



ORKUSTOFNUN

**Gagnavinnslukerfi vatnamælinga
Orkustofnunar og Landsvirkjunar. Varðveisla
á tímaraðagögnum**

Tómas Jóhannesson

Greinargerð TJ-92-04

GAGNAVINNSLUKERFI VATNAMÆLINGA ORKUSTOFNUNAR OG LANDSVIRKJUNAR

Varðveisla á tímaraðagögnum

1. INNGANGUR

Til varðveislu á tímaraðagögnum í úrvinnslukerfi Vatnamælinga Orkustofnunar og Landsvirkjunar verður notað bandarískt skrávinnslukerfi sem er sérstaklega hannað til geymslu á hvers kyns mæligögnum. Kerfið nefnist *netCDF* og er þróað af veðurfræðirannsóknarstofnuninni UCAR (University Corporation for Atmospheric Research) í Boulder í Colorado. Kerfið hefur verið í þróun síðustu 3-4 árin og hefur notkun þess aukist hratt á síðustu 1-2 árum. Er það nú m.a. notað til þess að geyma og vinna með mæligögn frá gervitunglum, hafrannsóknarbúum, veðurfræðilofbelgjum o.m.fl., auk reikni- niðurstaða úr straumfræði- og veðurfræðilökum. *netCDF* er notað í mörgum rannsóknastofnunum og háskólum í Bandaríkjunum auk UCAR og má þar meðal annarra nefna USGS (Bandaríska jarðfræðistofnunin), NASA (Bandaríska geimferðastofnunin) og NOAA (bandarísk hafrannsóknastofnun).

Helstu kostir *netCDF* eru eftirfarandi:

1. Kerfið keyrir á mörgum tölvutegundum og er notkun þess eins á þeim öllum. Meðal tölvutegunda sem *netCDF* keyrir á má nefna IBM PS/2 (MSDOS, OS/2), Sun 3 (SunOS), Sun SPARC (SunOS), DEC VAX (VMS, Ultrix), DECstation (Ultrix), IBM RS6000 (AIX), HP9000/700 og 800 (HP-UX) og CRAY YMP (UNICOS).
2. Gögn eru geymd í stýrikerfisskrám á svokölluðu XDR formi og hafa skrárnar sömu uppbyggingu á öllum tölvutegundum sem *netCDF* keyrir á. Gögnin má því lesa og skrifa yfir net og hægt er að flytja gagnaskrár milli tölva án þess að hafa áhyggjur af þýðingum. XDR formið nýtir diskpláss tiltölulega vel

og taka gögnin svipað pláss og ef þau væru geymd í hráum (unformatted) stýrikerfisskrám.

3. Aðgangur að gögnum er um sérstakt undirforritasafn sem tengist bæði C og FORTRAN forritum. Einnig fylgja kerfinu nokkur forrit sem notuð eru til þess að skoða gagnaskrár, þýða þær frá XDR formi yfir á sérstakt ASCII form og öfugt og búa til á grundvelli skrálýsinga C eða FORTRAN forrit sem skrifa *netCDF* skrár.
4. Gagnauppbygging er sveigjanleg og hentar vel fyrir hvers kyns mæligögn. Breytur má skilgreina sem 1, 2 og 4 bætu heiltölur og 4 og 8 bætu fleytitölur, auk strengja. Hægt er að skilgreina vektora af einni eða fleiri breytistærðum á þægilegan hátt.
5. Hver gagnaskrá geymir upplýsingar um þau gögn sem í skránni eru auk gagnanna sjálfra. Þessar upplýsingar eru geymdar fremst í skránni í e.k. færslu- lýsingu. Mögulegt er að skrá mæliein- ingar gagna, skilgreina sérstakt gildi sem merkir að gögn vanti, útskýra upp- runa gagnanna og gera grein fyrir þeirri úrvinnslu sem fram hefur farið.
6. Aðgangur að gögnum er hraður, þ.e. lestrar- og skrifarhraði er svipaður og ef gögnin væru geymd eru í hráum (unformatted) stýrikerfisskrám.
7. *netCDF* kerfið er ókeypis og er því dreift sem forritstexta. Notendur geta breytt kerfinu ef þeim hentar.

2. UPPBYGGING *netCDF* SKRÁA

netCDF skrár geyma eins og nefnt er að fram- an annars vegar *upplýsingar um gögnin* sem í skránni eru og hins vegar *gögnin sjálf*. Að auki er hægt að skrá ýmis konar athugasemdir

um gögnin í svokölluð *einkenni*. Upplýsingum um gögnin má skipta í tvennt. Annars vegar eru skilgreiningar á víddum (dimensions) en þær eru nefndar ákveðnum nöfnum og notaðar við skilgreiningar á breytum. Hins vegar eru skilgreiningar á breytunum sjálfum. Breytur í netCDF skrá geta verið af 6 mismundandi grunngerðum:

Nafn	Stærð (bætur)
char	1
byte	1
short	2
long	4
float	4
double	8

Breytur eru annað hvort skalarbreytur og hafa þá engar víddir eða þær eru skilgreindar með einni eða fleiri víddum.

Hefð er fyrir því meðal netCDF notenda að nefna breytur sem skilgreina gildi á hnitum hnitakerfa eða tímasetningu mælinga með sama nafni og viðkomandi vídd. Þannig svarar t.d. víddin "time" til tímabreytunnar "time" og tilsvareandi mælistærð er skilgreind $Q(\text{time})$ ef um rennslismælingar er að ræða.

Auk vídda og breyta er hægt að skilgreina einkenni, sem geta ýmist verið strengir eða tölur, til þess að geyma ýmsar athugasemdir um gögnin. Einkenni eiga ýmist við ákveðnar breytur eða eru almenn (global). Eftirtalin einkenni hafa staðlaða merkingu fyrir breytur í netcdf skrá:

Nafn	Merking
long_name	Full nafn á mælistærð (t.d. "Rennsli")
units	Mælieining (t.d. "m ³ /s")
scale_factor	Margföldunarstuðull
add_offset	Hliðrun
_FillValue	Skráð þegar gildi vantar

netCDF kerfið gefur kost á því að geyma gögn sem hafa tiltölulega lítið mælisvið þannig að geymslupláss nýtist mjög vel (gögnin eru í raun geymd sem 1 eða 2 bæta heiltölur eftir mælisviði). Þá eru skráðir með gögnunum margföldunarstuðull (einkennið *scale_factor*) og hliðrun (einkennið *add_off-*

set) og í stað mæligagnanna sjálfra eru geymdar (heilar) tölur sem hliðrunin hefur verið dregin frá og deilt í með margföldunarstuðlinum (þ.e. $(y-o)/s$ er geymt í stað y , þar sem y er gildi sem geyma þarf, o er hliðrun og s er margföldunarstuðull). Þegar gögnin eru lesin má á sjálfvirkan hátt umreikna tölurnar sem geymdar eru á disknum yfir í sjálfar mælitölurnar og öfugt þegar gögnin eru skrifuð út á disk til varðveislu.

