## 1. Johdanto tuotteeseen

Javist – jakaumien visualisointiyökalu on tarkoittettu Bassist–ohjelman tuottamien datamassojen reaaliaikaiseen visualisointiin. Lisäksi ohjelma tarjoaa työkalut datan tallentamiseen sekä kaavioiden tulostamiseen.

### *1.1 Oletettu käyttäjäprofiili ja käyttötapa*

Javistin kohdekäyttäjäryhmä muodostuu tietojenkäsittelyn ja tilastotieteen asiantuntijoista, joille tietokoneiden, tilastollisten mallien ja MCMC–simulaatiotulosten käyttö tutkimuksessa on arkipäivää.

Javist on vuorovaikutteinen työkalu Bassistin tuottamien pitkien lukujonojen visualisointiin ja yksinkertaiseen analyysiin. Javist visualisoi ja analysoi dataa monin keinoin: keskeisinä ovat mm. yhden parametrin tai kahden parametrin yhteisjakauman laskenta ja esitys eri muodoissa, paloittain vakioiden funktioiden käsittely, sekä tärkeimpien tilastollisten tunnuslukujen laskenta.

### *1.2 Käyttöympäristö*

Javist–ohjelma toimii pääasiassa nykytason mukaisessa LINUX–käyttöjärjestelmällä varustetussa PC–laiteympäristössä. Ohjelmaa oletetaan käytettävän 32 megatavun keskusmuistilla varustetuissa pentium–tason koneissa. Laite– ja ohjelmistoympäristö ei kuitenkaan ole rajattu pelkästään näihin, vaan kaikki riittävän tehokkaat java–ohjelmia tukevat ympäristöt tulevat kysymykseen.

### *1.3 Ohjeessa käytetyt lyhenteet*

*kvantiili* Kvantiilit ovat "luottamusvälejä" (esim. 2.5 – 97.5 prosenttia) kullakin ajan hetkellä. Esim. 2.5 prosentin kvantiilille saadaan arvo jättämällä simulaation tuloksesta pois 2.5/100 pienintä arvoa. Operaation jälkeen jääneistä arvoista ensimmäinen on kyseisen kvantiilin arvo.

*MCMC* Markov chain Monte Carlo –menetelmät ovat satunnaisuuteen perustuvia integrointialgoritmeja. Bassist –hankkeessa MCMC–menetelmiä sovelletaan otosten tuottamiseen tilastollisten mallien ns. posteriorijakaumista, jotka ovat usein hyvin monimutkaisia jakaumia. MCMC–menetelmä mahdollistaa laskennallisesti monien sellaisten mallien käytön, joiden posteriorijakaumia ei voida approksimoida millään toisella tunnetulla menetelmällä.

## 2. Yleiset rajoitteet

Javist on tarkoitettu Bassist –projektissa työskentelevien ohjelman tilanneiden asiakkaiden ja näiden parhaaksi katsomien henkilöiden käyttöön.

Tekninen rajoitus ohjelmalle on vähintään 32 megatavun keskusmuisti. Varsinaista syötteen maksimikokoa ei ole määrätty, mutta sen oletetaan mahtuvan keskusmuistiin.

## 3. Tuotteen käyttö

### *3.1 Ohjelman käytössä tarvittava apumateriaali*

Javist tarvitsee apumateriaalinaan Bassist –ohjelman tuottaman

syötteen. Syöte on mahdollista ohjata Javistille putken kautta tai vaihtoehtoisesti ladata tiedostosta.

### *3.2 Alustavat toimenpiteet*

Toimenpiteet ennen Javistin käynnistystä: java–ohjelmia tukevan ympäristön asennus ja Javist–ohjelman asennus (kts. luku 6). Lisäksi JAVIST ja CLASSPATH –muuttujat on LINUX–ympäristössä asetettava osoittamaan oikeisiin hakemistoihin komennoilla ˆsetenv JAVIST /home/javist/classes/kontrolli˜ ja ˆsetenv CLASSPATH $CLASSPATH:/home/javist/classes:$JAVIST˜ tässä järjestyksessä. Hakemistopolku ˆ/home/javist/˜ tulee korvata hakemistopolulla, jonne Javist on asennettu. Em. komennot on syytä sijoittaa johonkin automaattisesti suoritettavaan (esim. loginin yhteydessä) komentotiedostoon. Lisäksi käyttäjän bin–hakemistoon luodaan linkki komennolla ˆln –s $JAVIST/javist javist˜. Tällöin Javist voidaan käynnistää missä päin hakemistopuuta tahansa. Myös muissa käyttöympäristöissä käyttäjän on huolehdittava oikeiden hakupolkujen asettamisesta kunkin ympäristön käytännön mukaisesti.

Javist käynnistetään joko ohjaamalla Bassist –ohjelman tuottama syöte putken kautta Javistille tai vaihtoehtoisesti komentoriviltä itsenäisesti komennolla ˆjavist˜. Valmiina olevan syötetiedoston voi valita syötteeksi käynnistyksen yhteydessä komennolla ˆjavist < TIEDOSTO˜, jossa TIEDOSTO on kyseisen syötetiedoston nimi

### *3.3 Pikaopas*

– Parametrien nimet näytetään käyttäjälle ja käyttäjä voi valita

tarkasteltavat parametrit

– Käyttäjä voi pyytää poistamaan syötteen alusta jonkin määrän rivejä

– Käyttäjä voi pyytää ottamaan syötteestä mukaan vain joka n:nnen rivin.

– Eri kaaviotyyppien (kts. alla) tuottaminen, tulostaminen ja tallentaminen.

– Tilastollisten tunnuslukujen laskeminen: yhden muuttujan keskiarvo ja hajonta, kahden muuttujan välinen korrelaatio, yhden muuttujan arvojonon autokorrelaatio.

– Käyttäjän valikoimien arvojen ja tunnuslukujen tallennus.

Kaaviotyypit:

*Autokorrelaation seuranta:* Yhden muuttujan arvojonon autokorrelaatiota voidaan seurata omassa kaaviossaan. Käyttäjä voi itse määritellä autokorrelaation laskentasyvyyden.

Simulaation jäljittäminen: Simulaation kulkua voidaan seurata yhden parametrin osalta, jolloin vaaka–akselilla on simulaatiokierroksen numero ja pystyakselilla tarkasteltavan parametrin arvo kullakin kierroksella.

*Parametrin jakauma:* Parametrin saamien arvojen jakaumaa voidaan kuvata histogrammalla. Vaaka–akselilla ovat parametrin saamat arvot, pystyakselilla esiintymien lukumäärä (tai suhteellinen osuus) simulaation aikana. Käyttäjä pystyy vaikuttamaan pylväiden kokoon ja lukumäärään sekä antamaan haluamansa arvovälin, jolta histogramma piirretään koko arvovälin sijasta. Histogrammasta ilmenee myös jakauman approksimoitu odotusarvo. Myös halutun kvantiilin indikoiminen on mahdollista.

