

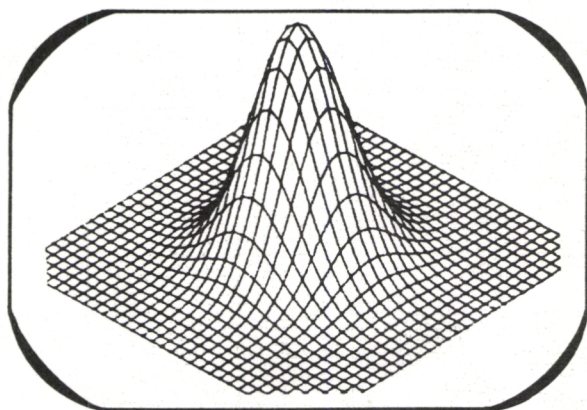
METSÄNTUTKIMUSLAITOKSEN
TIEDONANTOJA 269



Matemaattinen osasto

TILASTOLLINEN TIETOJENKÄSITTELY MIKROTIETOKONEILLA

KIMMO LINNILÄ



Helsinki 1987

METSÄNTUTKIMUSLAITOS
Kirjasto

TILASTOLLINEN TIETOJENKÄSITTELY MIKROTIETOKONEILLA

Kimmo Linnilä

SISÄLLYSLUETTELO

1. Johdanto 5
 2. Katsaus mikro-ohjelmistoihin 8
 3. Testatut ohjelmat 11
 - 3.1. STATGRAPHICS 13
 - 3.2. SAS 17
 - 3.3. SURVO 84C 22
 4. Yhteenvedo 27
- LIITE: Ohjelmistojen hinnat 29

ALKUSANAT

Tämä suppea katsaus tarkastelee tilastollista tietojenkäsittelyä Metsäntutkimuslaitoksessa yleensä, mutta erityisesti nykyään saatavilla olevien mikrotietokoneiden ja niille implementoitujen ohjelmistojen kannalta. Työ on tehty laitoksen matemaattisella osastolla epävirallisen MESTA (METSäntutkimuksen Statistical Analysis) -projektin nimissä. Työn alkuvaiheessa keskityttiin tarjolla olevien ohjelmistojen kartoittamiseen. Lisäksi työn tekijä syvensi tietojaan tilastollisessa tietojenkäsittelyssä sekä ohjelmoi ja testasi keinoitekoisia aineistoja generoivan tietokoneohjelman. Testattavien ohjelmien jyrkkään rajaamiseen on vaikuttanut pääosin työn tekijän käytettävissä olevan mikrotietokoneen (IBM -PC AT) lyhyt vuokra-aika. Mainittu laitteisto tai yhteensopiva tarjoaa tällä hetkellä nimittäin ainoan järkevän työskentelyympäristön tilastolliselle tietojenkäsittelylle mikrotietokoneilla. Huomattava osa laadukkaista ohjelmista (mukaanlukien tilasto-ohjelmistot) on tehty IBM - yhteensopiville koneille. Tutkimuslaitoksen tämänhetkisen mikrokannan ytimen muodostaville PC100 Rainbow- koneille ei kaiken kaikkiaan löydy kovinkaan paljon kohtuuhintaisia hyötyohjelmia. Tilanne laitoksessa korjaantunee sitä mukaa, kun a.o. laitteiden leasing-sopimusten umpeuduttua hankitaan laitoksen tarkoitukseen sopivampaa mikrotietokonekapasiteettia.

Tarkastelussa mukana olleita ohjelmistoja ei ole ollut alunpitäenkään tarkoitus pistää mihinkään ehdottomaan paremmuus-

järjestykseen, vaan pikemminkin esittää erilaisia vaihtoehtoja erilaisiin tarpeisiin erilaisille käyttäjille. Tämän esityksen toivotaan antavan osviittoja lähinnä niille Metlassa työskenteleville tutkijoille ja atk-henkilöille, jotka päättävät yksikkönsä ohjelmistohankinnoista.

1. Johdanto

Metsäntutkimuslaitoksen tilastollinen tietojenkäsittely alkoi 1960-luvun puolivälissä, jolloin yksittäisissä tutkimushankeissa käytettiin tietokonetta tutkimusaineistojen käsitteilyyn. Käyttö oli satunnaista, eikä sitä ollut varsinaisesti organisoitu tutkimuslaitoksen taholta. Tietokonepalveluja saatiin lähinnä Helsingin yliopistolta, opetusministeriön UNIVAC-laitteistola ja jonkin verran Valtion Tietokonekeskukselta.

Laitoksen oma atk-toiminta käynnistyi varsinaisesti vuonna 1977, jolloin keskusyksikköön hankittiin PDP 11/34-laitteisto. Sen kapasiteetti oli nykyisen mittapuun mukaan alhainen niin tiedonhallintamäärän kuin nopeudenkin suhteen. Laitteistoon ei ollut installoitu mitään yleistä tilastollisen tietojenkäsittelyn ohjelmistoa. Laitetta käytettiin lähinnä aineistojen analysointiin etupäässä laitoksen piirissä kehitettyjen ohjelmistojen LM, REKO, KPL-avulla. Tutkimusaineistojen tilastollinen analysointi hoidettiin noihin aikoihin pääosin opetusministeriön U1108-laitteistolla ja erityisesti Helsingin yliopiston laskentakeskuksessa kehitetyllä HYLPS-ohjelmistolla. Ohjelmisto toimi käytännössä eräajotyypisenä ja oli käyttäjän kannalta kankeahko johtuen osin päätetyöskentelyolosuhteista eräajoympäristössä. Kuitenkin HYLPS oli ohjelmistona monipuolinen ja riitti kapasiteettinsa puolesta suurienkin aineistojen käsitteilyyn. Esimerkiksi HYLPS:in varianssianalyysiohjelmisto hakee tänä päivänäkin monipuolisuudessaan vertaistaan. Toisaalta 70-luvun lopulla ei aidosti interaktiivisia tilastollisen tietojenkäsittelyjärjestelmiä

ollut muita kuin WANG 2200-pientietokoneelle implementoitu järjestelmä SURVO 76.