Fylligildi (einkennið *_FillValue*) er skráð til marks um að gildi viðkomandi mælistærðar sé óþekkt. Fylligildi fyrir hverja gerð breytistærðar eru sjálfgefin í netCDF kerfinu og þarf því ekki að gefa upp einkennið *_FillValue* nema notandi þurfi að nota annað gildi en hið sjálfgefna.

Auk breytueinkenna er mælt með því að almenna einkennið *history* sem tilgreinir uppruna og úrvinnslu gagnanna sé skilgreint í sérhverri netCDF skrá.

3. CDF- OG CDL-SKRÁR

netCDF skráavinnskluerfinu fylgja forrit sem breyta netCDF skráum yfir á skiljanlegt ASCII form. Skrár á netCDF formi eru nefndar cdf-skrár ("cdf" er skammstöfun fyrir "Common Data Form") og ASCII form þeirra kallast cdl-skrár ("cdl" er skammstöfun fyrir "Common Data form Language"). cdl formið er þægilegt að nota til þess að skilgreina netCDF skráaupbyggingu fyrir ný gögn og einnig til þess að skoða innihald cdf-skráa. Þegar cdl-lýsing fyrir ákveðin gögn hefur verið samin má nota forrit sem fylgja netCDF kerfinu til þess að búa á sjálfvirkan hátt til C eða Fortran forrit sem skrifa út cdf-skrána. Í viðauka aftast í þessari greinargerð er útprentun á cdl skilgreiningum fyrir tímaraðir sem lýst er hér að neðan.

4. MÆLIEININGAR

Með netCDF skráavinnskluerfinu er hægt að fá undirforritasafn sem nefnist *udunits* og vinnur með mælieiningar. Þetta undirforritasafn er hentugt til þess að breyta um einingar á mæligögnum og viðhalda á sjálfvirkan hátt réttum mælieiningum þegar keyrð eru úr-

vinnsluforrit sem reikna afleiddar stærðir svo sem summur.

5. TÍMARADAGÖGN

5.1 Almenn

Hver tímaröð geymir ákveðna mælistærð eða afleidda stærð, t.d. vatnshæð eða rennsli. Tímaröðinni verða að fylgja upplýsingar um *mælistað* og *raðnúmer* stærðarinnar á viðkomandi mælistað. Raðnúmerið er notað til aðgreiningar þegar ein og sama mælistærð er mæld með fleiri en einum hætti fyrir sama mælistað, t.d. þegar vatnshæð er skráð bæði á síritablöð og með stafrænum skráningartækjum, eða þegar samnefnd afleidd stærð er reiknuð á nokkra mismunandi vegu fyrir sama stað, t.d. nokkrar gráðudagaraðir fyrir mismunandi gildi á gráðudagaþröskuldi. Koma þarf fram hver *uppruni* gagnanna er, þ.e. hvort um er að ræða *mælda* eða *afleidda* stærð. Einnig þurfa að fylgja tímaröðinni upplýsingar um *gerð* raðarinnar en hún segir til um það hvort líta skuli á gildi raðarinnar sem augnabliksgildi (t.d. vatnshæð) eða hvort þau eiga við lengra tímabil (t.d. meðalrennsli) (sjá nánar að neðan). Að síðustu þurfa að fylgja upplýsingar um *úrvinnslu* gagnanna.

Tímaraðir í úrvinnslukerfinu verða af tveimur aðaltegundum, jafnt- og ójafnt safnaðar raðir. Fyrir jafnt safnaðar raðir þarf að geyma upphafs- og millitíma auk gagnanna sjálfra en fyrir ójafnt safnaðar raðir þarf að geyma tímasetningu með hverju gildi.

5.2 Mismunandi gerðir tímaraða

Gerð tímaraðar segir til um túlkun gildanna í röðinni og getur hún verið með þrennum hætti. Í fyrsta lagi að gildin séu augnabliksgildi og að leyfilegt sé að reikna út gildi milli uppgefna gildanna með línulegri brúun (l). Litið er á slíka tímaröð sem samfellt fall og ræðst fjöldi og tímasetning gilda raðarinnar af breytileika stærðarinnar sem liggur að baki tímaraðarinnar. Dæmi um slíkar tímaraðir eru mælingar á vatnshæð eða rennsli. Tímaröðin er teiknuð sem línustrik sem dregin eru milli gildanna í tímaröðinni. Í öðru lagi að litið sé á gildin sem augnabliksgildi en ekki sé leyfilegt að reikna út gildi milli uppgefna

gildanna (p). Í þessu tilfelli eru gildi fyrir aðra tíma en þá sem tilheyra röðinni óskilgreind. Dæmi um slíkar tímaraðir eru stakir aflestrar sem skráðir eru þegar gæslumenn heimsækja mælistaði og notaðir eru til þess að staðfesta þéttari mælingar sem skráðar eru með síritum eða stafrænum söfnunartækjum. Tímaröðin er teiknuð sem stakir punktar. Í þriðja lagi að gildin eigi við allt tímabilið frá síðasta uppgefnum tíma á undan til þess tíma sem fylgir gildinu (s). Dæmi upp slíkar raðir eru meðaltöl, summur, hámark og lágmark, svo og stærðir sem halda óbreyttu gildi milli þess sem gildið breytist í einu stökki, t.d. lokustaða fyrir stíflur. Þessar raðir eru teiknaðar sem stöplarit.