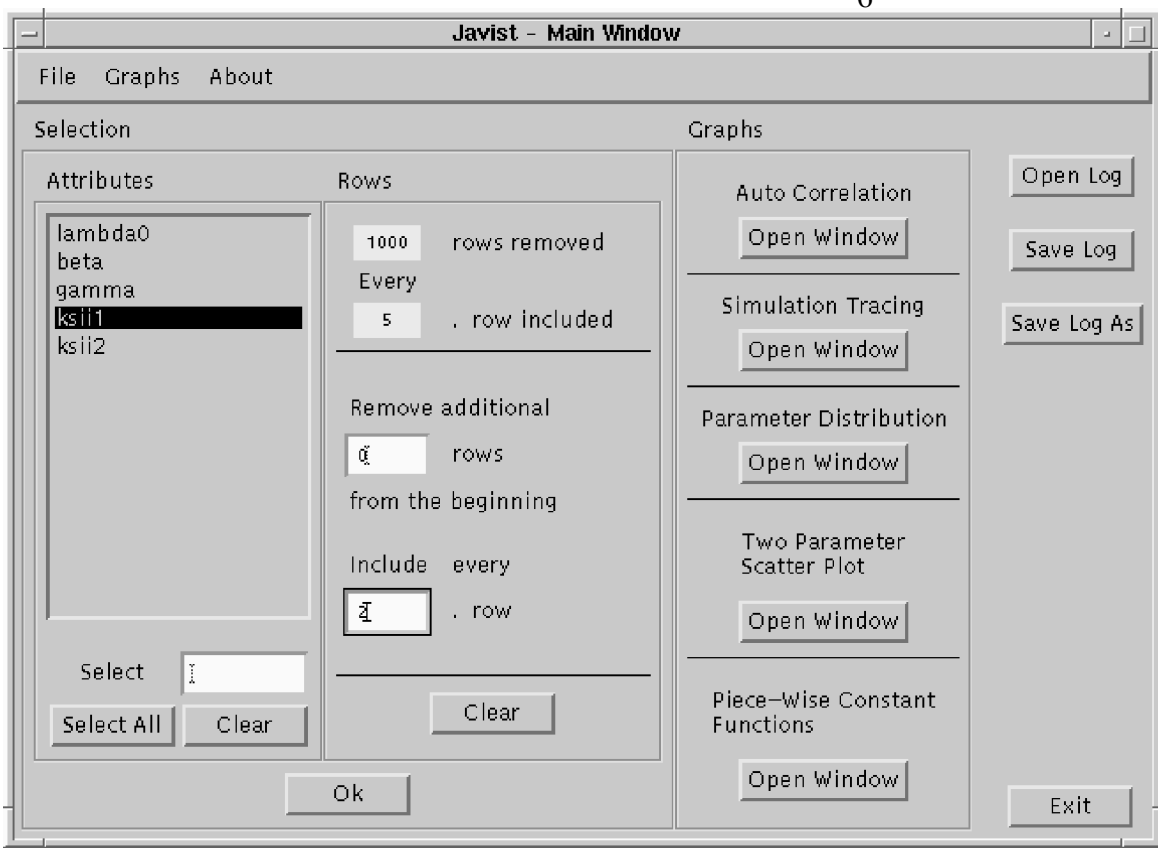
*Kahden parametrin arvojen tarkastelu:* Kahden parametrin arvoja voidaan tarkastella samanaikaisesti hajontakuviolla: Kutakin simulaatiokierrosta vastaa kaaviossa piste, jonka vaakakoordinaatti on toisen ja pystykoordinaatti toisen parametrin arvo.

*Paloittain vakiot funktiot:* Paloittain vakiot funktiot ovat tässä yhteydessä tyypillisesti, vaikkakaan eivät aina, ajan funktioita. Vaaka–akselilla on aika ja pystyakselilla funktion arvo, joka puolestaan tyypillisesti on ns. intensiteettiarvo, joka ilmaisee tapahtumien tiheyttä tai riskiä. Paloittain vaikoista funktioista keskeisintä on approksimoidun odotusarvon sekä kvantiilien esittäminen. Koska funktion muutoskohdat ovat tyypillisesti simulaation aikana vaihtuvia, keskiarvon laskeminen tapahtuu siten, että kiinnitetään ennalta niiden pisteiden lukumäärä, joissa halutaan keskiarvo laskettavaksi. Myös keskiarvo ilman kvantiileja voidaan esittää. Koska funktiossa voi olla simulaation aikana vaihteleva määrä paloja, koko simulaation tuloksen odotusarvon lisäksi ohjelma tuottaa tarvittaessa kuvan, jossa on ainoastaan sellaiset iteraatiokierrokset, joilla on haluttu määrä paloja. Kun palamäärä on kiinnitetty, ohjelma pystyy käsittelemään yksittäisiä hyppykohtia ja tasoja kuten mitä tahansa yksittäisiä parametrejä.

## *3.4 Ohjeet toimintojen suorittamiseksi*

### 3.4.1 Pääikkuna (Main Window)

Javistin käynnistyessä avautuu näytölle pääikkuna (kuva 3.1). Ikkunassa käyttäjä voi karsia dataa, aloittaa tallentamisen, avata kaavioikkunoita ja avata tiedoston datan lukemista varten.



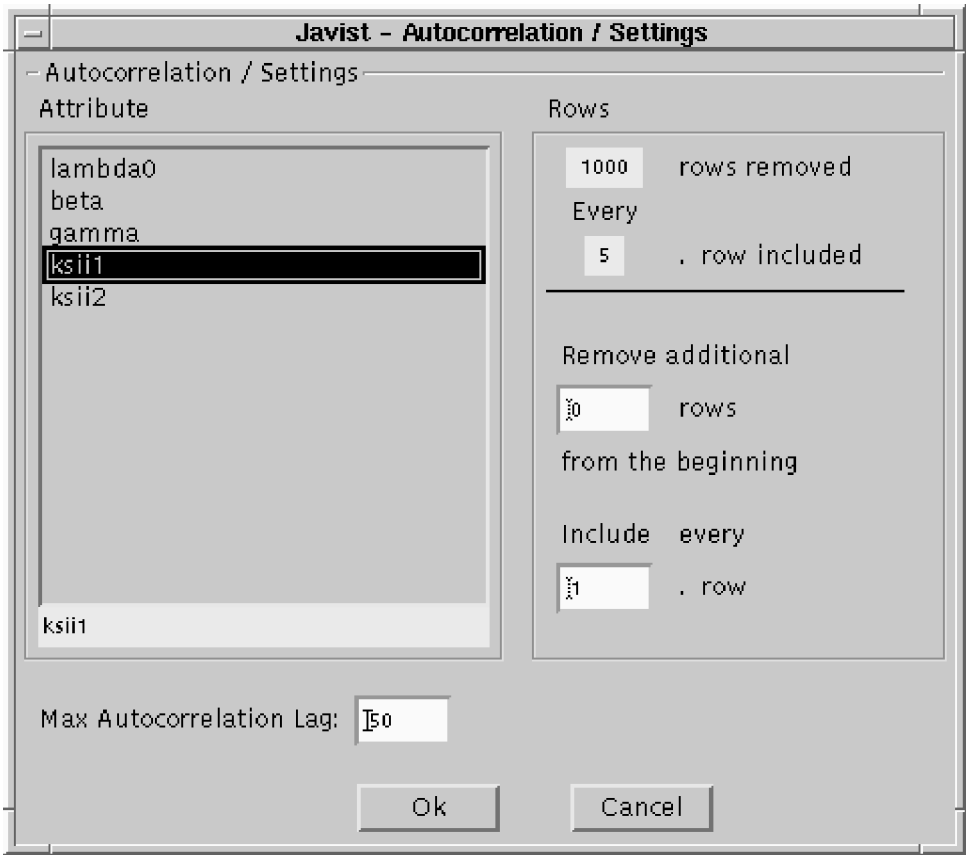

Kuva 3.1 Pääikkuna

Pääikkunassa tapahtuvat karsinnat ovat lopullisia ja edellisen karsinnan tulokset ovat pohjana seuraavalle karsinnalle. Esimerkki: Käyttäjä poistaa aluksi 1000 riviä syötteen alusta ja ottaa mukaan vain joka viidennen rivin. Tämän jälkeen käyttäjä poistaa uudestaan 1000 riviä syötteen alusta ja ottaa mukaan joka toisen rivin. Tällöin datasta on poistettu 1000+1000*5=6000 riviä alusta ja jäljellä on enää joka kymmenes rivi. Karsinnat toteutuvat vasta OK-nappulan painamisen jälkeen. Kun mahdolliset karsinnat on tehty, käyttäjä voi avata kaavioikkunoita (OPEN WINDOW). Kaavioikkunoita voidaan avata useita kappaleita kutakin kaaviotyyppiä kohti. Kaavioikkunoita voi myös avata GRAPHS-valikosta.

Halutessaan käyttäjä voi aloittaa datan tallentamisen tiedostoon painamalla SAVE LOG AS ja antamalla tiedostolle nimen. Tällöin tiedostoon ohjataan data tallentamishetken karsintaehtojen mukaisesti. Mahdolliset uudet karsinnat eivät vaikuta tiedostoon menevään dataan, vaan tiedostoon viedään data alkuperäisillä karsintaehdoilla. Pääikkunasta voidaan syöteaineistoksi valita tiedosto (OPEN LOG), jolloin mahdollinen datan lukeminen putkesta keskeytyy. Jos käyttäjä tiedostoa karsittuaan haluaa tallentaa sen vanhan version päälle, tapahtuu se nappulasta SAVE LOG. Tiedoston käsittely on mahdollista myös FILE-valikosta.