METLAN atk-olot paranivat oleellisesti vuonna 1980, jolloin laitoksen keskusyksikköön saatiin VAX 11/780- laitteisto koh- tuullisessa laite- ja ohjelmistokokoonpanossa. Tilastollista tietojenkäsittelyä edusti BMDP (Biomedical programs), joka on edelleen laitoksen pääasiallinen tilasto-ohjelmisto. Se on kansainvälisesti tunnettu ja usein referoitu ohjelmistoko- konaisuus.

BMDP on eräajotyyppinen, vaikka pienet työt voidaankiin ajaa osituskäyttöistunnon aikana. Sen ohjelmistovalikoima on mo- nipuolinen ja on riittänyt tutkimuslaitoksen useimpiin tar- peisiin. BMDP:n käyttökoulutuksesta on huolehtinut alunpi- täen laitoksen matemaattinen osasto järjestämällä säännöllis-esti kurssitusta. Lisäksi ohjelmiston käytön päiväkohtai- siin ongelmiin on matemaattinen osasto tarjonnut konsulttia- pua, joka on ollut osana yleisempää tilastollisten menetel- mien käytön opastusta ja opetusta. Näin BMDP:n käyttö ja ongelmien ratkaisu on ollut suhteellisen helppoa ja välitöntä niissä laitoksen yksiköissä, joissa BMDP on saatavilla. Lai- toksen keskusyksikön lisäksi BMDP on asennettu Rovaniemen tutkimusaseman VAX 11/750- laitteistolle. Edelleen joiltakin tutkimusasemilta on linjayhteys jompaankumpaan laitteistoon. Lisäksi Joensuun tutkimusasema hoitaa tietojenkäsittelytar- peensa Joensuun yliopiston laskentakeskuksen VAX- laitteiston välityksellä. Muissa kuin edellä mainituissa toimintayksi- köissä tilastollinen tietojen käsittely on ollut kirjavaa.

Metsäntutkimuslaitos hankki ensimmäiset mikrotietokoneensa v. 1983. Laitoksen nykyinen mikrotiekonekanta lienee runsaat 50 kpl, joista n. 40 on Digitalin PC 100- laitteita ja loput lähinnä IBM PC- yhteensopivia.

Siirtymiselle osittain hajautettuun tietojenkäsittelyyn on ollut moniakin syitä:

Keskuskoneen kapasiteettia ei ole voitu lisätä käytön lisäystä vastaavalla määrällä. Tämä on aiheuttanut ajoittain työskentelyn puuroutumista ja turhautumista.

Laitoksen atk- toiminnan kehittämisen tatyöasevoitteisiin kuuluu mien runsas likimallsääminen. Tähän toivotaan päästävän hanka ensisijaisesti mikrotietokoneita.

Käyttäjäkunnassa on esiintynyt kiinnostusta uuteen teknologiaan ja itsenäiseen keskuskooneesta riippumattomaan tietojenkäsittelyyn.

Sopivilla laite- ja ohjelmistoratkaisuilla tutkimuslaitoksen tietojenkäsittelyä voidaan hajauttaa ja poistaa sitä toimintayksikkökohtaista eriarvoisuutta, jota tällä hetkellä esiintyy.

2. Katsaus mikro-ohjelmistoihin

Ensimmäisten mikrotietokoneiden ohjelmistot olivat vaatimattomia konekohtaisia tuotteita, joita ei voinut siirtää laitteelta toiselle. Vasta käyttöjärjestelmien ilmaannuttua markkinoille alkoi ohjelmistotalojen kiinnostus mikrotietokoneita kohtaan lisääntyä. 8-bittisen CP/M:n alaisuuteen valmistettiin jo runsaasti ohjelmia mutta varsinaisen läpimurto kuitenkin tapahtui 16-bittisen MS DOS:n tulemisen myötä. Muodollisesti sama käyttöjärjestelmä ei ole valitettavasti aina taannut ohjelman toimivuutta eri laitteilla. Itse asiassa laitteiston nk. IBM-yhteensopivuus vasta takaa jokseenkin varmasti ohjelman toimivuuden. Tämä on johtanut melko yksipuolisesti IBM-yhteensopivien ohjelmatuotteiden tuottamiseen, mistä muiden laitteiden omistajat ovat joutuneet kärsimään. Esimerkiksi PC 100 Rainbow on mikrotietokone, jonka ohjelmistovalikoima on suppeahko ja kallis.

Mikrotietokoneohjelmistojen kehitys 80-luvulla on ollut huomattava niin määrällisesti kuin laadullisesti. Keskeisiä sovellusalueita ovat:

tekstinkäsittely

tietokannat- ja kortistot

grafiikka

taulukkolaskenta

teknistieteelliset, erityisesti CAD-,CAM- sovellukset

tilastollisen tietojenkäsittelyn ohjelmistot.

