachten.

In dit project is een langere periode is doorgerekend dan in 2004. In 2004 was het bezwaar dat de verontreiniging binnen de rekentijd zijn eindbestemming mogelijk niet heeft gehaald en daardoor de detectiekans kleiner lijkt dan hij in werkelijkheid is. Dat bezwaar is nu omzeild. De berekeningen zijn met intervallen van 30 minuten uitgevoerd en per uur gerapporteerd voor 40 dagen na de start van de verontreiniging.

Per verontreiniging zijn de concentratieverlopen van 960 uurmetingen op alle (meer dan 1700) mogelijke monsterlocaties (incl. productiebedrijf, reservoirs en de 44 woningen van het monsterprogramma 2002) berekend.

De verontreinigingen zijn op de volgende plaatsen gesimuleerd:

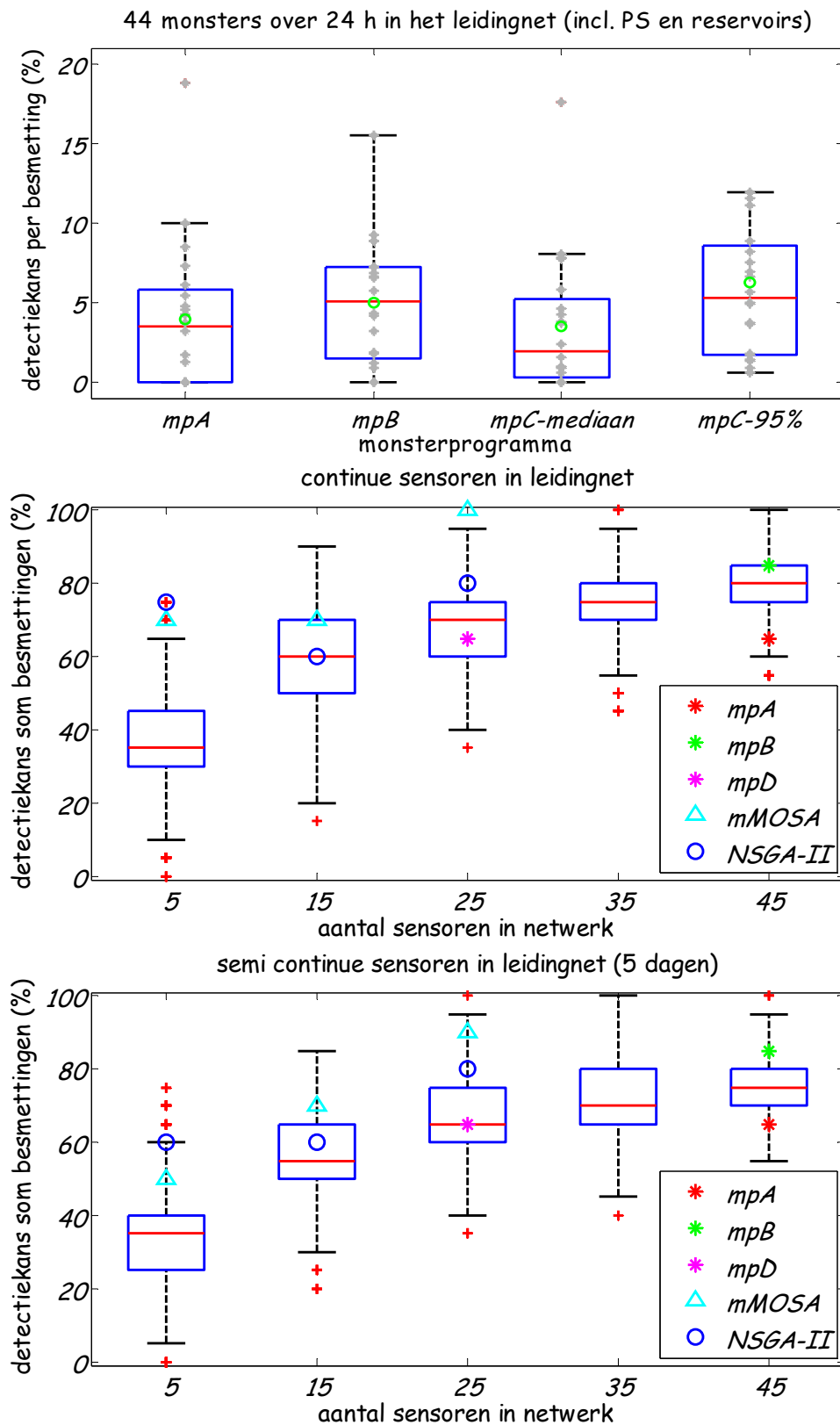
- Productiebedrijf Wierden, in de uitgaande leiding (gelijk aan 2004).
- Laagreservoir Bornestraat, direct in het reservoir (gelijk aan 2004).
- Hoogreservoir Almelo, direct in het reservoir (gelijk aan 2004).