Javist suljetaan painamalla nappia EXIT tai valitsemalla kohta FILE-valikosta.

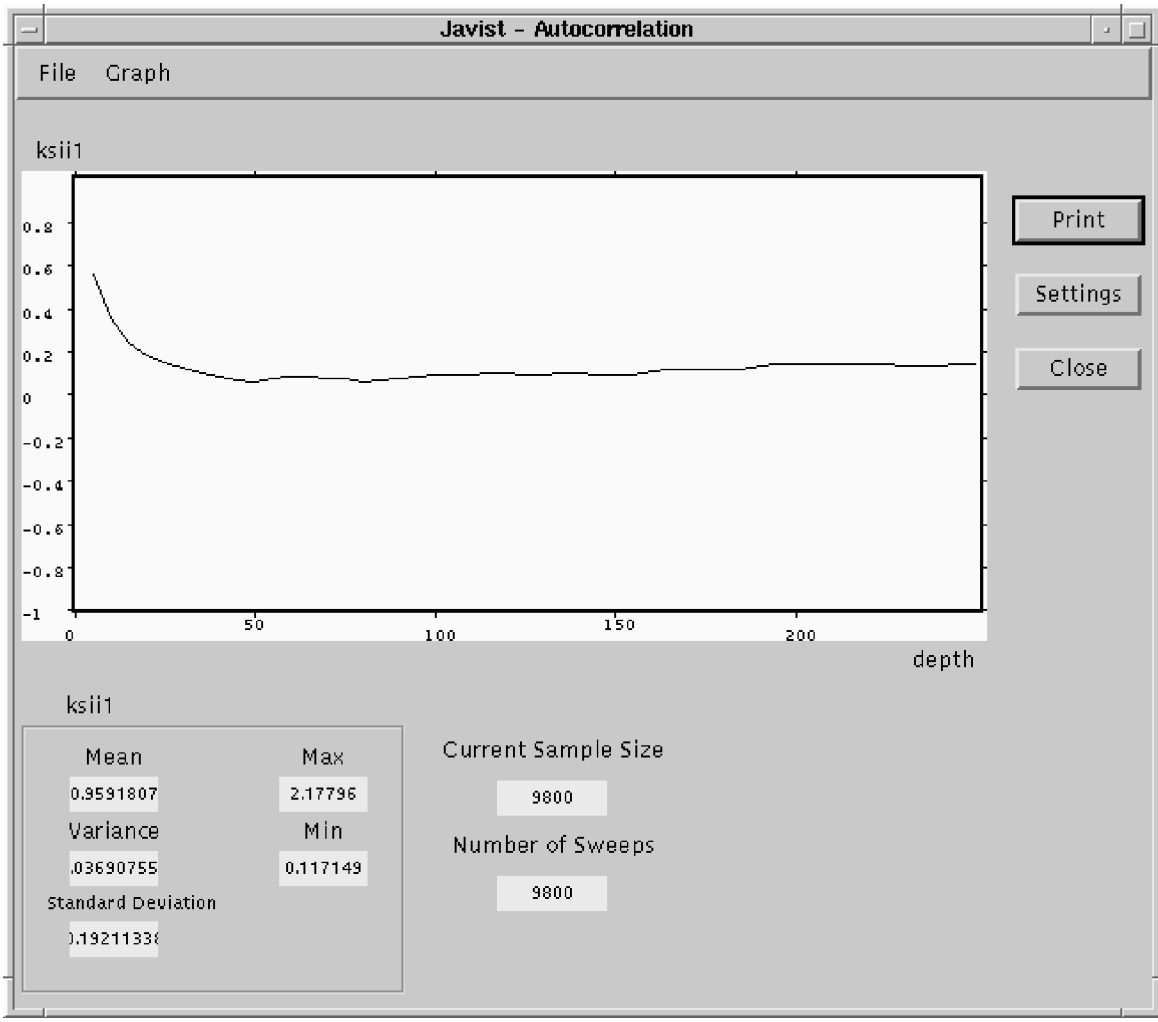
### 3.4.2 Autokorrelaation seuranta (Autocorrelation)




Kuva 3.2 Autokorrelaation seuranta, asetukset

Autokorrelaation seuranta – samoin kuin kaikki muutkin kaaviotyypit – avaa ruudulle ensin varsinaisen kaavioikkunan ja sen päälle asetusikkunan (kuva 3.2). Asetusikkunasta valitaan seurattava attribuutti (1 kpl) sekä voidaan karsia dataa lisää. Karsinnat eivät ole pysyviä siinä mielessä, että avattaessa uusi autokorrelaation seuranta –ikkuna on käytettävissä sama aineisto kuin ensimmäisessä vastaavassa ikkunassa ennen mahdollista karsintaa. On mahdollista vaihtaa autokorrelaation syvyys perusasetuksesta 50. Kaikki muutokset ja

karsinnat astuvat voimaan vasta painettaessa OK, jolloin asetusikkuna sulkeutuu ja kaavioikkunaan piirtyy kuvaaja kuvan 3.3 tyyliin. Kaikissa kaavioikkunoissa näkyy myös kuvaajan alapuolella kullekin kaaviotyypille ominaiset tilastolliset tunnusluvut



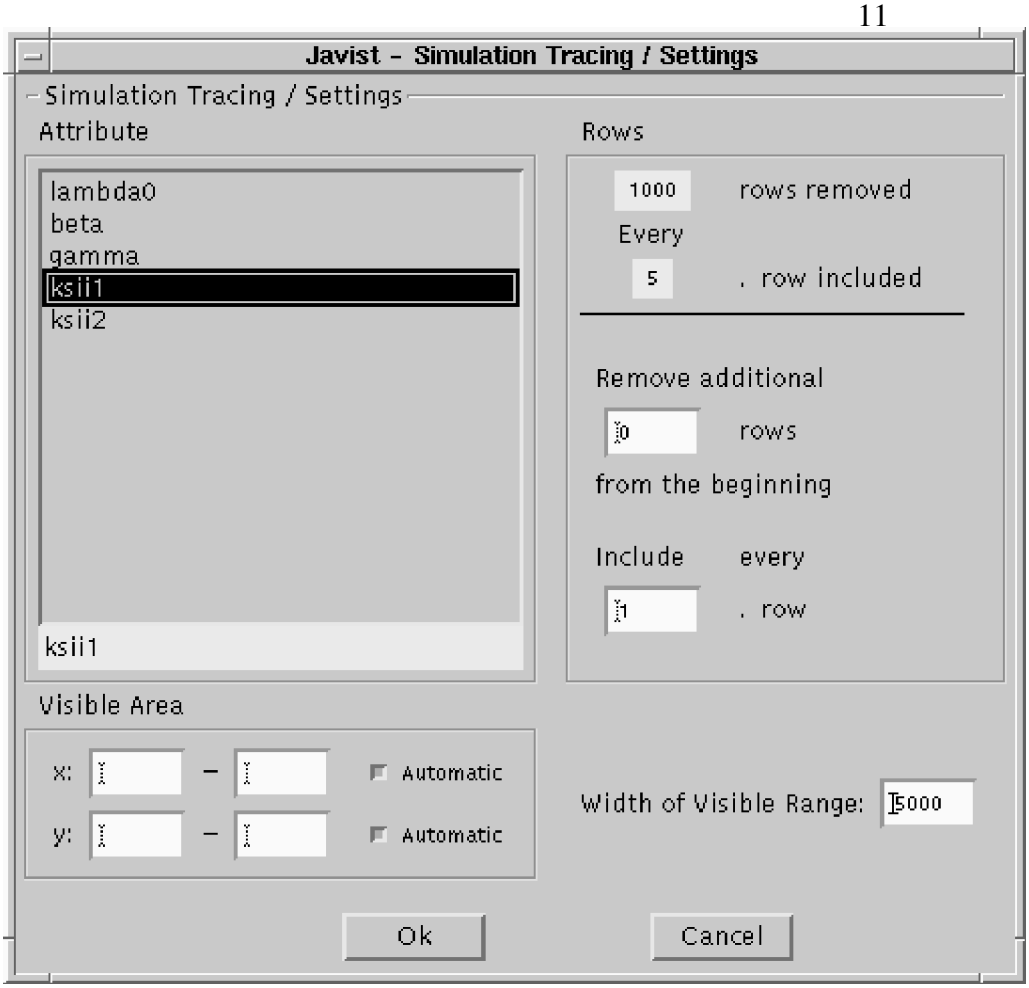

Kuva 3.3 Autokorrelaation seuranta, kaavioikkuna

Tässä tapauksessa tunnuslukuja ovat keskiarvo (Mean), varianssi (Varianssi), keskihajonta (Standard Deviation), maksimi (Max), minimi (Min), otoksen koko (Current Sample Size) ja aineiston koko (Number of

Sweeps).