Metsäntutkimuksen kannalta kaikki yllä mainitut alueet, kenties CAD-,CAM- sovelluksia lukuunottamatta ovat sovellettavia ja kiintoisia. Tekstinkäsittely laitoksessa hoidetaan voittopuolisesti keskuskoneen LEX- ohjelmistolla tai muilla vanhemmilla järjestelyillä. Mikrotietokoneen käyttäjistä vajaa puolet käyttää jotakin mikro-ohjelmistoa tekstinkäsittelyssään. Tarve tietynlaiseen yleisohjelmiston käyttöön on selvästi olemassa. Olisi luonnollisesti miellyttävämpää ja resurssien kannalta säästeliäämpää, jos useat tietojenkäsittelyn eri alueen toiminnot voitaisiin hoitaa integroidusti yhdestä ohjelmasta käsin. Markkinoilla olevat parhaat taulukkolaskentaohjelmistot ovat pyrkimys tähän tavoitteeseen. Esimerkiksi Lotuksen lippulaiva Symphony tarjoaa viisi eri toimintamoodia: taulukkolaskenta, tekstinkäsittely, grafiikka, tietokannat ja kommunikointi isäntäkoneen kanssa. Vaativa käyttäjä löytää valitettavasti ohjelman rajat melko nopeasti. Grafiikan käyttömahdollisuudet ovat melko rajoitetut, tekstinkäsittely ei tarjoa kaikkia niitä mahdollisuuksia, joita tavanomaisilla tekstinkäsittelyjärjestelmillä on jne. Lotus-tuotteiden kaltaisten, sinänsä erinomaisten ohjelmien perusongelma on vahvan tilastollisen analyysiarsenaalin puuttuminen. Ohjelman tarjoamat tilastolliset perustoiminnot ovat riittämättömät, eivätkä kaupan olevat lisäohjelmat täysin tätä puutetta poista.

Eräänlaisia yritelmiä integroinnin suunnalla edustavat Macintoshin käyttöjärjestelmä ja Microsoftin Windows. Ne toimivat kuitenkin eri periaatteella. Jälkimmäinen sallii tietyin edellytyksin ohjelmien moniajon, mutta ei välttämättä kokoa eri ajojen tuloksia yhteen. Kotimainen SURVO 84C on lienee ainoita ohjelmistoja, jossa yhdelle "työpöydälle" on koottu tiettyjä osia tieteisteknisestä tietojenkäsittelystä. Ohjelmisto onkin mukana edempänä esiteltävässä tarkastelussa.

3. Testatut ohjelmat

Tähän tilastollisten tietojenkäsittelyn järjestelmien "vertailuun" on valittu seuraavat ohjelmatuotteet: STATGRAPHICS, SAS, SURVO 84C. Valikoiman suppeuden syistä mainittiin alku-sanoissa. Kyseisten ohjelmistojen mukaanottoon ovat vaikuttaneet useat syyt. Ohjelmistojen saatavuus Suomessa helpottaa luonnollisesti lyhyellä varoitusaajalla niiden installointia koneelle testaustilanteessa. Maksuttoman kokeilun järjestäminen suoraan ulkomailta tuotteesta, jolla ei ole edustusta Suomessa, on hankalaa, jos edes aina mahdollistakaan. Metsäntutkimuslaitoksella on harkittu SAS- järjestelmän hankkimista keskuskoneelle, joten on luonnollista tutustua mikro SASiin. Ohjelman toimintahan on pääpiirteissään sama molemmilla laitteistotyypeillä. SURVO 84C on professori Seppo Mustosen Helsingin yliopiston tilastotieteen laitoksella suunnittelema ja toteuttama ohjelmisto, joka tutkijan yleistyökaluna lienee ainoa laatuaan. STATGRAPHICS valittiin mukaan tarkasteluun lähinnä melko monipuolisten graafisten ominaisuuksiensa ja erityisesti helppokäyttöisyytensä vuoksi. Kaikki ohjelmat ovat kansainvälisesti tunnettuja, SAS ilmeisesti laajimmin. Huomautettakoon, että valintaan ovat vaikuttaneet tietyt positiiviset ennakkokäsitykset.

Ohjelmiin tutustuttaessa on pyritty kiinnittämään huomiota pääasiassa seuraaviin näkökohtiin:

Laiteympäristö

Järjestelmän käyttöönotto, asennointi, käytön oppiminen ja työskentelyn mukavuus

Tiedon hallinta

Järjestelmään kuuluvat ohjelmistot

Analyysitulosten jatkokäsittely sekä raportointi ja tekstinkäsittely.

Data-aineistoina on käytetty ohjelmistojen mukana tulleita tiedostoja sekä lisäksi itse tietokoneella generoituja multi-normaalisesti jakautuneita keinotekoisia aineistoja. Käytössä ollut mikrotietokone oli IBM PC AT varustettuna 640 Kb:n keskuistilla ja 20 Mb:n kiintolevyllä, Enhanced Color Display (64 Kb:n EGA- kortti)- näytöllä. Kokoonpanoon ei kuulunut 80287- matematiikkaprosessoria. Se ei tutustumisen kannalta ole välttämätön, mutta on varmastikin suositeltava jatkuvassa käytössä, etenkin runsaasti laskentaa vaativissa proseduureissa. Mitään tarkkoja suoritusaikavertailuja eri ohjelmistojen suhteen ei ole tehty.

3.1. STATGRAPHICS

STATGRAPHICS, Statistical Graphics System on yhdysvaltalaisen STSC Inc.- nimisen yhtiön ohjelmatuote. Sen käyttöympäristövaatimukset ovat seuraavat:

- Vähintään 512 Kb keskusmuistia
- grafiikkasovitin ja vastaava värimonitori
- kaksi 2- puoleista levyasemaa tai kiintolevy
- käyttöjärjestelmä DOS, versio 2.0 tai myöhempi
- matematiikkaprosessori ja täysi keskusmuisti 640 Kb suositeltavia
- ohjelmisto tukee lukuisia graafisia kirjoittimia ja piirtureita

Järjestelmän käyttöönotto on helppoa. Ohjelmisto mahtuu 4 levykkeelle, joten asennointi tapahtuu nopeasti. Ohjelmointikielenä on APL, mikä osaltaan lienee vaikuttanut ohjelmis-

ton kokoon. Toisaalta lyhyt koodi lienee osaltaan vaikuttanut ohjelmiston auttamattomaan hitauteen. Statgraphics ei ole kopiosuojattu.

