**Datavalidatie algemeen**HDSR heeft 661 locaties waar geautomatiseerd de waterstand gemeten wordt. Op de weg van meting tot eindgebruiker kunnen op allerlei punten dingen mis gaan die tot foute data leiden. Dit is beschreven in het H2O artikel ‘[Datavalidatie: voorbeelden uit de praktijk van waterkwantiteitsmetingen](https://www.h2owaternetwerk.nl/images/2020/Juli/H2O-Online_200702_Datavalidatie.pdf)’.

Bij HDSR wordt het controleren en valideren van de metingen is deels geautomatiseerd en deels visueel gedaan. De visuele controle vindt circa 3 keer per jaar plaats, waardoor het soms enkele maanden kan duren voordat fouten worden opgespoord. Ook beginnen sommige afwijkingen klein en groeien ze langzaam, waardoor ze pas na langere tijd zichtbaar worden.

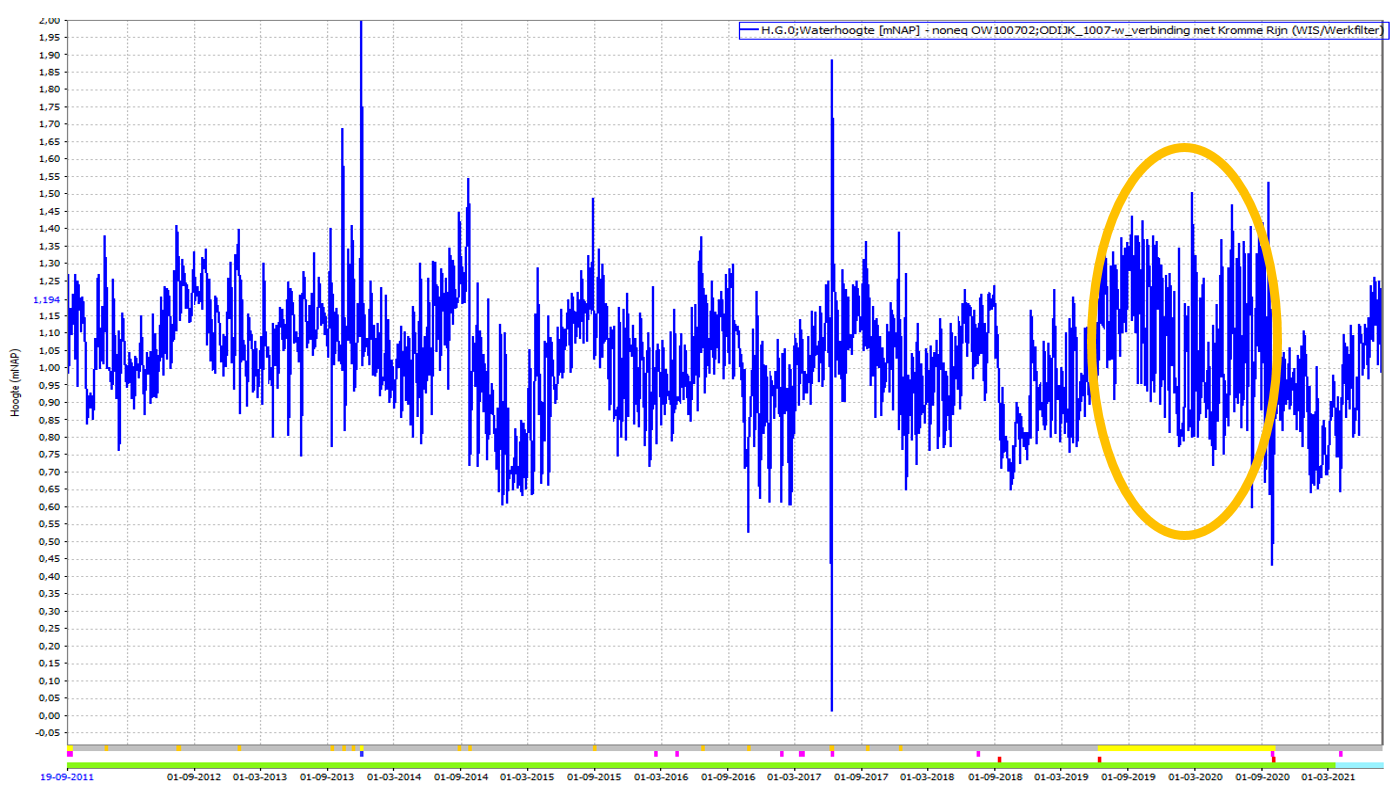
De wens is dat de visuele validatie door de datavalidator steeds verder wordt geautomatiseerd, te beginnen met het automatisch opsporen van vervuiling van de sensor met schelpdieren.

**Het probleem van de schelpdieren**Een van de typische fouten in waterstandsmetingen ontstaat door vervuiling van de sensor met schelpdieren. Op 27 locaties zijn 1 of meer periodes metingen afgekeurd vanwege (vermoeden van) vervuiling met schelpdieren. Vaak begint het klein en neemt de afwijking steeds verder toe. Afhankelijk van de locatie, het kunstwerk en het normale gedrag van de waterstand worden deze fouten eerder of later opgespoord.



