

Nota over nuttige tools bij software voor gegevensverwerking. IN.A.2003.2

Pieter Cabus, Marcel Voet

Aan de ‘onderzoeksgroep Landelijk Waterbeheer’ wordt sinds eind 2001 werk gemaakt van de ‘correctie en validatie’ van de historische meetreeksen van peil- en debiet in de limnigrafische stations van de Afdeling Water. Hierbij worden verschillende software pakketten gebruikt en een groot aantal tools. De belangrijkste hiervan zijn het pakket CODEAU© en het pakket WISKI-TV (<http://www.kisters.de>) van de firma Kisters©. In de volgende bespreking zal aangeduid worden, welke tools in welk pakket aanwezig zijn, of ze al dan niet voldoende functioneren, en desgevallend welke tools voorlopig nog in geen van beide pakketten aanwezig zijn. Naast deze pakketten is er ook nog het HYMOS©-pakket van WL-Delft (<http://www.wldelft.nl/soft/hymos>). Hier hebben wij geen ervaring mee. Andere software voor de visualisatie en wijziging van langere tijdreeksen is ons niet bekend.

In eerste instantie kan vermeld worden dat er bij de aanmaak van een databank rekening gehouden wordt met extra informatie per gegeven over de validatie‘toestand’. Bij het inlezen van de data zal dat een code zijn voor ‘on gevalideerd’, waarna deze data de validatieprocedure kunnen doorlopen. Voor de verschillende validatie-stappen zou dan de code kunnen gewijzigd worden. <validatiecode>. WISKI houdt systematisch in de file’s wijzigingen bij door middel van een standaard code. Niet-standaard coderen is mogelijk door het invoegen van een extra tekstkolom. Deze beide opties functioneren niet optimaal in de “stand-alone” WISKI-TV. Ook CODEAU codeert bewerkte gegevens, door middel van 16 aan/uit vlaggen. Deze codering loopt vlekkeloos. De 8 standaard codes worden gegenereerd door CODEAU, er is ruimte voorzien voor 8 extra gebruiker codes.

Algemeen kan ook gesteld worden dat een goede grafische interface essentieel is. De data moeten kunnen gevisualiseerd worden in een zo breed mogelijke waaier van combinaties (verschillende reeksen) en tijdspannes. Zowel WISKI als CODEAU laten toe de tijdreeksen grafisch voor te stellen, in te zoomen, ... Bovendien moet deze interface het toelaten data ‘on screen’ te bewerken, markeren, ... Dit is niet mogelijk bij CODEAU.

Bij de bespreking van de verschillende stappen zal dieper ingegaan worden op de tools die bij die bewerking nuttig zijn.

1. Input en Output

Gezien de diversiteit aan formaten, tijdstappen, ... is het essentieel dat het systeem aauitgebreide waaier van ASCII-formaten kan inlezen (en evt. wegschrijven) <inputmodule>. De inputmodule zal vanzelfsprekend niet in staat zijn de verschillende formaten waarin de historische data staan allemaal te herkennen (in het ideale geval probeert het programma dit natuurlijk wel, en geeft het formaat aan waarin de gegevens herkend worden). Het is daarom belangrijk dat er een interface is om het formaat van de gegevens te definiëren, of de voorgestelde definitie ervan te veranderen. Hierin kan bvb. aangeduid worden of er sprake is van kolommen van vaste breedte, dan wel of er een scheidingsteken tussen de kolommen staat en welk dit scheidingsteken is (“,”, “;”, “/”, “:”, “SPACE“, “TAB“, of een combinatie van deze), in welk formaat de datum staat, ... WISKI heeft geen universele inputmodule, kan enkel vooraf gedefinieerde en door KISTERS geprogrammeerde formaten lezen. WISKI

werkt ook op die file's zelf, maakt geen nieuwe WISKI-file aan. CODEAU heeft wel een zeer goede en universele inputmodule en maakt bij het inlezen van data eigen CODEAU-file aan. Ook naar output kan zo'n interface handig zijn <*outputmodule*>. De afdeling Water hanteert immers verschillende formaten: bvb. LIMNIMET-formaat, ISIS-PDM-formaat. Ook hier laat WISKI te wensen over. Slechts de formaten die KISTERS geprogrammeerd heeft kunnen door WISKI weggeschreven worden. Opnieuw heeft CODEAU een ruimere module voor het wegschrijven van output-file's.