Kaikissa kaavioikkunoissa on yhtenäiset toiminnot: Tulostaminen tapahtuu nappulalla PRINT tai FILE–valikosta kohdasta PRINT. Tällöin käynnistyy kaksivaiheinen tulostus, jossa ensin tallennetaan kuva GNUPlot–formaatissa ja toisena tilastolliset tunnusluvut. Suora kirjoittimelle tulostus ei siis ole mahdollista. Asetuksia pääsee muuttamaan SETTINGS–nappulasta tai GRAPHS–valikon vastaavasta kohdasta. Kaavio suljetaan nappulalla CLOSE tai FILE–valikon kohdasta CLOSE.

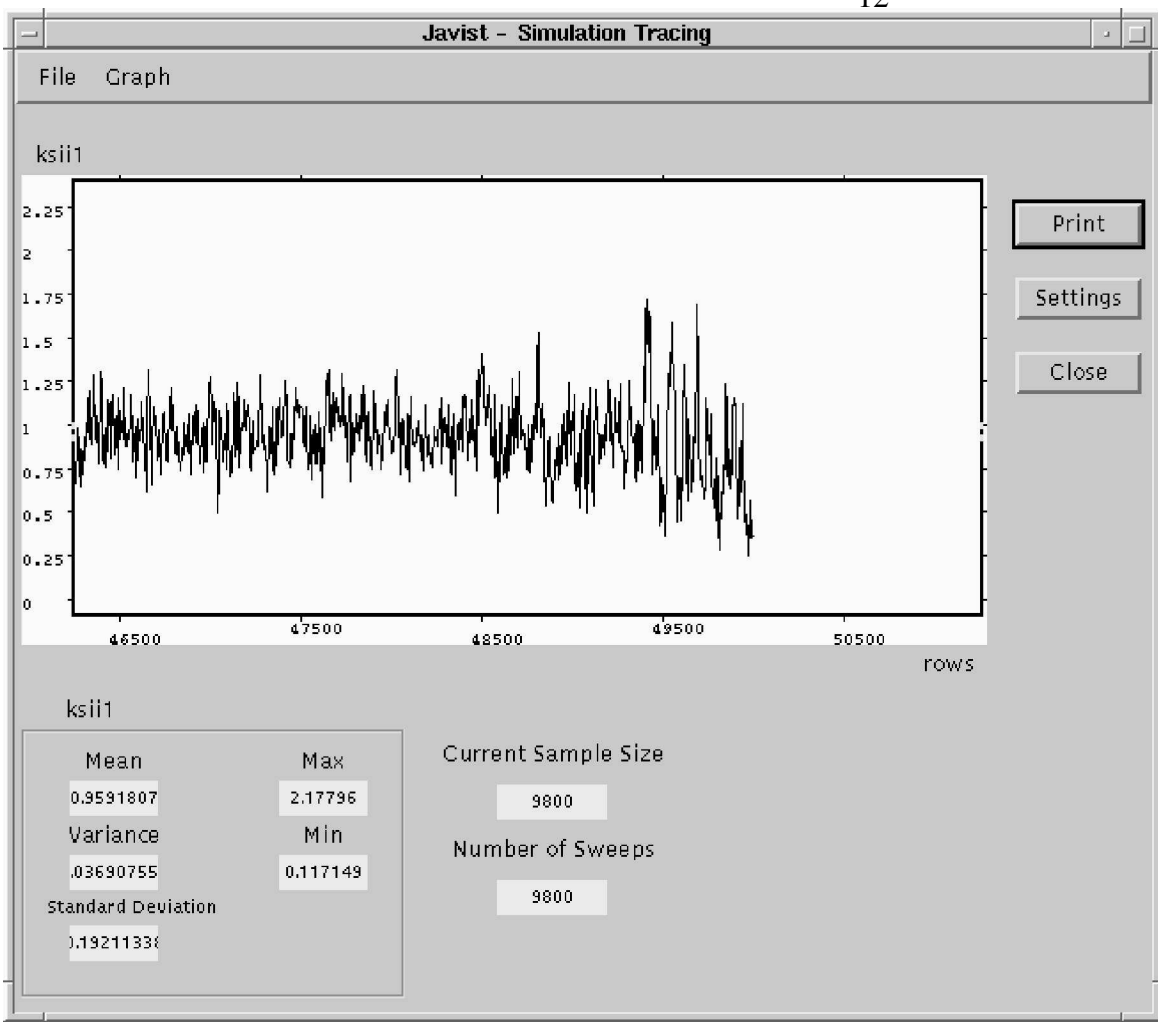
### 3.4.3 Simulaation jäljittäminen (Simulation Tracing)




Kuva 3.4 Simulaation jäljittäminen, asetukset

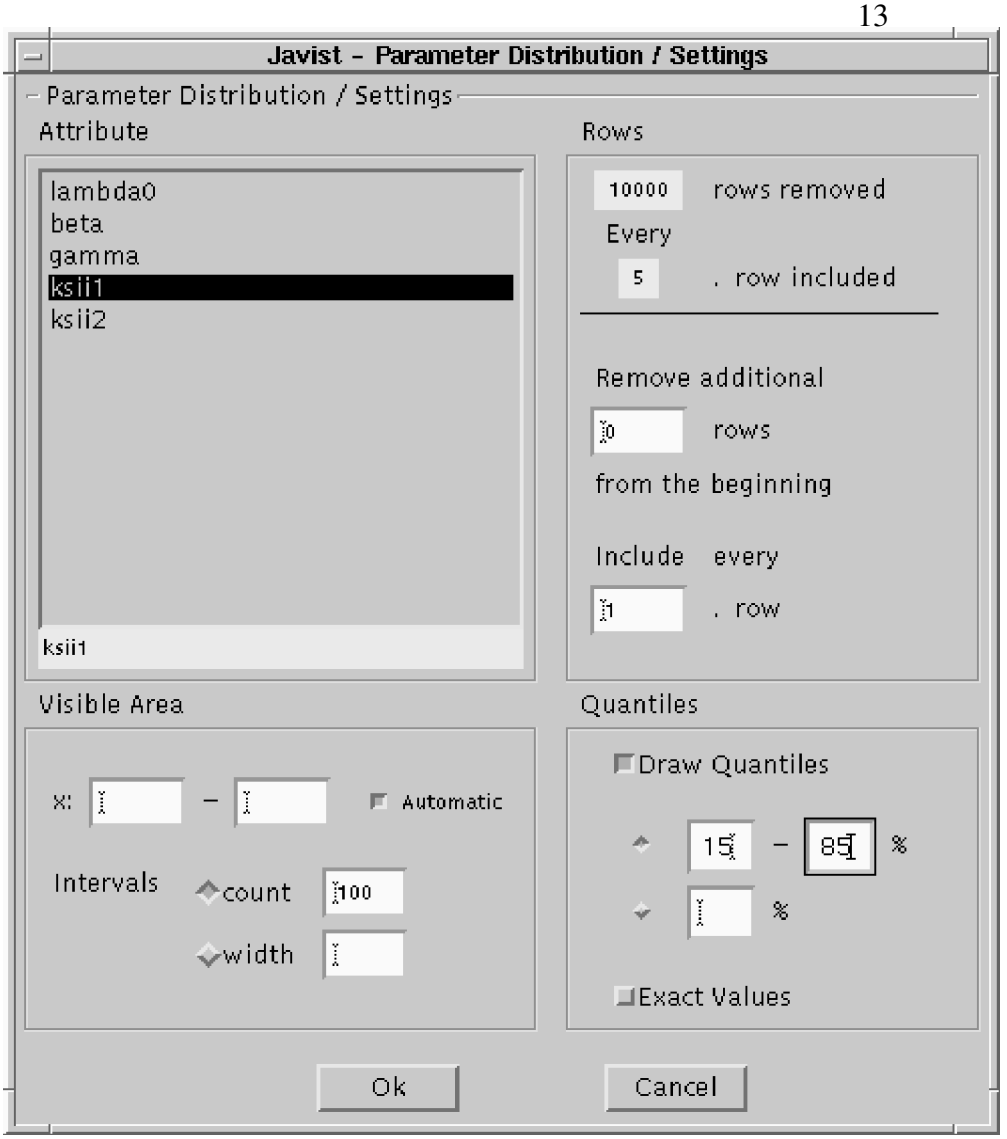
Simulaation jäljittämisen asetukset (kuva 3.4) ovat attribuutin valinnan ja rivien karsinnan osalta samanlaiset kuin autokorrelaation seurannassa (kuva 4.7). Lisäksi voidaan määrittää kaavion näkyvä alue joko automaattisesti tai kirjoittaa kullekin koordinaatille halutut arvot. Valittavana on myös näkyvän alueen leveys kuvaajassa.