Statgraphics on valikkopohjainen ratkaisu, jossa ohjelman käynnistyksen jälkeen kuvaruudulta on valittavissa 22 osaluetta ryhmiteltynä 6 pääkategoriaan. Kaikkiaan erilaisia toimintoja löytyy 175 kpl. Eri toimintojen käytössä avustavat Helpit. Ohjelmiston mukana seuraa melko runsas ja monipuolinen käsikirja. Käytön oppiminen kokeneelle tutkijalle, erityisesti tilastotieteilijälle ei tuota suurempaa tuskaa: työskentelyn voi aloittaa välittömästi; manuaalia ei tarvitse välttämättä kovinkaan usein. Tähän liittyykin eräs ongelma. Ohjelmat ovat loppuun asti räätälöityjä, eikä vaativalle käyttäjälle jää juuri "muovailuvaraa". Edelleen ohjelmien sirpaloiminen pieniin moduleihin saattaa joistakin tuntua rasitteelta erityisesti raportointia silmällä pitäen.

Tiedonhallinnan osalta Statgraphics noudattelee tilastollisen tietojenkäsittelyn mikroversioiden tyyppillistä linjaa. Ensiksi kannattaa mainita full screen- editori, jonka avulla voi välittömästi päästä kokeilemaan ohjelmiston tilasto-osia tai graafisia tulostusmahdollisuuksia. Editorilla voidaan luoda, korjata ja päivittää tiedostoja. Systemin ulkopuolisten tiedostojen käyttöönnotossa ei ole vaikeuksia tiettyä standardia noudattavien tiedostojen osalta. Merkkimuotoisen tiedoston siirtäminen systeemin datatiedostoksi oli ongelmatonta. Statgraphics tunnistaa esim. Lotus-, dBase III- ja Symphony

ohjelmilla tuotetut tiedostot. Samoin DIF (Data Interchange Format)- tiedostojen tuonti ja vienti sujuu vaivattomasti. Statgraphicsin havaintotiedostojen maksimikoko on 64 Kb, mikä saattaa aiheuttaa ongelmia. Toisaalta operointi isoilla havaintotiedostoilla saattaa ohjelmiston ja mikrotietokoneiden hitauden vuoksi olla ongelmallista sinänsä.

Systemin toiminnot jakautuvat 6 osa-alueeseen: datan käsittely ja systeemiohjelmistot, grafiikka ja kuvailevat tunnusluvut, varianssi- ja regressioanalyysi, aikasarjamenetelmät, "kehittyneet" menetelmät sekä matemaattiset proseduurit. Va liokoimaa voi pitää monipuolisena ja useimpiin jokapäiväisiin ongelmatilanteisiin riittävänä. Ohjelmisto on jossakin määrin ekonometrian suuntaan orientoitunut, mutta esim. monimuuttujamenetelmienkin puolelta löytyvät pääkomponentti- ja faktorianalyysit ja kanoniset korrelaatio etc. Ohjelmien ajoa luonnehtii huomattava hitaus. Esim. korrelaatiomatriisin tuottaminen aineistosta, jossa 15 muuttujaa ja 100 havaintoa, kestää tuskastuttavan kauan. Hitaus ei rajoitu pelkästään laskentaan ; sen huomaa myös uusien ohjelmamodulien lataamisessa ja graafiikassa, joka on koko ohjelmiston mainosvaltti. Statgraphics tunnetaan ohjelmistona, joka visualisoi tilastoaineistot erilaisina kuvina. Tässä on sinänsä onnistuttukin. Varsinkin aloittelijaa hurmaavat kolmiulotteiset frekvenssikuviot, jotka syntyvät melko vaivattomalla tehtävän määrittelyllä. Erilaisten funktioiden kuvaajien piirtäminen onnistuu helposti. Grafiikalle on ominaista myös valmiiden piirrosten tekstien editointi ja kuvien rotatointi on line-tilassa.

Ohjelmiston matemaattiselta puolelta kannattaa mainita lineaarisen ohjelmoinnin mahdollisuus , ominaisarvojen ja yhtälön juurien ratkaiseminen sekä numeerinen derivoiminen. Tällaiset piirteet tekevät ohjelmistosta luonnollisesti monikäyttöisemmän. Systeemin päävalikosta käynnistyy raportointi ja piirrosten uudelleen esitys (Report Writer and Graphics Replay), joka mahdollistaa tutkimustekstin tuottamisen ja esim. erilaisten taulukoiden sijoittelun tekstien lomaan. Valitettavasti järjestelmä ei tue skandinaavisia merkkejä, joten suomenkielisen tekstin tuottamisen kanssa tulee vaikeuksia.

Yleisvaikutelmana Statgraphicsista voi sanoa, että se helppokäyttöinen, kohtalaisen hyvin dokumentoitu tilastollisen tietojenkäsittelyn ohjelmisto, joka on tarkoitettu erityisesti tilastoaineistojen ja tilastollisten analyysitulosten graafiseen havainnollistamiseen. Analyysivalikoima riittänee useimpiin sovelluksiin. Kuten edellä mainittiin, ohjelmat ovat pitkälle räätälöityjä. Maahantuojan ilmoituksen mukaan systeemiin on mahdollista lisätä omia APL- kielisiä ohjelmia, jotka toimivat samaan hakemistoon installoidun APL- tulkin alaisuudessa. Tulkki merkinnee entistä huomattavampaa hitautta ohjelmien ajossa. Statgraphics maksaa Suomessa manuaaleineen 5800,-. Se on kertamaksu tuotteesta, josta maahantuoja ei usko tulevan päivityksiä lähiaikoina.

