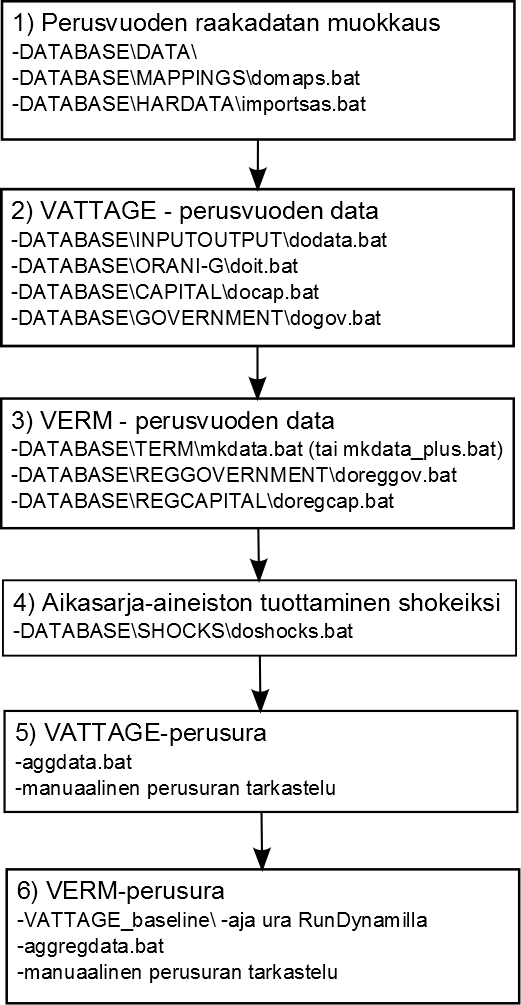
**VATT-mallien tietokanta**

VATT:ssa on käytössä kaksi Suomen taloutta kuvaavaa yleisen tasapainon mallia: koko maan malli VATTAGE ja aluemalli VERM. Mallien aineisto kootaan useasta lähteestä, mutta keskeisimpänä osana voidaan pitää Tilastokeskuksen tuottamia kansantalouden tilinpidon tietoja, jotka kuvaavat malleissa toimivien edustavien toimijoiden[[1]](#footnote-1) välisiä suhteita. Kokoaminen tapahtuu vaiheittain, ja tämän dokumentin tarkoituksena on kuvata prosessi siten, että aineiston kehittäjät ja käyttäjät löytävät siitä työssään tarvitsemansa asiat vaivatta. Tarjonta- ja käyttötaulut ovat ehkä keskeisin osa aineistoa, sillä niiden kuvaamat toimialojen väliset suhteet muodostavat mallien staattisen pohjan. Tätä pohjaa laajennetaan pääomakannan kehitystä kuvaavalla tiedolla, johon mallien keskeisin dynamiikka perustuu. Lisäksi mallit sisältävät yksityiskohtaisen kuvauksen julkisen sektorin toiminnasta, joka perustuu pääasiassa kansantalouden sektoritilinpidon tietoihin. Aluetason aineisto muodostetaan NUTS3-tason alueille eli maakunnille ja siinä keskeisimpänä aineistolähteenä ovat aluetilinpidon tiedot.

Aineiston kehitys on ollut aikaa vievä prosessi. Tämä kuvaus käsittelee prosessia sellaisena kuin se uuteen TOL 2008 toimialaluokitukseen siirtymisen yhteydessä muodostui. Kimmo Marttilan ja Mark Horridgen luoma prosessi, jolla koottiin TOL2002-luokitukseen perustuva aineisto, toimi pohjana päivitystyölle. Aikaisempi prosessi ei kuitenkaan ollut kattavasti dokumentoitu edellisen päivitystyön yhteydessä, minkä havaitsin ongelmalliseksi aineiston johdonmukaisen kehittämisen kannalta. Dokumentin olemassaolo toivottavasti mahdollistaa jatkossa aineiston joustavamman kehittämisen sekä useamman henkilön osallistumisen prosessiin. Tämä dokumentti ei tarkemmin erittele prosessin alkuperäisiä ja uusia vaiheita.

Päivityksen yhteydessä uudeksi perusvuodeksi valittiin vuosi 2008. Tämä sopi tarjolla olleista vuosista perusvuodeksi parhaiten, sillä se on suhteellisen tuore ja kuvastanee myös parhaiten ”normaalia” suhdannetta, johon talouskriisit ovat vaikuttaneet mahdollisimman vähän. Kaikki tarvittavat tiedostot löytyvät hakemistosta W:\Vmalli\MODEL\_REGTECH\.. Hakemistossa on kolme alihakemistoa DATABASE, VATTAGE\_baseline ja VERM\_baseline. Ideaalitapauksessa VATT-mallien perusurat eivät ole keskenään ristiriidassa vaan VERM:ssä käytetty perusura summautuu koko maan tasolla VATTAGEn perusuraan. DATABASE-kansiossa luodaan perusvuoden aineisto sekä VATTAGE:lle että VERM:lle ja osa VATTAGEn ja VERMin perusuralla käytetyistä shokeista. Kansiossa VATTAGE\_baseline luodaan VATTAGEn perusura ja kansiossa VERM\_baseline puolestaan VERMin perusura. VERMin perusuran luomiseen käytetään VATTAGEn perusuran tietoja. Kuvio 1 havainnollistaa prosessia.



Kuvio 1. Perusvuosien aineistojen ja perusuran luominen.

Prosessissa tarvittavat ohjelmat ovat SAS (raaka-aineiston käsittely), GEMPACK-ohjelmisto (HAR-tiedostot ja aineiston testaus) sekä Stat/Transfer (Excel-taulukoiden muuntaminen SAS:lle käyttökelpoiseen muotoon). Lisäksi TERM-vaiheessa tarvitaan Fortran-kokoaja mallitiedostojen kokoamista varten. Alueaineiston muokkaamisessa on käytetty myös R-tilasto-ohjelmaa. VATT:n malleja käytetään GEMPACK-ohjelmistolla, ja käytetty data tulee silloin olla HAR (Header ARray) muodossa. Keskeinen GEMPACK-ohjelma on Tablo, jolla ajetaan aineiston muokkaamiseen käytettäviä tab-tiedostoja. tab-tiedostoille tulee GEMPACK-ohjelmiston mukana oma editori TABMATE, jossa on monia tab- ja cmf-tiedotojen luomista ja muokkausta helpottavia ominaisuuksia. HAR-tiedostoissa data on kirjoitettu moniulotteisiksi matriiseiksi, joita on kätevää lukea ViewHAR-ohjelman avulla. Aineisto ja joukot on HAR-tiedostoissa ryhmitelty headereihin, joista ne GEMPACK-ohjelmilla luetaan. RunDynam-ohjelma tarvitaan VATTAGE:n perusuran luomiseen.

1. ***Raaka-aineisto – DATA-kansio***

VATT-mallien aineiston pohjalla oleva Tilastokeskuksen tuottama raaka-aineisto on tallennettu tämän kansion eri alihakemistoihin SAS-tiedostoina. Ensimmäinen tehtävä aineistoa päivitettäessä on kopioida uudet, päivitetyt tiedostot tähän kansioon. Aineisto koostuu seuraavista tiedostoista:

CAPITALSTOCK-kansio

pääomakanta (poma.sas7bdat)

toimintaylijäämä (toiminta\_ylijaama.sas7bdat)

investoinnit (inv.sas7bdat)

NATACCOUNTS-kansio

tuotantotilit (ttili.sas7bdat)

julkismenot (cofog.sas7bdat)

sektoritilit (stili.sas7bdat)

PRICEDATA-kansio

hyödykehinnat (ttase.sas7bdat)

toimialahinnat (Skt\_data1.sas7bdat)

työllisyystiedot (Skt\_data2 .sas7bdat)

REGCAPSTOCK-kansio

alueellinen pääomakanta (Pomazseutu7505\_k.sas7bdat)

REGSUPUSE-kansio

alueellinen tuotantotieto (tuottyol.sas7bdat)

alueellinen käyttötaulu (kaytto.sas7bdat)

alueellinen tarjontataulu (tarjonta.sas7bdat)

SUPPLYUSE-kansio

tarjonta- ja käyttötaulu (tkt\_2008.sas7bdat)

TAXREGISTER-kansio

verotiedot (verot9511.sas7bdat)

Uusimman aineiston perusvuotena käytetään siis vuotta 2008, ja suurin osa aineistosta onkin olemassa tälle vuodelle joitain poikkeuksia lukuun ottamatta. Poikkeuksen muodostavat aluetason aineiston tarjonta- ja käyttötaulut (kaytto.sas7bdat ja tarjonta.sas7bdat; vuosi 2002; TOL2002) sekä alueellinen pääomakanta (Pomazseutu7505\_k.sas7bdat; vuodet 1975–2004; TOL2002). Alueellisten tarjonta- ja käyttötaulujen tulevista päivityksistä ei ole varmuutta ja alueellinen pääomakanta on tuotettu kertaluontoisesti opinnäytetyönä, joten näiden aineistojen päivittyminen on jatkossakin hyvin epävarmaa. Todellisen aineiston puuttuessa joudummekin turvautumaan erilaisiin second best –ratkaisuihin. Monash yliopiston TERM-mallin aluejako tapahtuu painovoimamenetelmällä, joka luo alueelliset panos-tuotostaulut valtakunnallisista panos-tuotostauluista ja alueellisista tuotos- ja investointitiedoista. Alueellisen pääomakannan päivittämisessä voimme myös soveltaa edellä mainitulla tavalla tuotetun alueellisen panos-tuotosaineiston ja valtakunnallisen pääomakannan tietoja. Edellä lueteltujen aineistojen lisäksi malleissa käytetään tietoa ammattijakaumasta, joka on aikaisemmin saatu FLEED-aineistosta, mutta käsitelty dataprosessissa myöhemmässä vaiheessa. Myöhemmin FLEED-aineisto voidaan ottaa osaksi aineistoprosessia jo tässä vaiheessa.

Toimialatasolla (TOL2008) valtakunnallinen aineisto on olemassa 170 (pääomakanta), 172 (tarjonta- ja käyttötaulu ja hyödykehinnat) ja 182 (loput) toimialan jaoilla. Alueelliset tuotantotiedot ovat puolestaan 173 toimialatasolla saatavissa. Hyödykkeitä aineistossa on 761 (KTTL-luokitus).

1. ***Datan käsittely ja luokitusavaimet – MAPPINGS-kansio***

Mallien keskeisimmät luokitukset ovat toimialat ja hyödykkeet, joiden dimensioissa toimii suurin osa mallin yhtälöryhmistä. Mallien ratkeamisen vaatima laskenta-aika asettaakin keskeisimmän rajoitteen mallissa käytettävän toimiala- ja hyödykeluokituksen koolle. VATTAGEn peruslaskentatasoksi määritimme 93 toimialaa ja 147 hyödykettä. VERMin peruslaskentataso on pienempi – 49 toimialaa ja 49 hyödykettä. Aluemallissa 19 maakunnan kauppamatriisi nostaa mallin ratkaisuaikaa olennaisesti, jonka vuoksi aluemallin koko on luontevaa olla koko maan mallia selvästi pienempi. VATTAGEn laskentatason toimialamäärityksessä keskeisimpänä kriteerinä käytimme ennakointityössä tarvittavaa Opetushallituksen käyttämää toimialaluokitusta. Jotta VATT-malleja voitaisiin joustavasti käyttää peruslaskentatason lisäksi muilla toimiala- ja hyödykekokoonpanoilla, koostamme aineiston mahdollisimman pitkälle ns. MASTER-tasolla, josta se voidaan aggregoida halutun käyttötarkoituksen mukaisille tasoille. Laskentatason aineistokin saadaan vasta prosessin loppuvaiheessa aggregoimalla se MASTER-tason aineistosta. MASTER-tasolla on 242 toimialaa ja 296 hyödykettä.