Kaavioikkunan (kuva 3.5) toiminnot ja tunnusluvut vastaavat kuin kohdassa 3.4.2.




Kuva 3.5 Simulaation jäljittäminen, kaavioikkuna

### 3.4.4 Parametrin jakauma (Parameter Distribution)

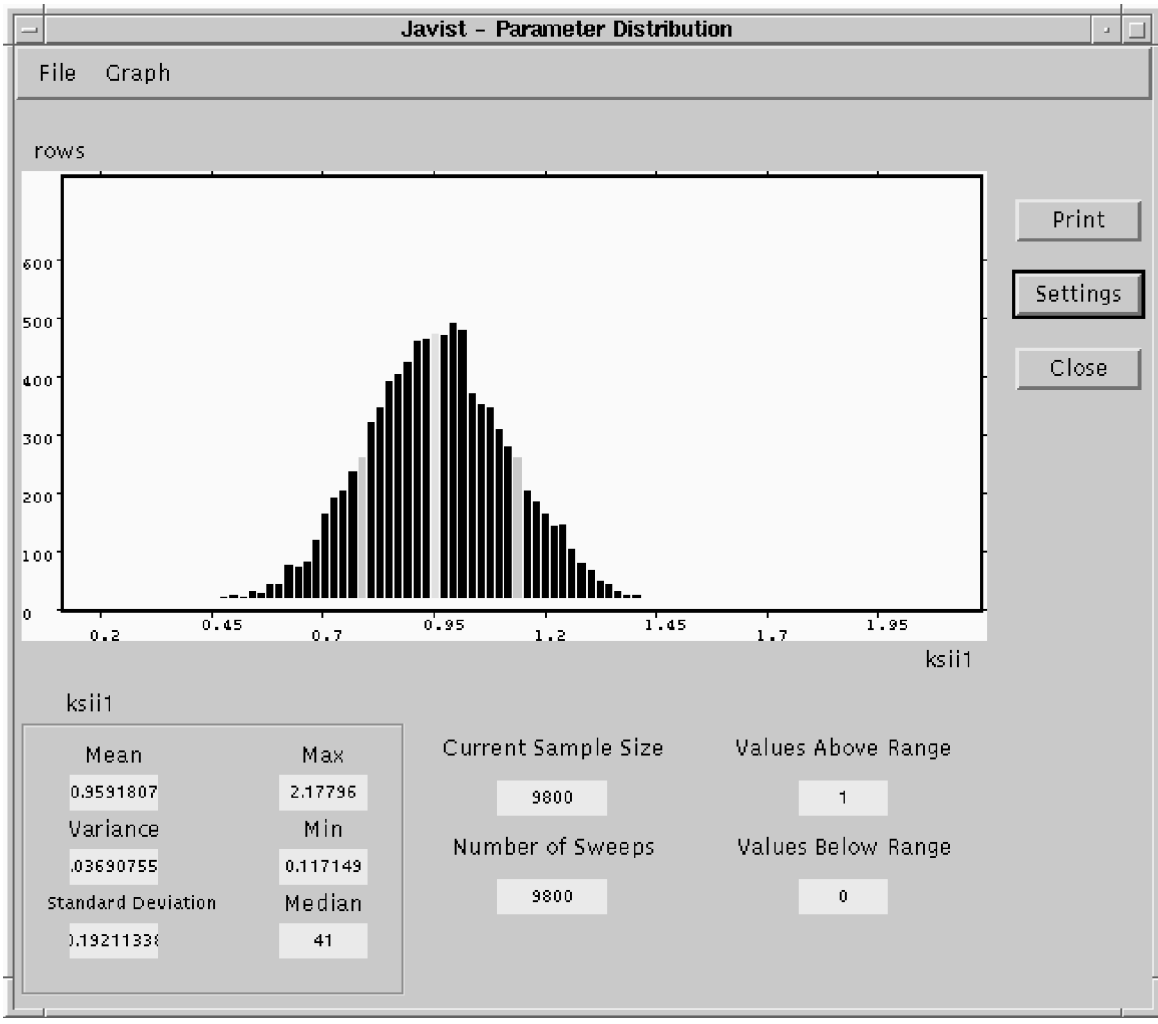



Kuva 3.6 Parametrin jakauma, asetukset

Attribuutin valinta ja rivien karsinta kuten kohdassa 3.4.2. Lisäksi voidaan asettaa kaavion näkyvä alue ja pylväiden lukumäärä/leveys. Haluttaessa kaavioon saadaan myös näkymään kvantiilit valitsemalla DRAW QUANTILES. Ne voidaan määrittää kahdella eri tavalla: esim. 15 – 85 % vastaa samaa asiaa kuin 70 %. Kvantiilien laskeminen voidaan tehdä tarkoilla arvoilla valitsemalla kohta EXACT VALUES kohdasta

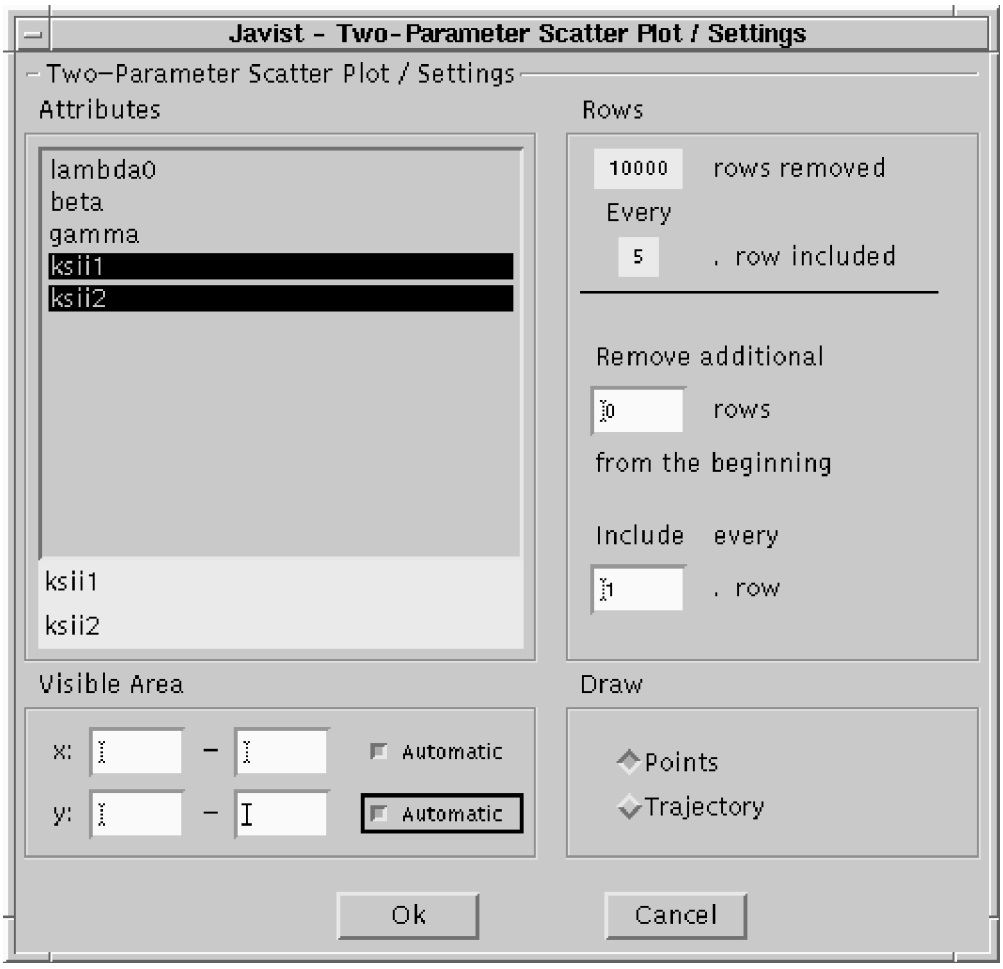
QUANTILES.