Hægt verður að tilgreina að mæli- eða afleitt gildi sé óþekkt eða óskilgreint með því að gefa upp gildið `_FillValue`, sem hægt er að skilgreina með viðeigandi einkenni (sjá að framan). Þegar um jafnt safnaðar tímaraðir er að ræða þá eru göt í röðinni tilgreind þannig að sérhvert gildi í gatinu er skráð óskilgreint. Göt í ójafnt söfnuðum röðum eru hins vegar nægilegt að tilgreina með því að gefa upp eitt óskilgreint gildi einhvers staðar í gatinu þegar um línulegar raðir (l) er að ræða en í aftari enda gatsins fyrir stöplaraðir (s). Ekki þarf að hafa áhyggjur af götum þegar um staka punkta er að ræða (p).

Þegar um stöplaraðir er að ræða þá er hægt að skilgreina gerð raðarinnar nánar á 6 mismunandi vegu:

Auðkenni	Stöplagerð
s	summa
m	meðaltal
h	hámark
l	lágmark
i	miðgildi
q	"quantile"

Ef stöplagerð er ekki skilgreind þá er litið svo á að um sé að ræða tímaröð með mæligildum sem breytist í stökkum en eru óbreytt þess á milli, t.d. lokustaða fyrir stíflur.

Þegar um "quantile" er að ræða þá þarf að geyma línurnar, sem miðað er við, í sérstöku

einkenni sem nefnist *level* og tekur gildi milli 0 og 1.

Fyrir hásmörk og lágsmörk þarf að vera hægt að geyma tímann þegar hámarkinu eða lágmarkinu var náð fyrir hvert gildi raðarinnar.

Þegar mæld vatnshæð er leiðrétt, t.d. vegna ístruflana, þá þarf að vera hægt að geyma upprunalegu mældu vatnshæðina til þess að hægt sé að endurskoða leiðréttingarnar síðar ef þurfa þykir.

5.3 Uppruni tímaraða, mældar og afleiddar raðir

Gera verður skýran greinarmun á mælingum annars vegar og afleiddum eða útreiknuðum gildum hins vegar. Þetta verður gert með því að tilgreina *uppruna* raða og verða til að byrja með notaðir þrjár flokkar til þess að skilgreina hann. Í fyrsta lagi verða mæligögn og verða rennslistímaraðir taldar þar með þó svo þær séu ekki eiginleg mæligögn vegna þess að notaður er rennslistykill til þess að reikna rennsli út frá vatnshæð. Í öðru lagi verða afleidd gögn sem reiknuð eru með einföldum formúlum út frá öðrum (mældum eða afleiddum) gögnum. Fyrir slík gögn verður geymd formúlan sem síðast var notuð við útreikninga á röðinni. Í þriðja lagi verða niðurstöður eiginlegra líkanreikninga (t.d. niðurstöður úr NAM2 eða HBV). Fyrir slík gögn verður geymt nafnið á líkaninu sem notað var við útreikninga á röðinni.

5.4 Tímaraðir fyrir borholur

Auk tímaraða sem eru ein breytistærð sem eingöngu er fall af tíma eins og lýst er hér að framan, er þörf á því að geyma mæligögn í borholum þar sem fyrir hvern tíma eru mörg mæligildi (t.d. vatnshiti) sem svara til mismunandi dýpis niður borholuna. Slíkar tímaraðir má geyma með sama hætti og að framan greinir fyrir ójafnt safnaðar tímaraðir nema hvað geymt verður dýpi fyrir hvert mæligildi auk tímasetningarinnar. Ekki verður hins vegar hægt að geyma jafnt safnaðar tímaraðir með hliðstæðum hætti og fyrir tímaraðir sem eingöngu eru fall af tíma. Tiltekinn mæliferill niður borholu einkennist þá af því að öll mæligildin hafa sömu tímasetn-

ingu. Tímarað mæligilda fyrir ákveðið dýpi sem fall af tíma verður hægt að fá fram með brúun í dýpi á hverjum mæliferli. Gerð tímaraðarinnar (sjá að framan) mun eftir sem áður segja til um hvernig gildi tímaraðarinnar tengjast í tíma (væntanlega oftast línuleg brúun ("l") eða stakir punktar ("p")) en gert verður ráð fyrir að beita megi línulegri brúun í dýpi til þess að reikna út gildi fyrir dýpi sem falla á milli dýpisgildanna sem mælingarnar eiga við.

6. ATHUGASEMDIR

Tímaraðir mæligilda munu alla jafna geyma eiginleg mæld gildi en þó er þörf fyrir að fylla í göt, leiðréttu augljóslega röng gildi o.fl.þ.h. Eðlilegast er að geyma slíkar viðbætur og leiðréttingar með mældu gildunum og þarf þá að merkja viðbæturnar og leiðréttingarnar sérstaklega til þess að aðgreina þær frá raunverulega mældum gildum. Þetta verður gert með svokölluðum athugasemdum sem einnig verða notaðar til þess að merkja atburði sem máli skipta fyrir úrvinnslu og túlkun gagnanna (t.d. ístruflanir eða jökulhlaup). Athugasemdir verða eingöngu geymdar með mæligögnum en ekki afleiddum gögnum.

Athugasemdir þarf að vera hægt að geyma fyrir hvert mæligildi og verða þær geymdar sem heilar tölur (16 bita) þar sem hver biti tölunnar hefur sérstaka merkingu. Athugasemdir eru oftast þær sömu fyrir mörg gildi í röð og er því eðlilegast að geyma þær þannig að á ákveðnum tíma sé "kveikt á" ákveðinni athugasemd, t.d. ístruflun, og "slökkt á" henni síðar þegar við á. Einnig þarf að vera hægt að skrá punktathugasemdir sem eiga við ákveðið augnablik. Athugasemdir verða því geymdar sem vektor af heilum tölum og tilgreindur tími fyrir hverja þeirra. Merking bitanna í athugasemdum verður sem hér segir:

Númer bita	Merking
1	Kveikt á athugasemd
2	Slökkt á athugasemd
3	Ístruflun
4	Áætlað, mat
5	Áætlað, yfirfært
6	Áætlað, lífkanreikningar
7	Jökulhlaup
8	Ónotað enn sem komið er
9-16	Númer rennslislykils eða númer athugasemdar