Naast deze multifunctionaliteit in formaat is het aangewezen *één formaat* te hanteren voor opslag en bewerking van reeksen. Dit formaat lijkt best gerelateerd aan de formaten uit de databank. CODEAU hanteert z'n eigen (binaire) formaat voor de opslag en behandeling van reeksen. Bij de interactie met een databank dienen de file's eerst geëxporteerd, waarna in CODEAU ingelezen, waarna geëxporteerd, waarna in de databank ingelezen. WISKI is een databank-gebaseerde software die geprogrammeerd wordt voor de databank van de gebruiker en werkt rechtstreeks op de databank of bij een 'stand-alone' op de ingelezen file's. Hierdoor is WISKI sterk te prefereren boven CODEAU bij de implementatie in een geïntegreerd systeem bij een gestructureerde databank.

2. Analyse van de peilreeks

Voor vele limnigrafische stations vormen de peilgegevens de basisgegevens. Eventuele outliers, ontbrekende waarden, definiëerbare onnauwkeurigheden kunnen automatisch verwijderd of gemarkeerd worden. <*scripts definiëren*><*tijdsafhankelijke reken-eenheid*>. Bij het inlezen in het CODEAU-formaat kunnen enkele parameters onderzocht worden (maximum, minimum, maximale en minimale gradiënt, periodiciteit). WISKI heeft een rekenmodule die toelaat om nieuwe reeksen te berekenen op basis van andere reeksen. Hierdoor is het ook mogelijk om dergelijke reekseigenschappen te onderzoeken of eventueel te verbeteren.

Daarna dienen de reeksen aan een visuele controle te worden ontworpen <*grafische interface*>. Belangrijk hierbij is dat de volledige reeks kan bekeken worden, waarna ingezoomd kan worden op interessante perioden <*uitgebreide zoom-functie*>. Beide pakketten kunnen tijdreeksen grafisch voorstellen en inzoomen. WISKI heeft hierbij het nadeel dat er automatisch ingezoomd wordt op zowel de tijds-as als de waarde-as. CODEAU zoomt automatisch (via de muis) in op de tijdas. De waarde-as blijft behouden en kan via het klavier gewijzigd worden. Hierdoor is het makkelijker om reeksen naast elkaar voor te stellen, bvb. met omgekeerde assen.

Voornamelijk onnatuurlijke peilverschuivingen en eventuele peilverschuivingen ten gevolge van plantengroei dienen hier gedetecteerd te worden. Zeer handig is dan ook dat een langere periode kan geselecteerd en gemarkeerd worden in de grafiek <*on screen selecteren*> <*on screen markeren*>. CODEAU laat enkel toe om waarde per waarde te markeren in de grafische mode. Voor de selectie van een langere periode dient de grafische mode verlaten te worden. Dit betekent een groot tijdsverlies. Begin- en einddatum dienen te worden genoteerd. WISKI laat toe langere perioden te selecteren. Automatisch markeren kan evenwel niet. Wel kan de geselecteerde range gekopieerd worden naar een andere (tijdelijke reeks) waardoor het als 'gemarkeerd' kan beschouwd worden.

3. debietkromme

De analyse van de debietkromme maakt essentieel deel uit van de validatie van een meetreeks. WISKI-TV is hierbij geen. Het volledige WISKI-pakket bevat wél een Q/H-editor. De performantie daarvan is ons niet bekend. CODEAU bevat een beperkte Q/H-editor. Deze werd door ons nog niet gehanteerd vanwege de beperkte mogelijkheden. Hieronder worden toch de tools geschetst die een dergelijke editor zou moeten bezitten.