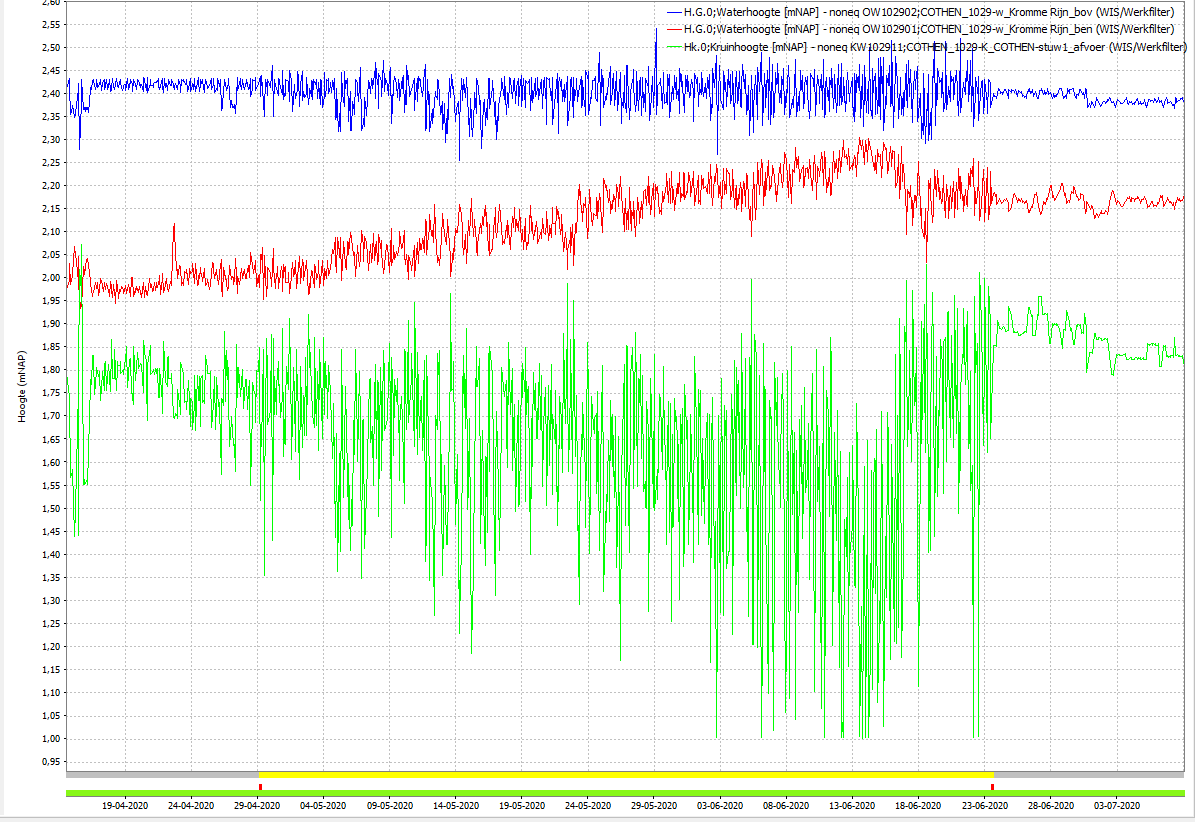
Figuur 1 Met schelpdieren vervuilde sensor

Op een locatie met sterk variabele waterstand zullen de pieken minder snel opvallen, vaak blijft de waterstand lange tijd binnen min of meer normale grenzen, zie figuur 2. Dit zorgt ervoor dat het lang kan duren voordat de fout in de metingen wordt opgespoord. Daarnaast is het moeilijk om te bepalen vanaf welk moment de metingen moeten worden afgekeurd.



Figuur 2: Periode met foute metingen geel gemarkeerd. Metingen zijn fout, maar blijven absoluut gezien redelijk binnen normale bandbreedte

In figuur 3 zien we wat er gebeurt als een kunstwerk stuurt op de foute waterstand. De stuw stuurt op het bovenpeil, doordat de vervuiling van de sensor ontstaan er pieken en daar reageert de stuw (groene lijn) op door de hele tijd omhoog en omlaag te sturen. Als gevolg van de afwijkende sturing wordt de waterstand benedenstrooms ook vreemd piekerig. Het is moeilijk om in een dergelijke situatie oorzaak en gevolg uit elkaar te houden. Door de schelpdieren geeft de bovenpeilsensor niet de werkelijke waterstand weer. Maar de pieken in de benedenwaterstand en de kruinhoogte hebben wel echt plaatsgevonden.



Figuur 3 Fout in bovenpeil (blauw) geeft sterke schommeling in kruinhoogte (groen) en daardoor ook in benedenpeil (rood)

**Beschikbare data**  
Op 27 locaties zijn 1 of meer periodes metingen afgekeurd vanwege (vermoeden van) vervuiling met schelpdieren. Van deze locaties zijn non-equidistante metingen beschikbaar en metingen per 15 minuten. De metingen kunnen uit WIS gehaald worden. Een deel van de data gebruiken als trainingsdata en een deel om te testen.

Er is een Word document met een beschrijving van de datavaldatie, hieruit kan voor de 27 locaties de periode van afwijking gehaald worden. Desgewenst wil ik hier een overzicht van maken.

**Gewenste opbrengst**Script waarmee foute metingen als gevolg van vervuiling met schelpdieren opgespoord kunnen worden. Implementatie van het script in WIS -> maandelijks laten draaien en als output verdachte locaties + verdachte periode.

Lijst met locaties:

|  |
| --- |
| OW100202 |
| OW100702 |
| OW102902 |
| OW103101 |
| OW104503 |
| OW104504 |
| OW108101 |
| OW108903 |
| OW109201 |
| OW210401 |
| OW212502 |
| OW214101 |
| OW215701 |
| OW217502 |
| OW218101 |
| OW320301 |
| OW431301 |
| OW431401 |
| OW431702 |
| OW433503 |
| OW434001 |
| OW434102 |
| OW437001 |
| OW439402 |
| OW440001 |
| OW440002 |
| OW463401 |