Fyrstu 2 bitarnir eru notaðir til þess að segja til um hvort verið sé að "kveikja á" athugasemd (sem tilgreind er nánar með næstu bitum í töluinni), "slökka á" henni eða hvort um augnabliksathugasemd er að ræða (þá er hvorki biti 1 né 2 settur). Næstu 6 bitar tilgreina 5 mismunandi athugasemdir sem geta náð yfir mislöng tímabil eftir hentugleikum. Biti 8 hefur ekki merkingu enn sem komið er. Bitar 9-16 skilgreina númer rennslislykils (1-255) eða númer athugasemdar (1-255). Númer rennslislykla eru notuð til þess að aðgreina rennslislykla þegar margir lykjar eru til fyrir sama rennslismælistað. Númer athugasemdar er notað til þess að vísa til ákveðinnar augnabliksathugasemdar um tímaröðina. Athugasemdin sjálf (t.d. upplýsingar um rennslismælingu) verður hins vegar ekki geymd með tímaröðinni sjálfri. Greint er á milli númera á athugasemdum og rennslislyklum þannig að rennslislyklar eiga ævinlega við tímabil (annaðhvort biti 1 eða 2 er settur) en númeraðar athugasemdir eiga við tiltekið augnablik (hvorki biti 1 né 2 er settur).

Athugasemdin "mat" þýðir að ekki er um beinar mælingar að ræða heldur hefur verið hnitað með sjónmati yfir göt í mælingum eða ístruflun. Athugasemdin "yfirfært" þýðir að upplýsingar um úrkomu, hita, vatnshæð eða rennsli á sama eða nálægum mælistað hefur verið haft til hliðsjónar þegar fyllt var í gat í mælingum eða þær leiðréttar. Athugasemdin "lífkanreikningar" þýðir að eiginlegir lífkanreikningar (t.d. NAM2 eða HBV) hafa verið notaðir til þess að fylla í göt.

Athugasemdin "ístruflun" verður eingöngu notuð þegar um vatnshæðartímaraðir er að

ræða. Athugasemdir um "áætlun" (þ.e. "mat", "yfirfært" og "lífkanreikningar") verður hins vegar hægt að nota fyrir hvaða mælistærð sem er. Þegar ístruflud vatnshæðarmæligildi eru leiðrétt þá verða upprunalegu mæligildin lögð til hliðar (sjá nánar að neðan) en ístruflaða tímabilið heldur athugasemdin þar um. Jafnframt verður viðeigandi athugasemd um "áætlun" sett fyrir það tímabil sem vatnshæðin var leiðrétt á.

Númer rennslislykla verða eingöngu skráð fyrir rennslitímaraðir. Þegar leiðréttar vatnshæðartímaraðir eru notaðar við útreikninga á rennsli þá er bæði skráð númer rennslislykils eins og við á fyrir hvert tímabil og einnig athugasemdir þar sem vatnshæð hefur verið áætluð til þess að fylla upp í göt eða leiðrétt ístruflun. Rennsli, sem áætlað hefur verið með því að áætla fyrst vatnshæð og beita rennslislykli síðan á hana, einkennist þannig af því að í athugasemdum er bæði skráður rennslislykill og einnig hvernig vatnshæðargildin hafa verið áætluð (þ.e. "mat", "yfirfært" eða "lífkanreikningar" eins og lýst er hér að framan). Rennsli, sem áætlað hefur verið beint án þess að vatnshæð sé áætluð fyrst, einkennist hins vegar af því að ekki er skráður neinn rennslislykill í athugasemdum.

Athugasemdin "jökulhlaup" verður hægt að skrá fyrir vatnshæðar og rennslitímaraðir og færast hún yfir á rennsliröðina þegar rennsli er reiknað út frá vatnshæð með rennslislykli.

Númeraðar athugasemdir verður hægt að skrá fyrir hvaða mæligögn sem er.

7. TÍMASETNINGAR

Tími mæligilda verður geymdur sem tímahliðrun frá miðnætti ákveðins upphafs dags (þ.e. kl 00:00). Þannig verður með hverri tímaröð skráður upphafsdagur sem allar tímasetningar í sambandi við þessa tímaröð miðast við. Tímasetningarnar verða að öðru jöfnu í dögum en hægt verður að gefa upp aðrar mælieiningar þegar betur fer á því. Ef tími er geymdur sem 4 bæta fleytitala þá er með þessum hætti hægt að geyma tímasetningar sem spanna 80-240 ára tímabil með 5-15 mínútna upplausn sem telst fullnægjandi

fyrir vatnamælingakerfið.

Tímaraðir mánaða- og árlegra gilda svo og tímaraðir á vatnsársgrunni verður litið á sem jafnt safnaðar raðir þó svo mánuðir og ár séu í raun mislöng og jafnað sé út eða sleppt dögum úr árinu þegar unnið er með vatnsár. Þetta verður tilgreint með sérstakri mælieiningu á tímasetningum fyrir slíkar tímarraðir (sjá nánar að neðan).

8. BREYTUR OG BREYTUNÖFN

Hefð hefur myndast um nöfn á helstu mæli- og afleiddum stærðum sem geymdar eru í netCDF skrá og verður þeim fylgt til þess að halda sem mestu samræmi við aðra netCDF notendur (nöfnin sem merkt eru með stjörnu ("*") er ekki að finna í listum yfir hefðbundin netCDF nöfn en nöfn sem merkt eru með "†" verða hins vegar líklega ekki notuð mikið í vatnamælingakerfinu). Hefð er fyrir því að nota upphafsstafi í nöfnum eiginlegra mælistærða en lágstafi í nöfnum breyta sem geyma tímasetningar o.fl.þ.h. auk sumra reiknaðra og afleiddra stærða.

Nafn	Merking
Q*	Rennsli
W*	Vatnshæð
H*	Útreiknuð vatnshæð
T	Lofthiti
Tmin [†]	Lágmarkshiti
Tmax [†]	Hámarkshiti
SST	Yfirborðshiti sjávar
WT*	Vatnshiti
DD*	Gráðudagar
PRECIP	Úrkoma
EVAP*	Uppgufun
PEVAP*	Gnóttargufun
RH	Rakastig
TD	Daggarmark
SPD	Vindhraði
DIR	Vindátt
P	Stöðvarþrýstingur
PSL	Þrýstingur við sjávarmál
E*	Orkuframleiðsla

Þegar um gráðudagatímaraðir er að ræða þá þarf að geyma gráðudagaþrýskuldinn, sem miðað er við, í sérstöku einkenni sem nefnist *threshold*.