De omzetting van peil- naar debietreeks gebeurt op basis van de debietkromme. Het lijkt dan ook aangewezen een interface te hebben voor de behandeling/selectie van goede debietkromme(n) <*interface debietkromme*>. Hierbij moet je de beschikbare ijkingspunten (manuele snelheidsmetingen, eventuele resultaten van continue snelheidsmetingen) kunnen selecteren, *krommen opstellen tot beperkte hoogten, geleidelijke peilverschuivingen in rekening kunnen brengen, ...*

Het zou mogelijk moeten zijn de vergelijkingen handmatig in te brengen van debietkromme(n) <*invoer vergelijking*>. Als er een module voorzien wordt die zelf met statistische methoden één of meerdere debietkromme(n) voorstelt op basis van de geselecteerde punten <*berekening vergelijking*>, dient deze de volgende eigenschappen te bezitten:

- willekeurige selectie van ijkingspunten op basis waarvan de kromme berekend wordt <*selectie punten*>, eventueel ook automatische selectie op basis van seizoen of jaartal
- de ijkingspunten moeten ook kunnen verschoven worden op basis van een formule, om bvb. peilwijzigingen in rekening te kunnen brengen. <*aanpassing ijkingspunten*>
- optimale (state of the art)-statistische methode's voor de automatische berekening van de kromme <*statistische methode's*>
- optimale (state of the art)-methode's voor de berekening van overgangskrommen tussen twee gegeven krommen, een punt en een kromme, ... <*berekening overgangskrommen*>
- verschillende mogelijke vergelijkingen (machts wetten, meerderegraadspolynomen, evt. klassieke overlaatvergelijkingen en zeker een vrij te definiëren functie) <*verschillende vergelijkingen*>
- statistische voorstelling van de resultaten <*correlatie-coëfficiënt, betrouwbaarheidsintervallen, significantie van parameters in vgl.*>

Eventueel kan het bovenstaande ook buiten het systeem uitgevoerd worden in bvb. een spreadsheet in combinatie met een statistisch pakket. Daarvoor dienen zowel de bestaande debietkromme als de ijkingspunten op een eenvoudige manier *geëxporteerd* te worden.

Uiteindelijk moet het systeem uitgaande van de debietkromme(n) die uiteindelijk geselecteerd werd(en) en een, eventueel deels verschoven, peilreeks de debietreeks kunnen bepalen <*rekenmodule met logische operatoren*>. Zowel CODEAU als WISKI laten toe de peilreeks om te zetten in een debietreeks. Beide bevatten een rekenmodule waarin nieuwe reeksen kunnen berekend worden op basis van bestaande reeksen en logische en mathematische operatoren. Bij CODEAU kan de input bestaan uit tot 5 verschillende reeksen. WISKI-TV laat ogenschijnlijk een onbeperkte combinatie van reeksen, voorwaarden, ... toe binnen z'n rekenmodule.

4. Reeksverbetering

Bij de uiteindelijke reeksverbetering is het van belang om zoveel mogelijk vrijheid te hebben bij de weergave, selectie en behandeling van de reeks.