Mallissa käytettävät toimiala- ja hyödyke-aggregoinnit on kuvattu tiedostossa INDMAPS.xls ja COMMAPS.xls. Näissä tiedostoissa tulee olla mukana sarakkeittain kaikkein disaggregoiduimmasta tasosta lähtien kaikki tarvittavien luokitusten aggregointiavaimet eli mappingit. Kaikkein disaggregoiduimmaksi tasoksi otettiin sekä toimialoilla että hyödykkeillä disaggregoiduin hyödyketaso. Vaikka tämä taso on huomattavasti disaggregoiduinta toimialatasoa disaggregoidumpi, niin kaikki toimialat voidaan periaatteessa jakaa hyödykkeiden tuotantosuhteilla pienempiin osiin. MASTER-tason aggregointiavain on sarakkeessa MASTER ja VATTAGEn peruslaskentatason aggregointiavain sarakkeessa VATTAGE. VERM-mallin laskentatasolle on myös oma avaimensa sarakkeessa VERM. Malleissa käytetään raportointia varten muutamaa pienempää toimiala-aggregointia, jotka ovat 28 ja 10 toimialaa. Jälkimmäinen on kuvattu tiedostossa INDMAPS2.xls. Uusia toimiala- ja hyödykejakoja voidaan jatkossa muodostaa tarpeen mukaan, mutta nykyisiä mapping-tiedostoja tulee muokata harkiten. Tiedostoon INDMAPS.xls ei tulisi koskea uusia aggregointeja tehdessä, vaan uusi aggregointi on hyvä tehdä mieluummin tiedostoon INDMAPS2.xls, jonka sarakkeista prosessissa on tällä hetkellä aktiivisessa käytössä vain viimeinen (10 toimialan aggregointia varten).

Tiedostot muutetaan Stat/Transfer-ohjelmalla xls-muodosta sas-muotoon ennen varsinaisen ohjelman ajamista. Excel-tiedostojen tallentaminen csv-muodossa ennen muuttamista sas-muotoon saattaa olla tarpeen virheiden välttämiseksi.

SAS-ohjelmalla **mappings.sas** muokataan raaka-aineisto mallien kannalta käyttökelpoiseen muotoon ja tehdään aggregointiavaimista TXT-tiedostot. Ensimmäisessä vaiheessa (1. STAGE) eri sas-tiedostojen muuttujien nimet yhtenäistetään. Yksittäisten toimialojen aggregointi tapahtuu tässä vaiheessa, jolloin saadaan seuraavaan vaiheeseen käyttökelpoinen toimialajako. Lähtökohtana on, että raaka-aineiston tyhjät toimialat poistetaan ja tarjonta- ja käyttötauluihin kuulumattomia toimialoja ei esiinny pienemmän toimialamäärän aineistoissa. Tässä vaiheessa toimialoja on 168 (pääomakanta), 172 (tarjonta- ja käyttötaulu ja hyödykehinnat) ja 178 (loput).

Seuraavassa vaiheessa toimialoittainen (2. STAGE) ja hyödykkeittäinen (3. STAGE) aineisto käydään läpi ja ylimääräiset tiedot poistetaan; kansantalouden tilinpidossa on mukana enemmän tietoa kuin mitä malleissa käytetään. Toimialoille luodaan omat joukot (esim. IND172) ja toimialojen TOL:n perustuvat nimikkeet muutetaan mallin käyttämään muotoon, jossa toimialakoodin eteen lisätään I\_ (esim. toimiala 011\_16 eli maatalous muutetaan muotoon I\_011\_16). Toimialasisältö vaihtelee hieman aineistoittain, ja luokitusavain kaikkien TOL-muotoisten nimekkeiden muuttamisesta IND-muotoon on kuvattu tiedostossa TOL2IND.sas7bdat. Hyödykkeille luodaan nimekkeet, jossa KTTL-luokituksen eteen lisätään C\_.

Mallissa osaa hyödykkeistä käytetään marginaaleina muulle tuotannolle (esim. kuljetukset ja kauppa). Jos alkuperäisen aineiston marginaaliluokituksia halutaan muokata (esim. energiamarginaalin luominen ja vesi- ja ilmakuljetusten lisääminen kuljetusmarginaaleihin), niin se onnistuu tässä vaiheessa. KTTL-luokituksen kolme viimeistä hyödykettä (999230, 999240 ja 999320[[2]](#footnote-2)) poistetaan, jolloin saadaan lopullinen suurin hyödykemäärä 758. Myös investointihyödykkeiden määrittely tapahtuu tässä vaiheessa ja niistä luodaan invcoms.sas7bdat-tiedosto. Investointihyödykkeiden joukkoa on mahdollista muokata tarpeen vaatiessa. Toimiala- ja hyödykemappingit sisältävät INDMAPS.sas7bdat- ja COMMAPS.sas7bdat-tiedostot muutetaan tässä vaiheessa txt-muotoon.

Tämän jälkeen tarkistetaan miltä vuosilta aineiston eri tiedostojen havainnot ovat olemassa (4. STAGE). SAS2CSVINP makro muuttaa seuraavassa vaiheessa (5. STAGE) datan csv-muotoon ja luo inp-tiedostot, jotka tarvitaan har-tiedostojen luomiseen CSV2HAR-ohjelmalla (csv-tiedostot sisältävät datan ja inp-tiedostot ohjeet csv-tiedostojen muuntamisesta har-tiedostoiksi). Lopuksi (6. STAGE) ohjelma luo yhteenvedot toimiala- ja hyödyketasoista html-muotoon.

1. ***Datan muuntaminen HAR-muotoon – HARDATA-kansio***

Edellisen askeleen tuotokset tallentuvat tähän kansioon, jossa csv-tiedostot muutetaan har-muotoon mappings.sas-ohjelmassa luodulla komentojonotiedostolla **importsas.bat**. Aineistoa voidaan nyt käsitellä GEMPACK-ohjelmistolla. Tämän askeleen voi jättää väliin jos perusaineisto ei ole muuttunut ja muutokset koskevat ainoastaan uutta toimiala- ja hyödykejakoa, sillä aggregointiavainten muuttaminen HAR-muotoon tapahtuu myöhemmin TXT2HAR-ohjelman avulla. Tilastokeskuksen aineisto tallennetaan HAR-tiedostoina HARDATA-kansioon seuraavasti:

***HAR-tiedosto Tietosisältö Toimialataso Muut luokitukset***

AGGPRI.har aggregaattituotanto - TRS (9), PRICE (2), YEAR (1975-2010)

CAPNAT.har pääomakanta IND168 TRS (5), YEAR (1975-2011), PRICE (2), CNME (12)

CAPREG.har pääomakanta (alue) IND91 (TOL2002) TRS (10), YEAR (1975-2004), REG (21), SEC (15), CRM (2)

COFOG.har julkiset menot - TRS (65), SEC (4), COFOG (65), YEAR (1990-2011)

CPRIBP.har hyödykkeet (perus) IND172+TOT COM (761), SEC (1), UTYPE (14), PRICE (2), YEAR (2008)

CPRIPP.har hyödykkeet (kuluttaja) IND172+TOT COM (761), SEC (1), UTYPE (14), PRICE (2), YEAR (2008)

EMPPRI.har työvoiman käyttö IND174 TRS (2), SRC (13), YEAR (1975-2010)

INDPRI.har tuotanto (1) IND175 TRS (2), SRC (13), PRICE (2), YEAR (1975-2010)

INVCOMS.har investointihyödykkeet - TALTO (12), ICOM (189)

INVDATA.har investoinnit (kaikki) IND177 SEC (17), TRS (3), YEAR (1975-2011), PRICE (2)

INVPRI.har investoinnit (yks.+julk.) - TRS (1), SRC (2), PRICE (2), YEAR (1975-2010)

MARGS.har marginaalit - MAR (18), MTYPE (4)

PRODIND.har tuotanto (2) IND177 TRS (38), YEAR (1975-2011), SEC (8), PRICE (2)

REGSUP.har tarjonta (alue) IND77 (TOL2002) COM (124), REG (21), YEAR (2002), UTYPE (1), PTYPE (3)

REGSUPI.har tarjonta (alue) TOT (2) COM (136), REG (21), YEAR (2002), UTYPE (5), PTYPE (3)

REGUSE.har käyttö (alue) IND77 (TOL2002) COM (127), REG (21), YEAR (2002), UTYPE (1), PTYPE (1)

REGUSEI.har käyttö (alue) TOT COM (143), REG (21), YEAR (2002), UTYPE (13), PTYPE (1)

SECDIAG.har diagnostiikkaa - SEC (25), SOURCE (9)

SECTORS.har sektoritilit - TRS (162), YEAR (1975-2011), SEC (11)

SU2008.har panos-tuotos IND172+TOT COM (761), UTYPE (14), OTYPE (5), TRS (9)

SURDATA.har toimintaylijäämä IND175 TRS (3), YEAR (1975-2011), PRICE (2)

TAXES.har verot - TRS (23), YEAR (1995-2011), SEC (3)

TRSDIAG.har diagnostiikkaa - TRS (182), SOURCE (9)

TUOTYOL.har tuotantotilit (alue) IND173 TRS (7), YEAR (2000-2010), REG (21)

HARDATA-kansioon tallennettavien tiedostojen nimet voivat jostain syystä olla korkeintaan seitsemän merkkiä pitkiä. Pidempien tiedostonnimien käyttäminen aiheuttaa virheen ja kyseiset tiedostot eivät tallennu ollenkaan!

Mappings-kansiossa on myös komentojonotiedosto **domaps.bat**, joka suorittaa MAPPING- ja HARDATA-kansioiden vaiheet yhdellä kertaa.

1. ***Panos-tuotosaineiston luominen – INPUTOUTPUT-kansio***

Tässä kansiossa luodaan kansallisen tason panos-tuotosaineisto edellisessä askeleessa luoduista HAR-tiedostoista. **dodata.bat**-komentojono suorittaa työn ajamalla kansiossa olevat tablo-ohjelmat (**stage1.tab-stage6.tab**), joilla muokataan aineistoa ja tehdään siitä yhdenmukainen seuraavassa askeleessa käytettävän staattisen ORANI-G-mallin kanssa. ORANI-G-mallilla voidaan puolestaan testata aineiston tasapaino ja käyttökelpoisuus staattisessa YTP-mallissa.

Ennen tablo-ohjelmien ajamista kopioidaan tarvittavat tiedostot HARDATA-kansiosta. Mapping-tiedostot muutetaan txt-muodosta har-muotoon TXT2HAR-ohjelmalla.

*Stage 1 – Labeling and sources*

Luettavat tiedostot:

* tarjonta-käyttötaulu (SUTS.har)
* toimiala- ja hyödykeavaimet (INDMAPS.har ja COMMAPT.har; rawdata-kansio)

Tuotettavat tiedostot:

* output-tiedosto (STAGE1.har)
* joukkotiedosto (SETS1.har)
* tarkistustiedosto (dgSTAGE1.har)

Tässä vaiheessa panos-tuotostaulut erotellaan käyttö- (”USE2”) ja valmistusmatriiseihin (”MAKE” ja ”MAKI” tuonnille). Käyttömatriisi on kuvaus toimialojen välituotekäytöstä. Valmistusmatriisi puolestaan kuvaa kuinka eri hyödykkeiden tuotanto jakautuu eri toimialojen kesken, jolloin yksi toimiala voi tuottaa useampaa kuin yhtä hyödykettä ja samaa hyödykettä voidaan tuottaa useammalla toimialalla. Marginaalit erotetaan käyttömatriisista omaksi headerikseen (”MARG”) ja sama tehdään hyödykeveroille (”CTAX”). Tällä tavalla saadaan muodostettua käyttötaulukko, joka sisältää ainoastaan sen välituotekäytön, jota suoraan käytetään hyödykkeen valmistuksessa.