Kaavioikkunan (kuva 3.6) toiminnot ovat vastaavat kuin kohdassa 3.4.2. Tunnuslukuina kohdasta 3.4.2 poiketen arvot määrittelyvälin ylä– ja alapuolella (Values Above/Below Range) ja mediaani (Median). Kaavioon on keltaisella merkitty mediaani ja harmaalla asetusikkunassa valitut kvantiilit.




Kuva 3.7 Parametrin jakauma, kaavioikkuna

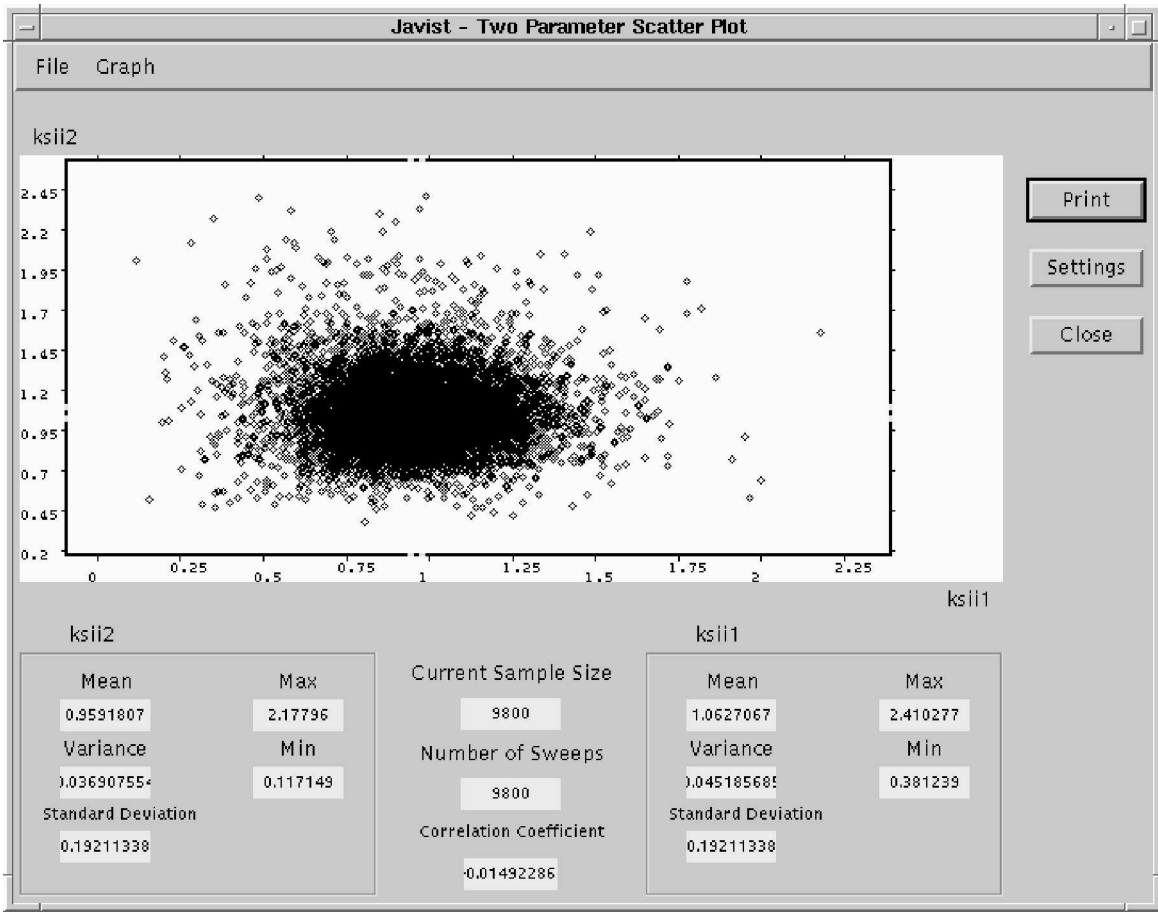
### 3.4.5 Kahden parametrin arvojen tarkastelu (Two–Parameter Scatter Plot)




Kuva 3.8 Kahden parametrin arvojen tarkastelu, asetukset

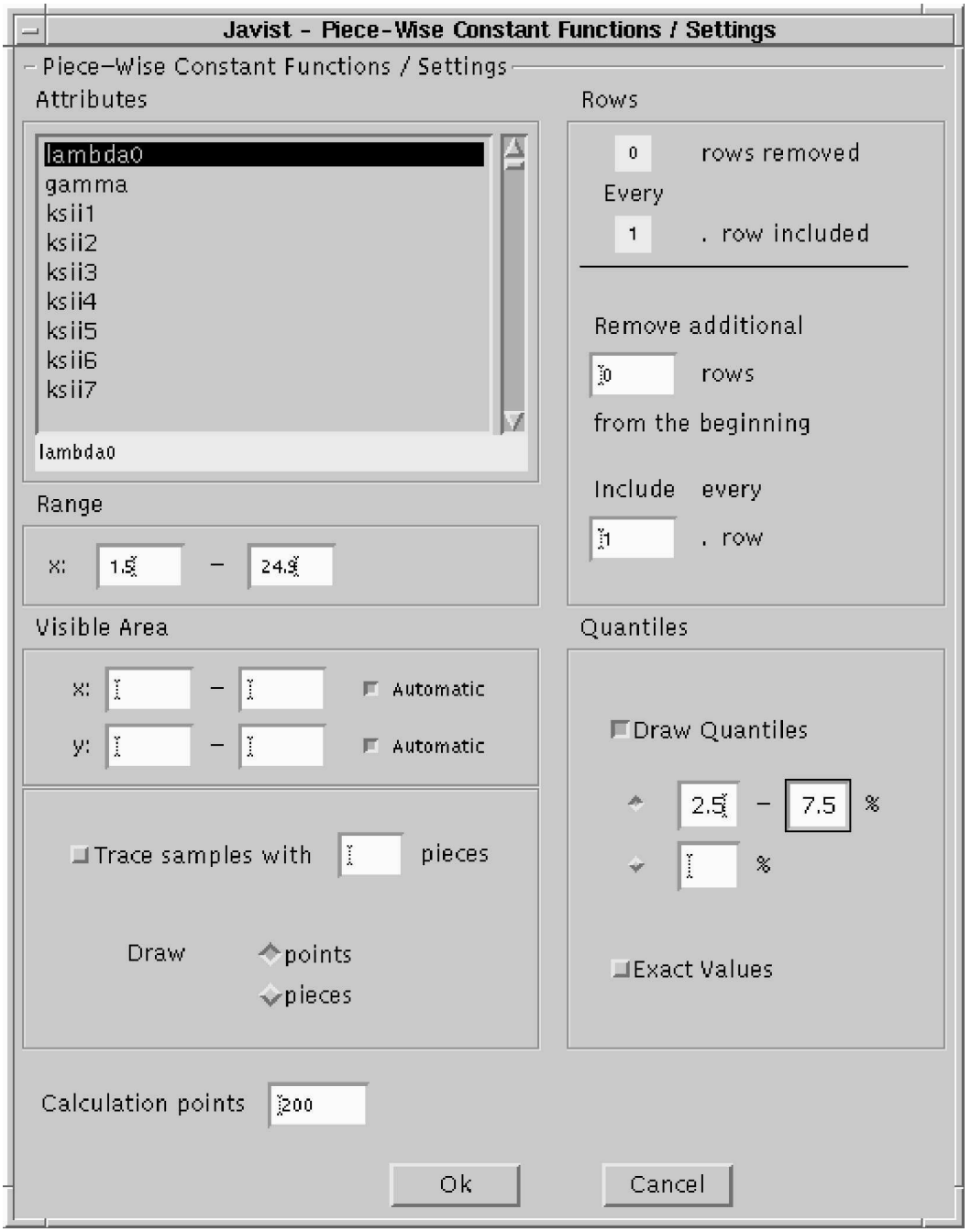
Rivien karsinta kuten kohdassa 3.4.2. Attribuutteja on nyt valittava kaksi kappaletta. Valitut attribuutit näkyvät listan alapuolella ja vaihtuvat siten, että kolmas valittu attribuutti menee ensimmäiseksi valitun tilalle ja neljäs vastaavasti toiseksi valitun tilalle jne. Myös voidaan määrittää kaavion näkyvä alue ja sen piirtotapa joko viivana (Trajectory) tai pisteinä (Points).