3.2 SAS

SAS, Statistical Analysis System, on yhdysvaltalaisen SAS Institute Inc.:n valmistama. Samanniminen ohjelmatuote on laajasti käytetty isommilla tietokoneilla. valmistaja suosittelee seuraavaa käyttöympäristöä:

- Vähintään 512 Kb keskusmuistia, josta 450Kb vapaana

- kiintolevy, josta mielellään 7Mb vapaana

- Käyttöjärjestelmäversio PC DOS 2.0 tai myöhempi varustettuna RESTORE- komennolla. MS DOS- version kanssa voi tulla installoidessa hankaluuksia

- käytettäessä SAS: in lisämoduleja (esim. SAS/STAT), täytyy perussAS in ja lisämodulin version olla sama

Järjestelmän mikroversiosta on tällä hetkellä saatavilla Base SAS, SAS/STAT sekä SAS/RTERM, joista viimeksi mainittu on lähinnä eräitä oheislaitteita emuloiva ohjelmisto. Tässä tarkastelussa keskitytään kahteen ensin mainittuun ohjelmis-

toon. Järjestelmän maahantuoja toimittaa asianomaisten levykkeiden ja käsikirjojen ohella asennointiooppaan, jonka avulla ohjelmien lataaminen kiintolevylle käy vaivattomasti. Ohjelmisto ei ole kopiosuojattu. Manuaaleja on kaikkiaan 4: johdanto-opas, SAS- kielen opas, SAS- proseduurien opas sekä SAS/STAT- käsikirja. Manuaalit ovat selkeitä ja tarjoavat vähemminkin tilastotiedettä hallitsevalle riittävästi tietoa itse ohjelmien toiminnasta ja käytettävästä menetelmästä. Erityisesti johdanto-oppaan värilliset kuvaruutunäytöt helpottavat vasta-alkajan tutustumista järjestelmään. SASilla on kaksi käyttötapaa: valikkopohjainen ja editoriaalinen. Edellinen on jonkin verran helpompi käyttää, mutta editoriaalinen työskentelytapa on selkeämpi ja tarjoaa mahdollisuuden uusien proseduurien tallentamiseen. Tämä käy päinsä tallentamalla editointialue tiedostoksi.

Työskentely SASilla tapahtuu ikkunoissa. Perusasetuksena on tulos-, LOG- ja editori-ikkunat. Liikkuminen ikkunasta toiseen tapahtuu kirjoittamalla asianomainen komento tai ohjelmalla tietyt näppäimet tähän tarkoitukseen. Erilaisia komentoja varten voidaan ohjelmoida kaikkiaan 92 erilaista näppäinsekvenssiä. Tulossikkunaan kirjoittuu ajon lopputuloksia, kuten esim. regressiokertoimet, korrelaatiomatriisi jne. LOG- ikkuna sisältää tietoja ajon kulusta ja mahdollisista virheistä, jotka SAS- ajon määrittelyssä on tehty. Varsinainen ohjelma kirjoitetaan editori-ikkunassa. SASohjelma koostuu DATA- askeleesta ja yhdestä tai useammasta proseduurista.

DATA- askel on yleensä jokaisen SAS- määrittelyn ensimmäinen vaihe. Siinä joko määritellään olemassaoleva SAS- tiedosto (SAS data set) tai kuvaillaan ulkopuolinen tiedosto, joka halutaan systeemiin mukaan. Havaintomatriisin voi antaa luonnollisesti manuaalisesti CARDS- optiota käyttäen. Myös full screen - editointi on mahdollinen. DATAaskel yhdessä lukuisien proseduurien kanssa muodostaa tehokkaan tiedonhallinnollisen työkalun, jolla havaintoaineisto voidaan vähällä vaivalla muokata haluttuun asuun. Lotus 1-2-3 ja Symphonytiedostot voidaan ottaa käyttöön DIF- tiedostoina. Niitä ja dBASE II ja III- tiedostoja varten on omat proseduurinsa. Havaintojen valintaan ja puuttuvan tiedon käsittelyyn on monipuoliset mahdollisuudet. Myös erityyppisen tiedon, esim. päivämäärätiedon, käsittelymahdollisuudet ovat omaa luokkaa.

SAS koostuu laajoista ohjelmamoduleista, joista Base SAS käsittelee tiedon hallinnan sekä joukon yleisiä ja tilastollisia prosedureja. Edelleen eräät proseduurit palvelevat raportointia. Base SAS sellaisenaan tuskin riittää tilastolliseen tietojenkäsittelyyn, vaan se kaipaa lisämoduliksi SAS/STATin. Sen tarjoaa jo huomattavasti laajemman analyysivalikoiman. Tosin esimerkiksi epälineaarinen regressioanalyysi puuttuu. SASin perusohjelmiston graafiset tulostusmahdollisuudet ovat jokseenkin vaatimattomat, suurin piirtein samaa luokkaa kuin isolle koneelle installoidun BMDPn. Valitettavasti SAS/GRAPH- moduli ei ole vielä valmis mikroversiona. Minikoneiden

SASKäyttäjien kokemukset nimenomaan SASin grafiikasta ovat olleet positiiviset. Mikromoduli tullee Suomen markkinoille vielä vuoden 1987 aikana.

SASia vaivaa huomattava hitaus. Se tulee ilmi jo ohjelmistoa käynnistettäessä, mutta jatkuu eri proseduurien alkuvaiheessa. Itse laskennassa se todennäköisesti on nopeampi kuin STATGRAPH. SAS käyttää myös huomattavasti keskusmuistia. Valmistajan ilmoituksen mukaan vähimmäisvaatimus on vähintään 450 Kb vapaata keskusmuistia. Tämän seurauksena esim. Side Kick- tyyppisen kalenterilaskuriohjelman samanaikainen mukanaolo ei ole mahdollista käynnistettäessä useimpia SASin proseduureja. Onko tämä suuri puute, on toinen kysymys. SASin käyttökynnys on selvästi korkeampi kuin STATGRAPHin jopa valikkopohjaisessa työskentelyssä. Kynnyksen yli päästään toki kohtuullisessa ajassa hyvien käsikirjojen avulla. Editoriaalisessa työskentelyssä, joka ehdottomasti on suositeltavampi, käyttäjä voi itse määrittellä työskentely-ympäristönsä mieleisekseen ikkunoiden sijoittelulla, tausta- ja tekstivärien määrittelyillä jne.