- weergave: Essentieel is een duidelijke en handige *grafische interface*, met zoommogelijkheden, mogelijkheid tot afbeelden van *verschillende reeksen* boven elkaar, *manipulatie van kleuren en assen*, Handig is ook dat een neerslagreeks op een *secundaire (onafhankelijke) Y-as* omgekeerd kan afgebeeld worden. Ook tijdens het *zoomen* zou deze moeten kunnen behouden blijven. Zowel CODEAU als WISKI-TV kunnen verschillende reeksen afbeelden. CODEAU is hierbij beperkt tot 2 assen, WISKI-TV kan per reeks een nieuwe as maken. De manipulatie van de formaten is in WISKI uitgebreider, de basisconfiguratie in CODEAU oogt wel beter dan deze in WISKI (overzichtelijker). Beide programma's laten toe een secundaire inverse as af te beelden. Bij het zoomen heeft WISKI-TV het nadeel (cf. supra) dat er zowel in de tijd als in de waarde gezoomd wordt, waardoor de secundaire as vaak van het beeld verdwijnt bij het inzoomen. Verder is het bij het inzoomen niet meer mogelijk de assen via het klavier te definiëren.
- Selectie: Het moet mogelijk zijn *verschillende punten tegelijk* te behandelen, punten uit een andere reeks grafisch te kopiëren in de te verbeteren reeks, ... Dit neemt niet weg dat individuele punten zowel grafisch als in de tabel moeten kunnen aangepast worden. CODEAU laat in de grafische interface enkel toe dat individuele punten behandeld worden. WISKI laat wel toe meerdere punten tegelijk te behandelen.
- Afgeleide reeksen bvb. leegloopcurven, debieten in mm/s (oppervlaktedegedeeld), dag-, maand- en jaargemiddelden, ... moeten eenvoudig berekend kunnen worden *<rekenmodule met logische en mathematische, tijdsafhankelijke operatoren>*. Ook *combinaties* van reeksen moeten hiermee kunnen berekend worden, bvb. vergelijking tussen neerslag-debiet, berekening concentratietijden. Ook voor de behandeling/validatie van neerslagdata is een dergelijke module essentieel voor bvb. berekening van Thiessen-neerslagen. Deze module dient ook rekening te houden met de *tijdsstap* van de verschillende reeksen (het al dan niet ontbreken van waarden, niet-equidistante tijdsreeksen, ...) voor de berekening. Hierbij dient opgemerkt dat een ontbrekende waarde in een neerslagreeks anders geïnterpreteerd dient te worden dan een ontbrekende waarde in een debietreeks *<kenmerken van hydrologische parameters>*. Idealiter kunnen deze afgeleide reeksen *continue berekend* worden zodat wijzigingen in basisdata onmiddellijk effect hebben op de afgeleide reeks. De rekenmodule van CODEAU kan tot 5 reeksen beschouwen, met verschillende tijdstappen. De mogelijke interpolaties worden per type van tijdreeks verschillend uitgevoerd. Er kunnen zowel mathematische als logische bewerkingen uitgevoerd worden in een hele waaier van mogelijkheden. Dit CODEAU-rekenmachine is gebaseerd op de HP-calculator. Verder heeft CODEAU enkele functie's voor de export van reeksen welke toelaten om gemiddelden, sommen, ... sneller te berekenen. De WISKI-TV-rekenmodule heeft enkele functie's die hierbij een hulp kunnen zijn. Noch de gebruikershandleiding noch persoonlijke testen gaven duidelijkheid over het functioneren van deze functie's. De continue berekening van tijdreeksen kan door geen van beide programma's uitgevoerd worden.

- Behandeling kan grafisch, in de tabel of automatisch. Grafisch moeten punten kunnen versleept worden (in Y-richting). Bij meerdere punten is het handig van verschillende modi te hebben: alle punten *evenveel verschuiven of proportioneel*. Ook moeten punten uit de ene reeks moeiteloos in een andere kunnen gebracht worden, al dan niet proportioneel of enigszins verschoven in de tijd *<reeks kopiëren>*. Bij manuele behandeling in de tabel kan voor elk punt een waarde ingevoerd worden. Eventueel kan een geselecteerde reeks automatisch lineair of via spline gevuld worden *<automatisch vullen>*. Een universele *reeks-rekenmachine* (cf. supra) vervolledigt de tools. Hierbij kunnen waarden uit reeksen geselecteerd, aan formule's onderworpen, reeksen vergeleken, correlatie's berekend, ... door middel van *logische, mathematische en evt. statistische operatoren*. Extra functionaliteit kan bekomen worden door de variabelen ook tijdsafhankelijk te maken. CODEAU kan waarden enkel via het klavier wijzigen. In de grafische modus van CODEAU dient hiertoe een punt geselecteerd waarna de eventuele nieuwe waarde of code via het klavier moet ingebracht worden. Verschuiven van punten en behandeling van meerdere punten kan enkel door met behulp van de rekenmodule een nieuwe reeks te berekenen, waarna deze nieuwe reeks in de oude kan gekopieerd worden. Ook opvullen van een reeks dient via de rekenmachine te gebeuren, wat niet zo eenvoudig is. WISKI bevat een veel uitgebreidere grafische interface waarbij het mogelijk is de reeks op het scherm te verslepen, al dan niet proportioneel, reeksen te kopiëren, waarden te wissen, in te voegen... Tegelijkertijd blijft het mogelijk om waarden via het klavier in de tabel te typen. Statistische functies zijn in WISKI beperkt aanwezig, maar functioneren niet (onvoldoende?). Kisters stelt dat in de "stand-alone" versie verschillende functies niet actief zijn, deze zouden bij het volledige WISKI-pakket wel actief zijn (mondeling mededeling A. Matamala, Kisters AG).