Kaavioikkunan (kuva 3.9) toiminnot vastaavat kuin kohdassa 3.4.2. Tunnuslukuna kohdasta 3.4.2 poiketen arvo kahden muuttujan korrelaatiokertoimelle (Correlation Coefficient).




Kuva 3.9 Kahden parametrin arvojen tarkastelu, kaavioikkuna

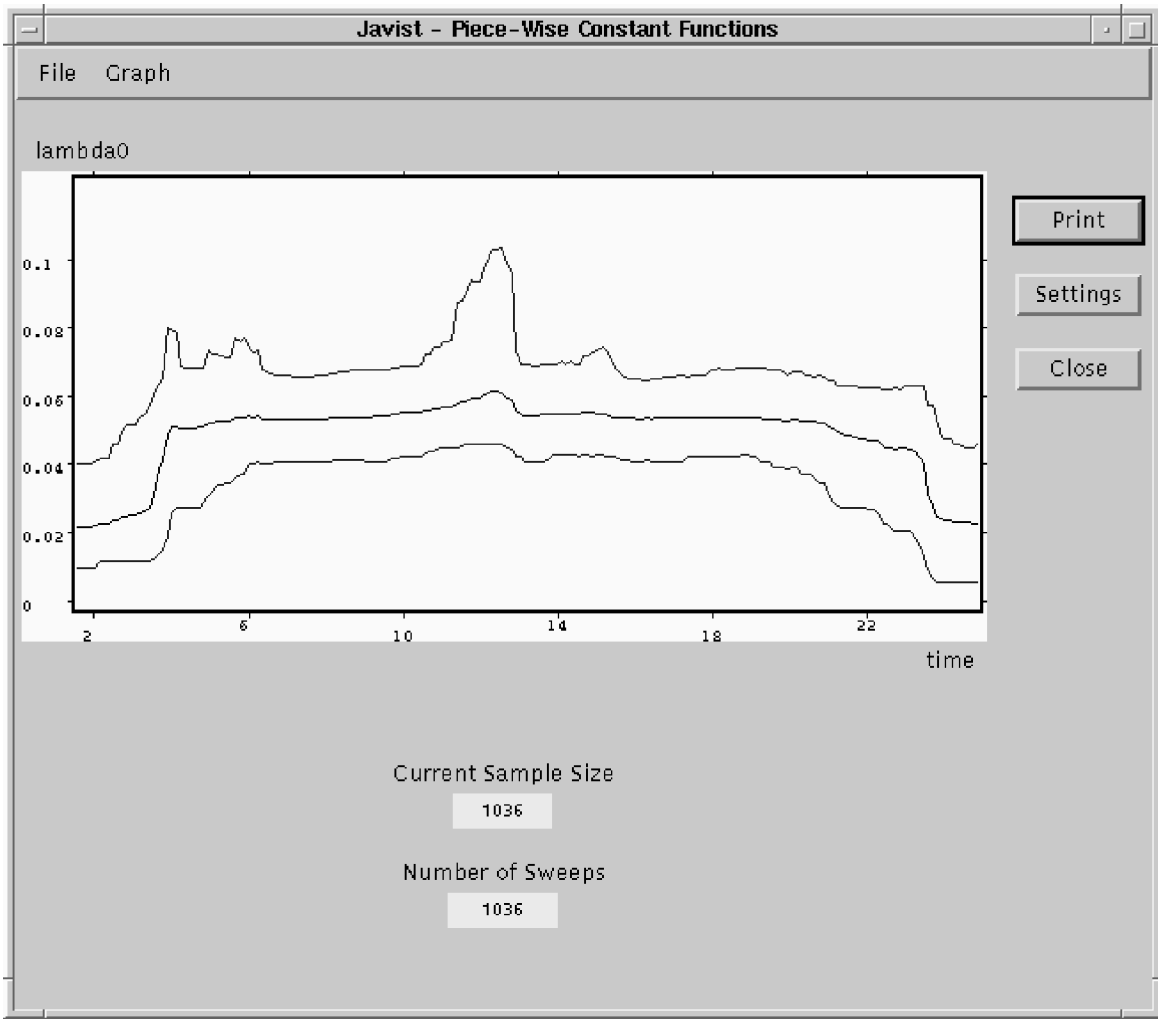
### 3.4.6 Paloittain vakiot funktiot (Piece–Wise Constant Functions)




Kuva 3.10 Paloittain vakiot funktiot, asetukset

Attribuutin valinta ja rivien karsinta kuten kohdassa 3.4.2. Lisäksi voidaan asettaa kaavion näkyvä alue, kvantiilit, määrittelyväli (Range) ja laskentapisteiden määrä (Calculation points). Tietyn lukumäärän omaavien funktioiden jäljittäminen on mahdollista valitsemalla kyseinen kohta ja antamalla seurattava palamäärä.

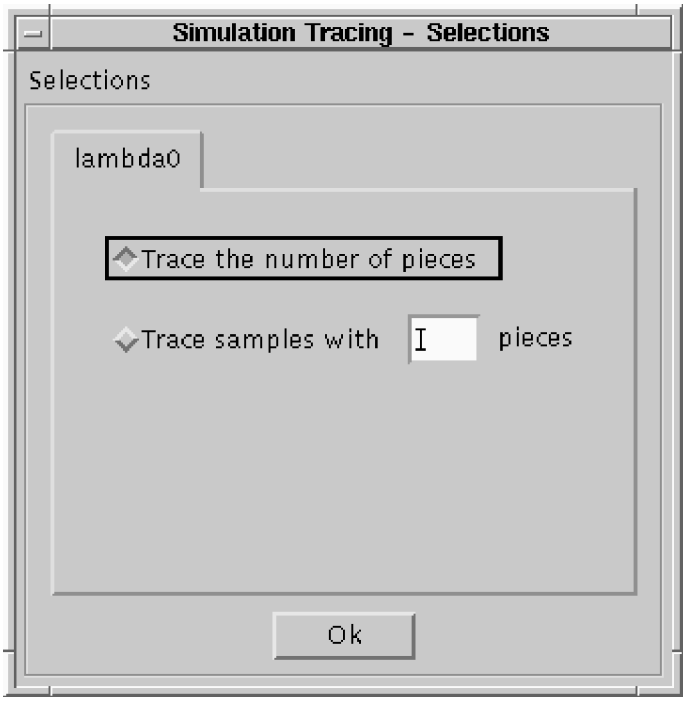
Kaavioikkunan (kuva 3.6) toiminnot vastaavat kuin kohdassa 3.4.2. Tunnuslukuina otoksen koko ja aineiston koko.