SASin hinnoittelu on lisenssipohjainen ja se suosii luonnollisesti suuria työyhteisöjä. Hinnat ovat vuosimaksuja. Esimerkkinä maksujen suuruudesta mainittakoon, että Base SASin 3 lisenssin maksu on ensimmäisenä vuotena 7000 mk ja seuraavina 5000 mk. Vastaavasti 50 lisenssin maksut ovat 12000 mk ja 10000 mk. SAS/STATin maksut ovat suurin piirtein samaa luok-

kaa. Joka tapauksessa nähdään, että esim. Metlan kokoiselle työyhteisölle SAS saattaisi olla suhteellisen edullinen ratkaisu.

3.3. SURVO 84C

SURVO 84C on professori Seppo Mustosen Helsingin yliopiston tilastotieteen laitoksella suunnittelema, kehittänyt ja toteuttama yleisohjelmisto. Järjestelmän laiteympäristövaatimukset ovat seuraavat:

- Vähintään 640 Kb keskusmuistia, 20 megatavun kiintolevy ja Enhanced Color Display- (EGA) näyttö. Mm. kuvaputkigrafiikkaan värivalikoiman kannalta on eduksi, jos EGA - kortti on varustettu 256 kilotavun muistilla. Vapaata levytilaa tulee ennen installointia olla ainakin 5 megatavua.
- ainoa tulostuslaite, jota SURVO 84C virallisesti tukee, on Canonin LBP-8 A2-laserkirjoitin (Centronics- liitännällä).
- Käyttöjärjestelmäversio PC DOS 3.2 tai sitä uudempi.

SURVO 84C:n installointi kovalevylevylle käy vaivattomasti esitteen "SURVO 84 tutkijan työvälineenä " ohjeita noudattaen. Asennettavia diskettejä on 20. SURVO 84C on suojattu

modulilla (software key), joka kytketään jatkokaapelin avulla RS-232porttiin. Ilman modulia ohjelma ei toimi. Toisaalta ohjelman moduleineen voi siirtää tarvittaessa toiselle laitteelle.

SURVO 84C:n käyttöoppaana e.m. esitteen lisäksi on SURVO 84 Käyttäjän opas, joka on laadittu SURVO:n aikaisemman, Basic-kieleisen version aikoihin, eikä ole kaikilta osin täysin yhteensopiva ohjelmiston nykyisen version kanssa. Tärkein tietolähde järjestelmässä onkin oma neuvontajärjestelmä, joka voidaan aina käynnistää HELP-näppäimen painalluksella. Toinen mahdollisuus on aktivoida toimituskentässä avainsana varustettuna kysymysmerkillä. Neuvontajärjestelmän antamat opastukset riittävät usein ongelman ratkaisemiseen. Erityisen tärkeitä SURVO:n käyttöä opiskeltaessa ovat lukuisat käyttöesimerkit, joita on Käytön oppaassa, neuvontajärjestelmässä sekä erityisissä opetusohjelmissä. Opetusohjelmia käyttäjä voi itsekin laatia n.k. opetusmoodissa, jolloin yhden istunnon toimenpiteet tallentuvat myöhempiä istuntoja varten. Tämä mahdollistaa opetusta, joka ei koske pelkästään SURVO:a, vaan saattaa liittyä esim. tilastotieteen opetukseen. Kaiken kaikkiaan voi todeta, että SURVO 84 C:n oppiminen ja oikeiden käyttötottumuksien omaksuminen vie huomattavasti enemmän aikaa kuin esim. muilla tässä tarkastelussa mukana olleilla ohjelmistoilla. SURVO 84C onkin ammattilaisen työkalu, jonka tarjoamat mahdollisuudet painavat vaa'assa enemmän kuin opiskeluvaiheen tuskat.

SURVO 84C on yleisohjelmisto, joka palvelee ensisijaisesti tilastollista laskentaa, grafiikkaa ja tutkimusraporttien tuottamista. Historiallisen taustansa puolesta se rakentuu tilastollisen tietojenkäsittelyn ympärille. Järjestelmä ei ole valikkopohjainen, vaan toiminta tapahtuu editoriaalisesti toimituskentässä, joka oletusarvoisesti on 100x100 merkin kokoinen. Toimituskenttiä voi olla rajaton määrä. Niistä yksi on kerrallaan aktiivi muiden ollessa talletettuna levyllä. Toimituskentän sisältö voi olla mitä tahansa merkkitietoa: tekstiä, lukuja, kaavoja jne. Toimituskentän osia voidaan aktivoita painamalla ESC- näppäintä. Aktivointi voi kohdistua matemaattisen lausekkeen arvon ratkaisemiseen, jonkin tilastollisen operaation suorittamiseen jne. Siirtymisen toimituskentästä toiseen voi organosoida suhteellisen näppäräksi.

Tiedonhallinnan yleinen periaate on seuraava: mitä tahansa ASCII- tyyppistä tekstitiedostoa voidaan selata toimituskentässä ja valita siitä sopivia osia toimituskentän osaksi esim. tilastollisia analyysejä tai tekstinkäsittelyä varten. Vastaavasti SURVO 84C:n tuottamia tulostaulukoita yms. voidaan ottaa toimituskentän osaksi ja siirtää tekstitiedostoon. valmiita rutiineja esim. Lotus 1-2-3 - ja Symphony- tiedostojen käsittelyyn ei ole, vaan tiedostot siirretään tekstitiedostoina käyttäen a.o. ohjelmistojen PRINTFILE- optiota. Varsinaisia tilastollisia havaintoaineistoja voidaan käsitellä kahdella tavalla. Pienet aineistot voidaan taltentaa suoraan

toimituskenttään. Suuremmat sen sijaan tallennetaan SURVO 84C:n havaintotiedostoiksi, jolloin haut, päivitykset, muuttujatransformaatiot ovat välittömästi toteutettavissa. Havaintojen suunnassa tiedoston kokoa rajoittaa vapaa levytila. Muuttujien lukumäärä voi olla tapauksesta riippuen huomattavan suuri.