Op 20 distributieleidingen, waarvan 9 op gelijke knopen als in 2004 en 11 handmatig gekozen locaties zodanig dat een goede spreiding over het net ontstaat.

De resultaten van 2004 en voorliggend onderzoek zijn niet één-op-één te vergelijken, omdat er enkele kleine verschillen in de berekening zitten.

De volgende figuren zijn van toepassing op

- Handmatige ALEID-simulatie;
- 20 verontreinigingen;
- detectielimiet = 100 KVD/L (vergelijkbaar met 10 KVD/L in onderzoek 2004 voor verontreinigingen in het net, niet op pompstation en in reservoirs).



Handmatig zijn slechts 20 verontreinigingen doorgerekend; dit geeft onvoldoende statistisch significante resultaten. Dat blijkt met name in de onderste twee van de bovenstaande figuren uit het feit dat relatief vaak een 100% detectiekans gevonden wordt. De bovenste figuur verandert te veel wanneer de simulatie nogmaals wordt gedaan met een verontreinigings-locatie meer of minder.

Automatische ALEID-dll-simulatie

Omdat bleek dat 20 verontreinigingslocaties een onvoldoende statistische betrouwbare uitkomst gaf ontstond de behoefte om meer verontreinigingen door te rekenen. Daarom was het gewenst om minder handmatig te rekenen en de uitkomsten te kopiëren en op te slaan, maar dit alles te automatiseren. Daarvoor is gebruik gemaakt van de in Exeter ontwikkelde en gebruikte ALEID-dll. De simulatie-instellingen zijn daardoor wel iets anders ten opzichte van de ALEID-simulatie van de vorige paragraaf.

Het grote verschil is dat met de handmatige simulaties (uit de vorige paragraaf en zoals toegepast in het onderzoek van 2004) een constante verontreiniging wordt toegevoegd (en de concentratie in het drinkwater daardoor varieert met de volumestroom – dus het verbruik - van het drinkwater), terwijl bij de geautomatiseerde berekeningen (van deze paragraaf, en verder toegepast in dit onderzoek) de concentratie van de besmetting in het drinkwater constant is. Om de resultaten van voorliggend onderzoek te kunnen vergelijken met de resultaten uit 2004 zijn een aantal instellingen in de berekening aangepast.