Nöfn á breytum sem eru sérstakar fyrir tímar- aðaskrár vatnamælingakerfisins þarf að velja án stuðnings af netCDF hefðum. Þau eru:

Nafn	Merking
time	Tímahliðrun mæligilda
t0dt	Upphaf- og millitími
depth	Dýpi niður borholu
hltime	Tímahliðrun há/lágmarka
a	Athugasemdir
atime	Tímahliðrun athugasemda
WW	Óleiðrétt vatnshæð
wwtime	Tímahliðrun WW

Í hverri skrá verður oftast ekki geymd nema ein mælistærð. Þó verður hægt að geyma nokkrar mælistærðir (t.d. frá sama mælistað) í einni sömu skrá ef allar aðrar upplýsingar í skránni (þ.e. tímaupplýsingar, athugasemdir, almenn einkenni o.s.fr.) eru þær sömu fyrir allar mælistærðirnar. Þetta hefur einkum þýðingu í sambandi við veðurgögn þar sem e.t.v. er eðlilegast að geyma flestar eða allar mælistærðir fyrir hverja veðurathugunarstöð í einni og sömu skrá.

Engar sérstakar takmarkanir verða á notkun netCDF breytugerðanna char, byte, short, long, float og double í tímaraðaskrá í vatnamælingakerfinu nema hvað breytutegundin char á einungis við strengi og verður því ekki notuð til þess að geyma töluleg gögn.

9. VÍDDIR BREYTIÐÆRÐA

Breytur í netCDF skrá er skilgreindar með ákveðnum víddum og er hefð fyrir því að láta víddirnar heita sömu nöfnum og nöfn breyta sem geyma hnit eða tímasetningar eins og nefnt er hér að framan. Þannig er eðlilegt að skilgreina víddirnar time, t0dt, atime og wwtime sem hver um sig tilgreinir fjölda gilda í tilsvarendi breytum. Víddin time tilgreinir fjölda mæligilda sem jafnframt er fjöldi dýpisgilda ef um borholutímarað er að ræða, t0dt er alltaf 2, atime tilgreinir fjölda athugasemda (lengd breytunnar a) og wwtime tilgreinir fjölda óleiðrétttra vatnshæðargilda (lengd breytunnar WW).

10. EINKENNI

netCDF gefur kost á að skrá einkenni með hverri breytu og einnig almenn (global) einkenni sem eiga við allar breyturnar í ákveðinni netCDF skrá. Rétt er að skilgreina fyrir allar breytur í tímaraðaskrá einkennin *long_name* og *units*, sem nefndir eru hér að framan, og einkennin *scale_factor*, *add_offset* og *FillValue* eftir því sem við á.

Þessu til viðbótar verður hægt að skilgreina eftirtalin *breytueinkenni* með mæli- eða afleiddu breytunni sjálfri:

Nafn	Merking
radnr	raðnúmer mæli- eða afleiddrar stærðar
type	gerð, "l", "s" eða "p"
stype	stöplagerð, "s", "m", "h", "l", "i", "q"
expression	formúla afleiddrar stærðar
model	nafn á líkani

Einkennið *radnr* þarf alltaf að skrá og geymir raðnúmer mæli- eða afleiddu stærðarinnar á mælistaðnum. Einkennin *type*, sem alltaf verður að skrá, og *stype* tilgreina gerð tímaraðarinnar eins og lýst er að framan. Einkennið *expression* geymir síðustu formúlu sem notuð var við útreikninga á breytunni og *model* geymir nafn á líkani sem notað var við útreikninga á breytunni. Aðeins verður hægt að skilgreina annað þessara síðastnefndu einkenna og er undanskilið að viðkomandi breyta sé afleidd stærð (en ekki mæld) ef þau eru skilgreind. Ef hvorugt þessara einkenna er skilgreint þá er gert ráð fyrir að um mæligögn sé að ræða. Að auki verður hægt að skrá önnur breytueinkenni eins og þörf krefur, t.d. einkennið *level* fyrir "quantile" breytur og *threshold* fyrir gráðudagabreytur sem lýst er að framan.

Eftirfarandi *almenn* einkenni verður hægt að skrá fyrir tímaraðir:

Nafn	Merking
fd	upphafs- eða viðmiðunardagur
stadur	tilvísun til mælistaðar
history	ýmsar upplýsingar um uppruna og úrvinnslu

11. MÆLIEININGAR

Hefð er fyrir því að skrá mælieiningar (einkennið *units*) með sérhverri breytistærð í netCDF skrá og verður þeirri reglu fylgt í skráningu tímaraðagagna. *udunits* undirforritasafnið gerir kleift að skrá tímaeiningu með öllum tímasetningum og vinna með þær einingar á þægilegan hátt. Þannig verður hægt að skrá tíma sem sekúndur, mínútur, klukkustundir eða daga eftir hentugleikum. Að öðru jöfnu verður notast við daga á fleytitöluforni.

Fyrir tímaraðir mánaða- og árlegra gilda og tímaraðir á vatnsársgrunni verður *units* skráð "MONTH", "YEAR", "2WEEK" og "WEEK" og verða úrvinnsluforrit að sjá um að bregðast rétt við þessum sértilfellum jafnt safnaðra tímaraða.

12. TENGING VIÐ SPLUS

Teikning, villuleit, leiðréttingar og ýmis önnur úrvinnsla tímaraðagagna í vatnamælingakerfinu verður unnin með teikni- og gagnaúrvinnsluforritinu *Splus*. Gagnavinnsla í *Splus* byggir á svokallaðri hlutbundinni aðferðafræði. Þá eru gögnin sem unnið er með skilgreind í tiltölulega stórum einingum sem nefnast hlutir (objects) og greinast í flokka (classes). Flokkun hluta getur verið á nokkrum þrepum þannig að margar skyldir hlutir teljist til yfirflokks sem hefur ýmsa almenna eiginleika en hann getur greinist í nokkra undirflokka sem hver um sig hefur einhverja sérstöðu. Helstu aðgerðir úrvinnslunnar eru sérsniðnar fyrir þá flokka sem unnið er með. Þessi aðferðafræði einfaldar alla gagnaúrvinnslu til mikilla muna. Í sambandi við tímaraðavinnslu í vatnamælingakerfinu er eðlilegt að skilgreina tímaraðaflokk sem endurspeglar þá uppbyggingu tímaraða sem að ofan er lýst. Með því móti verða tímaraðir jafn eðlilegar einingar til að vinna með og tölur eða vektorar. Þannig verði hægt að gefa skipanirnar

```
q <- q1 + q2 - q3
q <- 5.6*(w-1.0)^1.2
plot(2*q + 1)
```

til þess í fyrsta lagi að leggja saman tvær rennslitímaraðir og draga þá þriðju frá, í

öðru lagi að beita jöfnu til þess að reikna rennsli út frá vatnshæð og í þriðja lagi að teikna rennslisröð eftir að búíð er að margfalda rennslisgildin með 2 og leggja 1 við.