Yleisohjelmistona SURVO 84C sisältää oleellisia piirteitä tekstinkäsittely-, taulukkolaskenta-, tilastollisen tietojenkäsittely-, grafiikka- ja tiedonhallintaohjelmistoista. Tekstinkäsittelyn tavanomaiset toiminnot kuten siirrot, kopiointit, tasaukset, tavutukset yms. löytyvä vakiovarusteenä. Grafiikkaa voidaan tuottaa PLOT- ja GPLOT-komennoilla joko suoraan laserkirjoittimelle tai tallentaa nimellä kuvatiedostoiksi. Aikaisemmin tuotetut kuvat voidaan liittää osaksi raporttia tulostusvaiheessa. Kuvan yksityiskohtien määrittely on osittain taitoa vaativaa. Lopputuloksena saatava kuitenkin olla korkeatasoisinta grafiikkaa, mitä tämänkaltaisilla ohjelmilla on nähnyt tuotettavan. Kolmiulotteisten kuvien tekeminen onnistuu luonnollisesti.

Tilastollisten analyysimenetelmien joukko on jo tällä hetkellä mittava ja valikoimaa täydennetään jatkuvasti. Tavanomaisten menetelmien lisäksi ansaitsee mainita erityisesti todennäköisyyspaperit, matriisitulkki ja monimuuttujamenetelmät matriisitulkin välityksellä. Oman erikoisuutensa muodostavat aritmeettiset laskutoimitukset editoriaalisen tai kosketuslaskennan avulla. Edellisen avulla voidaan ratkaista

matemaattisen lausekkeen arvo keräämällä automaattisesti tarvittavat alkuarvot toimituskentästä. Jälkimmäisessä "kosketaan" toimituskentässä olevia lukuja kohdistamalla niihin jokin aritmeettinen laskutoimitus. Painamalla lopuksi näppäintä = , saadaan lausekkeen lopputulos selville.

SURVO 84C jättää itsestään monipuolisen ja poikkeuksellisen vaikutelman. Ohjelma levittää tutkijan työpöydälle työkalupakin, jonka sisältö riittää hyvin moniin ongelmatilanteisiin. Työkalut ovat välittömästi käytettävissä, sillä SURVO 84C:n nopeus verrattuna muihin tarkastelun kohteisiin on omaa luokkaansa. Erilliset ohjelmat, joita on yhteensä n. sata, on ohjelmoitu Microsoftin C-kielillä. C-kielihän tuottaa tunnetusti nopeaa koodia. Ohjelmointitaitoiselle käyttäjälle tarjoutuu mahdollisuus lisätä tiettyjen standardien puitteissa ohjelmistovalikoimaa ja räätälöidä ohjelmistoa osin omaan työympäristöönsä. SURVO 84C vaatii käyttäjältään paljon, mutta palkitsee vaivannäön laadukkaalla lopputuloksella.

4. Yhteenveto

Tarkastelun kohteena olleet järjestelmät olivat kaikki vähintään tyydyttäviä. Jokaiselle ohjelmistolle löytyisi Metsäntutkimuslaitoksen piiristä oma käyttäjäkuntansa. Esimerkiksi STATGRAPHICS helppoutensa puolesta sopisi laitoksen koko tutkijakunnalle koulutustaustasta riippumatta. Kuitenkaan ohjelmiston käytön helppous ei ole aina pelkästään myönteinen asia. Helpot ohjelmaratkaisut saattavat johtaa köykäiseen tutkimusongelman asetteluun, koska ohjelmisto ei tarjoa mahdollisuuksia monimutkaisempaan pohdiskeluun. Tuloksena voi olla mielikuvituksettomia stereotyyppioita, jotka eivät edistä tiedettä. Nimenomaan ohjelmien räätälöimismahdollisuus tarjoaa mahdollisuuden suorittaa tilastollista tietojenkäsittelyä tutkimuksen eikä ohjelmiston ehdoilla.

Tutkimuslaitoksen tulevia VAX 11:n ohjelmistoratkaisuja ajatellen SAS saattaisi tulla mikroilla kyseeseen. SASia on nimittäin harkittu BMDPn ohelle isojen aineistojen käsittelyyn. Ratkaisun tekemistä on osaltaan viivyttänyt VAXin ohjelaitteiden soveltumattomuus graafiseen tietojenkäsittelyyn. Tilanne korjautunee lähi vuosina. SASkoulutuksella hoidettaisiin samalla sekä ison koneen että mikrojen käyttäjien tarpeet. Suurella käyttäjäkunnalla (50 lisenssiä) SAS olisi taloudellinen ratkaisu, vaikka kukin vuosi maksaakin erikseen. Täydellä käyttäjämäärällä Base SAS täydennettynä 3 lisämodulilla maksaisi 1. vuotena 840 mk ja seuraavina vuo-

sina 660 mk yhtä mikroa kohden. Vie tosin aikansa, ennekuin laitoksella on 50 kpl vähintään IBM PC XT-tasoisia mikrotietokoneita.