5. Controle

Uiteindelijk moeten de effecten van de verbetering gecontroleerd worden. Ook hierbij kan het handig zijn over enkele tools te beschikken zoals *<berekening minima en maxima>*, *<berekening durlijnen en percentieldebieten>* en *<volume-berekening>*. Deze controle hoeft niet per definitie deel uit te maken van de reeksverwerkings-software. CODEAU bevat een uitgebreide set aan tools voor de verdere analyse van de reeksen, zoals de bepaling van waterbalansen op basis van de afvoer- en de neerslagreeks, splitsing van de afvoerreeks in de verschillende afvoer-golven, ... Specifieke tools voor de controle van effecten van wijzigingen zoals durlijnen en percentielen bevat CODEAU noch WISKI-TV. Enkele functie's binnen de WISKI-rekenmodule laten wel toe bijvoorbeeld minima en maxima te bepalen, spreiding te berekenen, ... Volume-berekening zou theoretisch ook mogelijk zijn, maar functioneert niet in de "stand-alone" versie (cf. supra).

6. Besluit

Essentieel voor de validatie van tijdreeksen is een goede grafische interface. Deze maakt het immers mogelijk fouten te detecteren, een eerste voorwaarde voor validatie. Automatische foutenopsporing kan hierbij een extra hulp betekenen.

Ook de analyse van debietkromme(n) maakt essentieel deel uit van de validatie. Een goede debietkromme is een voorwaarde voor een betrouwbare debietreeks.

Algemeen kan gesteld worden dat een kwalitatieve validatie van tijdsreeksen optimaal mogelijk is met een multifunctionele set van tools. Hoe breder de inzetbaarheid van tools,

(selectie van punten, reken-eenheid, ...) hoe meer mogelijkheden geboden worden aan de 'validator' om de validatie te onderbouwen met cijfers en statistiek, hoe groter ook de betrouwbaarheid van de validatie.

Hoewel de CODEAU-software verschillende functie's beter uitgebouwd heeft (in- en output, zoomfunctie, analyse), lijkt de WISKI-TV software een groot voordeel te hebben van z'n grafische interface die behandeling van de reeks op het scherm toelaat, waardoor voornamelijk de snelheid en de eenvoud van de reeksverbetering aanzienlijk stijgen. Ook de WISKI-rekenmodule heeft verschillende functie's welke niet door CODEAU aangeboden worden en een uitgebreidere analyse en berekening van afgeleide reeksen mogelijk is. Verschillende van deze functie's werken niet in de "stand-alone" versie, waardoor een evaluatie hiervan moeilijk is.

Tot slot dient opgemerkt dat deze nota enkel handelt over de validatie van peil- en debietreeksen en niet over de validatie van neerslagreeksen. Vele aspecten zijn gelijkaardig, maar door de unieke eigenschappen van neerslag en de methoden waarop neerslagdata verzameld worden is een eigen aanpak en extra functionaliteit nodig.