*Stage 2 – Margins and taxes*

Luettavat tiedostot:

* input-tiedosto (STAGE1.har)
* joukkotiedosto (SETS1.har)
* marginaalihyödykkeiden joukot (MARGS.har; rawdata-kansio)

Tuotettavat tiedostot:

* output-tiedosto (STAGE2.har)
* tarkistustiedosto (dgSTAGE2.har)

Tässä vaiheessa tarkistetaan edellisessä vaiheessa muodostettu marginaalien header ja muokataan sitä niin, että se on yhdenmukainen mallin käyttämän aineiston kanssa. Marginaalien käytön tulisi tarjonta-käyttötaulujen suhteen olla määritelmällisesti negatiivisia. Näin ei kuitenkaan aineistossa ole, vaan mukana on myös positiivisia arvoja. Tämä johtuu tilastollisista virheistä ja siitä, että marginaalihyödykkeiden joukossa saattaa olla mukana hyödykkeitä, jotka eivät todellisuudessa ole marginaaleja. Asia korjataan muuttamalla positiiviset marginaalit (MARGSMAR) negatiivisiksi. Muiden kuin marginaalihyödykkeiden (MARGSNON) tulee määritelmällisesti olla positiivisia. Tämä asia voidaan korjata asettamalla negatiiviset arvot positiivisiksi tai antamalla niille mahdollisimman pieni positiivinen arvo (TINY). Toteutimme korjauksen edellä mainitulla tavalla. Alkuperäisessä aineistossa marginaaleja on kolmea tyyppiä: tukkukauppa (WHOLE), vähittäiskauppa (RETAIL) ja kuljetus (TRANSP). Korjattu marginaaliaineisto jaetaan tässä vaiheessa vastaamaan kaikkein disaggregoiduinta käytössä olevaa hyödyketasoa (COM = 758) johon otetaan mukaan myös energiamarginaalit. Marginaalihyödykkeiden käytölle (MARGINS) sekä varastoinnin marginaalille (STOCKMAR) muodostetaan lopuksi omat headerit. Käyttötaulun headeriin (”USE5”) marginaalit päivitetään uudella luokituksella.

*Stage 3a – Investments by industries*

Luettavat tiedostot:

* input-tiedosto (STAGE2.har)
* investointiaineisto (INVS.har)
* pääoma-aineisto (CAPNAT.har)
* joukkotiedostot (INDMAPS.har ja COMMAPS.har; rawdata-kansio)
* investointihyödykkeiden joukot (INVCOMS.har; rawdata-kansio)
* perusvuosi-dummy (BASEYEAR.har)

Tuotettavat tiedostot:

* output-tiedosto (STAGE3a.har)
* output-tiedosto RAS-tasapainotusta varten (FORRAS.har; invests-kansio)
* tarkistustiedosto (dgSTAGE3a.har)

Tässä vaiheessa tuodaan mukaan tiedot investoinneista ja toimintaylijäämästä. Toimintaylijäämän suhteen tieto on valmiiksi sopivassa muodossa ja se täytyy ainoastaan lisätä mukaan. Investointien osalta aineisto on käyttötarkoitukseen nähden vajaampaa. Investoinneista tiedetään ainoastaan kaikkien investointihyödykkeiden kokonaiskäyttö ja eri toimialojen kokonaisinvestoinnit, mutta ei toimialoittaista investointihyödykkeiden käyttöä. Jotta investointien kysyntä toimialoittain ja hyödykkeittäin tulisi mallissa kuvattua mahdollisimman realistisesti, hyödynnämme muutamaa tietolähdettä ja omaa intuitiota aineiston koostamisessa. Yksinkertaisimmallaan voidaan olettaa kaikille toimialoille sama investointikysyntärakenne, mikä ei ole kovinkaan realistista. Päädyimme siksi muokkaamaan aineistoa omaan näkemykseen perustuvan a priori tiedon avulla esim. määrittämällä maataloushyödykkeiden kuten elävät eläimet investointikysynnän olevan peräisin ainoastaan maataloudesta, ja laivojen investointikysynnän keskittyvän vesiliikenteen toimialoille. Tämän lisäksi saimme apua pääoma-aineistosta, joka sisältää tiedot investointien jakautumisesta 12 varaan (CNME-luokitus), josta pystyimme arvioimaan tiettyjen pääinvestointihyödykkeiden, kuten esim. asuin- ja muut rakennukset, jakautumisen toimialoittain.

*Stage 3b – Balancing investments with RAS*

Luettavat tiedostot:

* input-tiedosto (FORRAS.har; invests-kansio)

Tuotettavat tiedostot:

* output-tiedosto (RASSED.har; invests-kansio)

Edellisessä askeleessa luotu investointiaineisto (FORRAS.har) on mahdollisesti päätynyt epätasapainoon ja se tasapainotetaan tässä vaiheessa **rasiter.tab**-ohjelmalla. Tuloksena saadaan RASSED.har-tiedosto. RAS:n komentotiedosto stage3b.cmf sisältää tarvittavat ohjeet tasapainotukselle kuten menetelmän (Euler) ja iteraatioiden määrän (300).

*Stage3c – Investment balance check*

Luettavat tiedostot:

* input-tiedosto (STAGE3a.har)
* investointiaineisto (RASSED.har; invests-kansio)
* joukkotiedostot (INDMAPS.har; rawdata-kansio)
* investointihyödykkeiden joukot (INVCOMS.har; rawdata-kansio)

Tuotettavat tiedostot:

* output-tiedosto (STAGE3c.har)
* tarkistustiedosto (dgSTAGE3c.har)

Tässä kohdassa tarkistetaan investoinneilla täydennetyn aineiston tasapaino. Aineistoon lisätään omat headerit investoinneille (”USEI”) ja investointien marginaaleille (”MARI”). Vuoden 2008 aineistossa oli mukana ongelmallinen suuri negatiivinen investointi risteilyaluksissa (hyödyke C\_301121), joka johtui risteilyaluksen ulosliputuksesta kyseisenä vuonna. Mukana oli myös muutama pienempi, negatiivinen investointi. Nämä disinvestoinnit siirrettiin kyseiselle toimialalle kyseisen hyödykkeen myyntituloksi, jolloin aineiston tasapaino säilytettiin johdonmukaisella tavalla. Erottelimme kaikki negatiiviset investoinnit omaksi headerikseen (”USED”).

Vaiheessa 4 aineistoon muodostetaan uusia toimialoja jakamalla alkuperäisiä toimialoja niiden hyödykeosuuksilla. Tämän vaiheen tarvitsema header kustannusosuuksille (”CSHR”) muodostetaan tablo-ohjelmalla **costshr.tab**, joka ajetaan **stage3c.tab**:n jälkeen. A priori tieto uusien toimialojen kustannusrakenteesta sisällytetään mukaan tässä vaiheessa ja se tulee tehdä muokkaamalla **costshr.tab**-ohjelmaa. Tällä hetkellä ohjelmassa on ainoastaan muokattu paperintuotannon puunkäyttöä siten, että kaikki raakapuu käytetään massan tuotannossa ja massaa puolestaan käytetään paperintuotannon panoksena. Muu välituotekäyttö skaalataan tämän jälkeen siten, että kokonaiskäyttö pysyy alkuperäisellä tasolla. Muiden toimialojen välituotekäyttöä tulisi muokata sitä mukaa kuin relevanttia tietoa löytyy. Syntyvä header ”CSHR” tallennetaan tiedostoon STAGE3cCSHR.har. Myös korkein käytettävä toimialataso (MASTER) määritetään seuraavassa vaiheessa ja sitä varten siirretään mapping omaksi headerikseen ”MIND” tiedostoon STAGE3cCSHR.har.

*Stage 4a – Splitting industries by commodities*

Luettavat tiedostot:

* input-tiedosto (STAGE3cCSHR.har)
* tarjonta-käyttötaulu (SUTS.har)
* joukkotiedostot (INDMAPS.har ja COMMAPS.har; rawdata-kansio)
* investointihyödykkeiden joukot (INVCOMS.har; rawdata-kansio)

Tuotettavat tiedostot:

* output-tiedosto (STAGE4a.har)
* tarkistustiedosto (dgSTAGE4a.har)

Joidenkin toimialojen jakaminen Tilastokeskukselta saatavaa aineistoa tarkemmalle tasolle voi olla tarpeen useissakin sovelluksissa. Jakaminen tapahtuu tässä vaiheessa, ja siinä hyödynnetään aineistoa hyödykkeiden tuotannosta useampaa hyödykettä tuottavien toimialojen kohdalla. Jakosuhteena käytetään siis toimialan tuottamien hyödykkeiden tuotossuhteita. **stage4a.tab** suorittaa jaon toivotulle toimialatasolle, joka tässä tapauksessa on edellä määritelty MASTER-aineiston taso. Maksimimäärä toimialoja on kuitenkin periaatteessa hyödykkeiden määrä eli 758 toimialaa. Ennen jakoa hyödykkeiden määrä tulee aggregoida MASTER-aineiston toimialatasolle, sillä jakaminen onnistuu ainoastaan neliömatriisilla. Hyödykkeille täytyy siis olla olemassa MASTER-tason toimialaluokitusta vastaava hyödykeluokitus valmiina COMMAPS.har tiedostossa. Jakaminen ei automaattisesti muuta jaettavien toimialojen kustannusrakenteita vaan ne jäävät identtisiksi, jolloin myös jaon hyödyt mallintamisen suhteen jäävät olemattomiksi. Jos kustannusrakenteiden erot jaetun toimialan sisällä tiedetään, voidaan ne sisällyttää aineistoon. Tämä tapahtuu edellisessä vaiheessa kuvatussa **costshr.tab**-ohjelmalla, joka tuottaa headerin ”CSHR”. Sen dimensioina ovat kaikki hyödykkeet ja ne toimialat, jotka jakamisen jälkeen muodostuvat uusina (joukko SPLITTEDIND). Tässä headerissä tulisi olla siis uusien toimialojen kustannusrakenteet esitettynä sikäli kun niistä on tietoa olemassa. Tieto kustannusrakenteista panos-tuotostaulukoita tarkemmalla toimialatasolla on vaikeasti saatavilla ja päivitys tapahtuukin lähinnä tapauskohtaisesti tutkimustarpeen ja tiedon saatavuuden mukaan. Tällä hetkellä kustannusosuuksia on muokattu puuta raaka-aineenaan käyttävien paperin tuotannon toimialojen osalta. Toimialoja voidaan jakaa myös hyödyketasoa alemmaksi, mutta tämä vaatii ulkopuolisen aineiston käyttöä, eikä se onnistu suoraan tässä kuvatulla menetelmällä. Jako tapahtuu vaiheittain ja jokaisen vaiheen jälkeen syntyvä STAGE4a.har tiedosto kopioidaan tiedostoksi STAGE3cCSHR.har, joka otetaan uudestaan stage4a.tab:n käsittelyyn. Tätä jatkuu niin kauan kunnes kaikki toimialojen jakamiset on suoritettu. Jos haluaa tietää miltä STAGE3cCSHR.har on näyttänyt ennen stage4a-vaiheen käynnistymistä, niin sen voi tarkastaa tiedostosta STAGE3c\_PRE4.har.

*Stage 4b – Imports data*

Luettavat tiedostot:

* input-tiedosto (STAGE4a.har)
* joukkotiedostot (INDMAPS.har ja COMMAPS.har; rawdata-kansio)

Tuotettavat tiedostot:

* output-tiedosto (STAGE4b.har)
* tarkistustiedosto (dgSTAGE4b.har)

Tässä vaiheessa muokataan valmistusmatriisia siten, että tuotteet, joita ei valmisteta kotimaassa, lisätään mukaan matriisiin tuontituotteina. Valmistusmatriisissa ei voi olla tyhjiä toimialoja, ja niitä on voinut syntyä edellisessä askeleessa tehdyssä toimialojen jakamisessa, jos jakoperusteena käytetyllä hyödykkeellä ei ole ollut kotimaista tuotantoa. Tämä tarkoittaisi, että toimialan koko tuotanto on ulkomaantuontia. Lisätty toimiala tulkitaan tuovan kaiken käyttämänsä välituotekäytön ja muuntavan sen lopputuotteeksi. Valmistusmatriisin (”MAKE”) lisäksi myös muut matriisit (”USEA”, ”USEI”, ”MARG”, ”MARI” ja ”MARS”) tulee päivittää siten, että ne eivät sisällä tyhjiä toimialoja. Tässä vaiheessa muodostetaan myös joukot tuontihyödykkeille ja –toimialoille.