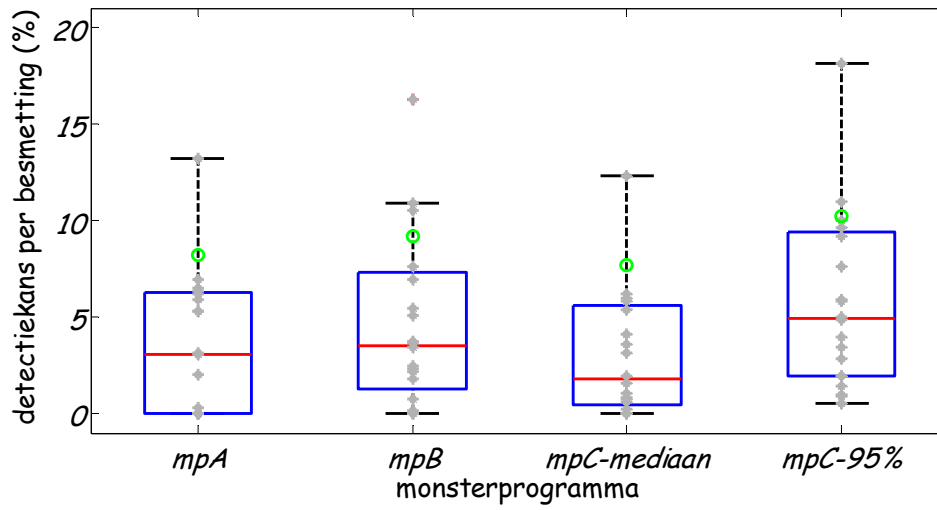
Voor de simulaties is gebruik gemaakt van een verontreiniging met een continue concentratie van 10^3 KVD *E. coli* per liter drinkwater, die gedurende 16 uur in het leidingnet aanwezig is, gestart om 00.00 uur (12.00 uur 's nachts). De berekeningen zijn met intervallen van 1 uur uitgevoerd en per uur gerapporteerd voor 40 dagen na de start van de verontreiniging. Per verontreiniging zijn de concentratieverlopen van 960 uurmetingen op alle (meer dan 1700) mogelijke monsterlocaties (incl. productiebedrijf, reservoirs en de 44 woningen van het monsterprogramma 2002) berekend.

De detectielimiet bij de ALEID-dll simulatie is afgestemd op de detectielimiet van de handmatige ALEID-simulatie zodanig dat de berekende detectiekansen van beide methoden vergelijkbaar zijn. Vergelijk hiertoe de figuren in deze bijlage.

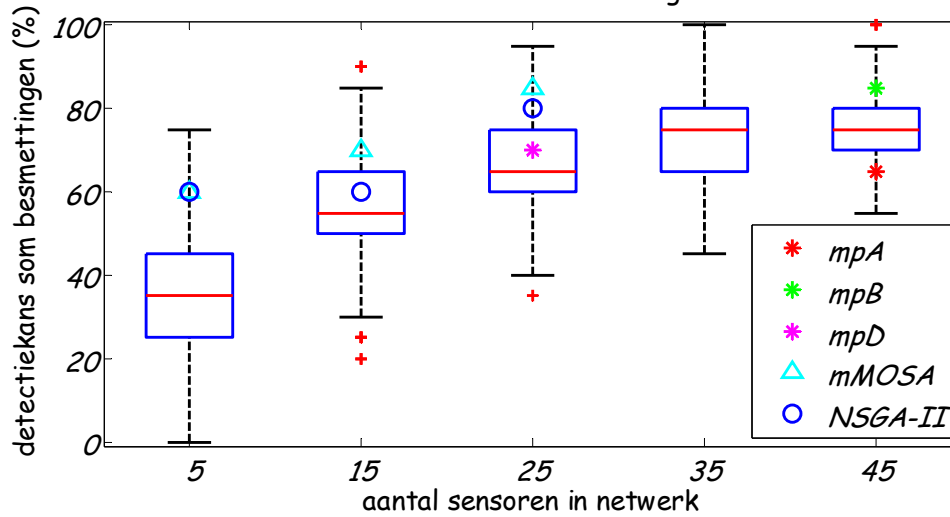
De volgende figuren zijn van toepassing op

- automatische ALEID-dll simulatie;
- 20 verontreinigingen (zelfde locaties als voorgaande figuren);
- detectielimiet = 5 KVD/L.

44 monsters over 24 h in het leidingnet (incl. PS en reservoirs)



continue sensoren in leidingnet



semi continue sensoren in leidingnet (5 dagen)