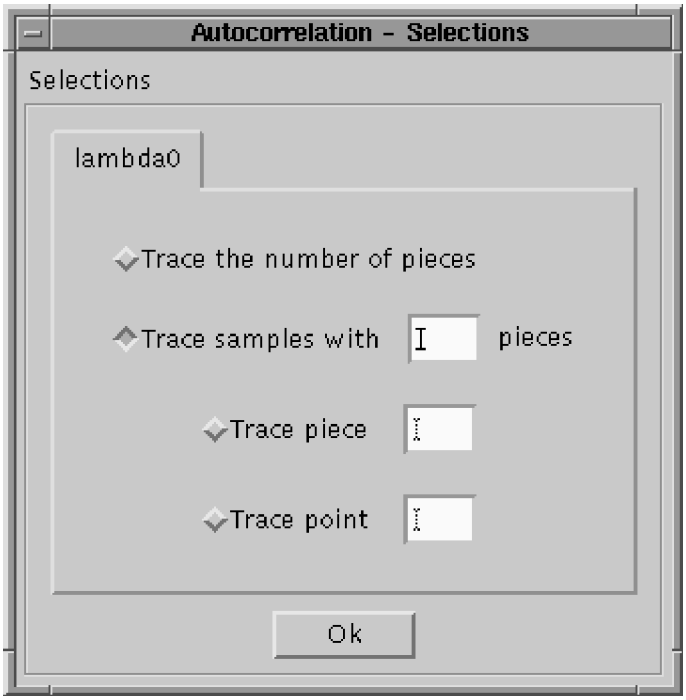

Kuva 3.11 Paloittain vakiot funktiot, kaavioikkuna

### 3.4.7 Funktioikkunat (Selections)

Käyttäjän valittua kaaviotyypistä riippumatta asetusikkunassa attribuutiksi funktion ja painettuaan OK–nappulaa ilmestyy ruudulle funktioikkuna (kuva 3.12). Paloittain vakioiden funktioiden tapauksessa käy aina näin, sillä asetusikkunassa oletetaan käyttäjän aina valitsevan funktion attribuuttien joukosta. Kahden parametrin tapauksessa kummallekin parametrille on ikkunassa oma `lehti´, josta kullekin parametrille ominaiset valinnat tehdään. Osa valinnoista on keskenään poissulkevia.




Kuva 3.12 (a) Funktioikkuna




Kuva 3.12 (b) Funktioikkuna

Funktioikkunassa mahdollisia vaihtoehtoja on jäljittää palojen lukumäärää tai vaihtoehtoisesti tietyn lukumäärän omaavia näytteitä. Tällöin käyttäjän valittavissa on jäljittääkö tiettyä kohtaa vai palaa.

### *3.5 Toimintojen peruutusmahdollisuudet*

Pääikkuna: Attribuuttien valinta voidaan peruuttaa painamalla CLEAR–nappulaa, samoin rivien karsinnassa (kummallekin siis oma CLEAR–nappula). Olennaista on, että tämä tapahtuu ennen OK–nappulan painamista.
Karsittu data on OK–nappulan painamisen jälkeen mahdollista palauttaa ainoastaan tallennetusta tiedostosta tai ajamalla Bassist uudestaan

Tiedoston lataus−/tallenusikkuna: Peruutus tapahtuu painamalla CANCEL.

Kaavioikkuna: Kaaviota varten asetusikkunassa karsittu data on mahdollista palauttaa pääikkunassa tehtyjen karsintojen mukaiseksi sulkemalla kyseinen kaavioikkuna ja avaamalla se pääikkunasta uudestaan.

Asetus − ja funktioikkunat: Asetusten ja valintojen peruuttaminen ennen OK−nappulan painamista painamalla CANCEL. Jos OK−nappulaa on jo painettu ja kyseinen ikkuna avattu uudestaan on arvot ja valinnat käsin palautettava haluttuihin arvoihin painettava OK−nappulaa.

### *3.6 Suorituksen tilan tallentaminen ja uudelleenkäynnistys*

Javistin käynnissä olevaa tilaa ei sellaisenaan voi tallentaa, mutta sen hetkisen pääikkunatasolla karsitun datan voi tallentaa tiedostoon ja ladata myöhemmin esim. uudelleenkäynnistyksen jälkeen tarkasteltavaksi.

## 4. Virheilmoitukset

Data error!
Header error!
Input error!
Choose a function!
Choose an attribute!
Choose two attributes!
Exit? Are you sure?

No attributes chosen!
No data!

*Data error!:* Javistin saama syöte joko Bassistilta tai tiedostosta on virheellisessä muodossa. Ilmoitus odottaa käyttäjän kuittausta (painamalla OK–nappulaa virheilmoitusikkunassa), jonka saatuaan Javist sulkeutuu.

*Header error!:* Javistin saaman syötteen otsikkotiedot ovat virheelliset. Ilmoitus odottaa kuittausta, jonka saatuaan ohjelma sulkeutuu.

*Input error!:* Käyttäjän odotetaan syöttävän tekstikenttiin numeerisia arvoja ja jos näin ei tapahdu ilmoittaa ohjelma tästä virheilmoituksella. Käyttäjän oletaan virheilmoituksen jälkeen kuittaavan ilmoitus painamalla OK:ta ja tarkistavan tekstikenttien arvot ja tekevän tarvittavat muutokset.

*Choose a function!:* Paloittain vakioiden funktioiden asetusikkunassa käyttäjä ei ole valinnut listalta funktiota. Ilmoitus odottaa käyttäjän kuittausta, jonka jälkeen voidaan valita valintaikkunan attribuuttilistalta funktio.

*Choose an attribute!:* Käyttäjä ei ole valinnut attribuuttilistasta tarvittavaa attribuuttia, jotta kaavio voitaisiin piirtää. Käyttäjän tulee kuitata ilmoitus ja valita attribuutti.

*Choose two attributes!:* Kahden parametrin arvojen tarkastelussa käyttäjän on valittava attribuuttilistasta kaksi attribuuttia ja jos näin ei tapahdu (käyttäjä valitsee vain yhden tai ei yhtään) on seurauksena kyseinen kehoitus. Käyttäjän tulee kuitata ilmoitus ja valita kaksi attribuuttia.

*Exit? Are you sure?:* Käyttäjä on valinnut ohjelmasta poistumisen, josta seurauksena varmistusilmoitus. Käyttäjä voi perua poistumisen painamalla CANCEL tai varmistaa poistumisen painamalla OK.

*No attributes chosen!:* Käyttäjä ei ole valinnut pääikkunnassa attribuuttien valintalistasta yhtään attribuuttia, mutta on silti painanut OK ikäänkuin valitakseen osan tarjolla olevista attribuuteista. Käyttäjä kuittaa ilmoituksen, jonka jälkeen hän voi valita attibuutteja tai siirtyä suoraan kaavioikkunoihin (jolloin kaikki attribuutit ovat mukana).

*No data!:* Javist ei ole saanut syötettä ja käyttäjä on ollut karsimassa tai valitsemassa jotakin kaaviotyyppiä. Käyttäjä kuittaa ja voi tämän jälkeen esim. avata valmiiksi tallennetun tiedoston ohjelman syötteeksi.

## 5. Toiminnalliset erot erilaisissa käyttöympäristöissä

Javist on kehitetty java–kieltä tukevissa ympäristöissä toimivaksi, joten sen toiminnassa ja toiminnoissa ei ole käyttöympäristöstä riippuvia eroja. Käyttöympäristöstä riippuen saattaa käyttöliittymän ulkoasussa esiintyä minimaalisia kosmeettisia muutoksia kuten hieman erilainen fontti käyttöliittymässä Windows– ja LINUX–ympäristöissä, mutta tällä ei ole suurempaa vaikutusta Javistin toimintaan tai käyttöön.

## 6. Asennusohje

Javist asennetaan LINUX–ympäristössä haluttuun hakemistoon purkamalla javist.zip –paketti komennolla "unzip javist.zip" tai käyttämällä jotain zip–yhteensopivaa purkuohjelmaa. Muissa

käyttöympäristöissä javist.zip voidaan purkaa zip-yhteensopivilla purkuohjelmilla kunhan niiden nimeämisjärjestelmä ei ole turhan rajoitettu (esim MS-DOS 8+3).
Mikäli asennus ei onnistu, tulee varmistaa että javist.zip tiedosto on ehjä, hakemistossa on riittävästi tilaa ja käytössä on tarpeeksi uusi versio zip-paketin purkuohjelmasta.

*Lähteet:*

Ehrlund et al. [Javist – jakaumien visualisointityökalu, Suunnitteludokumentti, 4.4.1998, Helsingin Yliopiston tietojenkäsittelytieteen laitos]