*Stage 5 – Factor payments*

Luettavat tiedostot:

* input-tiedosto (STAGE4b.har)
* joukkotiedostot (INDMAPS.har; rawdata-kansio)
* työllisyystiedosto (EMPPRI.har; hardata-kansio)
* tuotantotiedosto (PRODIND.har)
* perusvuositiedosto (BASEYEAR.har)

Tuotettavat tiedostot:

* output-tiedosto (STAGE5.har)
* tarkistustiedosto (dgSTAGE5.har)
* mapping vanhaan toimialatasoon (käytetään ORANI-vaiheessa) (OLDMAPS.har)

Tässä vaiheessa erotellaan toimialoilla muodostunut arvonlisä eri komponentteihinsa. Arvonlisä lasketaan erottamalla välituotekäyttö tuotoksesta. Jos välituotekäyttö on jollain toimialalla suurempi kuin tuotanto, määritetään arvonlisäykseksi pieni positiivinen luku. Arvonlisän tiedot kirjoitetaan headeriin ”INC”. Arvonlisä jaetaan komponentteihin työvoima ja toimintaylijäämä eli pääomatulot. Jakoon tarvittava aineisto saadaan kahdesta eri lähteestä: panos-tuotostauluista ja kansantalouden tilinpidosta. Panos-tuotostaulujen aineistosta on koottu tähän mennessä käsitelty aineisto. Kansantalouden tilinpidon aineisto haetaan puolestaan HARDATA-kansiosta (työllisyysaineisto EMPPRI.har ja palkkakustannusaineisto PRODIND.har). Vaikka molemmat aineistot sisältävät kuvauksen samasta asiasta, eivät ne ole täysin yhteneväiset. Kansantalouden tilinpidon tiedot katsotaan olevan luotettavammat työvoiman käytön suhteen ja niinpä työvoima muodostetaan tämän aineiston perusteella. Jäljelle jäävä osuus arvonlisäyksestä eli toimintaylijäämä joustaa tässä tapauksessa, ja kahden aineiston välinen erotus siirretään sinne. Sosiaaliturvamaksujen tiedot saadaan myös palkkakustannusaineistosta ja kirjoitetaan headeriin ”SSF”.

Lisäksi tässä vaiheessa on mahdollista luokitella työvoiman käyttö ammattiluokkiin. Aikaisemmissa versioissa luokitteluun käytettiin Tilastokeskuksen FLEED-aineistoa (yhdistetty työnantaja-työntekijä-aineisto), jossa työvoima jaettiin 28 ammattiluokkaan. Uuteen toimialaluokitukseen siirryttäessä ammattiluokituksen päivitys on vielä kesken. Tämä ei ole suuri puute, sillä mallin ammattiluokituslaajennus on niin ikään keskeneräinen.

Ongelmallista on että jotkut toimialat saavat raakadatan perusteella nollan työvoiman käytölle, mutta positiivisen arvon toimintaylijäämälle myöhemmin tapahtuvan jaon perusteella. Tämä ongelma on korjattu seuraavassa askeleessa (stage 6) toimialalle I\_701, jossa toimialojen I\_701 ja I\_702 työpanos jaetaan tasan kummankin kesken niiden toimintaylijäämien suhteessa. Työvoiman käytön vähimmäismääräksi määritellään samalla pieni positiivinen luku.

*Stage 6 – Creation of VATTAGE database*

Luettavat tiedostot:

* input-tiedosto (STAGE5.har)
* joukkotiedostot (INDMAPS.har; rawdata-kansio)

Tuotettavat tiedostot:

* output-tiedosto (STAGE6.har)
* tarkistustiedosto (dgSTAGE6.har)
* aggregointitiedosto (käytetään aineiston aggregointiin halutulle tasolle) (AGGSETS.har)

Tässä vaiheessa edellä muodostettu data kootaan ORANI-G-mallin käyttämään muotoon, jolloin sen homogeenisyys ja toimiminen staattisessa YTP-mallissa voidaan testata. Suurin osa aineistosta on jo valmiina ja se ainoastaan nimetään ORANI-G:n tunnistamalla tavalla. ORANI-G:ssä maa on mukana kolmantena tuotannontekijänä ja se muodostetaan tässä vaiheessa erottamalla maa-alan tuotto (V1LND) toimintaylijäämästä (V1CAP). Maata käyttäviä toimialoja on tässä vaiheessa ainoastaan maatalouden eri toimialat (I\_01111\_4, I\_01116, I\_01118, I\_01135, I\_011371, I\_01411, I\_01412 ja I\_0143\_5), raakapuuntuotanto (I\_02131, I\_02132, I\_02133, I\_02134, I\_02135, I\_02136, I\_02137) sekä turvetuotanto (I\_0892) mutta jatkossa maankäyttö voidaan laajentaa koskemaan myös muita toimialoja kuten esimerkiksi rakentamista. Maatalousmaan osuudet laskettiin Tilastokeskuksen MMYT-aineistosta, jossa maatalouden pääomakanta on jaettu rakennuksiin, koneisiin, osakkeisiin ja osuuksiin sekä maatalousmaahan kahdeksalla eri tuotantosuunnalla. Metsätalouden ja turvetuotannon käyttämän maan arvo taas saatiin tiedostosta CAPNAT.har, jossa varaluokituksessa (CNME) on omat varallisuusluokat maalle. Näiden osuus kokonaisvarallisuudesta otettiin maa-alan tuoton osuudeksi. Saadut osuudet vähennetään pääomatuloista V1CAP. Lisäksi aineisto tarkistetaan niin, että siellä ei esiinny negatiivisia lukuja siellä missä ei pitäisi ja että aineisto on tasapainossa. Negatiiviset luvut viedään varastojen muutokseen. Tämän jälkeen tarjonta ja käyttö tasapainotetaan. Lisäksi mallin tarvitsemille parametreille annetaan arvot.

ORANI-G:ssä on top-down aluejako, joka tarvitsee syötteeksi alueosuuksia. Tässä vaiheessa osuuksille annetaan arvot. Tarvittavat osuudet on lueteltu alla:

Header Tietosisältö Dimensiot Lähde

R001 tuotanto IND\*REG TUOTYOL.har

R002 investoinnit IND\*REG TUOTYOL.har

R003 kotital. kulutus COM\*REG

R004 ulkomaan vienti COM\*REG

R005 julkinen kulutus COM\*REG

R006 varastot COM\*REG

Toimialaluokituksessa olevat osuudet (tuotanto ja investoinnit) tulevat aluetilinpidosta (TUOTYOL.har) ja ne ovat aggregoidummalla toimialaluokituksella kuin mitä MASTER-tason aineistossa. Niinpä stage4a-vaiheessa hyödyketiedoilla jaetut toimialat jakautuvat samoissa suhteissa alueille kuin mitä ne aluetilinpidonaineiston aggregoidumalla tasolla tekevät. Tuotannon alueittaisesta jakautumisesta olevaa muuta tietoa voidaan kuitenkin hyödyntää ja se tapahtuu **stage6.tab**-ohjelman jälkeen ohjelmalla **regshr.tab**. Varsinaiset alueosuudet on koottu kansion ADDREGSHR tiedostoon ADDREGSHRS.har, joka luodaan useasta lähteestä komentojonotiedostolla **regshares.bat**. Joukolla APIND määritetään ne toimialat joiden kohdalta tietoa on olemassa. Tällä hetkellä tietoa on maatalouden eri tuotantosuunnista, fossiilisista polttoaineista sekä paperiteollisuudesta. Maataloudessa kasvinviljelyn toimialat on jaettu alueiden kokonaissatojen suhteen. Kotieläintuotannossa on käytetty eläinyksikkömäärien alueellista jakautumista. Turve on jaettu maakunnille tuotantoalojen ja hehtaarisaantojen avulla. Muut jalostamattomat fossiiliset polttoaineet on jaettu niiden tuontisatamien tuontimäärillä ja maakaasu on kohdistettu Kymenlaaksoon. Paperiteollisuus on jaettu alueittaisilla tuotantokapasiteeteilla. Massan, sanomalehtipaperin ja pehmopaperin osalta tämä tuottaa melko tyydyttävän aluejaon. Hienopaperit sen sijaan sisältävät lukuisia hyödykkeitä, joiden yksikköhintojen tunteminen voisi tarkentaa nykyistä aluejakoa. Käytetty tieto perustuu tyypillisesti johonkin tuotantomääräiseen tekijään ja sopii siksi alueittaisen tuotannon jakautumisen (header ”R001”) korjaamiseen. Investoinneista, jotka tulevat headeriin ”R002”, sen sijaan on vaikeampi löytää sopivaa tietoa, jonka vuoksi olen määritellyt ne samoiksi tuotannon jakautumisen kanssa.

Alueosuuksia käytetään myöhemmin myös TERM-vaiheessa kun aluemallin aineisto muodostetaan sekä jo ORANI-G-vaiheen top down -aluejaossa.

1. ***Aineiston testaaminen ORANI-mallilla – ORANI-G-kansio***

Tässä kansiossa edeltävässä vaiheessa luotu panos-tuotosaineistoa (base2008.har-tiedosto kansiossa INPUTOUTPUT\SAVHARS) testataan staattisen ORANI-G-mallin[[3]](#footnote-3) avulla. Aineistolle tehdään mallin avulla ns. homogeenisyystesti, jossa mallin numeraire-muuttujalle (tässä tapauksessa reaalinen vaihtosuhde phi) annettava shokki (tässä tapauksessa +10 %) tulisi aineiston ollessa homogeeninen aiheuttaa samansuuruisen muutoksen mallin endogeenisissä hintamuuttujissa. Tämä kaikki tapahtuu ajamalla **doit.bat**-komentojonotiedosto, joka käyttää kansiossa olevia tab-tiedostoja. ORANI-G:stä poiketen VATTAGE-mallissa vienti ja tuonti on jaettu EU-maiden ja EU:n ulkopuolisten maiden kesken. Tämän vuoksi SRC- ja DST-ulottuvuudet aggregoidaan summaamalla kaikki tuonnit ja viennit yhteen aivan komentojonotiedoston alussa. Summauksen säännöt on määritelty tiedostossa aggorani.har.

Haetut tiedostot:

* BASE2008.har (edellisessä vaiheessa luotu perusaineisto; INPUTOUTPUT\savhars)
* COMMAPT.TXT (hyödykeluokitukset; HARDATA)

Ajettavat Tablo-ohjelmat:

* trans.tab (datan tarkistus, parametrien lisääminen)
* otest.cmf (homogeenisyystestin ajaminen)
* transage.tab (VATTAGEn tarvitseman datan lisääminen)

**trans.tab**-ohjelma tarkistaa, että aiemmin luodussa aineistossa ei ole negatiivisia lukuja siellä missä niitä ei pitäisikään olla. Se lisää myös edellisessä vaiheessa luotuun aineistoon loput mallin tarvitsemat parametrit kuten joustot ja indikaattoriparametrit yksittäisille vientihyödykkeille (IsIndivExp), ainoastaan paikallisesti tuotettaville hyödykkeille (IsLocCom) ja paikallisia hyödykkeitä tuottaville toimialoille (IsLocInd)[[4]](#footnote-4), sekä etäisyydestä riippuvaisiin marginaaleihin (ts. kuljetukset; ”DMAR”-header). Parametrit haetaan tiedostoista PARAMETERS.har ja REGSUPP.har (rawdata-kansio). Molemmat tiedostot on tehty käsin edeltävissä versioissa käytetyistä tiedostoista, ja arvot on syytä päivittää kun uutta tietoa ilmaantuu. PARAMETERS.har sisältää ORANI-G:n tarvitsemat parametrit, ja ne lisätään perusaineistoon. Mukana ovat Armington-joustot eri tasoille (”1ARM” välituotekäyttöön, ”2ARM” investoinneille ja ”3ARM” kotitalouksille). Vientikysynnän joustolle on kaksi parametria toinen kokonaistasolla (”EXNT”) ja toinen hyödyketasolla (”P018”). Frisch-parametri, eli tulojen rajahyöty kotitalouksien kulutusfunktiossa, luetaan headerista ”P021”. Subsitituutio- ja transformaatiojoustoja luetaan myös (”P028” primäärituotannontekijöiden välillä, ”SCET” lopputuotteiden välillä ja ”SLAB” ammattiryhmien välillä). Kotitalouksien korvattavuusjousto eri hyödykkeiden välillä luetaan headerista ”XPEL”. REGSUPPOLD.har puolestaan sisältää myöhemmin alueellisen TERM-aineiston muodostamisessa käytettäviä parametreja. Näistä ”DIST” kertoo laskennallisen etäisyyden alueiden välillä ja ”PO01” alueiden väestömäärän. Alueiden välinen Armington-jousto luetaan headerista ”SGDD” ja tuonnin maakunnittainen jakauma headerista ”MSHR”. Jälkimmäinen perustuu aikaisemmin TOL 2002 –luokituksen kanssa käytettyihin jakaumiin ja sen päivittäminen on aiheellista kunhan uutta aineistoa vaan saadaan. Marginaalien osalta määritetään niiden marginaalien joukko, jotka ovat todennäköisesti omalla alueella tuotettuja (”LMAR”). Marginaalien välinen korvattavuusjousto luetaan headerista ”SMAR”. Näitä parametreja tarvitaan TERM-vaiheen painovoimafunktiota ja sen tasapainotusta varten. Lopuksi luodaan tiedosto SEC.AGG, joka sisältää joukot ja mappingit TERM-vaiheen aggregointia varten.