SURVO 84C tarjoaa vaativalle ja osaavalle käyttäjälle ehdottomasti eniten. Hinta on kohtalaisen korkea - 12000 mk - mutta muita ohjelmia ei juuri tarvitakaan. Ohjelma on kotimainen ja on todennäköistä, että sen kehittelyyn voi käyttäjäkunta vaikuttaa. Samoin opastus omien C-kielisten ohjelmodulien teossa on taattu. Tilastotieteen ja tutkimuksen opettamisen puolella SURVO 84C:llä olisi varmaan monipuoliset mahdollisuudet. Toistaiseksi valmiita opetusmoduleja ei tosin vielä ole runsaasti. Edelleen puhtaasti metsällisten menetelmien, puun ja metsikön tunnusten laskemisen integroiminen SURVOon on harkinnan arvoista. Lisäkokemuksien hankkiminen SURVO 84 C:stä Metsäntutkimuslaitoksen työympäristössä on joka tapauksessa suositeltavaa.

LIITE

Ohjelmistojen hinnat tammikuussa 1987

STATGRAPH	5800,-	
SAS	1. vuosi	seuraavat vuodet
Base		
3 lisenssiä	7000,-	5000,-
50 lisenssiä	12000,-	10000,-
SAS/STAT		
3 lisenssiä	5000,-	3500,-
50 lisenssiä	12000,-	9000,-
SAS/RTERM		
3 lisenssiä	3500,-	2600,-
50 lisenssiä	9000,-	7000,-
SAS/IML		
3 lisenssiä	3500,-	2600,-
50 lisenssiä	9000,-	7000,-
SURVO 84C	16000,-	
Valtion tutkimus- laitokset -25%	12000,-	

METSÄNTUTKIMUSLAITOS

Matemaattinen osasto

Osoite: PL 37, 00381 HELSINKI (Kornetintie 8) ja
Unioninkatu 40 A, 00170 HELSINKI

Puhelin: (90) 556 276 ja
(90) 661 401

Pekkonen, Timo, vs. professori

Klippi, Lea, tutkimussihteeri

Menetelmät

Häkkinen, Risto, matemaatikko

Heinonen, Jaakko, tutkija (Joensuun tutkimusasema)

Linnilä, Kimmo, tutkija

Sievänen, Risto, tutkija

Atk

Pöntinen, Jukka, atk-päällikkö

Herrala-Ylinen, Helena, tutkija

Kaila, Erkki, tutkija (Rovaniemen tutkimusasema)

Kinnunen, Hilikka, tutkija (Rovaniemen tutkimusasema)

Mäkinen, Markku, tutkija

Salmi, Veli-Pekka, atk-suunnittelija

Snellman, Carl-Gustaf, tutkija

Granlund, Hilikka, pääoperaattori

Palviainen, Pertti, tutkimusapulainen (Rovaniemen tutkimusasema)

Soimula, Maire, operaattori

Metsätilasto

Uusitalo, Matti, tutkija

Aarne, Martti, tutkija

Lehto, Kari, tutkija

Valli, Pasi, tutkija

Leppäkumpu, Tuula, toimistosihteeri

Kämäräinen, Paula, toimistosihteeri

Metsäverotus

Rauskala, Raimo, vanhempi tutkija

Kakkuri, Eero, tutkija

Kulju, Irma, toimistosihteeri

Mäkinen, Kaija, ohjelmoija

Sivulliset tutkijat

Hari, Pertti, dosentti

Heikinheimo, Lauri, emer. professori

Kallio, Markku, professori

Rytkönen, Antti, metsänhoitaja

Matemaattisella osastolla ilmestyneet Metsäntutkimuslaitoksen tiedonantoja -sarjan viimeisimmät julkaisut:

- nro 93 Eero Kakkuri. Ilomantsin luonnonsuojelualueiden taloudellinen merkitys puuntuotannolle ja matkailulle. 23 s. 1983.
- nro 128 Raimo Rauskala. Kunnittaiset kantohinnat ja puukuutiometrin bruttoarvot hakkuuvuonna 1982/83. 38 s. 1984.
- nro 149 Pertti Hari, Kullervo Kuusela, Pentti K. Räsänen, Risto Seppälä. Metsäntutkimukseen liittyvistä kehityssuunnista. 38 s. 1984.
- nro 152 Eero Kakkuri. Yksityismetsänomistajien puun kasvatuksen kulut vuosina 1981 ja 1982. 17 s. 1984.
- nro 157 Erkki Kaila ja Markku Taipale. Tutka-tiedonhallintaohjelmisto Tietokannan muodostus ja käyttö. 113 s. 1984.
- nro 176 Raimo Rauskala. Forest taxation and roundwood supply in Finland. 12 s. 1985.
- nro 183 Staffan Ringbom. Virkesproduktionens totala lönsamhet och dess mätning. 32 s. 1985.
- nro 191 Raimo Rauskala. Kunnittaiset kantohinnat ja puukuutiometrin bruttoarvot hakkuuvuonna 1983/84. 44 s. 1985.
- nro 194 Heinonen, J., Penttinen, A., Salminen, S., Tomppo, E. Spatiaalisen tilastotieteen soveltaminen metsäntutkimukseen. 129 s. 1985.
- nro 223 Raimo Rauskala. Kunnittaiset kantohinnat ja puukuutiometrin bruttoarvot hakkuuvuonna 1984/85. 55 s. 1986.
- nro 224 Eero Kakkuri. Puun hintojen vaihtelu kuntien sisällä hakkuuvuonna 1980/81. 22 s. 1986.
- nro 240 Eero Kakkuri. Yksityismetsänomistajien puun kasvatuksen kulut vuosina 1983 ja 1984. 22 s. 1986.
- nro 251 Eija Virtanen (toim.). BIB-viitetietokantaohjelmisto. Version 2.1 käyttöohje. 62 s. 1987.
- nro 252 Hilikka Kinnunen. Metlan sarjat viitetietokantana. 24 s. 1987.
- nro 254 Kari Lehto. Turvallisuuskohdista ja suojauksista Metsäntutkimuslaitoksen atk-järjestelmissä. 55 s. 1987.

ISBN 951-40-0841-3
ISSN 0358-4283

Helsinki 1987, VAPK Kampin VALTIMO