Gagnauppbygging Splus er að því leyti lík netCDF að hægt er að skilgreina einkenni fyrir allar breytur. Eðlilegast er að líta á tímaröð sem lista (búinn til af Splus skipuninni list()) með jafn mörgum stökum og breytur eru í viðkomandi netCDF skrá tímaráðarinnar. netCDF einkenni breytanna yfirferast á eðlilegan hátt yfir í Splus. Breytur og einkenni má nefna sömu nöfnum í netCDF skránni og Splus nema hvað þægilegast er að nefna aðalmælistærð tímaráðarinnar alltaf nafninu "y" í Splus og geyma eiginlegt nafn hennar í netCDF skránni (t.d. "Q", "W" eða "T") í Splus einkenninu "name". Til þess að vinna með tímaráðir í netCDF skrám í Splus þarf að skrifa undirforrit sem tengjast Splus og lesa og skrifa tímaráðir á netCDF formi.

Í Splus er þegar skilgreindur tímaráðaflokkur en hann reynist ekki henta til þeirrar úrvinnslu sem vatnamælingakerfið krefst og þarf því að skilgreina nýjan flokk tímaráða fyrir vatnamælingakerfið. Flokkar eru í Splus tilgreindir með einkenninu "class". Tímaráðir vatnamælingakerfisins (að borholutímaröðum fráöldum) munu tilheyra flokknum "its" sem greinast mun í undirflokkana "uts" fyrir ójafnt safnaðar tímaráðir og "ets" fyrir jafnt safnaðar raðir. Greint verður milli jafnt safnaðra og ójafnt safnaðra tímaráða í netCDF skrám þannig að jafnt safnaðar raðir munu hafa breytuna "t0dt" skilgreinda en ójafnt safnaðar raðir hins vegar breytuna "time".

Grunngerðir breyta í Splus eru færri en í netCDF (þ.e. character, integer, single og double) og þarf að breyta netCDF stærðum af gerðinni byte, short og long yfir í integer í Splus þegar tímaráðir eru lesnar inn og öfugt þegar þær eru skrifaðar út. Í þessu sambandi er rétt að skilgreina Splus einkennið "cdf_data_type" fyrir hverja breytu sem lesin er úr netCDF skrá til þess að hægt sé að skrifa gögnin út á sama formi og þau voru lesin inn á. Ef netCDF einkennin "scale_factor" og "add_offset" eru notuð þá er rétt að láta gerð þeirra segja til um það hvaða

breytu á að skila eftir hliðrun og margföldun þegar gögnin eru lesin og skilgreina Splus einkennið "ret_data_type" til þess að varðveita þessar upplýsingar milli þess sem tímaráðagögn eru lesin og skrifuð.

Ekki verður farið nánar út í lýsingu á tímaröðum í Splus hér, en forritun og prófun á þeim möguleikum sem hér hafa verið nefndir er komin nokkuð áleiðis.

13. PLÁSSPÖRF

Plássfrekustu gögn í gagnasafni Vatnamælinga Orkustofnunar eru tímaráðir af rennsli og vatnshæð (bæði rennslisgæfri vatnshæð og vatnsborði í borholum, lónum og stöðuvötnum). Einnig þarf að gera ráð fyrir að geymdar verði í úrvinnslukerfinu veðurgagnatímaráðir frá Veðurstofu Íslands sem geta hæglega orðið jafn frekar á pláss og vatnafræðilegu tímaráðirnar. Vatnafræðilegu tímaráðirnar þarf að vera hægt að geyma bæði sem augnabliksgildi (nokkur gildi á dag, t.d. vatnshæðarraðir) og einnig sem dagsgildi (t.d. summa rennslis yfir sólarhringinn). Veðurfræðilegu tímaráðirnar frá Veðurstofu Íslands verða væntanlega nær eingöngu geymdar sem dagsgildi (t.d. meðalhiti, hámarkshiti, o.s.fr.). Auk tímaráðagagna verða geymd ýmis hliðlæg gögn um rekstur mæla-kerfisins, staðsetningu mælistaða o.fl.þ.h. Umfang þessara gagna er mun minna en umfang tímaráðagagnanna og plásspörf vegna þeirra sleppt í yfirlitinu hér að neðan.

Til viðbótar við vatnafræðilegar tímaráðir og veðurgögn Veðurstofu Íslands mun á næstu árum verða vaxandi plásspörf á Vatnamælingum OS fyrir veðurgögn frá sjálfvirkum veðurstöðvum OS á hálendinu.

Til þessa hafa vatnafræðilegu tímaráðirnar einkum verið geymdar sem dagsgildi (þ.e. meðalrennsli dagsins og vatnsborð á miðnætti). Í framtíðinni verða auk dagsgildanna geymd hnitúð augnabliksgildi og bendir sú reynsla sem fengist hefur af SKUR hnitunar-kerfinu á OS til þess að gera megi ráð fyrir u.þ.b. 10 hnitúðum gildum á dag til þess að skilgreina breytingar innan dagsins með nægilegri upplausn.

Í núverandi kerfi eru auk dagsgilda geymd meðaltöl og útgildi mánaða og ára. Í nýja kerfinu verða þessi gildi reiknuð út frá tímaröðum yfir augnabliksgildi þegar á þarf að halda, en þau þarf engu að síður að geyma fyrir eldri gögn. Jafnframt þarf að geyma athugasemdir um mæligildi og úrvinnslu. Væntanlega þarf lítið pláss fyrir meðaltöl, útgildi og athugasemdir og er plássþörf fyrir þessi gildi sleppt í yfirlitinu hér að neðan.