**adjMSHR.tab**-ohjelmalla muokataan vielä tuonnin jakautumista maakuntiin (”MSHR” header). Aikaisemmin jakaumamatriisi muodostettiin ilmeisesti maakunnittaisen loppukäytön perusteella, vaikka TERM-mallin aluejaon perusajatuksena on hakea tästä matriisista tuonti sen mukaan minkä alueen kautta se on maahan tuotu. Ohjelma hakee tietoa useasta lähteestä ja muodostaa uuden jakauman tuonnille. Tarvittavat tiedostot ovat kansiossa VATT-MALLIT\_MASTER\_2008\TRADE\_DATA. Kansiossa oleva **trade.bat**-komentojonotiedosto luo tiedoston REGDATA.har, jossa on mukana Liikenneviraston aineisto merikuljetuksista satamittain. REGDATA.har:n perusteella hyödykkeiden tuonti jaetaan maakunnittain. Lisäksi tiedetään merikuljetusten osuus tietyillä hyödykkeillä kaikesta tuonnista sekä muiden kuljetusmuotojen avulla tapahtuneen tuonnin osuus. Nämä ylimääräiset tiedot on koottu käsin tiedostoon MODESHR.har. Rautateitse tapahtuva tuonti menee kaikki Etelä-Karjalan maakuntaan. Tiekuljetusten tuonti taas jaetaan maaraja-asemamaakunnille hyödykkeiden loppukäytön suhteessa (perusdatan header ”R003”). Lentorahdin osalta tuonti on jaettu Finavian aineiston perusteella (header ”ASHR”). Putkikuljetuksissa maakaasu kohdennetaan Kymenlaakson maakuntaan ja sähkö neljään maakuntaan Energiateollisuuden sähköntuontitilaston mukaan (header ”ESHR” vuodelle 2010). Palveluiden tuonti katsotaan jakautuvan kaikkiin maakuntiin niiden hyödykkeiden loppukysynnän mukaan.

**oranig03.tab** on itse mallitiedosto, josta homogeenisyystestiä varten ajetaan kondensoitu versio komentotiedostolla otest.cmf, jossa on määritelty testin closure ja testishokki. Testin onnistumista voi tarkastella otest.sl4 tiedostosta. Kun homogeenisyystesti on suoritettu toivon mukaan onnistuneesti, aineistoon lisätään vielä muita VATTAGE-mallin tarvitsemia parametrejä. Nämä liitetään mukaan **transage.tab**-ohjelmalla. transage.tab-ohjelmassa luodaan myös myöhemmin tarvittavia alijoukkoja hyödykkeille (SERVICE, palveluhyödykkeet) ja toimialoille (IMPIND, toimialat joiden tuotanto on pelkkää tuontia) sekä aikajoukko TIME. Aineisto tallennetaan eri kansioihin VATTAGEa ja VERMiä varten. TERM-kansioon tallennetaan VERM-aluemallin tarvitsemat tiedostot national.har, regsupp.har ja sec.agg. savhars-kansioon puolestaan tallennetaan VATTAGEn perusaineistot VATTBASE2008.har ja WEIGHTS2008.har. VATTAGE-mallin aineiston kokoaminen jatkuu CAPITAL-kansiossa pääomakannan luomisella (vaihe 5) ja VERM-mallin puolestaan TERM-kansiossa aineiston jakamisella alueittain (vaihe 10).

1. ***Pääomakannan luominen VATTAGE-malliin – CAPITAL-kansio***

ORANI-G-malli on staattinen malli ja dynaamista VATTAGE-mallia varten edellisessä vaiheessa luotu ORANI-G:n kanssa yhteensopiva aineisto tulee täydentää tiedoilla pääomakannasta. Tämä tapahtuu CAPITAL-kansiossa olevan **CAPBASE.bat**-komentojonotiedoston avulla.

STAGE 0: Tiedostojen hakeminen ja valmistelu

Haetut tiedostot:

* VATTBASE2008.har (perusaineisto; ORANI-G\savhars)
* WEIGHTS2008.har (parametrien painotettuun aggregointiin tarvittavat painot; ORANI-G\savhars)
* INDMAPS.TXT (toimialaluokitus; HARDATA)
* COMMAPT.TXT (hyödykeluokitus; HARDATA)
* CAPNAT.har (pääomakanta TK:n aineistosta; HARDATA)
* SURDATA.har (toimintaylijäämä TK:n aineistosta; HARDATA)

Ajettavat Tablo-ohjelmat:

* transfer.tab (aggregointitiedostojen luominen)

Valmistelevassa vaiheessa tarvittavat tiedostot haetaan ja ajetaan **transfer.tab**-ohjelma, joka luo aggregointitiedoston SUPS.har AGGHAR-ohjelmaa varten. Data aggregoidaan AGGHAR:lla sille tasolle, jolla pääomakanta on TK:n aineistossa tilastoitu. TOL 2008 luokituksen tapauksessa taso on 168 toimialaa.

STAGE 1: Datan kokoaminen yhteen

Ajettavat Tablo-ohjelmat:

* datacheck.tab (luetaan data)
* trend.tab (pääoman kasvutrendin muodostaminen)
* fixdata.tab (uusien headerien lisääminen aineistoon)
* adjustdata.tab (pääoma-aineiston ja panos-tuotosaineiston yhteensovitus)
* daggextra.tab (EXTRA-tiedoston disaggregointi)
* extraextra.tab (lisätään EXTRA-tiedostoon siitä vielä puuttuvia tietoja)
* extraextra2.tab (lisätään EXTRA2-tiedostoon siitä vielä puuttuvia tietoja)

**datacheck.tab**-ohjelmalla luetaan pääomakannan (CAPNAT.har) ja toimintaylijäämän (SURDATA.har) tiedot ja aggregoidaan ne halutulle tasolle. Pääomakanta on aineistossa tasolla 168 toimialaa ja toimintaylijäämä tasolla 175 toimialaa, jolloin toimintaylijäämän sisältämät tiedot aggregoidaan tasolle 168. Aineistosta haetaan omiksi headereiksi tiedot pääomakannasta, poistoista, investoinneista ja toimintaylijäämästä (nykyisillä ja kiinteillä vuoden 2000 hinnoilla). Nykyarvoisista poistoista huomioidaan kaksi eri arvoa, NATCAP.har ja SURDATA.har tiedostoista omansa. Tuotoksena syntyy CAP1.har tiedosto.

**trend.tab**-ohjelmalla muodostetaan pääomakannan aikasarjasta kasvutrendit toimialoittaiselle pääomakannan kasvulle. Trendi lasketaan halutulle ajanjaksolle, joksi tässä tapauksessa määritimme aineiston viimeiset kymmenen vuotta, 2000-2010. Uudessa toimialaluokituksessa on muutamia toimialoja, joiden tuotanto on vasta alkanut viimeisen kymmenen vuoden aikana. Nämä nollasta lähteneet toimialat ovat ongelmallisia kasvutrendin määrittämisen kannalta. Ongelmaa lievennettiin määrittämällä pääomakannan vähimmäisarvoksi 0,5 miljoonaa euroa. TK:n raaka-aineisto on yhden miljoonan tarkkuudella ja niinpä aineistossa esiintyvät nolla-arvot sijoittuvat todellisuudessa välille . Arvioimme, että alarajan määrittäminen 0,5 miljoonaan euroon antaa uusien toimialojen kasvulle sen alarajan. Ohjelma tuottaa tiedoston TREND.har, joka myöhemmin yhdistetään seuraavassa askeleessa tuotettavaan CAP2.har tiedostoon.

**fixdata.tab**-ohjelma lisää datacheck.tab-vaiheessa luotuun aineistoon uusia parametreja, joista myöhemmin kootaan mallin tarvitsemat EXTRA.har-tiedostot. Pääomakannan kasvulle muodostetaan neljä trendiä edellisessä vaiheessa luodun trendin lisäksi. Trendit lähtevät vuosista 1975, 1990, 1995 ja 2000 ja päätyvät perusvuoteen, joka tässä tapauksessa on siis 2008. Pääomakanta, toimintaylijäämät ja poistot ovat joillain toimialoilla nollia, mikä ei ole toivottavaa mallin kannalta. Nämä korjataan ohjelmassa TEMP- ja FIX-alkuisten parametrien avulla. Pääomakannalle tehdään myös oma header ”VCPM”, joka sisältää keskiarvot pääomakannasta vuoden alussa ja lopussa. Keskeisin mallin pääoman dynamiikkaan vaikuttava parametri on pääoman tuottoaste (rate of return, headerit ”ROR” ja ”ROR2”)[[5]](#footnote-5). Tuottoasteelle lasketaan kaksi vaihtoehtoista arvoa sen mukaan onko toimintaylijäämän laskettu mukaan sekatulo (B3N) (header ”ROR2”) vai ei (header ”ROR”). Investointien ja pääomakannan suhde lasketaan ja kirjoitetaan headeriin ”IKRT”. Ohjelman lopussa luetaan vielä työllisyysdata rawdata-kansiosta (EMPRI.har työvoimakustannukset 174 toimialan tasolla; LABSUP.har työvoimatilastosta henkilötyövuosittain). Työvoimatilaston tietoja muokataan siten, että ne ovat yhdenmukaiset työvoimakustannustietojen kanssa. Muokattu aineisto tallennetaan tiedostoon savhars\WORKFORCE.har, jota voidaan käyttää perusuran luomisessa. Tämän vaiheen jälkeen syntyvään CAP2.har-tiedostoon lisätään aiemmin luotu TREND.har, jolloin saadaan CAP3.har-tiedosto. Tässä vaiheessa on määritetty parametrin arvot reaalikorolle (”RINT”), inflaatioasteelle (”INF”), pääomatuloveroasteelle (”TXKR”) ja kuluttajahintaindekseille (”LCPI”,”LCPL” ja ”LCP2”). Viimeksi mainitut parametriarvot on selvitetty koko trendijaksolle ja tiedot on koottu tiedostoon BOFdata.har (rawdata-kansio). Reaalikorko ja kuluttajahintaindeksi tulevat Suomen Pankin tilastoista. Pääomatuloveroaste määritetään jakamalla korkotulotulojen lähdeveron vuosittainen kertymä (Tilastokeskuksen aineistot) toimintaylijäämän arvolla koko kansantalouden tasolla. Nämä arvot tulee päivittää tapauskohtaisesti uuden aineiston päivityksen yhteydessä.

**adjustdata.tab**-ohjelmalla varmistetaan, että edellisessä vaiheessa luotu dynamiikan kannalta olennainen aineisto sopii yhteen ORANI-G-vaiheessa luodun panos-tuotosaineiston kanssa. Osalla toimialoista (maa- ja metsätalous, rakentaminen) toimintaylijäämään (V1CAP) kuuluu paljon sekatuloa (esim. yrittäjän omalle työlle maksettava palkka), joka tässä vaiheessa allokoidaan työvoimakustannuksiin (V1LAB). Toimialoittainen poistoaste lasketaan tässä vaiheessa ja sitä rajataan niin, että se ei voi ylittää 20 %:ia. Poistoasteet kirjoitetaan headeriin ”DEPR”. Koska tietyt toimialat voivat saada aineiston karkeuden vuoksi poistoasteen arvoksi nollan, korvataan nämä nolla-arvot pitkän aikavälin poistoasteen keskiarvolla. Pääoman tuottoastetta muutetaan tässä vaiheessa siten, että sille asetetaan ylärajaksi 20 %[[6]](#footnote-6). Tämän jälkeen toimintaylijäämä (V1CAP) lasketaan uudestaan uusilla poisto- ja tuottoasteilla. Toimintaylijäämään syntyneet muutokset siirretään ”muihin kustannuksiin” (V1OCT, ”other cost tickets”), jolloin aineiston tasapaino saadaan säilytettyä. Mallinnuksen kannalta ajateltuna tällä tavalla siirretään aineistossa olevia mallin teorian kannalta ongelmallisia puhtaita voittoja talouden resursseja käyttämättömään luokkaan ”muut kustannukset”. Panos-tuotosaineiston sisältämästä tiedosta muutoksia tulee työvoimakustannuksiin (V1LAB), toimintaylijäämään (V1CAP) sekä ”muihin kustannuksiin” (V1OCT)[[7]](#footnote-7). Näillä headereilla korvataan IOAGG.har tiedostossa olevat vastaavat headerit. Uusi perusdata nimetään BASE-2008.har:iksi. Uuden poistoasteen mukaan lasketaan myös pääoman kasvuvauhti K\_GR, jonka muutosten myötä joudutaan muokkaamaan investointien määrää. Seurauksena syntyvät muutokset investoinneissa siirretään varastojen muutokseen. Jotta tämä siirto ei häiritse BKT:n tasapainoa, siirto kohdistetaan ainoastaan kotimaan investointihyödykkeiden kysyntään. Uusien poistoasteiden ja investointien staattisten tuottoasteiden perusteella lasketaan vielä uudestaan pitkän aikavälin ”luontainen” investointien tuottoaste RORN, joka ei päivity mallin perusuralla. Tämän parametrin arvojen järkevyyteen on syytä kiinnittää huomiota.

Edellisen vaiheen tuotoksesta CAP3.har:sta ja sen muokatuista headereista tehdään tiedosto EXTOUT.har, joka toimii pohjana MONASH-mallin EXTRA- ja ROREXT.har-tiedostoille. EXTRA-tiedostoon kirjoitettava pääomakannan kasvun trendi-parametri (”TRNK”) lasketaan tässä vaiheessa. Ajanjaksoa voi vaihdella tarpeen mukaan, ja tässä versiossa se on tehty perusvuotta edeltävän kymmenen vuoden mukaan (1998-2008). CAP3.har kopioidaan vielä kansioon savhars nimellä CAPITAL.har, josta siinä olevia tietoja pääomakannan yms. kehittymisestä voidaan käyttää perusuran rakentamiseen.

**daggextra.tab**-ohjelma disaggregoi EXTOUT.har-tiedoston uudeksi tiedostoksi DAGGEXTRA.har, jossa toimialaluokituksen taso on TOL 2008. Pääomakannan header ”VCAP” jaetaan painotetusti käyttäen disaggregoidun tietokannan VATTBASE2008.har toimintaylijäämää (”1CAP”) painona. Muut toimialatason headerit dublikoidaan, sillä ne sisältävät parametreja.

**extraextra.tab**-ohjelma lisää EXTRA-tiedostoon siitä vielä tässä vaiheessa puuttuvat mallin tarvitsemat headerit lukuun ottamatta muutamaa seuraavassa GOVERNMENT-vaiheessa lisättävää headeria. Hiilidioksidipäästöt sisältävä tiedosto on RAWDATA\CARBONS.har. Tämä tiedosto on tehty käsin käyttäen Tilastokeskuksen ja Öljy- ja Kaasualan Keskusliiton aineistoja ja tulee päivittää erikseen jokaista perusvuotta varten. Muut tarvittavat parametriarvot ovat sellaisia, joita ei ole voitu laskea perusdatasta sinänsä ja ovat lähinnä kovakoodattu. **extraextra2.tab**-ohjelma tekee saman EXTRA2-tiedostolle. Listaus tässä vaiheessa määritetyistä parametreista:

DUM\_YEAR1 – dummy ensimmäiselle vuodelle  
 BETA1 – parametri pehmeälle budjetille (????)  
 BETA2 – parametri pehmeälle budjetille (????)  
 LEV\_CPI\_L – viivästetyt kuluttajahinnat (taso)  
 DEFDIFSUM – kumulatiivinen erotus ensimmäisen vuoden alijäämälle  
 DEFDIFNPV pääomittunut kumulatiivinen erotus ensimmäisen vuoden alijäämälle  
 DISCFACT – diskonttaustekijöiden summa  
 R\_DEFGDP\_BO julkisen sektorin velka prosenttina BKT:sta  
 DIFF – suurin sallittu poikkeama trendikasvusta  
 FEMPADJ – työllisyyden sopeutumisen shifterimuuttuja  
CO2F – polttoainekohtaiset päästöt  
 f\_eeqror – odotetun tuottoasteen shifterimuuttuja  
 LEV\_CPI – kuluttajahintaindeksi (taso)  
 NAIRU – luonnollinen työttömyysaste  
 NATSAVSH – kansallisen säästämisen osuus BKTL:stä  
 PLAB – nimellinen palkkataso  
 PLAL – viivästetty nimellinen palkkataso  
 PSDATTGP0 – julkisen sektorin velka prosenttia BKT:stä  
 COMPRFAC – perusvuodesta kumuloitunut korkotaso  
 RINT – reaalikorkotaso  
 RALPH – RALPH l. voittojen verotuksen parametri  
 RINT\_L – viivästetty reaalikorkotaso  
 RWAGE – reaalipalkkataso (kuluttajahintadeflatoitu)  
 CO2T – hiilidioksidipäästöt (summa)  
 V3\_T – hiilidioksidipäästöt (summa)  
 YEAR – jonkinlainen vuosimuuttuja

STAGE 2: Uuden datan vieminen MONASH-mallin tiedostopohjaan

Tässä vaiheessa ainoastaan kopioidaan tarvittavat tiedostot paikasta toiseen, jonka jälkeen voidaan siirtyä seuraavaan vaiheeseen, jossa aineistoon lisätään julkisen sektorin tiedot.

1. ***Julkisen sektorin aineiston lisääminen VATTAGE-aineistoon – GOVERNMENT-kansio***

Tässä vaiheessa aineistoon tuodaan mukaan julkisen sektorin tiedot. Pohjana toimii ORANI-G- ja CAPITAL-vaiheiden aineistot, joita täydennetään aiemmin HARDATA-kansioon tallennetulla aineistolla. **doit.bat**-komentojonotiedosto ajaa tarvittavat tablo-ohjelmat. Tarvittavat tiedostot:

*Tiedosto Tietosisältö Sijainti*

VATTBASE2008.har panos-tuotos (IND 208) ORANI-G

DAGGBASE2008.har panos-tuotos (IND 208) ORANI-G

AGGBASE2008.har panos-tuotos (IND 168) CAPITAL

BASE-2008.har panos-tuotos (IND 168) CAPITAL

SUPS2008.har panos-tuotos (IND 208) CAPITAL

SECTORS.har sektoritilinpito HARDATA

PRODIND.har tuotantotilit HARDATA

TAXES.har julkisen sektorin verotulot HARDATA

INVDATA.har investoinnit HARDATA

COFOG.har julkiset menot HARDATA

INDMAPS.TXT toimiala-mappingit HARDATA

Ensimmäinen tablo-ohjelma **govscale.tab** kirjoittaa tiedoston GOVSCALE.har, joka sisältää headerit toimintaylijäämälle (”1CAP”) ja työpanoksen (”1LAB”) käytölle. Ohjelmalla muokataan näitä arvoja siten, että ne skaalataan samalle tasolle kuin ne ovat CAPITAL-vaiheen jälkeen tiedostossa BASE-2008.har, mutta mallin käyttämällä MASTER-toimialatasolla TOL 208. Koska muokkaus koski CAPITAL-vaiheessa ainoastaan toimintaylijäämää, työpanoksen käyttö V1LAB\_O ei välttämättä muutu lainkaan.

Seuraavaksi **govdata.tab**-ohjelma kokoaa yhteen tarvittavat tiedot julkisen sektorin lisäämiseksi ja kirjoittaa ne GOVDATA.har-tiedostoon. GOVDATA.har-tiedostossa aineisto kirjoitetaan seuraaviin headereihin:

*Header Tietosisältö Aineistolähde*

CDAT sosiaaliturva SOSTBLE.har (sosiaalitaulut-kansio; ei käytössä)

GDAT julkiset menot COFOG.har

IDAT tuotantotilit PRODIND.har

SDAT sektoritilit SECTORS.har

TDAT verotulot TAXES.har

VDAT investoinnit INVDATA.har

**govdagg.tab**-ohjelma muodostaa ensimmäisen disaggregoidun aineiston, jossa on mukana muun aineiston kanssa tasapainossa oleva julkinen sektori. Lähtökohtana on ORANI-G-vaiheessa luotu perusdata (VATTBASE2008.har), jota täydennetään edellisessä vaiheessa GOVDATA.har-tiedostoon kootuilla tiedoilla. Ohjelma tuottaa kolme uutta tiedostoa GOVDAGG.har, EXTRADAGG.har ja GOVFORC\_SHOCK.har. Kaksi edellistä tiedostoa muodostavat MONASH-tyyppisen mallin tarvitsemat julkisen sektorin syötteet. Viimeinen tiedosto taas sisältää julkisen sektorin menoihin liittyviä ennustetrendejä, joita mahdollisesti käytetään perusuran luomisessa. Tätä tiedostoa ei enää muokata tässä prosessissa.

**powedit.tab**-ohjelma tuottaa uuden tiedoston GOVSECSPLIT.har, joka sisältää tietoa edeltävässä askeleessa luoduista tiedostoista GOVDAGG.har ja EXTRADAGG.har sekä muutaman uuden headerin. Uudet headerit sisältävät power-muotoisia parametrejä palkkaveroista sekä julkisista investoinneista (tasot ja osuudet). Tämä tiedosto tallennetaan sellaisenaan savhars-kansioon ja sitä käytetään myöhemmin julkisen sektorin alueellisessa jaossa (REGGOVERNMENT-kansio; vaihe 10).

**govgeneric.tab**-ohjelma lukee myös GOVDAGG.har ja EXTRADAGG.har tiedostot ja tarkistaa, että ne ovat tasapainossa. Se luo näistä MONASH-tyyppisen mallin tarvitsemat tiedostot GOVACC.har ja GOVEXTRA.har. Ohjelma vaikuttaa keskeneräiseltä. Nykyisessä muodossaan se ainoastaan summaa aineiston yli julkisten sektoreiden. Tasapainoa tarkastellaan jo **govdagg.tab**-ohjelmassa. GOVEXTRA.har yhdistetään lopuksi edellisessä askeleessa luotuihin EXTRA1- ja EXTRA2-tiedostoihin MERGEHAR-ohjelmalla. Tämän jälkeen myös EXTRA-tiedostot sisältävät kaikki VATTAGEn tarvitsemat headerit.

**ADJdata.tab**-ohjelmalla varmistetaan vielä aineiston tasapainon säilyminen CAPITAL- ja GOVERNMENT vaiheessa tehtyjen muutosten jälkeen. Ohjelman lähtökohtana toimii Mark Horridgen kehittämä ADJUSTER.tab-ohjelma[[8]](#footnote-8), joka tasapainottaa aineiston hyödyke- ja toimialataseet sekä päivittää kulujoustot uuden aineiston mukaisiksi. Aineiston tasapaino on hyödyllistä tarkistaa vielä erikseen **balchk.tab**-ohjelmalla. Ohjelma laskee vaiheiden INPUTOUTPUT, ORANI-G, CAPITAL, GOVERNMENT ja ADJUSTER tuottamien mallitiedostojen BKT:t tulo- ja menopuolelta sekä hyödyke- ja toimialataseet. Tällä tavalla mahdollisten epätasapainojen jäljittäminen onnistuu aineistoprosessissa vaiheittain.

1. ***Perusuran shokkien luominen – SHOCKS-kansio***

Tilastokeskuksen aineistossa on paljon perusvuoden jälkeistä aineistoa, jota kannattaa hyödyntää perusuran shokkien luomisessa. Tässä kansiossa olevalla **doshocks.bat**-komentojonotiedostolla kootaan tämä data yhteen ja muutetaan MASTER-tasolle. Shokkeihin tarvittavat tasomuotoiset aikasarjat voidaan siten aggregoida halutulle sovellustasolle muun aineiston aggregoinnin yhteydessä. Aluksi tarvittavat tiedostot kopioidaan HARDATA-kansiosta rawdata-kansioon. Tässä vaiheessa ajetaan seuraavat tablo-ohjelmat:

* maps.tab luokitusavainten luominen
* macros.tab makroaikasarjojen muodostaminen
* empdagg.tab työllisyysaikasarjojen muodostaminen
* comdata.tab hyödykeaikasarjojen muodostaminen
* fixedprices.tab hinta-aikasarjojen muodostaminen hyödykeaineistosta
* industries.tab toimialoittaisten aikasarjojen muodostaminen
* impexp.tab tuonti- ja vientiaikasarjojen muodostaminen
* regdata.tab alueaikasarjojen muodostaminen
* COFOG???
* Mshocks.tab aikasarjojen kokoaminen yhteen

Tuloksena syntyvät tiedostot tallennetaan savhars-kansioon.