Vatnafræðilegu tímaraðarnar á OS skiptast í raðir frá rennslismælistöðum (u.þ.b. 100 talsins), vatnsborðsmælistöðum (u.þ.b. 20 talsins) og rafstöðvum (u.þ.b. 20 talsins). Fyrir rennslismælistöðvar þarf að geyma bæði rennsli og vatnshæð, fyrir vatnsborðsstöðvar þarf aðeins að geyma vatnsborð en fyrir rafstöðvar eru geymd nokkur mæligildi (líklega 7 gildi). Væntanlega verða eingöngu geymd dagsgildi fyrir rafstöðvar.

Augnabliksgildi verða geymd með mislöngu tímabili milli mæligilda og verður bilið milli gilda valið með hliðsjón af breytileika mælinganna á hverjum tíma. Hverju augnabliksgildi þarf því að fylgja tímasetning. Fyrir dagsgildi og önnur jafnsöfnuð gildi þarf hins vegar eingöngu að geyma upphafstíma mæliraðar en sæti mæligildis í röðinni segir til um tíma þess. Gera má ráð fyrir að 4 bætu pláss þurfi til þess að geyma hvort heldur sem er tíma eða mæligildi. Þannig þarf alls 16 bætur af geymsluplássi fyrir hvert augnabliksgildi vatnshæðar og rennslis frá rennslisstöðvum (tími, rennsli + tími, vatnshæð) en 8 bætur fyrir dagsgildi (rennsli og vatnshæð). Fyrir hvert augnabliksgildi frá vatnsborðsstöðum þarf 8 bætur (tími, vatnsborð) en einungis 4 bætur fyrir dagsgildin. Fyrir hvert dagsgildi frá rafstöðvum þarf væntanlega um $7 \cdot 4 = 28$ bætur (rafstöðvar verða geymdar sem nokkrar sjálfstæðar tímaraðir en það hefur ekki áhrif á plássþörfina).

Ekki er fjarri lagi að um 20 ár af dagsgildum séu að meðaltali til í gagnasafni Vatnamælinga fyrir þessar tímaraðir (heldur minna fyrir vatnsborðsstöðvar og rafstöðvar). Í yfirlitinu hér að neðan verður miðað við plássþörf til næstu 10 ára þannig að geyma þurfi bæði eldri mæligögn og þau gögn sem inn koma á

næstu 10 árum. Eitthvað verður endurnhitað af eldri mæligögnum eftir að úrvinnslukerfi sem býður upp á aukna upplausn í tíma verður tekið í notkun. Til þess að hafa eitthvað við að miða verður í yfirlitinu hér að neðan gert ráð fyrir því að 10 ár af eldri mæligögnum verði endurunnin (það er væntanlega full mikið í lagt en ætti í öllu falli ekki að leiða til vanmats á plássþörfinni). Samkvæmt þessu þarf kerfið að geta geymt 30 ár af dagsgildum og 20 ár af augnabliksgildum með hærri tímaupplausn.

Veðurfræðilegu gögnin eru frá um 140 veðurfarsstöðvum og er gert ráð fyrir að geyma þurfi 8 dagleg mæligildi fyrir hverja stöð í að meðaltali 40 ár.

Eftirfarandi tafla sýnir plássþörf vegna vatnafræðilegra gagna og veðurgagna frá Veðurstofu Íslands í nýja úrvinnslukerfinu í megabætum skv. ofangreindum forsendum.

Gögn	Augnabliksgildi	Dagsgildi	Alls
Rennslisstöðvar	117	9	126
Vatnsborðsstöðvar	12	1	13
Rafstöðvar	0	6	6
Veðurgögn	0	65	65
Samtals	129	81	210

Ef gert er ráð fyrir því til einföldunar að sjálfvirkar veðurstöðvar OS skrái 5 breytistærðir á klukkustundar fresti allan ársins hring þá er plássþörf vegna hvernar slíkrar stöðvar u.þ.b. 2 megabæti fyrir hvern áratug sem stöðin er rekin.

Samkvæmt töflunni taka tímaraðagögnin umtalsvert pláss og verða væntanlega mun meiri um sig en nokkur önnur gögn sem geymd eru á tölvum Orkustofnunar að frátöldum gögnum landupplýsingakerfisins Arc/Info þegar það verður komið í notkun. Umfang gagnanna er þó fjarri því að vera til vandræða. Gögnin munu öll komast fyrir á innan við helmingi af Vatnamælingahluta harða disksins sem Vatna- og Borholumælingar Orkustofnunar keyptu í sameiningu í fyrra. Gögnin munu jafnframt auðveldlega komast fyrir á

1 DAT-spólu eða á 1-2 segulbandspólum ($\frac{1}{2}$ '' band, 6250 bpi) þannig að afritataka og öryggisvarsla verður auðveld.

Margföldunarstuðulinn (`scale_factor`) og hliðrunina (`add_offset`) sem skilgreina má fyrir netCDF breytur má nota til þess að minnka plássþörf í sambandi við veðurgögnin um ca. helming frá því sem reiknað er með hér að ofan. Mælisvið rennslisgagna er of mikið til þess að hægt sé að geyma þau á þennan hátt og tæpast er ástæða til þess að geyma vatnshæð með þessum hætti heldur.