1. ***Maksutaseaineisto – BOP-kansio***

Tässä kansiossa tuotetaan tiedosto SAVHARS\TEST.har, joka sisältää tiedot maksutaseesta, jotka malli lukee BOPACC-2008.har-tiedostona. Kaikki TEST.har sisältämä data koostuu skalaareista. Aineiston luominen tapahtuu ajamalla komentojonotiedosto **doit.bat**, joka sisältää perusvuoden valitsemisen ja **bopdata.tab**-ohjelman ajamisen. bopdata.tab-ohjelma lukee tiedoston BALANCEOFPAYMENTS\BOPXLS2HAR.har, joka sisältää maksutasetileistä vuodesta 1997 vuoteen 2008 asti ja kokoaa ne VATTAGE:ssa käytettyyn muotoon. Ohjelma lukee myös mallin perusaineiston ja sektoritilinpidon aineiston, mutta ei käytä niitä prosessissaan. Ohjelman tarkoituksena on myös luoda shokkitiedosto BOP\_SHOK.har perusuraa varten, mutta tältä osin työ on keskeneräinen ja edellyttää myös aineiston päivitystä.

1. ***Sovelluskohtaisen aggregoinnin luominen – aggdata.bat ja MASTER-kansio***

Kun mallin käyttämät kaikki HAR-tiedostot ovat valmiina, voidaan ne aggregoida tarvittavaa käyttötarkoitusta varten **aggdata.bat**-komentojonolla. Yleensä tässä vaiheessa luotu hyvin disaggregoitu aineisto (l. MASTER-taso) ei sellaisenaan sovellu järkevien simulointien ajamiseen vaan sitä pitää aggregoida hyödykkeiden ja toimialojen suhteen. Ensin luodaan alikansio MASTER, johon master-tason tiedostot kopioidaan. Lisäksi tarvitaan tiedosto MASTER.agg, joka sisältää master-tason toimiala- ja hyödykejoukot, aggregoidun aineiston vastaavat joukot, sekä mappingit näiden välillä. Kansiossa ajetaan seuraavat tablo-ohjelmat:

* aggset.tab (MASTER.agg-tiedoston luominen toimiala- ja hyödykeavaimista)
* scoreset.tab (alijoukkojen parametrisointi)
* weights.tab (painot parametrien aggregointia varten)
* preaggBASE.tab (aggregointitiedoston luominen BASE-tiedostoa varten)
* preaggEXTRA1.tab (aggregointitiedoston luominen EXTRA1-tiedostoa varten)
* preaggEXTRA2.tab (aggregointitiedoston luominen EXTRA2-tiedostoa varten)
* preaggGOVACC.tab (aggregointitiedoston luominen GOVACC-tiedostoa varten)
* preaggROREXT.tab (aggregointitiedoston luominen ROREXT-tiedostoa varten)
* preaggSHOCKS.tab (aggregointitiedoston luominen SHOCKS-tiedostoa varten)
* preaggSCORESET.tab (aggregointitiedoston luominen SCORESET-tiedostoa varten)

Mallissa käytetään useita joukkojen alijoukkoja, jotka täytyy ottaa huomioon aggregoinnissa. Tätä varten alijoukot parametrisoidaan ohjelmassa **scoreset.tab**, joka luo tiedoston SCORESET.har, josta malli alijoukot tulee lukemaan. Tasomuotoisten headerien aggregointi on suoraviivaista, mutta erilaisia parametreja ei voi aggrgoida suoraan summaamalla, vaan summausta varten täytyy määrittää painot. Kaikki painoiksi tarvittavat headerit luetaan ja summataan aggregoitavien parametrien tasolle ohjelmassa **weights.tab**, joka tuottaa tiedoston WEIGHTS.har. **preagg**-alkuiset tablo-ohjelmat luovat jokaista aggregoitavaa tiedostoa vastaavan aggregointitiedoston (esim. AGGSUPP-BASE.har BASE-tiedostolle), joka sisältää diaggregoidut ja aggregoidut joukot sekä mappingit niiden välillä, tarvittavat painot sekä headerin ”WAGG”, joka kertoo mitä painoja minkäkin parametrin aggregointiin tulee käyttää. Tämän jälkeen mallin tiedostot aggregoidaan AGGHAR-ohjelmalla. **re-label-base.tab**-ohjelma tarvitaan vielä BASE-tiedostoa varten, sillä painotetun aggregoinnin ”WAGG” headeriin ei voida kirjoittaa numerolla alkavia headereita, joita mallissa on kolme (”1ARM”, ”2ARM” ja ”3ARM”), vaan ne täytyy ensin kirjoittaa jollain muulla nimellä ja nimetä uudestaan tässä vaiheessa. **VATTAGEshk.tab**- ja **COFOGshk.tab**-ohjelmilla lasketaan aggregoiduista shokkitiedostoista vuosittaiset prosenttimuutokset, joita voidaan sitten käyttää perusuran shokkien luomiseen.Mallin aggregointitasoa voidaan näin ollen muuttaa melko joustavasti. Lopuksi valmiit har-tiedostot kopioidaan kansioon VATTAGEDATA.

1. ***Panos-tuotosaineiston jako maakunnille – TERM-kansio***

Aikaisemmissa askeleissa luotu kansallisen tason aineisto (NATIONAL.har) jaetaan maakunnille. mkdata.bat ajaa tämän vaiheen kaikki ohjelmat. mkdata.bat suorittaa jonossa tarvittavat .bat-tiedostot. Jako maakuntiin tapahtuu painovoimamenetelmällä ja tasapainotus RAS-menetelmällä. Koska TERM-tyyppisen mallin aineistot ovat tyypillisesti hyvin laajoja, niiden .tab-tiedostot kootaan ennen ajamista FORTRANilla .exe-tiedostoiksi, mikä nopeuttaa prosessin suoritusta. mkdata.bat on jaettu kolmeen vaiheeseen. Ensimmäisessä vaiheessa syötettävä data tarkistetaan ja valmistellaan aluejakoa varten. Toisessa vaiheessa tapahtuu bottom-up alueaineiston muodostaminen. Kolmannessa vaiheessa aineisto testataan. Vaiheet perustuvat standardiseen TERM-dataprosessiin, joka on yksityiskohtaisesti kuvattu tässä kansiossa sijaitsevassa dataproc.doc tiedostossa.

Vuoden 2008 aineiston jako toteutettiin siten, että tasapainotukseen käytettiin ainoastaan perinteistä RAS-algoritmia (trdras.tab) eikä ollenkaan lineaarista RAS-algoritmia (raslin.tab) jälkimmäisestä johtuvien ongelmien vuoksi. trdras.tab ajettiin läpi Eulerin menetelmällä käyttäen 1000 askelta, joka takasi kohtuullisen konvergoitumisen.

Seuraavissa kahdessa vaiheessa luodaan dynamiikan ja julkisen sektorin kuvaukseen tarvittavat tiedot myös maakuntatasolle. Tässä hyödynnetään mahdollisimman paljon VATTAGE-malliin vaiheissa 5 ja 6 muodostettua valtakunnallista tietoa.

1. ***Julkisen sektorin jako maakunnittain – REGGOVERNMENT-kansio***

Tässä kansiossa julkisen sektorin taloustoimet jaetaan maakunnittain. **doitgov.bat**-komentojonotiedostolla ajetaan tarvittavat ohjelmat. Prosessissa on kolme vaihetta: 1) aineiston hakeminen eri lähteistä, 2) datan nimien yhdenmukaistaminen VERM-tietokannan kanssa sekä 3) aineiston tasapainon tarkastaminen.

Ensimmäisessä vaiheessa haetaan tarvittava aineisto. Komentojonotiedosto **rnsector\_p3.bat** (kansiossa Estimateddata) yhdistää alueellisen momenttiaineiston (tiedostossa PLLUMO2008.xlsx) ja valtakunnallisen julkisten taloustoimien momenttiaineiston (koottu käsin tiedostoon SECACC\_p3.CSV). Näistä kootaan tiedosto PLLUMOREG.har, jonka headerissä on kuvattu alueellinen momenttiaineisto taloustoimittain. Sen koostamisessa on oletettu, että momentit jakautuvat taloustoimiin samassa suhteessa kaikilla alueilla. Väestön maakunnalliset jakaumat ovat POPX5.har-tiedostossa (ESTIMATEDDATA-kansio). Menojen jakautumien julkisen sektorin sisällä eri sektoreihin (valtionhallinto, paikallishallinto ja sosiaaliturvarahastot) saadaan hakemalla tiedosto REGGOV\_2004.txt[[9]](#footnote-9) (..DATA\REGGOV\GOVACCOUNTS-kansio). Lisäksi tarvitaan VATTAGE-mallia varten luotu GOVDATA-2008.har-tiedosto (GOVERNMENT\savhars-kansio) sekä edellisessä vaiheessa luotu TERM-mallin kanssa yhdenmukainen aineisto.

Toisessa vaiheessa ajetaan ohjelma **reggov.tab**. Ohjelma ajetaan HAR-tiedostolle REGGOV.har ja simulointien ohjeet ovat tiedostossa reggov.cmf. Tuotoksena vaiheesta syntyy REGGOV3.har-tiedosto, jossa aineistolle on annettu oikeat labelit, sillä xls2head-ohjelmalla luoduissa HAR-tiedostoissa ei labeleita ole.

Kolmannessa vaiheessa ajetaan ohjelma **govregedit.tab**, joka jakaa VATTAGE-mallin julkisen sektorin aineiston maakunnille. Ohjelmassa tarvitaan seuraavat HAR-tiedostot:

*Tiedosto Kansio Kuvaus*

AGGMOD.har TERM TERM-tietokanta

REGGOV3.har DATA\REGGOV\GOVACCOUNTS Julkisen sektorin aluetilit (perusvuosi)

GOVDATA.har GOVERNMENT Kansantalouden tilinpidon julkinen sektoritili

GOVSECSPLIT.har GOVERNMENT\SAVHARS Julkisen sektorin jako kolmeen

PLUMO.har ESTIMATEDDATA Julkisen sektorin momenttitiedosto

POPX5.har ESTIMATEDDATA Julkiset investoinnit ja kulutus

POPUDATA.har RAWDATA Väestömäärät ja ennusteet

MIGDATA.HAR RAWDATA Julkiset menot alueittain ja ikäryhmittäin

REGEMPL.har RAWDATA Alueellinen työllisyys

GOVDATA.har-tiedostoa tarvitaan veroaineiston jakamiseen halutulle aggregointitasolle. Tiedostosta luetaan sektoritilinpidon aineisto (”SDAT”), cofog-aineisto (”GDAT”) ja verotiedot (”TDAT”), sekä eri veroasteet (”TLAB”, ”TLOB” ja ”TLUB”). REGGOV3.har-tiedosto sisältää alueelliseen jakoon tarvittavaa aineistoa taloustoimien jakautumisesta. GOVSECSPLIT.har tiedostosta saadaan tiedot julkisten sektoreiden välisistä siirroista (”PTRS”) sekä muista julkisista menoista (”OGIS”). POPX5.har-tiedostosta luetaan julkisen sektorin investointeja (”IN01”-”IN21”) ja kulutusta (”IA1”-”IL4”) koskevat tiedot. Näiden tietojen lähde on ollut hieman epäselvä ja julkiset investoinnit on nyt päivitetty siten, että ne luetaan GOVDATA.har-tiedostosta. Tiedostosta saadaan kaikki investoinnit sisältävä ”VDAT” header, josta voidaan laskea julkiset investoinnit per toimiala ja sektori halutulle vuodelle. PPLUMO-aineistosta saadaan taas investointien jakautuminen maakunnittain. Tällä tavalla koostetut julkiset investoinnit sisältävät tosin sen oletuksen, että investointien rakenne on samanlainen jokaisessa maakunnassa. Myöhemmin kun aluetilinpidon aineisto tulee saataville, voimme päivittää julkisen sektorin investointien jakaumat toimialoihin ja sektoreihin. Väestömäärät saadaan tiedostosta POPUDATA.har. Aineistossa väestö on jaettu kolmeen ikäluokkaan: 14-vuotiaat ja nuoremmat (”ADOL”), 15-64-vuotiaat (”WORK”) ja 65-vuotiaat ja vanhemmat (”AGED”). Aineisto on jaettu maakuntiin ja löytyy kahdesta headerista: RPOP (toteutuneet väestömäärät vuodet 2008-2012) ja RPOF (väestöennuste vuosille 2008-2040). Alueellinen työllisyys on Tilastokeskuksen tietoa ja se löytyy vuosille 1998-2012 tiedostosta REGEMPL.har (rawdata-kansio).

VERM-mallissa olevaa muuttoliikemoduulia varten tarvitaan tietoa julkisten menojen jakautumisesta maakunnittain. Tätä varten on koostettu tiedosto MIGDATA.har (rawdata-kansio), josta tarvittavat tiedot löytyy. [TÄYDENNÄ AINEISTOLÄHTEIDEN OSALTA KUN TYÖ ON VALMIS!!!]

Tuotoksena tästä vaiheesta syntyvät tiedostot REGGOVACC.har, joka sisältää tiedot aluemallin julkista sektoria varten sekä REGSHOCKS.har, joka sisältää alueittaiset julkisen sektorin shokit perusuran muodostamista varten.

1. ***Pääomakannan jako maakunnittain – REGCAPITAL-kansio***

Pääomakanta jaetaan maakunnittain tässä vaiheessa **doitext.bat**-komentojonotiedoston käyttämillä ohjelmilla. CAPITAL-vaiheessa luotua koko maan pääomakanta-aineistoa hyödynnetään mahdollisimman paljon. Työtä varten haetaan seuraavat HAR-tiedostot:

*Tiedosto Kansio Kuvaus*

AGGMOD.har TERM TERM-tietokanta

AGGSUPP.har TERM Aggregointiohjeet

REGGOVACC.har REGGOVERNMENT/savhars Julkinen sektori alueittain

EXTRA1-2008.har GOVERNMENT/savhars Koko maan EXTRA-tiedosto

ITER.har CAPITAL/savhars Iterointiohjeet (eteenpäin katsova dynamiikka)

ROREXT.har CAPITAL/savhars Pääoman tuottoasteen EXTRA-tiedosto

REGEMPL08.har rawdata Alueittainen työllisyys (vuosi 2008)

Koko maan tason EXTRA-tiedosto aggregoidaan ensin sille tasolle, jota aluemallissa käytetään. **preaggext.tab**-ohjelma muodostaa aggregointia varten ohjeet tiedostoon XTRAGGSUPP.har. Kaikki painotusta tarvitsevat parametriheaderit painotetaan pääomakannalla. Varsinainen ohjelma, jolla pääomakannan jako tapahtuu, on **fixregdata.tab**. Ohjelma jakaa pääoman ja kirjoittaa tiedot oikeisiin tiedostoihin siten, että ne ovat VERM-mallin tunnistamassa muodossa. Käytännössä tasomuotoinen aineisto jaetaan alueittaisilla osuuksilla ja muut parametrit oletetaan olevan alueellisella tasolla yhtenevät koko maan mallissa käytettyjen arvojen kanssa. TERM-vaiheen aineiston tuotannontekijätietoja muokataan niin, että se sopii yhteen alueille jaetun pääomakannan kanssa.

1. ***Valtionosuusjärjestelmän aineisto – VOSDATA-kansio***

VERM:ssä on kuvattuna valtionosuus-järjestelmä (vos) maakuntatasolla. Tässä kansiossa vos-aineisto aggregoidaan kuntatasolta maakuntatasolle. Tämä tapahtuu komentojonotiedostolla **dovos.bat**, joka ajaa tablo tiedoston **vosdata.tab**. Alihakemistossa rawdata on vos-raaka-aineisto VOSRAW.har, joka aggregoidaan kuntatasolta maakuntatasolle käyttämällä mappingeja. Kansiossa olevat xls-tiedostot ovat varsinaista raaka-dataa, joka on haettu kunnat.net[[10]](#footnote-10) sivustolta. VOSDATA.xlsx tiedostoon on koottu mallissa käytettävät vos-erät ja AGEDATA.xlsx tiedostoon puolestaan kunnitaisen ikärakenteen muutos. Tässä vaiheessa aineisto on kasattu vuoteen 2013 asti. Vuoden 2014 arvioitujen vos-maksujen päivitys on vielä tekemättä muutaman erän osalta (kts. VOSDATA.xlsx), jotka saadaan mukaan kun VOSTOT\_2013.xls tiedostoa vastaava vuoden 2014 tiedosto ilmestyy kunnat.net sivuille. Lisäksi täytyy huomioida vos-järjestelmän päivitys, joka on summattu tiedostoon Valtionosuuteen tehtävät vähennykset ja lisäykset vuonna 2014 (Kunnan peruspalvelujen valtionosuus 2014 välilehti, kunnat.net). [!![ TARKISTA TÄMÄ PÄIVITYSASIA TAMMIKUU 2014 ]!!]

Päivitettäessä vos-aineistoa on kätevintä päivittää ensin VOSRAW.xlsx, jonka jälkeen aineisto voidaan siirtää manuaalisesti VOSRAW.har-tiedostoon. Prosessissa syntyy alihakemistoon savhars kaksi tulostiedostoa vos-aineiston perusvuodelle (VOSDATA.har) ja aikasarjalle (VOS\_TS.har). Molemmat aggregoidaan seuraavassa vaiheessa käytettävälle aluetasolle ja aikasarjatiedostosta muodostetaan VERMin perusuran vos-shokit (..\REGMASTER\VERMshk.tab).

1. ***Sovelluskohtaisen aggregoinnin luominen aluetasolla – REGMASTER-kansio***

Myös VERM-aluemallin aineisto aggregoidaan sovellustasolle erillisellä ohjelmalla, joka on VATT-mallit-kansiossa oleva komentojonotiedosto **aggregdata.bat.** Sen periaate on samanlainen kuin edellä kuvatun **aggdata.bat**-ohjelman. VERM:n aggregointi tapahtuu REGMASTER-kansiossa ja tuotokset tallennetaan VERMDATA-kansioon. Toimialojen ja hyödykkeiden lisäksi voidaan aggregoida myös alueita. Mallin nykyisessä perusdatassa aggregoidaan Itä-Uudenmaan maakunta osaksi Uusimaata.[[11]](#footnote-11) Toimialojen ja hyödykkeiden aggregointiavaimet haetaan INDMAPS.har- ja COMMAPS.har-tiedostoista, mutta alueaggregointi tehdään tapauskohtaisesti tablo-ohjelmassa **regaggset.tab** (REGMASTER-kansio). VERMin perusuralla käytettävät shokit lasketaan tablo-ohjelmalla **VERMshk.tab**.

**aggregdata.bat** sisältää myös osion, joka tuottaa VERMin perusuraa varten tiedoston, johon on kerätty VATTAGEn perusuralla käytettyjä shokkeja (SHKS2VERM.har). Jotta SHKS2VERM.har sisältäisi VATTAGEn perusuran kanssa yhtenäiset shokit, täytyy ensin VATTAGE-perusura ajaa RunDynam ohjelmalla. CMBHAR ohjelmalla summataan yhteen ..\VATTAGE\_baseline kansiossa olevat vuosittaiset tulostiedostot. Jotta ohjelma toimii oikein, täytyy perusuran kolmikirjaiminen tunniste olla BCS. CMBHAR ohjelma saa ohjeensa tiedostosta **shks2verm.sti** ja se tuottaa nykymuodossaan shokit vuosille 2009-2030. VATTAGEn perusuralta otetaan myös muutama shokki tasomuuttujille, ja ne käsitellään rinnakkaisessa **shks2verm2.sti** prosessissa. Tässä kohtaa joudutaan olettamaan, että viimeisen vuoden tasomuutos shokki on yhtä suuri kuin toiseksi viimeisen vuoden tasoshokki, sillä tasomuuttujia ei voi käsitellä update-muuttujana tablo-koodissa, minkä vuoksi viimeinen vuosi jää niiden osalta päivittämättä. Lopulta tablo-ohjelmalla **shks2verm.tab** yhdistetään luodut HAR-tiedostot ja lasketaan niiden perusteella shokit VERM-perusuraa varten.

Kuvio 2. VATTAGEn aineistoprosessi

Kuvio 3. VERMn aineistoprosessi

PÄIVITYSTARPEET:

* työvoimatiedot (LABSUP.har; CAPITAL\rawdata)
* kauppatase (BOPXLS2HAR; BOP\BALANCEOFPAYMENTS)

1. Yleisen tasapainon mallit kuvaavat talouden toimintaa ns. edustavien toimijoiden käyttäytymisen kautta. Esimerkiksi tuotanto tapahtuu toimialoittain ja kulutus kotitalousryhmittäin. Toimialoja VATT-malleissa on vaihteleva määrä ja kuluttajia yksi edustava kuluttaja. [↑](#footnote-ref-1)
2. 999230 = suomalaisten kotitalouksien ostot ulkomailta, 999240 = ulkomaisten kotitalouksien ostot Suomesta ja 999320 = cif/fob-korjaus, vakuutus- ja rahtitulo. [↑](#footnote-ref-2)
3. Horridge, Parmenter & Pearson: ORANI-G: a general equilibrium model of the Australian economy. <http://monash.edu.au/policy/ftp/oranig/oranidoc.pdf> [↑](#footnote-ref-3)
4. IsLocCom ja IsLocInd ovat yhteneväiset ja jos niitä haluaa muokata, tulee ne muokata molemmat samalla kertaa PARAMETERS.har-tiedostossa. Arvio hyödykkeiden ja toimialojen paikallisuudesta on subjektiivinen ja melko karkea. Sama pätee myös parametriin IsIndivExp. [↑](#footnote-ref-4)
5. Mallissa on kaksi erillistä tuottoastetta, staattisten odotusten tuottoaste (ROR\_SE) ja historiallinen, normaali tuottoaste (RORN). Datan luomisen yhteydessä staattisten odotusten tuottoaste lasketaan aineiston perusteella ja sille annetaan adjustdata-vaiheessa yläraja (esim. 20 %). Normaali, historiallinen tuottoaste sopeutetaan saatuun uuteen staattisten odotusten tuottoasteeseen. [↑](#footnote-ref-5)
6. Tuottoasteen ylärajasta riippuu kuinka paljon toimintaylijäämää tarvitsee sopeuttaa. Mitä suurempi yläraja asetetaan, sen pienempi sopeutustarve ja sitä enemmän voidaan hyödyntää alkuperäistä aineistoa. [↑](#footnote-ref-6)
7. Investointien käyttöön (V2BAS), investoinneista maksettaviin veroihin (V2TAX) sekä investointien marginaaleihin (V2MAR) voidaan myös tehdä muutoksia tässä vaiheessa, jos se jostain syystä nähdään tarpeelliseksi. Vuoden 2004 datassa investointeja muutettiin metsäteollisuuden (I\_21) ja metallien jalostuksen (I\_27) osalta. Uudessa aineistossa investointihyödykkeiden käyttö on kuitenkin käyty tarkemmin läpi ja ylimääräinen sopeutus on tuskin tarpeen. [↑](#footnote-ref-7)
8. Horridge, Mark: Using levels GEMPACK to update or balance a complex CGE database. CoPS-archive (<http://www.monash.edu.au/policy/archivep.htm>): TPMH0058. [↑](#footnote-ref-8)
9. Tämä tiedosto on Jussi Ahokkaan gradu-työn aineistoa. Sen päivittäminen vuoteen 2008 olisi ideaalia, mutta epätodennäköistä. [↑](#footnote-ref-9)
10. http://www.kunnat.net/fi/asiantuntijapalvelut/kuntatalous/valtionosuudet/valtionosuuslaskelmat/Sivut/default.aspx [↑](#footnote-ref-10)
11. Mallin perusvuonna Itä-Uusimaa oli vielä olemassa ja se on myös mukana Tilastokeskukselta saadussa alueaineistossa. Vuonna 2011 se kuitenkin yhdistettiin Uusimaahan. [↑](#footnote-ref-11)