14. VIÐAUKI I — CDL-SKRÁ FYRIR ÓJAFNT SAFNAÐAR TÍMARÆÐIR

```

// netcdf skilgreining fyrir sjafnt safnapar tmmarapir
// 9.7.1992, tj

netcdf uts {

dimensions:
    time = 10;

variables:
    float Q(time);
        Q:long_name = "rennsli";
        Q:units = "m^3/s";
        Q:radnr = 1;
        Q:type = "s";
        Q:stype = "h";
    float time(time);
        time:long_name = "tmmahliprun";
        time:units = "day";
    float htime(time);
        htime:long_name = "tmmahliprun ha/lagmarka";
        htime:units = "day";
    :fd = 1992, 7, 9;
    :stadur = 12;
    :history = "~etta er cdl skilgreining a netcdf skra fyrir0,
        "sjafn safnapar tmmarapir0,
        "Hver stfrp (Q hir ap ofan) hefur akvepip nafn0,
        "sem asamt rapnzmerinu segir einhlmtt til um 0,
        "stfrpina sem um er ap rfpa.0,
        "Nokkrar aprar breytur sem hafa fastakvepin nvfn0,
        "(time,t0dt,depth,htime,a,atime,WW,wwtime)0,
        "geyma vipbstaruppl}singar um tmmasetningu o.fl.~h.0,
        "Hfgt er ap geyma nokkrar stfrpir m svmu skra ef0,
        "vipbstaruppl}singarnar eiga jafnt vip um ~fr allar.0,
        "Vmddina time (=fjvldi gilda m tmmarvpinni) ~arf alltaf0,
        "ap skilgreina, en atime bara ef athugasemdir fylgja.0,
        "Breytunum htime, a, atime ma sleppa ef vill.0,
        "Almennu einkennin stadur, fd og0,
        "history verpur alltaf ap skilgreina:0,
        " stadur: nzmer mflistapar,0,
        " fd:      vipmipunardagur,0,
        "          tmmahliprun mipast vip mipnftti ~ennan dag,0,
        " history: almennar uppl}singar.0,
        "Breyturnar geta allar verip af hvapa lvlegri netcdf0,
        "gerp sem vera skal nema char, ~.e. byte, short, long,0,
        "float epa double (nema a sem alltaf er long).0,
        "Breytueinkennin long_name og units verpur alltaf ap0,
        "skilgreina, nema fyrir a sem er einingalaus breyta.0,
        "Breytueinkennip radnr er notap til ~ess ap apgreina0,
        "mismunandi mfli- epa afleiddar stfrpir a sama stap.0,
        "Breytueinkennin type og stype skilgreina0,
        "gerp mfli- epa afleiddrar stfrpar:0,
        " type: l (lmnuleg brzun), p (stakir punktar),0,
        "          s (stvplabrzun fra vinstri)0,
        " stype: s,m,h,l,i,q (summa,mepaltal,hamark,lagmark,0,
        "          mipgildi, quantile),0,
        "Auk long_name og units ma fyrir hverja breytu0,
        "skilgreina eftirfarandi einkenni:0,
        " scale factor: margfvldun mep constant,0,
        " add_offset: hliprun (eftir margfvldun),0,
        " FillValue: gildi sem svarar til NA.0,
        "Fyrir mfli- epa afleiddar stfrpir ma0,
        "auk ~ess skilgreina breytueinkennin0,
        " expression: formzla sem reiknapi zt afl. stfrp.0,
        " model:      nafn lmkans sem reiknapi zt afl. stfrp.0,
        "Ap lokum ma fyrir vatnshfparbreytur (W) skilgreina0,
        "sleipritta vatnshfp (WW) mep vmddinni wwtime og tmmunum0,
        "wwtime(wwtime) en ap vpru leyti m samrfmi vip0,
        "skilgreiningarnar ap ofan.0;
}

```

15. VIÐAUKI II — CDL-SKRÁ FYRIR JAFNT SAFNAÐAR TÍMARÆÐIR

```

// netcdf skilgreining fyrir jafnt safnapar tmmarapir
// 9.7.1992, tj

netcdf ets {

dimensions:
    time = 10;
    t0dt=2;

variables:
    float Q(time);
        Q:long_name = "rennsli";
        Q:units = "m^3/s";
        Q:radnr = 1;
        Q:type = "s";
        Q:stype = "h";
    float t0dt(t0dt);
        t0dt:long_name = "tmmahliprun";
        t0dt:units = "day";
    float hltime(time);
        hltime:long_name = "tmmahliprun ha/lagmarka";
        hltime:units = "day";
:fd = 1992, 7, 9;
:stadur = 12;
:history = "^etta er cdl skilgreining a netcdf skra fyrir0,
"jafnt safnapar tmmarapir.0,
"Hver stfrp (Q hir ap ofan) hefur akvepip nafn0,
"sem asamt rapnzmerinu segir einhlmtt til um 0,
"stfrpina sem um er ap rfpa.0,
"Nokkrar aprar breytur sem hafa fastakvepin nvfn0,
"(time,t0dt,depth,hltime,a,atime,WW,wwtime)0,
"geyma vipbstaruppl)singar um tmmasetningu o.fl.~h.0,
"Hfgt er ap geyma nokkrar stfrpir m svmu skra ef0,
"vipbstaruppl)singarnar eiga jafnt vip um ~fr allar.0,
"Vmddina time (=fjvldi gilda m tmmarvpinni) og t0dt=20,
"~arf alltaf ap skilgreina, en atime bara ef athugasemdir0,
"fylgja. Breytunum hltime, a, atime ma sleppa ef vill.0,
"Almennu einkennin stadur, fd og0,
"history verpur alltaf ap skilgreina:0,
" stadur: nzmer mflistapar,0,
" fd: vipmipunardagur,0,
" tmmahliprun mipast vip mipnftti ~ennan dag,0,
" history: almennar uppl)singar.0,
"Breyturnar geta allar verip af hvapa lvglegri netcdf0,
"gerp sem vera skal nema char, ~e. byte, short, long,0,
"float epa double (nema a sem alltaf er long).0,
"Breytueinkennin long_name og units verpur alltaf ap0,
"skilgreina, nema fyrir a sem er einingalaus breyta.0,
"Breytueinkennip radnr er notap til ~ess ap apgreina0,
"mismunandi mfli- epa afleiddar stfrpir a sama stap.0,
"Breytueinkennin type og stype skilgreina0,
"gerp mfli- epa afleiddrar stfrpar:0,
" type: l (lmnuleg brzun), p (stakir punktar),0,
" s (stvplabrzun fra vinstri)0,
" stype: s,m,h,l,i,q (summa,mepaltal,hamark,lagmark,0,
" mipgildi, quantile),0,
"Auk long_name og units ma fyrir hverja breytu0,
"skilgreina eftirfarandi einkenni:0,
" scale factor: margfvldun mep constant,0,
" add offset: hliprun (eftir margfvldun),0,
" FillValue: gildi sem svarar til NA.0,
"Fyrir mfli- epa afleiddar stfrpir ma0,
"auk ~ess skilgreina breytueinkennin0,
" expression: formzla sem reiknapi zt afl. stfrp.0,
" model: nafn lmkans sem reiknapi zt afl. stfrp.0,
"Ap lokum ma fyrir vatnshfparbreytur (W) skilgreina0,
"sleipritta vatnshfp (WW) mep vmddinni wwtime og tmmunum0,
"wwtime(wwtime) en ap vpru leyti m samrfmi vip0,
"skilgreiningarnar ap ofan.0;
}

```

